

A Thesis Submitted for the Degree of PhD at the University of Warwick

Permanent WRAP URL:

<http://wrap.warwick.ac.uk/171008>

Copyright and reuse:

This thesis is made available online and is protected by original copyright.

Please scroll down to view the document itself.

Please refer to the repository record for this item for information to help you to cite it.

Our policy information is available from the repository home page.

For more information, please contact the WRAP Team at: wrap@warwick.ac.uk



Characterisation of *cis*-regulatory modules controlling gene expression within *sox17+* lineages during zebrafish embryonic development.

By

Randa Elsayed

Thesis

Submitted to the University of Warwick

For the degree of

Doctor of Philosophy

Warwick Medical School

31st May 2021

Table of Contents

List of figures.....	vii
List of tables	xi
Declaration.....	xiii
Acknowledgements.....	xiv
Abstract.....	xv
List of abbreviations	xvi
Chapter 1 Introduction.....	1
1.1 Endoderm, the innermost germ layer	1
1.2 Zebrafish as an animal model.....	2
1.3 Segregation of the three germ layers	3
1.4 Fate map of zebrafish embryo	5
1.5 YSL in zebrafish	6
1.6 Pathway to endoderm formation	7
1.6.1 Nodal signalling pathway	7
1.6.2 Maternally-contributed factors.....	10
1.6.3 From Nodal induction to endoderm specification	10
1.6.4 HMG-containing transcription factor, Sox32 and endoderm formation.....	14
1.7 Sox17 is a marker of DFCs and haematopoietic lineages	18
1.8 Beyond Sox32 and Sox17: the role of Forkhead factors	20
1.9 From endoderm formation to organogenesis	22
Identifying developmentally important <i>cis</i> -regulatory modules (CRMs).....	24
1.11 Classification of CRMs	25
1.11.1 Core promoter.....	25
1.11.2 Enhancers.....	26
1.11.3 Silencers	26
1.11.4 Insulators.....	27
1.12 Histone variants and chromatin accessibility	28

1.13 Zygotic genome activation: earliest event in chromatin reorganisation	30
1.14 Challenges in identifying enhancers	32
1.15 Enhancers identified in endodermal organs	36
1.16 Enhancer mutations underlie numerous developmental diseases.....	38
1.17 Objectives of this study	39
Chapter 2 Materials and Methods	41
2.1 Fish strains and transgenics	41
2.2 Generation of the triple transgenic line	41
Embryo injections	41
2.3	41
2.3.1 Labelling of DFCs using Rhodamine-dextran (RD) injection	41
2.3.2 <i>Sox32</i> mRNA injection	41
2.3.3 <i>Sox32</i> and standard morpholino injections.....	42
2.4 Imaging of zebrafish embryos	42
2.5 Embryo dissociation.....	42
2.6 Fluorescence-activated cell sorting (FACS).....	43
2.7 Real-time quantitative PCR (qRT-PCR).....	43
2.7.1 qRT-PCR data analysis	45
2.8 Assay for transposase-accessible chromatin using sequencing (ATAC- seq).....	45
2.8.1 ATAC-seq data analysis	46
2.9 RNA-seq.....	49
2.9.1 RNA-seq (6 h.p.f.): high input library.....	49
2.9.2 RNA-seq (28 h.p.f.): low input library.....	49
2.9.3 RNA-seq data analysis	50
Chapter 3 Overexpression of <i>sox32</i> induces ectopic formation of both endoderm and DFC fates.....	51
3.1 <i>Sox32</i> is required for endoderm formation	51

3.2 Studying the chromatin landscape	53
3.3 Objective of this study	55
3.4 Sox32 overexpression upregulates expression of sox17 in the embryo .57	
3.5 Sox32 overexpression induces opening of chromatin in regions proximal to endoderm-expressed genes	59
3.5.1 Sox32 OE has different accessibility profile to control.....	62
3.5.2 Overexpression of sox32 does not drive artificial accessibility at loci that are not accessible in control embryos.....	66
3.5.3 Sox32 OE results in enhanced chromatin accessibility proximal to genes associated with endoderm and DFC formation.	67
3.5.4 Sox32 OE DARs are enriched for binding motifs of transcription factors that play a role in the endoderm and DFCs	69
3.5.5 Sox32 OE DARs show concordance with Tbx16 and Tbx17 bound regions.....	71
3.6 Overexpression of sox32 results in upregulation of endoderm and DFC target genes.....	76
3.7 Sox32 OE upregulated transcripts significantly correlate with endoderm and DFC-enriched genes.....	79
3.8 Changes in the chromatin landscape significantly correlate with changes in gene expression	82
3.9 Overexpression of sox32 induces early expression of DFC markers prior to endoderm formation.....	84
3.10 Discussion.....	94
Chapter 4 Investigating the chromatin landscape at early organogenesis (28 h.p.f.).....	101
4.1 From gastrulation to organogenesis	101
4.2 Sox17 marks the endoderm and DFCs	102
4.2.1 KV, organ of laterality.....	103
4.2.2 Study of endoderm and DFCs by using the pan-endodermal line.....	104
4.3 Sox17 is also a marker of haemopoietic lineages	105

4.3.1 Transgenic lines used to study endothelial and blood lineages in zebrafish.....	107
4.4 Objective of this study	108
4.5 Characterising <i>Tg(sox17:EGFP)</i> expression pattern during early organogenesis	109
4.6 GFP+ cells in posterior notochord are derivatives of the KV	110
4.6.1 The GFP+ cells in the posterior notochord are derivatives of the DFCs/KV	112
4.7 Sox17+ endoderm cannot be isolated from contaminating mesoderm based on GFP expression.....	113
4.8 Successfully isolating GFP+ endoderm from contaminating mesoderm by in-crossing of transgenic lines	119
4.9 Validating the use of triple transgenic line in enriching for endoderm at 28 h.p.f.	121
4.10 ATAC-seq reveals endoderm-specific enhancers at 28 h.p.f.	125
4.10.1 The accessibility profile differs between all three cell populations	127
4.10.2 The accessibility profile of <i>sox17E</i> is similar to <i>sox17N</i> while <i>sox17M</i> is more distinct.....	132
4.10.3 <i>Sox17E</i> is enriched for terms associated with the endoderm.....	135
4.10.4 Known endoderm TF motifs are enriched in <i>sox17E</i> DARs	137
4.11 Endoderm-expressed transcripts are upregulated in <i>sox17E</i>.....	140
4.12 <i>Sox17E</i> comprises both KV transdifferentiated cells and endoderm progenitors	144
4.13 Changes in the chromatin accessibility correlates with changes in gene expression in <i>sox17+</i> endoderm	145
4.14 Discussion	147
Chapter 5 Investigating the chromatin landscape in the endoderm during development	154
5.1 Chromatin organisation accommodates gene expression changes	154
5.2 Objective of project	155
5.3 <i>Sox32</i> OE and <i>sox17E</i> have different accessibility profiles	157

5.3.1 <i>Sox32</i> OE and <i>sox17E</i> DARs are proximal to genes associated with the endoderm at different stages of development.....	160
5.3.2 <i>Sox32</i> OE and <i>sox17E</i> DARs are enriched with motifs for endoderm-expressed factors	162
5.4 Discussion.....	166
Chapter 6 Discussion.....	172
6.1 <i>Sox32</i> is a master inducer of endoderm and DFC fates.....	172
6.2 <i>Sox32</i> OE upregulates expression of Nodal ligands.....	173
6.3 Endoderm factors likely bind enhancers downstream of <i>sox32</i>.....	175
6.4 Separating <i>sox17+</i> endoderm from contaminating mesoderm at early organogenesis	176
6.5 Sorted GFP+ cells comprise both endoderm and KV transdifferentiated lineages at 28 h.p.f.....	177
6.6 The chromatin landscape changes in the endoderm during development	180
6.7 Future directions and experiments	182
The questions raised from this study are as follows:	182
References	185
Chapter 7 Supplementary.....	223

List of figures

FIGURE 1-1: ORTHOLOGUE GENES SHARED BETWEEN ZEBRAFISH AND HUMAN.....	3
FIGURE 1-2: NODAL SIGNALLING IS REQUIRED FOR MESENDODERM INDUCTION.....	4
FIGURE 1-3: FATE MAP OF THE ZEBRAFISH EMBRYO AT SHIELD STAGE (6 H.P.F.).....	5
FIGURE 1-4: TRANSCRIPTION FACTOR NETWORK IN ENDODERM SPECIFICATION IN ZEBRAFISH.....	7
FIGURE 1-5: NODAL SIGNALLING TRANSDUCTION PATHWAY.....	9
FIGURE 1-6: BINDING OF SOX FACTORS TO THE MINOR GROOVE RESULTS IN BENDING AND WIDENING OF THE DNA DOUBLE HELIX.....	15
FIGURE 1-7: SOX32 AND POU5F3 ARE ESSENTIAL FOR ENDODERM SPECIFICATION.....	16
FIGURE 1-8: CRMS REGULATE TRANSCRIPTION OF TARGET GENES.....	25
FIGURE 1-9: A SIMPLE SCHEMATIC OF CHROMATIN POSITIONING.....	28
FIGURE 1-10: LONG-RANGE ENHANCER-PROMOTER INTERACTIONS.....	30
FIGURE 1-11: MATERNAL-TO-ZYGOTIC TRANSITION IN ZEBRAFISH.....	31
FIGURE 1-12: SCHEMATIC OF A NUCLEOSOME CORE PARTICLE WITH FOUR CANONICAL HISTONES (H3, H4, H2A AND H2B).....	32
FIGURE 1-13: SCHEMATIC REPRESENTATION OF ATAC-SEQ LIBRARY PREPARATION.....	35
FIGURE 1-14: ENDODERM MAKES CONTRIBUTES TO THE EPITHELIAL LINING AND ASSOCIATED ORGANS OF THE GASTROINTESTINAL TRACT AND RESPIRATORY SYSTEM.....	37
FIGURE 3-1: FLUORESCENT IMAGES OF CONTROL AND <i>SOX32</i> MRNA INJECTED <i>TG(SOX17:EGFP)</i> EMBRYOS...	57
FIGURE 3-2: QRT-PCR ANALYSIS OF <i>SOX17</i> , <i>GFP</i> AND ENDOGENOUS <i>SOX32</i> IN CONTROL, <i>SOX32</i> KD AND <i>SOX32</i> OE EMBRYOS.....	58
FIGURE 3-3: SCHEMATIC OF EXPERIMENT PERFORMED AT 6 H.P.F. USING THE PAN-ENDODERMAL LINE <i>TG(SOX17:EGFP)</i>	59
FIGURE 3-4: INSERT SIZES FOR AMPLIFIED ATAC-SEQ FROM CONTROL, <i>SOX32</i> KD AND <i>SOX32</i> OE.....	60
FIGURE 3-5: DISTRIBUTION OF ATAC-SEQ PEAKS FROM CONTROL, <i>SOX32</i> KD AND <i>SOX32</i> OE ACROSS GENOMIC REGIONS WITH RESPECT TO THE NEAREST GENE.....	62
FIGURE 3-6: A CLUSTERED HEATMAP OF PEARSON CORRELATION COEFFICIENTS BASED ON MACS2 PEAK SCORES BETWEEN CONTROL, <i>SOX32</i> KD AND <i>SOX32</i> OE SAMPLES.....	63
FIGURE 3-7: A PEARSON CORRELATION COEFFICIENT HEATMAP BASED ON READ COUNTS THAT MAP EACH CONSENSUS REGION BETWEEN CONTROL, <i>SOX32</i> KD AND <i>SOX32</i> OE.....	64
FIGURE 3-8: A DIFFERENTIAL ACCESSIBILITY HEATMAP BETWEEN TWO SETS OF SAMPLE GROUPS.....	65
FIGURE 3-9: ATAC-SEQ READ DENSITIES SHOWING REGIONS OF ENHANCED ACCESSIBILITY PROXIMAL TO <i>CXCR4A</i> IN <i>SOX32</i> OE, WHICH IS A MARKER OF ENDODERM AND DFC FATES.....	66
FIGURE 3-10: HEATMAPS SHOWING READ DENSITIES ACROSS CONTROL, <i>SOX32</i> KD AND <i>SOX32</i> OE SAMPLES.....	67
FIGURE 3-11: OVER-REPRESENTED ANATOMICAL TERMS AT 6 H.P.F. FROM GREAT FOR <i>SOX32</i> OE AND CONTROL.....	69

FIGURE 3-12: HOMER MOTIF ANALYSIS SHOWING OVER-REPRESENTED TF MOTIFS IN <i>SOX32</i> OE OR CONTROL DARS.....	71
FIGURE 3-13: COMPARISON OF CONTROL AND <i>SOX32</i> OE DARS WITH NANOG CHIP-SEQ PEAKS.....	73
FIGURE 3-14: COMPARISON OF CONTROL AND <i>SOX32</i> OE DARS WITH TBX16 AND TBXTA CHIP-SEQ PEAKS. ...	75
FIGURE 3-15: COMPARISON OF CONTROL AND <i>SOX32</i> OE DARS WITH POU5F3 CHIP-SEQ PEAKS.	76
FIGURE 3-16: RNA-SEQ ANALYSIS COMPARISON OF GENES DIFFERENTIALLY EXPRESSED BETWEEN CONTROL AND <i>SOX32</i> OE EMBRYOS.	78
FIGURE 3-17: GSEA ENRICHMENT PLOTS.....	80
FIGURE 3-18: COMPARISON OF GENES ENRICHED IN THE DIFFERENT GERM LAYERS AT 6 H.P.F..	81
FIGURE 3-19: GSEA ENRICHMENT PLOTS FOR COMPARISON OF PROXIMAL GENES FROM CONTROL AND <i>SOX32</i> OE DARS TO RANKED GENE LISTS FROM RNA-SEQ ANALYSIS.	82
FIGURE 3-20: CHANGES TO THE CHROMATIN LANDSCAPE IN <i>SOX32</i> OE CORRELATES WITH GENE EXPRESSION.	83
FIGURE 3-21: ENDODERM MARKERS <i>SOX17</i> AND <i>FOXA2</i> ARE UPREGULATED IN <i>SOX32</i> OE EMBRYOS RELATIVE TO CONTROL.	85
FIGURE 3-22: DFC MARKER, <i>VGLL4L</i> IS UPREGULATED IN <i>SOX32</i> OE EMBRYOS RELATIVE TO CONTROL.	86
FIGURE 3-23: OVEREXPRESSION OF <i>SOX32</i> RESULTS IN UPREGULATION OF ENDOGENOUS <i>SOX32</i> EXPRESSION IN THE EMBRYO.....	88
FIGURE 3-24: OVEREXPRESSION OF <i>SOX32</i> UPREGULATES EXPRESSION OF <i>GATA5</i> IN <i>SOX32</i> OE RELATIVE TO CONTROL EMBRYOS.....	90
FIGURE 3-25: OVEREXPRESSION OF <i>SOX32</i> INDUCES EARLY AND UPREGULATED EXPRESSION OF NODAL LIGANDS IN <i>SOX32</i> OE RELATIVE TO CONTROL EMBRYOS.	92
FIGURE 4-1: MODEL FOR DFC INDUCTION.....	103
FIGURE 4-2: KV FUNCTION IN ZEBRAFISH.....	104
FIGURE 4-3: TIME COURSE OF <i>SOX17</i> EXPRESSION DURING THE FIRST 5 DAYS OF ZEBRAFISH DEVELOPMENT.	106
FIGURE 4-4: WHOLE-MOUNT <i>IN SITU</i> HYBRIDISATION SHOWING THE EXPRESSION PATTERN OF <i>SOX17</i> IN 12, 18 AND 26 H.P.F. ZEBRAFISH EMBRYOS.....	107
FIGURE 4-5: BRIGHTFIELD AND FLUORESCENCE MERGED MICROSCOPY IMAGES DEMONSTRATING THAT <i>SOX32</i> -KD LACK ENDODERM BUT DISPLAY GFP+ MESODERM.	110
FIGURE 4-6: IMAGING THE KV.	111
FIGURE 4-7: HIGH RESOLUTION IMAGES OF CONTROL AND <i>SOX32</i> KD EMBRYOS AT 24 H.P.F.	111
FIGURE 4-8: USING RHODAMINE-DEXTRAN (RD) TO ASSAY THE DEVELOPMENTAL FATE OF THE KV.....	112
FIGURE 4-9: SORTING OF GFP ^{HIGH} AND GFP ^{LOW} POPULATIONS FROM <i>TG(SOX17:EGFP)</i> EMBRYOS USING FACS.	114
FIGURE 4-10: SCHEMATIC OF ENDODERM SUB-REGIONS IN <i>TG(SOX17:EGFP)</i> EMBRYOS.	115
FIGURE 4-11: PAN-ENDODERMAL MARKERS USED FOR QRT-PCR.....	116
FIGURE 4-12: PHARYNGEAL ENDODERM MARKERS USED FOR QRT-PCR.....	117

FIGURE 4-13: POSTERIOR FOREGUT MARKERS USED FOR QRT-PCR.	118
FIGURE 4-14: MIDGUT AND HINDGUT MARKER USED FOR QRT-PCR.....	118
FIGURE 4-15: MESODERM-SPECIFIC MARKERS TESTED BY QRT-PCR.....	119
FIGURE 4-16: SCHEMATIC IMAGES DEMONSTRATING THE BREEDING PROCEDURES REQUIRED TO ATTAIN THE TRIPLE TRANSGENIC EMBRYOS AT 28 H.P.F.....	120
FIGURE 4-17: BRIGHTFIELD AND FLUORESCENCE MICROSCOPY IMAGES OF A TRIPLE TRANSGENIC EMBRYO AT 28 H.P.F.	121
FIGURE 4-18: SORTING CELLS FROM THE TRIPLE TRANSGENIC LINE INTO THREE CELL POPULATIONS (<i>SOX17N</i> , <i>SOX17E</i> AND <i>SOX17M</i>) USING FACS.....	122
FIGURE 4-19: QPCR ANALYSIS OF ENDODERM MARKERS ON <i>SOX17N</i> , <i>SOX17E</i> AND <i>SOX17M</i> CELL POPULATIONS.	123
FIGURE 4-20: QPCR ANALYSIS OF MESODERM MARKERS ON <i>SOX17N</i> , <i>SOX17E</i> AND <i>SOX17M</i>	124
FIGURE 4-21: SCHEMATIC OF EXPERIMENT PERFORMED USING TRIPLE TRANSGENIC EMBRYOS AT 28 H.P.F.	124
FIGURE 4-22: INSERT SIZES AS OBTAINED BY ATAC-SEQ ANALYSIS.	125
FIGURE 4-23: DISTRIBUTION OF GENOMIC LOCATIONS OF ACCESSIBLE REGIONS WITH RESPECT TO ANNOTATED GENES FOR <i>SOX17N</i> , <i>SOX17E</i> AND <i>SOX17M</i>	127
FIGURE 4-24: PEARSON CORRELATION HEATMAP BASED ON MACS2 SCORES.	128
FIGURE 4-25: A PEARSON CORRELATION HEATMAP BASED ON READ COUNTS.	129
FIGURE 4-26: DIFFERENTIAL ACCESSIBLE HEATMAPS SHOWING DIFFERENCES IN DARS BETWEEN TWO SETS OF SAMPLES.	131
FIGURE 4-27: AN EXAMPLE OF A GENE THAT IS HIGHLY ACCESSIBLE IN <i>SOX17E</i> COMPARED TO <i>SOX17N</i> AND <i>SOX17M</i> IS <i>SIX1B</i>	132
FIGURE 4-28: HEATMAPS SHOWING READ DENSITIES OF <i>SOX17N</i> , <i>SOX17E</i> AND <i>SOX17M</i> SURROUNDING <i>SOX17N</i> AND <i>SOX17E</i> DARS.	133
FIGURE 4-29: HEATMAPS SHOWING READ DENSITIES OF <i>SOX17N</i> , <i>SOX17E</i> AND <i>SOX17M</i> SURROUNDING <i>SOX17E</i> AND <i>SOX17M</i> DARS.	134
FIGURE 4-30: HEATMAPS SHOWING READ DENSITIES OF <i>SOX17N</i> , <i>SOX17E</i> AND <i>SOX17M</i> SURROUNDING <i>SOX17N</i> AND <i>SOX17M</i> DARS.....	135
FIGURE 4-31: BAR GRAPH SHOWING OVER-REPRESENTED ANATOMICAL TERMS ASSOCIATED WITH <i>SOX17E</i> , <i>SOX17N</i> AND <i>SOX17M</i>	136
FIGURE 4-32: A SELECTIVE NUMBER OF TF BINDING MOTIFS IN <i>SOX17E</i> DARS THAT ARE LIKELY CRITICAL DURING ENDODERM ORGANOGENESIS IN ZEBRAFISH.....	139
FIGURE 4-33: RNA-SEQ ANALYSIS FOR <i>SOX17E</i> VS <i>SOX17N</i>	143
FIGURE 4-34: GSEA ENRICHMENT PLOT COMPARING UPREGULATED TRANSCRIPTS FROM <i>SOX17E</i> AND <i>SOX17N</i> TO ENDODERM-ENRICHED GENES FROM DIFFERENT STAGES OF DEVELOPMENT (6, 8,10,14,18 AND 24 H.P.F) AND DFCS AT 8 H.P.F..	144

FIGURE 4-35: GSEA ENRICHMENT PLOT COMPARING UPREGULATED TRANSCRIPTS FROM <i>SOX17E</i> AND <i>SOX17N</i> SAMPLE GROUPS TO GENES PROXIMAL TO <i>SOX17N</i> , <i>SOX17E</i> AND <i>SOX17M</i> DARS OBTAINED FROM GREAT ANALYSIS.	145
FIGURE 4-36: AN EXAMPLE OF A GENE WITH UPREGULATED EXPRESSION IN <i>SOX17E</i> IS <i>EPCAM</i>	146
FIGURE 5-1: PEARSON CORRELATION HEATMAP BASED ON MACS2 SCORES.	157
FIGURE 5-2: PEARSON CORRELATION HEATMAP BASED ON READ COUNTS.....	158
FIGURE 5-3: A DIFFERENTIAL ACCESSIBLE HEATMAP BETWEEN <i>SOX32</i> OE AND <i>SOX17E</i>	159
FIGURE 5-4: ATAC-SEQ READ DENSITIES AT THE <i>FOXA3</i> LOCUS USING INTEGRATED GENOME VIEWER (IGV) SHOWING DIFFERENT PUTATIVE ENHANCERS AT 6 AND 28 H.P.F.....	160
FIGURE 5-5: BAR GRAPHS DEMONSTRATING OVER-REPRESENTED ANATOMICAL TERMS ASSOCIATED WITH <i>SOX32</i> OE AND <i>SOX17E</i> AT 6 AND 28 H.P.F., RESPECTIVELY.	162
FIGURE 5-6: MOTIF ANALYSIS DEMONSTRATING TF BINDING SITES IN <i>SOX32</i> OE AND <i>SOX17E</i> DARS.....	164
FIGURE 5-7: EXAMPLES OF GENES WHICH EXHIBIT DIFFERENTIAL EXPRESSION IN THE ENDODERM AT DIFFERENT STAGES OF DEVELOPMENT (6 VS 28 H.P.F.).	165
FIGURE 7-1: TOP GO TERMS ASSOCIATED WITH GENES FOUND PROXIMAL TO COMMON REGIONS BETWEEN (A) <i>SOX32</i> OE OR (B) CONTROL DARS AND NANOG CHIP-SEQ PEAKS ACCORDING TO GO ANALYSIS USING DAVID.	249
FIGURE 7-2: TOP GO TERMS ASSOCIATED WITH GENES FOUND PROXIMAL TO COMMON REGIONS BETWEEN (A) <i>TBX16</i> OR (B) <i>TBXTA</i> CHIP-SEQ PEAKS AND <i>SOX32</i> OE DARS ACCORDING TO GO ANALYSIS USING DAVID.	264
FIGURE 7-3: TOP GO TERMS ASSOCIATED WITH TRANSCRIPTS THAT WERE UPREGULATED IN <i>SOX32</i> OE VS CONTROL ACCORDING GO ANALYSIS USING DAVID.	290
FIGURE 7-4: BAR GRAPH DISPLAYING OVER-REPRESENTED ANATOMICAL TERMS ASSOCIATED WITH <i>SOX17E</i> AT 28 H.P.F.	308
FIGURE 7-5: TOP GO TERMS ASSOCIATED WITH TRANSCRIPTS THAT WERE UPREGULATED IN <i>SOX17E</i> VS <i>SOX17N</i> ACCORDING TO GO ANALYSIS USING DAVID.	348

List of tables

TABLE 2.1: OLIGONUCLEOTIDES USED FOR QRT-PCR	44
TABLE 2.2: TOTAL NUMBER OF MAPPED READ COUNTS PER BIOLOGICAL REPLICATE	47
TABLE 3.1: NUMBER OF CALLED PEAKS FOR EACH BIOLOGICAL REPLICATE PER CONDITION.....	60
TABLE 3.2: REPRESENTATIVE PEAKS FROM EACH SAMPLE CONDITION	61
TABLE 3.3: TOTAL NUMBER OF MAPPED READ COUNTS.....	64
TABLE 4.1: NUMBER OF CALLED PEAKS FOR EACH BIOLOGICAL REPLICATE PER CONDITION.....	126
TABLE 4.2: NUMBER OF CALLED PEAKS FROM MACS2.....	126
TABLE 4.3: TOTAL NUMBER OF MAPPED READS	128
TABLE 4.4: THE NUMBER OF DARS BETWEEN TWO SETS OF SAMPLE GROUPS	130
TABLE 7.1: LIST OF DARS BETWEEN <i>SOX32</i> KD AND CONTROL RANKED BY FDR FROM DIFFBIND.	223
TABLE 7.2: LIST OF TOP 200 DARS BETWEEN <i>SOX32</i> OE AND CONTROL RANKED BY FDR FROM DIFFBIND.	223
TABLE 7.3: A LIST OF 25 GENES THAT ARE FOUND PROXIMAL TO <i>SOX32</i> OE VS CONTROL DARS.....	229
TABLE 7.4: A LIST OF 25 GENES THAT ARE FOUND PROXIMAL TO CONTROL VS <i>SOX32</i> OE DARS.....	230
TABLE 7.5: TF BINDING MOTIFS FOUND OVER-REPRESENTED IN <i>SOX32</i> OE VS CONTROL DARS.....	231
TABLE 7.6: TF BINDING MOTIFS FOUND OVER-REPRESENTED IN CONTROL VS <i>SOX32</i> OE DARS.....	232
TABLE 7.7: A LIST OF GENES FOUND PROXIMAL TO <i>SOX32</i> OE DARS AND NANOG CHIP-SEQ COMMON REGIONS ACCORDING TO GREAT.....	233
TABLE 7.8: A LIST OF GENES FOUND PROXIMAL TO CONTROL DARS AND NANOG CHIP-SEQ COMMON REGIONS ACCORDING TO GREAT.	245
TABLE 7.9: A LIST OF GENES FOUND PROXIMAL TO <i>SOX32</i> OE DARS AND TBX16 CHIP-SEQ COMMON REGIONS ACCORDING TO GREAT.	249
TABLE 7.10: A LIST OF GENES FOUND PROXIMAL TO CONTROL DARS AND TBX16 CHIP-SEQ COMMON REGIONS ACCORDING TO GREAT.	255
TABLE 7.11: A LIST OF GENES FOUND PROXIMAL TO <i>SOX32</i> OE DARS AND TBXTA CHIP-SEQ COMMON REGIONS ACCORDING TO GREAT.	255
TABLE 7.12: A LIST OF GENES FOUND PROXIMAL TO CONTROL DARS AND TBXTA CHIP-SEQ COMMON REGIONS ACCORDING TO GREAT.	263
TABLE 7.13: A LIST OF UPREGULATED TRANSCRIPTS IN <i>SOX32</i> OE VS CONTROL EMBRYOS.	264
TABLE 7.14: LIST OF TOP 200 DARS BETWEEN <i>SOX17E</i> VS <i>SOX17N</i> RANKED BY FDR FROM DIFFBIND.	290
TABLE 7.15: LIST OF TOP 200 DARS BETWEEN <i>SOX17N</i> VS <i>SOX17M</i> RANKED BY FDR FROM DIFFBIND.	296
TABLE 7.16: LIST OF TOP 200 DARS BETWEEN <i>SOX17M</i> VS <i>SOX17E</i> RANKED BY FDR FROM DIFFBIND.	302
TABLE 7.17: A LIST OF 25 GENES THAT ARE FOUND PROXIMAL TO <i>SOX17E</i> VS <i>SOX17N</i> DARS.....	308
TABLE 7.18: A LIST OF 25 GENES THAT ARE FOUND PROXIMAL TO <i>SOX17M</i> VS <i>SOX17N</i> DARS.	309
TABLE 7.19: A LIST OF 25 GENES THAT ARE FOUND PROXIMAL TO <i>SOX17N</i> VS <i>SOX17E</i> DARS.....	310
TABLE 7.20: TF BINDING MOTIFS FOUND OVER-REPRESENTED IN <i>SOX17E</i> VS <i>SOX17N</i> DARS.	312
TABLE 7.21: A LIST OF UPREGULATED TRANSCRIPTS IN <i>SOX17E</i> VS <i>SOX17N</i> SAMPLE POPULATIONS.	313

TABLE 7.22: LIST OF TOP 200 DARS BETWEEN <i>SOX32</i> OE AND <i>SOX17E</i> RANKED BY FDR FROM DIFFBIND.	349
TABLE 7.23: A LIST OF 25 GENES THAT ARE FOUND PROXIMAL TO <i>SOX32</i> OE VS <i>SOX17E</i> DARS.	355
TABLE 7.24: A LIST OF 25 GENES THAT ARE FOUND PROXIMAL TO <i>SOX17E</i> VS <i>SOX32</i> OE DARS.	355
TABLE 7.25: TF BINDING MOTIFS FOUND OVER-REPRESENTED IN <i>SOX32</i> OE VS <i>SOX17E</i> DARS.	357
TABLE 7.26: TF BINDING MOTIFS FOUND OVER-REPRESENTED IN <i>SOX17E</i> VS <i>SOX32</i> OE DARS.	358

Declaration

This thesis is my original work and has been written by me, and all sources or collaboration have been acknowledged where applicable.

This thesis has not been submitted for consideration of any degree in any other university previously.

Acknowledgements

First and foremost, I would like to thank my primary PhD supervisor, Dr Andrew Nelson, for all his help and guidance and for supporting me during the last ~4 years. He was the primary source for getting my scientific questions answered and has been instrumental in helping with my PhD and I definitely wouldn't be here if it wasn't for his support. He was always very approachable and understanding even when stacked with teaching assignments. I have learnt so much from him and I am grateful for all his mentoring throughout.

I am also grateful to my second supervisor Dr Sascha Ott for providing the support I needed to analyse my genomics data. I also have to thank Dr Mark Walsh for providing the help I need with analysing RNA-seq and for being the "FACS" guy and a good friend.

I would like to extend my gratitude to my advisory panel Dr Andrew Blanks and Dr Kristen Panfilio for their advice throughout my PhD project and also providing me with ideas for experiments. Also want to thank Professor Karuna Sampath for telling everyone about the pizza story but also for letting me use her microscopes.

I would like to give special thanks to my husband Aasim Galal for being my rock in the last few years of my PhD and for constantly pushing me to keep going even when things got tough. I also want to thank my Warwick University friends Iago, Marco, and Magda who have supported me throughout the years. My PhD would certainly not have been as enjoyable as it was without them.

Lastly I want to thank my family, especially my mum, for their endless support and love. They always encouraged me to keep going and see the bigger picture and always believed in me.

Abstract

Precise spatial and temporal regulation of gene expression is critical for embryonic development. This is achieved in part via transcription factor binding to *cis*-regulatory modules (CRMs). However, in contrast to other germ layers, the gene regulatory network involved in endoderm formation is relatively understudied in vertebrates. In this study we use genomic approaches to study the chromatin landscape that governs endoderm specification and early endoderm organ formation in zebrafish embryos. Since the endoderm is a minor cell population, we employed two distinct methods to find to endoderm-specific CRMs: 1) to capture CRMs that regulate early endoderm-specific genes during early specification, we performed overexpression of the master regulator of endoderm formation, *sox32*, 2) to capture CRMs that function during early organogenesis, we have generated a three-colour transgenic line to specifically FACS isolate endoderm cells and study endoderm-specific CRMs at 28 hours post-fertilisation. Our results show that *sox32* overexpression not only induces ectopic endoderm formation but also ectopic formation of progenitors of the zebrafish left/right organiser known as dorsal forerunner cells (DFCs), a phenomenon not previously discussed in literature. We also identified 16, 000 regions that exhibit enhanced accessibility by ATAC-seq in embryos injected with *sox32* mRNA compared to control and that a large number of these regions harboured motifs of transcription factors that play important roles during endoderm specification. Our findings also suggest that sorted GFP⁺ cells from the triple transgenic line comprise both endoderm and transdifferentiated Kupffer's vesicle cells, which are implicated in left/right asymmetry in zebrafish. We also show by ATAC-seq analysis that a large number of accessible regions are likely shared by GFP⁺ cells and cells that are devoid of fluorescence. Finally, we show that changes in the chromatin landscape correlate with changes in gene expression at both stages of endoderm development through comparison of data from ATAC-seq and RNA-seq, but also that the chromatin landscape changes between endoderm specification and early organogenesis. Our results provide inside into how the chromatin landscape changes in the endoderm during development and also reveals a number of potentially important CRMs that may help shed light into how endoderm is formed during embryogenesis.

List of abbreviations

Abbreviations for model organism

<i>Xenopus</i>	<i>Xenopus laevis</i> , <i>Xenopus Africana</i>
Mice/Mouse	<i>Mus musculus</i>
Zebrafish	<i>Danio rerio</i>
Medaka	<i>Oryzias latipes</i>

Further abbreviations

3'UTR	3' untranslated region
ATAC-seq	assay for transposase-accessible chromatin using sequencing
bp	base pairs
BSA	bovine serum albumin
cDNA	complementary DNA
Cas9	CRISPR-associated protein 9
CDS	coding sequence
ChIP-seq	chromatin immunoprecipitation sequencing
CRISPR	clustered regularly interspaced short palindromic repeats
CRMs	<i>cis</i> -regulatory modules
DARs	differential accessibility regions
DFCs	dorsal forerunner cells
DNA	deoxyribonucleic acid
EDTA	ethylenediaminetetraacetic acid
TF	transcription factor
TFBS	transcription factor binding site
GFP	green fluorescent protein
GREAT	genomic regions enrichment of annotations tools
GO	gene ontology
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
HOMER	hypergeometric optimization of motif enrichment
h.p.f.	hours post-fertilisation
d.p.f	days post-fertilisation

ICM	Intermediate cell mass of mesoderm
kb	kilobase
KV	Kupffer's vesicle
LPM	Lateral plate mesoderm
mRNA	messenger RNA
MZ	maternal zygotic
MZT	maternal zygotic transition
PBS	phosphate buffered saline
PCR	polymerase chain reaction
qRT-PCR	quantitative reverse-transcription polymerase chain reaction
RNA	ribonucleic acid
RNA-seq	RNA sequencing
RPM	reads per million
<i>sox32</i>	SRY-box transcription factor 32
<i>sox32</i> OE	<i>sox32</i> overexpressing
<i>sox32</i> KD	<i>sox32</i> knockdown
<i>sox17</i>	SRY-box transcription factor 17
<i>sox17</i> N	<i>sox17</i> negative
<i>sox17</i> E	<i>sox17</i> + endoderm
<i>sox17</i> M	<i>sox17</i> + mesoderm
Tg	Transgene
<i>Tdgf1</i>	Teratocarcinoma-derived growth factor 1
TPM	transcripts per million
TADs	Topologically associating domains
YSL	yolk syncytial layer
ZGA	zygotic genome activation

Chapter 1 Introduction

1.1 Endoderm, the innermost germ layer

All tissues and organs in the human body consist of many diverse specialized cells that are formed during embryonic development (Kiecker, Bates and Bell, 2016). The primary cell types that contribute to these individual organs are known as the germ layers. The germ layers arise during a process called gastrulation (Kiecker, Bates and Bell, 2016), where undifferentiated embryonic blastula cells become segregated to three primary germ layers: ectoderm, mesoderm and endoderm (Pander, 1817). Endoderm is the innermost germ layer of all eumetazoan embryos, surrounded by mesoderm (middle layer) and ectoderm (outside layer). The study of the germ layers began in 1817 by the German biologist and embryologist, Christian Heinrich Pander, while investigating early chick development (Pander, 1817; Zorn and Wells, 2009). Traditionally, the endoderm was very difficult to study particularly due to its inaccessibility and location within the embryo, which made it impossible to visualise perturbed and normal development, but also because it comprises a small number of cells at any given stage (Nowotschin, Hadjantonakis and Campbell, 2019). For example, the endoderm represents approximately 3.6% of all cells within the mouse embryo at mid-gestation (Nowotschin *et al.*, 2019). However, identification of robust endoderm markers have tremendously improved our understanding of this tissue.

The endoderm gives rise to highly specialized epithelial cells that line the respiratory and digestive systems and contribute to the development of associated organs which include; thyroid, thymus, lungs, liver, biliary system and pancreas (Wlizla and Zorn, 2014) but also the pituitary gland (Fabian *et al.*, 2020). Disturbances to the formation of these organs are associated with many life-threatening diseases affecting millions of patients (Yiangou *et al.*, 2018), including diabetes (Spence and Wells, 2007). For example, mutations in pancreatic and duodenal homeobox 1 (PDX1), an essential gene for pancreas formation, has been found to cause pancreatic agenesis and neonatal diabetes in humans (Stoffers *et al.*, 1997; Piper *et al.*, 2002; Thomas *et al.*, 2009). Thus, diseases affecting early endoderm development are likely to be detrimental to organ formation and represents a need for tissue replacement therapy.

The use of diverse animal models has aided our understanding of endoderm development and how this is perturbed in diseases. Mammalian models such as the mouse (*Mus musculus*)

were historically exploited as model organisms due to their high degree genetic similarity to humans (>80%), highly conserved anatomical and physiological features and a diverse array of genetic targeting strategies to recapitulate human disease phenotypes (Waterston *et al.*, 2002; Devoy *et al.*, 2012). However, systematic mutagenesis studies were difficult due to the long generation times, large space requirements and laborious maintenance and breeding of these animals (Haffter *et al.*, 1996; Devoy *et al.*, 2012). By contrast, invertebrate models such as nematode worm (*Caenorhabditis elegans*), or the fruit fly (*Drosophila melanogaster*) offer inexpensive alternatives and were proven to be powerful models in identifying orthologs of human genes through genetic manipulations (Antoshechkin and Sternberg, 2007; Wangler, Yamamoto and Bellen, 2015). However, only a small number of genes are shared between worm and fly vs human (43% and 61%, respectively) (Lander *et al.*, 2001). Due to a high scale degree of homology between genes of different vertebrates, it was imperative that an animal model was chosen that enables large-scale mutagenesis screens to be performed. For this reason, many mutagenesis screens were performed on zebrafish, which in turn led to the characterisation of more than 400 genes required for morphogenesis, pattern formation, organogenesis, and differentiation (Mullins *et al.*, 1994; Solnica-Krezel, Schier and Driever, 1994; Driever *et al.*, 1996). These studies also led to the characterisation of 23 genes affecting liver, intestine and kidney formation in zebrafish (Chen *et al.*, 1996). Thus, these studies augmented the importance of using zebrafish as an animal model of vertebrate embryogenesis.

1.2 Zebrafish as an animal model

Studies in a number of model organisms, including mouse, frog, chick and fish have greatly enhanced our understanding of vertebrate endoderm development (Zorn and Wells, 2009). Although mice are evolutionary more similar to humans, the freshwater aquarium zebrafish, a relative newcomer to the list of animal models, have become a popular model for studying human development and disease. Firstly, zebrafish are genetically similar to humans; approximately 70% of human genes have functional homologs in zebrafish (Langheinrich, 2003; Howe *et al.*, 2013), (Figure 1-1). In addition, zebrafish have high organ system homology to humans, high fecundity and produce large clutches of embryos per spawning, which are not only optically transparent but also develop rapidly *ex utero*, with major organs forming by 24 hours post-fertilisation (h.p.f) (Santoriello and Zon, 2012; Seth, Stemple and Barroso, 2013). This makes them an ideal animal model for studying early development (Dawid, 2004). However, perhaps the most popular reason for using zebrafish is the fact that transgenic reporter lines can easily be generated to allow for genetic screening and identification of regulatory elements.

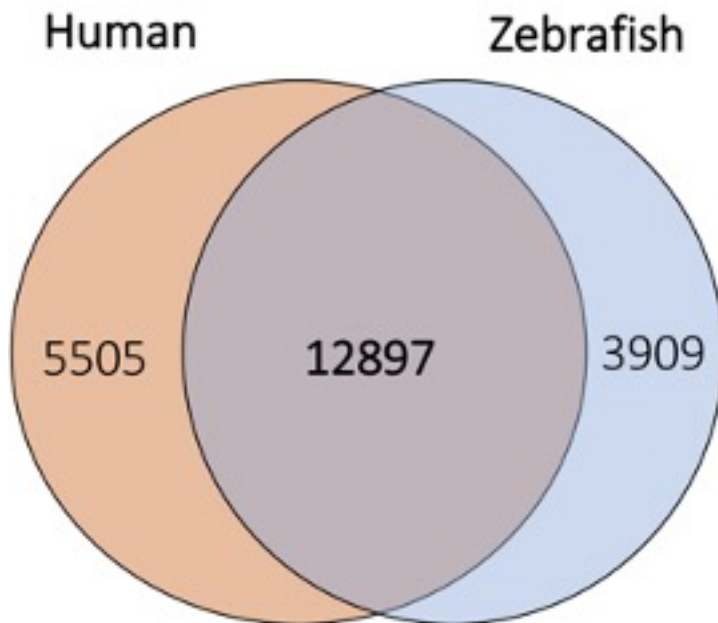


Figure 1-1: Orthologue genes shared between zebrafish and human. Gene numbers extracted from: (Howe *et al.*, 2013).

Zebrafish were first identified to be genetically tractable during the 1980s which led to their immense usage in high throughput phenotypic screenings and discovery of hundreds of mutants with defects in embryogenesis during the 1990s (Mullins *et al.*, 1994; Solnica-Krezel, Schier and Driever, 1994; Driever *et al.*, 1996). Their suitability for mutagenesis screening makes zebrafish a powerful model for studying human embryogenesis, development and disease (Haramis *et al.*, 2006).

1.3 Segregation of the three germ layers

Segregation of the three primary germ layers occurs during gastrulation and marks one of the earliest cell fate decisions that occurs during development (Warga and Nüsslein-Volhard, 1999). In zebrafish, the endoderm and mesoderm are derived from cells near the blastoderm margin (Kimmel, Warga and Schilling, 1990) and both involute towards the margin during gastrulation to form the hypoblast, whereas ectoderm is formed from cells that remain in the layer farther from the margin called the epiblast (Warga and Kimmel, 1990). Initially, while the mesoderm and endoderm derivatives within the hypoblast are morphologically indistinguishable, studies have shown that time of involution is related to cell fate; earlier involuting cells predominantly form the endoderm while cells that involute later form the mesoderm (Warga and Kimmel, 1990). By labelling single cells at blastula margin, it was revealed that mesoderm and endoderm arise from a common precursor cell population, known as the mesendoderm (Warga and Nüsslein-Volhard, 1999). By mid-gastrulation

however, endoderm cells adopt a unique morphology and are located in the deepest layer of the hypoblast while, the mesoderm arise from cells located between the epiblast and the endoderm cells (Kimmel, Warga and Schilling, 1990).

Genetic studies have shown that endodermal cell fate induction is controlled by Nodal signalling, which consists of Nodal growth factors but also a group of core downstream transcription factors (Zorn and Wells, 2009). Nodal signalling acts as a morphogen; it elicits concentration-dependent responses in its target cells and can act at a distance from its site of production (Chen and Schler, 2001; Meno *et al.*, 2001; Schier, 2009). In support of this, high levels of Nodal growth is required for endoderm specification, while low levels induce mesoderm development (Schier, 2003; Schier and Talbot, 2005; Shen, 2007), (Figure 1-2), and this dose-dependent activity is highly conserved in fish, mouse and humans (D'Amour *et al.*, 2005; Shen, 2007; Zorn and Wells, 2007). In zebrafish, β -catenin-dependent signal(s) in the extraembryonic yolk syncytial layer (YSL) stimulates high levels of Nodal-related transcription within the first six cell tier in the dorsal-anterior mesendoderm (Fan *et al.*, 2007; Zorn and Wells, 2009; van Boxtel *et al.*, 2015) resulting in endodermal structures primarily arising from the dorsal margin (Warga and Nüsslein-Volhard, 1999).

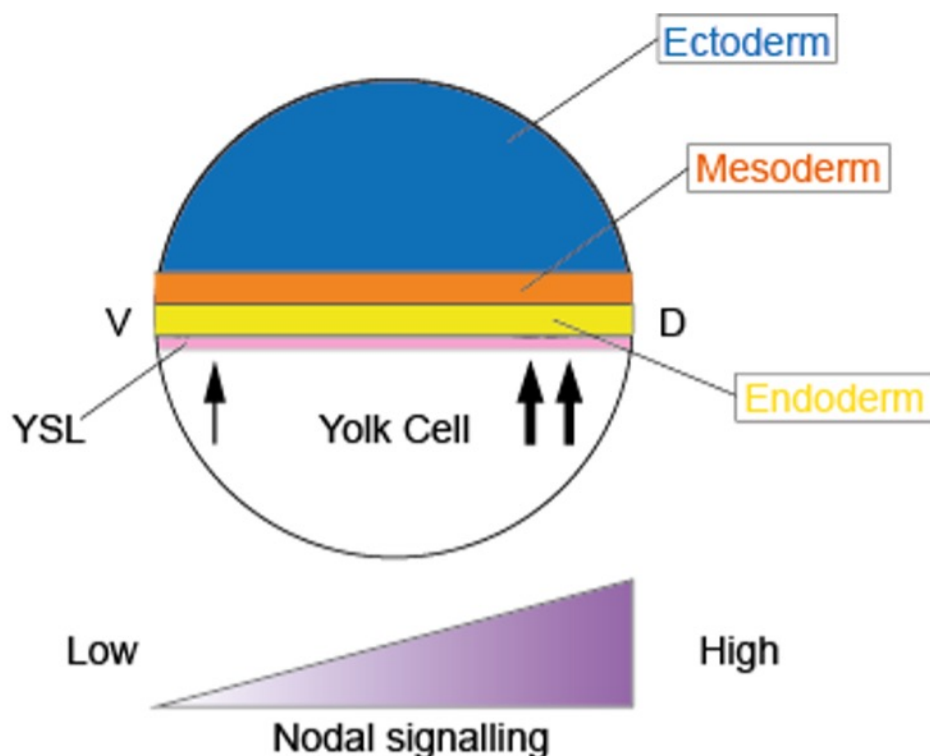


Figure 1-2: Nodal signalling is required for mesendoderm induction. Nodal signalling is highest on the dorsal side of the embryo from which endoderm progenitors arise, while low Nodal signalling is required for mesoderm specification. Arrows represent Nodal signalling. V and D represent Ventral and Dorsal axis, respectively.

1.4 Fate map of zebrafish embryo

At early gastrulation stage, roughly around shield stage (6 h.p.f.), the different regions of the final body can be roughly mapped onto the embryo (Figure 1-3). The precursors of the different germ layers can be arranged along the animal-vegetal axis. The endoderm and mesoderm are intermingled along the marginal blastomeres (Kimmel, Warga and Schilling, 1990). Early mesodermal precursors are also arranged along the dorsal-ventral axis (Robert, 1992). For instance, cells that contribute to the slow muscle, notochord and head are located at the dorsal side of the embryo, while cells located at the ventral side contribute to blood, kidney, somite and heart (Kimmel, Warga and Schilling, 1990). On the other hand, cells furthest from the margin, located at the animal pole, adopt an ectodermal fate.

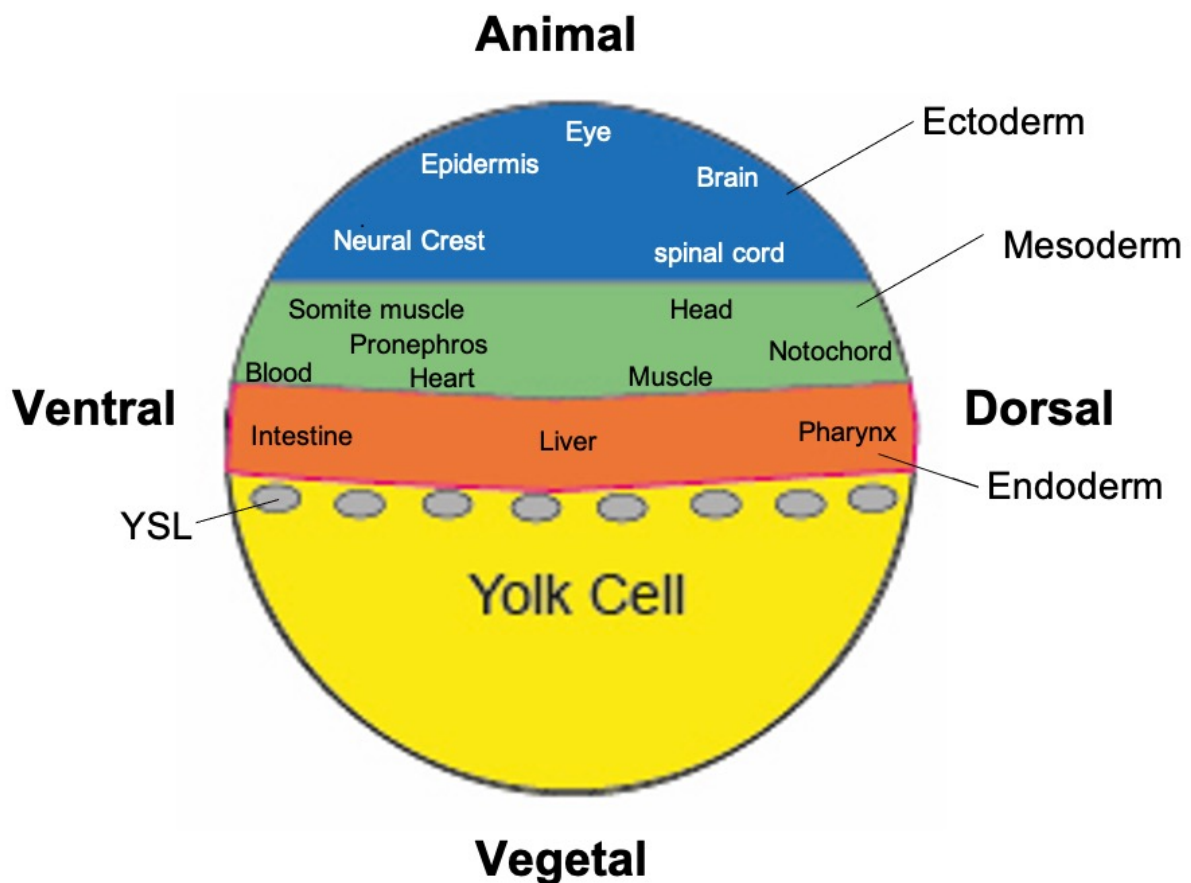


Figure 1-3: Fate map of the zebrafish embryo at shield stage (6 h.p.f.). Cells closest to the margin adopt an endoderm fate. Cells furthest from the margin adopt an ectoderm fate.

1.5 YSL in zebrafish

In teleosts, the YSL is an extraembryonic and multinucleated structure located at the margin that forms between the yolk and blastoderm. Formation of the YSL begins at late cleavage stages when the marginal cells are confluent with the yolk cell cytoplasm (Carvalho and Heisenberg, 2010). Between 512 and 1k-cell stage just after mid-blastula transition (3 h.p.f.), and at the onset of zygotic transcription, some marginal blastomeres undergo acute membrane collapse and deposit their content into the cytoplasmic cortex of the yolk cell, forming the YSL (Kimmel, Warga and Schilling, 1990; Carvalho and Heisenberg, 2010). The nuclei in the YSL, called yolk syncytial layer nuclei, primarily line up in a single row along the marginal blastomeres and undergo mitosis without cytokinesis until they eventually become post-mitotic prior to gastrulation (Kane, Warga and Kimmel, 1992; Trinkaus, 1993; Kimmel *et al.*, 1995). Thus, the YSL is a syncytium containing a large number of yolk syncytial layer nuclei, which exhibit high transcriptional activity (Williams *et al.*, 1996).

Nodal signalling from the YSL is thought to be required for mesendoderm specification. In zebrafish, two of the three Nodal ligands, *ndr1* and *ndr2* are expressed in both the YSL and surrounding marginal blastomeres and are both required and sufficient for dorsal mesoderm formation (Erter, Solnica-Krezel and Wright, 1998; Feldman *et al.*, 1998; Rebagliati, Toyama, Haffter, *et al.*, 1998; Sampath *et al.*, 1998). Furthermore, *ndr1* and *ndr2* expression in the YSL induce expression of *ndr1* in the neighbouring marginal blastomeres creating stable Nodal signalling that is required for endoderm and head mesoderm fate induction (Fan *et al.*, 2007). This observation led to the speculation that the YSL shares a common evolutionary origin with the mouse anterior visceral endoderm (AVE), which also plays a critical role in early head patterning (Foley and Stern, 2001; Martinez-Barbera and Beddington, 2001). In relation to this, targeted depletion of Nodal signalling from the YSL impaired both endoderm and head mesoderm formation (Fan *et al.*, 2007). Similarly, mice that lack *nodal* function in the visceral endoderm exhibit defects in the prechordal plate, which affects anterior neural structures (Varlet, Collignon and Robertson, 1997; Brennan *et al.*, 2001). Thus, these results likely suggest that the YSL is functionally equivalent to the mouse AVE.

The function of the YSL in endoderm and mesoderm induction was highlighted by transplantation experiments where biotin-labelled yolk cells, including the YSL, from an embryo was grafted onto the animal-pole region of an unlabelled host embryo at mid-blastula stage resulting in ectopic expression of dorsal markers, including mesodermal makers *tbxta* and *gsc*, in the host animal pole (Mizuno *et al.*, 1996). Moreover, the YSL appears to be required for ventrolateral mesoderm since injection of RNase into the yolk cell after 1k-cell

stage eliminated ventrolateral expression of markers including *tbxta*, *gsc*, *ndr1* and *ndr2*, while dorsal expression of these markers remained unaffected (Chen and Kimelman, 2000). On the other hand, dorsal mesendoderm induction requires stabilization of β -catenin in the dorsal marginal blastomeres and YSL, with evidence that embryos treated with lithium chloride (LiCl) which induces β -catenin stabilization across the embryo, resulted in ectopic expression of dorsal markers, *dharma* and *gsc* (Chen and Kimelman, 2000). Thus, signals from the YSL are necessary and sufficient for mesendoderm formation.

1.6 Pathway to endoderm formation

1.6.1 Nodal signalling pathway

Nodal signalling is the main signalling pathway involved in endoderm specification, as mentioned above, and it begins with maternally contributed factors required for Nodal induction and ends with expression of the zebrafish master inducer of endoderm fate, *SRY* (*sex determining region Y*)-box 32 *sox32* (Figure 1-4).

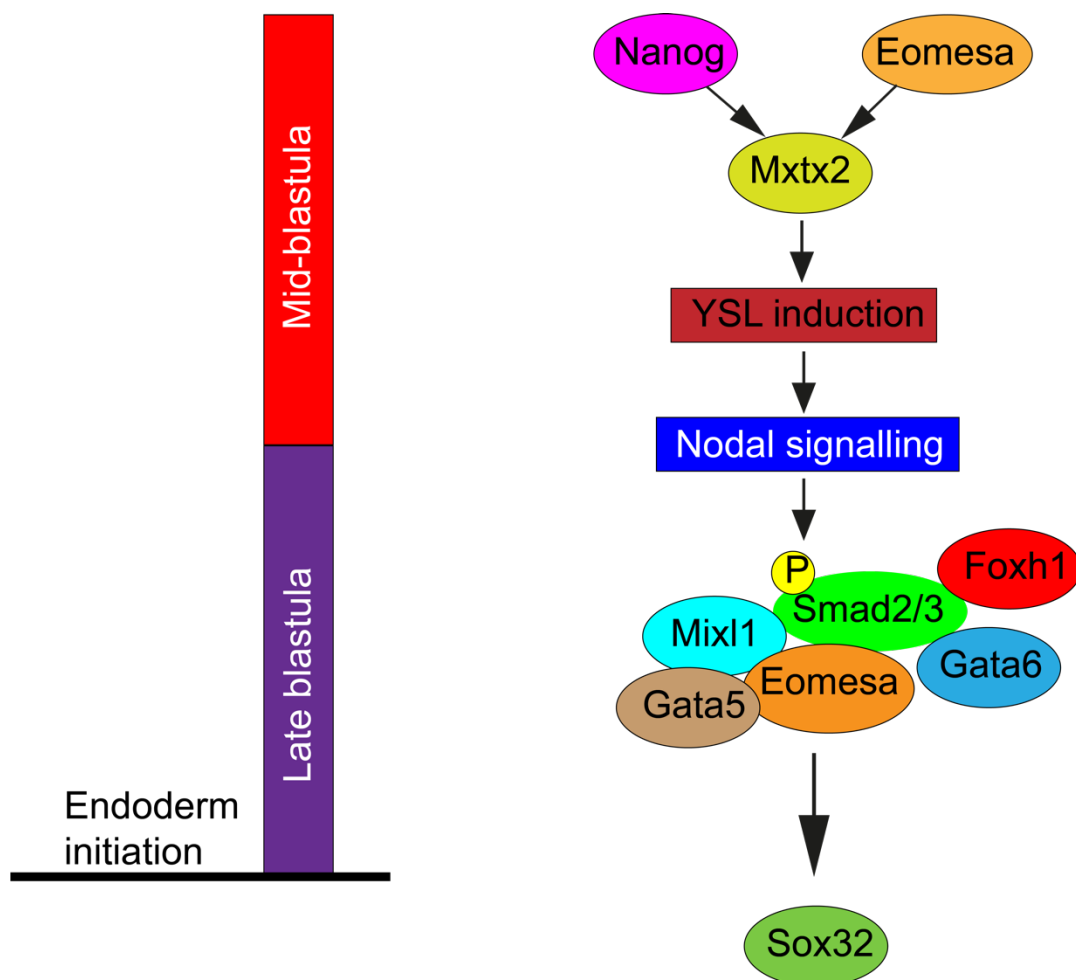


Figure 1-4: Transcription factor network in endoderm specification in zebrafish. At mid-blastula stage (~3 h.p.f.), maternal factors Nanog and Eomesa are required to induce expression of *mxtx2*, a key regulator of YSL formation from which *ndr1* and *ndr2* are expressed. Once Nodal signalling is activated, phosphorylated Smad2 binds to a

number of transcription factors, including Foxh1, Eomesa, Mixl1 and Gata6, which results in the expression of endodermal regulators. Eomesa also physically interacts with Mixl1 and Gata5 to initiate expression of *sox32*. Once *sox32* is expressed, endoderm specification begins.

Nodal ligands are members of the transforming growth factor β (TGF β) family of secreted growth factors (Zorn and Wells, 2009). In zebrafish, there are three nodal-related genes: *ndr1* and *ndr2*, which are required for mesendoderm formation (Feldman *et al.*, 1998), while *southpaw* is expressed after gastrulation and is required for left/right patterning establishment (Long, Ahmad and Rebagliati, 2003). Like other TGF β proteins, Nodal proteins are translated as precursors, consisting of a long N-terminal prodomain (Le Good *et al.*, 2005) and a mature ligand domain (Schier, 2003). Processing of the Nodal precursor by proprotein convertases is essential for activation of the Nodal signalling pathway in zebrafish (Zorn and Wells, 2007). Nodal ligands signal via a heteromeric receptor complex consisting of type I (Alk4 or Alk7) and type II (ActRIIA or ActRIIB) activin transmembrane serine-threonine kinase receptors as well as an EGF-CFC coreceptor, Teratocarcinoma-derived growth factor 1 (Tdgf1) (Gritsman *et al.*, 1999; Schier, 2009). In embryos that lack both *ndr1* and *ndr2* or maternal and zygotic *tdgf1* (MZ*tdgf1*) activity, marginal cells fail to involute and exhibit a dramatic reduction in subsequent endoderm and mesoderm (Feldman *et al.*, 1998; Gritsman *et al.*, 1999), suggesting a critical role for Nodal signalling in mesendoderm induction.

In zebrafish, *ndr1* is maternally contributed and is expressed during oogenesis, but its function at this stage is not very well understood (Gore and Sampath, 2002; Gore *et al.*, 2005; Schier, 2005; Hagos, Fan and Dougan, 2007). However, by four-cell stage, its localisation predicts the embryonic dorsal axis (Gore *et al.*, 2005). *Ndr1* expression begins at after mid-blastula stage (3 h.p.f.) in the blastoderm margin and this is believed to be dependent on the function of maternal dorsal determinant β -catenin (Dougan *et al.*, 2003; Bellipanni *et al.*, 2006). Following this, *ndr1* expression extends into the YSL and enveloping layer (EVL) (Erter, Solnica-Krezel and Wright, 1998; Feldman *et al.*, 1998). During later blastula stages, *ndr1* activates expression of *ndr2* at the dorsal margin and this co-expression is required for mesendoderm induction since *ndr1:ndr2* double mutants lack all derivatives of mesoderm and endoderm (Feldman *et al.*, 1998; Dougan *et al.*, 2003). *Ndr1* and *ndr2* at late blastula stage are thought to specify mesendoderm formation by forming a signalling gradient in the first six tiers from the margin in zebrafish. Upon activation of Nodal signalling, Alk4/7 phosphorylates the cytoplasmic signal transducer protein Smad2 and/or Smad3, leading to their interaction with Smad4 and the translocation of the Smad2/3-Smad4 complex to the nucleus (Schier, 2003), (Figure 1-5). Once in the nucleus, Smad2/3-Smad4 complexes act together with a number of transcription factors, including Forkhead box h1 (Foxh1), Eomesodermin homolog

A (Eomesa) and Mix paired-like homeobox (Mix1) to regulate expression of key mesendoderm genes (Chen, Rubock and Whitman, 1996; Chen *et al.*, 1997; Germain *et al.*, 2000; Kunwar *et al.*, 2003; Picozzi *et al.*, 2009; Nelson *et al.*, 2014). Interestingly, TGF β also activates non-Smad2 pathways including c-Jun N-terminal kinase (JNK), which regulates nuclear movement towards the YSL in the most laterally located marginal cells, independently of the Smad2 pathway, and as a result, aids in the transduction of the Smad2 signal towards the nucleus during endoderm specification (Hozumi, Aoki and Kikuchi, 2017). Thus, Nodal promotes two signalling pathways during endoderm specification: one mediated by phosphorylated Smad2, and the other mediated by JNK.

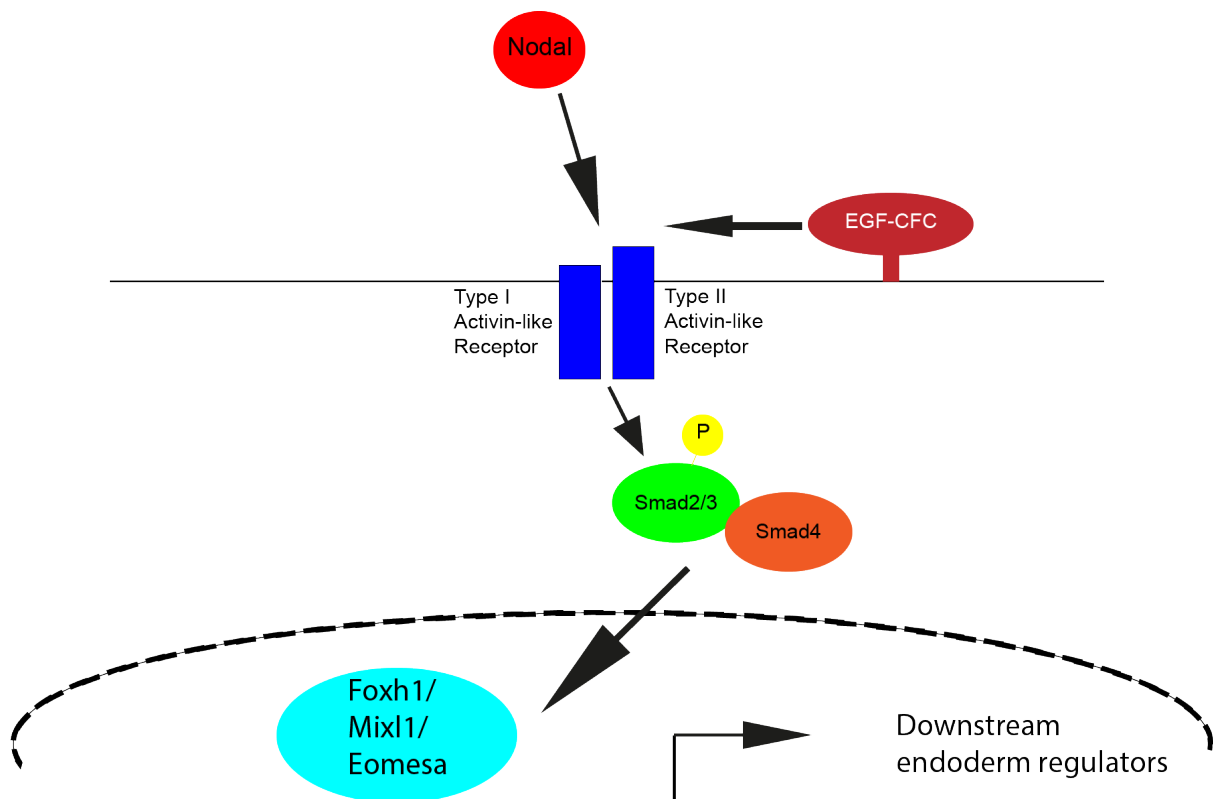


Figure 1-5: Nodal signalling transduction pathway. Binding of the Nodal ligands to the ActRIIB receptors and EGF-CFC coreceptors results in activation of Smad transcription factors and their subsequent translocation into the nucleus where they activate genes required for endoderm specification.

However, other TGF β members are also believed to be critical for Nodal signalling, including growth differentiation factor 3 (Gdf3). In zebrafish, Nodal ligands require Gdf3 to activate Nodal receptors in order to induce mesendoderm and without Gdf3, embryos fail to form endoderm and mesoderm (Montague and Schier, 2017). Also, *gdf3* mutants, strongly resemble Nodal or *tdgf1* mutants (Feldman *et al.*, 1998; Gritsman *et al.*, 1999) which suggests that Nodal signalling requires Gdf3 and that Nodal factors are not the sole mesendoderm inducers.

1.6.2 Maternally-contributed factors

Aside from β -catenin, induction of *ndr1* and *ndr2* at the blastoderm margin is controlled by multiple other maternally-contributed factors including Nanog and Eomesodermin homolog A. (Eomesa). Broader expression of Nodal factors depend on Nanog and Eomesa, which activate the homeobox transcription factor, Mxtx2, a key regulator of YSL formation (Bruce *et al.*, 2005; Hong *et al.*, 2011; Xu *et al.*, 2012). Knockdown of *mxtx2* results in a lethal defect in epiboly and a reduction in mesendoderm gene expression which includes *ndr1* and *ndr2* (Hong *et al.*, 2011). This was further supported by morpholino studies, which revealed that *nanog* morphants exhibit reduced *mxtx2* expression and fail to develop the YSL which produces Nodal required for endoderm formation (Xu *et al.*, 2012; Gagnon, Obbad and Schier, 2018). Similarly, knockdown of Eomesa led to partial or incomplete loss of *mxtx2* (Du *et al.*, 2012). In addition to this, Eomesa also promotes expression of zygotic *ndr1* and *ndr2* incorporation with Mxtx2 (Xu *et al.*, 2014). Thus, maternal factors Eomesa, Nanog as well as zygotic Mxtx2 are required to induce expression of Nodal ligands, which in turn are essential for endoderm induction in the zebrafish embryo.

However, these maternal factors also play other essential roles. For instance, in mammals Nanog alongside other key pluripotent factors including Pou5f1 and Sox2 control expression of target genes required to maintain embryonic stem cell pluripotency (Onichtchouk and Driever, 2016). In zebrafish, Nanog and the closest fish homolog of Pou5f1, called Pou5f3 and Sox2, called Sox19b, are essential for zygotic genome activation, where zygotic gene transcription begins and maternal proteins are degraded such that the embryo solely depends on zygotic transcription (Lee, Bonneau and Giraldez, 2014). This was validated by morpholino injections against Pou5f3, Nanog and Sox19b which was correlated with reduced degradation of maternal factors and failure to activate zygotic transcription likely due to decreased chromatin accessibility around zygotic target genes (Pálffy *et al.*, 2020). In relation to this, in humans, NANOG regulates activity of EOMES which actively directs the differentiation of embryonic stem cells towards a definitive endoderm fate (Teo *et al.*, 2011). However, in zebrafish, these maternal factors appear to mediate Nodal signalling in aiding endoderm formation, as discussed below.

1.6.3 From Nodal induction to endoderm specification

Smad2/3-Smad4 have poor or no affinity for DNA and thus, once in the nucleus, activated Smad2-Smad4 complex must associate with other transcription factors such as the winged-helix Fork head box h1 (Foxh1/FAST1) to mediate TGF- β -mediated mesendoderm

transcriptional activation (Chen, Rubock and Whitman, 1996). Foxh1 is expressed maternally and zygotically and binds to active Smad complexes using its highly conserved Smad interaction Motif (SIM), which is present in the C-terminal region (Germain *et al.*, 2000; Randall *et al.*, 2002). Foxh1 SIM binds to the MH2 domain present in the C-terminal region of phosphorylated Smad2, which in turn binds to Smad4 in the nucleus (Germain *et al.*, 2000; Randall *et al.*, 2002). Knockdown of *foxh1* activity results in a partial loss of endoderm and axial mesoderm, while injection of Foxh1 RNA rescues early endoderm regulators in *MZtdgf1* mutants (Pogoda *et al.*, 2000), suggesting a role of Foxh1 in endoderm induction. Moreover, in *MZfoxh1* (also known as *MZsur*) mutants, expression of Nodal genes *ndr1* and *ndr2* are initiated but not maintained and thus, it is likely that Foxh1 is required to maintain expression of Nodal ligands during endoderm induction (Pogoda *et al.*, 2000). However, *MZfoxh1* mutants exhibit only minor defects in endoderm formation which likely suggests that Nodal signalling is mediated by the interaction of Smad2/3 with other transcription factors during endoderm formation (Schier, 2003).

Another interacting partner of Smad2/3 is Eomes, a T-box factor required for endoderm specification in mouse (Arnold *et al.*, 2008). In zebrafish, there are two homologues of Eomes, *eomesa* and *eomesb*, however, unlike *eomesa* which is maternally and zygotically expressed (Bruce *et al.*, 2003), *eomesb* is not expressed during early development and does not participate in endoderm specification (Figiel, Elsayed and Nelson, 2021). Overexpression of Eomesa induces expression of endodermal markers at early gastrula stages (Bjornson *et al.*, 2005). However, maternal and zygotic *eomesa* (*MZeomesa*) mutants exhibit only a slight loss of endodermal genes, suggesting that Eomesa is not absolutely required for endoderm specification (Du *et al.*, 2012).

Interestingly, Eomesa and Foxh1 are believed to act redundantly in mediating Nodal signalling. In zebrafish, Foxh1 and Eomesa single mutants appear to exhibit only a partial loss of endoderm, however, morpholino knockdown of Foxh1 in *MZeomesa* completely abolishes expression of *SRY* (*sex determining region Y*)-*box 32* (*sox32*)-the master inducer of endoderm fates in zebrafish (Nelson *et al.*, 2014). In relation to this, inhibition of Eomesa activity in *foxh1* mutants effectively phenocopies the phenotypes observed upon complete loss of Nodal signalling including loss of endoderm markers (Slagle, Aoki and Burdine, 2011). Given that overexpression of *foxh1* can partially rescue expression of *sox32* in *MZeomesa* (Nelson *et al.*, 2014), it is therefore likely that Eomesa and Foxh1 are partially redundant but also act cooperatively to induce endoderm specification.

There are therefore discrepancies between mammalian Eomes and zebrafish Eomesa since Eomes is absolutely required for endoderm specification with evidence that *Eomes* mutants lack expression of all markers of definitive endoderm (Arnold *et al.*, 2008; Teo *et al.*, 2011). This is likely because in zebrafish, Eomesa is believed to be functionally redundant in the presence of other members of the T-box family such as Tbx16 and Tbx16. Eomesa, Smad2 as well as Tbx16 and Tbx16 also appear to share common genomic regions and have been shown to regulate expression of key endodermal modulators including *mixl1* during early gastrula stages (Nelson *et al.*, 2017). Thus, it is likely that in zebrafish, *eomesa* mutants exhibit mild defects due to functional redundancy amongst all T-box factors. Interestingly Tbx16 appears to be missing in most mammals except in marsupials and monotremes (Ahn, You and Kim, 2012). In zebrafish, loss of Tbx16 or Eomesa function alone results in a slight decrease in *mixl1* expression, while knockdown of Tbx16 in *eomesa* mutants results in a severe loss of *mixl1* (Talbot *et al.*, 2020). This likely suggests that Tbx16 is able to compensate for the loss of *eomesa* in MZeomesa mutants in zebrafish (Du *et al.*, 2012).

In zebrafish, Tbx16 and Tbx16 also appear to share partially overlapping functions. Though Tbx16 is required for notochord formation (Halpern *et al.*, 1993; Schulte-Merker *et al.*, 1994; Martin and Kimelman, 2008) and Tbx16 is key in regulating migration of mesodermal progenitors and trunk somite formation (Kimmel *et al.*, 1989; Ho and Kane, 1990; Griffin *et al.*, 1998), they both also appear to collectively regulate expression of key modulators of endoderm proliferation including *chemokine (C-X-C motif) ligand 12b (cxcl12b)* (Nelson *et al.*, 2017). Also, whole-mount *in situ* hybridisation revealed that endoderm cells were substantially reduced in *tbx16* and *tbx16* double morphants but were only moderately reduced in *tbx16* morphants and barely reduced in *tbx16* morphants (Nelson *et al.*, 2017). In relation to this, liver, gut, and pancreas were also severely compromised in *tbx16* and *tbx16* double morphants compared to single *tbx16* and *tbx16* morphants, which suggests that both Tbx16 and Tbx16 act redundantly in liver, pancreas, and gut development but also that Tbx16 appears to have a greater influence on endoderm formation.

T-box factors including Tbx16, Tbx16 and Eomesa are characterised by the presence of a T-box DNA binding domain, named after the founding member Brachyury also known as T, which was originally identified as a mouse mutant that displayed a short-tail phenotype (Dobrovolskaia-Zavadskaja, 1927). There are currently 17 members of the T-box containing transcription factors (Ghosh, Brook and Wilsdon, 2017). Genome-wide chromatin profiling of multiple T-box factors including Brachyury, Eomes and Tbx6 in mice, Humans, *Xenopus* and zebrafish were shown to most frequently recognise variants of an eight to nine base pair core sequence of (T)TVRCACT, which likely suggests that this T-box DNA recognition sequence

is somewhat conserved amongst different T-box factors and across different species (Morley *et al.*, 2009; Teo *et al.*, 2011; Nelson *et al.*, 2012, 2014, 2017; Gentsch *et al.*, 2013; Lolas *et al.*, 2014; Faial *et al.*, 2015; Tsankov *et al.*, 2015; Windner *et al.*, 2015). This therefore likely enables different T-box members to bind the same genomic site. Thus, it is likely that Eomesa, Tbx16 and Tbx1a exhibit redundant activities, by binding the same partners (Smad2) and regulating the same proximal genes (*mixl1*).

At the onset of zygotic genome activation, many of the Smad2/3 target genes have been shown to also bind Smad2 as part of a feedforward mechanism. For instance, Mix paired-like homeobox (*Mixl1*), seems to bind the same genomic regions as Eomesa and Smad2 as revealed by CHIP-seq and that these regions are proximal to key regulators of endoderm formation (Nelson *et al.*, 2017). These include *tbx16*, *sox32* and chemokine (C-X-C motif) receptor 4a (*cxcr4a*), a cell surface receptor that binds to *cxcl12a/b* to promote correct endoderm migration and proliferation (Stückemann *et al.*, 2012). Like Foxh1, *Mixl1* also contains a SIM in its C-terminal, which is necessary for its interaction with phosphorylated MH2 domain of Smad2 (Germain *et al.*, 2000) to aid in the expression of mesendoderm targets. Incidentally, *Mixl1* SIM is believed to inhibit the binding between Foxh1 and activated Smad2/3-4 complex (Germain *et al.*, 2000). This indicates that *Mixl1* SIM can compete with Foxh1 for its interaction with phosphorylated Smad complex. Although Foxh1 is not required for endoderm formation, it is required for maintenance and enhanced expression of *mixl1* (Kunwar *et al.*, 2003) since *mixl1* expression was significantly reduced in *Foxh1* mutants (Slagle, Aoki and Burdine, 2011). In zebrafish, *Mixl1* plays a crucial role in endoderm formation as *mixl1* mutants exhibit a substantial decrease in endoderm cells (Kikuchi *et al.*, 2000). However, the cells that do form endodermal precursors differentiate normally which means that *mixl1* is not required for later stages of endoderm development. In addition, overexpression of *mixl1* was shown to increase the number of endodermal cells in *tdgf1* mutants, (Alexander and Stainier, 1999) as well as in *ndr1* and *ndr2* double mutants (Kikuchi *et al.*, 2000), where endoderm formation is abolished due to lack of Nodal signalling. This means that *Mixl1* is a crucial regulator of endoderm formation but also that other factors are likely required for maintenance of endoderm progenitors. In relation to this, overlap of Nanog, Mxtx2, Pou5f3 and *Mixl1* CHIP-seq peaks revealed that they regulate expression of *cxcl12a*, *tbx16*, and *tbx1a* (Nelson *et al.*, 2017). This suggests that *Mixl1*/Smad2/Eomesa may work cooperatively with Nanog, Mxtx2 and Pou5f3 to induce endoderm specification in the embryo.

In zebrafish another Mix family member known as *sebox* was previously shown to act in parallel with *mixl1* but unlike *mixl1*, it does not have the core sequence motif to bind to Smad2 (Germain *et al.*, 2000). Similar to *mixl1*, *sebox* is also a target of Nodal signalling since its

expression was completely abolished in *ndr1* and *ndr2* double mutants (Alexander and Stainier, 1999; Poulain and Lepage, 2002). In relation to this, the phenotypes observed in *MZtdgf1* mutants were partially rescued by mRNA injection of *sebox* or *mixl1* into embryos (Alexander and Stainier, 1999; Poulain and Lepage, 2002). The function of *sebox* appears to be partially redundant with *mixl1* since injection of *sebox* mRNA into *mixl1* mutants partially restored some endodermal markers but inhibition of *sebox* function in *mixl1* mutants resulted in the loss of all endoderm cells (Poulain and Lepage, 2002). This suggests that activities of both *mixl1* and *sebox* are required for correct endoderm specification.

Another notable example of a factor that binds Smad2 during mesendodermal induction is *GATA binding protein 6 (gata6)*. Gata6 is believed to be important in endoderm specification since knockdown of *gata6* decreases expression of *sox32*, the master inducer of endoderm formation in zebrafish but also upstream factors including *sebox* and *tbx16* (Tseng *et al.*, 2011). However, it is also believed to act redundantly in the presence of another member of the GATA family of transcription factors known as *GATA binding protein 5 (gata5)* and that both Gata5 and Gata6 are activated by Orthodenticle homeobox 2b (Otx2b) which is also implicated in endoderm specification (Tseng *et al.*, 2011). Gata5 mutants have reduced number of endoderm cells and also appear to have reduced expression of *sox32* (Reiter, Kikuchi and Stainier, 2001). However, *gata5/gata6* single morphants exhibit minor defects in endoderm numbers compared to *gata5* and *gata6* double morphants where *sox32* was significantly downregulated (Tseng *et al.*, 2011). This suggests that both *gata5* and *gata6* act redundantly in activating *sox32* expression. However, endoderm numbers are not completely lost in the double mutant which likely means that GATA transcription factors in combination with other endoderm regulators are required to induce endoderm formation. Previous experiments have shown that Eomesa acts synergistically with Mixl1 and Gata5 to induce endoderm formation by driving expression of *sox32* (Bjornson *et al.*, 2005). Thus, this incomplete loss of *sox32*-expressing endoderm progenitor cells in *gata5* and/or *gata6* morphants may be due to functional redundancies of different factors including Mixl1 and Eomesa, in the regulation of *sox32* expression.

1.6.4 HMG-containing transcription factor, Sox32 and endoderm formation

Sox32 is a member of the SoxF subgroup of Sox (SRY-related HMG-box) family transcription factors, which are evolutionary conserved and play crucial roles during embryonic development. There are approximately 30 Sox family members which are defined by the presence of the highly conserved HMG-box which binds the minor groove of DNA and induces a kink the DNA helix (Pevny and Lovell-Badge, 1997; Hou, Srivastava and Jauch, 2017),

(Figure 1-6). The bending of the DNA by the HMG-box is proposed to contribute to the assembly of specific transcriptional initiation complexes, and, in turn, to the regulation of gene expression (Hou, Srivastava and Jauch, 2017). However, beyond transcription initiation, Sox factors may also elicit transcriptional repression activities in the early stages of transcription (Zhang and Hou, 2021). Interestingly, Sox family of transcription factors elicit their transcriptional changes by forming complexes with partner transcription factors (Kamachi and Kondoh, 2013). For instance, Sox32 binds to Pou5f3, the closest zebrafish homolog of mammalian Pou5f1 and induces expression of another member of SoxF family member called *sox17* (Zhao *et al.*, 2013) as well as Forkhead box A2 (*foxa2*) which mark all endodermal progenitor cells (Alexander and Stainier, 1999) (Figure 1-7). Maternal-zygotic Pou5f3 mutants (also called *MZspg*) fail to induce endoderm while zygotic Pou5f3 mutants exhibit reduced number of *sox17*- and *foxa2*-expressing cells (Lunde, Belting and Driever, 2004; Reim *et al.*, 2004). In addition, *sox32* overexpression in *MZspg* mutants fail to rescue *sox17* expression while overexpression of Pou5f3 rescues endodermal tissues in *MZspg* mutants (Lunde, Belting and Driever, 2004). Moreover, *sox32* mutants also fail to induce expression of *sox17* and *foxa2* (Alexander and Stainier, 1999; Alexander *et al.*, 1999). These results suggest that interaction between Pou5f3 and Sox32 is critical for endoderm formation.

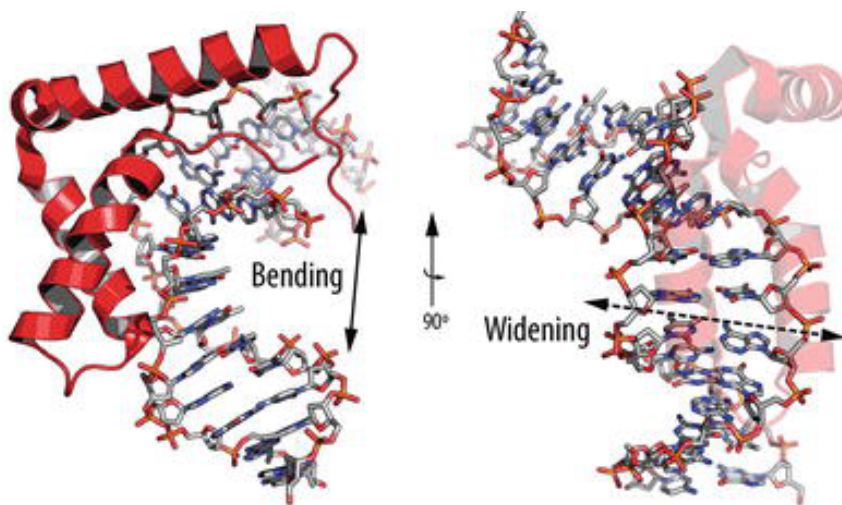


Figure 1-6: Binding of Sox factors to the minor groove results in bending and widening of the DNA double helix. Figure from: (Ángel *et al.*, 2017).

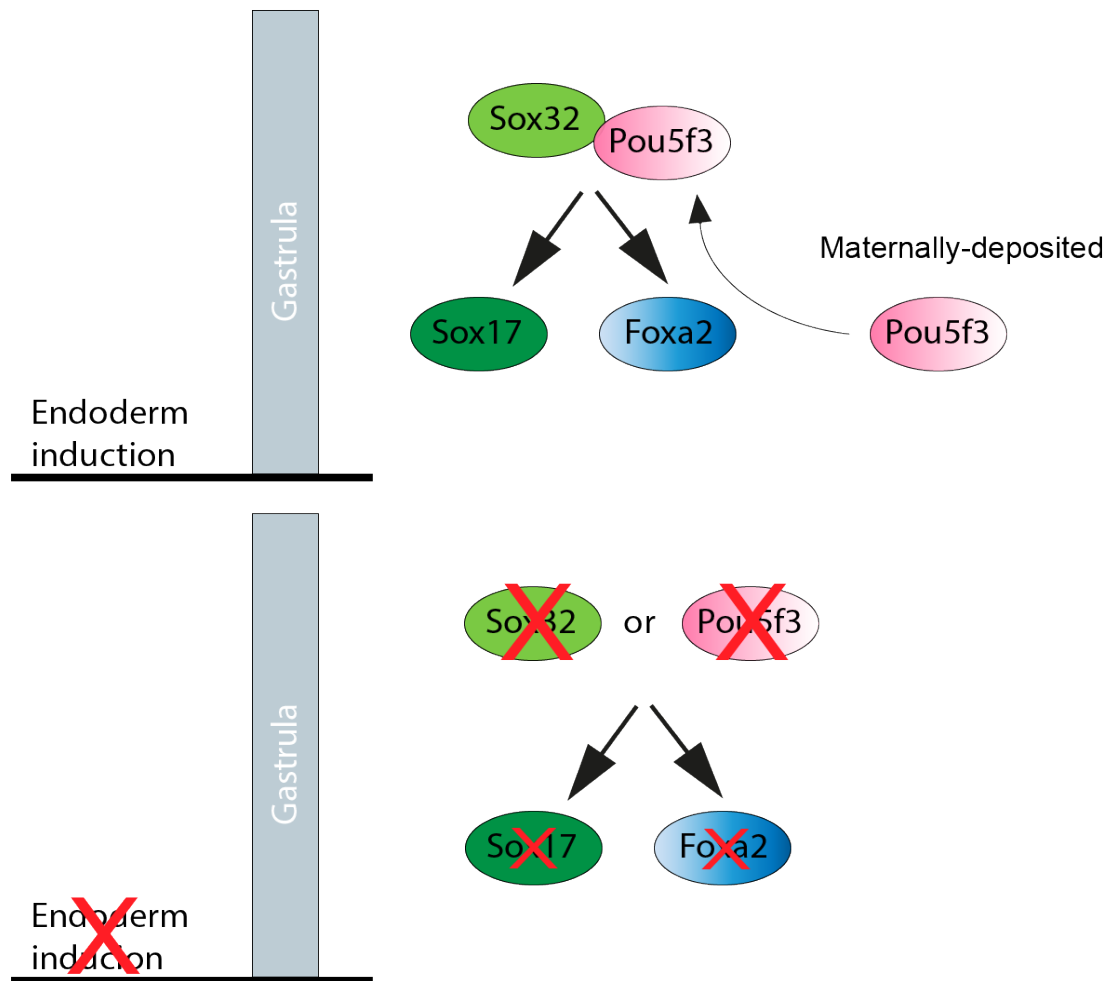


Figure 1-7: Sox32 and Pou5f3 are essential for endoderm specification. During gastrulation, Pou5f3 binds to Sox32 and induces expression of *sox17* and *foxa2* which mark endoderm progenitors. However, loss of either Pou5f3 or Sox32 results in a failure to induce endoderm and a loss of all endoderm-associated genes, including *sox17* and *foxa2*.

It is notable that Sox32 also physically interacts with Nanog, disrupting Nanog-Pou5f3 interactions in the endoderm along the dorsoventral axis and restricting the complex to the ventrolateral mesendoderm (Perez-Camps *et al.*, 2016). However, the relevance of Sox32-Nanog interaction is not yet known. Though loss of *nanog* results in endodermal defects, it is not clear whether this is attributable to loss of *mxt2* expression (Xu *et al.*, 2012), as previously mentioned, or to its combinatorial role with Sox32. In addition, loss of endoderm in *nanog* morphants resemble the phenotypes of *MZspg* mutants, which fail to express *sox17* and maintain endoderm fate (Lunde, Belting and Driever, 2004; Xu *et al.*, 2012) which likely suggests that both Nanog and Pou5f3 participate in the same developmental programme albeit at different stages of development.

Sox32 also plays a role in normal heart development. Under normal conditions, migrating cardiac precursor cells migrate toward the midline between the YSL and the endoderm to form the cardiac tube (Alexander *et al.*, 1999). However, failure to express *sox32* results in the presence of two independent hearts in bilateral positions, a condition known as *cardia bifida* (Alexander *et al.*, 1999). In addition to *cardia bifida*, heart oedema, reduced mobility (Chen *et al.*, 1996) as well as defects in blood circulation can be observed by 24 h.p.f (Nicolas B David *et al.*, 2002), highlighting the link between endoderm formation and proper heart development.

While *sox32* is fundamental for endodermal specification, it appears to be missing in other vertebrates. This is likely due to a small scale duplication event that occurred during evolution of teleost fishes that resulted in duplication of *sox17*, giving rise to two genes; *sox17* and *sox32*, both of which are still retained due to sub-functionalization and are localised on the same chromosome (Voldoire *et al.*, 2017). However, *sox17* which functions downstream of *sox32*, is highly conserved and is a critical endoderm regulator in other vertebrates. Indeed the function of *sox17* in endodermal formation was first discovered in *Xenopus*, where overexpression was shown to induce expression of endoderm markers in the animal cap, which normally contributes to ectoderm fate, while loss of *Sox17* failed to induce endoderm formation (Hudson *et al.*, 1997). Consistent with this, loss of *Sox17* in mice results in defective gut development especially after embryonic day 9.5 and perturbed definitive endoderm which was also accompanied by loss of biliary structures, pancreatic tissue and gut endoderm (Kanai-Azuma *et al.*, 2002). In relation to this, expression of *Sox17* was sufficient to induce the differentiation of human and mouse embryonic stem cells (ESC) into stable, proliferative early endoderm progenitor cells (Séguin *et al.*, 2008; Niakan *et al.*, 2010). While loss of *Sox17* expression in mice and *Xenopus* results in a dramatic reduction in endoderm progenitor cells, in zebrafish, targeted knockdown of *sox17* results in aberrant left/right asymmetry establishment but not a dramatic loss of endoderm (Aamar and Dawid, 2010). However, *sox17* is exclusively expressed in all endoderm progenitor cells during gastrulation and thus, is often used as an early endoderm marker during development.

Sox32 is a potent regulator of endoderm formation, which means overexpression of *sox32*, results in ectopic *sox17* expression in both *sox32* mutants and wild-type embryos in contrast to the normal salt-and-pepper distribution of *sox17*-endodermal cells in un-injected embryos (Kikuchi *et al.*, 2001). On the other hand, *sox32* mutants lack all *sox17* expression (Alexander and Stainier, 1999). Interestingly, endoderm progenitors differentiate into mesoderm derivatives, which suggests a role for *Sox32* in fate decisions between endoderm and mesoderm (Dickmeis *et al.*, 2001). Notably, overexpression of *sox32* in *MZtdgf1* mutants results in ectopic expression of *sox17* (Dickmeis *et al.*, 2001; Kikuchi *et al.*, 2001) which means

that *sox32* does not require Nodal signalling to induce endoderm formation. Similar to zebrafish, physical interaction between Pou5f1, the closest mammalian homolog of zebrafish Pou5f3, and Sox17 in mammals also orchestrate endoderm differentiation (Aksoy *et al.*, 2013). All together, these studies suggest that *sox32* is functionally equivalent to *Sox17* in other species and that *sox32* is absolutely essential for endoderm formation in zebrafish.

1.7 Sox17 is a marker of DFCs and haematopoietic lineages

In zebrafish, *sox17* marks the endoderm during gastrulation, but prior to gastrulation, its expression is also observed in dorsally located marginal cells, which correspond to the non-involuting endocytic cells that later form dorsal forerunner cells (DFCs) (Cooper and D'Amico, 1996), which localise outside the blastoderm at the dorsal margin (Melby, Warga and Kimmel, 1996). *Sox17* expression is maintained in DFCs throughout gastrulation where they reside at the leading edge and migrate ahead of the dorsal blastoderm, hence the name, until the end of gastrulation where they are engulfed by the blastoderm (Kupffer, 1868; Cooper and D'Amico, 1996; Warga and Kane, 2018). At tailbud stage (~10 h.p.f.), approximately 30 DFCs form the lining of the Kupffer's Vesicle (KV), a transiently ciliated fluid-filled sphere with a critical role in laterality determination (Kupffer, 1868; Essner *et al.*, 2005). Indeed fate mapping experiments demonstrated that labelled DFCs eventually transform into KV which eventually disperse and contribute to the tail notochord and muscle (Melby, Warga and Kimmel, 1996), and laser ablation of DFCs or loss of *sox17* disrupt KV formation and result in abnormal left/right patterning defects (Essner *et al.*, 2005; Amar and Dawid, 2010). This is also relevant in mice where embryos lacking *Sox17* exhibit aberrant left/right asymmetry (Viotti *et al.*, 2012). Thus, *sox17* in zebrafish, similar to other species, is required for correct left/right patterning and KV formation.

Vital organs including the heart, brain and gastrointestinal tract develop functional asymmetries along the left/right axis and perturbation of this often leads to congenital diseases (Kosaki and Casey, 1998). There is evidence that ciliated cells play a crucial role in establishing left/right patterning. In mouse, monocilia protruding from cells in pit-like structure known as the ventral node has been proposed to generate left-forward flow that results in left/right asymmetry establishment in the embryo (Nonaka *et al.*, 1998). Cilia-driven flow has also been described in rabbit, chick, medaka fish and *Xenopus* (Selleck and Stern, 1991; Okada *et al.*, 2005; Schweickert *et al.*, 2007) suggesting that ciliated cells play a conserved role in left/right patterning. Thus, left/right patterning in all species is dependent on the function of a laterality organ. In zebrafish, KV is analogous to the mouse ventral node pit and uses its cilia to exhibit left/right asymmetry establishment by creating a directional, counter-clockwise

fluid flow within the lumen of the KV (Kramer-Zucker *et al.*, 2005; Okada *et al.*, 2005). This fluid flow is critical for KV function and results in asymmetric expression of genes on the embryo's left and right side. In relation to this, knockdown of *sox17* in the DFCs impairs cilia formation and as a result, impairs laterality establishment in the embryo (Aamar and Dawid, 2010). For this reason, KV was later categorized as the "organ of asymmetry" in zebrafish (Essner *et al.*, 2005).

In the absence of Nodal signalling, endoderm and DFCs fail to develop and therefore there are many similarities between the two lineages since they are both derived from the dorsal edge of the blastoderm and require high Nodal-signalling to differentiate (Alexander and Stainier, 1999; Warga and Nüsslein-Volhard, 1999). Though Nodal signalling is required to induce DFC formation, upstream endoderm regulators *mixl1* and *gata5* are not expressed nor are they required to induce DFC formation. Indeed *gata5* and *mixl1* mutants exhibit reduced endoderm progenitor cells while the number of DFCs appear unaffected (Kikuchi *et al.*, 2000; Reiter, Kikuchi and Stainier, 2001). However, other Nodal-dependent genes are involved in KV formation. For instance, *Tbx16* and *Tbxta* act cooperatively to aid in KV morphogenesis and KV fails to form in the absence of both factors (Amack, Wang and Yost, 2007).

Aside from *sox17*, a number of additional factors are also believed to be critical for KV formation and function, which include vestigial like family member, *Vgll4l* and TEA domain DNA-binding of transcription factors (TEAD) 1a and 3a. Knockdown of *vgll4l* or *tead* expression results in significant downregulation of *ndr1*, *sox32*, *sox17*, *tbxta*, *tbx16* and other KV-expressed genes but also results in laterality defects and reduction in KV size (Fillatre *et al.*, 2019), suggesting that *Vgll4l* and *Tead* factors are essential for KV formation.

However, *sox17* also appears to be implicated in the regulation of haematopoiesis. Indeed, mammalian *Sox17* is highly expressed in foetal and neonatal but not adult haemopoietic stem cells (HSCs) and *Sox17*-deficient mice exhibit severe foetal hemopoietic defects including a loss of definitive haematopoiesis (Kim, Saunders and Morrison, 2007). In zebrafish, *sox17* marks the endoderm during the gastrulation however, beyond gastrulation, *sox17* is no longer expressed in the endoderm but in mesodermal lineages. At 12 h.p.f. *sox17* was observed in the anterior and posterior lateral plate mesoderm (LPM) which contributes to endothelial and blood cells and heart, while at 18 h.p.f. expression can be detected in the intermediate cell mass of mesoderm (ICM), which gives rise to haematopoietic lineages (Chung *et al.*, 2011). By 26 h.p.f., expression of *sox17* was detected in the ICM, neural tube and duct of Curvier (Chung *et al.*, 2011). Knockdown of *sox17* results in a significant reduction of erythroid cells

and reduction of ICM markers, which includes *GATA-binding protein 1 (gata1a)* (Chung *et al.*, 2011). Thus, the role of *sox17* in haematopoiesis appears to be conserved in vertebrates.

1.8 Beyond Sox32 and Sox17: the role of Forkhead factors

Forkhead factors, are also expressed in endodermal progenitors during gastrulation (Weigel *et al.*, 1989). They belong to the winged helix family of transcription factors which are characterised by a conserved 100 -amino acid domain called the “forkhead box” (Golson and Kaestner, 2016). Forkhead genes were first named after the *Drosophila melanogaster* gene *fkh* (fork head), which is necessary for normal gut development, and its absence results in developmental defects with a forked head appearance in adult flies (Weigel *et al.*, 1989; Golson and Kaestner, 2016). Soon after the discovery of *fkh*, Forkhead factors were also identified in multiple organisms ranging from yeast to humans (Golson and Kaestner, 2016).

There are currently nine members of the Forkhead family of transcription factors in zebrafish, and three of these factors have been found to play an important role during early endoderm development (Strahle *et al.*, 1993; Dirksen and Jamrich, 1995; Odenthal and Nüsslein-Volhard, 1998; Ober, Field and Stainier, 2003). The first factor is Forkhead A2 (*Foxa2*). In zebrafish, expression of *foxa2* is observed in the scattered endodermal cells around the marginal domain during both late-blastula and early gastrula stages (Kikuchi *et al.*, 2004). *Foxa2* mutants exhibit mild phenotypes and are characterised by a failure of floor plate differentiation (Norton *et al.*, 2005). Although *foxa2* mutants do not show severe endodermal defects, *foxa2* is likely involved in liver and pancreas formation in zebrafish (Shin *et al.*, 2008). In mouse, *Foxa2* expression can be detected in the primitive streak by embryonic day (E) 6.5, and by E7.5, its expression is observed in the anterior mesendoderm, node, notochord, definitive endoderm and floor plate (Ang *et al.*, 1993; Monaghan *et al.*, 1993). Targeted deletion of *Foxa2* abolishes node formation (left/right organiser), which in turn results in dorsal-ventral patterning defects of the neural tube and a complete lack of notochord. In addition to this, the primitive streak fails to elongate while the definitive endoderm, which is initially specified, fails to form the gut tube (Ang and Rossant, 1994; Weinstein *et al.*, 1994). Interestingly, in mice, *Foxa2* has been described as a “pioneer factor”, which has the ability to bind condensed chromatin and has been shown to be important for proper development of endoderm-derived organs such as the liver, pancreas, lung, and prostate. As a result, *Foxa2* is considered a master regulator of endoderm formation (Bahar Halpern, Vana and Walker, 2014).

The second Forkhead factor is Forkhead A1 (*Foxa1*). In zebrafish, *foxa1* becomes expressed in endodermal progenitors following gastrulation (Odenthal and Nüsslein-Volhard, 1998). In contrast to *foxa2*, single knockdown of *foxa1* had no effect on axial structures (Dal-Pra, Thisse and Thisse, 2011). In mouse, *Foxa1* follows *Foxa2* expression in the definitive endoderm and anterior mesendoderm but after a short temporal delay. *Foxa1* is expressed in the primitive streak by E7.0 and then is later expressed in the notochord, neural plate (Odenthal and Nüsslein-Volhard, 1998) and neural tube (Monaghan *et al.*, 1993). As opposed to *Foxa2*, *Foxa1* null mice contain a normal notochord and intervertebral discs but die shortly after birth as a result of hypoglycaemia and dehydration (Ang and Rossant, 1994). *Foxa1* together with *Foxa2* are required for endoderm lineage establishment and morphogenesis of endoderm-derived organs (Bahar Halpern, Vana and Walker, 2014).

The third factor is Forkhead A3 (*Foxa3*). In zebrafish, *Foxa3* expression begins at dome stage (4.3 h.p.f.) and upon gastrulation, *foxa3* appears in all cells of the axis including the prechordal plate region and outside the axis, *foxa3* is expressed in more lateral cells that were later identified to be endodermal progenitor cells (Warga and Nüsslein-Volhard, 1999). In addition, the prechordal plate is properly induced in *foxa3* zebrafish morphants but fails to form the hatching gland (Dal-Pra, Thisse and Thisse, 2011). In mouse, *Foxa3* expression is first detected at E8.5 and is more restricted than *Foxa1* and *Foxa2* since it is specific to the foregut endoderm (Golson and Kaestner, 2016). While *Foxa1*- and *Foxa2*-null mice exhibit perinatal and embryonic defects, *Foxa3*-deficient mice are viable with no obvious abnormalities (Shen *et al.*, 2001).

All three Forkhead factors continue to be expressed in endoderm progenitor cells until organogenesis stage, where they exhibit partially overlapping but also distinct patterns in a variety of tissues (Friedman and Kaestner, 2006). In zebrafish, *foxa2* expression was observed in the pharyngeal endoderm and notochord at 26 h.p.f., while *foxa1* is expressed in floor plate and gut tube and *foxa3* is expressed in the hatching gland, posterior notochord and gut (Odenthal and Nüsslein-Volhard, 1998). In mouse, *Foxa2* appears in a endoderm-derived tissues including the pancreas, liver, thyroid, prostate and lung (Friedman and Kaestner, 2006). On the other hand, *Foxa1* is more widely expressed in a number of tissues compared to *Foxa2*, including the respiratory system, brain and gastrointestinal tract and is also expressed in the renal pelvis, ureters and male reproductive organs whereas *Foxa2* is not (Friedman and Kaestner, 2006; Golson and Kaestner, 2016). Also, *Foxa3* expression is maintained in the liver, as well as the pancreas and intestine at later stages of development and adulthood (Monaghan *et al.*, 1993).

The role of Forkhead factors have been genetically analysed in a number of species including mouse and zebrafish. In zebrafish, in the absence of *foxa2* and *foxa3*, the pancreas, liver, and part of the intestinal bulb and intestine are correctly induced but fail to differentiate by 96 h.p.f, while inactivation of *foxa2* or *foxa3* single mutants only affect floor plate or hatching gland, respectively, but not the endoderm (Dal-Pra, Thisse and Thisse, 2011). Similarly, midgut and hindgut formation is perturbed in *Foxa2*-deficient mice (Ang *et al.*, 1993; Weinstein *et al.*, 1994; Dufort *et al.*, 1998). In addition, *Foxa1* and *Foxa3* mutants do not show early endoderm phenotype, likely due to functional redundancies between the Forkhead factors and due to overlap of expression patterns (Kaestner, Hiemisch and Schütz, 1998). Thus, Forkhead factors play essential roles during endoderm formation during development and this role appears to be conserved across vertebrate species.

1.9 From endoderm formation to organogenesis

During organogenesis, the endoderm forms the epithelial lining of the respiratory and digestive system and contributes to the development of a number of associated organs such as the thyroid, thymus, liver, biliary system, and pancreas. The endoderm is also essential for the morphogenesis and patterning of neighbouring structures, such as the heart and the head, as discussed previously (Peyri eras, Str ahle and Rosa, 1998; David and Rosa, 2001; Nicolas B. David *et al.*, 2002). These endoderm-derived organs have vital functions including: digestion, blood clotting, detoxification, gas exchange and glucose homeostasis (Zorn and Wells, 2009). Perturbation of these organs are the underlying cause of many life-threatening diseases. However, despite the physiological importance of these organs, our knowledge of the factors involved in their morphogenesis is currently limited. Understanding the factors involved in endoderm organ formation will be invaluable for understanding the underlying cause of many congenital diseases but also for future transplantation-based therapies (Spence and Wells, 2007).

During gastrulation in zebrafish, endoderm progenitor cells appear morphologically distinct from the surrounding mesoderm cells: they eventually flatten and obtain filopodial processes (Warga and Nusslein-Volhard, 1999). During this time, the flattened endodermal cells converge towards the dorsal side of the embryo and distribute along the anterior-posterior (A-P) axis. Towards the end of gastrulation and early somitogenesis (~14 h.p.f.), the flattened endoderm cells form a sparse but regular monolayer directly overlaying the yolk cell (Aronson, Stapleton and Krasinski, 2014). By 20 h.p.f., the endoderm migrates towards the midline to form a multicellular gut rod, which later cavitates to produce the gut tube that is surrounded by mesoderm. During this period, the gut tube becomes regionalised along the A-P axis into

foregut, midgut and hindgut (Dalgin and Prince, 2021). The foregut gives rise to a number of organs including the thyroid, liver, and pancreas, while the midgut and hindgut give rise to the small and large intestine respectively. The A-P endoderm patterning is believed to be dependent on a combination of signalling factors, mainly wingless/ integrated (Wnt), fibroblast growth factor (Fgf) and bone morphogenetic protein (BMP) at the posterior end of the endoderm (Zorn and Wells, 2009). For instance, Wnt signalling inhibits anterior fate and must be excluded to maintain foregut progenitors and for subsequent liver and pancreas development but is necessary and sufficient to promote hindgut development in zebrafish (McLin, Rankin and Zorn, 2007; Goessling *et al.*, 2008). Moreover, experiments on zebrafish demonstrated that BMP signalling is required to promote posterior endoderm development (Tiso *et al.*, 2002; Wills *et al.*, 2008). In relation to this, prior to organogenesis retinoic acid (RA) plays an essential role in the A-P patterning of the endoderm and regulates respiratory versus pharyngeal fate in the anterior endoderm, and midgut versus hindgut in the posterior endoderm (Kelly and Drysdale, 2015; Rankin *et al.*, 2018). The endoderm at this stage can be identified by the presence of a number of transcription factors including Haematopoietically expressed homeobox (Hhex), Pancreatic and duodenal homeobox 1 (Pdx1) and Caudal type homeobox (Cdx) factors. These factors are important for regional identity: Hhex marks the presumptive foregut, Pdx1 marks the posterior foregut/anterior midgut and Cdx factors mark the hindgut (Zorn and Wells, 2009).

During development, the foregut forms first at early somite stage, and is shortly followed by the formation of the hindgut, which forms at around 26 h.p.f. As development proceeds, the midgut also forms within the developing embryo which completes the development of the gut tube by 34 h.p.f. (Wallace and Pack, 2003). Interestingly, some of the factors involved in regional identity including Hhex, Pdx1, and Foxa2 play a role in later organ formation. For instance, in zebrafish Foxa2 is found in the anterior endoderm including pharyngeal endoderm and is implicated in liver and pancreas formation, Hhex is essential for liver and pancreas development while Pdx1 is essential for pancreas development since loss of *pdx1* results in severe pancreatic hypoplasia (Shin *et al.*, 2008; Kimmel *et al.*, 2011; Gao *et al.*, 2019). Thus, these studies highlight the importance of the different signalling pathways in cell fate decisions during endoderm organogenesis.

1.10 Identifying developmentally important *cis*-regulatory modules (CRMs)

Genetic screening has revealed a number of different transcription factors required for the differentiation of different cell lineages. However, a detailed analysis of the regulatory programme involved in the expression of developmentally important genes in distinct cell types remains challenging.

Transcriptional regulation during development is fundamental for the function and emergence of different cell types during development. Genes are regulated by a variety of different transcription factors and chromatin remodelling proteins which bind to specific DNA regions of open chromatin known as *cis*-regulatory modules (CRMs) (Suryamohan and Halfon, 2015), and via the involvement of cofactors and coactivators, recruit RNA polymerase II (Andersson, 2015). CRMs are short sequences of DNA with transcriptional regulatory activity, and these include but are not limited to enhancers, silencers, insulators, and promoters (Figure 1-8). Genes transcribed by RNA polymerase II typically depend on two distinct types of CRMs: a) promoter, which comprises a core promoter and nearby proximal regulatory regions, and b) distal regulatory regions which can be enhancers, silencers or insulators. These CRMs contain recognition sites for trans-acting transcription factors that either enhance or repress gene expression (Maston, Evans and Green, 2006).

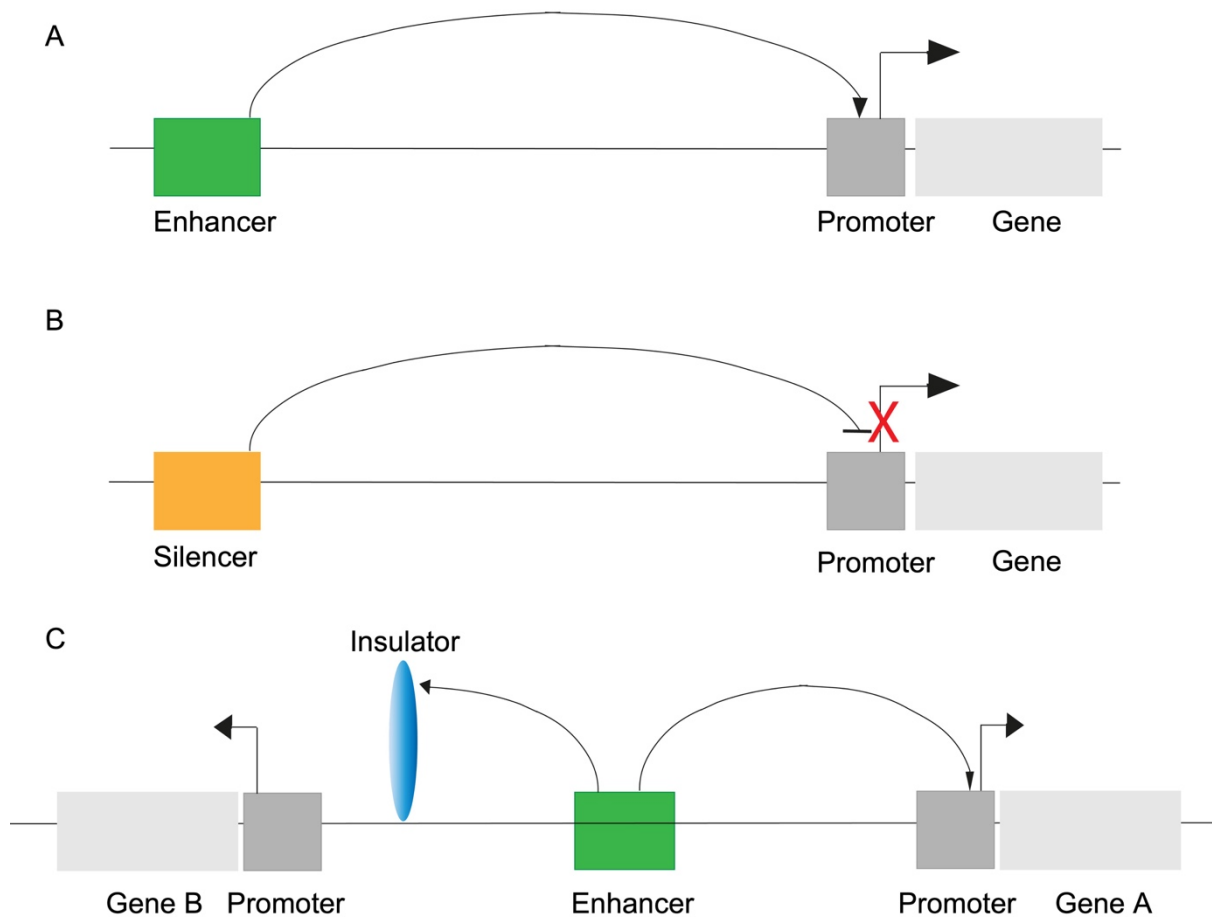


Figure 1-8: CRMs regulate transcription of target genes. The image illustrates three categories of regulatory elements, a) enhancer, b) silencer, c) insulator. An enhancer affects positively the transcription of the target gene; the silencer silences transcription of a gene, while an insulator constitutes a barrier between an enhancer/ or silencer to target gene, gene A. The promoter is located upstream of the gene coding region, and this is where the process of transcription of a gene begins.

1.11 Classification of CRMs

1.11.1 Core promoter

The core promoter is a specialised DNA sequence of 50-100 bp that contains the transcription start site and serves as a docking station for the basic transcriptional machinery, including RNA polymerase II and general transcription factors (GTFs) (Suryamohan and Halfon, 2015). In eukaryotes, the core promoter usually contains a TATA box, which is a binding site for the TATA-binding protein subunit of the general transcription factor TFIID. Alongside the TATA box, the core promoter comprises numerous other elements: Initiator element (Inr), downstream promoter element (DPE), downstream core element (DCE), TFIIB-Recognition Element (BRE) and Motif Ten Element (MTE) (Maston, Evans and Green, 2006). Aside from BRE, which is only recognised by TFIIB, the components mentioned above are TFIID-

interaction sites: DPE is recognised by TAF6 and TAF9, Inr is recognised by TAF1 and TAF2 and DCE is recognised by TAF1 (Lim *et al.*, 2004).

1.11.2 Enhancers

Enhancers were first discovered as regions within the simian virus 40 (SV40) genome which could markedly enhance transcription of a heterologous gene containing a promoter (Tong Ihn Lee and Young, 2000). Transcriptional enhancers were soon also discovered in immunoglobulin genes and other cellular genes (Schaffner, 2015). Studies on enhancers within the last 20 years have shown that they regulate transcription of genes in a spatial-or-temporal manner, and they typically function independent of orientation, distance or placement (5'/3') relative to the target gene (Maston, Evans and Green, 2006). Enhancers are typically composed of dense clusters of short transcription factor binding sites (TFBSs) that work collectively to enhance transcription and as a result, enhancer activity is produced by a combination of these TFB motifs also known as modules (Maston, Evans and Green, 2006). Due to their modular nature, a single gene can be regulated by multiple enhancer elements at different times, in different tissues, or in response to different stimuli (Koshikawa, 2015; Krijger and De Laat, 2016). Unlike promoters, enhancers typically regulate genes over long distances from their target genes and in *trans* (McCord *et al.*, 2011; Li *et al.*, 2012; Dean, Larson and Sartorelli, 2021). Inactive enhancers exist within closed chromatin conformation resulting in no regulatory activity and inaccessible sequence in most cells (Bozek and Gompel, 2020; Panigrahi and O'Malley, 2021). On the other hand, active enhance regions are characterised by open chromatin and high level of transcription factor binding .

1.11.3 Silencers

In contrast to enhancers, silencers are elements that silence or decrease transcription of a target gene (Segert, Gisselbrecht and Bulyk, 2021). While enhancers have been studied extensively, only a few silencers have been identified and they remain poorly understood. Silencers tend to have similar properties to enhancers. For instance, silencers are also modular and they function independently of distance and orientation, although some position-dependent silencers have been encountered (Maston, Evans and Green, 2006; Segert, Gisselbrecht and Bulyk, 2021). They can be situated proximal to their target genes, or as part of distal enhancers or even as an independent distal regulatory module. Like enhancers, silencers also contain TFBSs that recruit regulatory factors called transcriptional repressors to the promoter and thus, silencers must be accessible to these repressive-DNA binding proteins in order to silence genes (Maston, Evans and Green, 2006). However, relative to silencers, enhancers contain motifs from a more diverse collection of factors (Friedman *et al.*,

2021). A number of models have been proposed to explain the function of repressors. In some cases, repressors function by blocking the binding of activating factors called activators at nearby enhancer elements while in other cases, repressors have been found to compete for the same site as enhancers (Li *et al.*, 2004; Harris, Mostecky and Rothman, 2005). Alternatively, repressors silence genes by modifying the chromatin landscape either through the recruitment of histone-modifying factors or chromatin-stabilizing factors (Srinivasan and Atchison, 2004). Finally, repressors may block gene expression by inhibiting transcription-preinitiation complex assembly (Maston, Evans and Green, 2006).

1.11.4 Insulators

Insulators are another class of regulatory elements which block the action of enhancers or silencers on target genes. Insulators have two main properties; firstly they are able to block the action of an enhancer on a target promoter when inserted between these regulatory elements (Kyrchanova *et al.*, 2008). This may be explained by the observed interaction between insulators and activating factors which were found to block enhancer-activity (Defossez *et al.*, 2005). Moreover, an insulator may act as a barrier to prevent the advance of adjacent condensed chromatin that otherwise silences gene expression and also protects gene expression from the negative or positive chromatin effects surrounding the gene (Brasnet and Vaury, 2005). This heterochromatin-barrier activity is likely explained by the recruitment of histone-modifying or gene-activating factors (Maston, Evans and Green, 2006). Insulators typically function in a position-dependent and orientation-independent manner. The number of insulator elements in the human genome is unknown, but a known factor that mediates this insulator activity in vertebrates is known as CCCTC-binding factor (CTCF), which plays a crucial role in gene regulation and three-dimensional (3D) chromatin organisation (Maston, Evans and Green, 2006). In humans and mammals, CTCF binding sites are enriched at topologically associated domains (TADs) which contribute to gene regulation by blocking chromatin interactions of regulatory regions, such as enhancers, with their target genes (Downen *et al.*, 2014; Lupiáñez *et al.*, 2015; Krefting, Andrade-Navarro and Ibn-Salem, 2018; Despang *et al.*, 2019). Disruption of TADs may result in altered gene expression and is associated with a number of genetic diseases and cancers (Krefting, Andrade-Navarro and Ibn-Salem, 2018). In contrast to humans, CTCF binding is not enriched at TADs in developing zebrafish embryos, (Pérez-Rico, Barillot and Shkumatava, 2020), and thus CTCF is not essential for embryogenesis in zebrafish.

1.12 Histone variants and chromatin accessibility

Gene regulation is controlled by transcription factors, which recognise and bind to a specific set of DNA sequences known as “binding motifs” and switch genes on and off in a spatial and temporal manner (Huang *et al.*, 2018). However, in eukaryotes, genomic DNA is tightly coiled by an octamer of the four core histone proteins (H2A, H2B, H3 and H4) to form an array of nucleosomes, which are the basic unit of chromatin (Chereji and Morozov, 2015). The formation of nucleosomes helps to package DNA inside the nucleus, but also plays an important role in transcription since nucleosomes occlude the binding sites of transcription factors (Bai and Morozov, 2010; Voong *et al.*, 2017). Although nucleosomes are DNA-protein complexes, they are not static but dynamically regulated in different cell types and different environments, creating an accessibility continuum which ranges from closed chromatin to highly dynamic, accessible or open chromatin (Klemm, Shipony and Greenleaf, 2019). With the exception of a few transcription factor known as pioneer factors, which are capable of binding to target sites on nucleosomal DNA often leading to the local opening of chromatin, the majority of transcription factors bind to accessible chromatin almost exclusively (Thurman *et al.*, 2012; Iwafuchi-Doi, 2019). Thus, nucleosome positioning at promoters and other regulatory regions can negatively regulate gene transcription by preventing transcription factor binding to genomic DNA (Shlyueva, Stampfel and Stark, 2014), (Figure 1-9).

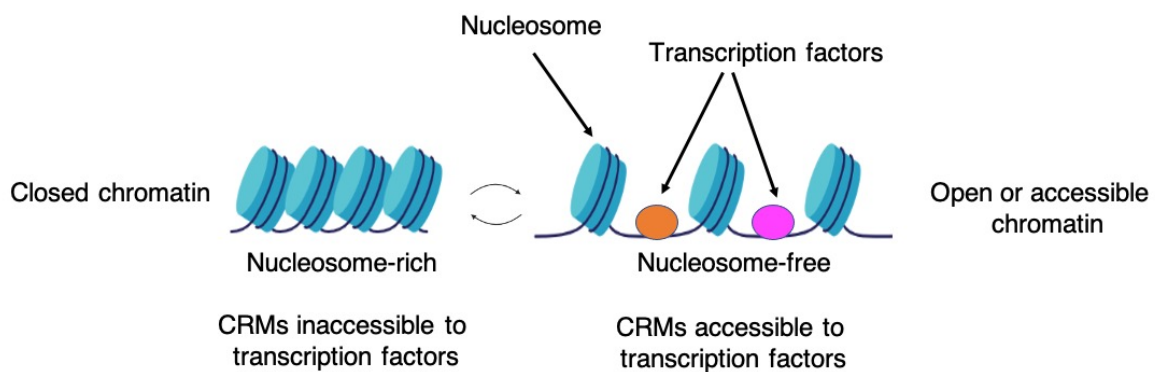


Figure 1-9: A simple schematic of chromatin positioning. In the presence of nucleosomes, transcription factor binding sites are not accessible, while nucleosome organisation enables transcription factors to bind DNA and induce gene expression. Thus, nucleosome positioning regulates DNA accessibility. Image adapted from BioRender.com.

In multicellular animals, CRMs can be very difficult to study since they do not always regulate expression of nearby genes and can be spread over a large distance away from the transcriptional start site of the target gene (Narlikar and Ovcharenko, 2009). Moreover, a single gene can be regulated by multiple CRMs (Nelson and Wardle, 2013) and that deletion of a single CRM can have no observable effect on the regulation of gene expression (Xiong, Kang and Raulet, 2002; Ghiasvand *et al.*, 2011) while in other cases, deletion of a single CRM

may underlie many developmental defects. Also, a single CRM can regulate expression of multiple genes (Nelson and Wardle, 2013), which further highlights the complexity of CRMs in a vertebrate embryo.

The core promoter is believed to contain the information required to initiate transcription by the RNA polymerase II machinery (Butler and Kadonaga, 2002). However, many core promoters are not sufficient to drive gene expression alone without input from other non-coding regulatory regions. For instance, enhancers are believed to physically interact with the target promoter to upregulate gene expression. On the other hand, the interactions between silencers and promoters results in a decrease in gene expression. Enhancers and silencers are located in nucleosome-free regions and comprise binding sites for transcription factors which aid in controlling expression of genes in a spatial and temporal specific manner (Narlikar and Ovcharenko, 2009; Spitz and Furlong, 2012). Despite their importance in gene regulation, silencers are much less characterised than enhancers (Cai *et al.*, 2021). Enhancers will be the main focus of this report.

Historically, identification of enhancers has proved to be challenging for many reasons. Firstly, enhancers can be spread across the 98% of the human genome that does not encode proteins, resulting in a large search space (~98% of 3×10^9) (Pennacchio *et al.*, 2013). Secondly, enhancers are known to be *cis*-acting, they can activate transcription irrespective of their orientation and location relative to the target gene and thus, they can also be located many hundred kb upstream, downstream, or internal to their target gene (Nelson and Wardle, 2013; Erokhin *et al.*, 2015; Suryamohan and Halfon, 2015). Another fundamental property of enhancers is that they do not always regulate expression of genes most proximal to them, and in some cases, they may also regulate expression of genes located on different chromosomes (Gorkin, Leung and Ren, 2014; Ibragimov, Bylino and Shidlovskii, 2020). Furthermore, activity of enhancers can be restricted to a particular cell-type, a developmental time point, or to specific physiological, pathological or environmental conditions (Pennacchio *et al.*, 2013). Finally, unlike promoters which are proximal to protein-coding genes, enhancers do not contain any discernible sequences, making it difficult for them to be identified by computational predictions (Abdulghani, Jain and Tuteja, 2019). Thus, these properties hinder the discovery and functional annotations of enhancers in the genome.

Despite extensive studies of enhancers, there is limited knowledge on their mechanism of action and how they spatiotemporally control transcription. However, activity of enhancers is believed to be influenced by high-order chromatin organisation or chromatin looping, which brings enhancers in close proximity to target gene promoters (De Laat and Duboule, 2013)

(Figure 1-10). This chromatin organisation is believed to facilitate chromatin accessibility of enhancers, which upon binding to activators, promotes gene expression within a given cell-type. Thus, changes to the chromatin landscape is essential for enhancer activity.

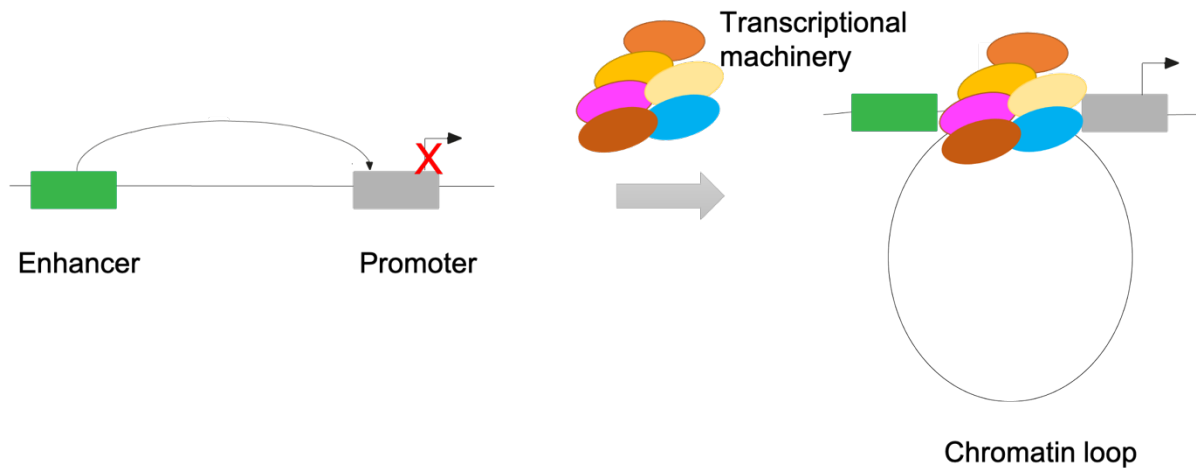


Figure 1-10: Long-range enhancer-promoter interactions. An enhancer is able to activate transcription from a target gene promoter through direct interaction over a long distance by creating a chromatin loop, which may span distances up to 100 kilobases of DNA. This enables the transcriptional machinery to be recruited to the promoter which in turns results in the transcription of target genes.

1.13 Zygotic genome activation: earliest event in chromatin reorganisation

During early development, the genome is known to undergo intense reorganisation including large-scale changes to the chromatin and histone modifications, to enable growth of a new organism. Following fertilisation, the differentiated germ cells, the egg, and the sperm must fuse together to form a zygote. The zygotic genome is initially transcriptionally quiescent which enables it to be reprogrammed to a totipotent state that can later be used to drive the differentiation of all diverse cell types in the body (Schulz and Harrison, 2019). Initially, the development of the embryo relies exclusively on maternally deposited RNA and protein from the egg cytoplasm (Lee, Bonneau and Giraldez, 2014). Subsequently, modifications to the chromatin landscape results in zygotic genome activation (ZGA) and degradation of maternal transcripts in a process known as maternal-to-zygotic transition (MZT) (Tadros and Lipshitz, 2009; Lee, Bonneau and Giraldez, 2014). Thus, following this stage, the zygote relies on its own transcription. Though the timing of this transition varies between organisms, MZT is a highly conserved process across animal models. In zebrafish activation begins at 64-cell stage (2h.p.f.) with initiation of miR-430 gene cluster. MicroRNAs from this cluster in turn result in activation of some zygotic genes and also play a key role in degradation of maternally deposited transcripts during MZT (Giraldez *et al.*, 2006; Heyn *et al.*, 2014; Hadzhiev *et al.*, 2019). However, ZGA is not a single event but a gradual wave of events with different genes

transcribed at various stages of development. Bulk transcription of ZGA begins at 1 k-cell stage (3 h.p.f.) also called mid-blastula transition (MBT) which is characterised by slowing of the cell cycle, divisions of neighbouring cells becoming asynchronous and cells becoming motile (Kane and Kimmel, 1993). MBT is subsequently followed by onset of gastrulation, where cells in the embryo start to adapt different cell fates, either endoderm, mesoderm and ectoderm, through a combination of cell ingression, migration and invagination (Tadros and Lipshitz, 2009).

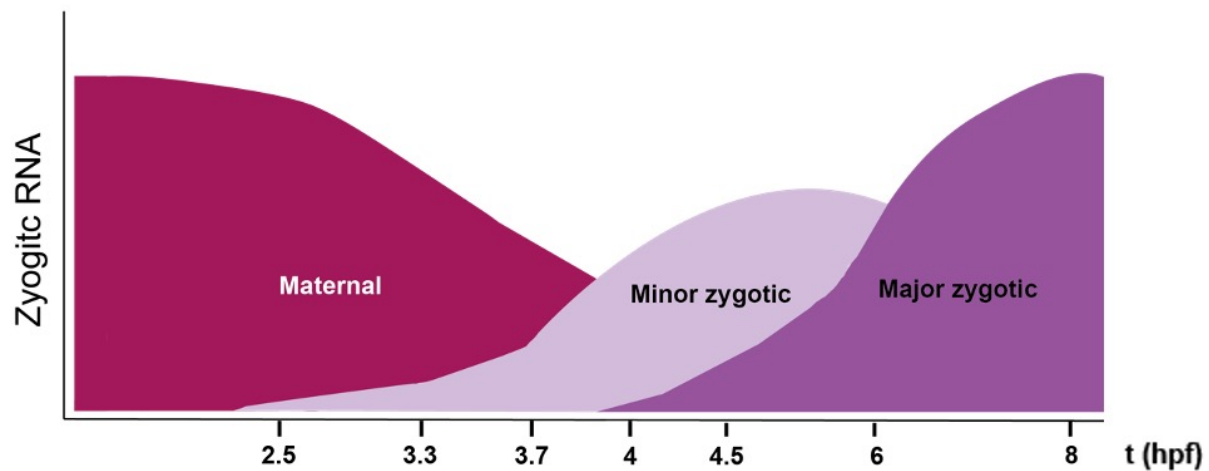


Figure 1-11: Maternal-to-zygotic transition in zebrafish. Red area represents the maternal transcriptional profile, violet represents the minor wave of zygotic transcription and purple represents the major wave of zygotic transcription.

The interplay between the transcriptional machinery and reorganization of nucleosome positioning likely regulates the timing of ZGA. Each nucleosome is composed of DNA wrapped around proteins called histones, and post-translational modification to these histones is key for chromatin to transition from a transcriptionally quiescent state to an activated state which is capable of expressing zygotic genes (Schulz and Harrison, 2019), (Figure 1-12). This chromatin reorganisation enables transcription factors to bind DNA, which is essential for regulation of gene expression and thus, chromatin landscape serves as a proxy for gene expression. As mentioned previously, loss of maternal factors including Pou5f3 and Nanog result in wide-spread decrease in chromatin accessibility and failure to induce expression of zygotic genes (Xu *et al.*, 2012; Lee *et al.*, 2013; Leichsenring *et al.*, 2013; Gagnon, Obbad and Schier, 2018; Liu *et al.*, 2018; Pálffy *et al.*, 2020). Thus, gene expression is influenced by chromatin accessibility and provides the basis for understanding cell-type specific gene regulation events during other stages of zebrafish development.

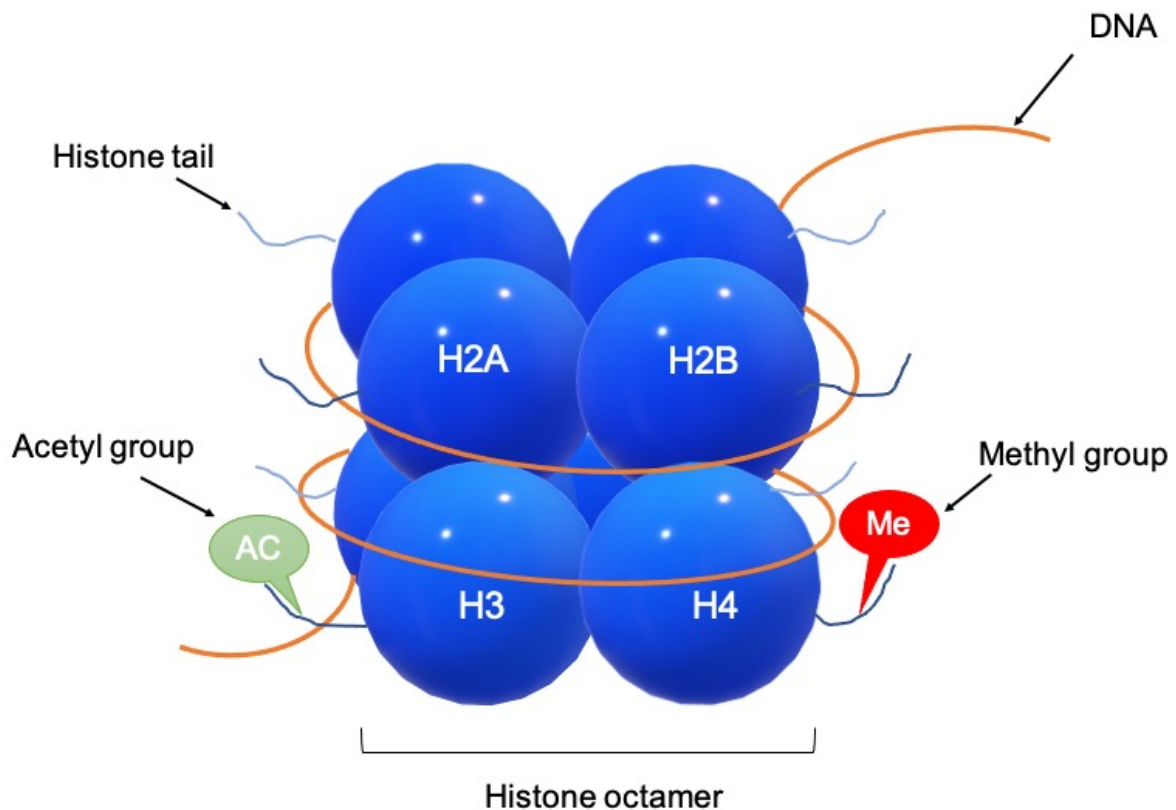


Figure 1-12: Schematic of a nucleosome core particle with four canonical histones (H3, H4, H2A and H2B). The covalent post-translational modifications of the histone tails, for example, methylation [ME] and acetylation [AC] are indicated.

1.14 Challenges in identifying enhancers

Identification of putative enhancers and their location within the genome has been facilitated by a number of different procedures. Initially, this was first aided by comparative genomics, in which genomic sequences from different vertebrate organisms were used to identify conserved non-coding regions of the genome (Nelson and Wardle, 2013; Pennacchio *et al.*, 2013). Comparative testing of humans and mouse loci revealed a number of highly conserved non-coding regions which often harbour enhancer activity (Pennacchio *et al.*, 2006; Visel *et al.*, 2008). In relation to this availability of the zebrafish genome has also enabled the identification of developmentally conserved enhancers across multiple vertebrate genomes (Howe *et al.*, 2013). However, there are drawbacks to using conservation studies alone in identifying enhancers. For example, many conserved genomic regions show no detectable enhancer activity and often do not have conserved functions but also that functionally important enhancers do not always have conserved sequences across species (Nelson and Wardle, 2013; Pennacchio *et al.*, 2013). However, the most challenging aspect appears to be that a large fraction of enhancers are not conserved across species, which again hinders this approach.

In more recent years, the emergence of next-generation sequencing resulted in genome-wide characterisation of enhancers across the genome, independently of sequence conservation. One example is ChIP-seq (for chromatin immunoprecipitation with high-throughput sequencing), a technique which enables genome-wide discovery of protein-DNA interactions, including transcription factor bindings and histone modifications, which involves post-translational modifications of histone proteins (Furey, 2012). Analyses of histone modification hallmarks through ChIP-seq has proven to be a more robust method in distinguishing enhancers from other CRMs on a genome-wide scale, enabling cell-type specific annotations (Coppola, Ramaker and Mendenhall, 2016). Studies on various cell lines and animal models including zebrafish (Aday *et al.*, 2011) have shown that enhancers are flanked by histone modification hallmarks including monomethylated H3K4 (K3K4me1), which is associated with inactive, poised, and active enhancers (Coppola, Ramaker and Mendenhall, 2016), and also histone 3 acetylated at lysine 27 (H3K27ac), which is used to distinguish active from poised enhancers (Zhou, Goren and Bernstein, 2011; Calo and Wysocka, 2013). However, other markers of enhancers also appear to exist in mouse ESCs, which include H3K122ac and H3K64ac, though they have yet to be identified in zebrafish (Pradeepa *et al.*, 2016). ChIP-seq against these histone hallmarks have previously been used to identify CRMs within the zebrafish genome at four different stages of development, dome (4.33 h.p.f.), 80% epiboly (~8.5 h.p.f.), Prim-5 (24 h.p.f.) and Long-pec (48 h.p.f.). This genome-wide analysis helped identify ~50,000 putative CRMs that may be important during the first two days of zebrafish development (Bogdanović *et al.*, 2012).

Though ChIP-seq has enabled the identification of putative enhancers in zebrafish, this was not without its limitations. Firstly, this work was limited to chromatin isolated from whole embryos, which means that it is near impossible to locate where enhancers likely drive gene expression based on histone hallmarks. Also, it can be difficult to identify enhancers in a minor cell population in the context of the whole embryo. In relation to this, a large number of embryos (1000~5000 cells according to the development stage) (Bogdanović *et al.*, 2012) or cells (at least 1-2 million cells) (Gilfillan *et al.*, 2012) are required thus, ChIP-seq is not feasible to be used on small populations of cells. Though methods have emerged that generated high-throughput libraries from limited number of cell types, these techniques were proven to be challenging especially when cell numbers are reduced to less than 1000 (Adli and Bernstein, 2011) and these techniques have yet to be performed on zebrafish embryos.

Over recent years, several other methods have been developed to profile chromatin accessibility. One of the earliest methods for probing accessible regions within the chromatin is FAIRE-seq, for Formaldehyde-Assisted Isolation of Regulatory Elements combined with

next-generation sequencing, which relies on formaldehyde cross-linking of histones to DNA and subsequent shearing of chromatin, while unbound regions were deemed to be accessible (Klein and Hainer, 2020). However, there are drawbacks to using this approach. Firstly, nucleosome-free regions bound by transcription factors or regions that are actively transcribed are also cross-linked (Klein and Hainer, 2020). Also, FAIRE-seq has lower signal-to-noise ratio compared to other chromatin profiling assays, which makes it difficult to computationally interpret data (Tsompana and Buck, 2014).

In relation to this, studies into gene regulation revealed that accessible CRMs which includes both enhancers and promoters are highly sensitive to cleavage by DNase I (Wu, 1980; Liu, Bergman and Zaret, 1988), and more recently, this technique was exploited to identify DNase I hypersensitive regions on a genome-wide scale which gave rise to DNase-seq (Crawford *et al.*, 2006). DNase-seq performed on several vertebrate genomes has demonstrated that there is a strong correlation between DNase I hypersensitive regions and gene expression, and has also enabled the identification of accessible regions that were later classified as CRMs (Crawford *et al.*, 2006; Boyle *et al.*, 2008). However, there are also drawbacks to using this technique. DNase-seq requires a large number of cells and may result in DNA loss since it involves multiple sample preparations and enzyme titration steps which further limits its sensitivity (Tsompana and Buck, 2014).

An alternative technique is MNase-seq, short for micrococcal nuclease digestion with deep sequencing, which is a widely used technique for mapping nucleosome positions and occupancy. It involves digestion of chromatin by the enzyme micrococcal nuclease (MNase) isolated from *Staphylococcus aureus*, an endo- and exo-nuclease that digests naked DNA between nucleosomes, releases the nucleosomes from chromatin and enriches for nucleosome-bound DNA, which are shielded from digestion by MNase (Chereji, Bryson and Henikoff, 2019; Klein and Hainer, 2020). However, during longer incubation times, MNase may also start to over-digest and destroy intact nucleosomes, resulting in fragments smaller than nucleosomal DNA (Noll and Kornberg, 1977; Sun *et al.*, 1986). Also, similar to DNase-seq, MNase-seq requires a large number of cells (Tsompana and Buck, 2014), making it unfeasible for minor cell populations.

Among these techniques, ATAC-seq has emerged as a rapid and sensitive alternative for mapping accessible regions (Buenrostro *et al.*, 2015). ATAC-seq (for Assay for transposase accessibility and high-throughput sequencing) involves the use of a hyperactive yeast Tn5 transposase to fragment and integrate sequencing adapters into accessible regulatory regions (Buenrostro *et al.*, 2013), (Figure 1-13). ATAC-seq has been exploited to map open chromatin

across a number of eukaryotic species including zebrafish (Quillien *et al.*, 2017; Pawlak *et al.*, 2019). Though the original protocol enabled the identification of accessible regulatory regions across vertebrate genome (Buenrostro *et al.*, 2015), it was prone to mitochondrial contamination, since it is more accessible to adapters due to lack of chromatin packaging (Yan *et al.*, 2020). For this reason, several different methods of ATAC-seq have been developed over the last few years to reduce mitochondrial reads in ATAC-seq (Klein and Hainer, 2020). This includes Omni-ATAC-seq, which differs in the reagents used to lyse cells and involves an additional washing step using detergents to remove mitochondria and improve library complexity (Corces *et al.*, 2017). Similar to standard ATAC-seq, Omni-ATAC-seq also uses phosphate-buffered saline in the transposition reaction to increase signal-to-noise ratio and to reduce background (Sun, Miao and Sun, 2019). However, unlike standard ATAC-seq, Omni-ATAC-seq eliminates mitochondria interference and thus, decreases background noise generated by mitochondrial DNA to obtain high quality data of chromatin accessibility (Corces *et al.*, 2017). Moreover, Omni-ATAC protocol can generate high-quality data on chromatin accessibility from frozen tissues, including brains (Corces *et al.*, 2017), unlike the standard ATAC-seq protocol which requires the transposition reaction to be performed on fresh and slowly cooled cryopreserved cells (Milani *et al.*, 2016).

All in all, unlike methods mentioned above, ATAC-seq is technically very simple, involves a fast two-step protocol, and is also highly sensitive since it can be applied to a small number of cells (500 to 50,000 cells), including single cells (Tsompana and Buck, 2014).

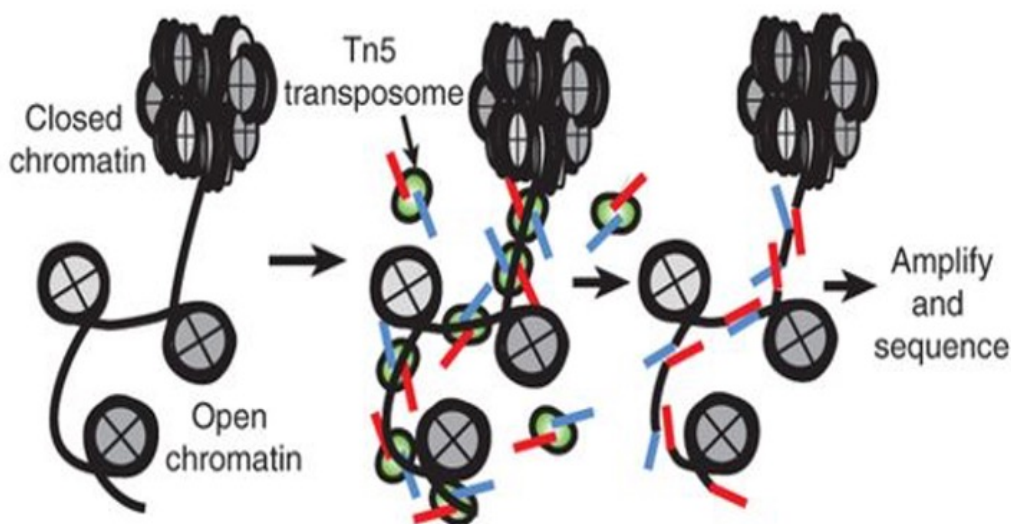


Figure 1-13: Schematic representation of ATAC-seq library preparation. ATAC-seq employs a hyperactive form of Tn5 transposase to identify accessible regions within the genome. Tn5 simultaneously cleaves and adds sequencing adapters to nucleosome-depleted regions to generate the ATAC-seq libraries. Following purification, the library is amplified by PCR using barcoded primers. The resulting library can be analysed by next-generation sequencing. Image from: (Buenrostro *et al.*, 2015).

ATAC-seq has already been exploited to identify developmental active enhancers in the genome and has enabled the identification of more than 5,000 active endothelial enhancers since they were all associated with histone hallmarks H3K4me1 and H3K27ac. Endothelial cells comprise less than 5% of the whole embryo (Quillien *et al.*, 2017), and thus a major advantage of ATAC-seq is that it can be used to identify active enhancers in minor cell populations. It has also recently been exploited to look at heart regulatory regions driving key heart morphogenesis at three stages of development and demonstrated that the chromatin landscape changes throughout heart development (Pawlak *et al.*, 2019).

Although a number of important CRMs have already been identified, a broad characterisation of developmentally important enhancers is still lacking. More specifically, a comprehensive list of enhancers required to regulate expression of key genes during organ formation is still missing.

1.15 Enhancers identified in endodermal organs

Following gastrulation, the endoderm differentiates into the primitive gut tube which later subdivides into broad foregut, hindgut and midgut domains which differ in gene expression. The foregut contributes to the oesophagus, trachea, stomach, lungs, thyroid, liver and biliary systems and pancreas, while hindgut and midgut give rise to the large and small intestine, respectively (Zorn and Wells, 2009), (Figure 1-14).

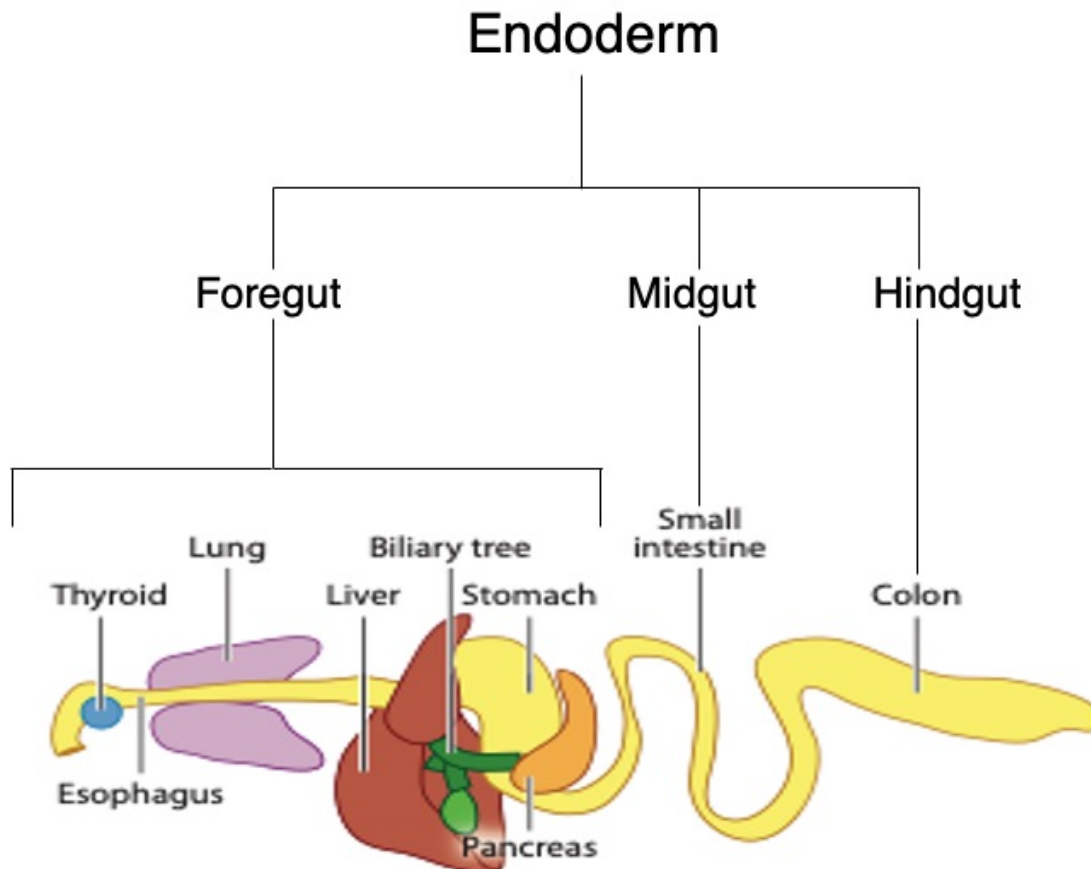


Figure 1-14: Endoderm makes contributes to the epithelial lining and associated organs of the gastrointestinal tract and respiratory system. Schematic adapted from: (Zorn and Wells, 2009).

In humans, ~ 100,000 enhancers have already been identified during the five different stages of pancreatic development; human ESCs, definitive endoderm, primitive gut tube, posterior foregut and pancreatic endoderm (Wang *et al.*, 2015). Interestingly, during differentiation of the endoderm to the primitive gut, enhancers acquire a poised state prior to activation, as marked by histone mark H3K4me1, and can later respond to inductive cues that help drive differentiation of ESCs to multiple gut tube organs, including pancreas, liver and lung lineages (Wang *et al.*, 2015). Consistent with this, studies in mice by ATAC-seq reveal that early midgut endoderm is also primed for heterologous cell fates and that binding of the intestinal transcriptional Cdx2 at intestinal enhancers at around mid-gestation, activates intestinal fate while loss of Cdx2 prior to embryonic day 14, stabilises foregut enhancers and imposes foregut identity (Banerjee *et al.*, 2018). However, information on developmentally important enhancers required to drive the formation of other endoderm-specific organs, including thymus, thyroid and pituitary glands is still lacking.

1.16 Enhancer mutations underlie numerous developmental diseases

Perturbation to the endoderm is the underlying cause of many developmental diseases and thus, there is an unmet need for understanding the underlying mechanisms of these conditions in order to find novel therapies. Historically, genome sequence analysis mainly focused on protein variants (Karnuta and Scacheri, 2018) however genetic variants constitute 25% of all developmental disorders. Instead, the majority of cases can be explained by mutations in non-coding regions of the genome (Short *et al.*, 2017). Genome-wide analysis has demonstrated that mutations in non-coding regions of the genome can be the most common case of discrete phenotypes. A notable example of this is pancreatic agenesis/hypoplasia which is a monogenic disorder caused by mutations in a long-range enhancer of a gene known as Pancreas Associated Transcription factor 1a (PTF1A), which is essential for early pancreas development (Miguel-Escalada, Pasquali and Ferrer, 2015). Linkage and whole-genome sequence analysis have previously uncovered six different mutations in a previously uncharacterised putative enhancer region ~25kb downstream of PTF1A in 10 families with pancreatic agenesis. Interestingly, the use of zebrafish transgenic reporter assays have identified a novel *ptf1a* enhancer, that when mutated, results in pancreatic agenesis in zebrafish, equivalent to the human disease phenotype (Bordeira-Carriço *et al.*, 2020). Though this region was not entirely analogous to the human enhancer sequence, the region contained the same transcription factor binding motifs for key pancreatic regulators including, FOXA2 and PDX1 (Pashos *et al.*, 2013; Weedon *et al.*, 2014; Bordeira-Carriço *et al.*, 2020) which suggests that they are functionally equivalent. However aside from pancreatic agenesis, there is still a lack of data that links other developmental diseases to causative enhancer mutations. In addition, though a small number of regulatory regions have been linked to pancreas development; this information is still missing for other endoderm-derived tissues. Thus, the use of zebrafish presents a powerful model for further characterisation and identification of risk disease variants in regulatory regions of other developmental disorders.

Genome-wide studies have also unveiled several single nucleotide polymorphisms in non-coding regions of the DNA associated with Type-2 diabetes (T2D), also called Type-2 diabetes mellitus (Saxena *et al.*, 2007; Sladek *et al.*, 2007; Flannick *et al.*, 2019). T2D is a chronic condition which affects around 300 million people world-wide and is characterised by insufficient β -cell function coupled with insulin resistance in a number of organs, including pancreas, liver, skeletal muscle, kidneys, brain, small intestine and adipose tissue (DeFronzo, 2009; Chatterjee, Khunti and Davies, 2017). Recently *in vivo* transgenic reporter assays were performed on zebrafish to test the functional relevance of risk variants associated with human T2D that were enriched for active enhancer hallmarks and contain transcription factor binding

motifs for pancreatic regulators, including FOXA2 and PDX1. The reporter assays demonstrated that a large majority of tested enhancers were able to drive expression of genes in β -cells and that single nucleotide mutations not only inhibited transcription factor binding but also completely impaired enhancer activity (Eufrásio *et al.*, 2020).

These studies highlight the advantage of using zebrafish for enabling our understanding of the consequences of enhancer modifications and disease progression. It is also apparent that despite of a lack of sequence conservation of enhancers between human and zebrafish, the regulatory mechanisms between them seems to be highly conserved. Thus, this provides a basis for using zebrafish to understand the regulatory network involved in early organ formation and how they may be perturbed in diseases.

1.17 Objectives of this study

Though the regulatory network involved in zebrafish endoderm formation is largely understood prior to gastrulation, there is a lack of information regarding how genes in the endoderm are regulated beyond endoderm specification. However, the endoderm constitutes a minor cell population making it difficult to find enhancers that regulate gene expression in the endoderm during development.

To enable the study of endoderm-specific CRMs at endoderm at specification, *sox32* overexpression was performed since it induces ectopic endoderm formation. We also created a new fish line to enrich for endoderm at later stages of development and to aid in finding CRMs in the various endodermal organs. Finally, we used ATAC-seq and RNA-seq at the two different stages of development to identify potentially important enhancers but also to investigate how the chromatin landscape changes in the endoderm during development. Understanding this will further aid our understanding of how the endoderm is regulated differently at different stages of development. Also, knowing how endoderm genes are regulated during development will further aid our understanding of how enhancer mutations lead to deleterious phenotypes.

The objectives of this study are as follows:

- 1) To find potentially important CRMs and motifs that drive endoderm formation downstream of *sox32* using both ATAC-seq, motif analysis and RNA-seq.
- 2) To find potentially important CRMs and motifs at the onset of organogenesis using both ATAC-seq and RNA-seq.

- 3) Investigate how the chromatin landscape changes within the endoderm by comparing CRMs from two different stages of endoderm development.

Chapter 2 Materials and Methods

2.1 Fish strains and transgenics

AB and *Tg(sox17:EGFP)^{ha01Tg}* fish (Mizoguchi *et al.*, 2008) were reared as described previously (Westerfield, 1995). Embryos were incubated at 28.5°C and staged according to (Kimmel *et al.*, 1995). Ages of development were expressed as hours post-fertilisation (h.p.f.). All zebrafish studies complied fully with the UK Animals (Scientific Procedures) Act 1986 as implemented by University of Warwick.

2.2 Generation of the triple transgenic line

Tg(kdr1:mCherry)^{ci5} (Proulx, Lu and Sumanas, 2010) were crossed with *Tg(gata1a:dsRed)^{sd2}* (Fish *et al.*, 2008) to make double heterozygous progeny, which were subsequently in-crossed and screened for double homozygous embryos before they were outcrossed to *Tg(sox17:EGFP)* embryos to make the triple transgenic line, which expresses all three transgenes. Triple transgenic embryos were screened at 28 h.p.f. for mCherry and dsRed expression and were reared as described previously (Westerfield, 1995).

2.3 Embryo injections

2.3.1 Labelling of DFCs using Rhodamine-dextran (RD) injection

Zebrafish embryos were injected with rhodamine B isothiocyanate-dextran (RD) (Gift from Professor Karuna Sampath group, University of Warwick) into the yolk sac of *Tg(sox17:EGFP)* embryos at mid-blastula stage (3 hours), to label the DFCs for imaging of the KV. The embryos were injected using Nikon SMZ800N stereo zoom microscope and an INTRACEL IM 300 Microinjector.

2.3.2 Sox32 mRNA injection

The coding region of *sox32* was cloned into pCS2 vector to make capped mRNA for overexpression analysis (Kikuchi *et al.*, 2001). Capped mRNA was synthesized using the mMMESSAGE mMACHINE SP6 (polymerase) Transcription Kit (Ambion). Sox32 overexpression was performed by injecting 150 pg of *sox32* mRNA into one cell-staged *Tg(sox17:EGFP)* embryos in a volume of 1nl.

2.3.3 Sox32 and standard morpholino injections

Injections of morpholino antisense oligonucleotides (MOs) (Gene Tools) were carried out as described previously (Nasevicius and Ekker, 2000). Sox32 MO was used to target the translation initiation site to block transcription of the gene and was injected at the one-cell stage of *Tg(sox17:EGFP)* embryos (5'-CAGGGAGCTACCGGTCGAGATACAT-3') (Dickmeis *et al.*, 2001) at 0.5 pmol each or, equivalent quantities of standard control MO (5'-CCTCTTACCTCAGTTACAATTTATA-3') (GeneTools).

2.4 Imaging of zebrafish embryos

For labelling of the KV, *Tg(sox17:EGFP)* embryos were mounted at 6-somite stage (12 h.p.f.) in low melting point agarose (1%, wt/vol) (Geneflow Limited). The chorions were first removed before mounting of embryos. 40 µl of agarose was positioned onto the glass coverslip of the glass bottom 25-mm culture dishes (MatTek corporation) and embryos were positioned onto the dish using a glass Pasteur pipette (150-mm, VWE). *Tg(sox17:EGFP)* embryos were imaged at 10x objective using Zeiss LSM 880 with Airyscan at 12 h.p.f. while Leica SP5 confocal microscope at 10x objective was used for imaging the posterior notochord at 24 h.p.f..

High resolution images of sox32 OE and control embryos at 6 h.p.f. were obtained using Nikon ECLIPSE Ni stereo microscope with an ORCA-flash4.0LT camera with Nikon Plan Fluor 4x objective. Embryos at 6 h.p.f. were first imbedded in low melting agarose (1%, wt/vol) (Promega) prior to imaging using the NIS ELEMENTS imaging software version 5.

For embryos older than 6 h.p.f., embryos were imaged using Nikon SMZ18 stereo zoom microscope and a Photometrics COOLSNAP HQ² CCD camera. Prior to imaging, chorions were manually removed, and the embryos were anesthetized with ethyl 3-aminobenzoate methane sulfonate salt solution (Mesab, Sigma).

All images were processed using Fiji (ImageJ) (Schindelin *et al.*, 2012).

2.5 Embryo dissociation

Embryos were first dechorionated in 1 mg/ml Protease from *streptomyces griseus* (Pronase, SIGMA) in egg water (60µg/ul ZM salt in distilled water). Once the chorion was successfully removed, the embryos were washed extensively in egg water to remove any residual pronase solution and were subsequently resuspended and dissociated in calcium free Ringers buffer with 0.5 M EDTA (GIBCO) (116mM NaCl (Sigma), 2.9 mM KCl (Sigma), 5 mM HEPES (Gibco).

Embryos at 28 h.p.f. were then subsequently dissociated using 25 mg/ml Collagenase (SIGMA) in 0.05% Trypsin (Gibco), 1 mM EDTA (Sigma), 1X Hank's Balanced Salt solution (HBSS) (Gibco). The reaction was stopped by the addition of 50 μ l Foetal Bovine Serum (FBS) (Sigma) in 1X HBSS+/+ solution (10X HBSS (Gibco), 250 mg BSA (Sigma), 1 ml HEPES (Gibco)).

Embryos were later centrifuged for 5 min at 500 xg and the pellet was gently resuspended in 1X HBSS+/+ solution. Dissociated cells derived from 28 h.p.f. embryos were further filtered through a 70 μ M cell strainer (Millex-GP) and resuspended in 1X HBSS+/+ solution prior to fluorescence-activated cell sorting (FACS).

2.6 Fluorescence-activated cell sorting (FACS)

Prior to FACS, cells from 28 h.p.f. embryos were placed into sterile FACS 5 ml polypropylene conical tubes (BD Biosciences) that were previously coated for at least one hour on ice with 1X PBS supplemented with 5% FBS. Samples were adjusted by dilution in 1X HBSS+/+ solution to assure optimal counting. FACS was performed by either Dr Andrew Nelson or Dr Mark Walsh using Becton Dickinson FACS Aria Fusion Cell Sorter, in the WISB cell cytometry lab. The technical settings relevant for this study were as follows: A 100 μ m nozzle was used with a sheath fluid pressure of 25 pounds per square inch (psi). Forward scatter (FSC) and side scatter (SSC) characteristics were used to select for single cells and cells were sorted based on GFP fluorescence (FITC, bandpass filter 530/30nm) and/or dsRed/mCherry fluorescence (PE-Texas Red-A, bandpass filter 610/20 nm). BD FACSDiva™ Software was used for both operating and analysing the flow cytometry data.

The total population of cells were distinguishable from background noise in a bivariate dot plot using the area of FSC (FSC-A) vs. area of SSC (SSC-A), and a gate was established around this population. Green and red fluorescent cells were gated based on fluorescence by using the area of FITC (FITC-A) vs area of PE-Texas red (PE-Texas Red-A). Sorted fractions were collected into 5 ml polypropylene conical tubes and maintained on ice and were later subsequently used for ATAC-seq (Assay for transposase-accessible chromatin using sequencing) and RNA-seq library preparation.

2.7 Real-time quantitative PCR (qRT-PCR)

Prior to qPCR, dissociated cells in 1X HBSS+/+ were centrifuged and the cell pellet was resuspended and lysed in 350 μ l RLT (Lysis buffer from the QIAGEN RNAeasy mini kit)-

containing a final concentration of 0.5% (v/v) 2-mercaptoethanol (SIGMA). RNA samples were extracted using QIAGEN RNAeasy Plus kit and cDNA synthesis was performed on 10 ng mRNA using First Strand cDNA Synthesis kit (Invitrogen), both according to the manufacturer's instructions. qRT-PCRs were performed using Power SYBR™ Green PCR Master Mix (Thermofisher) on an Agilent Technologies Stratagene Mx30005P and Ct values were exported and analysed using MxPro software. Primer sequences are reported below (Table 2.1).

Table 2.1: Oligonucleotides used for qRT-PCR

qPCR	Sequence 5' - 3'
Name	
<i>egfp_F</i>	AAGGGCATCGACTTCAAGGA
<i>egfp_R</i>	GGCGGATCTTGAAGTTCACC
<i>foxa1_F</i>	CTCAGCACAGCTCCATGAAC
<i>foxa1_R</i>	GCGGGTAACTCCTCCTGAA
<i>foxa2_F</i>	ATGCTCGGTGCTGTCAA
<i>foxa2_R</i>	CAAGTCCAGTGTTTCATGTTGC
<i>gata5_F</i>	AGAGACGGAACCGGACACTA
<i>gata5_R</i>	TCTCCTCCACAGTGTTGTCG
<i>18S_F</i>	TCGCTAGTTGGCATCGTTTATG
<i>18S_R</i>	CGGAGGTTCTGAAGACGATCA
<i>sox17_F</i>	CACAATGCGGAGCTGAGTAA
<i>sox17_R</i>	ATCGCTTGTTTCGTTTCACC
<i>pdx1_F</i>	GGACCAGCCAAATCTTACCG
<i>pdx1_R</i>	CCTCGGCCTCGACCATATA
<i>ins_F</i>	GCCCAACAGGCTTCTTCTACAAC
<i>ins_R</i>	GCAGATTTAGGAGGAAGGAAACCC
<i>sox32 UTR_F</i>	AGTACCTGTGTTAAAGTAAGAAATTGT
<i>sox32 UTR_R</i>	CACAGACTTTGGACCACAGC
<i>kdr1_F</i>	CCAGATCACGGTGGATACAT
<i>kdr1_R</i>	AAGAGGAGGAAGAGCAAGAG
<i>gata1a_F</i>	TCTGAGCCTTCTCGTTGGGTGTC
<i>gata1a_R</i>	TCCTGGAGCCTGGGACTGTCTTT
<i>sox32_F</i>	CCGTACATGCAAGAAGCAGA
<i>sox32_R</i>	ATCAAAGGTGGCATTGAG

<i>vgl4l_F</i>	ATCTCCACCAAGCAGTCCAA
<i>vgl4l_R</i>	TGGTTGGACTGCTCGAATCT
<i>cxcr4a_F</i>	TGGCTTATTACGAACACATCG
<i>cxcr4a_R</i>	GAGCCGAATTCAGAGCTGTT
<i>ndr1_F</i>	GCGAGCTGAACTTCGCATT
<i>ndr1_R</i>	TCAATTAGCCCAACCGCAAG
<i>ndr2_F</i>	AATGGGACCCAGAAACACAG
<i>ndr2_R</i>	CACGAGAACACGTCACCACA

2.7.1 qRT-PCR data analysis

For each primer set, qPCR efficiencies and specificity were first determined using standard curves of diluted cDNAs and melting curve analysis. Two pipetting replicates were used for each condition and calculations were performed using the $\Delta\Delta C_t$ method (Livak and Schmittgen, 2001). Error bars represent the standard deviation of at least two biological replicates and statistics were based on Student's t test.

2.8 Assay for transposase-accessible chromatin using sequencing (ATAC-seq)

50,000 cells from 6 h.p.f. embryos and 27,000 cells from 28 h.p.f. embryos were used for ATAC-seq as described previously (Corces *et al.*, 2017). Paired-end ATAC-seq libraries from *Tg(sox17:EGFP)* embryos were generated according to Corces *et al.* (2017). Briefly, cells were washed with ice cold PBS prior to centrifugation at 500 xg for 5 minutes at 4 °C. Following this, the cell pellets were then resuspended in cold lysis buffer (10 mM Tris-HCL, pH 7.5, 10 mM NaCl, 3 mM MgCl₂, 0.1% NP-40, 0.1% Tween-20, and 0.01% Digitonin) and incubated on ice for 3 minutes. Following this, the nucleic acid was washed with 1 ml wash buffer (10 mM Tris-HCL, pH 7.5, 10 mM NaCl, 3 mM MgCl₂ and 10% Tween-20) and the tubes were inverted to mix. The nuclei were subsequently centrifuged for 10 minutes at 500 xg at 4 °C and then resuspended in 50 μ l transposition mix (25 μ l 2X Tagment DNA buffer, 26.5 μ l 1X PBS, 0.5 μ l 10% Tween-20, 0.5 μ l 1% Digitonin, 2.5 μ l Tagment DNA Enzyme 1 and 5 μ l nuclease-free water) and the reaction was mixed by pipetting before incubating at 37 °C for 30 minutes. The reaction was cleaned using MinElute Reaction Cleanup kit (Qiagen). The remainder of the ATAC-seq library preparation were performed as described previously (Buenrostro *et al.*, 2015). All libraries were amplified for 6 PCR cycles. The libraries were sequenced by the Genomics Facility at the University of Warwick with Illumina NextSeq 500 using the High Output Kit v2.5 (FC-404-2002) according to the manufacturer's instructions, while libraries

from triple transgenic embryos were sequenced by Novogene (UK) Company Limited using Novaseq 6000-S4-type flow cell. To allow multiplexing of several ATAC-seq libraries in the same sequencing lane, different barcoded reverse primers were used for each sample. The quality of the ATAC-seq libraries were determined by Agilent Bioanalyzer 2100 to ensure that they all exhibit a pattern of fragment size periodicity with intervals of around 200 bp. Libraries were quantified by Dr Mark Walsh using JetSeq Library Quantification kit (Bioline). Libraries were diluted serially and in triplicates, according to the manufacturer's instructions. Samples were run on RotorGeneQ (Professor John McCarthy group, University of Warwick) with 10 μ l volumes using recommended cycling conditions. Library concentrations were estimated based on average fragment sizes reported by the bioanalyzer.

2.8.1 ATAC-seq data analysis

For ATAC-seq, data were analysed using a computational protocol as composed by Katherine C Woolley-Allen (Sascha Ott group, University of Warwick). Firstly, FastQC was used to check the quality of the sequenced ATAC-seq libraries (Anders, 2010). Adapter sequences were removed using Trimmomatic (Bolger, Lohse and Usadel, 2014) and reads were mapped to zebrafish reference genome, GRCz11.97, using Bowtie2 (Langmead and Salzberg, 2012). Default settings were modified to allow for mapping of paired-end fragments up to 2 kb. Reads mapping to mitochondrial DNA were excluded from the analyses together with low quality reads (MAPQ<22) and duplicated reads using Picard (<http://broadinstitute.github.io/picard/>). Mapped reads were visualised using Integrated Genome Viewer (version 2.3.10) (Thorvaldsdóttir, Robinson and Mesirov, 2013). For peak calling, samples were compared to genomic DNA control, where all proteins have been removed, to ensure reads are not aligned unambiguously as a result of Tn5 sequence-specific signature bias (Li *et al.*, 2019). Approximately, 50 million reads sample were mapped to genomic DNA, as shown in Table 2.2. Model-based Analysis of ChIP-seq (MACS2 (version 2.1, (Zhang *et al.*, 2008)) was used to call enrichment with the following settings: callpeak -g 1.4e9 -qvalue 0.05 -bw 300 -mfold 5 50. Peaks were also merged and these representative peaks were used to display putative enhancers within the endoderm at different stages of development. Merging was performed by SAMtools (Li *et al.*, 2009). Called peaks were displayed using Integrated Genome Viewer. Narrow peak parameters were utilised since they were found to be easier to distinguish from background noise compared to broad peaks parameters. Peak calling cut-off of 0.05 was used to remove background noise.

To identify differential accessible regions (DARs), Diffbind version 3.0.15 was used (Stark and Brown, 2011). Diffbind was used to determine DARs from peaks called by MACS2 in at least

two samples from all the conditions tested. DESeq2 option during Diffbind analyses was then subsequently used to normalise read counts and to help identify DARs between two sets of sample groups. DARs for *sox32* OE/ *sox32* KD vs control or *sox17E*/*sox17M* vs *sox17N* were identified by using \log_2 fold change in read density ≤ 1 and ≥ 1 , respectively with a false discovery rate (FDR) cut-off of ≤ 0.05 and Pvalue of ≤ 0.02 . Differential accessibility heatmaps were plotted based on FDR to show statistically significant DARs between the two sample groups.

Table 2.2: Total number of mapped read counts per biological replicate

Sample	Replicate number	Total number of mapped read counts
6 h.p.f. control	1	57,278,982
	2	52,489,324
	3	66,103,240
6 h.p.f. <i>sox32</i> knockdown (KD)	1	48,248,090
	2	47,630,398
	3	61,603,986
6 h.p.f. <i>sox32</i> overexpression (OE)	1	49,732,812
	2	48,286,218
	3	75,497,540
28 h.p.f. <i>sox17</i> + negative population (<i>sox17N</i>)	1	115,198,342
	2	44,640,558
28 h.p.f. <i>sox17</i> + endoderm population (<i>sox17E</i>)	1	14,120,110
	2	57,803,282
28 h.p.f. <i>sox17</i> + mesoderm population (<i>sox17M</i>)	1	11,761,930
	2	45,117,334

To compare DARs across multiple samples, *k*-means clustering was performed using SeqMINER version 1.3.4 (Krebs *et al.*, 2010) by Dr Andrew Nelson. 50 million reads per BAM file were generated for each merged sample group using Picard and these were used as input for SeqMINER. DARs from each sample group were used as genome coordinates. The read densities from each dataset were collected in a window of +/- 5 kb around each DAR centre and the collected values were subjected to *k*-means clustering (using linear normalisation method). The number of expected clusters was 10 by default. Visualisation of the results was achieved through a heatmap.

Genomic distribution and enrichment of ATAC-seq called peaks were estimated using ChIPseeker (Yu, Wang and He, 2015). Peaks were annotated with respect to the transcriptional start site (TSS) where peaks within -3kb to +3 kb were defined as promoters. Frequencies of distribution of peaks around the TSS were plotted using the “annotatePeak” function of the R package “ChIPseeker”. ChIPseeker genomic annotation is based on the following hierarchy: promoter, 5’UTR, 3’UTR, exon, intron, downstream of a gene end and intergenic.

For ChIP-seq analysis, ChIP-seq data were downloaded from GEO: Nanog (GEO: GSE34683; (Xu *et al.*, 2012)) and Pou5f3 (GEO: GSE39780 (Leichsenring *et al.*, 2013)). Tbxta and Tbx16 ChIP-seq coordinates were downloaded from Nelson *et al.* (2017). Downloaded ChIP-seq data were mapped to danRer11 and peaks were called using MACS2, as stated above, while Tbxta and Tbx16 coordinates were converted to danRer11 genome build using LiftOver (<https://genome.ucsc.edu/cgi-bin/hgLiftOver>). BEDTools (Quinlan and Hall, 2010) was subsequently used to screen for overlaps between DARs and ChIP-seq coordinates. Differences in overlap between *sox32* OE or control DARs with ChIP-seq coordinates were compared using Chi-squared test.

To look for over-represented anatomical terms associated with DARs, Genomic Regions Enrichment of Annotations Tool (GREAT) (McLean *et al.*, 2010) analysis was performed using the web interface (<http://bejerano.stanford.edu/great/public/html/>). DAR coordinates from each dataset were first converted to danRer7 genome build using LiftOver (as mentioned above) before using the GREAT web interface version 3.0.0 to look for enriched anatomical terms in the tested dataset. The single nearest gene parameter was chosen for all analysis to ensure that regions are specified to one gene. GREAT assigns each gene a regulatory domain that extends to 1000 kb in both directions to the midpoint between TSS of one gene and TSS of a neighbouring gene. Each DAR is associated with all genes in whose regulatory domain it lies. GREAT uses both a binomial and a hypergeometric test to discover significant associations between DARs from each dataset and anatomical terms associated with proximal genes. Zebrafish anatomical terms were identified by viewing “zebrafish Wildtype Expression” in GREAT. Only gene classes with a false discovery rate of less than 0.05 were considered.

Gene Ontology with ‘Protein Analysis THrough Evolutionary Relationships’ (PANTHER, <http://PANTHERdb.org/>) (Thomas, Campbell, *et al.*, 2003; Thomas, Kejariwal, *et al.*, 2003) or ‘The database for annotation, visualisation and integrated discovery’ (DAVID, <https://david.ncifcrf.gov/tools.jsp>) (Huang, Sherman and Lempicki, 2009b, 2009a) were also used for further functional annotations of genes proximal to GREAT.

Over-represented transcription factor sites in DARs were identified using Hypergeometric Optimization of Motif EnRichment (HOMER; (Heinz *et al.*, 2010)) version 4.11. Motif analysis was performed using findMotifsGenome.pl function and the size parameter was set to 200 (-size 200). In brief, HOMER uses a library of transcription factor binding motifs (in the form of position weight matrices) and scans an input set of genomic regions (in this case DARs) for statistical enrichment of each position weight matrix relative to GC matched background sequences generated by HOMER. HOMER then ranks all enriched transcription factor binding motifs according to their Pvalues. Motifs with Pvalues less than 0.01 were considered in this report. Transcription factor motifs were chosen based on either their function or presence in the endoderm/Kupffer's vesicle.

2.9 RNA-seq

2.9.1 RNA-seq (6 h.p.f.): high input library

Total RNA was extracted using the QIAGEN RNeasy Mini kit with on-column DNase I treatment (QIAGEN) following the manufacturer's protocol. The quality of the RNA samples was checked with Agilent Bioanalyzer 2100 and all RNA samples had RNA integrity numbers > 7. Library concentrations were based on Qubit values. For 6 hour-staged embryos from *Tg(sox17:EGFP)* line, libraries were prepared by Genomics facility, University of Warwick using the Illumina TruSeq stranded total RNA kit according to the manufacturer's protocol starting with 500 ng total RNA and sequenced with Illumina NextSeq 500 using the High Output Kit v2.5 (FC-404-2002). Briefly, rRNA-depleted RNAs (Total RNA kit) were fragmented and converted to cDNA with reverse transcriptase. The resulting cDNA were converted to double stranded cDNAs and subjected to end-repair, A-tailing, and adapter ligation.

2.9.2 RNA-seq (28 h.p.f.): low input library

For RNA sample preparations from triple transgenic line at 28 h.p.f., 10 ng RNA was served as the starting material for each sample. Libraries were constructed by using the NEBNext Ultra II Directional RNA Library Prep kit for Illumina according to the manufacturer's instructions and were sequenced using Novaseq 6000-S4-type flow cell. In brief, ribosomal RNA was depleted with residues randomly cut into short fragments. First-strand cDNA was produced with random hexamer primer and second-strand was synthesised using DNA polymerase I and a reaction buffer containing deoxyuridine triphosphates (dUTPs). The 5' overhangs were changed into blunt ends with exonuclease and polymerase activity. Following adenylation of 3' DNA ends and hairpin structure ligation, libraries were prepared by selecting for cDNA fragments of < 200 bp. Thereafter, PCR was conducted with High-Fidelity DNA

polymerase, Universal PCR primers and Index (X) Primer. Prior to sequencing, the PCR samples were purified using Agencourt AMPure XP beads and the library quality was assessed by Agilent Bioanalyzer 2100.

2.9.3 RNA-seq data analysis

For RNA-seq analysis, reads were mapped onto zebrafish genome, danRer11 using STAR (Dobin *et al.*, 2013). RNA-seq data quality was checked by FastQC prior to downstream analysis (Anders, 2010). HTSeq was used to count aligned reads per gene (Anders, Pyl and Huber, 2015) and the raw count matrix was imported into DESeq2 R package (Love, Huber and Anders, 2014) and normalised by log₂ transformation. Genes with read counts of less than 10 were removed from the analysis. Genes with FDR < 0.05 were defined as differentially expressed. To check the gene expression changes between samples, genes were sorted by log fold change and clustered by hierarchical clustering using DESeq2's regularised-logarithm transformation (rlog), which transforms the read counts to the log₂ scale to help normalise the data with respect to library size and to minimise differences between samples for genes with low counts. For visual representation of differential expressed genes (DEGs) between the different sample groups, significant endoderm-expressed genes with FDR ≤ 0.05 were chosen and the read counts were plotted by heatmaps with accompanying dendrograms using ggplot2 package in RStudio.

To investigate a relationship between genes from two different data sets, Gene set enrichment analysis (GSEA), version 4.0.3 was applied (Subramanian *et al.*, 2005). Genes enriched in the endoderm, mesoderm and ectoderm were obtained from published single-cell RNA-seq data at six different stages of development (6,8,10,14,18, and 24 h.p.f.) as well as DFCs at 8 h.p.f. (Wagner *et al.*, 2018), in which single cell transcriptomics were collected using InDrops during the first 24 hours of embryonic development. Genes were obtained using the web interface (https://kleintools.hms.harvard.edu/tools/springViewer_1_6_dev.html?cgi-bin/client_datasets/fish_embryo_timecourse/full). GSEA was used to compare DFC, endoderm, mesoderm, and ectoderm-enriched genes from published single-cell RNA-seq data to genes pre-ranked on statistic from RNA-seq analysis based on 1,000 permutations. The cut-off value for GSEA was set at 0.05 for both FDR and Pvalue.

Chapter 3 Overexpression of *sox32* induces ectopic formation of both endoderm and DFC fates

3.1 *Sox32* is required for endoderm formation

Endoderm formation in zebrafish requires specific transcription factors (TFs) that function downstream of Nodal signalling but upstream of *sox17*, including Mix1, Gata5 and Sox32 (Kikuchi *et al.*, 2001). While mutations in *mix1* and *gata5* disrupt endoderm formation and reduce the number of endoderm progenitors (Kikuchi *et al.*, 2001), *sox32* mutants display more dramatic phenotypes; they lack all *sox17* expression in the endoderm and dorsal forerunner cells (DFCs), fail to form a gut tube (Kikuchi *et al.*, 2001) and exhibit cardia bifida by 24 h.p.f. (Alexander *et al.*, 1999) and these phenotypes were also recapitulated by using morpholino antisense oligonucleotides against *sox32* (Sakaguchi, Kuroiwa and Takeda, 2001). Additionally, in the absence of *sox32*, endodermal progenitors differentiate into mesoderm structures (Dickmeis *et al.*, 2001), while overexpression of *sox32* results in ectopic expression of *sox17* as well other endoderm markers including *ndr1* and *ndr2* throughout the blastoderm (Kikuchi *et al.*, 2001). However, in contrast to *sox32* mutants, *sox32*-overexpressing embryos fail to undergo normal gastrulation, lack mesoderm and as a result, die by 9 h.p.f. (Stafford *et al.*, 2006; Kinkel and Prince, 2009). This suggests a key role of *sox32* in fate decision between endoderm and mesoderm, where it induces an endoderm-specific transcriptional programme within the embryo, and as a result, *sox32* is considered a master regulator of endoderm fate (Alexander and Stainier, 1999; Dickmeis *et al.*, 2001; Kikuchi *et al.*, 2001; Ober, Field and Stainier, 2003).

Sox32 expression initiates in the dorsal yolk syncytial layer (YSL) at early blastula stage and is later expressed during gastrulation in all endodermal cells prior to expression of *sox17* (Kikuchi *et al.*, 2001). *Sox32* is believed to play a central role in endoderm formation since its expression can induce pan-endodermal markers, *sox17* and *Forkhead box A2 (foxa2)* in the absence of Nodal signalling (Dickmeis *et al.*, 2001; Kikuchi *et al.*, 2001). Though, *sox32* plays a central role in endoderm formation, it is specific to teleosts and likely arose as a result of small-tandem duplication of *sox17* during evolution, hence it is located adjacent to *sox17* in the zebrafish genome (Voldoire *et al.*, 2017). Interestingly, *Sox17*, which functions downstream of *sox32* in zebrafish, is a master inducer of endoderm fate in other vertebrates. For instance, in *Xenopus*, knockdown of *Sox17* results in a loss of endoderm in the vegetal blastomeres, while overexpression of *Sox17* in conjunction with *mixer*, a homologue of

zebrafish *mixl1*, results in expression of endoderm-specific genes in the animal cap (Hudson *et al.*, 1997; Henry and Melton, 1998). In relation to this, *Sox17* overexpression induces endoderm fate in human and mouse embryonic stem cells (Séguin *et al.*, 2008; Niakan *et al.*, 2010). Therefore, *Sox32* appears to be functionally equivalent to *Sox17* in other vertebrates since it is necessary for endoderm formation (Dickmeis *et al.*, 2001; Kikuchi *et al.*, 2001; Sakaguchi, Kuroiwa and Takeda, 2001; Aoki *et al.*, 2002).

In zebrafish, *Pou5f3*, a zebrafish pluripotency factor is believed to play a central role in endoderm formation by acting synergistically with *Sox32* and inducing expression of *sox17* in endoderm precursor cells (Alexander *et al.*, 1999; Kikuchi *et al.*, 2001; Lunde, Belting and Driever, 2004; Reim *et al.*, 2004). *Sox32* can also interact with *Nanog*, another core pluripotency marker, (Mitsui *et al.*, 2003) required for mesendoderm differentiation (Xu *et al.*, 2012) and restricts *Pou5f3*-*Nanog* complexes to the ventrolateral mesendoderm (Perez-Camps *et al.*, 2016). While the functional relevance of *Sox32*-*Nanog* complexes is not yet known, *Nanog* is known to act upstream of *sox32* where it induces expression of endoderm regulators, including Nodal factors, *ndr1* and *ndr2* but also *mixl1*, *gata5* and *sox32* in mesendodermal cells (Xu *et al.*, 2012; Veil *et al.*, 2018). In addition, loss of endoderm in *nanog* morphants resemble the phenotypes of maternal and zygotic (MZ) *pou5f3* mutants, also known as *MZspg* mutants, which fail to express *sox17* and maintain endoderm fate (Lunde, Belting and Driever, 2004; Xu *et al.*, 2012) which likely suggests that both *Nanog* and *Pou5f3* participate in the same developmental programme albeit at different stages of development. Aside from binding to *Sox32*, both *Pou5f3* and *Nanog* also play essential roles during zygotic genome activation (ZGA) by opening up chromatin and priming genes (Stafford *et al.*, 2006; Kinkel and Prince, 2009). Indeed anti-sense knockdown of both factors results in a wide-spread decrease in chromatin accessibility and wide-spread decrease in zygotic gene expression. Thus, both pluripotent factors are likely essential during different stages of development: ZGA and endoderm specification.

T-box transcription factors *Tbx16* and *Tbx18* were also found to play essential roles during early mesendoderm development as well as left/right asymmetry establishment. Indeed, while *Tbx18* is required for notochord formation and *Tbx16* is an essential regulator of paraxial mesoderm development (Ho and Kane, 1990; Amacher *et al.*, 2002), together they are also required for Kupffer's vesicle (KV) specification with evidence that *tbx18* and *tbx16* mutants exhibit distinct KV and left-right asymmetry defects (Amack, Wang and Yost, 2007). However, *Tbx18* and *Tbx16* also play important roles during early endoderm specification. Comparison of ChIP-seq peaks from both factors demonstrated that they not only bind common genomic regions but that they also regulate essential modulators of endoderm formation, including

Mixl1 and Gata5 as well as extrinsic regulators of endoderm proliferation, Cxcl12a/b (Nelson *et al.*, 2017). Moreover, *sox32* expression is severely compromised in *tbxta/tbx16* double morphants (Nelson *et al.*, 2017), which suggests that endoderm specification requires both T-box factors Tbx16 and Tbx16.

3.2 Studying the chromatin landscape

Transcriptional regulation is fundamental for development and it is vital that it is controlled spatially and temporally (Nelson and Wardle, 2013). Genes are regulated by a variety of TFs and chromatin remodelling proteins which bind to DNA sequences known as *cis*-regulatory modules (CRMs) (Suryamohan and Halfon, 2015), which have transcriptional activity and as a result, genes are either expressed or silenced. CRMs can function as promoters, enhancers silencers and insulators and despite their importance in regulating gene expression, CRMs are poorly annotated, and the vast majority are not yet characterised. This is perhaps because unlike genes and promoters, which have recognisable features within the DNA sequence, no similar properties are known for other classes of CRMs, especially enhancers and silencers which enhance and repress target gene expression, respectively. Another challenge is that often CRMs are located far from the target genes they regulate. Availability of zebrafish genome sequences (Howe *et al.*, 2013), and multi-species alignment of genomic sequences (Wasserman *et al.*, 2000) has helped identify a number of conserved CRMs which often harbour enhancer activity since they were found to enhance transcription of target genes (Göttgens *et al.*, 2002; Komisarczuk, Kawakami and Becker, 2009). However, evolutionary conserved CRMs are not always functionally conserved. Also, most CRMs are not universally conserved across species. Thus, there is still a need to find CRMs that regulate gene expression on a genome-wide scale during embryonic development.

Zebrafish is a powerful model for investigating gene regulatory networks since the embryos are transparent, grow externally, and its genes are really easily manipulated (Ferg *et al.*, 2014). While detailed analysis of the transcriptional network in humans (Gerstein *et al.*, 2012), nematode *Caenorhabditis elegans* (Gerstein *et al.*, 2010) and *Drosophila* (Roy *et al.*, 2010) have benefited from collaborative large-scale genomic effort, zebrafish was not included in this work, highlighting that there is still a need to find developmentally important CRMs in zebrafish.

Previous genome-wide analyses have revealed that active enhancers are associated with specific histone modifications which include histone 3, trimethylated at lysine 4 (H3K4me1), histone 3 acetylated at lysine 27 (H3K27ac) (Heintzman *et al.*, 2009; Visel *et al.*, 2009;

Creyghton *et al.*, 2010). For this reason, antibodies against these enhancer-associated marks through chromatin immunoprecipitation (ChIP) with Next-generation sequencing (ChIP-seq), is widely used to identify genome-wide enhancers in a number of tissues and model organisms (Visel *et al.*, 2009, 2013; Vandermeer *et al.*, 2014), including zebrafish (Bogdanović *et al.*, 2012). However, this work was conducted on the chromatin that was extracted from whole embryos which makes it difficult to predict the location of active enhancers. Moreover, it is also difficult to identify enhancers that are active within a minor cell population in the context of the whole embryo, and as a result will likely only contribute a negligible signal in a ChIP-seq analysis. ChIP-seq also requires a large number of cells, at least, 1-2 million cells (Gilfillan *et al.*, 2012), making it difficult to apply to minor cell populations from isolated zebrafish embryos. Though methods have emerged that enable ChIP-seq on small numbers of cells (Adli and Bernstein, 2011), these techniques are challenging and have not been implemented on zebrafish cells.

Over recent years, several methods have been developed to profile chromatin accessibility, including FAIRE-seq, DNase-seq and ATAC-seq (Song and Crawford, 2010; Simon *et al.*, 2012; Buenrostro *et al.*, 2015). Among these techniques, ATAC-seq has emerged as the leading technique for mapping open accessible regions (Wei *et al.*, 2018). Assay for Transposase-Accessible Chromatin with high-throughput sequencing (ATAC-seq) identifies accessible regions by probing open chromatin with a hyperactive derivative of the Tn5 transposase that inserts adapters into open regions of the genome and concurrently, the DNA is sheared by the transposase activity (Buenrostro *et al.*, 2013). ATAC-seq is highly sensitive and can be applied to a small number of cells. ATAC-seq has already been exploited to look at gene regulation in endothelial cells, which comprise less than 5% of all cells in the embryo, and through fluorescence-activated nuclei sorting on cells isolated from zebrafish, has enabled the identification of more than 5,000 active endothelial enhancers (Quillien *et al.*, 2017). More recently, transcriptome profiling as well as ATAC-seq were used to study the gene regulatory network driving heart development (Pawlak *et al.*, 2019). Interestingly, ATAC-seq has also been combined with other techniques such as RNA-sequencing (RNA-seq), which together offer a powerful approach to understanding the role of the chromatin structure in regulating gene expression. The combination of ATAC-seq and RNA-seq on isolated human α - and β -cells has already helped locate open chromatin surrounding novel genes that were never shown to be expressed in these cell types (Ackermann *et al.*, 2016). More importantly, these genomic regions were found to harbour binding sites for known islet transcription factors, including PAX6, and FOXA2 and single nucleotide polymorphisms within these regions were identified as risk loci for Type-2 diabetes. Interestingly, a comparison of zebrafish, human and mouse genomic regions has enabled the identification of both

evolutionary conserved but also species-specific regions and gene regulatory networks. A recent study has identified that approximately 60-90% of zebrafish enhancers whose sequences were conserved in humans were also predicted to be enhancers in the same tissue in humans which includes the liver, intestine, heart and brain (Yang *et al.*, 2020). Thus, this highlights the importance of using zebrafish as an important system to study human regulatory networks.

3.3 Objectives of this study

Though ChIP-seq analysis and other procedures have enabled the identification of developmentally important CRMs prior to endoderm specification (Stainier, 2002; Zorn and Wells, 2007, 2009; Tremblay, 2010; Nelson *et al.*, 2014, 2017), CRMs that are bound by factors downstream of *sox32* are not well understood. The endoderm constitutes a minor cell population (Figiel, Elsayed and Nelson, 2021) and thus challenging to identify endoderm-specific enhancers in the context of the whole embryo. To enable the study of CRMs within the endoderm, *sox32* was overexpressed in *Tg(sox17:EGFP)* embryos, which marks all endoderm progenitor cells. This procedure overcomes this challenge since overexpressing *sox32* results in ectopic endoderm formation and therefore facilitates the finding of endoderm-specific regulatory regions.

The main objectives of this study

- 1) To investigate how *sox32* influences the chromatin landscape in the embryo to induce endoderm fate. This will be investigated by using ATAC-seq and comparing read densities across differential accessible regions from control vs *sox32*-expressing embryos.
- 2) To understand the effect of *sox32* overexpression on DFCs. This will be investigated by qRT-PCR, using DFC-specific markers on *sox32*-overexpressing vs control embryos. Moreover, RNA-seq analysis on *sox32*-overexpressing embryos vs control embryos will also be used to further investigate the effects of *sox32* overexpression on DFC-specific transcripts on a genome-wide scale.
- 3) To identify important CRMs that likely function downstream of Sox32 induction. This will be investigated by the use of ATAC-seq on *sox32*-overexpressing vs control embryos during endoderm specification (~ 6 h.p.f.), which is after endogenous *sox32* is induced within the embryo. This will also be followed by the use of Genomic Regions Enrichment of Annotations Tool (GREAT) to identify open chromatin regions within the genome.

To aid in the identification of active enhancers and gene expression in the endoderm, ATAC-seq and RNA-seq were applied to *sox32*-overexpressing embryos. Since embryos overexpressing high levels of *sox32* do not undergo normal gastrulation and die by 9 h.p.f. (Stafford *et al.*, 2006), we aimed to identify active CRMs during early stages of endoderm specification (~6 h.p.f.).

3.4 Sox32 overexpression upregulates expression of sox17 in the embryo

One of the major challenges of identifying active enhancers in the endoderm during development is that the endoderm constitutes a minor cell population within the embryo (Figiel, Elsayed and Nelson, 2021). Previously, overexpression of the endoderm master regulator, *sox32* was shown to induce ectopic expression of *sox17* in zebrafish (Kikuchi *et al.*, 2001). Also, transplanted cells from donor embryos injected with *sox32* mRNA into *sox32* morphants, which completely lack endoderm, restored expression of endoderm-specific markers and the development of endoderm tissues (Stafford *et al.*, 2006). Thus, *sox32* is necessary for endoderm formation.

To aid in the identification of endoderm-specific enhancers at specification stage (~6 h.p.f.), *sox32* mRNA was injected into 1-cell staged *Tg(sox17:EGFP)* embryos to induce ectopic endoderm formation. This is because the endoderm is a minor cell population, and around 6 h.p.f., the GFP from the transgene is not yet expressed in un-injected embryos, and thus fluorescence activated cell sorting (FACS) cannot be used to isolate endoderm from whole embryos. This makes it challenging to identify enhancers that function within the endoderm since the endoderm cannot be isolated by this method during early specification. However, *sox32* overexpression induces ectopic endoderm formation which can be used to enrich for endoderm and facilitate the finding of endoderm-specific enhancers. Consistent with previous data, overexpression of *sox32* resulted in ectopic GFP expression, which likely represents the endoderm (Figure 3-1)

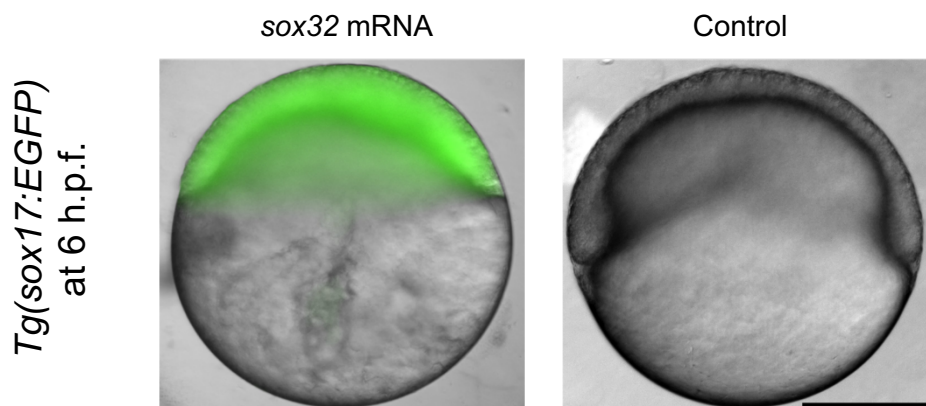


Figure 3-1: Fluorescent images of control and *sox32* mRNA injected *Tg(sox17:EGFP)* embryos. Embryos were imaged laterally at 6 h.p.f. with animal pole cells towards the top. GFP expression can be observed in *sox32*-overexpressing embryo compared to time-matched control. Scale bar: 250 μ M.

To study the effects of *sox32* overexpression (*sox32* OE) on the chromatin landscape, *sox32* overexpressing (hereafter referred to as *sox32* OE) embryos were compared to *sox32*-deficient embryos, which were subjected to morpholino-mediated knockdown (KD) and

therefore lack endoderm, a phenotype that is also recapitulated in *sox32* mutants (Sakaguchi, Kuroiwa and Takeda, 2001). *Sox32* OE and *sox32* morphants (hereby referred to as *sox32* KD) were collected at 6 h.p.f. and compared to *Tg(sox17:EGFP)* embryos injected with control morpholino (hereby referred to as control), prior to ATAC-seq and RNA-seq.

To validate the experiment and to ensure that *sox32* is overexpressed or reduced in *sox32* OE and *sox32* KD embryos, respectively, relative to control, qRT-PCR was performed on three independent biological samples against known endoderm markers (Figure 3-2). These include, the endoderm marker *sox17*, the transgene marker, *gfp* and endogenous master regulator *sox32*, where primers were used to specifically target 3' UTR which is absent in exogenous *sox32* mRNA. As demonstrated by qRT-PCR (Figure 3-2), *sox17*, *gfp* and endogenous *sox32* were significantly upregulated in *sox32* OE embryos compared to control. In contrast to this, the expression level of *sox17* and *gfp* were significantly downregulated in *sox32* KD embryos, while no significant difference was observed in the expression of endogenous *sox32* relative to control. Moreover, comparison of *sox32* KD and *sox32* OE embryos revealed that the expression level of all three markers were enriched in *sox32* OE embryos. Thus, these results are consistent with previous studies (Kikuchi *et al.*, 2001) that show that endoderm is more abundant in *sox32* OE embryos, while *sox32* KD results in reduction of endoderm markers.

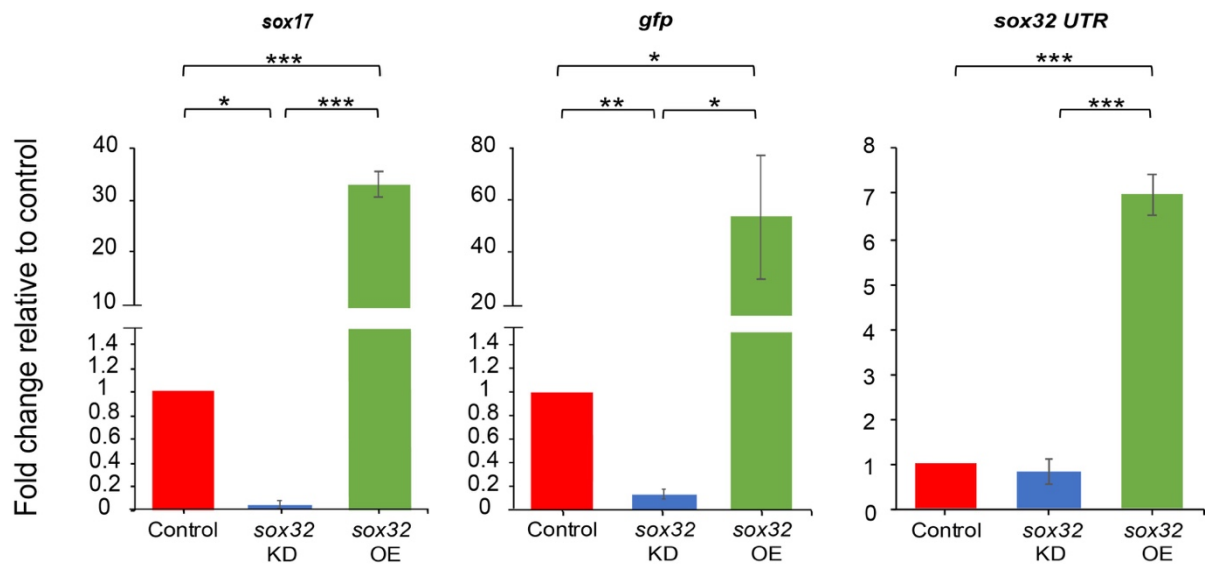


Figure 3-2: qRT-PCR analysis of *sox17*, *gfp* and endogenous *sox32* in control, *sox32* KD and *sox32* OE embryos. All genes were significantly upregulated in *sox32* OE while *sox17* and *gfp* were significantly downregulated in *sox32* KD; *** $p \leq 5 \times 10^{-5}$; ** $p \leq 5 \times 10^{-3}$; * $p \leq 5 \times 10^{-2}$; Student's t test. Data was normalised to *18S*. $n = 3$ per condition. Error bars represent standard deviation.

Therefore, the results are in accordance with published data which provided assurance for performing ATAC-seq and RNA-seq. A schematic of the experimental plan can be shown below (Figure 3-3).

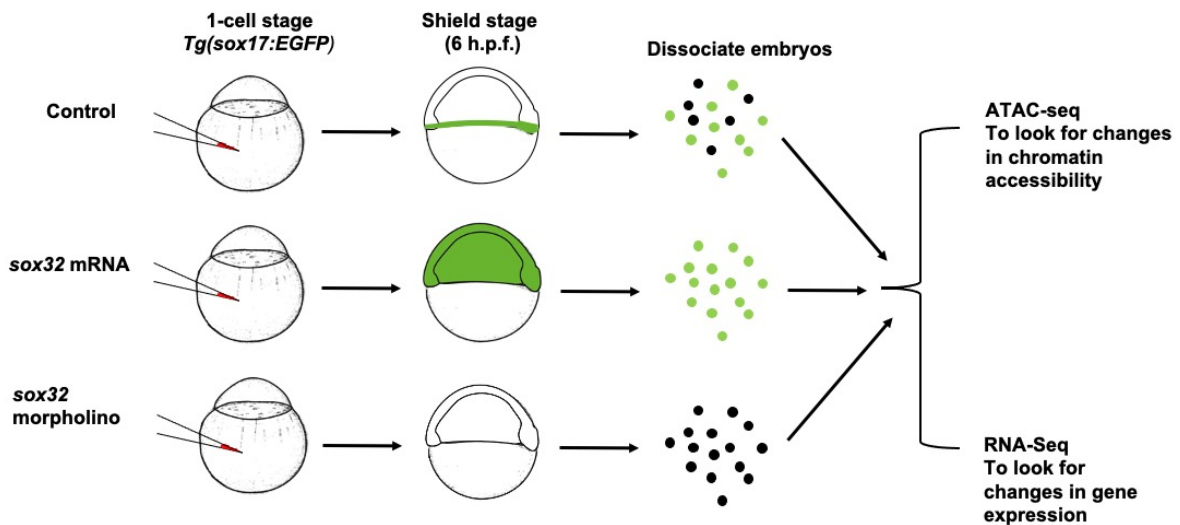


Figure 3-3: Schematic of experiment performed at 6 h.p.f. using the pan-endodermal line *Tg(sox17:EGFP)*. *Sox32* OE and *sox32* KD embryos were compared to control for qRT-PCR analysis prior to performing ATAC-seq and RNA-seq.

3.5 *Sox32* overexpression induces opening of chromatin in regions proximal to endoderm-expressed genes

In order to identify endoderm-specific enhancers at 6 h.p.f., ATAC-seq was performed in triplicate on cells isolated from *sox32* OE, *sox32* KD and control embryos. Insert size distributions were subsequently obtained from the ATAC-seq analysis of control, *sox32* KD and *sox32* OE sample groups and these were found to follow the expected trend, as shown in Figure 3-4 and were therefore considered of good quality.

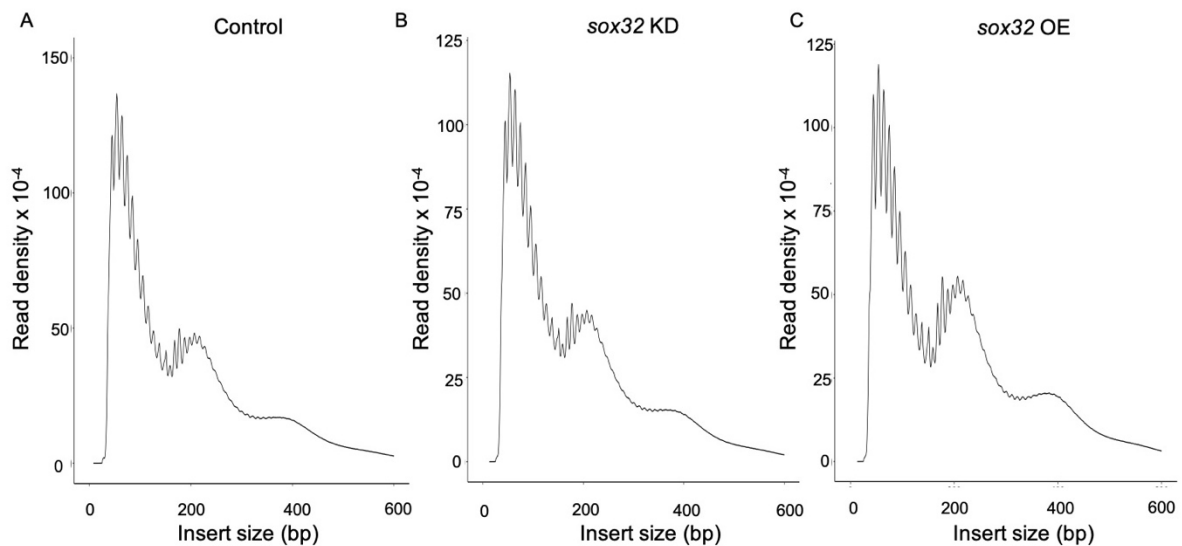


Figure 3-4: insert sizes for amplified ATAC-seq from control, *sox32* KD and *sox32* OE. Insert size distributions were determined by high-throughput sequencing of libraries from (A) control, (B) *sox32* KD, and (C) *sox32* OE embryos. A typical insert size profile is described in Buenrostro *et al.*, (2015), which is a histogram plot of insert sizes measured by mapped read densities. To determine whether our ATAC-seq data contained a typical distribution of fragment sizes, a distribution of all the mapped reads for each sample was plotted. This profile shows the 10.4 base pair periodicity, along with three nucleosomal regions along the distribution; the first demonstrating the nucleosomal depleted reads, the second mononucleosomal reads, i.e.: reads that span a single nucleosome (~200 bp peak) and third dinucleosomal reads, i.e.: reads that span two nucleosomes (~400 bp). Insert sizes were extracted from merged replicate samples for each condition.

We next sought to investigate whether the samples exhibit similar patterns of chromatin accessibility relative to genomic features. To find regions of accessible chromatin for each sample, the peak caller Model-based Analysis of ChIP-seq (MACS2) was used. Three biological replicates were used for each condition (Table 3.1). Following peak calling, individual peaks from each biological replicate were merged and these were used as a representative of the total number peaks for each condition. These merged peaks are shown in Table 3.2 and were used to identify putative enhancer locations relative to the nearest gene and were also used to visualise the accessibility of endoderm specific putative enhancers using integrative genome viewer (IGV) in all three sample groups.

Table 3.1: Number of called peaks for each biological replicate per condition

Sample	Number of called peaks
Control 1	92,493
Control 2	83,667
Control 3	115,119
<i>sox32</i> KD 1	67,451
<i>sox32</i> KD 2	80,221

sox32 KD 3	94,218
sox32 OE 1	81,848
sox32 OE 2	83,466
sox32 OE 3	113,188

Table 3.2: Representative peaks from each sample condition

Sample	Number of called peaks
Control	211,994
sox32 KD	193,889
sox32 OE	196,412

ATAC-seq signals which correspond to the level of chromatin accessibility can be used to identify regions of open chromatin on a genome-wide scale. As observed below, called peaks from all three sample groups appeared to follow a similar distribution relative to genomic features (Figure 3-5). These results suggest that *sox32* KD has no effect on the chromatin accessibility landscape since it is expressed in a minor population of cells, and its affects are likely negligible in the context of the whole embryo. But the results also suggest that *sox32* OE is sufficient but not necessary for the enhanced chromatin accessibility observed in the endoderm and likely suggests that another factor is involved in inducing ectopic endoderm formation within the embryo.

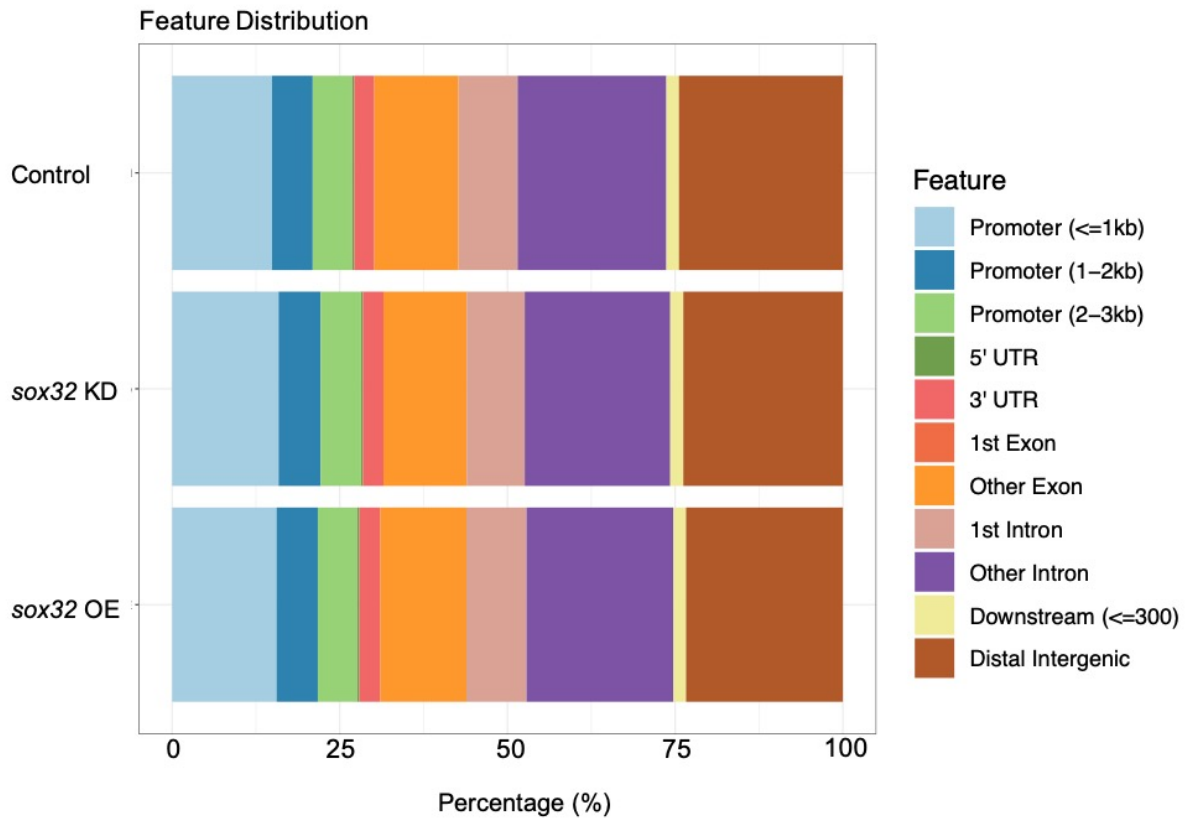


Figure 3-5: Distribution of ATAC-seq peaks from control, *sox32* KD and *sox32* OE across genomic regions with respect to the nearest gene. Promoters are located within -3 kb to + 3 kb of transcription start site (TSS). Peaks were annotated based on the following priority: promoter, 5'UTR, 3'UTR, exon, intron, downstream and intergenic, where downstream is defined as the downstream of gene end.

3.5.1 *Sox32* OE has different accessibility profile to control

The aim of the project is to identify active enhancers that play a role in endoderm formation and one way this can be achieved is through identifying accessible regions in endoderm progenitors. In order to identify endoderm-specific regulatory regions, we sought to look for regions that exhibit greater accessibility in *sox32* OE compared to control. The identification of differential accessible regions (DARs) between *sox32* OE and control is essential in determining the differential activity of regulatory elements that play a role in the endoderm. To find DARs, a programme called Diffbind was used (Stark, R and Brown, 2016). Diffbind uses mapped read numbers and called peaks from MACS2 to find DARs between two sets of sample groups. Firstly, called peaks from all sample groups are read into Diffbind. Then Diffbind merges all overlapping peaks and derives a single set of unique peaks which covers all supplied peaks. Thus, the presence of a peak in multiple samples means it is likely an indication that it is a real peak. Finally, Diffbind draws a heatmap based on MACS2 scores for each peak-set within all samples. The more similar the MACS2 scores, the stronger the correlation between the samples, and thus, this peak caller score helps cluster samples based

on similarity between all called peaks. As shown below, there is a strong correlation between *sox32* KD and control, while *sox32* OE is more distinct since each peak has a different MACS2 score (Figure 3-6).

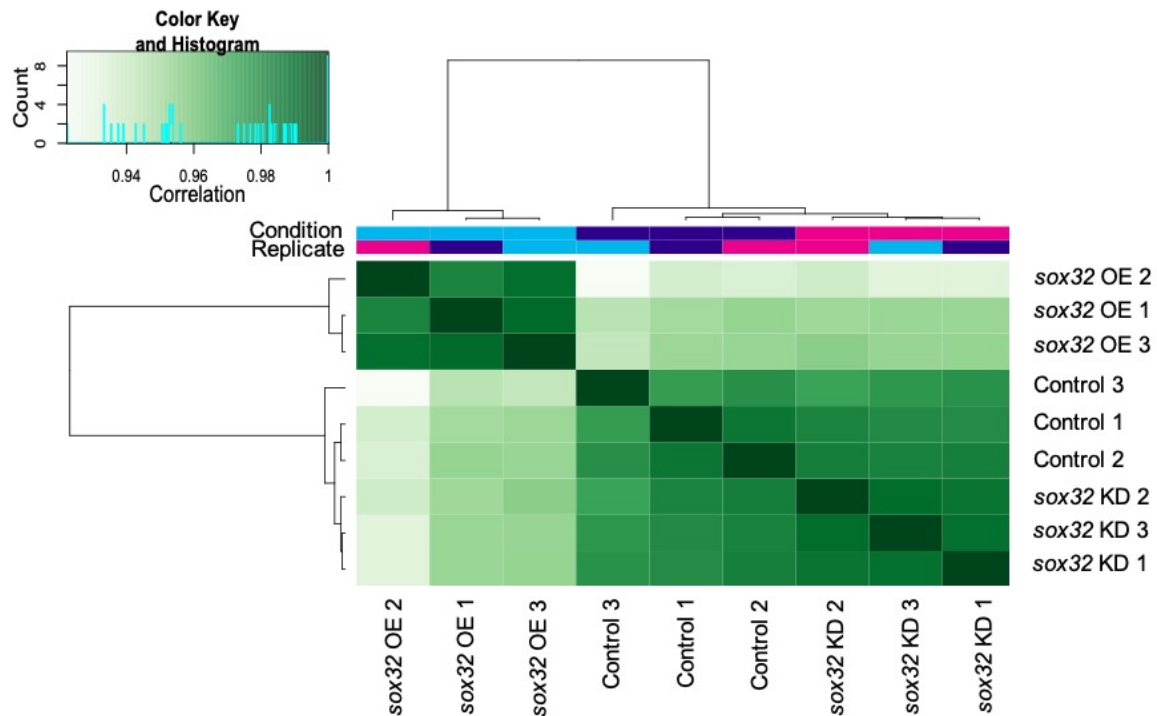


Figure 3-6: A clustered heatmap of Pearson correlation coefficients based on MACS2 peak scores between control, *sox32* KD and *sox32* OE samples. The correlation is depicted by varying colour intensities and the samples are grouped based on hierarchical clustering. The darker the green, the more similar the MACS2 scores between samples.

Following this, another correlation heatmap is plotted from the total number of mapped reads for each peak-set within a sample (Figure 3-7), rather than confidence scores for only those peaks called by MACS2 in each sample (Figure 3-6). For this step, Diffbind uses the peaksets that were obtained from Figure 3-6 and calculates the number of reads that map each peakset within each sample group. These regions are referred to as consensus peak sets, and the number of consensus peaksets was identified to be 97,111. Following this, a new correlation heatmap is drawn based on the number of reads that map each region. The number of mapped reads can be found in Table 3.3. Differences in the number of reads likely suggests differential accessibility between the different sample groups. Thus, this step helps identify statistically significant differential accessible regions (DARs) between the different samples but also samples that have similar accessibility profiles. For instance, based on hierarchical clustering, the number of reads that map each consensus region is similar between all *sox32* KD and control biological replicates while the *sox32* OE samples appear to be distinct (Figure

3-7). This means that *sox32* KD and control have similar accessibility profiles since they cluster together, while *sox32* OE does not and as a result, appears to cluster alone.

Table 3.3: Total number of mapped read counts

Sample	Replica number	Total number of mapped read counts
Control	1	57,278,982
	2	52,489,324
	3	66,103,240
<i>sox32</i> KD	1	48,248,090
	2	47,630,398
	3	61,603,986
<i>sox32</i> OE	1	49,732,812
	2	48,286,218
	3	75,497,540

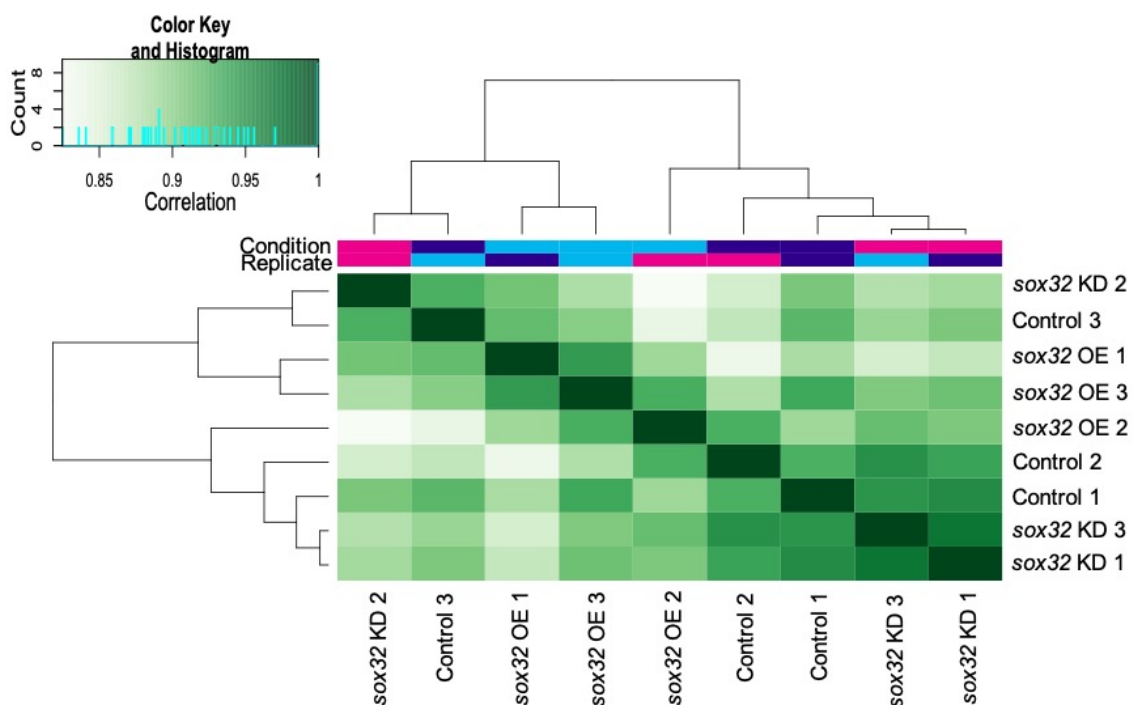


Figure 3-7: A Pearson Correlation coefficient heatmap based on read counts that map each consensus region between control, *sox32* KD and *sox32* OE. Correlation is depicted by varying colour intensities and the samples are grouped based on hierarchical clustering. The darker the shade of green, the more similar the number of reads that map to each genomic interval within the sample.

The core function of Diffbind is to identify statistically significantly differentially accessible regions between two sets of sample groups. The differential analysis is carried out by first normalising the data using DESeq2. It does this by normalising the number of reads extracted from each BAM file so that a normalisation factor is computed to weight each sample so that all samples have the same library size. For DESeq2 this is accomplished by dividing the number of reads by the mean library size. Thus, DESeq2 corrects for differences in library size between all samples. This analysis revealed six DARs between control and *sox32* KD (Figure 3-8A and Table 7.1), and 18,237 DARs between control and *sox32* OE (Figure 3-8B), some of which can be viewed in Table 7.2 . Furthermore, out of 18,217 regions, 16,120 regions exhibited higher accessibility in *sox32* OE compared to control. This therefore highlights the similarity in the accessibility profile between control and *sox32* KD, while *sox32* OE exhibits greater accessibility relative to control.

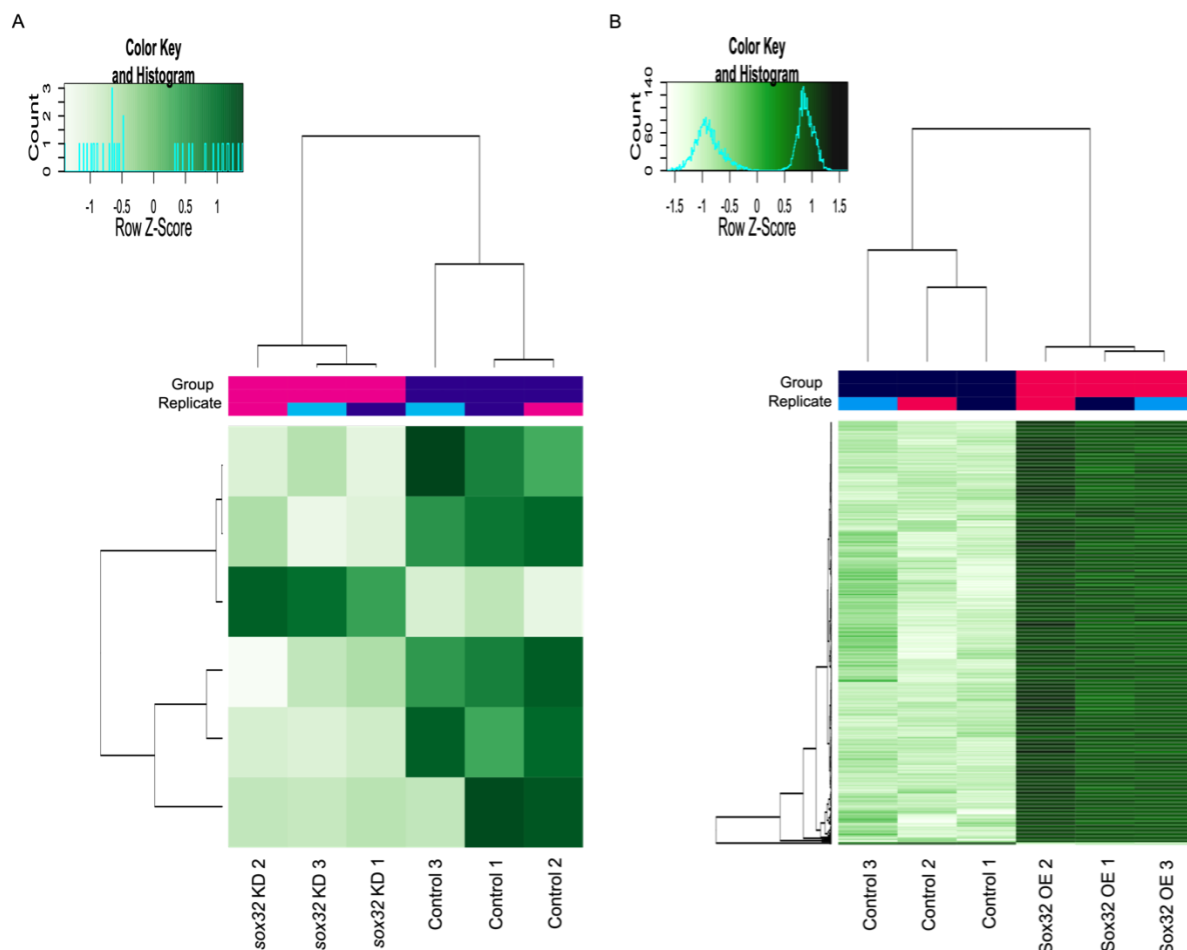


Figure 3-8: A differential accessibility heatmap between two sets of sample groups. A differential heatmap can be used to illustrate the differences in DARs between (A) *sox32* KD and control and (B) *sox32* OE and control. All reads from each sample were first normalised by DESeq2 which calculates a geometric mean and the reads in each sample are divided by this mean, which therefore corrects for library size. Each genomic region is assigned

a Pvalue and FDR indicating confidence that they are differentially accessible in each sample. FDR cut-off of 0.01 was used for *sox32* OE vs control to reduce list of accessible regions to a more manageable number for downstream analysis. However, changing input parameters to a number of FDR cut-offs (0.05 to 0.0001) did not change final results. The FDR cut-off was 0.05 for *sox32* KD vs control. The Pvalue cut-off for *sox32* KD vs control was 2×10^{-6} and for *sox32* OE vs control was 0.01. Heatmaps were clustered on samples (columns) and accessible regions (rows), and rows were scaled based on Z-scoring. Regions of higher accessibility are shown in Dark green, and light green regions represent low accessibility.

An example of regulatory elements that show enhanced accessibility in *sox32* OE compared to *sox32* KD and control is shown below for *cxcr4a* (Figure 3-9), which has established roles in the endoderm and DFCs (Mizoguchi *et al.*, 2008; Stückemann *et al.*, 2012; Liu *et al.*, 2019).

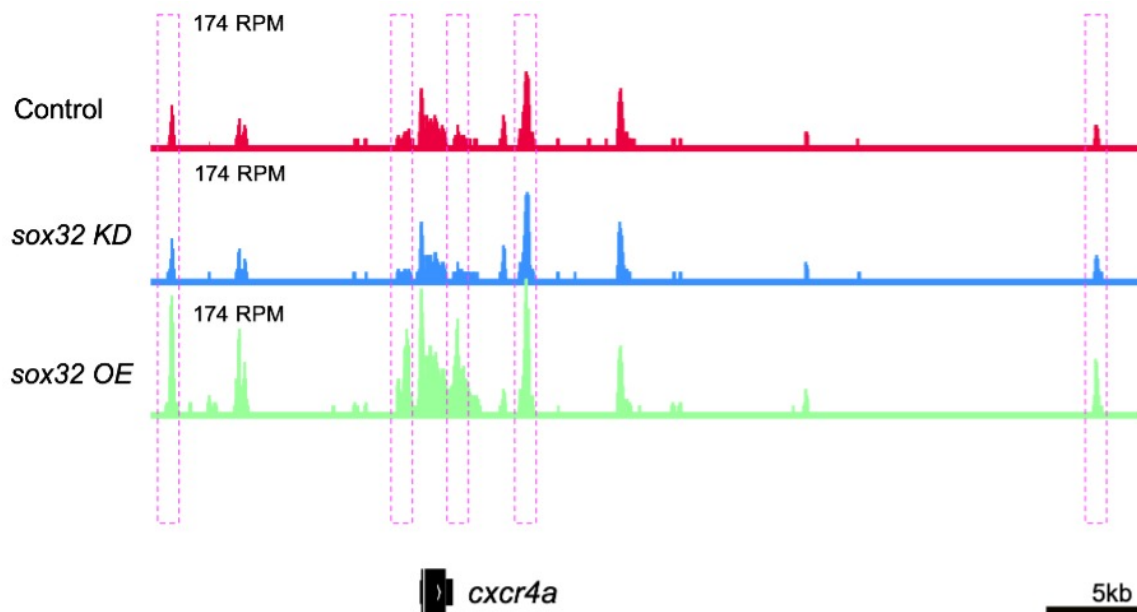


Figure 3-9: ATAC-seq read densities showing regions of enhanced accessibility proximal to *cxcr4a* in *sox32* OE, which is a marker of endoderm and DFC fates. Reads were scaled to adjust for library size. Purple dashed regions are likely putative enhancers of *cxcr4a*. RPM = reads per million.

3.5.2 Overexpression of *sox32* does not drive artificial accessibility at loci that are not accessible in control embryos

For this experiment, *sox32*-overexpressing *Tg(sox17:EGFP)* embryos were used to help identify endoderm-specific CRMs. To ensure that overexpression of *sox32* does not induce artefacts, i.e.: lead to opening of chromatin in regions that are not accessible in control or expression of genes that are not usually expressed in the endoderm, a program called SeqMINER was used (Krebs *et al.*, 2010). SeqMINER allows qualitative and quantitative comparisons between an input set of genomic coordinates and multiple high throughput datasets. For this analysis, we compared mapped reads from control, *sox32* KD and *sox32* OE across control and *sox32* OE DARs. SeqMINER first normalises the number of mapped

reads from each sample group and subsequently uses a clustering procedure (*k*-means) to organise DARs presenting similar read densities within a user-defined window, which in this case was +/- 5kb surrounding each DAR centre. 10 clusters were chosen for this analysis. As observed for *sox32* OE and control DARs, the majority of regions that exhibit higher accessibility in *sox32* OE were also observed in control and *sox32* KD (Figure 3-10). This suggests that *sox32* OE opens chromatin in more cells in the same regions that are also accessible in control, but also suggests that it does not drive artificial accessibility at loci that are not accessible in control embryos. Thus, based on this result, it is likely that *sox32* is sufficient but not necessary to induce genome-wide changes to the chromatin landscape during endoderm formation.

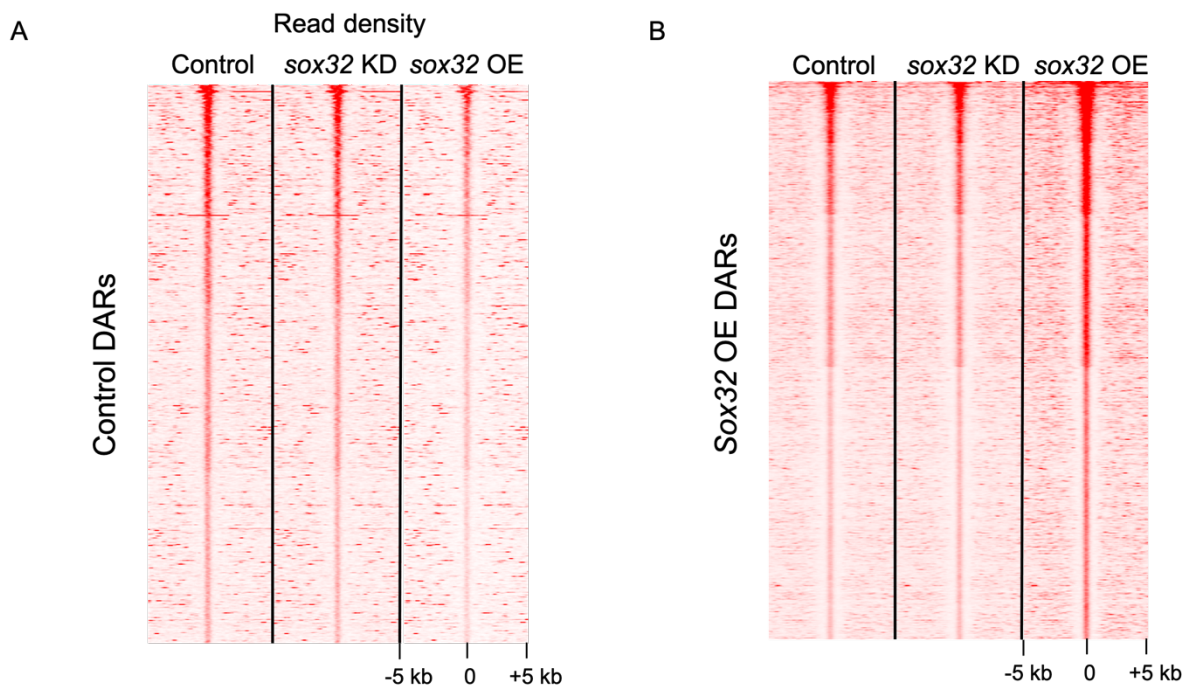


Figure 3-10: Heatmaps showing read densities across control, *sox32* KD and *sox32* OE samples. (A) Heatmaps display mapped reads from control, *sox32* KD and *sox32* OE surrounding control DARs. (B) Heatmaps display mapped reads from control, *sox32* KD and *sox32* OE surrounding *sox32* OE DARs. Tag or read densities with a window of +/- 5kb across each DAR centre was collected for each ATAC-seq dataset. Plots are from Dr Andrew Nelson.

3.5.3 *Sox32* OE results in enhanced chromatin accessibility proximal to genes associated with endoderm and DFC formation.

Though extensive effort has been used to study the transcriptional network governing endoderm formation, the regulatory elements required to induce endoderm gene expression is largely unknown. To explore the relationship between accessible regions and adjacent genes and to find enhancers of endoderm formation, Genomic Regions Enrichment of Annotations Tool (GREAT) was used (McLean *et al.*, 2010). GREAT identifies proximal genes

using an input of putative CRMs. Using GREAT, Sox32 OE DARs were found to be proximal to genes involved in DFC and endoderm formation and thus were associated with terms involved in the formation of both lineages (Figure 3-11A) vs control (Figure 3-11B). Some of these genes can be viewed in Table 7.3. Upon further observation, terms such as mesoderm and hypoblast were also over-represented in *sox32* OE. This is likely because only a few endoderm specific genes exist, and many endoderm expressed genes also play a role in the mesoderm population since they share common precursor cells (Tada *et al.*, 2005). In addition, hypoblast gives rise to definitive endoderm and mesoderm populations (Kimmel *et al.*, 1995). Examples of endoderm genes that also associated with mesoderm population include *tbx16*, *gata5*, and *mixl1*, which are Nodal-induced and required for endoderm formation (Henry and Melton, 1998; Reiter, Kikuchi and Stainier, 2001a; Nelson *et al.*, 2017) and *cxcr4a* which is required for correct endoderm migration and proliferation and KV formation (Stückemann *et al.*, 2012; Liu *et al.*, 2019). Additional endoderm genes which are also associated with the hypoblast include Nodal ligands *ndr1* and *ndr2*, but also *prdm1a* which is involved in endoderm formation (Bennett *et al.*, 2007; Tseng *et al.*, 2011) and *foxa2*, a marker of endoderm progenitors (Alexander *et al.*, 1999). An additional overrepresented anatomical term in *sox32* OE is yolk syncytial layer (YSL). This is likely because *sox32* expression also appears in the YSL (Kikuchi *et al.*, 2001).

By contrast, control DARs are proximal to genes expressed in the ectoderm and therefore not where endogenous *sox32* is expressed (Figure 3-11B). Some of these genes can be viewed in Table 7.4. No significant terms were annotated near *sox32* KD DARs. These results suggest that *sox32* OE DARs are proximal to genes that play a role in endoderm and/or KV formation.

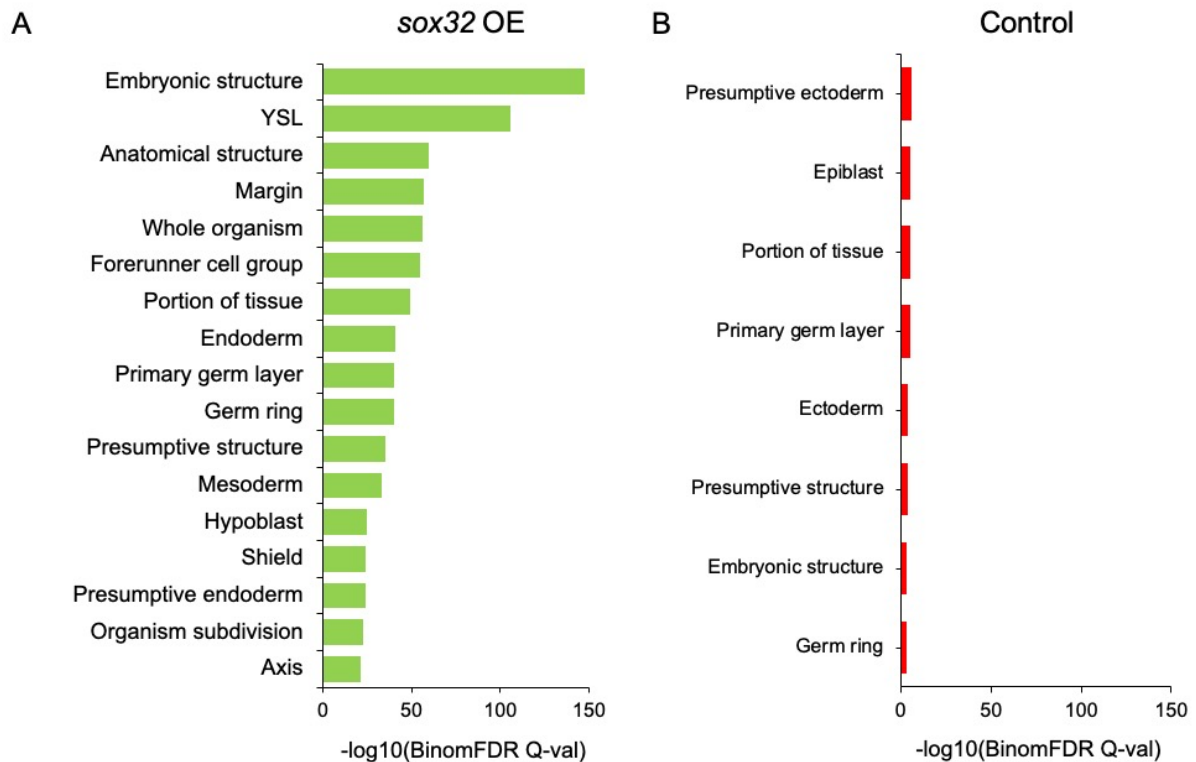


Figure 3-11: Over-represented anatomical terms at 6 h.p.f. from GREAT for *sox32* OE and control. Anatomical terms are represented as a bar group for (A) *sox32* OE and (B) control. GREAT assigns each gene a regulatory domain that extends in both directions to the midway point of TSS of two genes, but no more than 1Mb. Any genomic coordinates that fall into the regulatory domain of the gene will be assigned to that gene. GREAT calculates statistics by associating DARs with proximal genes and applying gene annotations to DARs. DAR coordinates were converted to danRer7 genome build for this analysis, despite this, genes expected to be expressed in the endoderm were over-represented in *sox32* OE. FDR = False discovery rate.

3.5.4 *Sox32* OE DARs are enriched for binding motifs of transcription factors that play a role in the endoderm and DFCs

To explore the functional significance of putative enhancers identified for *sox32* OE and to validate their relationship to endoderm and DFC formation, transcription factor (TF) binding motifs were investigated using Hypergeometric Optimization of Motif Enrichment (HOMER) (Heinz *et al.*, 2010). HOMER was used to identify statistically significant over-represented TF motifs present in *sox32* OE DARs relative to control DARs. Analysis of the enriched motifs in *sox32* OE revealed binding motifs of TFs with known roles in the endoderm and DFCs, and these include Sox, T-box, Forkhead, GATA and Homeobox proteins (Figure 3-12A). These motifs are also shown in Table 7.5. On the other hand, control DARs were enriched for binding motifs of pluripotent factors such as OCT4-SOX2-TCF-NANOG (Figure 3-12B) compared to *sox32* OE DARs, which play important roles in maintaining embryonic stem cells in a self-renewing and pluripotent state (Kim *et al.*, 2008). These motifs can also be viewed in Table 7.6.

Interestingly, Sox17 binding sites were also present in control DARs, though it was more enriched in *sox32* OE DARs. One hypothesis for this is that Sox17 may be playing a role in inhibiting expression of endodermal genes in non-endodermal lineages. But another possibility is that the motif for Sox17 could be representing other Sox members. Indeed HOMER does not have the position weight matrix for several Sox members including Sox5, yet it plays an essential role in the ectoderm since it is required for controlling laminar position, migration and layer-specific pattern of subplate and deep-layer neocortical neurons (Kwan *et al.*, 2008).

Notably, motifs of pluripotent factors Oct4, the closest mammalian homologue of Pou5f3, and also Nanog exhibit similar enrichment in both *sox32* OE and control DARs. This may suggest that Pou5f3 and Nanog play important roles in driving formation of endoderm lineages in *sox32* OE embryos and also driving other lineages in control embryos.

Taken together our results suggest that *sox32* OE DARs are enriched for binding motifs of TFs that either play a role in the endoderm or KV formation, while control DARs are enriched for motifs of pluripotency factors that together help drive the formation of other lineages during early zebrafish development.



Figure 3-12: HOMER motif analysis showing over-represented TF motifs in *sox32* OE or control DARs. (A) Over-represented TF binding motifs in *sox32* OE DARs were associated with endoderm and DFCs as opposed to control DARs (B). TF Motifs in *sox32* OE DARs were selected based on their role in the endoderm or DFCs. Motifs are ranked by statistical significance.

3.5.5 Sox32 OE DARs show concordance with *Tbxta* and *Tbx16* bound regions

Though there is extensive information regarding factors that act upstream of *sox32* (Stainier, 2002; Zorn and Wells, 2007, 2009; Tremblay, 2010; Nelson *et al.*, 2014, 2017), there is little knowledge on the factors that bind to enhancers downstream of *sox32*. To investigate transcription factors that may bind to *sox32*-regulated regions, *sox32* OE DARs were compared to published ChIP-seq data for Nanog at 3.3. h.p.f. (Xu *et al.*, 2012), *Tbxta* and *Tbx16* at 8-8.5 h.p.f. (Nelson *et al.*, 2017) and Pou5f3 at 5 h.p.f. (Leichsenring *et al.*, 2013). All the above factors induce endoderm and/or DFC formation upstream of *sox32* and were chosen since there are ChIP-seq data available for them. While these ChIP-seq experiments were performed on embryos from different stages of development, they help determine whether endoderm and DFCs/KV TFs act on regions that are likely regulated by *sox32* during development.

To begin, coordinates from *sox32* OE and control DARs were compared to Nanog ChIP-seq. (Liu *et al.*, 2016). Nanog is maternally deposited and is present in all embryonic and extra-embryonic cells (Xu *et al.*, 2012). It is also required for YSL formation (Gagnon, Obbad and Schier, 2018), which plays a critical role in cell fate decisions and morphogenesis of the early germ layers (Carvalho and Heisenberg, 2010). As shown below, Nanog peaks were found to overlap more with control DARs compared with *sox32* OE DARs (10% vs 4%) (Figure 3-13). Though only 4% of *sox32* OE DARs overlap with Nanog peaks, these common regions were found proximal to genes that are either expressed in the endoderm or DFCs, as shown in Table 7.7. For example, *vgll4l* is expressed in DFCs at 6 h.p.f. (Xue *et al.*, 2018) and is essential for KV formation since loss of *vgll4l* expression results in significant reduction in the expression of genes involved in DFC/KV formation, including *tbxta*, *tbx16*, *sox32* and *sox17* (Fillatre *et al.*, 2019). Also, *fscn1a* is required for Nodal signalling and loss of the gene disrupts endoderm formation in zebrafish and mammals (Liu *et al.*, 2016). However, these common regions were also proximal to genes that play a role in nervous system development, and these include *ntn1a*, *gbx2* and *irx6a* as observed by Gene Ontology (GO) analysis using DAVID (Huang, Sherman and Lempicki, 2009b, 2009a) (Figure 7-1A). In contrast, common regions between Nanog peaks and control DARs were proximal to genes that play a role in nervous system and tissue development, including, *foxg1a*, *isl1* and *rock2a* but no endodermal terms were observed by GO analysis (Huang, Sherman and Lempicki, 2009b, 2009a) (Figure 7-1B). These genes can also be viewed in Table 7.8. Thus, since there is a considerable overlap between Nanog peaks and control DARs, it is likely that Nanog plays a role in driving formation of cell types that are more abundant in control compared to *sox32* OE embryos. However, the results also suggest that Nanog plays a role in *sox32* OE by driving endoderm formation but also may play a role in silencing non-endodermal tissue formation including the formation of the central nervous system.

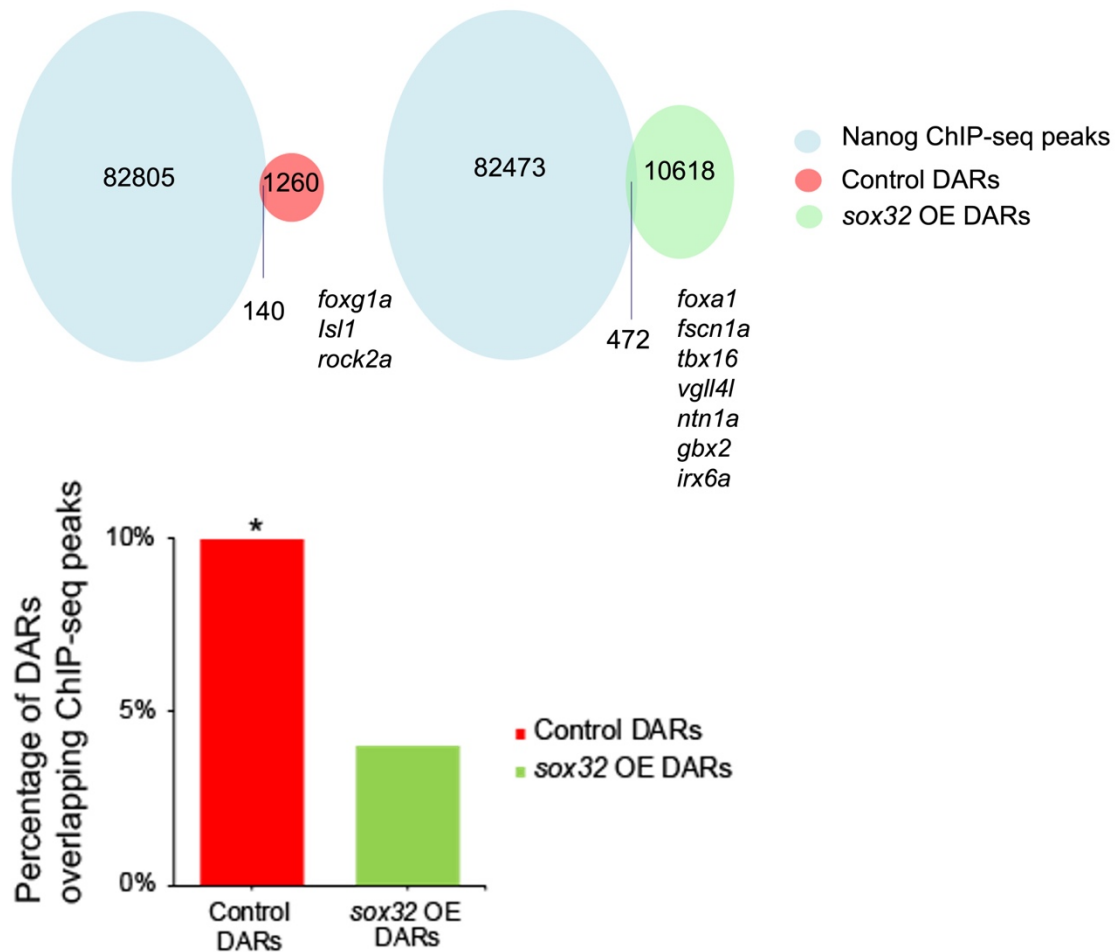


Figure 3-13: Comparison of control and *sox32* OE DARs with Nanog ChIP-seq peaks. Significant correlation was observed between Nanog ChIP-seq peaks and control DARs. Nanog ChIP-seq was performed at (3.3. h.p.f.). Bar graph represents percentage of DARs overlapping ChIP-seq peaks. * $p = 1 \times 10^{-4}$, chi-square test. Genes proximal to common regions between Nanog ChIP-peaks and control/ *sox32* OE DARs are displayed.

On the other hand, comparison of coordinates from *sox32* OE and control DARs with *Tbx16* and *Tbx16a* ChIP-seq peaks revealed that *sox32* OE DARs exhibit more overlap with *Tbx16* ChIP-seq sites compared with control DARs (4% vs 1%). There is also more overlap between *sox32* OE DARs and *Tbx16a* ChIP-seq peaks than with control DARs (5% vs 2%) (Figure 3-14A-C). In addition, these common regions were found proximal to genes that either regulate endoderm or DFC/KV formation in zebrafish including *gata5*, an essential endoderm regulator (Reiter, Kikuchi and Stainier, 2001a), but also *foxa1*, 2 and 3 which play a role in liver formation and glucose homeostasis. These common regions were also found proximal to *cxcr4a*, which is required for migration and proliferation of endoderm cells (Stückemann *et al.*, 2012) as well as laterality determination (Liu *et al.*, 2019), and *yap1* which is essential for endoderm cell survival (Fukui *et al.*, 2014) and may combine with other factors to regulate laterality in zebrafish (Fillatre *et al.*, 2019). These gene lists can be viewed in Table 7.9 and Table 7.11. However, these common regions were also proximal to genes that play a role in the formation

of other tissues. For instance, common regions between *sox32* OE DARs and Tbx16 ChIP-seq peaks were proximal to genes such as *efnb1* and *ntn1a* which are involved in nervous system development but also *nr2f2* and *scube3* which are involved in tissue development according to GO analysis using DAVID (Huang, Sherman and Lempicki, 2009b, 2009a) (Figure 7-2A). Similarly, common regions between *sox32* OE DARs and Tbx16 ChIP-seq peaks were proximal to genes involved in tissue and heart morphogenesis, which includes *smad5* and *ntn1a* but were also proximal to genes that are involved in the development but also in the negative regulation of the central nervous system, such as *sema3ab*, *sema3fa* and *sema3fb* as observed by GO analysis (Huang, Sherman and Lempicki, 2009b, 2009a) (Figure 7-2B). No GO terms were observed when controls DARs were compared to Tbx16 and Tbx16 peaks (Table 7.10 and Table 7.12). These results suggest that Tbx16 and Tbx16 bind accessible regions downstream of *sox32* and that these regions are likely important for expression of important endodermal genes, but also that they may be important for inhibiting non-endodermal tissue formation.

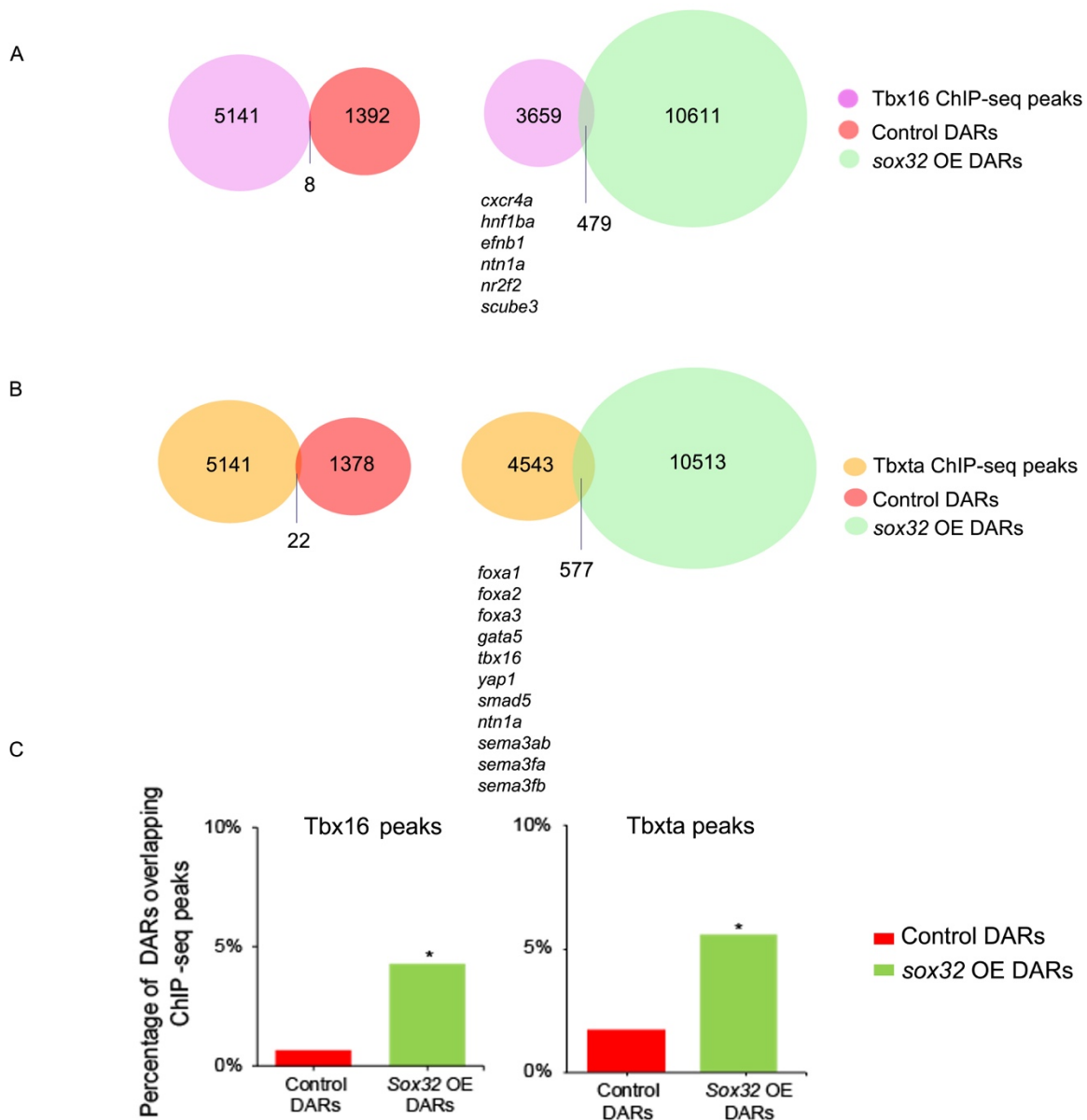


Figure 3-14: Comparison of control and *sox32* OE DARs with Tbx16 and Tbx16 ChIP-seq peaks. Significant correlation was observed between (A) Tbx16 and (B) Tbx16 ChIP-seq peaks with *sox32* OE DARs. Tbx16 and Tbx16 ChIP-seq were performed at 8-8.5 h.p.f.. (C) Bar graph represents percentage of DARs overlapping ChIP-seq peaks. * $p = 1 \times 10^{-4}$, chi-square test. Genes proximal to common regions between Tbx16/ Tbx16 ChIP-peaks and *sox32* OE DARs are displayed.

By contrast, there was no difference in the number of common regions between control or *sox32* OE DARs with Pou5f3 peaks (Figure 3-15), a surprising result given that Sox32 requires Pou5f3 to induce expression of *sox17*. This is perhaps because Pou5f3 may not be required to maintain endoderm downstream of Sox32. Thus, though Pou5f3 is essential for endoderm formation, once endoderm is induced and *sox17* is expressed, Pou5f3 may no longer be required and that other factors are needed to maintain endoderm.



Figure 3-15: Comparison of control and *sox32* OE DARs with Pou5f3 ChIP-seq peaks. No correlation was found between control or *sox32* DARs with Pou5f3 bound peaks.

Our finding suggests that *Tbx16*, *Tbx1a* and *Nanog* may induce expression of endoderm genes by binding to accessible regions downstream of *sox32*.

3.6 Overexpression of *sox32* results in upregulation of endoderm and DFC target genes

Based on the ATAC-seq analysis, *sox32* OE DARs show enhanced accessibility proximal to genes that play a role in endoderm and DFC formation but also other cell types. To test if this is a true reflection of the biology, RNA-seq was performed. RNA-seq libraries were generated from *sox32* OE and control embryos and genes that were upregulated for both sample groups were identified. Only significant genes were selected for all RNA-seq downstream analysis (FDR < 0.05) and comparison between the two sample groups revealed that the expression profile differs between control and *sox32* OE (Figure 3-16A), suggesting that these represent two different cell identities. RNA-seq was also used to investigate expression of a number of endoderm and DFC genes. These genes were identified by performing GO analysis using DAVID (Huang, Sherman and Lempicki, 2009b, 2009a) on transcripts that were upregulated in *sox32* OE embryos and looking at genes that were associated with endoderm and/or DFCs development. The total list of upregulated transcripts can be in Table 7.13 in which genes were ordered based on Wald-statistics, which is defined as the $\log_2\text{foldchange}$ divided by its standard error and is used as part of RNA-seq analysis for computing Pvalues.

Based on GO analysis, Kupffer's vesicle development was the most over-represented term while the regulation of endodermal cell fate specification was the fifth most over-represented term (Figure 7-3). Some of the genes associated with these terms can be viewed in Figure 3-16B, and these were selected based on their importance in the development of endoderm and left/right asymmetry establishment and also that they were significantly upregulated in *sox32* OE vs control embryos and these include *sox17*, *cxcr4a* and *sox32*, as previously mentioned. Other genes associated with endoderm development based on GO analysis

include *foxa2* and *ackr3b*. Interestingly, although *ackr3b* transcript is upregulated in *sox32* OE and is believed to act in parallel with *Cxcr4a* to promote correct migration and proliferation of endoderm cells (Tobia *et al.*, 2019), based on GO analysis, *ackr3b* is also believed to be important for neuron migration and angiogenesis. This suggests that genes important for endoderm formation likely play important roles in other lineages. On the other hand, genes associated with Kupffer's vesicle development and left/right asymmetry establishment based on GO analysis include *foxj1a* (Stubbs *et al.*, 2008; Yu *et al.*, 2008), *dnmt3bb.1*, *cftr* and *vgll4l* (Fillatre *et al.*, 2019). While a number of upregulated transcripts in *sox32* OE were associated with the endoderm and/or DFCs/KV, other transcripts were associated with other processes. For instance, genes including *htra1a* which is highly enriched (i.e.: found within top 50 upregulated transcripts) was associated with apoptotic process while *sema3fb* was associated with neural crest cell migration according to GO analysis. However, despite the fact that genes not associated with endoderm and KV/DFCs were also upregulated in *sox32* OE, the top three genes were *sox17*, *cxcr4a* and *vgll4l*, which are essential for either endoderm migration or KV formation, while other essential genes including, *sox32*, *dnmt3bb.1*, *foxj1a*, *ackr3b*, *rock2b*, *tbx16* and *foxa2* were in the top 150 upregulated transcripts for *sox32* OE vs control. Therefore, upregulated transcripts for *sox32* OE are in accordance with ATAC-seq analysis and that *sox32* OE is associated with endoderm and DFC/KV transcripts compared to control.

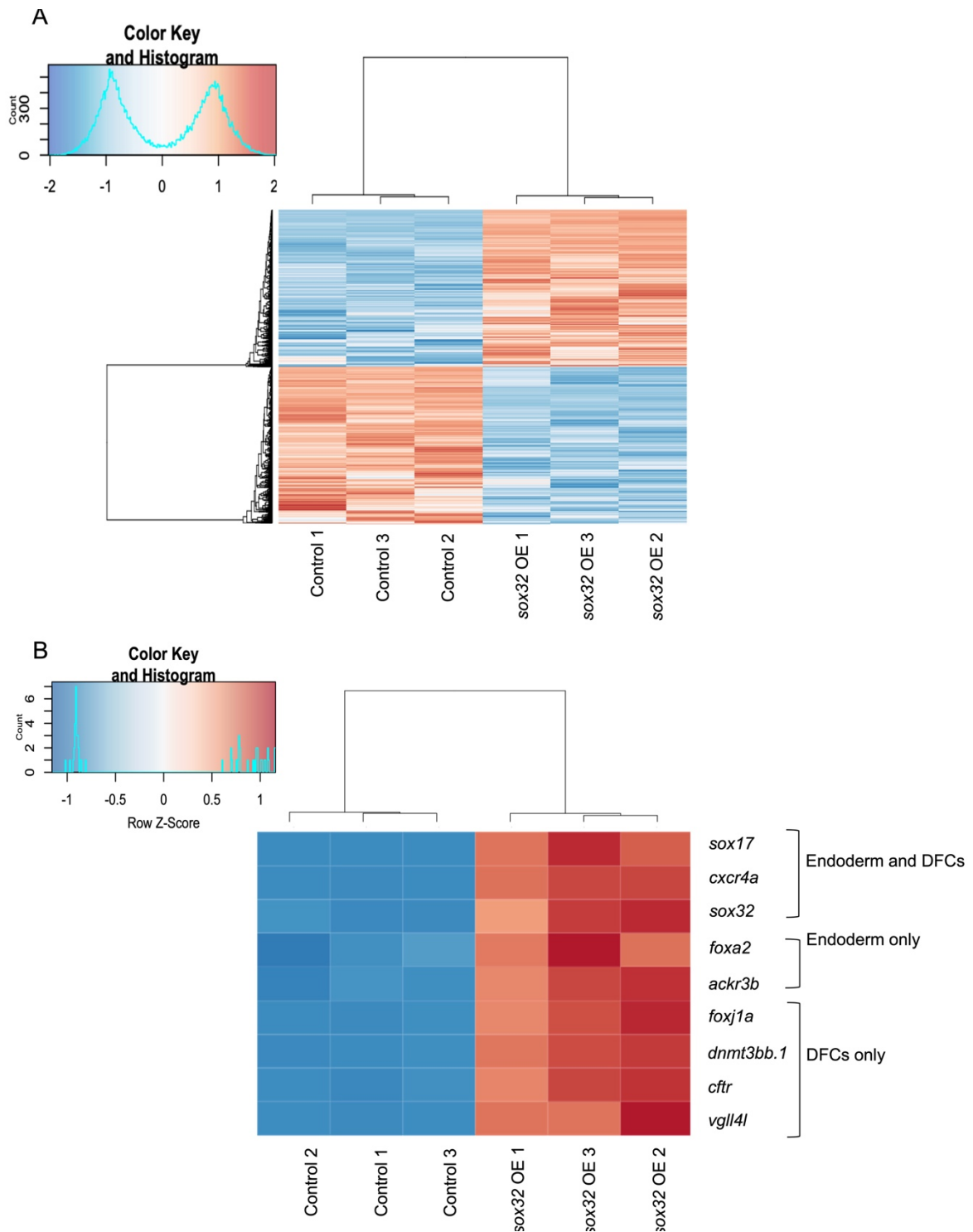


Figure 3-16: RNA-seq analysis comparison of genes differentially expressed between control and *sox32* OE embryos. A) Heatmap of significantly upregulated transcripts (FDR < 0.05) in control and *sox32* OE. B) Heatmap displays top ranked genes which were found to be either associated with the endoderm, DFCs, or both. 3,188 genes were upregulated in control while, 3,176 were upregulated in *sox32* OE. Hierarchical clustering shows that biological replicates cluster together while the gene expression profile is different between control and *sox32* OE sample. Three biological replicates were generated for each sample.

3.7 Sox32 OE upregulated transcripts significantly correlate with endoderm and DFC-enriched genes

According to GO and RNA-seq analysis, *sox32* OE was shown to be associated with DFCs and endoderm compared to control. However, while it is already known that overexpression of *sox32* results in ectopic endoderm formation, it is not yet known whether it also drives ectopic DFC formation. Moreover, the presence of mesoderm and ectoderm fates have yet to be explored in *sox32* OE embryos. Based on RNA-seq analysis, genes associated with KV formation were most enriched in *sox32* OE. To further confirm whether overexpression of *sox32* drives ectopic endoderm formation, and to explore the effects of *sox32* overexpression on the other germ layers and DFCs, genes upregulated in *sox32* OE embryos were compared to genes enriched in the ectoderm, mesoderm, ectoderm from different stages of development (6, 8, 10, 14, 18, 24 h.p.f.), but also DFCs (8 h.p.f.), from published single-cell RNA-seq data (Wagner *et al.*, 2018). As observed by Gene Set Enrichment Analysis (GSEA), genes upregulated in *sox32* OE were highly correlated with endoderm-enriched genes particularly at 6,8,10 (FDR $q \leq 5 \times 10^{-5}$, Pvalue $\leq 5 \times 10^{-5}$) and 14 h.p.f. (FDR $q \leq 5 \times 10^{-4}$, Pvalue $\leq 5 \times 10^{-4}$) but also DFCs at 8 h.p.f. (FDR $q \leq 5 \times 10^{-5}$, Pvalue $\leq 5 \times 10^{-5}$) (Figure 3-17A). The results therefore suggest that overexpression of *sox32* induces ectopic endoderm but also DFC formation.

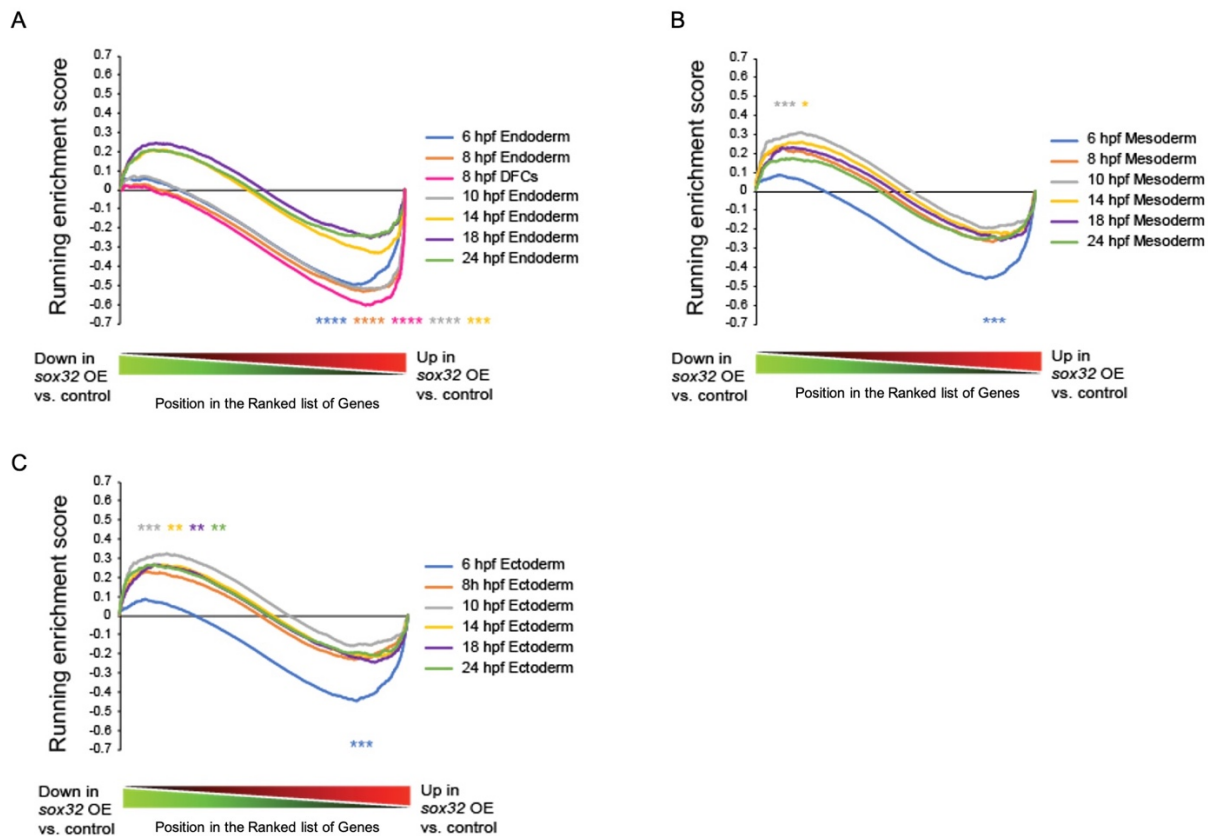


Figure 3-17: GSEA enrichment plots. GSEA was used for comparison of transcripts upregulated in *sox32* OE and control to genes enriched in (A) the endoderm and DFCs, (B) mesoderm and (C) ectoderm at different stages of development (6,8,10,14,18, 24 h.p.f.) from published single-cell RNA-seq data (Wagner *et al.*, 2018). ****False discovery rate (FDR) $q \leq 5 \times 10^{-5}$, Pvalue $\leq 5 \times 10^{-5}$; ***FDR $q \leq 5 \times 10^{-4}$ Pvalue $\leq 5 \times 10^{-4}$, **FDR $q \leq 5 \times 10^{-3}$, Pvalue $\leq 5 \times 10^{-3}$, *FDR $q \leq 5 \times 10^{-2}$, Pvalue $\leq 5 \times 10^{-2}$.

However, *sox32* OE upregulated genes also significantly correlated with mesoderm and ectoderm-enriched genes at 6 h.p.f. (FDR $q \leq 5 \times 10^{-4}$, Pvalue $\leq 5 \times 10^{-4}$) (Figure 3-17B-C), while control genes significantly correlated with mesoderm and ectoderm-enriched genes from later stages of development, particularly at 8 and 10 h.p.f. for mesoderm and 8, 10, 18 and 24 h.p.f. for ectoderm. To investigate the reason behind the significant correlation between *sox32* OE enriched genes and genes enriched in mesoderm and ectoderm genes at 6 h.p.f., genes enriched in endoderm were compared to mesoderm and ectoderm genes from single-cell RNA-seq data. This analysis revealed that ~80% and ~70% of genes associated with the endoderm were also associated with the mesoderm and ectoderm at 6 h.p.f. (Figure 3-18A-B), respectively, and are therefore not endoderm specific. Interestingly, approximately 84% and 80% of genes that are contributing to significant results in mesoderm and ectoderm, respectively, were found to be associated with the endoderm as observed by single-cell RNA-seq data, and these include *tbx16* and *cxcr4a*. This means that genes associated with the

endoderm at 6 h.p.f. are not unique to endoderm since they are also expressed in ectoderm and mesoderm and thus, may be required for early specification of all three germ layers.

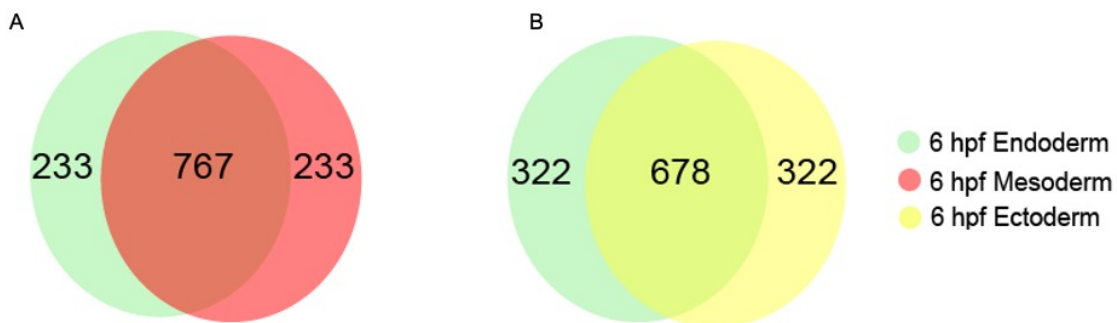


Figure 3-18: Comparison of genes enriched in the different germ layers at 6 h.p.f.. Venn diagrams showing overlap between (A) endoderm and mesoderm-enriched genes and (B) endoderm and ectoderm-enriched genes. Enriched genes were derived from published single-cell RNA-seq data (Wagner *et al.*, 2018).

The results demonstrate that *sox32* OE upregulated transcripts are highly concordant with genes associated with the endoderm and DFCs at 6 h.p.f. Since it has already been demonstrated that *sox32* OE induces ectopic endoderm formation, this may suggest that *sox32* OE also likely induces ectopic DFC formation although further downstream experiments may need to be performed to support this theory.

3.8 Changes in the chromatin landscape significantly correlate with changes in gene expression

To determine whether the change in chromatin landscape is correlated with changes in gene expression in *sox32* OE embryos, upregulated transcripts from control and *sox32* OE were compared to computationally-derived gene lists from GREAT, which were found proximal to control and *sox32* OE DARs. As observed by GSEA, genes upregulated in *sox32* OE significantly correlated with genes proximal to *sox32* OE DARs (FDR $q \leq 5 \times 10^{-5}$, Pvalue $\leq 5 \times 10^{-5}$) (Figure 3-19). In addition, genes upregulated in control embryos significantly correlated with genes proximal to control DARs (FDR $q \leq 5 \times 10^{-5}$, Pvalue $\leq 5 \times 10^{-5}$). Since there is a significant association between upregulated genes from *sox32* OE and markers of endoderm and DFCs from published single-cell RNA-seq data (Wagner *et al.*, 2018), and a significant association was observed between ATAC-seq and RNA-seq, this likely suggests that overexpression of *sox32* induces a genome-wide change in the chromatin landscape to aid in the induction of endoderm and DFC fates.

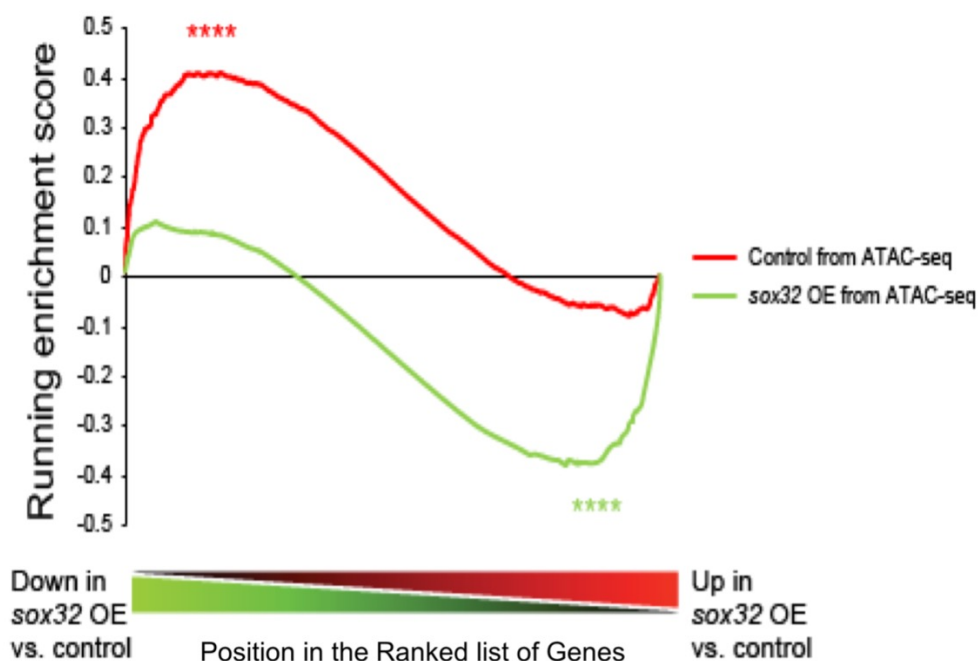


Figure 3-19: GSEA enrichment plots for comparison of proximal genes from control and *sox32* OE DARs to ranked gene lists from RNA-seq analysis. ****FDR $q \leq 5 \times 10^{-5}$, Pvalue $\leq 5 \times 10^{-2}$.

Example of genes showing enhanced accessibility in *sox32* OE and control embryos is shown below (Figure 3-20). *Cxcr4a* expression is upregulated in *sox32* OE embryos (Figure 3-20A), while ectoderm marker, *her3*, which is located in the hindbrain and midbrain (Thisse and Thisse, 2005), is highly expressed in control embryos (Figure 3-20B). Moreover, regions of

higher accessibility can also be observed proximal to *cxcr4a* locus for *sox32* OE and proximal to *her3* locus for control embryos, which are likely to be putative enhancers. This further suggests that changes in the chromatin landscape are causative of changes in gene expression in the different lineages.

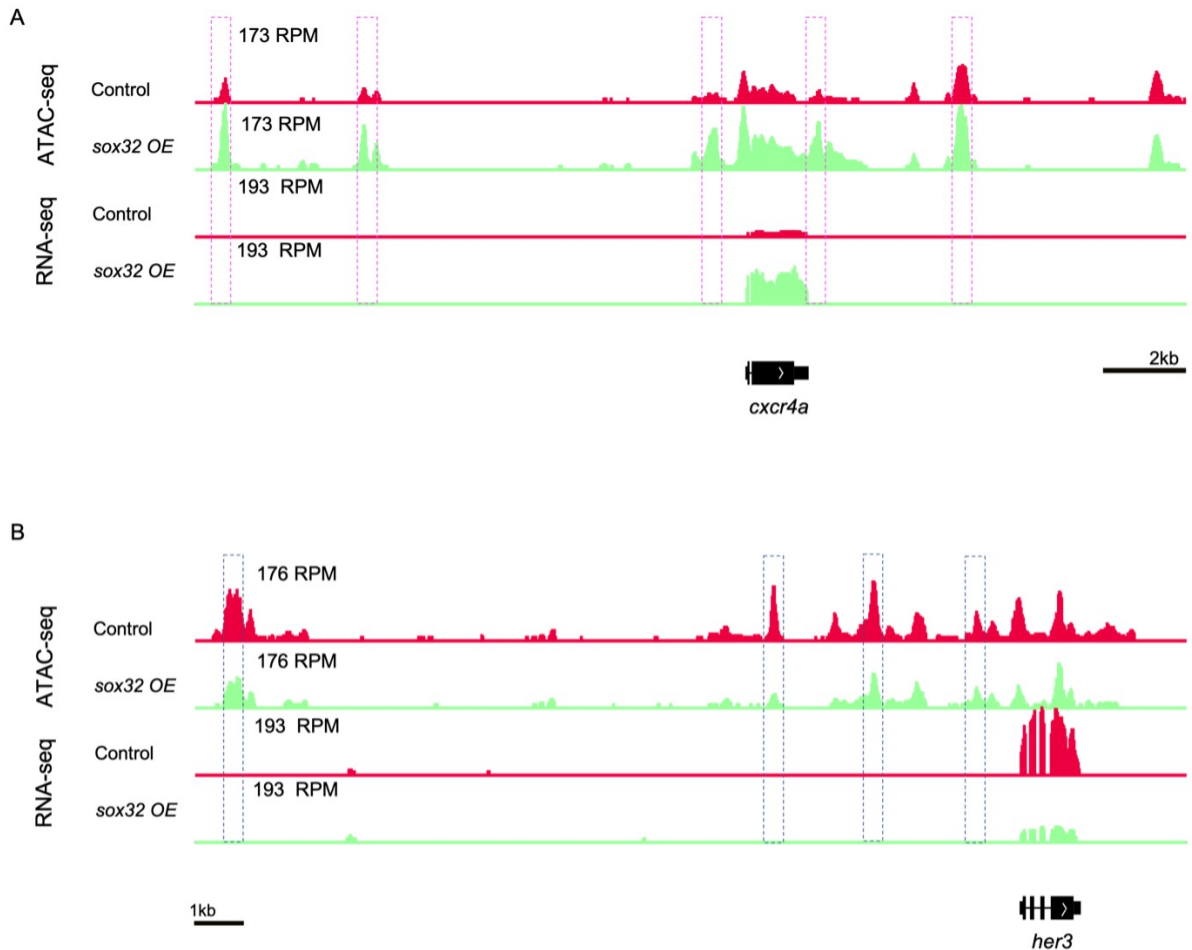


Figure 3-20: Changes to the chromatin landscape in *sox32* OE correlates with gene expression. (A) RNA-seq and ATAC-seq showing enhanced gene expression and increased chromatin accessibility around *cxcr4a* in *sox32* OE embryos which is an endoderm and DFC marker. (B) RNA-seq and ATAC-seq showing enhanced gene expression and increased chromatin accessibility around *her3* in control embryos, which is a marker of the nervous system. Putative enhancers for *cxcr4a* and *her3* are shown in pink and blue dashed lines, respectively. Reads were scaled to adjust for library size. RPM = reads per million.

3.9 Overexpression of *sox32* induces early expression of DFC markers prior to endoderm formation

The results thus far suggest that overexpression of *sox32* elicits changes to the chromatin landscape which results in ectopic endoderm and DFC formation, however the time point at which endoderm and DFCs are induced is not yet known. Previously, *sox32* OE was shown to significantly upregulate expression of Nodal genes, *ndr1* and *ndr2* (Kikuchi *et al.*, 2001), however, this was demonstrated by *in situ* hybridisation at one developmental time point, ~5.3 – 6 h.p.f, (Kikuchi *et al.*, 2001). In zebrafish, Nodal-signalling is believed to be activated after mid-blastula stage (~3 h.p.f) (Hagos and Dougan, 2007). The density array heatmaps shown in Figure 3-10 demonstrated that *sox32* OE results in enhanced accessibility in regions that are also accessible in control and *sox32* KD, which likely means that *sox32* is sufficient but not required for the enhanced accessibility observed around those regions. This therefore likely suggests that another factor or factors are required to induce ectopic endoderm and DFC formation. Thus, since it was previously shown that *sox32* OE significantly upregulates expression of *ndr1* and *ndr2* (Kikuchi *et al.*, 2001), it is likely that this enhanced accessibility and ectopic endoderm as well as DFC formation could be a result of enhanced expression of Nodal factors in *sox32* OE embryos.

To investigate the time point at which endoderm and DFCs are induced in *sox32* OE embryos, qRT-PCR analysis was performed on control and *sox32* OE embryos against a number of different endoderm and DFC markers at different stages of development. These stages include high stage (3.3 h.p.f), which is at the time of late zygotic genome activation, dome stage (4.3 h.p.f.) and shield stage (6 h.p.f.), which is the stage at which ATAC-seq and RNA-seq were performed. To further validate the experiment and ensure that *sox32* OE results in significant upregulation of endodermal markers (Kikuchi *et al.*, 2001), markers downstream of *sox32* were also investigated.

To validate the experiments, qRT-PCR was performed against markers downstream of *sox32*, including *foxa2*, the pan-endodermal and dorsal axial mesoderm marker and *sox17*, the marker of both endoderm and DFCs. *Foxa2* appears to be expressed at dome stage (4.3 h.p.f.) while *sox17* is detected between dome and 50% epiboly (5.25 h.p.f.) (White *et al.*, 2017) (Figure 3-21A). Overexpression of *sox32* results in upregulation of both *sox17* and *foxa2* at high stage relative to control, and both continue to be significantly upregulated in *sox32* OE throughout the three stages of development (Figure 3-21B). Therefore, the results show that overexpression of *sox32* results in earlier expression of *sox17* and *foxa2*. Since *sox32* is required for endoderm and DFC formation but not axial mesoderm, the results likely suggests

that *sox32* OE induces ectopic endoderm and DFC formation at an earlier time point than shield stage (6 h.p.f.).

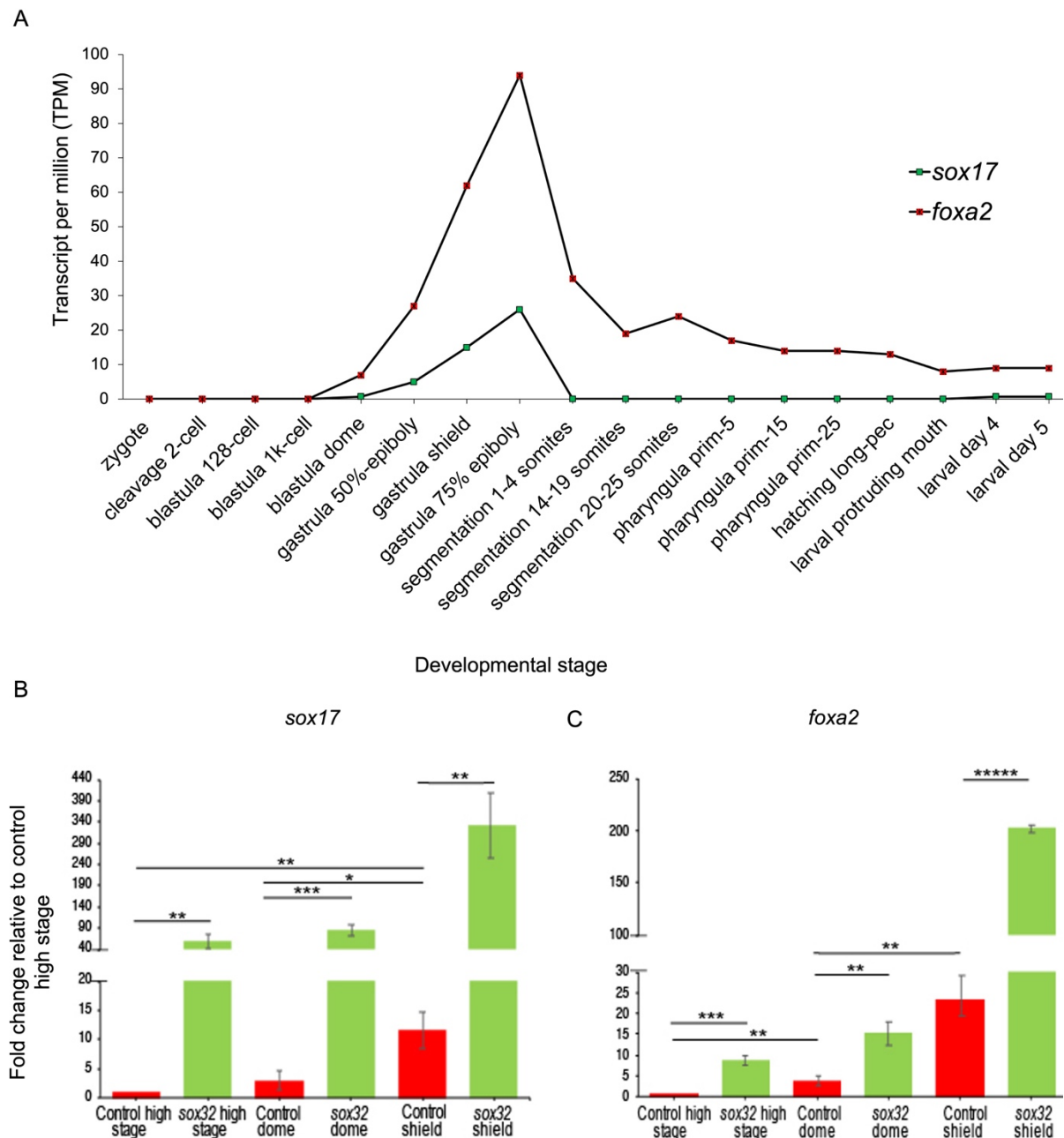


Figure 3-21: Endoderm markers *sox17* and *foxa2* are upregulated in *sox32* OE embryos relative to control. (A) RNA-seq data for *sox17* and *foxa2* during embryogenesis. (B) qPCR analysis of *sox17* and *foxa2* in control and *sox32* OE embryos at high stage (3.3 h.p.f.), dome stage (4.3 h.p.f.) and shield stage (6 h.p.f.). **** $p \leq 5 \times 10^{-7}$; *** $p \leq 5 \times 10^{-4}$; ** $p \leq 5 \times 10^{-3}$; * $p \leq 5 \times 10^{-2}$. Student's t test. Error bars represent standard deviation. Three biological replicates were generated for each sample. Data was normalised to *18S*. RNA-seq data was obtained from White *et al.* (2017).

To further investigate the effects of *sox32* overexpression on DFC fates, qRT-PCR for *vgll4l* was also performed. *Vgll4l* is expressed in the embryo at 1 k-cell stage (~3 h.p.f.) (White *et*

al., 2017) (Figure 3-22A) and is later expressed in the ectoderm and DFCs at shield stage (6 h.p.f.) (Xue *et al.*, 2018). High expression of *vgll4l* was observed at high stage in *sox32* OE compared to control embryos (Figure 3-22B). Moreover, expression of *vgll4l* in control and *sox32* OE embryos across the three different stages mirror the RNA-seq time course shown in Figure 3-22A, wherein the highest expression of the gene is observed at dome stage and its expression gradually decreases until the embryo reaches gastrula stages. The results therefore suggest that *sox32* OE induces DFC fate at an earlier stage relative to control, although more downstream experiments will be required to further prove this.

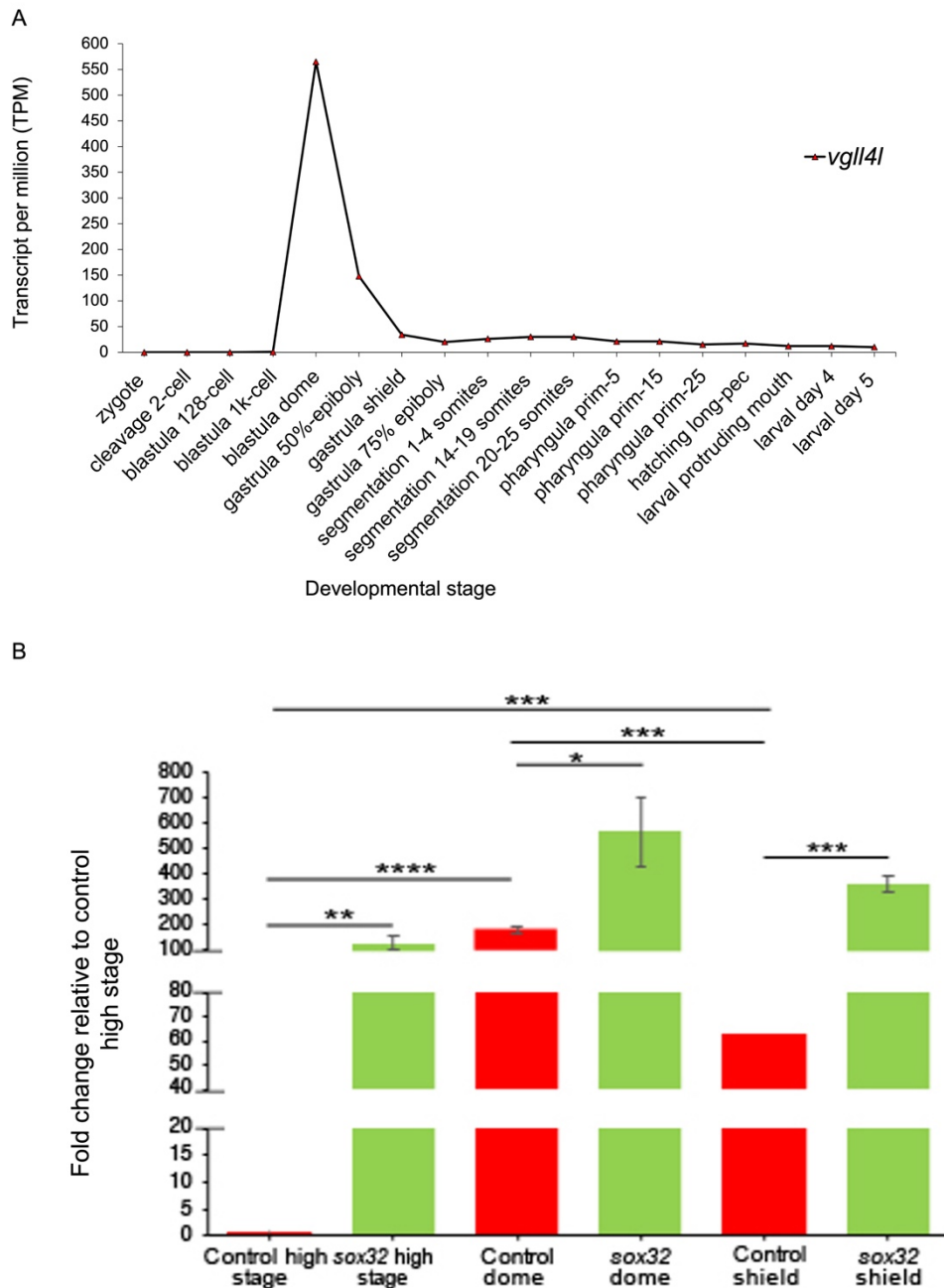


Figure 3-22: DFC marker, *vgll4l* is upregulated in *sox32* OE embryos relative to control. (A) RNA-seq data for *vgll4l* during embryogenesis. (B) qPCR analysis of *vgll4l* in control and *sox32* OE embryos at high stage (3.3 h.p.f.), dome stage (4.3 h.p.f.) and shield stage (6 h.p.f.). **** $p \leq 5 \times 10^{-5}$; *** $p \leq 5 \times 10^{-4}$; ** $p \leq 5 \times 10^{-3}$; * $p \leq 5 \times 10^{-2}$.

Student's t test. Error bars represent standard deviation. Three biological replicates were generated for each sample. Data was normalised to *18S*. RNA-seq data was obtained from White *et al.* (2017).

To test the effects of *sox32* OE on endogenous *sox32*, primers were targeted against the 3'UTR of *sox32*, which is absent in the exogenous *sox32* mRNA used to generate *sox32* OE embryos. Primers were also used to target the *sox32* coding sequence (CDS) which is present in both the injected *sox32* mRNA and endogenous *sox32*. Thus, the difference in expression between *sox32* UTR and CDS will help determine the effects of *sox32* OE on endogenous and exogenous *sox32*. Endogenous *sox32* expression begins at dome stage (White *et al.*, 2017) (Figure 3-23A). As shown below, overexpression of *sox32* results in the upregulation of endogenous and exogenous *sox32* at high stage in *sox32* OE relative to control embryos (Figure 3-23B). The level of expression of *sox32* decreases over time which is likely due to degradation of exogenous *sox32* mRNA. Though *sox32* CDS level decreases (Figure 3-23B), expression level of endogenous *sox32* increased over time, and was significantly upregulated at 6 h.p.f, relative to control embryos (Figure 3-23C). This likely suggests that *sox32* autoregulates its expression in the embryo at early gastrula stages. However, there is a substantial delay between exogenous *sox32* injection and expression of endogenous *sox32*, which suggests that endogenous *sox32* is not an immediate target. Altogether, the results suggest that *sox32* requires other factors to upregulate its own expression in the embryo.

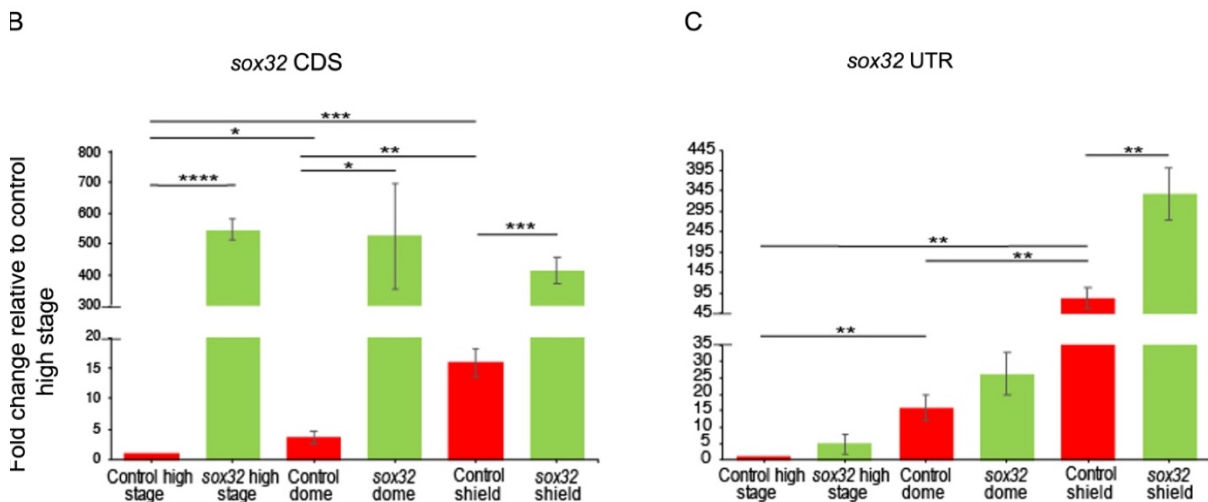
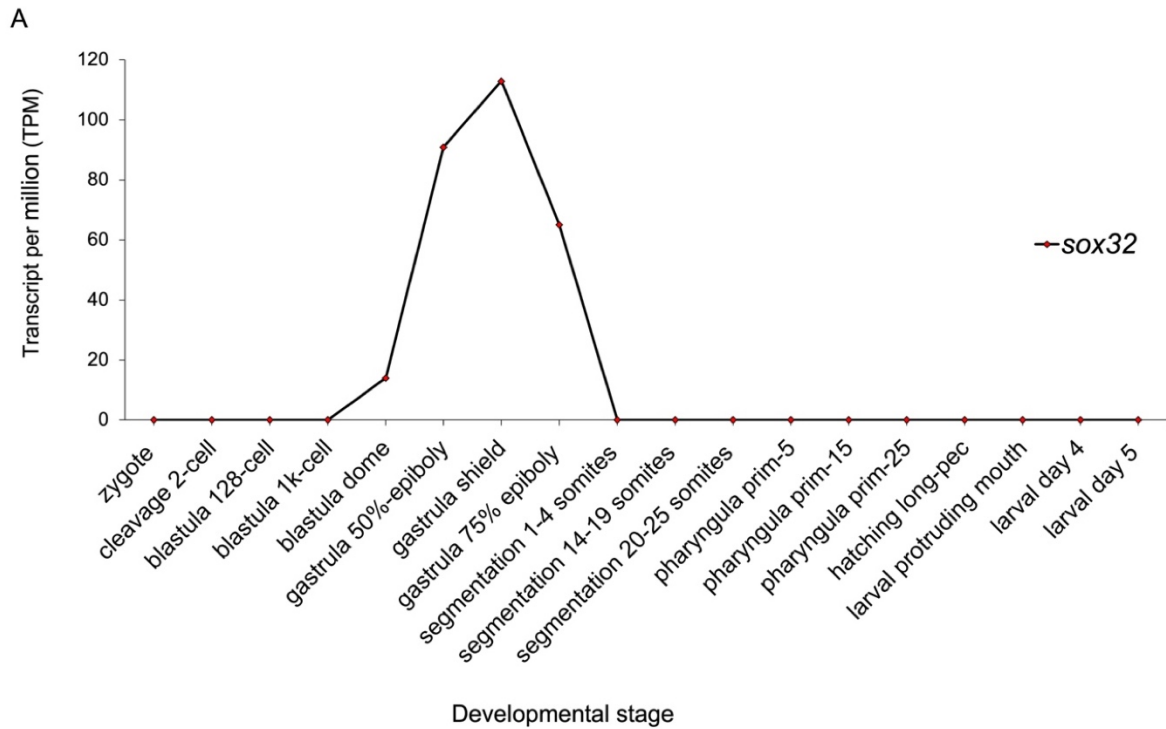


Figure 3-23: Overexpression of *sox32* results in upregulation of endogenous *sox32* expression in the embryo. (A) RNA-seq data for *sox32* during embryogenesis. (B) qPCR analysis for *sox32* coding sequence (CDS) in control and *sox32* OE embryos, which represents both endogenous and exogenous *sox32*. (C) qPCR analysis for *sox32* 3'untranslated region (UTR) in control and *sox32* OE embryos, which represents endogenous *sox32*. qRT-PCR were performed at three different stages of development high stage (3.3 h.p.f.), dome stage (4.3 h.p.f.) and shield stage (6 h.p.f.). **** $p \leq 5 \times 10^{-5}$; *** $p \leq 5 \times 10^{-4}$; ** $p \leq 5 \times 10^{-3}$; * $p \leq 5 \times 10^{-2}$. Student's t test. Error bars represent standard deviation. Note: difference in transcript expression between *sox32* CDS and *sox32* UTR in control embryos at shield stage may be due to differences in the regulation of the *sox32* 3'UTR and CDS during development. Three biological replicates were generated for each sample. Data was normalised to *18S*. RNA-seq data was obtained from White *et al.* (2017).

Though exogenous *sox32* was injected at 1-cell stage, endogenous *sox32* expression was upregulated at shield stage (6 h.p.f.). This delay likely means that *sox32* indirectly regulates its own expression through other factors acting upstream of the gene. Previously, *sox32* OE was demonstrated to induce expression of *ndr1* and *ndr2* (Kikuchi *et al.*, 2001), and for this reason expression of *ndr1* and *ndr2* were also investigated. In relation to this, it was previously stated that *sox32* OE in *MZtdgf1* embryos had no effect on *gata5* expression, which suggests that *sox32* cannot directly activate *gata5*. Thus, to test whether *sox32* OE directly or indirectly induces expression of *gata5*, qRT-PCR was also performed against *gata5* itself (Kikuchi *et al.*, 2001). *Gata5* acts downstream of Nodal signalling but upstream of *sox32* (Bjornson *et al.*, 2005). As shown below, expression of *gata5* commences after 1 k-cell stage in wild-type embryos (~4.3 h.p.f.) (Figure 3-24A). No significant difference was observed in *gata5* expression at high stage and dome stage for both control and *sox32* OE embryos (Figure 3-24B). However, *gata5* expression was significantly upregulated at shield stage (6 h.p.f.) in *sox32* OE relative to control embryos. Therefore, since *gata5* expression was induced later, it is likely that *gata5* is not an immediate target of *sox32* and its expression is induced by other factors.

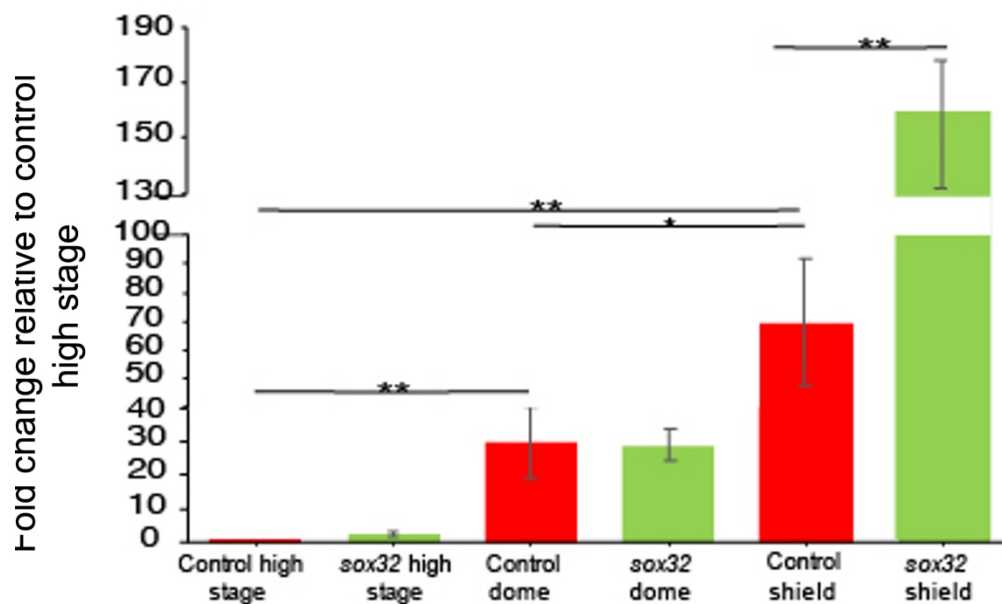
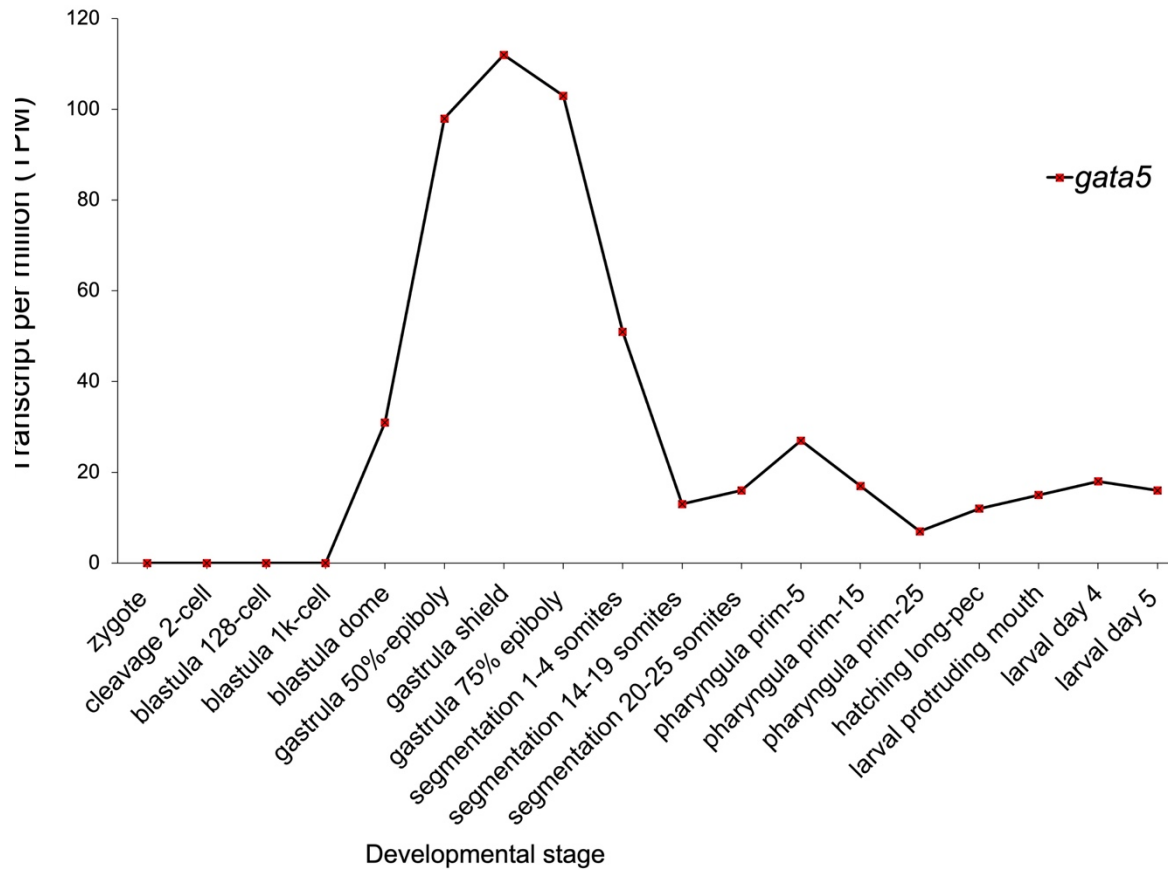


Figure 3-24: Overexpression of *sox32* upregulates expression of *gata5* in *sox32* OE relative to control embryos. (A) RNA-seq data for *gata5* during embryogenesis. (B) qPCR analysis of *gata5* in control and *sox32* OE embryos at high stage (3.3 h.p.f.), dome stage (4.3 h.p.f.) and shield stage (6 h.p.f.). ** $p \leq 5 \times 10^{-3}$; * $p \leq 5 \times 10^{-2}$. Student's t test. Error bars represent standard deviation. Three biological replicates were generated for each sample. Data was normalised to *18S*. RNA-seq was obtained from White *et al.* (2017).

In order to investigate whether *sox32* indirectly regulates expression of *gata5* through Nodal signalling, expression of Nodal ligands *ndr1* and *ndr2* were also investigated. Nodal signalling is required for endoderm and DFC formation since mutations in *ndr1* and *ndr2* inhibit formation of both lineages in the embryo (Feldman *et al.*, 1998; Sampath *et al.*, 1998). Previously, *sox32* OE was shown to induce expression of *ndr1* and *ndr2*, however, this was observed at 5.3 -6 h.p.f. (Kikuchi *et al.*, 2001), thus, we sought to investigate the stage at which Nodal genes are upregulated in *sox32* OE embryos. *Ndr1* is maternally contributed (Gore *et al.*, 2005) while *ndr2* is expressed later at 1 k-cell stage (White *et al.*, 2017) (Figure 3-25A). Moreover, *ndr1* expression at the 4-cell stage is required for dorsal localisation (Gore *et al.*, 2005), and by late blastula stage, *ndr1* accumulates in cells around the dorsal margin (Feldman *et al.*, 1998), the location of mesendodermal progenitors, before it is restricted to the DFCs by early gastrula stages (Rebagliati, Toyama, Fricke, *et al.*, 1998). Though *ndr2* expression is localised to the blastoderm margin at early blastula stage, at early gastrula stages it is restricted to the axial hypoblast of the shield, and by late gastrulation it is expressed in the notochord, prechordal plate and overlying anterior neuroectoderm (Rebagliati, Toyama, Fricke, *et al.*, 1998). Interestingly, *ndr2* expression at the blastula margin is believed to be induced by activation of *ndr1* and maternal factor β -catenin at late blastula stages and high levels of *ndr2* is crucial for formation of mesoderm and endoderm in the embryo (Dougan *et al.*, 2003). Thus, regulation of *ndr2* by *ndr1* is necessary for endoderm specification.

Analysis by qRT-PCR showed higher expression of *ndr1* in *sox32* OE embryos at high stage compared to control (Figure 3-25B). In addition, *ndr1* expression in *sox32* OE mirrors the results seen in control embryos, where the highest expression is observed at dome stage, but its expression decreases beyond this stage. This high expression at dome stage is likely necessary to aid in the specification of the endoderm. However, no difference was observed at shield stage between *sox32* OE and control, contrary to the *in situ* hybridisation data observed in Kikuchi *et al.* (2001).

Notably, no significant difference in expression was observed for *ndr2* between *sox32* OE and control embryos at high stage, dome, and shield stage (Figure 3-25C). Unlike *ndr1*, expression of *ndr2* increased over time in control embryos compared to *sox32* OE embryos. This is also contrary to what was observed in Kikuchi *et al.* (2001), where both *ndr1* and *ndr2* expression were induced by *sox32* OE at shield stage. However, Kikuchi *et al.* (2001) observed upregulation of Nodal genes at ~5.3 h.p.f. but did not investigate their expression prior to this stage. It is noteworthy that this expression pattern was observed by *in situ* hybridisation and not quantitatively by qRT-PCR. Nevertheless, our results demonstrate that *sox32* OE induces earlier expression of Nodal relative to control.

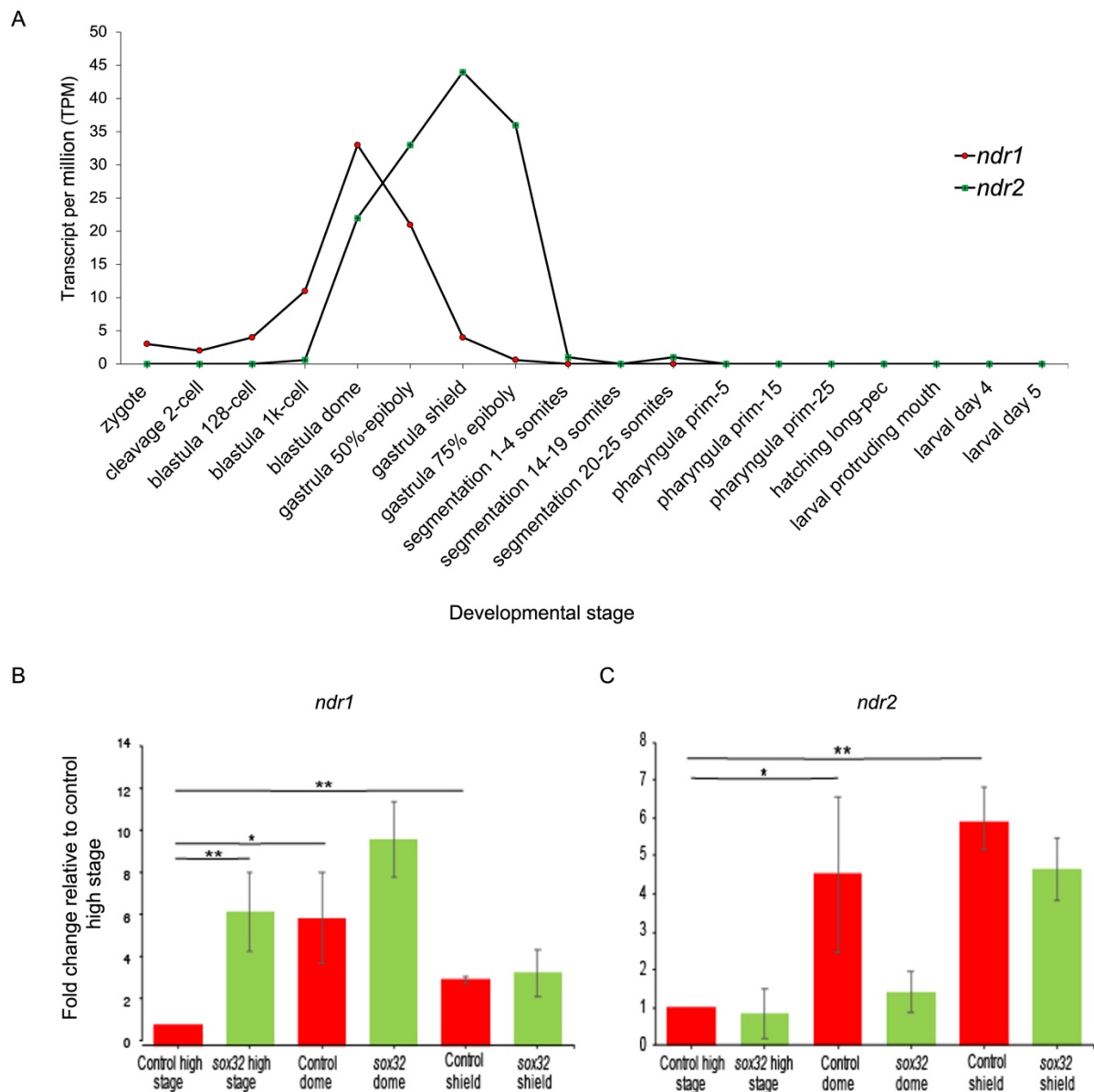


Figure 3-25: Overexpression of *sox32* induces early and upregulated expression of Nodal ligands in *sox32* OE relative to control embryos. (A) RNA-seq data for *ndr1* and *ndr2* during embryogenesis. (B) qPCR analysis of *ndr1* and *ndr2* in control and *sox32* OE embryos at high stage (3.3 h.p.f.), dome stage (4.3 h.p.f.) and shield stage (6 h.p.f.). ** $p \leq 5 \times 10^{-3}$; * $p \leq 5 \times 10^{-2}$. Student's t test. Error bars represent standard deviation. Three biological replicates were generated for each sample. Data was normalised to *18S*. RNA-seq was obtained from White *et al.* (2017).

Overall, our results demonstrates that *sox32* OE induces both ectopic endoderm and DFC formation and that this is likely due to early induction of Nodal signalling. In addition, the results also suggest that *sox32* OE does not induce artificial accessibility in regions that are not accessible in control embryos. Also, the data suggests that *Tbx16* and *Tbx18* likely aid in the maintenance of endoderm and DFCs by binding to *sox32*-regulated regions and regulating expression of genes involved in the formation of both lineages. Finally, we have identified a

number of putative enhancers and sequence motifs that may be implicated in either the formation of endoderm and DFCs or their function during early development.

3.10 Discussion

Although much effort has gone into studying the gene regulatory network in the endoderm during development, the network that acts downstream of *sox32* is still yet to be investigated (Stainier, 2002; Zorn and Wells, 2007, 2009; Tremblay, 2010; Nelson *et al.*, 2014, 2017). However, finding endoderm-specific enhancers is challenging since the endoderm is a minor cell population (Figiel, Elsayed and Nelson, 2021). The function of *sox32* as a master regulator of endoderm formation has already been established (Alexander and Stainier, 1999; Dickmeis *et al.*, 2001; Kikuchi *et al.*, 2001), and overexpression of *sox32* was shown to induce ectopic endoderm formation while loss of *sox32* inhibits endoderm induction (Kikuchi *et al.*, 2001). Therefore, to identify new active CRMs that play a role in endoderm formation, *sox32* was overexpressed in early zebrafish embryos. Our results show that *sox32* OE not only induces ectopic endoderm formation but also ectopic DFCs, a phenomenon not previously discussed in literature. Our results also show that *sox32* OE induces early expression of *ndr1* and also that markers of DFC formation are expressed prior to markers of endoderm formation. In addition to this, GO analysis on genes upregulated in *sox32* OE vs control showed that Kupffer's vesicle formation was the most enriched GO term compared to endoderm formation and that it was associated with genes that play essential roles in left/right asymmetry establishment and KV formation. For instance, *foxj1a* is essential for KV and motile cilia function, *cfr* controls lumen expansion and function of the KV, while depletion of the methyltransferase *dnmt3bb.1* has been shown to affect left/right asymmetry establishment (Yu *et al.*, 2008; Navis, Marjoram and Bagnat, 2013; Fillatre *et al.*, 2019). Thus, the data supports the notion that *sox32* OE induces upregulated expression of essential KV/DFC markers. Previously, *sox32* expression was shown to be either dramatically reduced or lost in both *MZtdgf1* and *Ztdgf1* mutants (Aoki *et al.*, 2002), while DFCs were unaffected in *Ztdgf1* but absent in *MZtdgf1* (Dickmeis *et al.*, 2001; Aoki *et al.*, 2002). Thus, though endoderm and DFCs require Nodal induction, endoderm formation is likely dependent on a greater Nodal signalling response compared to DFCs.

A common problem is that overexpression of genes can result in gene expression artefacts (Prelich, 2012) which means that such experiments can result in the expression of genes that are not a true reflection of normal conditions. To test whether this was the case, read densities obtained from ATAC-seq analysis of control, *sox32* KD and *sox32* OE embryos were mapped to control and *sox32* DARs. As shown in Figure 3-10, *sox32* OE exhibits higher accessibility in regions also accessible in control, suggesting that overexpression of *sox32* does not drive artificial accessibility at loci that are not accessible in control embryos.

Though it is known that *sox32* overexpression results in ectopic endoderm formation, little is known about TFs that bind enhancers downstream of *sox32* to help facilitate this. To investigate the motifs that likely aid in driving ectopic endoderm and DFCs in *sox32* OE embryos, motif analysis was performed on *sox32* OE and control DARs. Motif analysis revealed that *sox32* OE DARs were enriched for Sox, T-box, Forkhead, GATA and Homeobox protein binding motifs, all of which are known to play an important role in endoderm/endoderm organ and DFC/KV formation (Alexander *et al.*, 1999; Reiter, Kikuchi and Stainier, 2001a; Thisse *et al.*, 2001; Clotman *et al.*, 2002; Kim *et al.*, 2005; Amar and Dawid, 2010; Tseng *et al.*, 2011; Dal-Pra, Thisse and Thisse, 2011; DeLaForest *et al.*, 2011; Aksoy *et al.*, 2013; Nelson *et al.*, 2017), (Figure 3-12A). For instance, the most enriched TF binding motifs in *sox32* OE DARs are for mammalian T-box factors Tbx21 and Eomes, orthologues of these TFs have been identified in other species including zebrafish (Takizawa *et al.*, 2007; Mitra *et al.*, 2010). The zebrafish homologue of Eomes, Eomesa is required for normal expression of early endoderm markers which includes *sox32* (Du *et al.*, 2012), while Tbx21 is reported to be expressed in the liver, gut and gills (Mitra *et al.*, 2010). However, since all T-box factors bind the same consensus sequence, these enriched T-box factor motifs are likely to represent other members of the T-box factors that play a role in endoderm formation such as Tbx16 (Nelson *et al.*, 2017). Other enriched motifs present in *sox32* OE DARs include binding motifs for Sox17, which marks both endoderm and DFC lineages and Brachyury, the mammalian homologue of Tbx16, which plays a cooperative and redundant role with Tbx16 in regulating KV formation (Amack, Wang and Yost, 2007; Amar and Dawid, 2010). In addition, binding motifs involved in endoderm specification were also over-represented, and these include Gata6 and Otx2 which regulate expression of Nodal target genes, *gata5*, *sox32* and *sox17* (Tseng *et al.*, 2011) but also Gata4, which is required along with other GATA factors Gata5 and Gata6 to mediate Nodal-dependent endodermal gene expression in *Xenopus* (Molkentin, 2000). GATA TFs bind the same consensus sequences and are evolutionary conserved among animals, plants and fungi (Tremblay, Sanchez-Ferras and Bouchard, 2018). Thus, it is likely that these motifs also represent other GATA members such as Gata5 which is required for formation of the gut tube in zebrafish (Reiter, Kikuchi and Stainier, 2001a). Though T-box and GATA TF motifs were over-represented in *sox32* OE DARs, motifs for Gata5 and Tbx16, which are upstream of *sox32* and have established roles in endoderm formation were not present (Reiter, Kikuchi and Stainier, 2001a; Nelson *et al.*, 2017). However, considering that GATA and T-box factors bind consensus sequences (A/T)GATA(A/G) (Ko and Engel, 1993; Merika and Orkin, 1993) and TCACACCT (Kispert and Herrmann, 1993; Wilson and Conlon, 2002; Garnett *et al.*, 2009) respectively, and also that the position weight matrix is missing for Gata5 and Tbx16 in HOMER, it is likely that these over-represented binding motifs in *sox32* OE also includes both these factors.

Sox32 OE DARs are also enriched for binding motifs of hepatocyte nuclear factors such as HNF6 and HNF4 α (Figure 3-12A). In humans, HNF4 α is essential for specification of hepatic progenitors from pluripotent cell progenitors (DeLaForest *et al.*, 2011). On the other hand, in mice, HNF6 is expressed in hepatoblasts and in the gallbladder primordium and *Hnf6* knockout mice display abnormal pancreas phenotypes (Clotman *et al.*, 2002). Interestingly, the role of *Hnf6* is believed to be evolutionary conserved since knockdown of zebrafish *hnf6*, much like in mice, perturbs biliary development (Matthews *et al.*, 2004). Moreover, mammalian Pou5f1/Oct4-Sox17 binding motifs were also enriched in *sox32* OE DARs (Figure 3-12A). In mammals, Pou5f1 is known to bind to Sox17 to induce endodermal fate (Aksoy *et al.*, 2013). However, while *sox17* is expressed in zebrafish, it is Sox32 that binds to Pou5f3, a paralog of mammalian Pou5f1/Oct4 and as a result, induces endoderm formation (Lunde, Belting and Driever, 2004; Reim *et al.*, 2004; Chan *et al.*, 2009). Thus, mammalian Pou5f1-Sox17 binding motifs likely reflects the binding between Sox32 and Pou5f3. In relation to this, Nanog binding motif was also enriched in *sox32* OE DARs. In mammals, Nanog is known to work alongside other pluripotent factors including Oct4 and Sox2 to regulate expression of genes involved in embryonic stem cell pluripotency (Loh *et al.*, 2006). However, in recent years, Sox32 in zebrafish was also found to physically interact with Nanog *in vivo*, disrupting the interaction between Nanog and Pou5f3 at the dorsal endoderm (Perez-Camps *et al.*, 2016). Thus far, the functional relevance of this interaction is not yet known although it is believed to aid in restricting Pou5f3-Nanog complexes to the ventrolateral mesendoderm. However, the data in Figure 3-13 suggests that Nanog may also regulate important endoderm and DFC-specific genes by binding to Sox32-regulated regions. This contradicts previous data experiments which showed that *MZnanog* results in a substantial loss of endoderm but not DFCs (Veil *et al.*, 2018). Therefore, further experiments will need to be performed to investigate the functional relevance of Nanog-Sox32 interactions on the endoderm and DFCs.

Notably, *sox32* OE DARs also prominently feature Forkhead binding motifs specifically for Foxa1, 2 and 3 where Foxa1 binding sequence appears to be conserved since it matches the consensus sequence defined by experiments on human samples. Forkhead factors are highly conserved in vertebrates and play redundant roles in endoderm formation in zebrafish (Dal-Pra, Thisse and Thisse, 2011). In mice, loss of *Foxa2* perturbs development of the anterior definitive endoderm which results in the loss of all foregut and midgut structures, however, these structures can be compensated by later expression of Foxa1 and Foxa3 (Burtscher and Lickert, 2009), suggesting that Forkhead factors also play redundant roles in other species. In addition to this, Foxa3 is required for glucose homeostasis during a prolonged fast, with evidence that *Foxa3*-null mice have significantly lower blood glucose levels compared to

control groups and exhibit dramatic reduction of the glucose transporter GLUT2 (Shen *et al.*, 2001). On the other hand, in the absence of *Foxa2* and *Foxa3* in zebrafish, the pancreas, liver and intestine fail to differentiate during development (Dal-Pra, Thisse and Thisse, 2011). Thus, Forkhead factors are critical for endoderm organ formation. Interestingly, Forkhead factors are characterised as pioneer factors (Cirillo *et al.*, 1998; Zaret *et al.*, 2008), and thus, one possibility is that Forkhead factors may be the cause of this enhanced accessibility observed in *sox32* OE. However, this is unlikely to be the case since Forkhead factors are not expressed nor required for DFC or KV formation. Also, Forkhead factor *foxa2* acts downstream of *sox32* at the endoderm induction cascade and that expression of *foxa2* is lost in *sox32* mutants (Alexander *et al.*, 1999). In relation to this, *sox32* OE induced early and significant expression of *foxa2* as observed by qRT-PCR (Figure 3-21C), suggesting that *foxa2* is an immediate target of Sox32. Since the density array heatmaps were similar between control and *sox32* KD (Figure 3-10) and given that *foxa2* is an immediate target of Sox32, it is therefore unlikely that Forkhead factors are inducing ectopic endoderm and DFC formation in *sox32* OE.

In contrast to *sox32* OE, control DARs were enriched for binding motifs of OCT4-SOX2-TCF-NANOG TF complex (Figure 3-21B). TFs Oct4, Sox2, TCF and Nanog play an important role in maintaining embryonic cells in a self-renewing, pluripotent state (Cole *et al.*, 2008; Kim *et al.*, 2008). Since motifs of these pluripotent TFs are more over-represented in control embryos, this likely suggests that multiple fates are induced in control compared to *sox32* OE embryos, which in comparison is largely comprised of endoderm and DFC lineages.

Sox members play a widespread role in development but are dependent on the activity of their binding partners for their function (Kamachi and Kondoh, 2013). One binding partner of Sox32 is Pou5f3, which is known to synergistically bind to Sox32 to induce endoderm fate formation (Chan *et al.*, 2009). For this reason, one hypothesis is that ectopic endoderm formation is caused by upregulation of Pou5f3 in *sox32* OE embryos. However, TF binding motif for mammalian Pou5f1, a paralogue of Pou5f3, was not more enriched in *sox32* OE DARs relative to control (Figure 3-12A). In relation to this, there was no difference in the number of regions that overlap between *sox32* OE or control DARs with Pou5f3 ChIP-seq peaks (Figure 3-15). The reason behind this however is likely because Pou5f3 ChIP-seq was performed at 5 h.p.f. (Leichsenring *et al.*, 2013) which is prior to endoderm specification (Hagos and Dougan, 2007). There are therefore two problems with this, firstly, since endoderm constitutes a minor cell population, the signal for Pou5f3 in endoderm cells are likely lost in the context of the whole embryo. In addition, Pou5f3 ChIP-seq was performed at 5 h.p.f. which is prior to endoderm specification and thus, endoderm is unlikely to be observed at this stage. Therefore, one cannot conclude that Sox32 requires Pou5f3 to induce ectopic endoderm and DFC

formation in *sox32* OE embryos without performing ChIP-seq for Pou5f3 beyond 6 h.p.f., i.e.: once endoderm is specified in the embryo.

It is already established that *sox32* overexpression induces ectopic endoderm formation (Kikuchi *et al.*, 2001), however the effects of *sox32* overexpression on mesoderm, ectoderm and DFCs has not been previously investigated. In the absence of Sox32, endoderm progenitors differentiate into mesoderm structures (Dickmeis *et al.*, 2001) but it is not known whether mesoderm is entirely lost in *sox32* OE embryos. In addition, Sox32 is required for DFC formation (Cooper and D'Amico, 1996; Essner *et al.*, 2005), but the effects of *sox32* overexpression on DFC formation was never previously discussed. Comparison of upregulated transcripts from control and *sox32* OE embryos to genes enriched in endoderm, mesoderm, ectoderm from various stages of development, as well as DFCs, showed that *sox32* OE transcripts are more significantly concordant with endoderm and DFC-enriched genes (Figure 3-17A). In relation to this, *sox32* OE genes were not only significantly correlated with the endoderm at 6 h.p.f but also at later stages of development, between 8 and 14 h.p.f. This is likely because *sox32* mRNA was injected into one-cell staged embryos and were therefore exposed to the effects of *sox32* at an earlier stage compared to control. Though there was a significant overlap between mesoderm and ectoderm-enriched genes with *sox32* OE upregulated genes, this is likely because most of the genes at 6 h.p.f. are not restricted to the endoderm but also other germ layers (Figure 3-18B-C). The results are therefore in line with *sox32* OE being largely composed of endoderm and DFC lineages and that this is likely a causative effect of a change in the chromatin landscape in the embryo (Figure 3-19). However, we cannot completely exclude the possibility that some mesoderm and ectoderm lineages are present in *sox32* OE without further downstream analyses, for instance, FACS sorting of GFP+ cells from *sox32* OE *Tg(sox17:EGFP)* embryos at 6 h.p.f. will help determine whether all GFP+ cells are endoderm and/DFCs, or whether this includes other lineages.

Our results demonstrate that *sox32* OE drives enhanced accessibility in regions that are also accessible in control and *sox32* KD and also that the density array heatmaps look near identical between control and *sox32* KD (Figure 3-10). This is likely because *sox32* KD and control have very few *sox32*+ cells and thus, it is likely that the signal is lost in the context of the whole embryo. Despite this, since the signal is not lost in *sox32* KD vs control, the results suggest that *sox32* is sufficient but not necessary for this enhanced accessibility. However, it is also noteworthy that one cannot fully rule out that *sox32* is not responsible for this enhanced accessibility since it is expressed in a minor cell population and its signal is likely negligible in the context of the whole embryo.

It was previously demonstrated that *sox32* OE upregulates expression of Nodal ligands *ndr1* and *ndr2* as observed by *in situ* hybridisation at 5.3 – 6 h.p.f. (Kikuchi *et al.*, 2001), which are necessary for endoderm induction. However, Nodal signalling typically begins after mid-blastula stage (~3 h.p.f.) (Hagos and Dougan, 2007) and therefore, if *sox32* OE induces early expression of Nodal, it is therefore likely that endoderm and DFCs are induced earlier in the embryo and that Nodal signalling may be required to induce ectopic formation of both fates. For this reason, qRT-PCR was performed on markers upstream of *sox32*, but the experiment was first validated by looking at markers that act downstream of the gene. Expression of endoderm markers *foxa2* and *sox17*, which are typically expressed after dome stage (~4.3 h.p.f.) in wild-type embryos, were upregulated at an earlier stage in *sox32* OE embryos relative to control (Figure 3-21). In relation to this, expression of DFC maker, *vgl14l* which is normally expressed in the embryo at 1 k-cell stage (White *et al.*, 2017), was significantly upregulated at high stage (~3.3 h.p.f.) in *sox32* OE compared to control embryos (Figure 3-22). These results suggest that endoderm and DFCs are induced at an earlier stage in *sox32* OE compared to control.

In addition to the above, other markers were also investigated by qRT-PCR, including endogenous *sox32*. The results suggest that *sox32* regulates its own expression in the embryo since overexpression of exogenous *sox32* induced expression of endogenous *sox32*, and that endogenous *sox32* was significantly upregulated by the time *sox32* OE embryos reached shield stage (6 h.p.f) relative to control (Figure 3-23B-C). Though *sox32* expression was upregulated, there was a noticeable delay between exogenous and endogenous *sox32* expression in the embryo, which suggests that other factors are required to induce expression of endogenous *sox32*. Like *sox32*, there was also a noticeable delay between *sox32* OE and upregulation of *gata5* expression in *sox32* OE embryos (Figure 3-24B) which further suggests the need for other factors to induce their expression. This is agreement with Kikuchi *et al.* (2001) which claimed that *sox32* OE cannot induce expression of *gata5* in *MZtdgf1* mutants. Previously *sox32* OE was shown to upregulate expression of *ndr1* and *ndr2* (Kikuchi *et al.*, 2001), and thus, markers against both Nodal genes were also investigated. The qRT-PCR data demonstrates that *ndr1* expression is upregulated in *sox32* OE relative to control at high stage (~3.3 h.p.f.), but no difference was observed at shield stage (6 h.p.f.), contrary to the *in situ* hybridisation result in Kikuchi *et al.* (2001), (Figure 3-25). In addition, no difference was observed in *ndr2* expression between *sox32* OE and control embryos at high stage, dome and shield stage (Figure 3-25C), which is also contrary to Kikuchi *et al.* (2001). This could be due to that fact that *in situ* hybridisation identifies the location of transcripts but does not quantify the expression levels of genes per cell across the embryo. Thus, we cannot conclude that the data is different until further experiments have been performed. Nevertheless, the data

suggests that *sox32* OE significantly upregulates expression of *ndr1* at high stage, which in turn increases expression of *gata5* and *sox32* at later stages of development. Additional experiments can also be performed to further investigate the role of Nodal in endoderm and DFC induction in *sox32* OE embryos. For instance, density array analysis surrounding DARs from *sox32* OE in *MZtdgf1* mutants vs control embryos will help determine whether Nodal signalling is responsible for the enhanced accessibility observed in *sox32* OE embryos.

All in all, the results suggest that overexpression of *sox32* induces ectopic formation of both DFCs and endoderm. However, given early expression of DFC marker, *vgll4l*, but late expression of endoderm regulators *sox32* (endogenous) and *gata5*, our data likely suggests that DFCs are induced earlier than endoderm in *sox32* OE embryos. However, additional experiments would need to be performed to confirm this. For instance, single-cell RNA-seq and single-cell ATAC-seq can be performed at different stages of development to observe the time-point at which both endoderm and DFCs are induced in *sox32* OE. Other additional experiments can also be performed to test the presence of DFCs in *sox32* OE. For instance, transplantation of endocytic dorsal cells from *sox32* OE into host WT embryos will show whether these cells become KV. Our data also suggests the presence of important transcription factor binding motifs that might be important in driving endoderm and DFC formation and one way their role could be investigated is through the use of transient reporter assays. Finally, ATAC-seq experiments on *sox32* OE in *MZtdgf1* mutants vs control embryos will help determine whether this enhanced accessibility is the result of upregulation of early Nodal signalling.

These additional experiments will help further our understanding on the importance of Nodal signalling on the induction of endoderm and DFC formation but also on the transcription factor binding motifs that are likely important for their formation during development.

Chapter 4 Investigating the chromatin landscape at early organogenesis (28 h.p.f.)

4.1 From gastrulation to organogenesis

Organogenesis begins after gastrulation and involves dynamic regulation of gene expression changes to aid in the formation of the body's tissues and organs from rudimentary progenitor cells. The endoderm makes contributions to the respiratory and gastrointestinal tracts and all associated organs such as thyroid, lungs, liver, gall bladder and pancreas (Wlizla and Zorn, 2014) as well as the pituitary gland (Fabian *et al.*, 2020). During gastrulation, the endoderm progenitor cells involute first, migrate anteriorly under the epiblast and eventually flatten and form a sparse but regular monolayer against the yolk (Warga and Nüsslein-Volhard, 1999; Aronson, Stapleton and Krasinski, 2014). By early somite stages, these flattened cells converge on the dorsal midline to form a rod of cells from which organs eventually emerge and which later cavitates to form the gut tube (Aronson, Stapleton and Krasinski, 2014). During this period, the gut tube becomes regionalised along the dorsal-ventral and anterior-posterior axes into foregut, midgut and hindgut domains which can be identified by Hhex, Pdx1 and Cdx factors, respectively. The foregut gives rise to the oesophagus, trachea, stomach, lungs, thyroid, liver, biliary system and pancreas; while the midgut forms the small intestine and hindgut forms the large intestine (Zorn and Wells, 2009). These processes are believed to be controlled by many growth factor pathways including FGF, BMP, Wnt and retinoic acid (RA), which play essential stage-specific roles during endoderm organogenesis (Zorn and Wells, 2009).

Perturbation of these organs are associated with known defects, such as cystic fibrosis, chronic hepatitis and diabetes, which affect millions of people each year (Spence and Wells, 2007). Interestingly, though 25% of developmental disorders can be explained by genetic variation within protein-coding regions (McRae *et al.*, 2017), the majority of cases are caused by mutations in non-coding regions of the genome (Short *et al.*, 2017). These genetic variations are believed to reside in *cis*-regulatory modules (CRMs) that regulate developmental gene expression and therefore may play a critical role in organ formation at the onset of organogenesis. However, aside from rare examples such as pancreatic agenesis (Weedon *et al.*, 2014), there is still a lack of human data that provides a link between defective enhancers and developmental diseases. Moreover, although many animal models have aided in the identification of ultra-conserved developmental regulatory regions (Woolfe *et al.*, 2005;

Dickel *et al.*, 2018), a major drawback to these studies is that often sequence conservation provides no information regarding the developmental stage and tissue location of active putative enhancers. While only a small number of studies have identified regulatory elements that are crucial for pancreatic formation (Weedon *et al.*, 2014; Cebola *et al.*, 2015), this still remains unresolved for other endodermal organs.

4.2 Sox17 marks the endoderm and DFCs

Before the onset of gastrulation, *sox17*, which acts downstream of endodermal inducer *sox32*, can be detected in a dorsally located group of marginal cells. These cells correspond to a group of non-involuting endocytic cells known as Wilson cells which later form forerunner cells, also called dorsal forerunner cells (DFCs). The expression of *sox17* in the DFCs is maintained during gastrulation but in contrast to other dorsal cells, DFCs do not involute but remain at the leading edge of the blastoderm margin (Melby, Warga and Kimmel, 1996). During early somitogenesis, DFCs migrate deep into the embryo and form the lining of Kupffer's vesicle (KV), highly spherical, fluid-filled ciliated organ that is unique to the developing teleost tail bud (Essner *et al.*, 2005), (Figure 4-1) and soon hereafter, the KV ceases to express *sox17*.

On the other hand, endodermal expression begins soon after gastrulation. As cells of the marginal zone involute, a subpopulation were found to express *sox17* and were located in close proximity to the yolk: these cells were later identified to be endodermal progenitor cells (Warga and Nüsslein-Volhard, 1999). Expression of *sox17* continues to be expressed in the endoderm during gastrulation but disappears early in somitogenesis and is not detected again until 48 h.p.f. in a group of cells located in the left upper trunk (Alexander and Stainier, 1999).

Notably there are many similarities between the endoderm and DFCs. For example, they are induced by high-level Nodal signalling (Dickmeis *et al.*, 2001; Essner *et al.*, 2005) and when expression of Nodal ligands are absent, DFC and endoderm formation are impaired (Dickmeis *et al.*, 2001). Moreover, DFCs are derived from a group of cells known as Wilson cells, a group of marginal enveloping layer cells that are cytoplasmically connected to the yolk (Warga and Kane, 2018). At the onset of gastrulation, these Wilson cells ingress beneath the enveloping layer and become DFCs. However, prior to ingressing the Wilson cells were found to exhibit endodermal characteristics since they also require activation of endodermal-specific genes, as well as high-levels of Nodal signalling (Warga and Kane, 2018). Thus, it is likely that DFCs and endoderm share the same or similar transcriptional programmes during development.

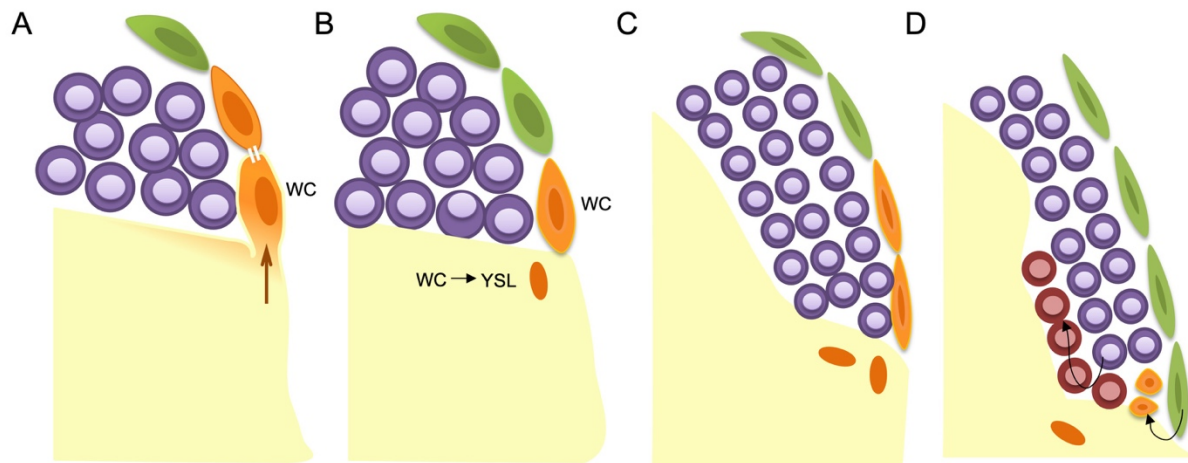


Figure 4-1: Model for DFC induction. A) DFC induction require a signal (shown in orange) from the yolk cell that is passed on to a Wilson cell that is cytoplasmically continuous with the yolk and to its daughter cell (both shown in orange) who share a midbody (white bars). Because of the coupling, one cell always becomes a Wilson cell while the other cell (the daughter cell in this case) becomes uncoupled and a part of the blastoderm, hereafter referred to as Wilson cell descendent. Exterior enveloping layer are shown in green, deep domain cells are represented in purple and the yolk cell is in yellow. B) By 1K-cell stage (3.3 h.p.f.), Wilson cells collapse into the yolk, forming the yolk syncytial layer (YSL; orange nucleus), leaving their siblings, the Wilson cell descendants, as part of the blastoderm. C) By dome stage (4.3.h.p.f), the Wilson cell descendants, the cells that form part of the blastoderm, divide one or two times in the enveloping layer. D) When deep cells (purple) ingress to form mesendoderm (red), dorsal enveloping layer Wilson cell descendants ingress to become DFCs.

4.2.1 KV, organ of laterality

In zebrafish, DFCs give rise to the KV, which is located ventrally to the forming notochord in the tailbud and is adjacent to the yolk cell (Essner *et al.*, 2005; Kramer-Zucker *et al.*, 2005). The KV lumen first form at around 2-somite stage (~10.33 h.p.f.) and the lumen is fully expanded by 8-somite stage (11.66 h.p.f.) (Amack, Wang and Yost, 2007; Wang, Manning and Amack, 2012; Gokey *et al.*, 2016). In the KV, ciliated cells generate a counter-clockwise fluid flow within the lumen of the KV, and this flow is critical for normal left/right patterning and asymmetric gene expression (Essner *et al.*, 2005; Kramer-Zucker *et al.*, 2005). Cilia-driven leftward flow aiding establishment of left/right asymmetry has also been described in other species, including mice, rabbit and *Xenopus* (Nonaka *et al.*, 1998; Okada *et al.*, 2005; Schweickert *et al.*, 2007) which suggests that ciliated cells play a conserved role in development in all vertebrates. The vertebrate left/right organiser is also believed to be lined with immotile cilia which are thought to act as sensors of leftward nodal flow: Mice with paralyzed motile cilia but intact sensory cilia retain the ability to sense nodal flow (McGrath *et al.*, 2003). Interestingly, cell labelling and transplantation experiments in mouse showed that cells of the left/right organiser called the node, gives rise to the axial mesoderm which includes the trunk notochord (Sulik *et al.*, 1994; Kinder *et al.*, 2001). Similar to the mouse, the left/right

organiser in chick known as Henson's node contributes to the head and notochord and later makes contributions to the posterior notochord (Selleck and Stern, 1991). This is also consistent in zebrafish because, following its role in left/right patterning, KV cells disperse and contribute to the notochord or muscle fates in the tail (Melby, Warga and Kimmel, 1996). This conserved role in establishing left/right asymmetry led to the categorization of the KV as an embryonic "organ of asymmetry" in zebrafish (Essner *et al.*, 2005).

In zebrafish, *sox17* is expressed in DFCs and KV, and is also expressed in the area of the node in mouse and chick (Chapman *et al.*, 2007; Hassoun, Püschel and Viebahn, 2010). *Sox17* expression is critical for DFC and KV formation with evidence that knockdown of *sox17* results in defective KV and abnormal left/right patterning (Aamar and Dawid, 2010), as highlighted below (Figure 4-2). This is equivalent to *Sox17* function in the mouse node, where loss of *Sox17* results in either no expression or significantly reduced expression of left/right asymmetric genes including *Nodal* (Viotti, Foley and Hadjantonakis, 2014). However, loss of *sox17* has little-to-no effect on endoderm formation in zebrafish unlike in other species (Aamar and Dawid, 2010). Thus, *sox17* is critical for KV function and left/right patterning in zebrafish.

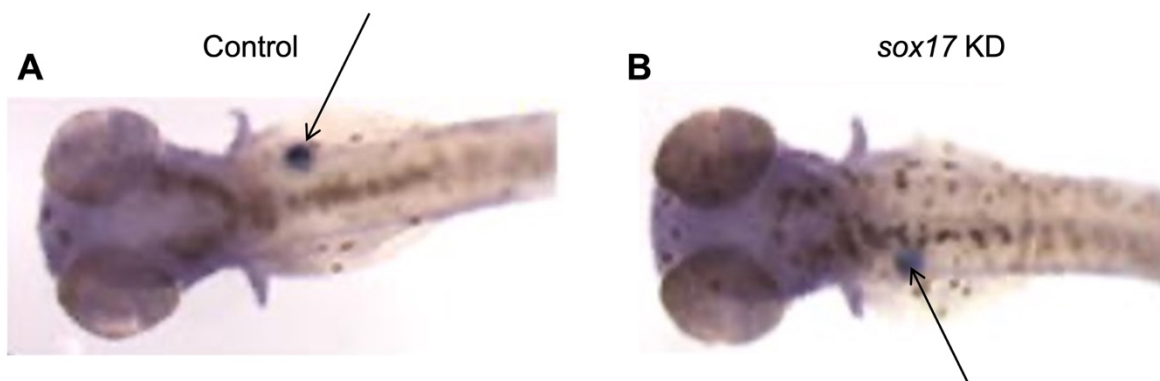


Figure 4-2: KV function in zebrafish. Disruption of the KV results in randomised expression of *insulin (ins)* which is exclusively expressed on the right as shown in panel A. (B) Anti-sense knockdown of *sox17* not only results in fragmented pancreas but also altered localisation of the gene with some embryos expressing the gene on the left side of the embryo. KD= knockdown. Figure adapted Aamar and Dawid, (2010).

4.2.2 Study of endoderm and DFCs by using the pan-endodermal line

To enable the study of the endoderm and DFCs, the transgenic zebrafish reporter line *Tg(sox17:EGFP)* is often used: this endoderm-specific line carries an *EGFP* reporter driven by *sox17* promoter (Mizoguchi *et al.*, 2008). Although the expression of *sox17* begins just after gastrulation, the earliest detectable *EGFP* fluorescence in the endoderm is around 60% epiboly, which is approximately 7 h.p.f. and this continues until 2 days post-fertilisation (d.p.f.), and thus, the reporter expression persists long after endogenous *sox17* is silenced in the

endoderm (Mizoguchi *et al.*, 2008). This also means that fluorescence-activated cell sorting (FACS) can be used to study the endoderm within the first 48 h.p.f. by sorting for GFP+ cells. For instance, FACS sorted GFP+ cells from *Tg(sox17:EGFP)* were previously used to quantitatively assess the levels of prostaglandin receptors *ep2a* and *ep4a*, at six-to eight-somite (~12 h.p.f.) stages, which are important regulators of endoderm specification (Nissim *et al.*, 2014). However, since *sox17* also marks the DFCs and KV, *Tg(sox17:EGFP)* has also been used to locate DFCs and KV during left/right asymmetry development. For instance, FACS was previously used to sort for DFCs at bud stage which express high strong levels of *GFP* at this stage compared to endoderm (Tavares *et al.*, 2017).

4.3 Sox17 is also a marker of haemopoietic lineages

However, emerging evidence has demonstrated that *sox17* is also expressed in haemopoietic cell lineages. Studies on mice revealed that *Sox17* marks foetal and neonatal hematopoietic stem cells and germline deletion of *Sox17* results in severe foetal hematopoietic cell defects including loss of definitive hematopoietic stem cells (Kim, Saunders and Morrison, 2007). Since then, expression of *sox17* was also observed in hemopoietic lineages in zebrafish (Chung *et al.*, 2011), suggesting that genetic programmes governing development of hemopoietic lineages is conserved among vertebrates.

In zebrafish, *sox17* is very strongly expressed during early stages of development, but barely detectable beyond gastrulation (White *et al.*, 2017), (Figure 4-3).

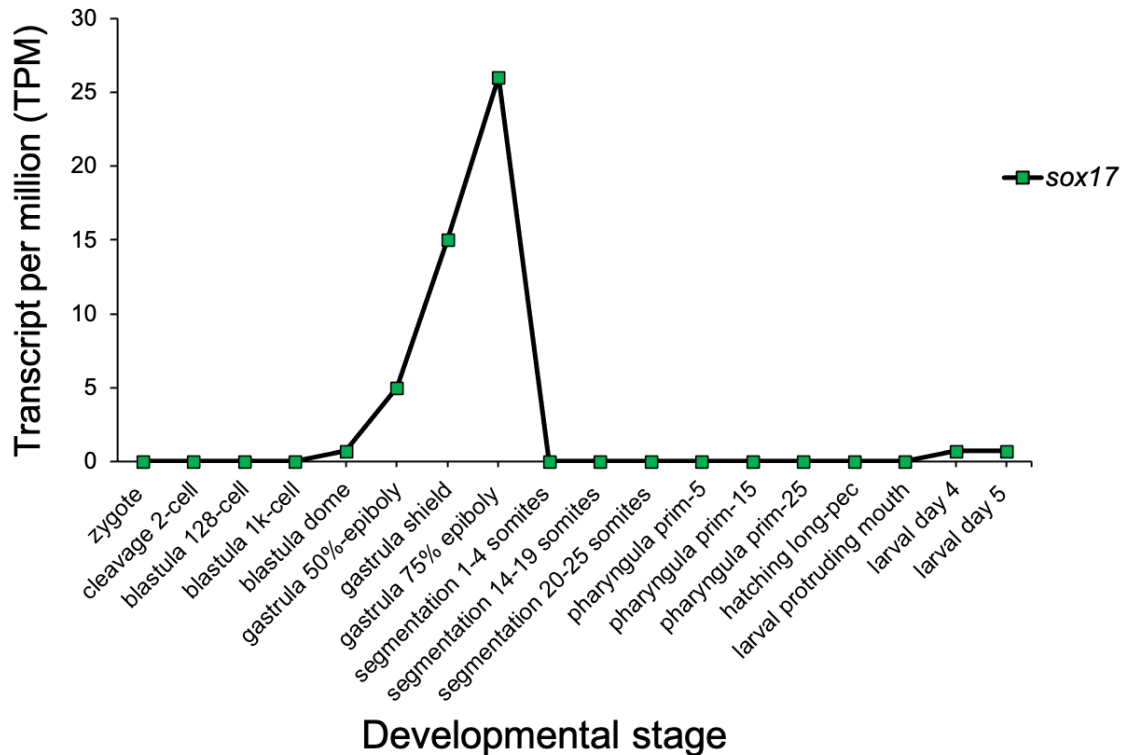


Figure 4-3: Time course of *sox17* expression during the first 5 days of zebrafish development. RNA-seq data for zebrafish *sox17* plotted over the course of development (1-cell to 5 d.p.f.), at its highest during gastrulation and then almost completely undetectable. Data extracted from White *et al.* (2017).

While *sox17* marks the endoderm during gastrulation (5.25-10.33 h.p.f), Chung *et al.* (2011) showed that post-gastrulation, *sox17* is no longer expressed in the endoderm but in several other lineages; at 12 h.p.f. *sox17* is expressed in the left and right lateral plate mesoderm (LPM), which give rise to endothelial cells and KV, and by 18 h.p.f. *sox17* marks the intermediate cell mass of mesoderm (ICM), which give rise to blood cells, while at, 26 h.p.f. *sox17* is expressed in ICM, neural tube and the duct of Curvier (Chung *et al.*, 2011) (Figure 4-4). Consistent with this, morpholino knockdown of *sox17* results in a significant reduction in haemopoietic cell markers such as *GATA binding protein 1a* (*gata1a*) in the ICM and reduced erythroid cell numbers at 18 h.p.f. (Chung *et al.*, 2011). Thus, endoderm marker, *sox17* is crucial for hemopoiesis and KV function in zebrafish.

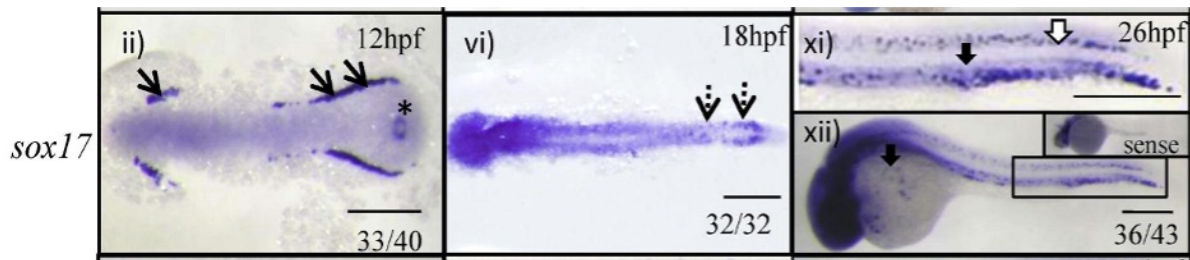


Figure 4-4: Whole-mount *in situ* hybridisation showing the expression pattern of *sox17* in 12, 18 and 26 h.p.f. zebrafish embryos. At 12 h.p.f., *sox17* was expressed in the anterior and posterior LPM, as indicated by arrows and KV, as indicated by asterisks (left panel). At 18 h.p.f., expression of *sox17* was present in the ICM as marked by broken arrows (middle panel). At 26 h.p.f., *sox17* was observed in hematopoietic lineages, as indicated by closed vertical arrows, as well as neural tube, as indicated by open vertical arrows. Figure from Chung *et al.* (2011).

4.3.1 Transgenic lines used to study endothelial and blood lineages in zebrafish

To facilitate the study of endothelial and haemopoietic lineages at around 24 h.p.f., a number of transgenic lines are commonly used. To enable the study of haemopoietic lineages in zebrafish, the *Tg(gata1a:dsRed)* line is often used, which expresses the reporter gene *dsRed* in blood cells under the control of the *gata1a* promoter (Fish *et al.*, 2008). Moreover, the study of endothelial cells is enabled by the *Tg(kdrl:mCherry)* transgenic line, which uses the reporter gene *mCherry* to mark blood vessels under the control of *kdrl* promoter, where *kinase insert domain receptor like (kdrl)* encodes a highly vascular endothelial- specific transcription factor (Jung *et al.*, 2017).

4.4 Objectives of this study

Though extensive knowledge is known about CRMs that regulate gene expression prior to endoderm specification, little information is known about potentially important enhancers that are crucial during endoderm organogenesis in zebrafish. To enable the study of the endoderm during development, *Tg(sox17:EGFP)* is often used, which marks all endoderm progenitors within the embryo. However, during endoderm organogenesis, *sox17* is no longer a marker of endoderm progenitors but of haemopoietic lineages. This means that the study of the endoderm is enabled by GFP expression but this also means that GFP marks both endoderm and mesoderm lineages at later stages of development.

The main objectives of this study:

- 1) To investigate whether it is possible to isolate *sox17+* endoderm from contaminating *sox17+* mesoderm based on GFP fluorescence from *Tg(sox17:EGFP)*. This will be investigated by first using FACS to sort cells based on GFP fluorescence in *Tg(sox17:EGFP)* embryos. In addition to this, qPCR against endodermal markers will be used to investigate the validity of this approach.
- 2) To identify potentially important CRMs that are likely required to drive endoderm organogenesis during zebrafish development. This will be investigated by performing ATAC-seq and RNA-seq on endoderm cells at the onset of organogenesis.

To enable the study of endoderm-specific CRMs at early organogenesis, ATAC-seq and RNA-seq were performed on endodermal cells isolated from embryos at the onset of organogenesis. Studying the endoderm at this stage will aid our understanding of potentially important endoderm-specific CRMs that may be required during endoderm organogenesis.

4.5 Characterising *Tg(sox17:EGFP)* expression pattern during early organogenesis

The pan-endodermal line *Tg(sox17:EGFP)*, which expresses EGFP under the control of a 5.0 kb region of the *sox17* promoter (Mizoguchi *et al.*, 2008), marks all endoderm progenitor cells during development. Beyond gastrulation, *sox17* is no longer a marker of the endoderm and therefore the study of the endoderm is enabled by the *EGFP* reporter gene which persists after endogenous *sox17* is silenced in zebrafish. However *sox17* post-gastrulation marks mesoderm progenitors, particularly ones that contribute to endothelial and hematopoietic lineages. This means that the transgene marks *sox17*-positive (*sox17*⁺) endoderm and *sox17*⁺ mesoderm at later stages of development. In order to characterise the expression pattern of the transgene during early endoderm organogenesis, we generated *sox32*-deficient *Tg(sox17:EGFP)* embryos by using a morpholino which targets *sox32*, the master regulator of endoderm formation, and imaged embryos at primordium (Prim)-4 (24 h.p.f.) and Long-pec stage (48 h.p.f.).

Consistent with *sox32* mutants (Dickmeis *et al.*, 2001), *sox32* morphants (*sox32*-KD) display a complete lack of endoderm (Figure 4-5C&F). For example, GFP-positive (GFP⁺) liver and pancreas, which are clearly defined in *Tg(sox17:EGFP)* embryos are absent in morphants (Figure 4-5B-F). Though *sox32*-KD lack endoderm, they display GFP⁺ ICM at 24 h.p.f., and pericardial oedema (Figure 4-5C&F), in accordance with published data (Alexander *et al.*, 1999). Moreover, GFP⁺ cells in the posterior notochord in *Tg(sox17:EGFP)* (as shown in Figure 4-5B&E) appear to be missing in *sox32*-KD (Figure 4-5C&F). Given that fate mapping experiments demonstrated that the KV contributes to the tail notochord (Melby, Warga and Kimmel, 1996), it is therefore likely that these GFP⁺ cells are derivatives of KV cells.

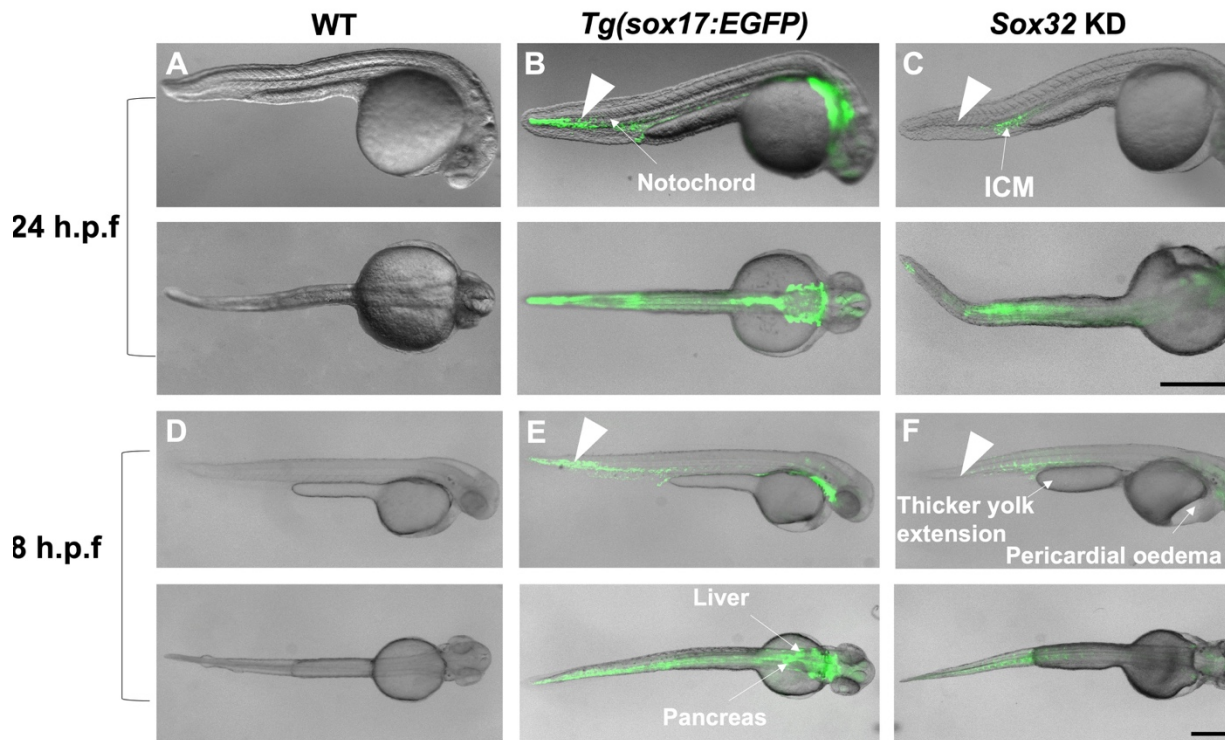


Figure 4-5: Brightfield and fluorescence merged microscopy images demonstrating that *sox32*-KD lack endoderm but display GFP+ mesoderm. Embryos were imaged at 24 h.p.f. (A-C) and 48 h.p.f. (D-F). *Sox32* KD (C&F) lack *sox17*+ endoderm, including liver and pancreas (clearly labelled in panel E) but display GFP fluorescence in *sox17*+ mesoderm in comparison to *Tg(sox17:EGFP)* embryos which display GFP fluorescence in both *sox17*+ lineages (B&D). *Sox32* KD also display pericardial oedema and thicker yolk extension and lack GFP+ cells in the posterior notochord, as evident by the white arrowhead. Embryos were imaged laterally (top panel) and dorsally (bottom panel). Notochord, ICM, liver and pancreas are clearly labelled. Scale bar, 0.5 mm.

4.6 GFP+ cells in posterior notochord are derivatives of the KV

The images in Figure 4-5 show that GFP+ cells in the posterior notochord are absent in *sox32* KD, and based on fate mapping experiments described in Melby, Warga and Kimmel, (1996), these cells are likely to be derivatives of the KV. Thus, we aimed to investigate the developmental fates of DFCs and KV. However, prior to this, it was vital that we first ensure that the KV is clearly visible under the microscope. As previously highlighted in Kimmel *et al.* (1995), the KV appears in the embryo at 6-somite stage (12 h.p.f.) (Figure 4-6A), and this was also evident by live confocal imaging of *Tg(sox17:EGFP)* embryos as shown in Figure 4-6B-D.

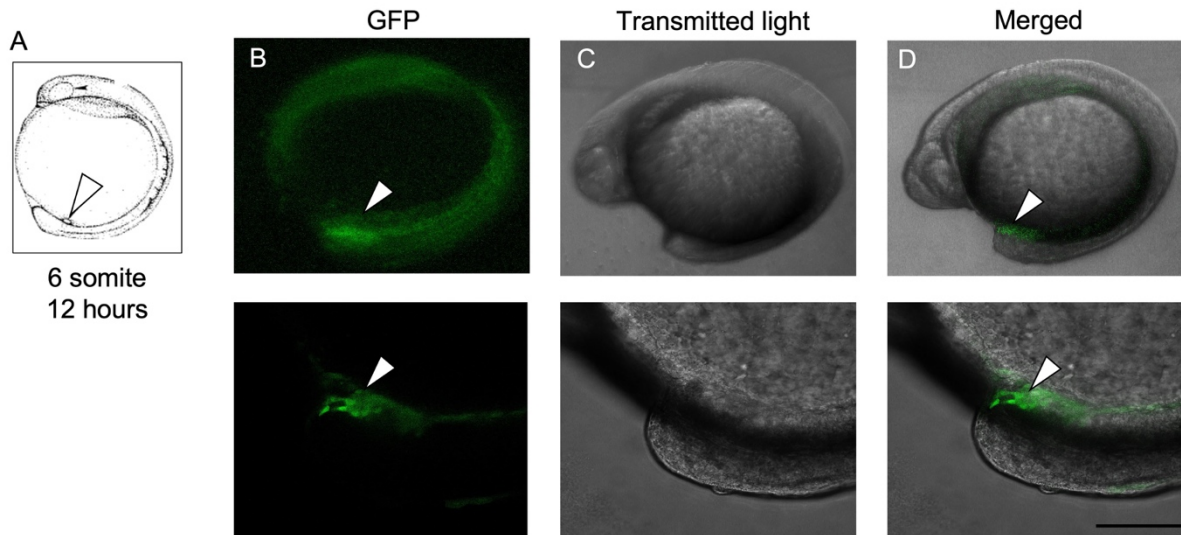


Figure 4-6: Imaging the KV. A) Schematic image of an embryo at 6-somite stage. B) GFP expression in *Tg(sox17:EGFP)* was used to locate the KV within the embryo at 6-somite stage. C) Transmitted light was used to locate the embryo under the objective. D) Merged image showing the location of the KV in a *Tg(sox17:EGFP)* embryo. Bottom panel shows the KV at higher magnification. White arrow indicates the location of the KV. Images were taken by Dr Andrew Nelson. Schematic image in (A) was adapted from: (Kimmel *et al.*, 1995). Scale bar, 250 μm .

We next sought to obtain a clearer image of the GFP+ posterior notochord using high resolution microscopy. As demonstrated below, knockdown of *sox32* results in the loss of GFP+ posterior notochord cells, suggesting that these cells are *sox32*-dependent (Figure 4-7).

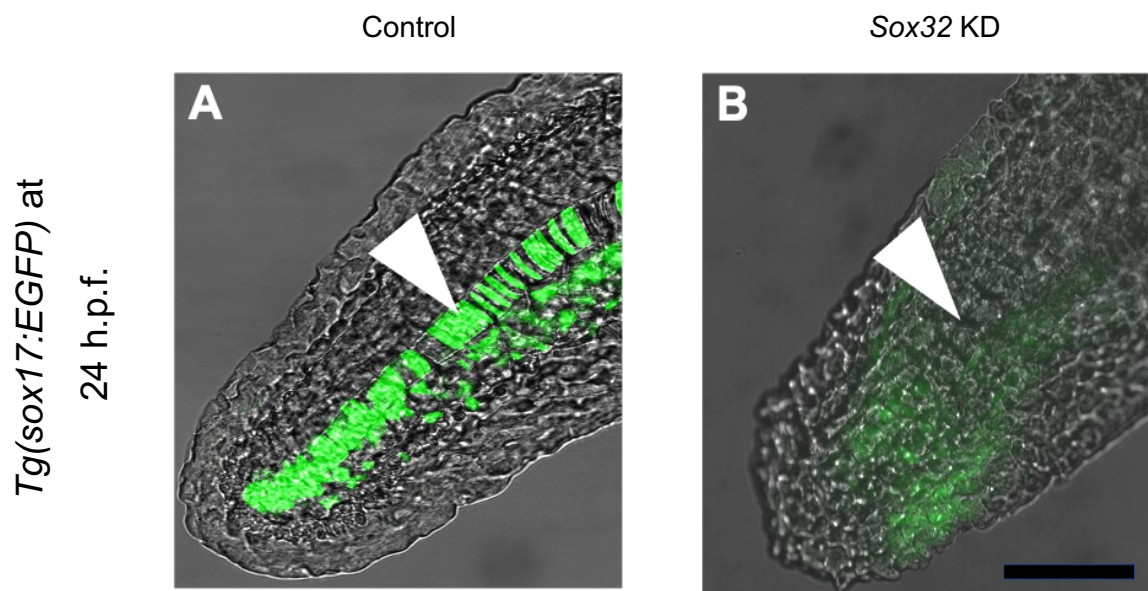


Figure 4-7: High resolution images of control and *sox32* KD embryos at 24 h.p.f. Images demonstrate the lack of GFP+ notochord in *sox32* KD embryos (B) compared to control (A), suggesting that GFP+ cells are *sox32*-dependent. The posterior notochord are shown as an overlay of fluorescence and brightfield images. Scale bar, 0.25 mm.

4.6.1 The GFP+ cells in the posterior notochord are derivatives of the DFCs/KV

In order to investigate the developmental fates of the DFCs and to explore whether these cells contribute to the posterior notochord, a red-fluorescent dye known as Rhodamine-dextran (RD) was injected into the mid-blastula stage (~3 h.p.f.) of *Tg(sox17:EGFP)* embryos to double label the DFCs and KV (Figure 4-8A). At mid-blastula stage, the cytoplasmic bridges are closed between the yolk sac and the blastula cells, but remain open between the yolk sac and dorsal Wilson cells which give rise to DFCs (Amack and Yost, 2004; Warga and Kane, 2018). As demonstrated below, the KV was clearly labelled with RD and GFP at 14-somite stage (16 h.p.f.) (Figure 4-8C-E), and by Prim-22 (35 h.p.f.), the posterior notochord was double labelled with RD and GFP (Figure 4-8G-L). Thus, the data suggests that the GFP+ cells in the posterior notochord are derivatives of DFCs/KV, in accordance with published data (Melby, Warga and Kimmel, 1996).

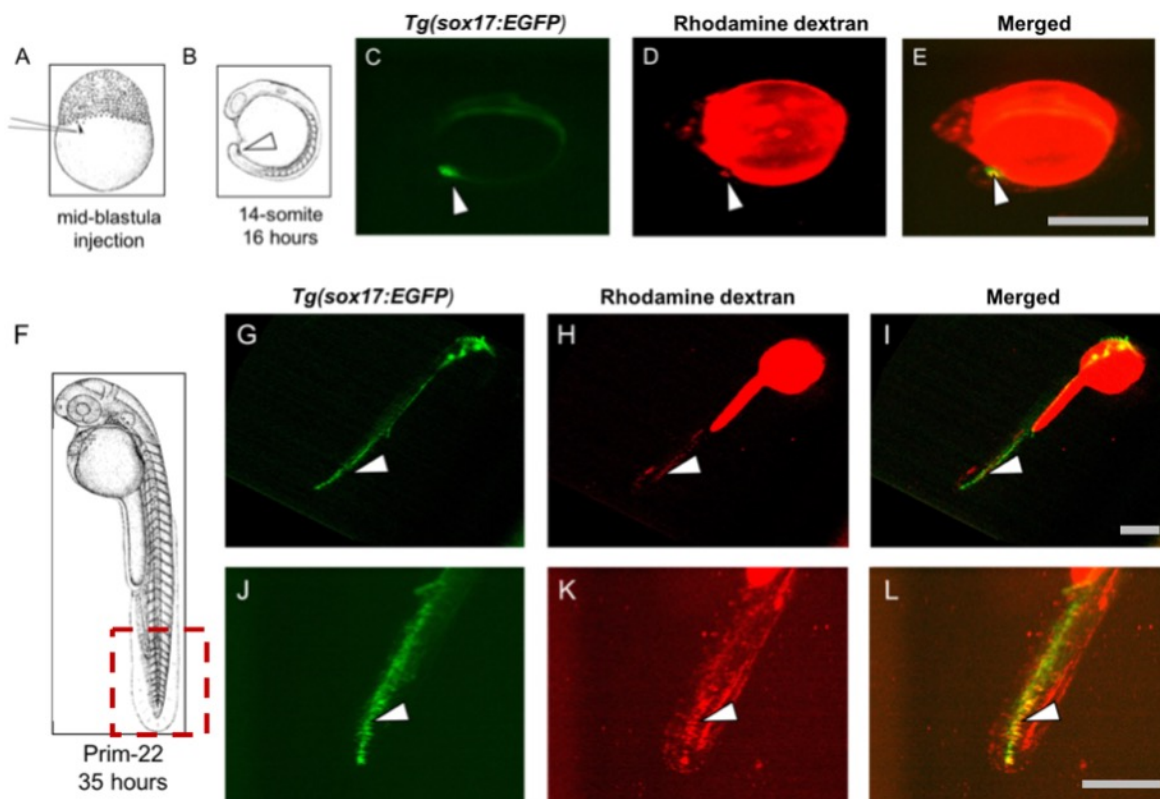


Figure 4-8: Using Rhodamine-dextran (RD) to assay the developmental fate of the KV. (A) Schematic of an embryo injected with RD at the mid-blastula stage (~3 h.p.f.). (B) Schematic of an embryo at 14-somite stage (16 h.p.f.). (C-E) Fluorescent images of the KV double labelled with RD and GFP from *Tg(sox17:EGFP)* transgene. (F) Schematic of an embryo at primordium-22 stage (35 h.p.f.) referred to as “Prim” stage or primordium of the posterior lateral line. Red box highlights the location of the posterior notochord. (G-L) Fluorescence images of the posterior notochord double labelled with RD and GFP. Location of the KV is indicated by white arrow. Images J-L show higher magnification of the posterior notochord. Schematics A,B and F were modified from Kimmel *et al.* (1995). Scale bar, 0.5 mm.

4.7 Sox17+ endoderm cannot be isolated from contaminating mesoderm based on GFP expression

After gastrulation, *sox17* no longer marks the endoderm but marks mesodermal progenitors, particularly cells that contribute to endothelial and hemopoietic lineages. This is demonstrated in Figure 4-5, where GFP appears to mark the endoderm but also ICM, which encompasses blood cells (Chung *et al.*, 2011). Thus, the transgene from *Tg(sox17:EGFP)* marks both the *sox17+* endoderm and *sox17+* mesoderm. There was substantial evidence that led to the hypothesis that *sox17+* endoderm can be isolated from mesoderm lineages based on their differences in GFP expression. For instance, whole mount *in situ* hybridisation of post-gastrula staged embryos revealed that *sox17* is weakly expressed in both ICM and LPM (Chung *et al.*, 2011). In addition to this, RNA-seq expression data from whole embryos revealed that *sox17* is highly expressed in the embryo during gastrulation but is barely detectable beyond this stage (White *et al.*, 2017), (Figure 4-3). Finally, anti-GFP-immunostaining of *Tg(sox17:EGFP)* embryos at 24 h.p.f. (Nelson *et al.*, 2017) and fluorescence images from (Figure 4-5) show strong signal for GFP in the endoderm but weak signal in LPM. Therefore based on the evidence, it was speculated that endoderm expresses high levels of GFP while mesoderm has low GFP expression. For this reason, a trial experiment was performed where GFP+ cells were sorted into two gates using fluorescence activated cell sorting (FACS), where GFP^{high}, which are cells that express high levels of GFP, were speculated to be endoderm and cells with low signal for GFP were speculated to be mesoderm (GFP^{low}) (Figure 4-9).

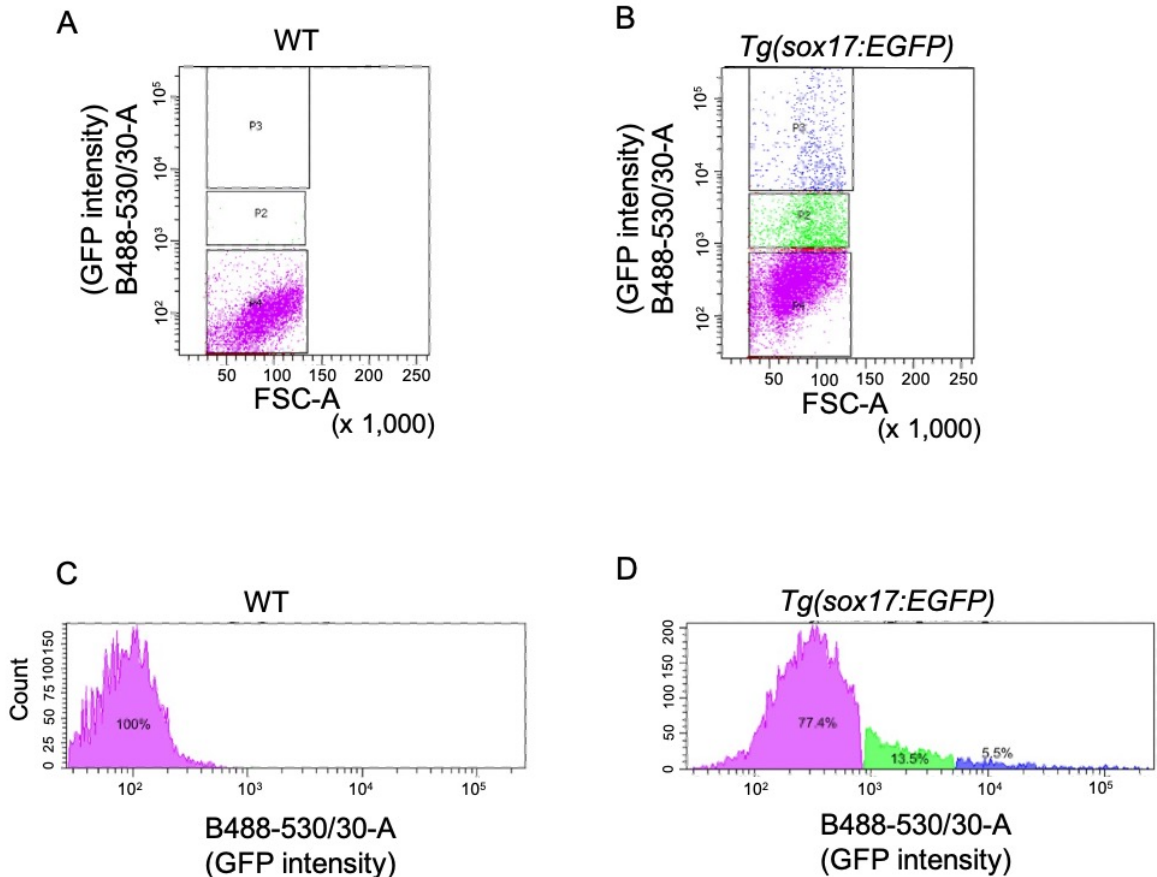


Figure 4-9: Sorting of GFP^{high} and GFP^{low} populations from *Tg(sox17:EGFP)* embryos using FACS. (A) FACS dot plot showing forward scatter for dissociated cells from wild-type embryos at 24 h.p.f.. (B) FACS plot showing forward scatter for dissociated GFP⁺ cells from *Tg(sox17:EGFP)* embryos at 24 h.p.f. Fluorescence scatter was used to separate cells based on GFP intensity. P4, P2 and P3 represent GFP^{negative}, GFP^{low} and GFP^{high} cells respectively. Gates were set to avoid cross-contamination of the respective cell populations. (C) Histogram showing relative fluorescence intensity of wild-type cells; 100% of cells express no GFP. (D) Histogram showing relative fluorescence intensity of GFP⁺ cells. GFP^{negative} cells represent 77.4% of all cells in *Tg(sox17:EGFP)* embryos, while GFP^{low} and GFP^{high} comprise 13.5% and 5.5% of all cells, respectively. Pink represents GFP^{negative}, green represents GFP^{low}, and blue represents GFP^{high} cell populations. FACS performed by Dr Andrew Nelson. FACS figures also from Dr Andrew Nelson.

To test these gating parameters, markers of endoderm and mesoderm lineages were selected for qRT-PCR. The endoderm can be divided into different subpopulations: pharyngeal endoderm, posterior foregut, midgut, and hindgut. Markers of the endoderm were chosen based on their expression in the endoderm, i.e.: whether they are broadly expressed in the endoderm or expressed in the different sub-regions. A schematic of the endoderm sub-regions at 24 h.p.f. is shown below (Figure 4-10).

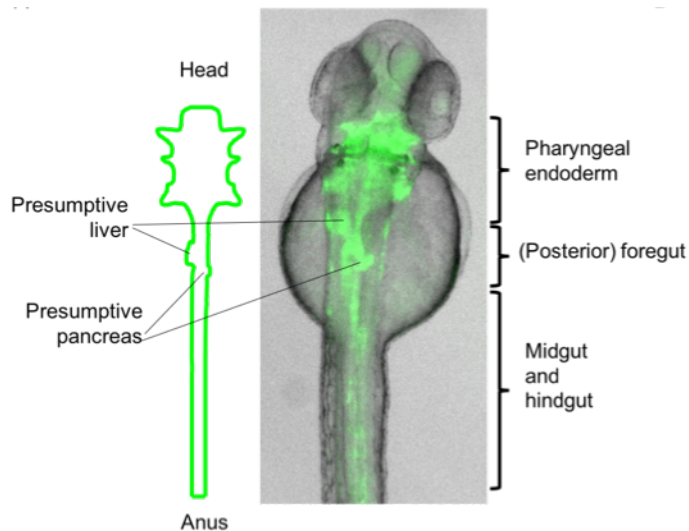


Figure 4-10: Schematic of endoderm sub-regions in *Tg(sox17:EGFP)* embryos. Schematic was based on Image of *Tg(sox17:EGFP)* at 48 h.p.f. from (Figure 4-5) since it provides a clear representation of all endoderm subregions. Pharyngeal endoderm gives rise to thymus and thyroid. Posterior foregut gives rise to liver and pancreas. The hindgut becomes the distal large intestine and the regions in between while the midgut becomes the intermediate zone between the same small and large intestine.

The first markers used for qRT-PCR analysis were the pan-endodermal markers; *egfp* and *foxa2*. The transgene, *egfp*, not only marks the endoderm but also marks ICM, LPM and posterior notochord, as demonstrated previously (Figure 4-5). Based on qRT-PCR, GFP^{high} exhibits high expression of *egfp* compared to GFP^{low} population, which is expected given the gating parameters (Figure 4-11). *Foxa2*, on the other hand, is not only expressed in the endoderm but also in the notochord (Odenthal and Nüsslein-Volhard, 1998), which is mostly devoid of GFP expression except for the posterior notochord cells which are likely derived from KV cells. Similar to *egfp*, *foxa2* expression was also greater in GFP^{high} compared to GFP^{low} (Figure 4-11) in accordance with the hypothesis.

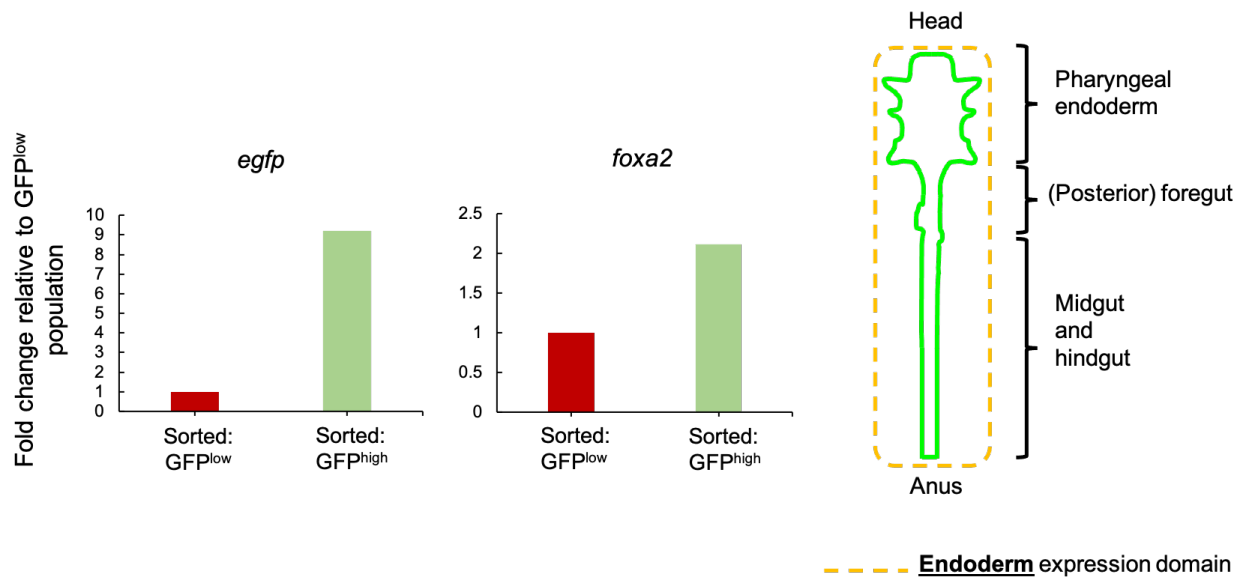


Figure 4-11: Pan-endodermal markers used for qRT-PCR. High level of expression of *egfp* and *foxa2* were observed for GFP^{high} vs GFP^{low}. Data was normalised to *18S*. Fold change was relative to expression in GFP^{low} population. Dotted region represents the expression pattern of *egfp* and *foxa2* in relation to the endoderm. Schematic from Dr Andrew Nelson.

The next markers investigated by qRT-PCR were markers of the pharyngeal endoderm. *T-box1* (*tbx1*) is not only a pharyngeal endoderm marker but also marks both the otic vesicle and heart tube (Choudhry and Trede, 2013; Maier and Whitfield, 2014), which lack GFP expression. The next marker tested by qRT-PCR is *foxa1*, which marks the medial pharyngeal endoderm (Piotrowski and Nüsslein-Volhard, 2000), but also liver and gut (Pézeron *et al.*, 2008; Liu *et al.*, 2016). However, *foxa1* is also expressed in hypochord and floorplate (Manfroid *et al.*, 2007), which are devoid of GFP expression. As shown below, *tbx1* and *foxa1* are highly expressed in GFP^{high} compared to the GFP^{low} cell population (Figure 4-12), which again, is in agreement with the hypothesis as discussed above, i.e.: that it is possible to separate endoderm from mesoderm based on GFP expression pattern.

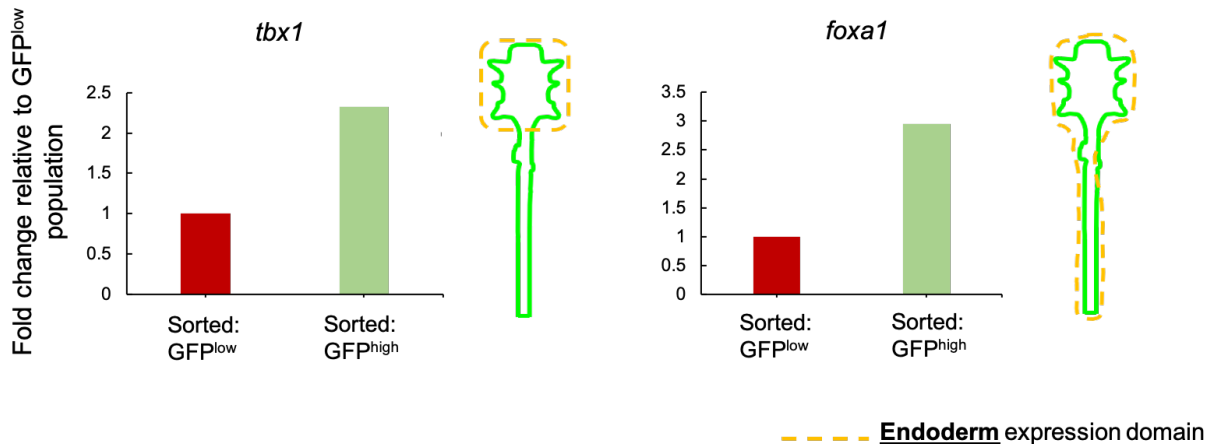


Figure 4-12: Pharyngeal endoderm markers used for qRT-PCR. Expression of *tbx1* and *foxa1* were upregulated in GFP^{high} compared to GFP^{low} population, Data was normalised to *18S*. Fold change was relative to expression in GFP^{low} population Dotted region represents the expression patterns of *tbx1* and *foxa1* in relation to the endoderm. Schematics from Dr Andrew Nelson.

Markers of the posterior foregut were also tested, and these include *gata5*, *haematopoietically expressed homeobox (hhex)*, *foxa3*, *ins* and *pancreatic and duodenal homeobox 1 (pdx1)*. *Gata5* is expressed in the gut, intestinal bulb, liver and pancreas but is also a marker of the heart (Holtzinger, 2005), which is not *sox17+*. Moreover, *ins* and *pdx1* are very specific markers of the endoderm since *ins* is expressed in the pancreas and *pdx1* marks the gut and pancreas (Chung, Shin and Stainier, 2008). Expression of all three markers were enriched in GFP^{high} relative to the GFP^{low} population (Figure 4-13), in line with the hypothesis. In addition, *hhex* is expressed in the liver and pancreas bud (Liu *et al.*, 2016) but also marks the dorsal aorta (Dooley, Davidson and Zon, 2005), ICM and notochord (Farooq *et al.*, 2008) at 24 h.p.f.. Although *hhex* was highly expressed in GFP^{high} compared to GFP^{low}, it is also a marker of the ICM at 24 h.p.f., which means that we cannot conclude that the GFP^{high} cell population is entirely composed of endodermal progenitors.

Another marker tested by qRT-PCR was *foxa3* which marks the liver, gut and pancreas (Farooq *et al.*, 2008) but also the posterior notochord and hatching gland (Stafford, 2006). Though *foxa3* is fundamentally an endoderm marker and therefore expected to be enriched in GFP^{high} population, the results did not align with our hypothesis. One possible theory for this could be due to the presence of GFP⁺ cells in the posterior notochord as a result of transdifferentiation of the KV. Thus, this result highlights the drawback to using this approach with the aim of separating *sox17+* endoderm from contaminating mesoderm based on GFP expression.

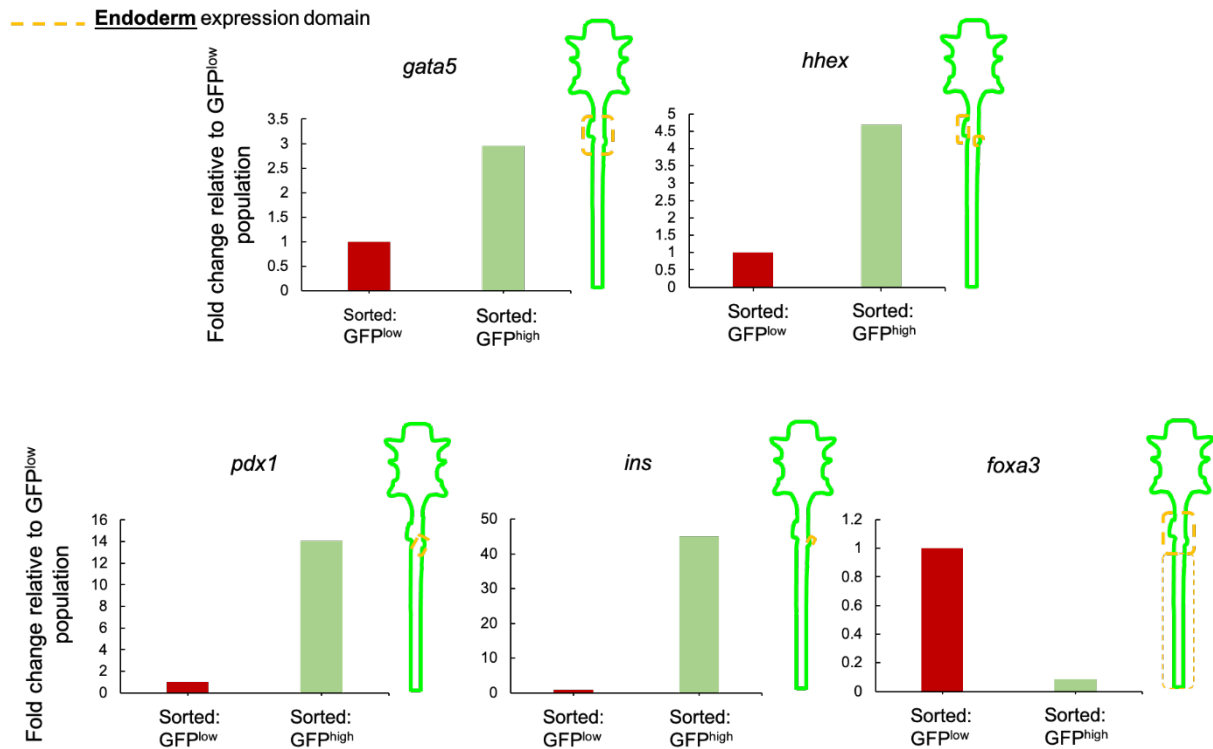


Figure 4-13: Posterior foregut markers used for qRT-PCR. High expression level was observed for *gata5*, *hex*, *pdx1* and *ins* for GFP^{high} vs GFP^{low}, while *foxa3* was more enriched in GFP^{low} population. Data was normalised to 18S. Fold change was relative to expression in GFP^{low} population. Dotted region represents the expression pattern of all markers relative to the endoderm. Schematics from Dr Andrew Nelson.

The final endoderm marker test is *growth hormone receptor b (ghrb)*. *Ghrb* marks the liver and intestine but also hindbrain, myotomes, spinal cord, lens and notochord (relatively low) (Di Prinzio *et al.*, 2010). As observed below, the expression of *ghrb* was greater in GFP^{high} compared to GFP^{low} (Figure 4-14), in line with the hypothesis.

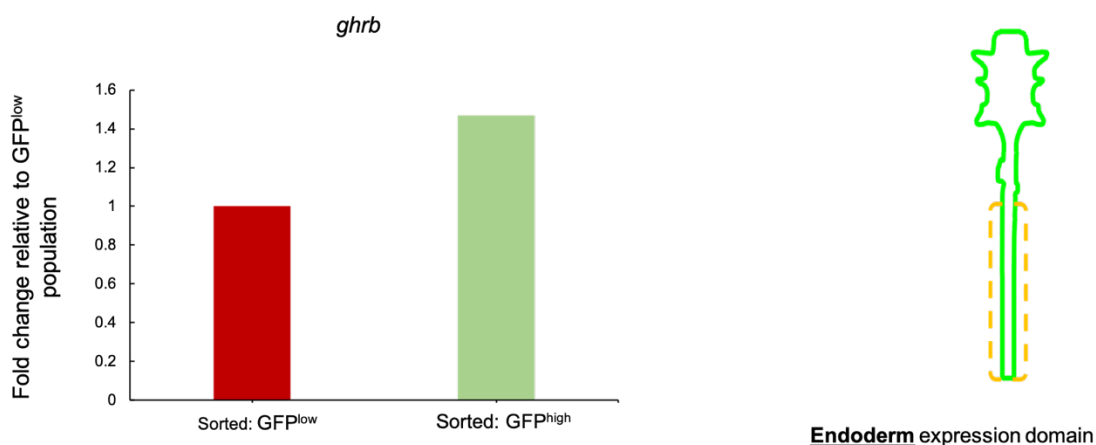


Figure 4-14: Midgut and hindgut marker used for qRT-PCR. *Ghrb*, was more expressed in GFP^{high} compared to GFP^{low} population. Data was normalised to 18S. Fold change was relative to expression in GFP^{low} population.

Dotted region represents the expression pattern of *ghrb* relative to the endoderm. Schematic from Dr Andrew Nelson.

In addition to the above markers, genes expressed in the hemopoietic, and endothelial cell lineages were also tested by qRT-PCR. These include, *sox17*, *gata1a* and *T-cell acute leukaemia 1 (tal1)*. At 24 h.p.f., *sox17* marks the LPM, ICM and cardinal vein (Chung *et al.*, 2011), while *gata1a* and *tal1* are expressed in the ICM (Thisse and Thisse, 2005). Based on the hypothesis, we expect that these markers to be enriched in GFP^{low} population. Whilst this holds true for *tal1*, *sox17* and *gata1a* showed surprising results; *sox17* was highly expressed in GFP^{high} population, while no difference was observed for *gata1a* (Figure 4-15).

In conclusion, while gating on GFP expression pattern showed some promising results, it was far from perfect. For instance, qRT-PCR analysis suggests that some endoderm cells likely express low levels of GFP while some mesodermal progenitors express high levels of GFP. This result therefore justifies the need for an alternative approach to separating *sox17*+ endoderm from *sox17*+ mesoderm.

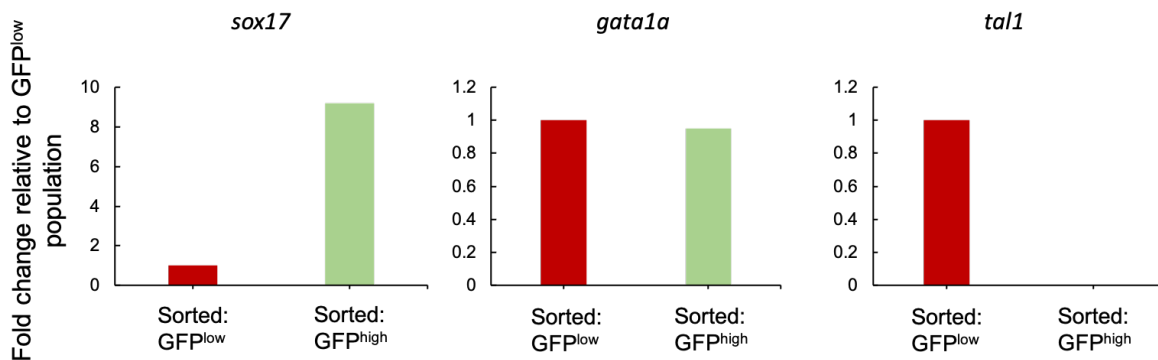


Figure 4-15: Mesoderm-specific markers tested by qRT-PCR. *Tal1* was enriched in GFP^{high} compared to GFP^{low} population, *sox17* was more highly expressed in the GFP^{high} population, while barely any difference was observed for *gata1a* between the two different populations. Data was normalised to *18S*. Fold change was relative to expression in GFP^{low} population.

4.8 Successfully isolating GFP+ endoderm from contaminating mesoderm by in-crossing of transgenic lines

An alternative route to isolating *sox17*+ endoderm from contaminating mesoderm is to use fluorescent proteins to mark the different germ layers. For this reason, we crossed *Tg(kdrl:mCherry)^{ci5}* fish line (Proulx, Lu and Sumanas, 2010) with *Tg(gata1a:dsRed)^{sd2}* line (Fish *et al.*, 2008) and *Tg(sox17:EGFP)* line (Mizoguchi *et al.*, 2008) to make a transgenic line

which expresses *GFP* in *sox17*+ cells, *mCherry* in endothelial cells derived from LPM and *dsRed* in haemopoietic lineages derived from ICM. This line, hereafter referred to as the triple transgenic line, was derived by in-crossing *Tg(kdr1:mCherry)^{ci5}* to *Tg(gata1a:dsRed)^{sd2}* to make double heterozygous progeny, which were then in-crossed to make double homozygous embryos before they were subsequently outcrossed to *Tg(sox17:EGFP)* to make the triple transgenic line (Figure 4-16). Because *dsRed* and *mCherry* were evident under fluorescence microscopes at 28 h.p.f., this stage was chosen for all downstream analyses.

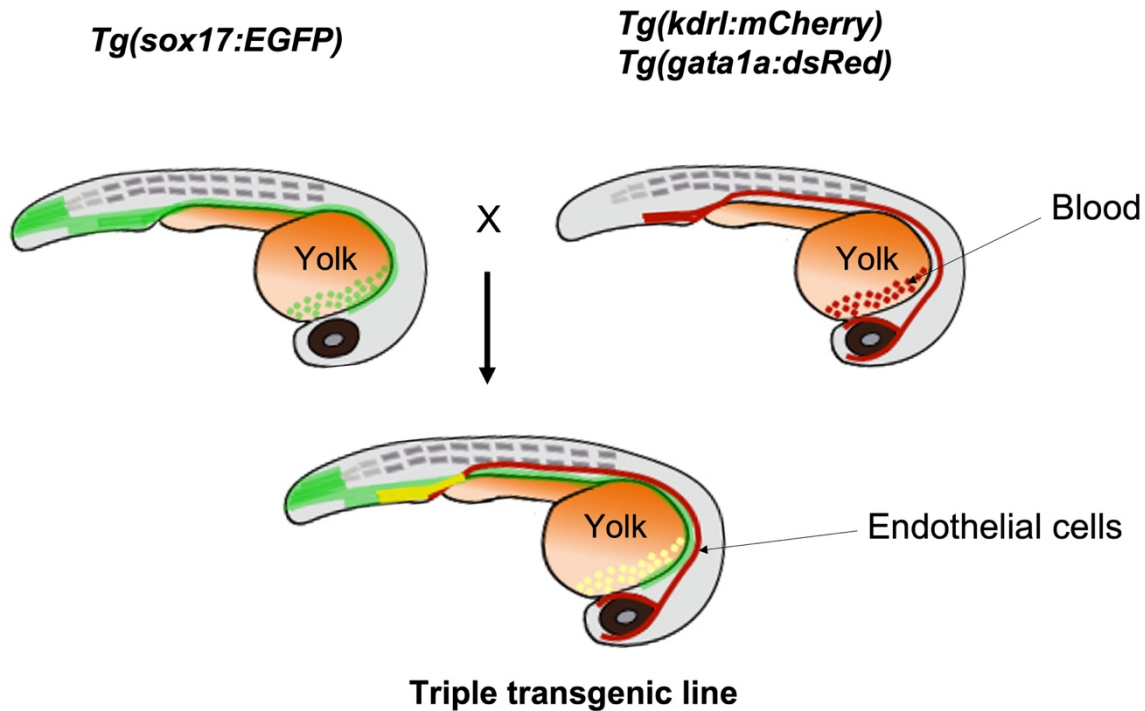


Figure 4-16: Schematic images demonstrating the breeding procedures required to attain the triple transgenic embryos at 28 h.p.f.. GFP+ cells represent either *sox17*+ endoderm or *sox17*+ mesoderm. Red fluorescent cells represent either the endothelial (*kdr1* + cells) or blood cells (*gata1a* + cells). Yellow cells represent the mesoderm population which express both *sox17* and either *gata1a* or *kdr1*.

Therefore, cells that express both GFP and *dsRed/mCherry* are likely ICM and LPM progenitors while GFP only cells represent endoderm and KV derivatives. An image of a triple transgenic embryo can be viewed below (Figure 4-17).

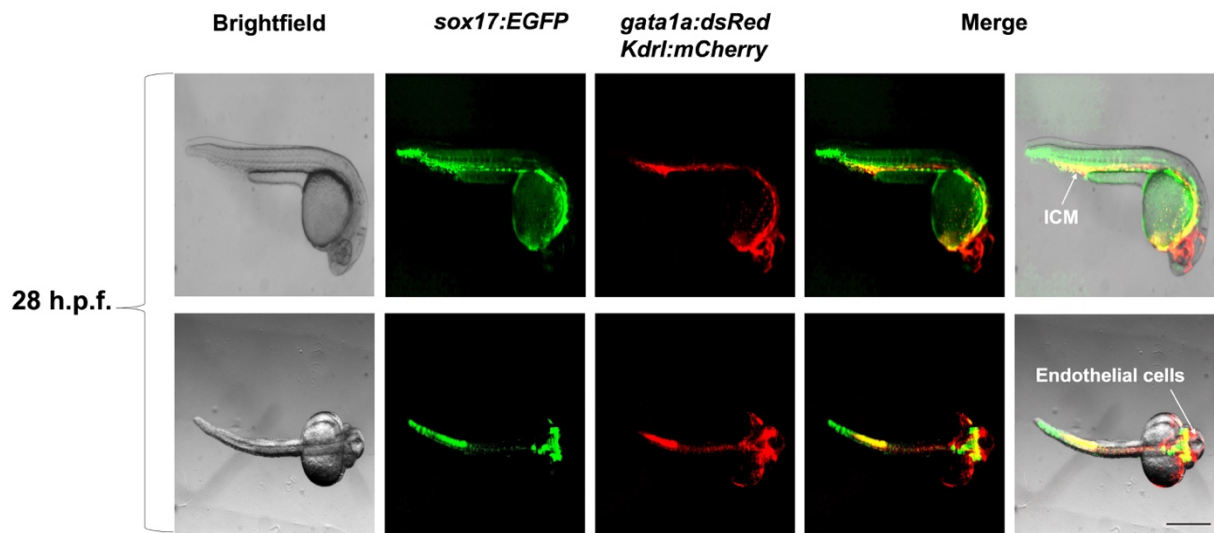


Figure 4-17: Brightfield and fluorescence microscopy images of a triple transgenic embryo at 28 h.p.f.. The embryos was imaged laterally (top panel) and dorsally (bottom panel). The ICM and endothelial cells are indicated by white arrows. Scale bar: 0.5 mm.

4.9 Validating the use of triple transgenic line in enriching for endoderm at 28 h.p.f.

Prior to performing any downstream analyses, we first wanted to validate the use of the triple transgenic line in isolating *sox17*⁺ endoderm from mesoderm. We firstly used FACS to gate cells into three different cell populations: *sox17*⁺ endoderm (*sox17E*), which express only *GFP*, *sox17*⁺ mesoderm (*sox17M*), which express both *GFP* and *mCherry/dsRed*, and *sox17* negative (*sox17N*), which are devoid of *sox17* expression but may express *mCherry/dsRed* (Figure 4-18). We then validated the different cell populations by using endoderm and mesoderm markers for qRT-PCR.

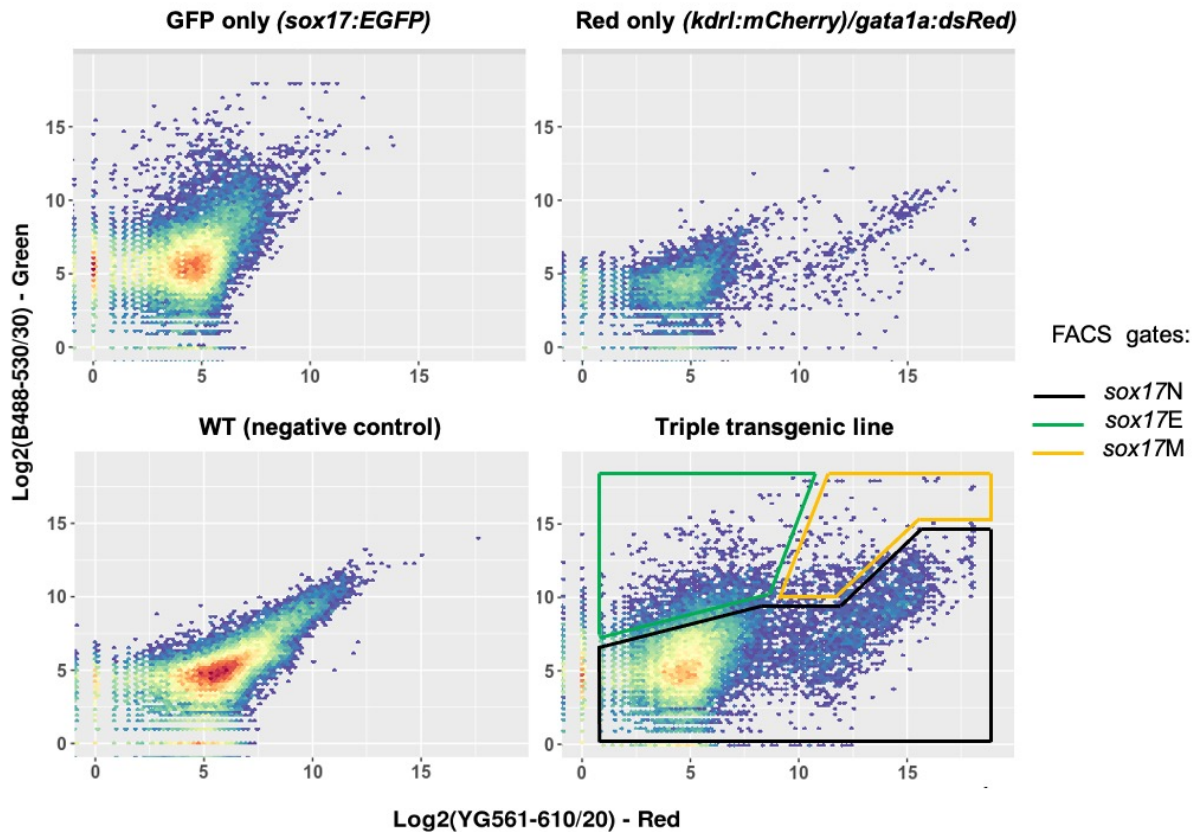


Figure 4-18: Sorting cells from the triple transgenic line into three cell populations (*sox17N*, *sox17E* and *sox17M*) using FACS. Cells from WT, *Tg(sox17:EGFP)* and *Tg(kdr1:mCherry/gata1a:dsRed)* embryos were used as gating controls. Cells were sorted on GFP fluorescence (bandpass filter 530/30nm) and/or dsRed/mCherry fluorescence (bandpass 610/20nm). Gates were set to avoid cross-contamination of the respective cell populations. Sorting gates for the different cell populations are indicated. FACS was performed by Dr Mark Walsh. FACS plots also from Dr Mark Walsh.

As shown by qRT-PCR, expression of endoderm markers which includes *foxa1*, *pdx1* and *ins* were highly expressed in *sox17E* compared to *sox17M* and *sox17N*. *Ins* and *pdx1* are specific markers of the endoderm while *foxa1* is a pharyngeal endoderm marker but is also expressed in hypochord and floorplate, as previously mentioned (Manfroid *et al.*, 2007), which are devoid of GFP expression and thus, *foxa1* was expected to be enriched in *sox17E*. On the other hand, *foxa2* was highly expressed in *sox17N* compared to the other cell populations, an expected result given that *foxa2* also marks the notochord which is mostly devoid of GFP. Therefore, the qRT-PCR results are in line with our hypothesis and reveal that endoderm markers are enriched in *sox17E* population (Figure 4-19)

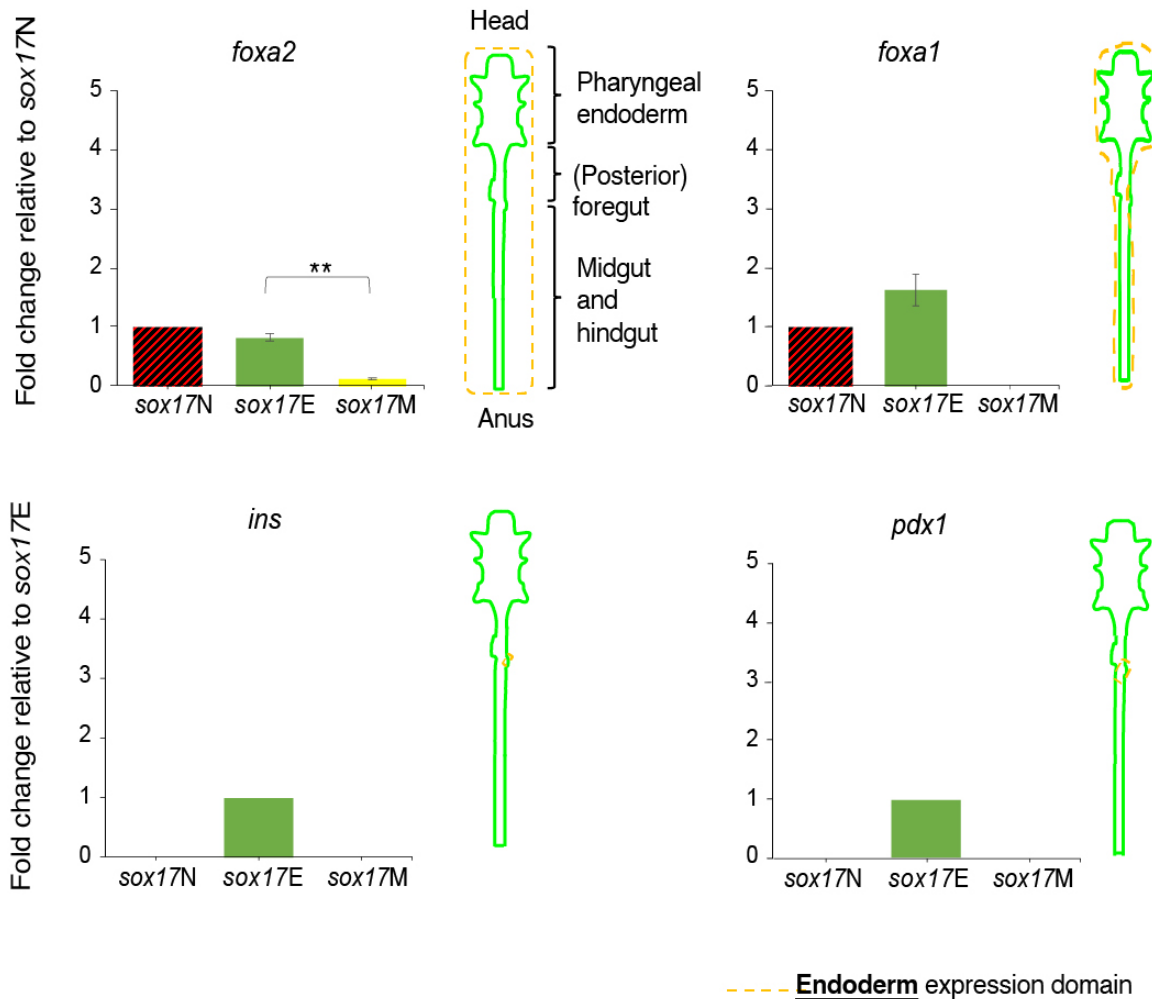


Figure 4-19: qPCR analysis of endoderm markers on *sox17N*, *sox17E* and *sox17M* cell populations. *Foxa1*, *ins* and *pdx1* were enriched in *sox17E* compared to the other cell populations; ** $p = 0.003$. Error bars represent standard deviation. Data was normalised to *18S*. Two biological replicates were generated for each sample. Fold change was relative to expression in *sox17N* or *sox17E* population as indicated by bar graphs. Dotted region represents the expression pattern of each gene relative to the endoderm

On the other hand, mesoderm markers, *sox17* as well as *gata1a* and *kdrl*, which are markers of blood and endothelial cells, respectively, were highly expressed in *sox17M* compared to *sox17N* and *sox17E*, in line with the hypothesis (Figure 4-20)

Therefore, the result demonstrate that the triple transgenic line can be used to isolate *sox17+* endoderm from contaminating mesoderm, and thus, can be used to study the endoderm during early organogenesis.

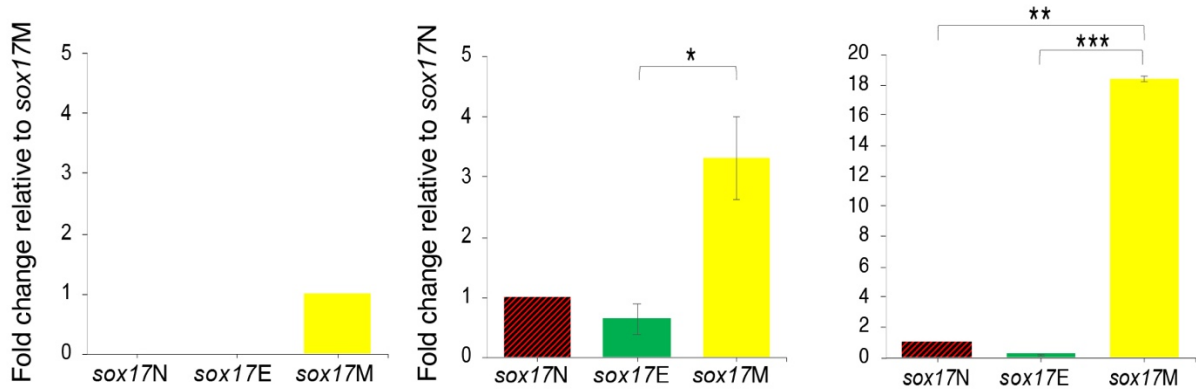


Figure 4-20: qPCR analysis of mesoderm markers on *sox17N*, *sox17E* and *sox17M*. All markers were enriched in *sox17M* compared to the other cell populations;*** p = 0.0001, **p = 0.002, *p=0.04. Error bars represent standard deviation. Data was normalised to *18S*. Two biological replicates were generated for each sample. Fold change was relative to expression in *sox17N* or *sox17M* population as indicated by bar graphs.

Given that it was possible to isolate *sox17+* endoderm from contaminating mesoderm using the triple transgenic line, we proceeded to perform ATAC-seq and RNA-seq on the different cell populations mentioned above. A schematic of the experimental procedure is highlighted below (Figure 4-21).

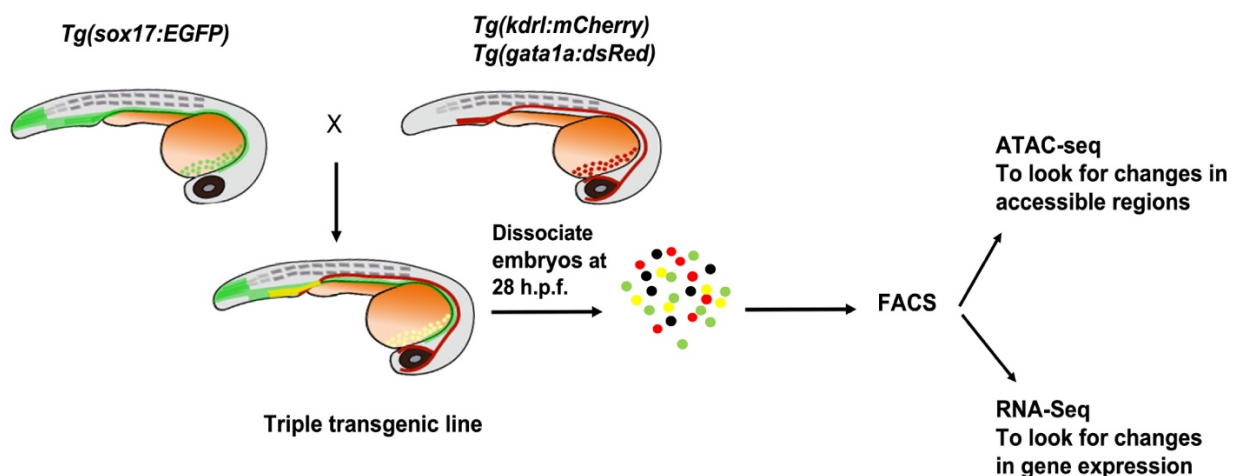


Figure 4-21: Schematic of experiment performed using triple transgenic embryos at 28 h.p.f. ATAC-seq and RNA-seq were performed on separate sorts. Green cells represent either *sox17+* endoderm or *sox17+* mesoderm. Red cells represent either endothelial or blood lineages. Non-fluorescent or “black cells” represent GFP/dsRed/mCherry negative cells. Cells were sorted based on fluorescence: Green only cells likely represent the endoderm/ KV derivatives (*sox17E* population), green and red cells (i.e.: yellow cells) represent mesoderm lineages (*sox17M* population), i.e.: endothelial and blood cells that express both GFP and dsRed/mCherry while red/black cells, likely represent cells that lack *sox17* expression (*sox17N* population).

4.10 ATAC-seq reveals endoderm-specific enhancers at 28 h.p.f.

To enable the study of the chromatin landscape and to characterise endoderm-specific enhancers at the onset of organogenesis, ATAC-seq was performed on cells isolated from the triple transgenic line: *sox17N*, *sox17E* and *sox17M*. Two biological replicates were used, and the insert size distributions obtained for all three cell types can be viewed below (Figure 4-22). The insert size distributions of all three samples follow the expected pattern shown in Buenrostro *et al.* (2015) and are therefore considered to be good quality.

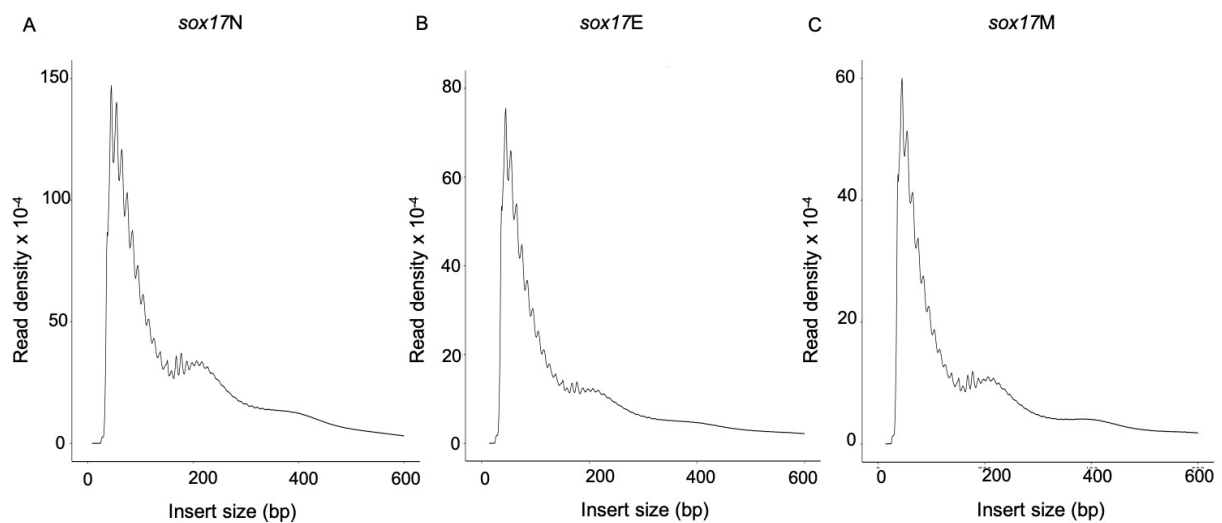


Figure 4-22: Insert sizes as obtained by ATAC-seq analysis. Insert size distributions of mapped reads in (A) *sox17N*, (B) *sox17E*, and (C) *sox17M*. Data was filtered to remove duplicated reads. Insert sizes were extracted by merging biological replicates.

We next sought to investigate the location of all accessible regions for each sample relative to genomic features, as previously discussed. To find regions of accessible chromatin for each sample, the peak caller MACS2 was used. Two biological replicates were used for each condition (Table 4.1). Following peak calling, individual peaks from each biological replicate were merged and these were used as a representative of the total number peaks for each condition (Table 4.2) and were used for all subsequent images using IGV as shown below. *Sox17M* constitutes a minor cell population compared to *sox17+* endoderm and *sox17N* cells and thus, only a small number of peaks were called.

Table 4.1: Number of called peaks for each biological replicate per condition

Sample	Number of called peaks
sox17N 1	105,686
sox17N 2	56,071
sox17E 1	39,410
sox17E 2	65,599
sox17M 1	25,704
sox17M 2	42,252

Table 4.2: Number of called peaks from MACS2

Sample	Number of called peaks
sox17N	131,480
sox17E	75,854
sox17M	46,414

As demonstrated below, annotated peaks from each sample group exhibit different distribution patterns relative to genomic features (Figure 4-23). For *sox17N*, more annotated peaks were located in distal intergenic regions compared to the other sample groups, while *sox17M* has more annotated peaks in proximal promoter regions, which is located within 3 kb of transcription start site (TSS). This may suggest that most *sox17N* and *sox17E* CRMs are located at distal regions with respect to genes. However, this difference in peak distribution may also be due to discrepancies in called peak numbers. The number of called peaks in *sox17M* is far less than the other cell populations, and this may be due to ICM and LPM progenitors exhibiting different accessibility profiles which results in a loss of signal due to ensemble averaging. On the other hand, *sox17N* has more called peaks which may suggest that it constitutes a more complex, heterogeneous population of cells.

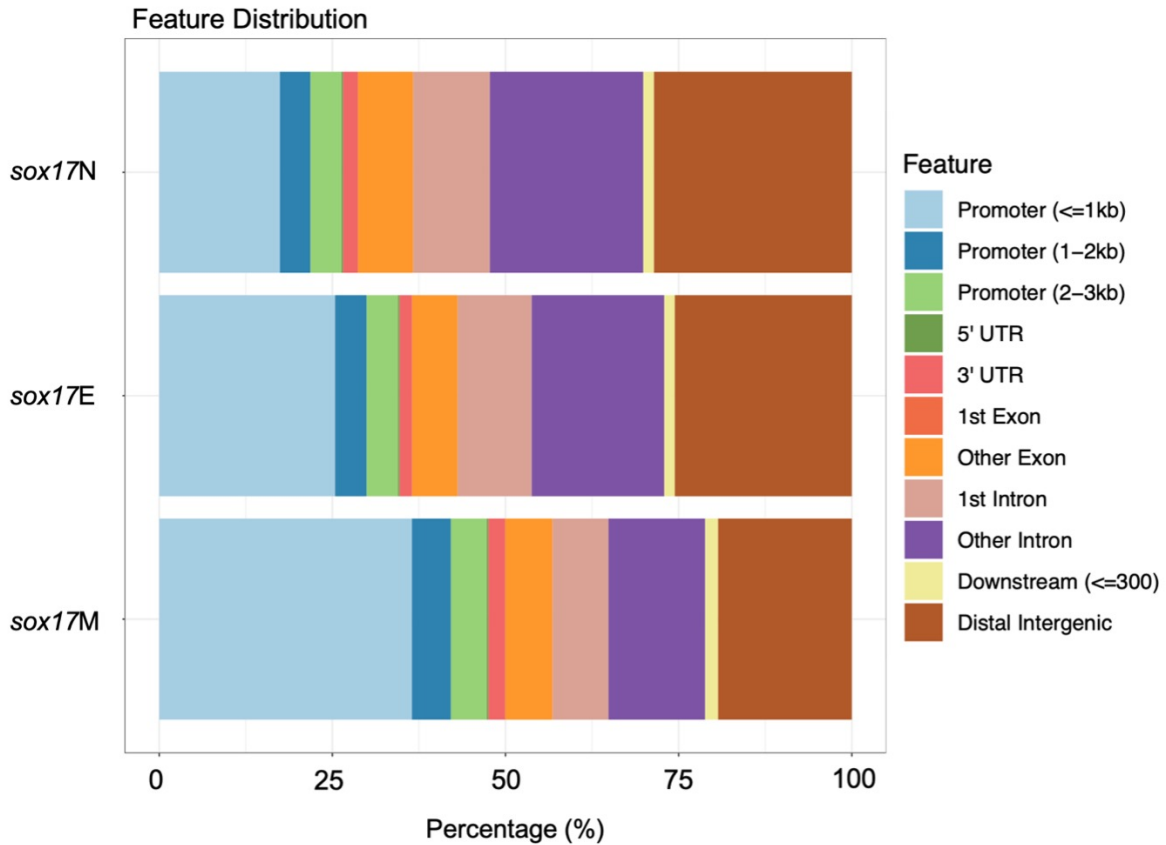


Figure 4-23: Distribution of genomic locations of accessible regions with respect to annotated genes for *sox17N*, *sox17E* and *sox17M*. Promoters are located within -3 kb to +3kb of TSS. Distribution is based on merged biological replicates for each sample group. Peaks were annotated based on the following priority: promoter, 5'UTR, 3'UTR, exon, intron, downstream and intergenic, where downstream is defined as the downstream of gene end.

4.10.1 The accessibility profile differs between all three cell populations

In order to find regions that exhibit greater accessibility in the endoderm compared to the other germ layers, Diffbind was used. A detailed account of Diffbind has already been described in (3.5.1). A strong correlation was observed between *sox17N* and *sox17E*, which means that the MACS2 scores for each peak-set is similar between the two sample groups, while *sox17M* is distinct since each called peak has a different MACS2 score compared to peaks from the other cell populations (Figure 4-24).

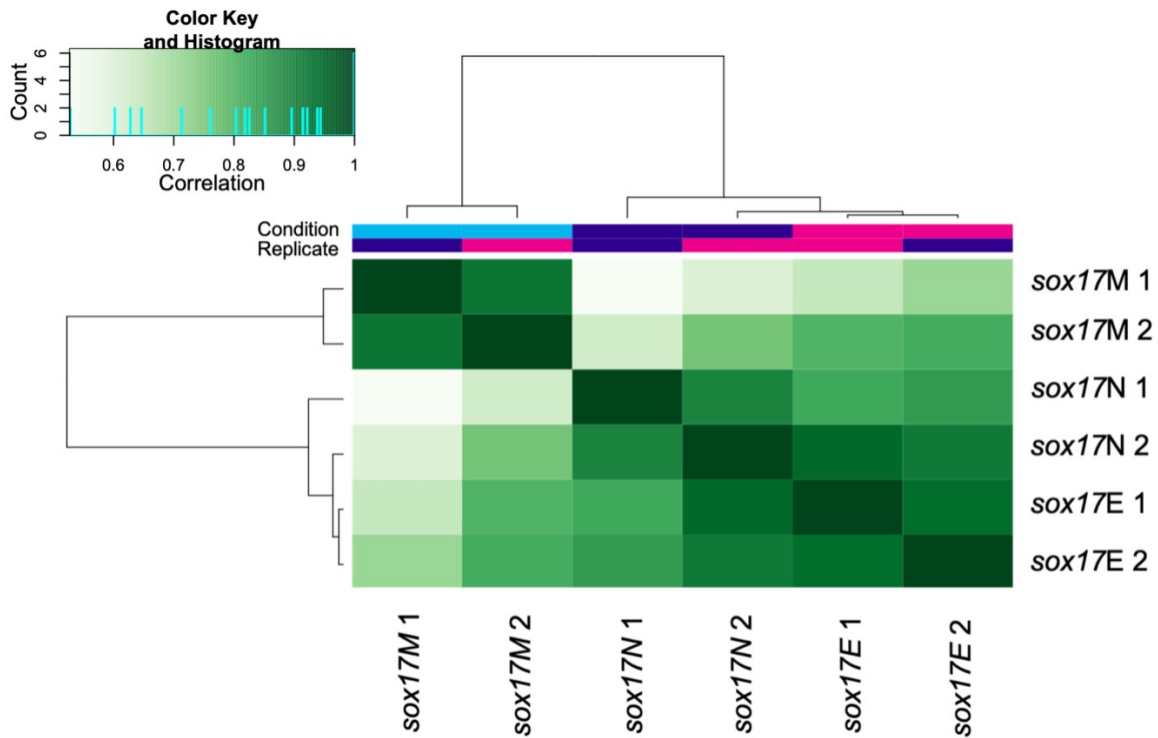


Figure 4-24: Pearson correlation heatmap based on MACS2 scores. According to hierarchical clustering, peaksets in *sox17E* and *sox17N* exhibit similar MACS2 scores while *sox17M* is distinct from the other two cell populations. Dark green and light represent high and low similarity in MACS2, respectively.

Following this, an alternative Pearson correlation heatmap is plotted based on the total number of mapped reads per genomic interval as opposed to (Figure 4-24). For this step, Diffbind uses the peaksets that were obtained from Figure 4-24 to represent all regions that are accessible within all sample groups. The number of consensus regions across all sample groups is 51,817. The number of total mapped reads can be viewed below (Table 4.3).

Table 4.3: Total number of mapped reads

Sample	Replica number	Total number of mapped read counts
sox17N	1	115,198,342
	2	44,640,558
sox17E	1	14,120,110
	2	57,803,282
sox17M	1	11,761,930
	2	45,117,334

As demonstrated by hierarchical clustering, *sox17E* 2 and *sox17N* 2 cluster together while *sox17M* 1 clusters with *sox17M* 2, suggesting that they have similar numbers of reads that

map each consensus region. In addition, the number of mapped reads were also similar between *sox17N* 1 and 2. However, *sox17E* 1 was distinct from the other samples since it was placed in its own cluster and was more similar to *sox17M* 1 and 2 than *sox17E* 2. Thus, this likely suggests that there might be discrepancies in reads counts at each genomic interval between *sox17E* 1 and 2.

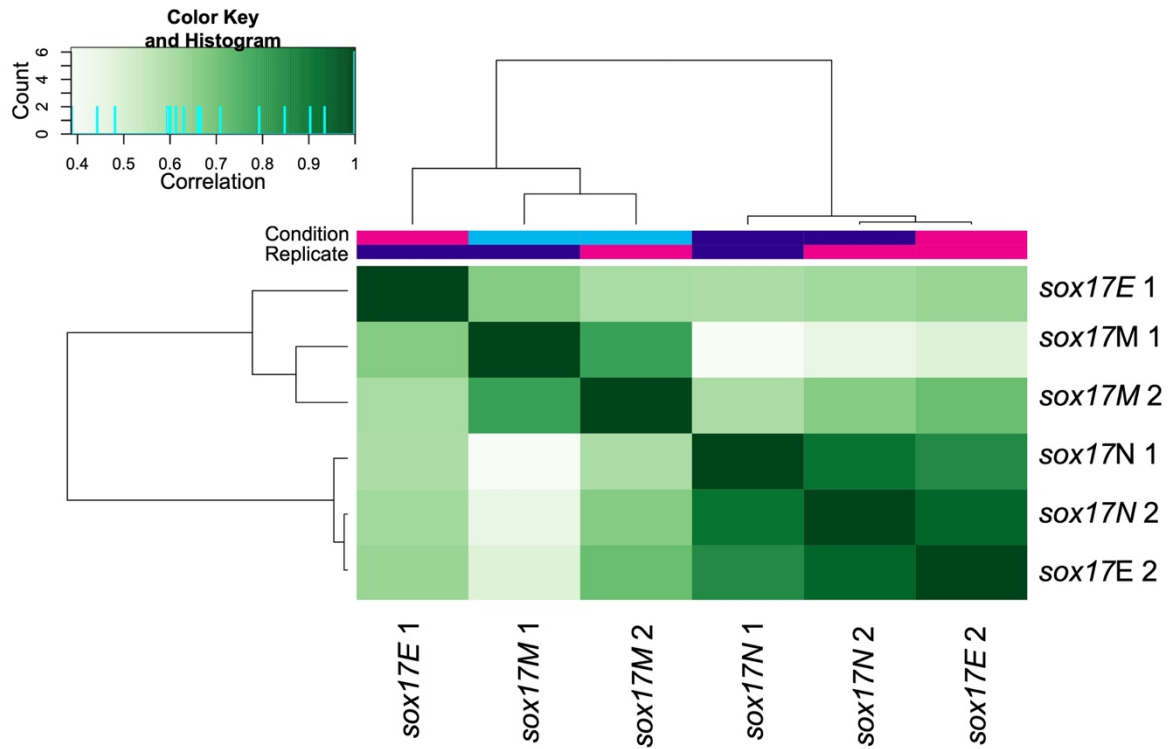


Figure 4-25: A Pearson correlation heatmap based on read counts. The number of reads that map each genomic interval is similar between *sox17N* 1 & 2 and *sox17E* 1, and also between two *sox17M* 1 and *sox17M* 2. However, there are differences in the number of mapped reads between *sox17E* 1 and 2, which may suggest that they are not equivalent to each other. Dark and light green represent high and low similarity in read counts, respectively.

To find endoderm-specific CRMs, a differential accessible heatmap was used to compare statistically significant DARs between two sets of sample groups: *sox17N* vs *sox17E*, *sox17E* vs *sox17M*, and *sox17N* vs *sox17M* (Figure 4-26). This was performed by first normalising the data by DESeq2 to correct for differences in library size across all sample groups, as mentioned previously. The differences in DARs between two sets of sample groups can be viewed in Table 4.4 and is also displayed by differential heatmaps as shown in Figure 4-26. A list of top DARs can also be viewed in Table 7.14, Table 7.15 and Table 7.16.

Table 4.4: The number of DARs between two sets of sample groups

Sample group comparison	Total number of DARs	DARs of each sample group
<i>sox17N vs sox17E</i>	2,544	<i>sox17N</i> = 2,164 <i>sox17E</i> = 380
<i>sox17E vs sox17M</i>	17,413	<i>sox17E</i> = 12,021 <i>sox17M</i> = 5,392
<i>sox17N vs sox17M</i>	20,190	<i>sox17N</i> = 15,700 <i>sox17M</i> = 4,490

The differential heatmaps reveal major differences in the accessibility profile between all three sample groups. This suggests that each cell population relies on different CRMs for controlling gene expression during development.

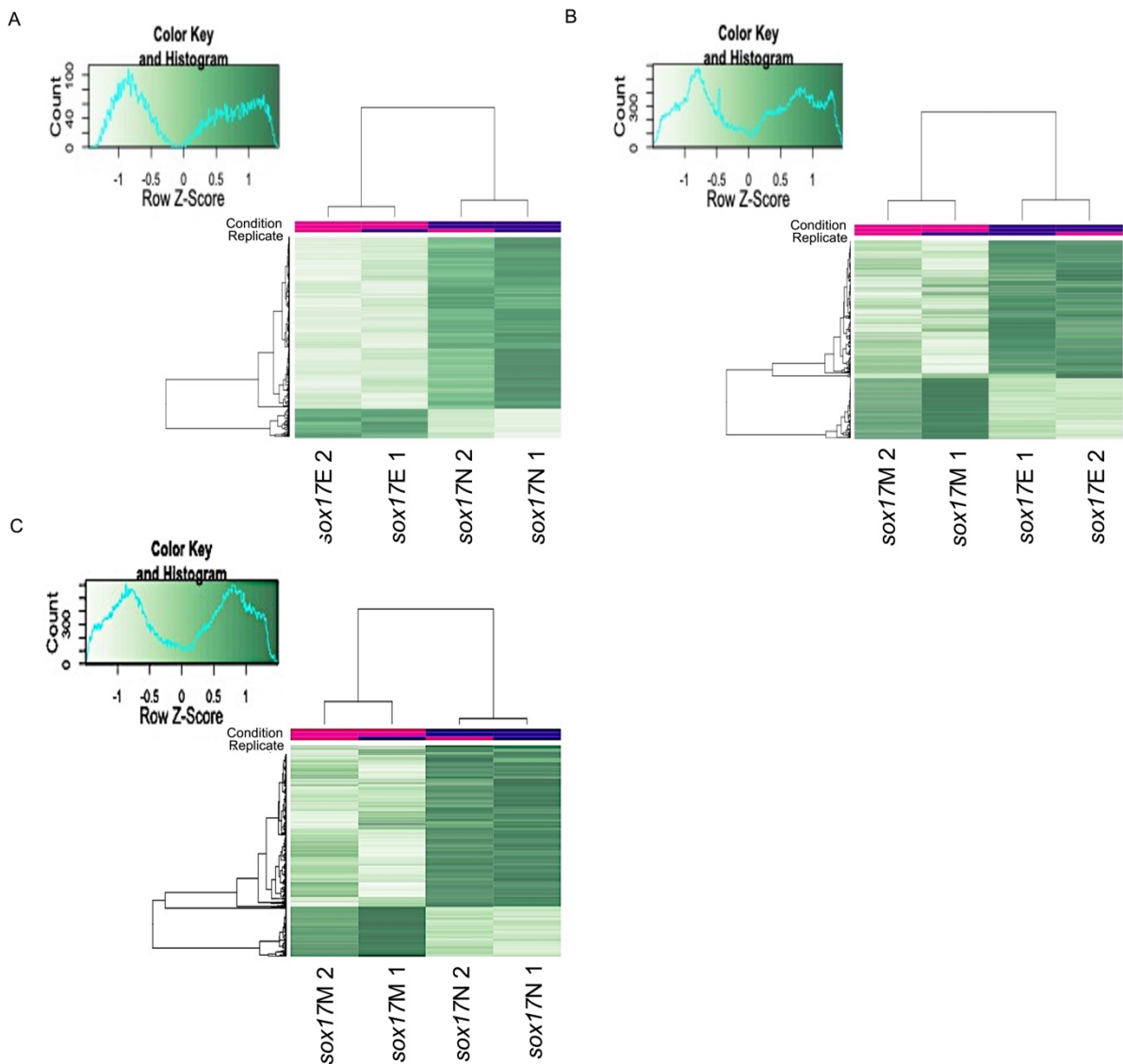


Figure 4-26: Differential accessible heatmaps showing differences in DARS between two sets of samples. Differential heatmaps can be viewed for (A) *sox17E* and *sox17N*, (B) *sox17E* and *sox17M* and (C) *sox17M* and *sox17N*. FDR cut-off of 0.05 was used for all comparisons between sample groups. However, changing input parameters to a number of FDR cut-offs (0.01 to 0.0001) did not change the final result. The Pvalue cut-off for *sox17E* vs *sox17N* was 0.00355, for *sox17N* vs *sox17M* was 0.0199 and for *sox17E* vs *sox17M* was 0.0168. Heatmaps were clustered on samples (columns) and accessible regions (rows), and rows were also scaled based on Z-score. Regions of higher accessibility are shown in Dark green, and light green regions represent low accessibility.

An example of a gene which exhibits higher accessibility in *sox17E* compared to *sox17N* and *sox17M* is *six1b* (Figure 4-27). In zebrafish, there are two close homologs of mammalian *Six1*: *Six1a* and *six1b*. In mammals, *Six1* is required for patterning of the pharyngeal endoderm (Talbot *et al.*, 2019). Although their role is not yet established in the zebrafish endoderm, *six1a/b* appear to be expressed in the pharyngeal endoderm at 24 h.p.f., according to single-cell RNA-seq data (Wagner *et al.*, 2018).

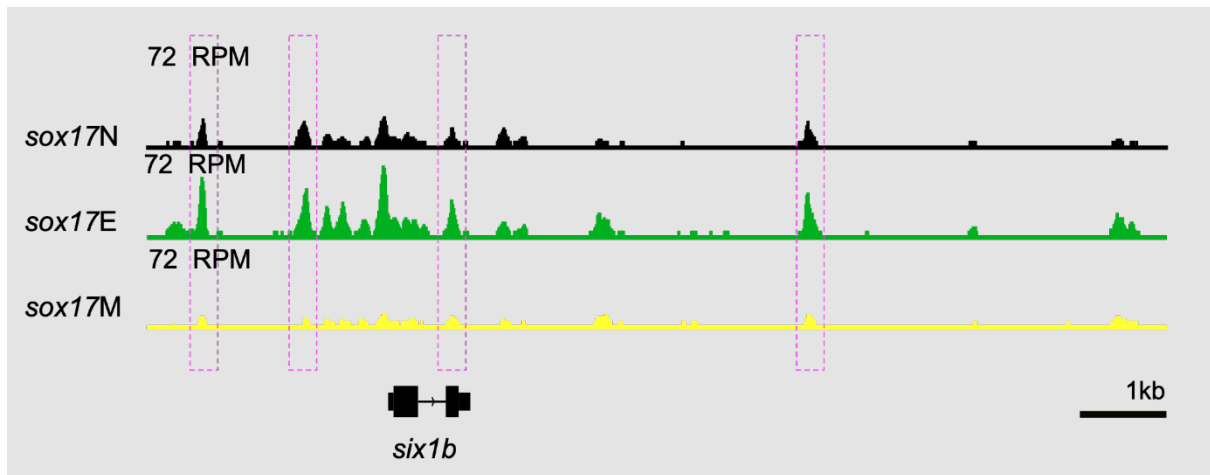


Figure 4-27: An example of a gene that is highly accessible in *sox17E* compared to *sox17N* and *sox17M* is *six1b*. Reads were adjusted to correct for library size. Dashed purple regions likely represent putative enhancers of *six1b* in the endoderm. RPM = reads per million.

4.10.2 The accessibility profile of *sox17E* is similar to *sox17N* while *sox17M* is more distinct

Next we sought to investigate the similarity in the accessibility profiles between all three sample groups: *sox17N*, *sox17E* and *sox17M*. This is to help determine whether the same regions are accessible within the different sample groups or whether they each have their own distinct CRMs during development. This was achieved by using SeqMINER which was used to compare read densities from *sox17N*, *sox17E* and *sox17M* across DARs from all three sample groups. The first analysis compares regions of higher accessibility in *sox17E* vs *sox17N* to mapped reads from all three samples. As observed by heatmaps below, *sox17E* exhibits higher accessibility for *sox17E* DARs, but these regions were also accessible in all three cell populations (Figure 4-28A). On the other hand, *sox17N* exhibits greater accessibility for *sox17N* DARs, but *sox17N* DARs were also accessible in *sox17E* but not in *sox17M* (Figure 4-28B). The results suggest that different CRMs are accessible in *sox17+* mesoderm compared to the other cell populations while *sox17+* endoderm and non-*sox17+* cells share putative enhancers during development.

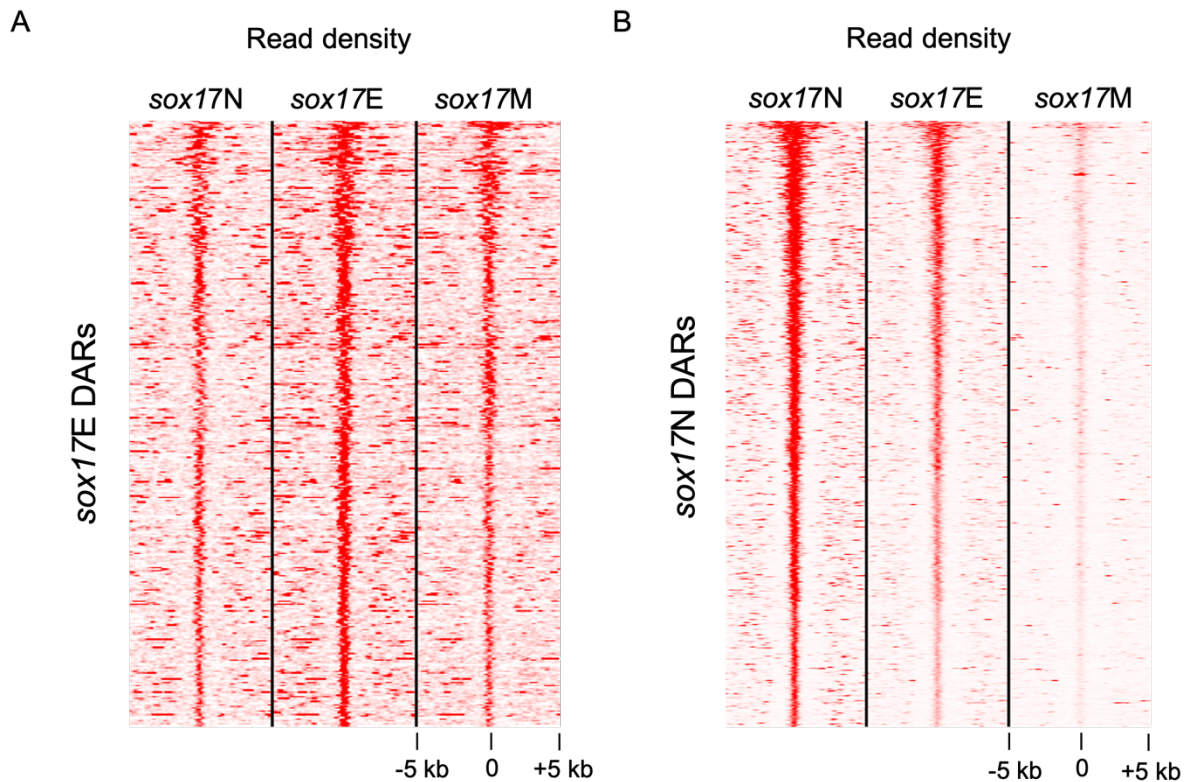


Figure 4-28: Heatmaps showing read densities of *sox17N*, *sox17E* and *sox17M* surrounding *sox17N* and *sox17E* DARs. (A) Heatmaps displaying mapped reads from all three sample groups across (A) *sox17E* DARs and (B) *sox17N* DARs. Read densities with a window of +/- 5kb across each DAR centre was selected for this analysis. Plots from Dr Andrew Nelson.

The next analysis was used to compare read densities of all three sample groups across *sox17M* vs *sox17E* DARs. This analysis was used to investigate the similarities in the accessibility profile between *sox17+* mesoderm and *sox17+* endoderm population. As demonstrated in Figure 4-29A, *sox17N* and *sox17E* were both accessible for *sox17E* DARs, while *sox17M* was barely accessible. On the other hand, heatmaps in Figure 4-29B demonstrate that *sox17M* exhibits high accessibility for *sox17M* DARs while only some *sox17M* regions were also accessible in *sox17N* and *sox17E*. The results suggest that endoderm-specific enhancers are also accessible in non-*sox17+* cells, and that some *sox17+* mesoderm CRMs are accessible in other cell populations. This means that non-*sox17+* cells and *sox17+* endoderm likely share similar regulatory regions while the regulatory programme is distinct for *sox17+* mesoderm during development.

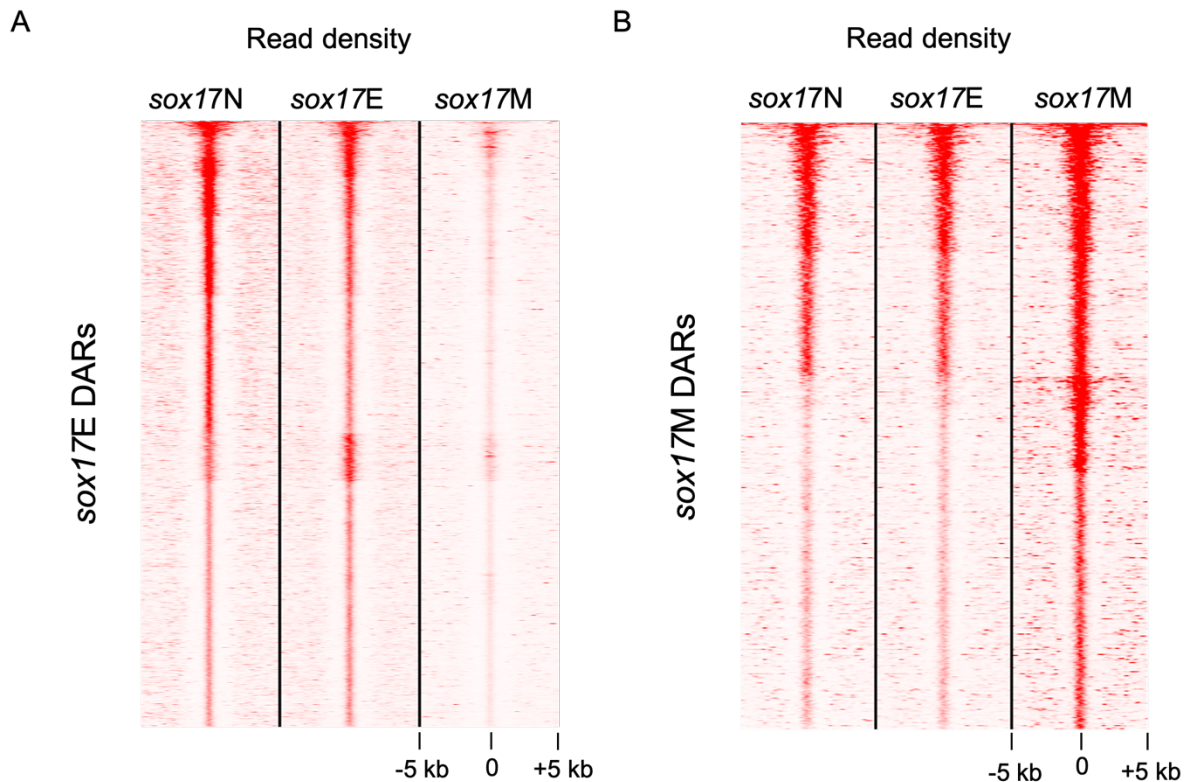


Figure 4-29: Heatmaps showing read densities of *sox17N*, *sox17E* and *sox17M* surrounding *sox17E* and *sox17M* DARs. (A) Heatmaps displaying mapped reads from all three sample groups across (A) *sox17E* DARs and (B) *sox17M* DARs. Read densities with a window of +/- 5kb across each DAR centre was selected for this analysis. Plots from Dr Andrew Nelson.

The last analysis was used to compare read densities from all three sample groups across *sox17M* vs *sox17N* DARs. This analysis was used to investigate the similarities in accessibility profile between *sox17+* mesoderm and *sox17* negative population. As demonstrated below, *sox17N* exhibit high accessibility for *sox17N* DARs but said DARs are also accessible in *sox17E* but not *sox17M* (Figure 4-30A). On the other hand, only some *sox17M* DARs were also accessible in *sox17N* and *sox17E* (Figure 4-30B). Thus, in line with previous results, *sox17+* endoderm CRMs are also accessible in non-*sox17+* cells while most of the *sox17+* mesoderm regulatory regions are not accessible in the other cell populations.

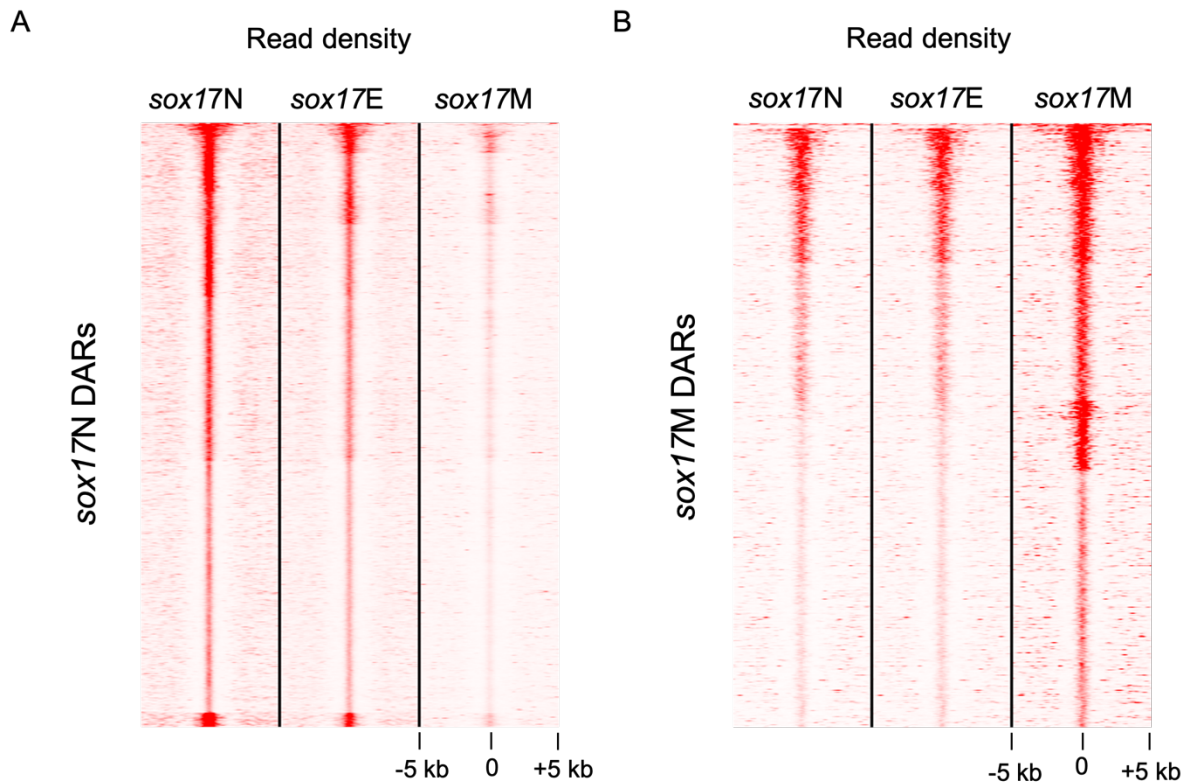


Figure 4-30: Heatmaps showing read densities of *sox17N*, *sox17E* and *sox17M* surrounding *sox17N* and *sox17M* DARs. (A) Heatmaps displaying mapped reads from all three sample groups across (A) *sox17N* DARs and (B) *sox17M* DARs. Read densities with a window of +/- 5kb across each DAR centre was selected for this analysis. Plots from Dr Andrew Nelson.

4.10.3 *Sox17E* is enriched for terms associated with the endoderm

Though a number of studies have found developmentally important enhancers during pancreatic or midgut development in humans and mouse, respectively (Wang *et al.*, 2015; Banerjee *et al.*, 2018), CRMs required to drive the morphogenesis of other endoderm-specific organs is still lacking. To find potentially important endoderm-specific CRMs at the onset of organogenesis, GREAT was used. A detailed account of GREAT has already been explained in 3.5.3. In addition to finding potentially important CRMs, GREAT was used to look at over-represented terms associated with genes proximal to DARs from all three sample groups: *sox17N*, *sox17E* and *sox17M*. This was used to further validate the use of the triple transgenic line in separating *sox17+* endoderm from the other cell lineages. As demonstrated below, DARs from *sox17E* were associated with the endoderm (Figure 4-31). *Sox17E* DARs were also proximal to genes that function in the central nervous system (CNS) and mesoderm, in line with previous data (4.10.2) but these terms were not as over-represented in the sample as anatomical terms representing endoderm (Figure 7-4). Some of these genes can be seen in Table 7.17. Though the vast majority of DARs from *sox17N* were associated with genes expressed in the CNS, the digestive system and pharynx were also over-represented in the

sample group. This is likely because most of the genes are not specific to the digestive system and pharynx but are also expressed in the brain and CNS. These include *T-box transcription factor (tbx20)* (Thisse *et al.*, 2001), *FGF receptor-like protein 1a (fgfr1a)* (Hall *et al.*, 2006), *insulin gene enhancer protein isl-1 (isl1)* (Binot *et al.*, 2010) and *protein jagged-1b (jag1b)* (Porazzi *et al.*, 2012). In addition, the pan-endodermal marker, *foxa2*, is also expressed in the CNS during development (Odenthal and Nüsslein-Volhard, 1998). Some of the genes proximal to *sox17N* can be viewed in Table 7.18. On the other hand, no endoderm terms were found for *sox17M* cell population, but instead, *sox17M* DARs were proximal to genes of the blood and endothelial lineages, some of which can be viewed in Table 7.19. The results therefore confirm the use of the triple transgenic line in enabling the finding of endoderm-specific enhancers at the onset of organogenesis. However, the results also suggest that endoderm-expressed genes seem to play a role in the CNS during development, while CRMs in *sox17+* mesoderm are unique to blood and endothelial lineages.

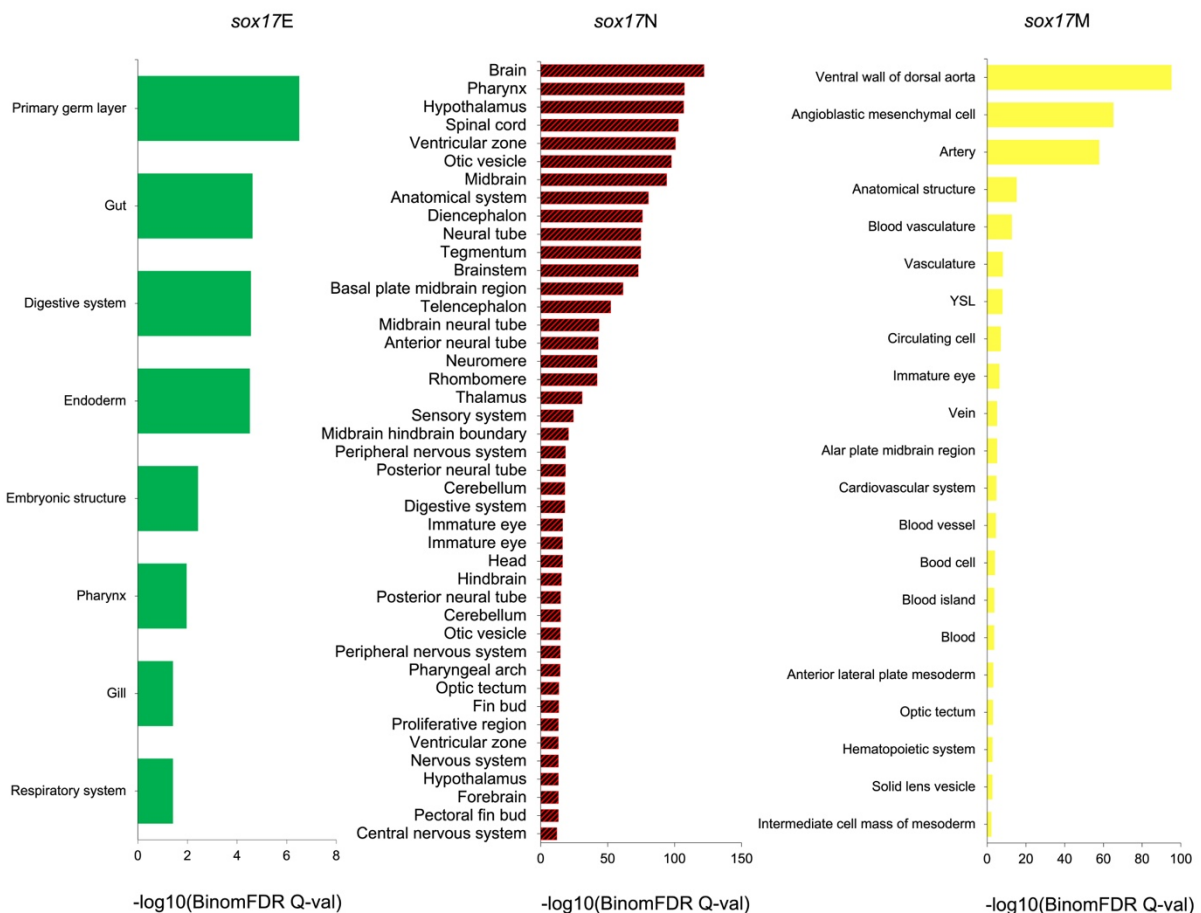


Figure 4-31: Bar graph showing over-represented anatomical terms associated with *sox17E*, *sox17N* and *sox17M*. Selected endoderm terms are shown for *sox17E*. DARs from *sox17M* vs *sox17N* were proximal to genes associated with the haemopoietic lineages while DARs from *sox17N* vs *sox17E* were proximal to genes associated with the CNS. Though the digestive system and pharynx were also over-represented in *sox17N*, this is likely because endoderm-expressed genes are not specific to the endoderm since they also play a role in the CNS.

4.10.4 Known endoderm TF motifs are enriched in *sox17E* DARs

We next sought to look for TF binding motifs in *sox17E* DARs that are likely bound by factors that play a role in driving endoderm organogenesis. To look for enriched TF binding motifs, HOMER was used. Our analysis reveal that *sox17E* DARs were enriched for known TFs that play a role in endoderm formation during development (Figure 4-32 and Table 7.20). Amongst the most enriched motifs belong to Tead proteins 1, 3 and 4. In the nucleus, Yes-associated protein 1 (Yap1), a final effector of HIPPO signalling interacts with TEA domain (Tead) binding family (Tead1-4) TFs (Zhao *et al.*, 2008) and together regulate cell proliferation, differentiation and survival (Harvey, Pflieger and Hariharan, 2003; Zhao *et al.*, 2007). In zebrafish, Tead/Yap1 interaction is critical for endoderm maintenance and formation with evidence that perturbation to Tead/Yap1 signalling results in defects in anterior endoderm formation due to increased apoptosis of endoderm progenitors (Fukui *et al.*, 2014). Notably, Tead1 and 3 are also expressed in the DFCs and KV. Knockdown of Tead1a and Tead3a in the DFCs result in a reduction in KV size at 12 somite stage (~14 h.p.f) and DFC number by bud stage (10 h.p.f.) but also reduction in cilia organisation and motility by 10 somite stage (12 h.p.f.) and laterality defects by 20 h.p.f. (Fillatre *et al.*, 2019). Though *sox17E* likely includes sorted transdifferentiated cells from the KV, it is unlikely that Tead proteins play a role in the KV or laterality establishment at 28 h.p.f..

Though Tead factors are within the top ten most enriched motifs, the most enriched motifs in *sox17E* DARs belong to factors that are not specific to the endoderm or that are critical in other lineages. Binding sites of AP-1 family of transcription factors which consist of Jun (c-Jun, JunB and JunD) and Fos (c-Fos, Fra-2 and FosB) proteins were among the most enriched motifs in *sox17E* DARs (Table 7.20). Indeed, motifs for Fosl2, Fra2, Jun-AP1 and JunB were the most enriched in *sox17E* DARs and these factors are critical in various developmental, physiological, and pathological processes. For instance, the binding site of Fosl2, also known as Fra2, is the most enriched motif in *sox17E* DARs and it is believed to be implicated in the regulation of cardiomyocyte differentiation in zebrafish (Jahangiri *et al.*, 2016). In mice, Fra2/Fosl2 is involved in a wide variety of processes including proliferation or differentiation, and elevated Fra2/Fosl2 expression has been described in several chronic lung diseases including pulmonary fibrosis and asthma (Birnhuber *et al.*, 2019). In addition, *junb* gene which has two orthologs in zebrafish-*junba* and *junbb*, has been shown to be a critical regulator of blood vessel development in zebrafish (Kiesow *et al.*, 2015). Similarly, in mice, inactivation of *Junb* results in embryonic lethality caused by impaired vasculogenesis and angiogenesis in the extra-embryonic tissues (Passegué *et al.*, 2002). Studies in mice have also shown that mice lacking *Jun* die during embryonic development and show defects in heart morphogenesis

and liver formation, while liver specific knockout of *Jun* results in defective liver regeneration (Passegué *et al.*, 2002; Sadler *et al.*, 2007). The role of *Jun* appears to be consistent in zebrafish since upregulation of *jun* was observed following partial hepatectomy and during liver regeneration (Sadler *et al.*, 2007). In zebrafish, while *Jun* is not liver specific, it is required in the liver for outgrowth and regeneration (Sadler *et al.*, 2007). Though it is likely that some of the most enriched motifs in *sox17E* DARs bind to factors that play important roles during embryonic development, in particular during liver formation, they are not specific to the endoderm since they are required for other processes including angiogenesis and cardiomyocyte differentiation.

Other enriched motifs in *sox17E* DARs include Sine oculis homeobox (SIX) family of transcription factors, which are involved in the patterning of the different germ layers and also pharyngeal endoderm during development (Zou *et al.*, 2006). In mice, *Six1* is co-expressed with *Six4* in the pharyngeal endoderm and loss of both proteins results in a complete loss of thymus and parathyroid markers (Zou *et al.*, 2006). However, there are currently no information on the roles of *Six1*, *Six2* and *Six4* in the zebrafish endoderm though zebrafish orthologues of *Six1*, 2 and 4 are expressed in the pharyngeal endoderm at 24 h.p.f, based on single-cell RNA-seq data (Wagner *et al.*, 2018). Thus, our data reveals a number of putative CRMs that may be required for thymus and parathyroid organogenesis in zebrafish.

Sox17E DARs were also enriched for binding motifs of TFs which are critical for pancreas development and these include *Pdx1* and *Ptf1a* (Decker *et al.*, 2006). In humans, mutations in *PTF1A* results in pancreatic agenesis (Sellick *et al.*, 2004), a congenital disease where the pancreas fails to develop during embryonic growth, and the function of *Ptf1a* seems to be highly conserved in zebrafish since *ptf1a* morphants exhibit agenesis of the exocrine pancreas (Jiang *et al.*, 2008). Likewise, in humans, *PDX1* mutations result in endocrine pancreas dysfunction and hyperglycaemia (Babu, Deering and Mirmira, 2007) and thus, *PDX1* plays an essential role in pancreas development and β cell function and survival (Fujimoto *et al.*, 2009). In accordance with this, null mutations of *pdx1* in zebrafish results in perturbation of β cell number, insulin levels and acinar differentiation, with mutants also exhibiting disrupted glucose homeostasis (Kimmel *et al.*, 2015). Thus, our data suggests the presence of CRMs that may be important for pancreas maintenance and maturation during development.

All in all, our data suggests the presence of a number of different TF binding motifs that are likely required for the morphogenesis of a number of different endodermal-specific organs in zebrafish. Additionally, a number motifs in *sox17E* DARs are also likely bound by factors that are not specific to the endoderm since they function in other lineages. However, the data also

suggests the presence of motifs that are likely bound by factors that play essential roles in endoderm organ formation in other species, but little information is known about their role in the endoderm in zebrafish, which includes all Six factors. Thus, this suggests a potential novel function of these TFs in endoderm organ formation in zebrafish during development.

sox17E

Motif	Name	P-value
	TEAD1(TEAD)	1e-14
	TEAD4(TEAD)	1e-12
	TEAD3(TEAD)	1e-12
	Six4(Homeobox)	1e-10
	Six1(Homeobox)	1e-9
	Six2(Homeobox)	1e-9
	FOXA1(Forkhead)	1e-5
	Gata4(Zf)	1e-3
	Foxa3(Forkhead)	1e-3
	Foxa2(Forkhead)	1e-3
	Ptf1a(Basic helix-loop-helix)	1e-2
	Gata6(Zf)	1e-2
	Pdx1(Homeobox)	1e-2

Figure 4-32: A selective number of TF binding motifs in *sox17E* DARs that are likely critical during endoderm organogenesis in zebrafish. TF motifs in *sox17E* DARs were first selected based on GO analysis on previously identified over-represented genes using GREAT in 4.10.3. TF motifs were subsequently selected based on their role in the endoderm in zebrafish or other species during early organogenesis. Motifs are ranked by statistical significance.

4.11 Endoderm-expressed transcripts are upregulated in *sox17E*

Thus far, ATAC-seq analysis has demonstrated that *sox17E* DARs are proximal to genes that play important roles in the endoderm during organogenesis. However, this gene list is computationally-derived and thus, further experiments are required to test whether they are a true reflection of the biology. To test this, RNA-seq was performed on *sox17E* and *sox17N* cell populations in duplicates as demonstrated below. RNA-seq libraries were generated from *sox17N* and *sox17E* cell populations and genes that were upregulated for both sample groups were identified. Only significant genes were selected for all RNA-seq downstream analysis (FDR < 0.05) and comparison of the genes revealed that *sox17N* and *sox17E* represent two different cell identities (Figure 4-33A). RNA-seq analysis was also used to investigate genes that play essential roles during endoderm organogenesis. This was achieved by GO analysis using DAVID on *sox17E* upregulated transcripts (Huang, Sherman and Lempicki, 2009b, 2009a) to identify over-represented terms associated either endoderm formation or endoderm organogenesis, some of these terms can be viewed Figure 7-5. Genes associated with endodermal terms can be viewed in Figure 4-33B. All upregulated transcripts for *sox17E* can be viewed in Table 7.21.

Based on GO analysis genes associated with embryonic liver development include genes such as neuron navigator 3 (*nav3*), which regulates liver organogenesis during development, *prostaglandin-endoperoxide synthase 2a* (*ptgs2a*) which is expressed in the gills, gut and liver and *HNF1 homeobox Ba* (*hnf1ba*) which is required for proper β -cell numbers, pancreas as well as hepatic specification (Ishikawa *et al.*, 2007; Lokmane *et al.*, 2008; Klein *et al.*, 2011; Lancman *et al.*, 2013). Other genes associated with liver formation include *epithelial cell adhesion molecule* (*epcam*) which binds and stabilises the Wnt receptor *kringle containing transmembrane protein 1* (*kremen1*) to aid in hepatic development (Lu *et al.*, 2013). In relation to this, according to GO analysis, genes associated with early endoderm development include *fraser extracellular matrix complex subunit 1* (*fras1*), which is required for late forming portion of the first endodermal pouch and also *cadherin 1, type 1, E-cadherin* (*cdh1*) which is expressed in the liver during development (Siew *et al.*, 2006; Talbot *et al.*, 2012). However, despite *cdh1* being expressed in the liver, high levels of *cdh1* are also detected in non-neural ectoderm, prechordal plate and enveloping layer (Vannier *et al.*, 2013), suggesting that a number of endodermal-expressed genes function in other cell lineages. In addition to the above, GO terms including exocrine pancreas development and type β pancreatic cell differentiation were over-represented in *sox17E* and these terms were associated with genes such as *insulin* (*ins*) which is used to control blood sugar levels and *lipolysis stimulated*

lipoprotein receptor (lsr) which is expressed in the pancreatic buds at 34 h.p.f. and is later expressed in the intestine, liver, pharynx, otocyst and pronephros. *Lsr* is believed to be essential for endoderm formation since anti-sense knockdown of the gene results in a smaller liver, intestine and pancreas as well as a scattered β -cell phenotype (Dokmanovic-Chouinard *et al.*, 2008).

While GO terms corresponding to endoderm or endoderm organ formation were enriched, these terms were not as over-represented as other terms which were also associated with important endodermal genes. Indeed, GO analysis on genes upregulated in *sox17E* vs *sox17N* revealed that the most enriched terms were associated with genes involved in tissue development, developmental process, and animal organ development, including *hnf1ba*, *prdm1a*, *epcam*, *kremen1* and *lsr*, which are all important during endoderm formation or endoderm organogenesis, as mentioned previously. However, these terms were also associated with genes that do not play a role in the endoderm, including, *myosin*, *heavy polypeptide 2*, *fast muscle specific (myhz2)*, which is the most enriched transcript in *sox17E* and is essential for fast muscle organisation within the tail and *keratocan (kera)* which is essential for the maintenance of corneal transparency and structure in zebrafish (Yeh *et al.*, 2008; Nord *et al.*, 2014). In addition, terms including muscle structure development, heart development as well as blood vessel development were also over-represented in *sox17E* vs *sox17N*. Whilst these terms do not correspond to the endoderm, they were found associated with a number of genes that play a role in the endoderm or KV. For instance, the GO term muscle structure development was associated with *prdm1a* which is involved in endoderm formation and *cxcr4a* which required for endoderm migration and proliferation but also KV formation (Tseng *et al.*, 2011; Stückemann *et al.*, 2012; Liu *et al.*, 2019). Similarly, the GO term heart development was associated with genes including *tbxta*, *rock2b* and *vgl14l* which are important for KV formation and left/right asymmetry establishment, while *tbxta* also plays an essential role in mesendoderm formation (Amack, Wang and Yost, 2007; Wang *et al.*, 2011; Fillatre *et al.*, 2019). In addition, blood development GO term was associated with *cxcr4a*, *cxcl12b* and *kremen1*. Nevertheless, genes not associated with endoderm nor endoderm organ formation were also upregulated in *sox17E* vs *sox17N* and were found significantly enriched compared to many important endodermal genes. For instance, *myhz2*, the most enriched transcript in *sox17E*, is a myofiber-specific gene, while *ECRG4 augurin precursor b (ecrg4b)*, which is in top 30 most enriched genes, is expressed in the presumptive epidermis (Devakanmalai, Zumrut and Özbudak, 2013; Kimelman *et al.*, 2017). One possibility for this is that genes that are expressed in endodermal cells may also play roles in other cell identities and vice versa. But an alternative hypothesis is that *sox17E* may not be entirely composed of endoderm or KV transdifferentiated cells and may contain cells from the other cell populations:

sox17N or *sox17M*. One way this can be investigated is through the use of single-cell RNA-seq and investigating genes which are upregulated in the different cell populations.

All in all, while numerous *sox17E* upregulated transcripts play a role in the endoderm, some of these genes were also found to play important roles in other lineages. In addition to this, genes which do not play a role in the endoderm were also upregulated in *sox17E* and were found associated with the formation of other cell lineages including muscle and heart. Alternative experiments including single-cell RNA-seq will need to be performed to investigate the reason why a number of non-endodermal-expressed genes were upregulated in *sox17E* vs *sox17N*.

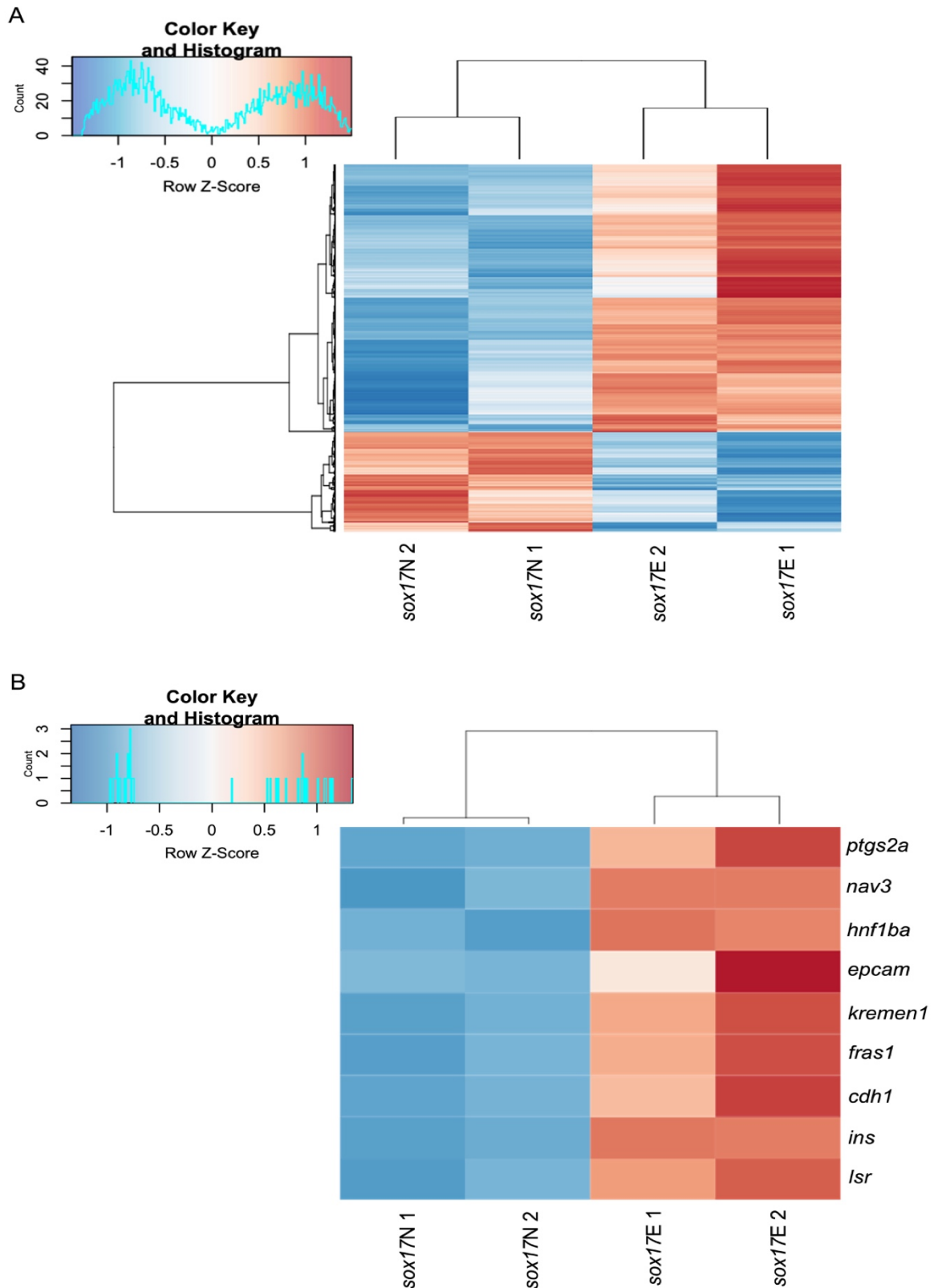


Figure 4-33: RNA-seq analysis for *sox17E* vs *sox17N*. (A) Heatmap displaying statistically significant genes (FDR < 0.05) which are either upregulated or downregulated in *sox17E*. (B) Heatmap displaying a selective number of genes based on GO analysis. 752 transcripts were upregulated in *sox17E* while 354 transcripts were upregulated in *sox17N*. Two biological replicates were generated for each sample.

4.12 Sox17E comprises both KV transdifferentiated cells and endoderm progenitors

So far, *sox17E* sorted cells from the triple transgenic line has helped identify putative CRMs proximal to genes that may play a role in endoderm organ formation. However, it is not yet known whether *sox17E* genes are concordant with known endoderm-expressed genes at early organogenesis stage. To test this, *sox17E* upregulated genes were compared to enriched endoderm genes at 6, 8, 10, 14, 18 and 24 h.p.f. as well DFCs at 8 h.p.f., obtained from published single-cell RNA-seq data (Wagner *et al.*, 2018). As demonstrated by GSEA, *sox17E* genes significantly correlated with endoderm-enriched genes at later stages of development, including 14,18 and 24 h.p.f. but also DFCs at 8 h.p.f. (FDR 5×10^{-4} , Pvalue 5×10^{-4}) (Figure 4-34). This means that upregulated transcripts in *sox17E* are typically expressed in the endoderm at later stages of development but also during early organogenesis. However, the results also support the idea that *sox17E* comprises DFC progeny, which further suggests the presence of KV transdifferentiated cells in the sample group.

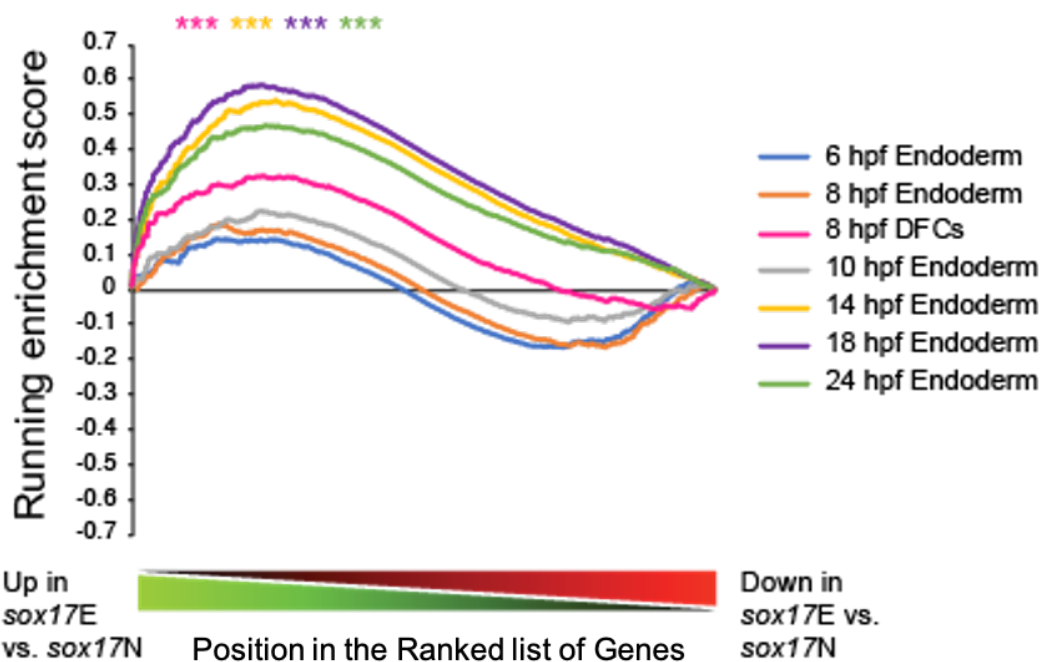


Figure 4-34: GSEA enrichment plot comparing upregulated transcripts from *sox17E* and *sox17N* to endoderm-enriched genes from different stages of development (6, 8,10,14,18 and 24 h.p.f) and DFCs at 8 h.p.f.. Enriched genes were obtained from Wagner *et al.* (2018). *** FDR 5×10^{-4} , Pvalue 5×10^{-4} .

4.13 Changes in the chromatin accessibility correlates with changes in gene expression in *sox17*⁺ endoderm

Our results so far suggest that endoderm-expressed genes are upregulated in *sox17E*. We next sought to investigate whether chromatin accessibility correlates with gene expression. To test this, we compared the gene lists obtained from GREAT, which were found proximal to DARs from *sox17N*, *sox17E* and *sox17M* DARs, to genes upregulated in *sox17E* and *sox17N* based on RNA-seq analyses. As evident by GSEA, *sox17E* transcripts were highly correlated with genes proximal to *sox17E* DARs (FDR $q = 2 \times 10^{-4}$, Pvalue 5×10^{-4}) (Figure 4-35). This means that changes in the chromatin accessibility correlates with changes in gene expression. Given that *sox17E* transcripts are concordant with endoderm and DFC-enriched genes, this further suggests that *sox17E* comprises both endoderm and KV-transdifferentiated cells.

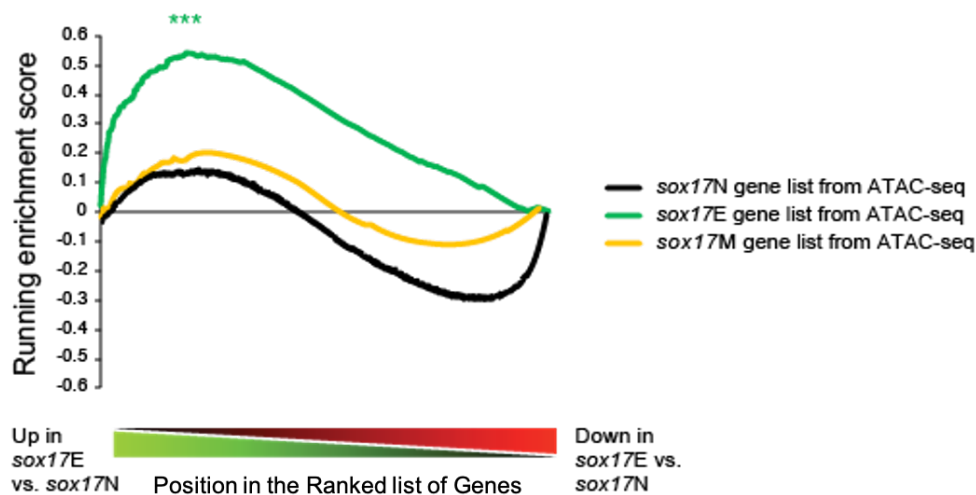


Figure 4-35: GSEA enrichment plot comparing upregulated transcripts from *sox17E* and *sox17N* sample groups to genes proximal to *sox17N*, *sox17E* and *sox17M* DARs obtained from GREAT analysis. *** FDR $q = 2 \times 10^{-4}$, Pvalue 5×10^{-4} .

An example of a gene which is upregulated in *sox17E* is *epcam*, which is important during hepatogenesis. ATAC-seq analyses revealed a number of putative enhancers proximal to *epcam* while RNA-seq analyses revealed that *epcam* was highly upregulated in *sox17E* compared to *sox17N* (Figure 4-36).

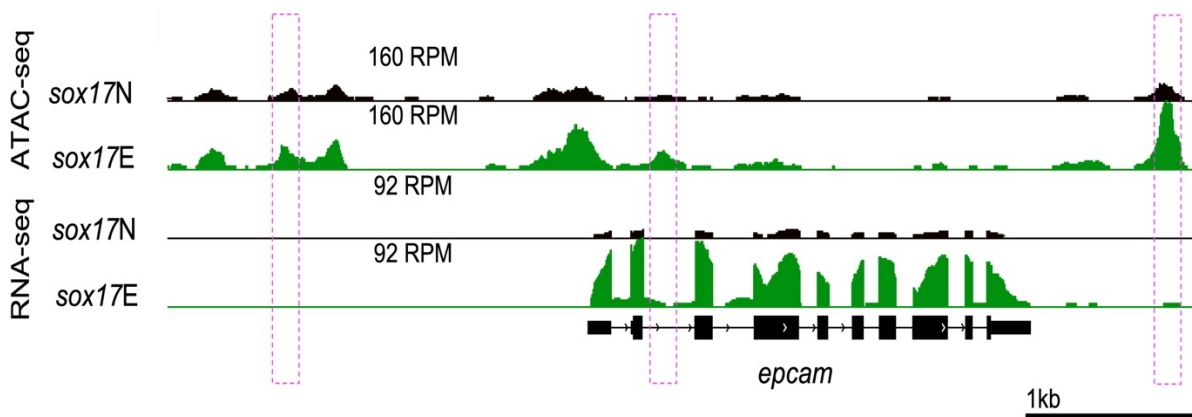


Figure 4-36: An example of a gene with upregulated expression in *sox17E* is *epcam*. Regions of higher accessibility were also found proximal to *epcam* which are likely putative enhancers of the gene. Reads were scaled to adjust for library size. Putative enhancers are indicated by pink dashed lines. RPM = reads per million.

To conclude, the results demonstrate that *sox17+* endoderm cannot be separated from contaminating mesoderm based on *GFP* expression at early organogenesis stage, particularly since markers of the endoderm including *foxa3* were highly upregulated in GFP^{low} relative to GFP^{high} cells, while markers of the mesoderm population, including *sox17*, exhibit upregulated expression in GFP^{high} relative to GFP^{low} cells. As opposed to this, our results demonstrate that *sox17+* endoderm was successfully isolated from *sox17+* mesoderm by using the triple transgenic line. We also show that KV cells make contributions to the posterior notochord which are likely sorted by FACS when using the triple transgenic line. Moreover, a number of putative enhancers were enriched for binding sites for TFs that play important roles in pancreas formation. For example *Ptf1a* and *Pdx1* are both crucial for pancreatic development and thus, mutagenesis studies may be used to understand their role in the context of pancreas formation. However, endoderm-specific CRMs also reveal binding motifs for TFs with unknown roles in the zebrafish endoderm, including Six TFs that play important roles in thymus and parathyroid formation in other species. Finally, we identify a number of different CRMs which likely function in the different *sox17+* populations. These results confirm the use of the triple transgenic line in studying the regulatory mechanisms in the endoderm during early organ formation.

4.14 Discussion

Previous studies have used *Tg(sox17:EGFP)* line, which expresses *EGFP* in *sox17+* lineages, to study the endoderm during development. However, beyond gastrulation (~5-10 h.p.f.), *sox17* is no longer expressed in the endoderm but in haemopoietic and endothelial lineages as well as KV derivatives (Chung *et al.*, 2011). This was further validated by *sox32*-KD which showed a lack of GFP expression in the endoderm but highlighted GFP expression in the ICM, which give rise to blood cells (Figure 4-5) as well as endothelial cells. Notably, GFP+ cells in the posterior notochord were also absent in *sox32*-KD. Fate-mapping experiments revealed that DFC progeny contribute to the notochord or muscle fates in the tail (Melby, Warga and Kimmel, 1996), suggesting that these cells are KV derivatives. To investigate this, RD was injected into mid-blastula staged *Tg(sox17:EGFP)* embryos and showed that by 35 h.p.f. the posterior notochord was double labelled with GFP and RD (Figure 4-8). This suggests that GFP+ posterior notochord cells are derivatives of the KV, in accordance with published data (Melby, Warga and Kimmel, 1996) and are therefore dependent on *sox32* expression in zebrafish.

Since the study of the endoderm beyond gastrulation is enabled by the transgene from *Tg(sox17:EGFP)* which at this stage, also marks *sox17+* mesoderm, this raises a problem; how can we study the endoderm when GFP marks both endoderm and mesoderm progenitors at later stages of development?

One possibility is that *sox17+* endoderm can be isolated from contaminating mesoderm by sorting for GFP expression. This was speculated based on the following; *sox17* expression pattern (White *et al.*, 2017), anti-GFP immunostaining (Nelson *et al.*, 2017) and *Tg(sox17:EGFP)* fluorescence images (Figure 4-5). To test this hypothesis, a number of endoderm and mesoderm markers were used for qRT-PCR. While most endoderm markers were enriched in GFP^{high} population, *foxa3*, a posterior foregut marker, was instead enriched in the GFP^{low} population. Also, *sox17* which marks the haemopoietic and endothelial lineages, was highly expressed in GFP^{high} and not in GFP^{low} cells while no difference was observed in the expression of blood cell marker *gata1a* in both cell populations. Thus, our results suggest that *sox17+* endoderm cannot be isolated from *sox17+* mesoderm based solely on GFP expression. Additionally, this technique has further caveats in that only a handful of markers can be used for qRT-PCR at a time and that these markers are based on *in situ* hybridisation results. This makes it difficult to interpret whether genes are expressed more extensively in the endoderm compared to other cell populations.

Alternatively, fluorescent proteins can be used to mark and therefore isolate the different cell populations. To enable the study of the endoderm at the onset of organogenesis, we generated a triple transgenic zebrafish line that expresses distinct fluorescent proteins in *sox17+* lineages (GFP), endothelial cells (mCherry) and blood lineages (dsRed), and therefore one way to isolate GFP+ endoderm from contaminating *sox17+* mesoderm progenitors is by sorting for GFP only populations.

We first validated the use of the triple transgenic line by selecting for markers of the endoderm and mesoderm for qRT-PR analyses. By quantifying the relative expression of these markers across *sox17N*, *sox17E* and *sox17M*, we demonstrated that most endodermal markers were highly expressed in the *sox17E* (Figure 4-19) while blood and endothelial markers were upregulated in *sox17M* cell populations (Figure 4-20). Thus, these results validate the use of the line in successfully separating *sox17+* endoderm from contaminating mesoderm.

The use of the triple transgenic line was further validated by looking for enriched anatomical terms in each cell population. As observed by GREAT, *sox17E* DARs were proximal to genes associated with the endoderm compared to *sox17N* and *sox17M* DARs which are more associated with CNS and blood cell lineages, respectively. However, genes proximal to *sox17E* DARs were also found to be associated with non-endodermal anatomical terms including mesoderm, immature eye, and renal system (Figure 7-4) although these terms were not as over-represented as the endodermal terms. This therefore likely suggests that many endodermal-expressed genes are also present in other cell lineages including mesoderm. Based on this analysis, computational methods can be utilised to not only identify DARs that are likely shared between the different lineages, but also identify genes that are proximal to these CRMs. This would further confirm whether CRMs within the endoderm are also shared by other lineages including mesoderm and CNS.

As demonstrated by GREAT, no endoderm terms were enriched in *sox17M*, however terms including the digestive system and pharynx were over-represented in *sox17N*. This is likely because endoderm-expressed genes are not unique to endoderm and play a role in other cell lineages. For instance, the pan-endodermal marker, *foxa2* also plays a role in the CNS (Odenthal and Nüsslein-Volhard, 1998). Consistent with this, both *sox17E* and *sox17N* DARs were highly accessible in *sox17E* and *sox17N* cell populations but not *sox17M*, suggesting that endoderm-specific CRMs are also shared by non-*sox17+* cell populations (Figure 4-28- Figure 4-30).

To find potentially important endodermal DARs, Motif analysis was performed on DARs derived from *sox17E* sample group. This analysis is crucial for identifying potentially important TF binding motifs that may a role in the endoderm at the onset of organogenesis. Motif analysis was performed using HOMER and revealed a number of TF binding motifs (Figure 4-32). The most enriched motifs belong to members of the AP1 factors which are expressed in many organs and are believed to play important roles in a variety of processes, for instance, Jun is believed to be required for liver formation and regeneration in zebrafish and mouse, while *Fosl2* is required for cardiac myocardial differentiation in zebrafish (Passegué *et al.*, 2002; Sadler *et al.*, 2007; Jahangiri *et al.*, 2016). Other enriched motifs belong to Tead proteins 1, 3 and 4. In zebrafish, Tead/Yap1 interaction is critical for endoderm maintenance and formation with evidence that perturbation to Tead/Yap1 signalling results in defects in anterior endoderm formation due to increased apoptosis of endoderm progenitors (Fukui *et al.*, 2014). In humans, interactions between TEAD1 and YAP1 are required to activate expression of key mediators of pancreatic development through binding to pancreatic progenitor CRMs (Cebola *et al.*, 2015). Consistent with this, mutation in Tead1 binding sites in zebrafish results in marked reduction of pancreatic progenitors, suggesting that Tead1 is a key regulator of pancreas development (Cebola *et al.*, 2015). In agreement with this, in mice, perturbations to Tead1 and 2 result in the down-regulation of *Sox17+* cells in the posterior definitive endoderm during development (Sawada *et al.*, 2008). Though, Tead3 and 4 are also over-represented in (Figure 4-32), Tead3 knockout mice appear normal (Sawada *et al.*, 2008) while Tead4 is dispensable for establishment of pluripotent ESCs (Nishioka *et al.*, 2008) though it is also present in lungs in new-born and adult mice (Hsu *et al.*, 1996). This likely suggests that Tead proteins play redundant activities during development. In zebrafish, only four Tead proteins have been reported; Tead1a, Tead1b, Tead3a, Tead3b (Fillatre *et al.*, 2019). Although there are currently no information on their individual roles in the endoderm in zebrafish, all four Tead members appear to be expressed in the pharyngeal endoderm in accordance to published single-cell RNA-seq (Wagner *et al.*, 2018). Interestingly, Tead2 motif was also over-represented in *sox17E* DARs, however, there are currently no data on the role of Tead2 in zebrafish and likely represents the presence of binding sites for other Tead factors. Given that Tead members contain a highly conserved DNA-binding domain and bind the same consensus M-CAT-like DNA sequence (5'-CATTCCA/T-3') (Anbanandam *et al.*, 2006), it is therefore likely that all Tead proteins play redundant roles in the endoderm during development.

Interestingly, Tead1 and 3 are expressed in the DFCs and KV and their loss results in a reduction of KV size, motility and causes laterality defects (Fillatre *et al.*, 2019). Though it is likely that KV transdifferentiated cells within the posterior notochord are also sorted as part of

the *sox17E* population, it is unlikely that Tead proteins play a role in the KV or laterality establishment at 28 h.p.f.

Other enriched motifs in *sox17E* DARs include Sine oculis homeobox (SIX) family of transcription factors. In mice, *Six1* is critical for patterning of the third endodermal pouch into thymus and parathyroid since both organs fail to develop in *Six1* mutant mice. In addition to this, *Six1* is co-expressed with *Six4* in the pharyngeal endoderm and loss of both proteins results in a complete loss of thymus and parathyroid markers (Zou *et al.*, 2006). In humans, SIX2 is important for the maturation of β -cells into glucose-secreting cells in the pancreas and loss of SIX2 results in a dramatic reduction of glucose-stimulated insulin secretion (Velazco-Cruz *et al.*, 2020). In zebrafish, orthologues of mammalian *Six1* and *Six4* are required for cell motility in the migrating muscle precursors and for partitioning two muscles in the fin bud (Talbot *et al.*, 2019). However, there are currently no information on the roles of *Six1*, *Six2* and *Six4* in the zebrafish endoderm though zebrafish orthologues of mammalian *Six1,2* and *4* are expressed in the pharyngeal endoderm at 24 h.p.f, based on single-cell RNA-seq data (Wagner *et al.*, 2018). Thus, our data reveals a number of putative CRMs that may be required for thymus and parathyroid organogenesis in zebrafish.

Sox17E DARs were also enriched for binding motifs of Gata4 and Gata6. In mice, Gata4 is expressed in visceral endoderm and is required for normal morphogenesis of the primitive gut tube during embryonic folding (Molkentin *et al.*, 1997). The function of Gata4 seems to be well conserved in vertebrates, since expression of organ-specific markers of differentiated liver, intestine and pancreas are either absent or reduced in Gata4 zebrafish morphants (Holtzinger and Evans, 2005). In relation to this, Gata6 plays an important role in lung development and is later required for the differentiation of the lung epithelium in mice (Koutsourakis *et al.*, 2001; Liu, Morrissey and Whitsett, 2002; Yang *et al.*, 2002). Gata6 is also expressed in the developing pancreas and is believed to aid in pancreas specification and differentiation (Decker *et al.*, 2006). In zebrafish, Gata6 works either in parallel or downstream of Nodal signalling to aid in the regulation of endoderm transcription factors, including *sox32*, *tbx16*, *foxa2* and *foxa3* which have established roles in endoderm development (Tseng *et al.*, 2011). However, little is known about the role of Gata6 at later stages of development, although it was found to be expressed in the pancreas primordium at 24 h.p.f. based on single-cell RNA-seq data (Wagner *et al.*, 2018). Our data reveals a number of CRMs with binding motifs for Gata4 and 6 which may be important for the morphogenesis of a number of different endoderm-specific organs during development.

Additional over-represented motifs in *sox17E* DARs include TF binding sites for Pdx1 and Ptf1a which are both critical for pancreas development (Decker *et al.*, 2006). In humans, mutations in PTF1A results in pancreatic agenesis (Sellick *et al.*, 2004), a congenital disease where the pancreas fails to develop during embryonic growth. In mice, *Ptf1a* RNA is detected as early as 9.5 days post coitum and loss of *Ptf1a* results in a failure of exocrine pancreas or acinar development. This role seems to be highly conserved in vertebrates since, *ptf1a* zebrafish morphants exhibit agenesis of the exocrine pancreas (Jiang *et al.*, 2008).

However, the vast majority of pancreatic cell type require the function of *pdx1*, which is one of the earliest pancreatic genes. In mice, knockdown of *Pdx1* leads to failure of pancreas tissue development and as a result, mutants lack both mature exocrine pancreas and endocrine cells (Jonsson *et al.*, 1994). In humans, PDX1 mutations results in endocrine pancreas dysfunction and hyperglycaemia (Babu, Deering and Mirmira, 2007) and thus, PDX1 plays an essential role in pancreas development and β -cell function and survival (Fujimoto *et al.*, 2009). In accordance with this, null mutations of *pdx1* in zebrafish results in perturbation of β cell number, insulin levels and acinar differentiation, with mutants also exhibiting disrupted glucose homeostasis (Kimmel *et al.*, 2015). Notably, Pdx1 also regulates expression of other transcription factors that are required for early pancreatic fate decision, including Ptf1a, with evidence that its expression was completely absent in Pdx1 knockout mice (Marty-Santos and Cleaver, 2016). Other evidence also suggests that Ptf1a regulates expression of Pdx1 in acinar and endocrine progenitor cells by binding to a conserved region of the Pdx1 promoter while also autoregulating its own expression (Wiebe *et al.*, 2007; Masui *et al.*, 2008). Thus, our data suggests the presence of CRMs that may be important for pancreas maintenance and maturation during development.

Finally, motifs of Forkhead factors were also over-represented in *sox17E* DARs. Important roles of Forkhead TFs have already been discussed in 3.5.4. In mice, multiple *Pdx1* enhancers are co-occupied by both Foxa1 and Foxa2 TFs and a loss of both factors in the pancreas primordium results in pancreatic agenesis as well as loss of *Pdx1* expression and failure in the expansion of the pancreatic primordium (Gao *et al.*, 2008). Thus, in addition to Pdx1 and Ptf1a, Forkhead factors are also likely required to bind endodermal-specific CRMs to drive pancreatic development.

All in all, our data suggests the presence of a number of different TF binding motifs that are likely required for the morphogenesis of a number of different endoderm-specific organs in zebrafish. However, the data suggests the presence of motifs that are likely bound by factors that play essential roles in endoderm organ formation in other species, but little information is

known about their role in the endoderm in zebrafish, which includes all Six factors. This therefore suggests a potential novel function of these TFs in endoderm organ formation in zebrafish during development. The roles of these factors can be further investigated through various methods including transgenic reporter assays.

To further confirm the upregulation of endodermal-expressed genes within *sox17E* cell populations compared to other cell populations, RNA-seq was used. Our RNA-seq analysis revealed a number of endodermal-genes which either play a role in the endoderm in zebrafish or other species (Figure 4-33B), including *hnf1ba*, which is required for proper β -cell numbers and pancreas specification and *nav3*, which is required for liver formation in zebrafish (Klein *et al.*, 2011; Lancman *et al.*, 2013). In relation to this, *epcam* and *fras1* transcripts were also upregulated in *sox17E*, which are required for hepatic development and first endodermal pouch formation, respectively (Talbot *et al.*, 2012; Lu *et al.*, 2013). Moreover, GO analysis revealed that a large number of genes upregulated in *sox17E* vs *sox17N* were associated with endodermal organ formation including pancreas and liver development. Some of these can be viewed in Figure 4-33B. However, according to GO analysis using DAVID, many of the genes upregulated in *sox17E* were not associated with endoderm. Indeed, muscle, blood vessel as well as heart development were highly over-represented compared to endodermal terms. Despite this, these terms were associated with numerous endodermal genes, including *epcam* and *cxcr4a*. However, in addition to this, these terms were also associated with genes that do not play a role in the endoderm, including *myhz2*, which is a myofiber-specific marker (Devakanmalai, Zumrut and Özbudak, 2013). While the results suggest that endodermal genes likely play a role in other cell lineages, it may also suggest a slight contamination of *sox17E* from *sox17N* and *sox17M* cells. An alternative method to use to ensure that there is no contamination from the other sorted populations is to use single-cell RNA-seq and identify transcripts that are highly upregulated in *sox17E* cell population. This will help determine whether genes including *myhz2* play novel roles in the endoderm during development.

While a number of endodermal genes were upregulated in *sox17E* vs *sox17N*, the heatmap shown in Figure 4-33B also suggests that there is slight discrepancies between the two *sox17E* biological replicates in the expression of the various upregulated genes. Indeed, *epcam* appears to be strongly expressed in *sox17E* 2 vs *sox17E* 1. However, despite this difference, the largest variability is attributed to differences between *sox17E* and *sox17N* and likely suggests that these represent different cell identities.

Comparison of upregulated genes from *sox17E* to genes enriched in the endoderm at different stages of development from single-cell RNA-seq data (Wagner *et al.*, 2018) revealed that

sox17E significantly correlates with endoderm genes that are expressed at later stages of development (14,18 and 24 h.p.f.), and thus is in line with the developmental stage of the embryos used in this study. GSEA analysis also revealed that *sox17E* significantly correlates with genes enriched in the DFCs at 8 h.p.f. (Wagner *et al.*, 2018). This likely correlates with the idea that KV transdifferentiated cells are also sorted into *sox17E* population when using the triple transgenic line.

In conclusion, our results reveal that *sox17E* from the triple transgenic line can be used to study the endoderm during early organogenesis, and that the genes expressed at this stage are in line with genes enriched at later stages of endoderm development according to published single-cell RNA-seq. Our results also reveal a number of putative CRMs with binding motifs of TFs that play crucial roles in the endoderm in other species but their roles are not yet established in zebrafish. Thus, transgenic reporter assays can be used to explore the importance of these motifs in relation to endoderm development. Finally, density array heatmaps revealed that CRMs are likely shared between endoderm and CNS but not mesoderm, and this was also in line with GREAT analysis which revealed that many genes are shared between the two lineages. Using computational analysis to identify CRMs that are shared by the two lineages will help reveal which endodermal expressed genes play roles in other lineages.

Chapter 5 Investigating the chromatin landscape in the endoderm during development

5.1 Chromatin organisation accommodates gene expression changes

Gene regulatory networks that involve the interaction between transcription factors and *cis*-regulatory modules (CRMs), which include enhancer and promoter elements, underlie all developmental processes and cell-lineage identities (Wassef and Margueron, 2017). Enhancers are regulatory DNA sequences that when bound by transcription factors, enhance the expression of an associated gene at the promoter. Transcription factors can bind to enhancers located upstream, downstream, internal to their target gene or even on different chromosomes to the target gene, resulting in enhancement of expression of the related gene (Nelson and Wardle, 2013; Erokhin *et al.*, 2015; Suryamohan and Halfon, 2015). Enhancers are known to be *cis*-acting and can activate transcription irrespective of their orientation and location and thus, enhancer sequences can be located thousands of bases away from the transcription site of the gene being regulated (Gorkin, Leung and Ren, 2014; Ibragimov, Bylino and Shidlovskii, 2020). They are believed to be influenced by high-order chromatin organisation known as chromatin looping, which brings enhancers in close proximity to the target gene promoter. Interestingly, high order chromatin organisation plays an important role in the regulation of gene expression, since accessible regulatory regions coincide with active gene expression while heterochromatic regions inactivate transcription. Although numerous developmental CRMs have already been discovered through the use of various animal models, a comprehensive list of developmentally important CRMs is still lacking. Thus far, a number of human CRMs have already been identified across the different stages of pancreatic differentiation; embryonic stem cells (ESCs), definitive endoderm, primitive gut tube, posterior foregut and pancreatic endoderm (Wang *et al.*, 2015). In this study, a total of 119, 795 enhancers were identified across the five stages of pancreatic differentiation, where the majority were marked by histone modification H3K4me1 (inactive enhancer state), with only a small fraction marked by H3K27ac (active enhancer state). In addition, the enhancers that were marked by H3K27ac were found to be associated with master regulators of pancreatic induction including *PDX1*, *SOX9*, *PTF1A* and *NKX6.1*. Furthermore, poised enhancers, which have the ability to respond to inductive signals during gut tube formation, were found enriched for binding motifs of transcription factors FOXA, GATA, HNF4A and HNF1 which regulate the development of endodermal organs including pancreas, lung, and liver. Thus, acquisition of a

poised state at enhancers play a critical role in terminal differentiation of progenitor cells into functional cell types.

In relation to this, ATAC-seq studies on mice during early mouse development have also helped identify putative CRMs within the developing and neonatal gut (Banerjee *et al.*, 2018). Similar to humans, the mouse early midgut harbours the ability to activate foregut genes and by extension, the differentiation of the oesophagus and forestomach and that these enhancers were marked by H3k27ac. However, once intestinal genes were expressed, open chromatin and H3k27ac at foregut enhancers were abrogated and this is believed to be controlled by intestinal-specific transcription factor CDX2. Thus, thousands of foregut enhancers are found to be active within the early intestine and are later decommissioned upon activation of midgut-specific enhancers.

However, a list of putative CRMs is still missing for other endodermal organ systems. Knowing that perturbation to non-coding regions of the DNA is associated with human diseases, including diabetes and pancreatic agenesis, identification of important CRMs may help shed light on potentially novel therapeutic targets and may lead to new treatment opportunities including regenerative approaches for diseases associated with the endoderm.

Many developmental processes are conserved between vertebrate species, therefore fish studies help provide insight into human biology (Zon, 1999; Ton *et al.*, 2000). For instance, studies on zebrafish zygotic genome activation (ZGA), a process that enables the zygotic genome to replace the maternal factors that initiated development, has revealed that this transcriptional change is highly conserved in vertebrates and is heavily reliant on chromatin remodelling which converts the genome from a transcriptionally quiescent state to an activated state (Schulz and Harrison, 2019). Thus, due to conservation of developmental processes, studies in zebrafish will likely give great insight into lineage-specific enhancers elements in humans.

5.2 Objectives of project

Extensive research has already demonstrated that changes in the chromatin landscape is accompanied by gene expression changes. However, there is still a lack of information on developmentally important enhancers that are critical in driving the morphogenesis of all endodermal organs from endoderm progenitor cells.

Thus the main objectives of this study are as follows:

- 1) To explore whether the chromatin landscape changes in the endoderm between specification and early organogenesis. This will be investigated by comparing differential accessible regions (DARs) from endoderm cells at 6 h.p.f. and 28 h.p.f..
- 2) To identify potentially important CRMs at the two different stages of endoderm development. This will be investigated by first finding genes proximal to DARs from 6 h.p.f. and 28 h.p.f. endoderm cells and by performing GO analysis on these lists of genes.
- 3) To find transcription factor binding motifs within accessible chromatin regions from endoderm cells at 6 and 28 h.p.f.. This will be investigated by performing motif analysis on DARs attained from endoderm cells at 6 and 28 h.p.f..

Investigating these objectives will provide a better understanding of how the chromatin landscape changes at the two different stages of endoderm development: specification and onset of organogenesis. However, this study may also help identify potentially important endoderm-specific CRMs as well as transcription factor binding motifs that are required to drive expression of key lineage-specific genes at the two different endoderm stages.

5.3 Sox32 OE and sox17E have different accessibility profiles

In order to investigate whether the chromatin landscape changes during endoderm development, accessible regions identified in the endoderm at 6 h.p.f. were compared to accessible regions identified in the endoderm at 28 h.p.f.. Indeed, called peaks in *sox32* OE (6 h.p.f.) were compared to peaks identified in *sox17E* cell populations, i.e.: *sox17+* endoderm sorted cells from the triple transgenic line. To compare the two datasets, Diffbind was used which first clusters samples based on MACS2 scores. The more similar the MACS2 scores are, the stronger the correlation between the samples. Based on hierarchical clustering, the peak-sets within replicate samples cluster together which means that they have similar MACS2 scores while the scores of the peak-sets are different between *sox32* OE and *sox17E* sample groups and as a result, the samples do not cluster together (Figure 5-1).

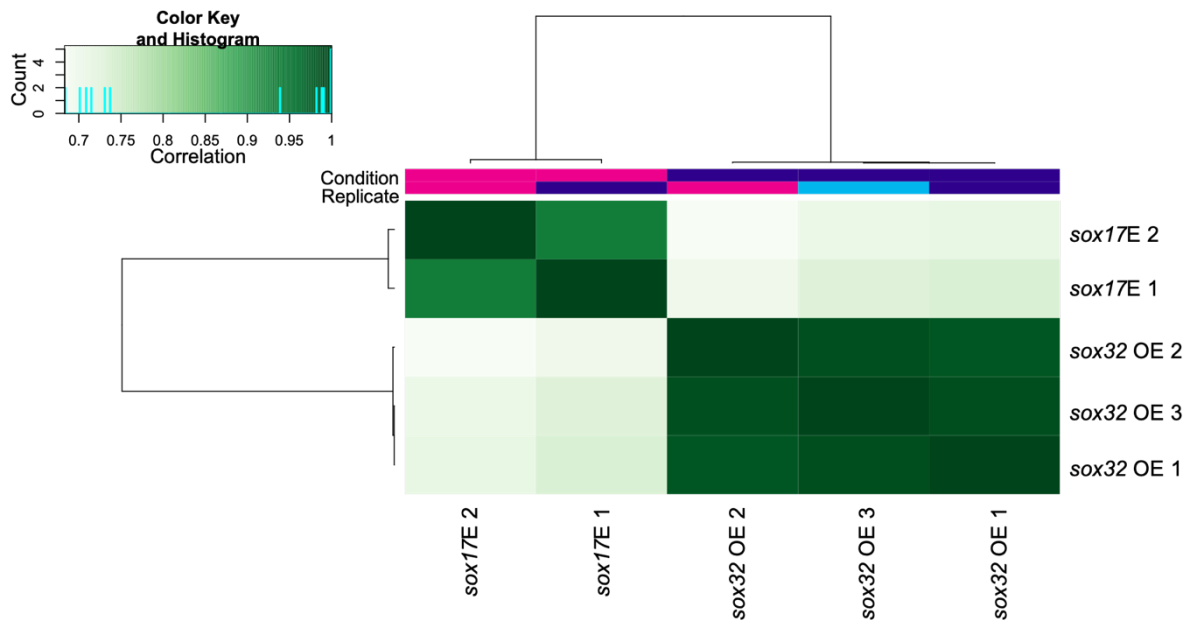


Figure 5-1: Pearson Correlation heatmap based on MACS2 scores. Dark green represents high similarity while light green represents low similarity. According to hierarchical clustering, biological replicates from 6 and 28 h.p.f. data set cluster together and thus the peak-sets from these samples have similar MACS2 scores, while MACS2 scores of the peak-sets are different between the two sample groups.

Diffbind then subsequently draws another heatmap based on read counts that map to each accessible genomic regions within a sample. For this analysis, Diffbind uses the peaksets that were obtained earlier in Figure 5-1 and calculates the number of reads that map each peakset within each sample group. The number of consensus regions across all samples is 68,764.

As shown below, the number of reads that map to each genomic interval appears to be similar within replicate samples of *sox32* OE and *sox17E* but different between sample groups (Figure 5-2).

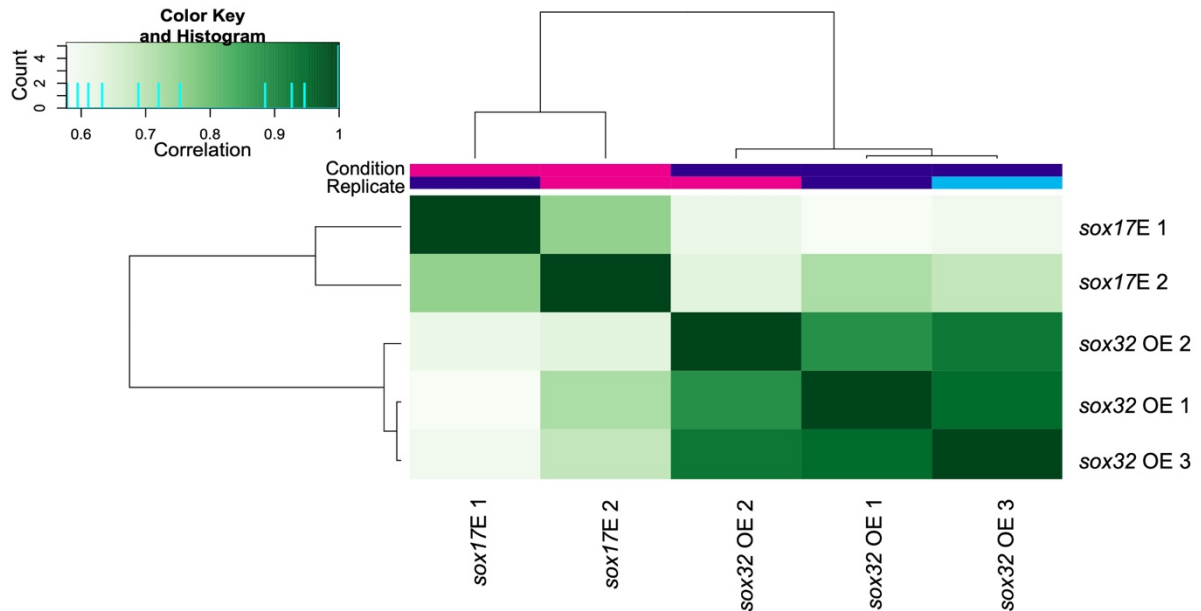


Figure 5-2: Pearson correlation heatmap based on read counts. The number of read counts that map to each genomic interval is similar between biological replicates but different between *sox32* OE and *sox17E* sample groups. Dark green represents high similarity while low similarity is represented by light green.

Lastly Diffbind was used to look for differential accessible regions (DARs) between *sox32* OE and *sox17E* sample groups. This was achieved by first normalising the data using DESeq2 and drawing a differential accessible heatmap to compare DARs from the different sample groups. This analysis revealed a total of 41,761 DARs between the two sample groups, where 34,942 regions exhibit a greater accessibility in *sox32* OE (Figure 5-3). Some of these DARs can be viewed in Table 7.22.

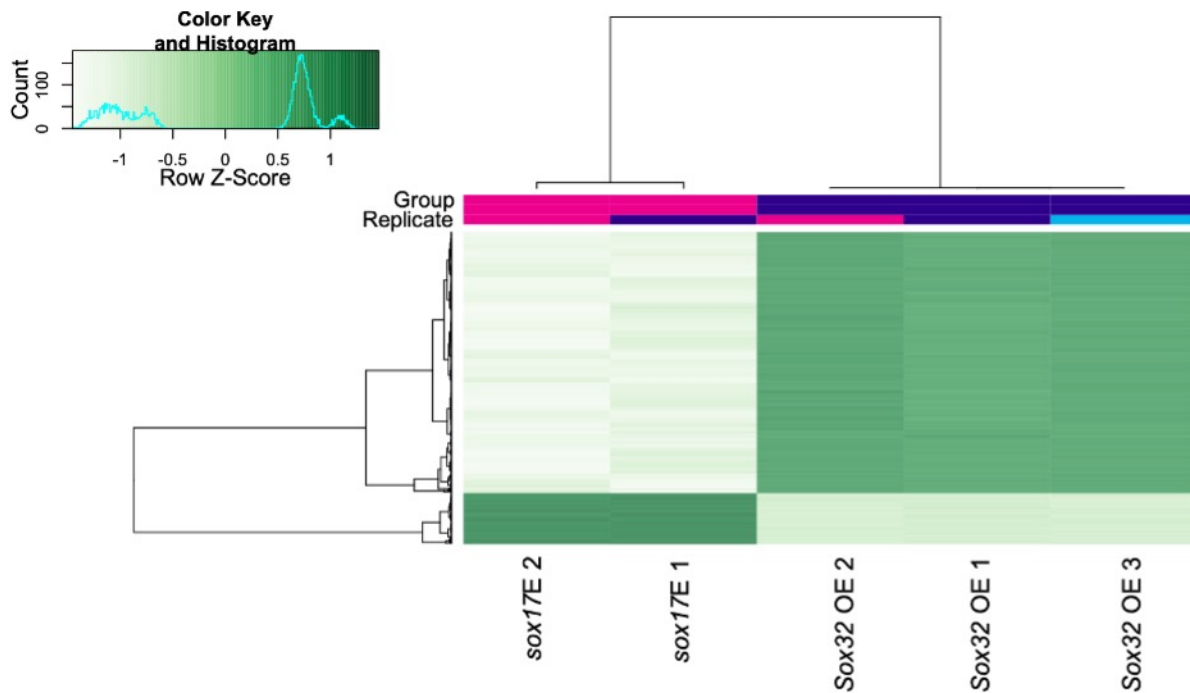


Figure 5-3: A differential accessible heatmap between *sox32* OE and *sox17E*. FDR cut-off of 5×10^{-4} was used to reduce list of accessible regions to a more manageable number for downstream analysis. The Pvalue cut-off was 1.8×10^{-4} . Heatmap was clustered on samples (columns) and accessible regions (rows), and rows were scaled on Z-score. Regions of higher accessibility are shown in dark green, and light green regions represent low accessibility.

The major differences in accessibility profile between 6 and 28 h.p.f likely implies that the chromatin landscape changes in the endoderm during development, particularly between specification and early organogenesis. This also means that different CRMs are adopted to control gene expression at various stages of endoderm development.

An example of a gene which demonstrates this change in accessibility between *sox32* OE and *sox17E* is *foxa3*. *Foxa3* is expressed in the endoderm throughout development including shield (6 h.p.f.) and early organogenesis stage (Dal-Pra, Thisse and Thisse, 2011). However, *foxa3* is not a specific endodermal marker since it is also expressed in the hatching gland and posterior notochord and therefore also expressed in the *sox17N* population as shown below. By visualising *foxa3* locus at 6 and 28 h.p.f., it is likely that different CRMs are required to control expression of the gene at different stages of endoderm development (Figure 5-4).

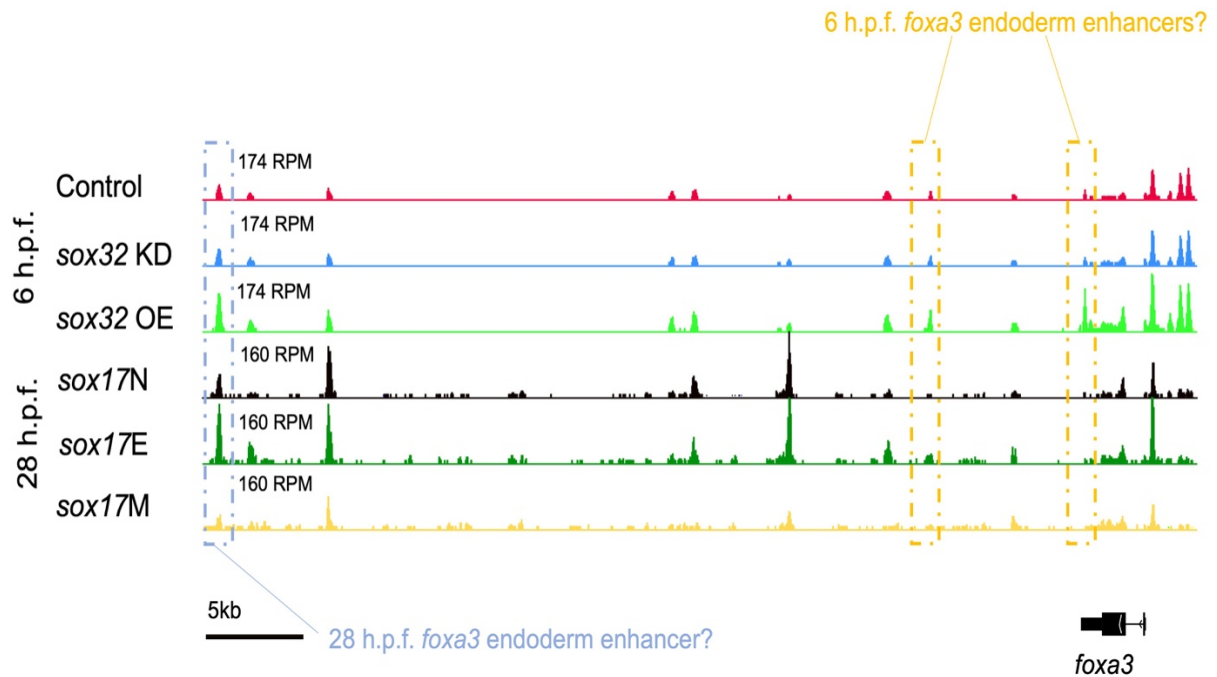


Figure 5-4: ATAC-seq read densities at the *foxa3* locus using Integrated Genome Viewer (IGV) showing different putative enhancers at 6 and 28 h.p.f.. Read numbers were scaled to adjust for library size. Blue and orange box surrounds putative enhancers at 28 and 6 h.p.f., respectively. Putative enhancers were also accessible in *sox17N* since *foxa3* is expressed in the hatching gland and posterior notochord. Two biological replicates were used for 28 h.p.f. samples and three biological replicates were used for 6 h.p.f. samples. RPM = reads per million.

5.3.1 *Sox32* OE and *sox17E* DARs are proximal to genes associated with the endoderm at different stages of development

The results thus far demonstrate that the accessibility profile differs between *sox32* OE and *sox17E*. Next we sought to explore the relation between *sox32* OE and *sox17E* DARs and adjacent genes by using GREAT, which was used to look at over-represented anatomical terms associated with both sample groups. Both *sox32* OE and *sox17E* DARs were proximal to genes associated with the endoderm but were also expressed in other cell identities (Figure 5-5). Some of these genes can be viewed in Table 7.23 and Table 7.24. To begin, for *sox32* OE, terms such as Margin and YSL were over-represented in the sample and this is likely because *sox32* expression appears in the YSL before it is later expressed in the dorsal marginal blastomeres from which endoderm progenitors arise (Kikuchi *et al.*, 2001). *Sox32* OE DARs were also proximal to genes associated with Forerunner cell group or DFCs which also express *sox32* during gastrulation (Alexander and Stainier, 1999). However, terms associated with epiblast, mesoderm and hypoblast were also associated with *sox32* OE. This is likely because, based on Gene Ontology (GO) analysis on the GREAT gene list using PANTHER (Thomas, Campbell, *et al.*, 2003; Thomas, Kejariwal, *et al.*, 2003), genes required during endoderm specification, including *bmp4*, *tbxta* and *bmp2b*, were also associated with

epiblast and brain. In relation to this, GO analysis revealed that genes expressed in the endoderm, including *osr1*, *yap1*, *fscn1a*, *gata5*, *dusp1* and *fgf8a* were associated with hypoblast and mesoderm. Notably, mesoderm genes including *tbxta*, and *tbx16* were also associated with KV formation, where KV is the progeny of DFCs. Thus, *sox32* OE DARs were proximal to genes that are required during endoderm and DFC specification but which also function in other cell identities.

In relation to this, *sox17E* DARs were also proximal to genes associated with the endoderm but at later stages of development (Figure 5-5). For instance anatomical terms for digestive system, pharynx, respiratory system, and gill which represent endoderm organ and organ systems, were all over-represented in *sox17E*. However, *sox17E* was also associated with terms including CNS and mesoderm. Based on GO analysis using PANTHER, genes associated with CNS including *jag1b*, *fgfr2* and *rargb* are also crucial for liver and pancreas development. Moreover, genes corresponding to CNS are also required for left/right asymmetry established and these include *chd*, *fgfr2* and *eomesa*, where *eomesa* has been found to induce ectopic DFC formation when overexpressed (Bjornson *et al.*, 2005). Also, many genes expressed in mesoderm lineages are crucial for left/right asymmetry establishment and liver development, according to GO analysis. Thus, *sox17E* DARs were proximal to genes associated with the endoderm but were also proximal to genes that are shared by both the KV and mesoderm.

Our results therefore suggest that genes proximal to *sox32* OE DARs are associated with early endoderm and DFC specification, while genes proximal to *sox17E* DARs are associated with early endoderm organogenesis and left/right asymmetry establishment.

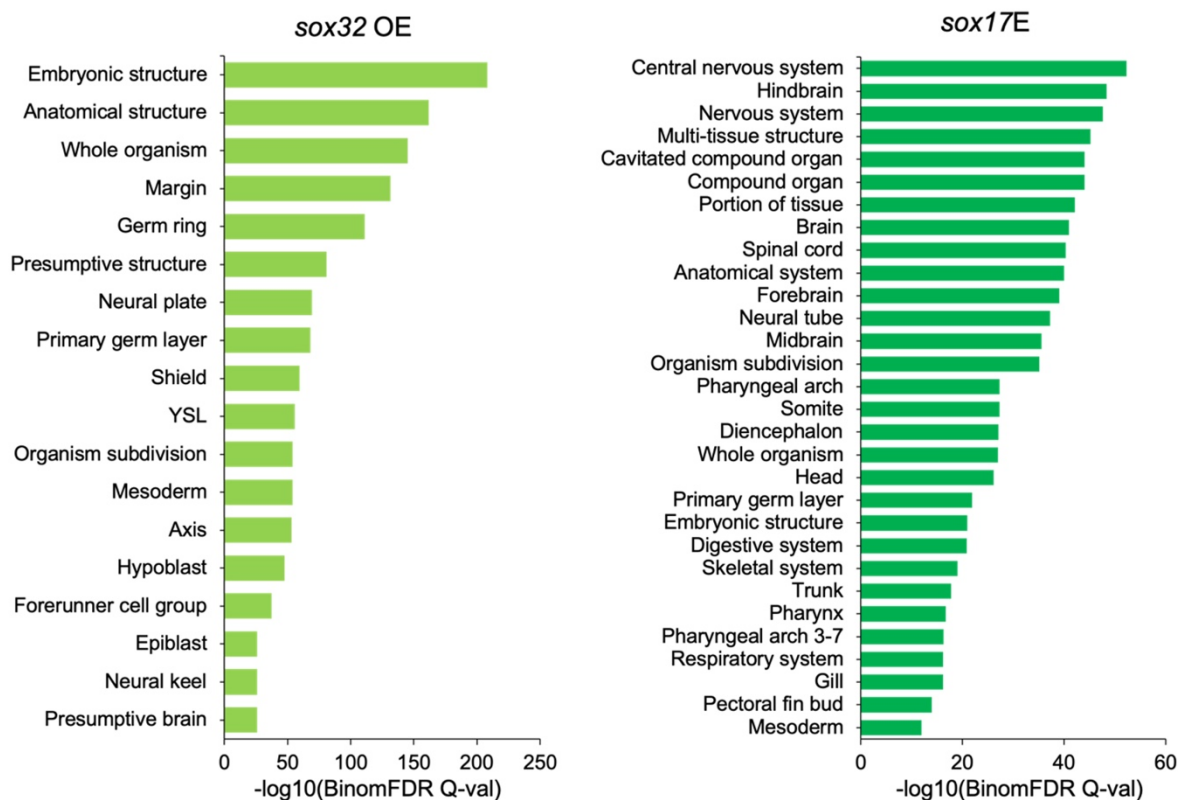


Figure 5-5: Bar graphs demonstrating over-represented anatomical terms associated with *sox32* OE and *sox17E* at 6 and 28 h.p.f., respectively. FDR = False discovery rate.

5.3.2 *Sox32* OE and *sox17E* DARs are enriched with motifs for endoderm-expressed factors

Next we sought to look at enriched TF binding motifs in *sox32* OE and *sox17E* DARs that likely control differential gene expression at different stages of development, and this was achieved by using HOMER (Heinz *et al.*, 2010). HOMER was used to identify motifs within *sox32* OE and *sox17E* DARs, which are likely to play a role in the endoderm at the different stages of development. Motif analysis revealed binding motifs of known TFs that are crucial during endoderm development in both *sox32* OE and *sox17E* DARs, as shown in Figure 5-6, Table 7.25 and Table 7.26.

The most over-represented TF-binding motif in *sox32* OE DARs was for CTCF, a multifunctional TF that plays a conserved role in gene regulation and organisation (Delgado-Olguín *et al.*, 2011; Kim, Yu and Kaang, 2015) (Figure 5-6A). Moreover, binding motif of Brn1 was also present in *sox32* OE DARs, which plays a crucial role in distal tube formation and kidney function in mammals (Nakai *et al.*, 2003). Though there are currently no information on Brn1 function in zebrafish, Brn1 appears to be expressed and required for midgut specification in sea urchin embryo *Strongylocentrotus purpuratus* (Yuh, Dorman and Davidson, 2005).

Thus, more downstream analysis is required to decipher the role of Brn1 in the zebrafish endoderm. In addition to this, binding sites for various T-box factors were also over-represented in *sox32* OE DARs. For instance, Tbx6 binding motif was significantly enriched within *sox32* OE DARs. Tbx6 is a member of the Tbx6/Tbx16 subfamily of T-box factors and has been shown to be essential for endoderm and KV formation, although it is missing in most mammals, except in marsupials and monotremes (Ahn, You and Kim, 2012; Nelson *et al.*, 2017). Given that T-box factors recognise the same core consensus sequence, it is likely that the Tbx6 motif represents binding sites of other T-box family members including Tbx16.

On the other hand, the most enriched motifs within *sox17E* DARs correspond to members of the Sox family of transcription factors. Indeed, based on motif analysis, the binding site for Sox21 was the most enriched motif within *sox17E* DARs which has previously been shown to act as a transcriptional repressor of neuronal differentiation within the developing nervous system (Argenton *et al.*, 2004; Whittington *et al.*, 2015). The binding sites for Sox2 and Sox3 were also over-represented in *sox17E* DARs, which are implicated in the maintenance of neural progenitor fate (Dee *et al.*, 2008). In addition, *sox17E* DARs were enriched for Sox6 and Sox10 binding motifs which are implicated in muscle fiber type specification and neural crest development, respectively, in zebrafish (Kelsh, 2006; Jackson *et al.*, 2015). However, Sox factors are known to bind the same consensus sequence (Dong, Wilhelm and Koopman, 2004) and thus, it is likely these over-represented Sox motifs are bound by factors that play essential roles during early endoderm organ formation. In addition, TR4 TF binding motif was also enriched in *sox17E* vs *sox32* OE DARs. In mice, Tr4 is believed to play essential roles in many processes, including aging, male/female fertility, neuron and brain development but also in glucose and lipid metabolism (Lin *et al.*, 2014). Motif analysis also revealed binding motifs of known TFs that are crucial during endoderm organ formation in *sox17E* DARs and these include Tead factors. In early zebrafish development, Tead/Yap1 signalling is crucial for endoderm maintenance and migration however, at later stages of development, Tead/Yap1 has been shown to be crucial for pancreatic development (Fukui *et al.*, 2014; Cebola *et al.*, 2015).

Thus, both *sox32* OE and *sox17E* DARs are enriched for binding sites for TFs that function in the endoderm during different stages of development. However, DARs from both sample groups are also enriched for binding sites for TFs that either play roles in other lineages or represent important TFs that function within the endoderm in other species. Alternatively, many of the enriched TF binding motifs may represent other factors that play a major role in the endoderm.

A Sox32 OE			B sox17E		
Motif	Name	P-value	Motif	Name	P-value
	CTCF(Zf)	1e-219		Sox21(HMG)	1e-76
	Brn1(POU, Homeobox)	1e-216		TR4(NR)	1e-62
	Tbx6(T-box)	1e-208		Sox3(HMG)	1e-61
	Oct6 (POU, Homeobox)	1e-191		Sox10(HMG)	1e-57
	BORIS(Zf)	1e-189		COUP-TFII(NR)	1e-46
	Eomes(T-box)	1e-178		TEAD3(TEA)	1e-45
	Tbet(T-box)	1e-171		TEAD1(TEA)	1e-41
	Sox17(HMG)	1e-129		Sox6(HMG)	1e-36
	Oct11(POU, Homeobox)	1e-119		Sox17(HMG)	1e-34
	Zic2(Zf)	1e-99		TEAD4(TEA)	1e-32
	Zic3(Zf)	1e-95			
	Oct2(POU, Homeobox)	1e-92			
	Oct4 (POU, Homeobox)	1e-64			

Figure 5-6: Motif analysis demonstrating TF binding sites in *sox32* OE and *sox17E* DARs. These motifs were selected since they were significantly over-represented in (A) *sox32* OE DARs or (B) *sox17E* DARs. Some of these binding sites are likely bound by factors that either play a role on the endoderm in zebrafish or in other species. However, some of these over-represented motifs are likely bound by factors that do not play a role in the endoderm but represent other factors that do. Motifs are ranked by statistical significance.

An example of a gene that is highly expressed in *sox32* OE compared to *sox17E* is *cxcr4a* (Figure 5-7A), which is maternally contributed and plays an essential role in migration and proliferation of endoderm progenitors (Mizoguchi *et al.*, 2008; Stückemann *et al.*, 2012) and KV formation (Liu *et al.*, 2019). However, *cxcr4a* is also expressed in the branchial arches which are derived from all three germ layers which may explain why its expression is also upregulated in *sox17N* and *sox17E* (Thisse *et al.*, 2001). An example of a gene that is highly expressed in *sox17E* vs *sox32* OE is *six1b* (Figure 5-7B), a zebrafish orthologue of mammalian *Six1*, which is required for thymus and parathyroid development (Zou *et al.*, 2006). Expression of *Six1* in thymus and parathyroid is evident at embryonic day 14.5 in mice, which is equivalent to ~48 h.p.f. in zebrafish (Kulkeaw and Sugiyama, 2012). Although the role of

six1 in the zebrafish endoderm is not yet known, motif analysis reveals the presence of Six1 binding motifs in *sox17E* and not *sox32* OE DARs (Table 7.26). Thus, this coincides with its function at later stages of endoderm development in comparison to *cxcr4a*.

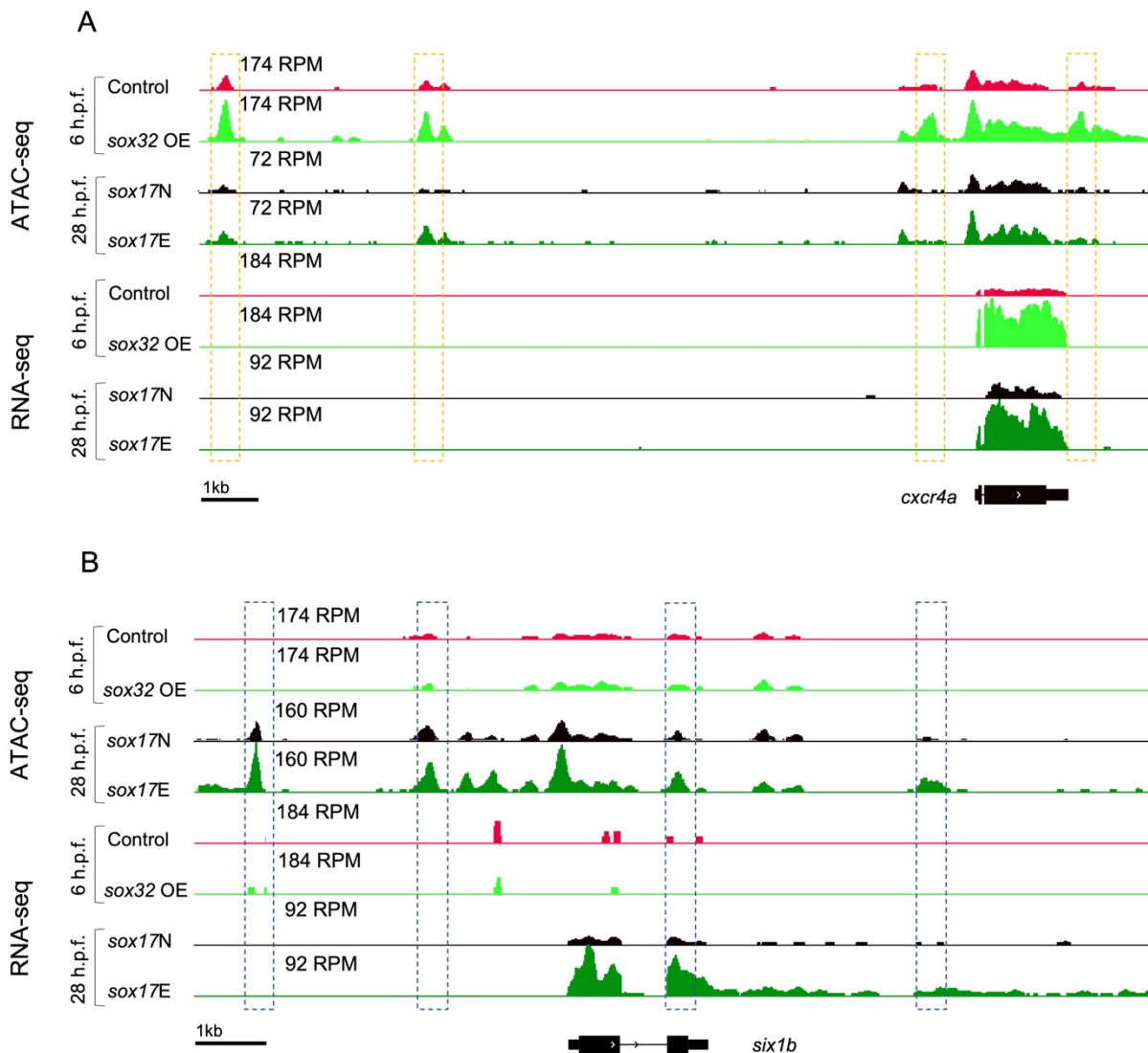


Figure 5-7: Examples of genes which exhibit differential expression in the endoderm at different stages of development (6 vs 28 h.p.f.). (A) *Cxcr4a* expression is upregulated in *sox32* OE compared to *sox17E*. (B) Expression of *six1b* is upregulated in *sox17E* compared to *sox32* OE. Read counts were scaled to adjust for library size. Putative enhancers of genes are indicated by dashed lines. RPM = reads per million.

Our results reveal that *sox32* OE is associated with terms that correspond with early endoderm development while *sox17E* coincides with terms associated with early organogenesis. We also show that the chromatin landscape changes in the endoderm during development and that endoderm markers including *foxa3*, appear to be regulated differently at different stages of development. Finally, we show that *sox32* OE and *sox17E* DARs harbour binding motifs of TFs with unknown functions in the endoderm but also reveal the presence of motifs of TFs that play crucial roles during endoderm development and these include, Sox and Tead factors.

5.4 Discussion

Changes in the chromatin landscape are required to induce ZGA which is characterised by a number of transcriptional waves that begin at 64-cell stage and end when cells within the embryo adopt different cell fates during gastrulation (Giraldez *et al.*, 2006; Heyn *et al.*, 2014; Hadzhiev *et al.*, 2019; Pálffy *et al.*, 2020). However, beyond ZGA, how the chromatin landscape changes in the endoderm between gastrulation and early organogenesis in zebrafish is not well documented. Previous studies have shown that TF binding at CRMs play an important role in directing cell differentiation of ESCs to a wide variety of cell types by driving lineage-specific gene expression (Heinz *et al.*, 2015). In human embryogenesis, over ~100,000 enhancers have been identified to aid in the differentiation of human ESCs from definitive endoderm and gut tube formation to pancreatic endoderm, and that many of these enhancers were found to regulate stage-specific gene expression (Wang *et al.*, 2015). Wang *et al.* (2015) also identified a number of poised enhancers during gut tube formation, which are primed to induce expression of genes involved in the formation of all primitive gut tube organs, including pancreas, lung, and liver. These enhancers were also enriched for binding motifs of known regulators of endoderm organogenesis, such as FOXA, HNF4A, GATA and HNF1. In mice, a number of gut-specific putative CRMs have also been identified and, similar to the human pancreatic study, many CRMs were primed to activate the differentiation of foregut and midgut fate (Banerjee *et al.*, 2018). Though a number of putative enhancers have been identified to aid in the differentiation of gut tube organs as well as the intestine from primitive states, this is still missing for other endodermal organs such as the thymus and parathyroid. To find developmentally important CRMs that aid in the differentiation of endoderm lineages from a primitive state, ATAC-seq analysis of endoderm cells from *sox32* OE (6 h.p.f.) and *sox17E* (28 h.p.f.) were compared. Our analyses reveal that the chromatin accessibility differs between both samples groups and demonstrated that ~80% of DARs exhibit higher accessibility in *sox32* OE compared to *sox17E* (Figure 5-3). This is likely because at 6 h.p.f., *sox32* OE likely represents a more homogenous cell group, while at 28 h.p.f., *sox17E* represents a more heterogeneous cell group with different accessibility profiles. However, this difference in the accessibility profile can also be because three biological replicates of *sox32* OE were used while only two biological replicates were attained for *sox17E*. Moreover, based on hierarchical clustering, there appears to be a difference in mapped read numbers at each consensus region between the two *sox17E* replicate samples. Thus, an additional replicate sample for *sox17E* will likely lead to more accurate and refined analysis. Despite this, it is clear that there is a difference in CRMs that function at each stage of endoderm development, in accordance with previous datasets in humans and mice (Wang *et al.*, 2015; Banerjee *et al.*, 2018).

GREAT also revealed that many of the over-represented anatomical terms in *sox32* OE were associated with early endoderm and DFC formation (Figure 5-5). Indeed, terms including presumptive brain, mesoderm and hypoblast, were associated with genes including *tbxta* which works cooperatively with *tbx16* to induce endoderm formation (Nelson *et al.*, 2017), while mesoderm and hypoblast were linked to genes such as *yap1*, *fscn1a*, *gata5*, which are required for early endoderm specification and migration (Reiter *et al.*, 1999; Reiter, Kikuchi and Stainier, 2001b; Fukui *et al.*, 2014; Liu *et al.*, 2016). GO analysis using PANTHER (Thomas, Campbell, *et al.*, 2003; Thomas, Kejariwal, *et al.*, 2003) also revealed that genes associated with the mesoderm, including *tbxta* and *tbx16*, were also expressed in the DFCs and are required for KV formation (Amack, Wang and Yost, 2007). Thus, while terms such as hypoblast, mesoderm and presumptive brain were over-represented in *sox32* OE, GO analysis revealed that many of the genes are not unique to these tissues since they are required for early endoderm and DFC/KV formation. Similar to *sox32* OE, genes proximal to *sox17E* DARs play a role in endoderm organogenesis and left/right asymmetry formation (Figure 5-5). GO analysis also revealed that genes required for liver and pancreas formation including *jag1b*, *fgfr2* and *rargb* (Stafford and Prince, 2002; Lorent *et al.*, 2004; Manfroid *et al.*, 2007; Kan, Junghans and Belmonte, 2009) were associated with mesoderm and CNS. Notably mesoderm and CNS were also associated with genes required for left/right asymmetry establishment which includes *chd*, *fgfr2* (Aamar and Dawid, 2010; Liu *et al.*, 2011) and *eomesa* which has been shown to induce ectopic formation of DFCs when overexpressed (Bjornson *et al.*, 2005). Thus, genes proximal to *sox32* OE DARs were associated with endoderm and KV formation, while genes proximal to *sox17E* DARs are involved in endoderm organogenesis and laterality establishment. However, these genes are also not unique to endoderm since they are expressed in other cell identities during development.

In relation to GREAT, motif analysis also revealed binding motifs of TF in *sox32* OE and *sox17E* DARs that play important roles in the endoderm at different stages of development either in zebrafish or in other species (Figure 5-6). For example, the most over-represented TF-binding motif in *sox32* OE DARs was for CTCF, a multiple zinc-finger protein which plays an integral part in regulating endoderm formation at the expense of ectoderm fate (Kaspi *et al.*, 2013). In mammals, loss of CTCF-binding site at super enhancer of miR-290-295, which is required for ESC maintenance and survival, results in downregulation of endoderm genes including *Gata6* and *Sox17* and upregulation of ectodermal lineage genes (Kaspi *et al.*, 2013; Downen *et al.*, 2014). In zebrafish, CTCF is ubiquitously expressed at early stages and by 48 h.p.f., it becomes restricted to the gut and branchial arches (Pugacheva *et al.*, 2006; Delgado-Olguín *et al.*, 2011). However, the role of CTCF will need to be further investigated in zebrafish since it was previously shown that CTCF is not essential for zebrafish embryogenesis and its

binding is not enriched at topologically associating domains (Pérez-Rico, Barillot and Shkumatava, 2020). In addition to this, the binding site of BORIS, a paralog of CTCF was also enriched in *sox32* OE DARs. It is believed that BORIS, also known as CTCF-like (CTCFL), arose due to duplication of CTCF in early mammals. Interestingly, compared to CTCF, in humans, *BORIS* was shown to be transcribed only in certain parts of the developing and adult testes (Hore, Deakin and Marshall Graves, 2008). No orthologues of *BORIS* could be detected in the zebrafish genome using nucleotide BLAST. It is therefore likely that this motif may represent the binding site of other similar factors including CTCF which may play roles in the endoderm during development. Moreover, binding motif of Brn1 was also enriched in *sox32* OE DARs, which plays a crucial role in distal tube formation and kidney function in mammals (Nakai *et al.*, 2003). Although, the function of Brn1 in the zebrafish endoderm is still lacking, Brn1 appears to be expressed and required for midgut specification in sea urchin embryo *Strongylocentrotus purpuratus* (Yuh, Dorman and Davidson, 2005). This result may suggest a possible link between endoderm and kidney function.

The presence of T-box factors were also enriched within *sox32* OE vs *sox17E* DARs. As mentioned in 3.5.4, T-box factors including Eomes (the mouse homologue of Eomesa), is required for expression of *sox32*, the master regulator of endoderm fate (Du *et al.*, 2012). Aside from the Eomes binding site, binding motifs of other T-box factors were also over-represented in *sox32* OE vs *sox17E* DARs. Indeed, binding sites for Tbx6, Tbx21, and Tbet were highly enriched in *sox32* OE DARs. Interestingly, in zebrafish, the *tbx6* belongs to the *tbx6/tbx16* subfamily of genes which are known for their evolutionary relationship to the *Tbx6/Tbx16* subfamily genes in tetrapods (Ahn, You and Kim, 2012). Additionally, *tbx6*, *tbx16* and *tbxta* are also expressed in overlapping domains that regulate gene expression and developmental fate within the zebrafish mesoderm and *tbx6* was also shown to effectively compete with *tbxta* to regulate T-site-dependent transcription (Goering *et al.*, 2003). On the other hand, Tbx21, previously known as Tbet, is believed to be expressed in a wide variety of tissues including the gut and liver but also spleen, kidney, and skin of a healthy fish and therefore may be required for the formation of these organs (Mitra *et al.*, 2010). Since T-box factors bind the same consensus sequence and given that the position weight matrix is missing for Tbx16 and zebrafish Tbx21, it is therefore likely that these T-box motifs represent binding sites for other T-box factors including Tbx16 which plays an essential role during endoderm formation (Nelson *et al.*, 2017).

Binding sites for Oct6 were also over-represented in *sox32* OE DARs. Oct6, also known as Pou3f1 is believed to be essential for neural differentiation of mouse embryonic stem cells in mice (Zhu *et al.*, 2014). This role is likely to be conserved in zebrafish since Pou3f1 is believed

to be involved in the dorsoventral patterning of the central nervous system (Hauptmann and Gerster, 1996). Binding sites for other Oct family members were also over-represented. For instance Oct2, Oct4 and Oct11 binding sites are also significantly enriched within *sox32* OE DARs. In mammals, Oct2 and Oct11 are part of the same POU class II subfamily and are associated with B-cell maturation and epidermal keratinocyte differentiation, respectively (Tantin, 2013; Zhao, 2013). In contrast to this, the interaction between the pluripotency factor Oct4 and Sox17 is required to drive endoderm formation in mammals (Aksoy *et al.*, 2013). However, there are currently no information on the presence of Oct2 and Oct11 in zebrafish, while the nearest zebrafish homolog to mammalian Oct4 is Pou5f3 (Lunde, Belting and Driever, 2004). Given that POU factors bind the same consensus region, known as the POU domain (ATGCA/TAAT) (Zhao, 2013), it may be possible that these POU binding sites represent the binding motifs of other members of the POU family of transcription factors, including Pou5f3. Pou5f3 is essential for endoderm formation in zebrafish, but its position weight matrix is missing in HOMER, and thus, it is likely that these binding sites correspond to the Pou5f3 binding motif within *sox32* OE DARs.

Sox32 OE DARs were also enriched for Zic family of zinc-finger transcription factors, Zic2 and Zic3. In zebrafish Zic2 is implicated in craniofacial and ocular morphogenesis (Sedykh *et al.*, 2017). Notably, development of the craniofacial region is believed to be derived from interactions between cranial neural crest cells, and adjacent surface ectoderm, neuroectoderm and pharyngeal endoderm (Cordero *et al.*, 2011). Thus, it may be possible that the presence of Zic2 binding site within *sox32* OE DARs is required to drive craniofacial morphogenesis within zebrafish, for which endoderm is required. The other member of the Zic family, Zic3 is believed to be essential during vertebrate development, particularly for early embryonic patterning. Indeed, among the targets of Zic3 include the Nodal and Wnt pathways which are known to regulate gastrulation but also left/right asymmetry establishment and neurogenesis, and that loss of its function results in defects in left/right asymmetry (Winata *et al.*, 2013). Thus, it is likely that Zic factors play essential roles during early zebrafish development.

Interestingly, multiple Sox binding sites were significantly over-represented in both *sox32* OE and *sox17E* DARs. Indeed, the most over-represented motif within *sox17E* DARs corresponded to Sox21, which is believed to act as a transcriptional repressor of neural differentiation within the developing nervous system in a number of species including zebrafish (Argenton *et al.*, 2004; Ohba *et al.*, 2004; Sandberg, Källström and Muhr, 2005). However, binding sites for Sox2, 3, 6, 10 and 17 were also over-represented. The same binding motifs were also significantly over-represented in *sox32* OE DARs. Sox proteins bind the same

conserved consensus sequence (A/T)(A/T)CAA(A/T)G using their HMG domain, and are known to play essential regulatory roles throughout development (Dong, Wilhelm and Koopman, 2004). For instance, the binding site for Sox17 was significantly enriched within *sox32* OE DARs and this is likely because Sox17 is not only an endodermal factor but is also necessary for DFC/KV function. However, since Sox factors bind the same consensus sequence, it may be possible that the presence of the Sox17 binding motif reflects the binding site of other members of the Sox family. For instance, the master inducer, Sox32 is crucial for endoderm and KV formation and given that the position weight matrix is missing for Sox32 in HOMER, it is likely that this Sox17 binding site reflects the binding site of Sox32 within *sox32* OE DARs. Similarly, the binding site for Sox17 was also over-represented in *sox17E* DARs, however, given that Sox17 does not function in the endoderm at 28 h.p.f., this binding site may therefore correspond to the binding site of other Sox family members, for instance Sox4b, which is required for the differentiation of alpha cells within the pancreas (Mavropoulos *et al.*, 2005)..

In addition to these factors, *sox17E* DARs harbour binding motifs for factors that are crucial during endoderm organ formation and these include Tead factors. During early zebrafish embryogenesis, Tead/Yap1 signalling is required for migration and maintenance of endodermal progenitor cells (Fukui *et al.*, 2014), however, it is likely that this signalling pathway is also essential during later stages of development. As mentioned previously in 4.10.4, TEAD/YAP1 signalling in human is crucial for regulating expression of key pancreatic genes during development. This is also consistent in zebrafish since mutations in binding sites of Tead factors resulted in a marked reduction in pancreatic progenitor cells (Cebola *et al.*, 2015). Similarly, *Sox17E* DARs were also enriched for binding motifs for COUP-TFII, which is believed to be located in endodermal-derived tissues, particularly the liver and pancreas, but has also been implicated in the development of a number of tissues and organs, including the heart, blood vessel, muscles and limbs in mammals (Pereira *et al.*, 1999; Zhang *et al.*, 2002; Lee *et al.*, 2004). This role of COUP-TFII seems to be evolutionary conserved since COUP-TFII is crucial for lymphatic and venous development in zebrafish and *Xenopus* (Aranguren *et al.*, 2011). However, little is known about the function of COUP-TFII within the endoderm and thus, further downstream analysis is necessary to decipher the role of COUP-TFII in the liver and pancreas.

Thus, our results reveal a number of putative enhancers that may regulate endoderm formation at different stages of development. Our results also suggest the presence of TF binding motifs that are present in the endoderm but their role within the endoderm is not yet established. Thus, investigating the roles of these putative enhancers through various

experiments, including transgenic reporter assay may shed light into identifying developmentally important CRMs that are critical during endoderm specification and early organogenesis. This can further aid in our understanding of CRMs that are required to regulate stage-specific gene expression at the two different stages of endoderm development.

Chapter 6 Discussion

6.1 Sox32 is a master inducer of endoderm and DFC fates

Several studies have already helped identify a number of developmentally important CRMs that are implicated in endoderm specification and that act upstream of *sox32*, as discussed in Figiel, Elsayed and Nelson, (2021), however, little is known about developmentally important CRMs that act downstream of *sox32*. Previously, comparison of the *sox17* genomic sequence to orthologous genomic regions from nine other species revealed important CRMs regulating *sox17* gene during zebrafish development (Chan *et al.*, 2009). This study also revealed binding sites for Sox32 and Pou5f3 within the *sox17* CRMs, which are known to work synergistically to regulate *sox17* expression (Lunde, Belting and Driever, 2004). However, other developmentally important *sox32*-regulated CRMs are still not known. For this reason, we sought to investigate CRMs that become accessible downstream of *sox32* and that are likely implicated during endoderm formation. A major challenge of studying endoderm-specific CRMs is that endoderm constitutes a minor cell population (Figiel, Elsayed and Nelson, 2021). Previously, overexpression of *sox32* has been found to induce ectopic endoderm formation in zebrafish embryos while knockdown of *sox32* completely abolished endoderm fate (Kikuchi *et al.*, 2001), and thus, *sox32* is considered a major inducer of endoderm fate. Thus, for this study, since endoderm is a minor population and that *sox32* is a potent inducer of endoderm formation, we sought to find *sox32*-regulated CRMs by overexpressing *sox32*.

In zebrafish, *sox32* induces formation of endoderm and DFCs and knockdown of *sox32* results in a complete lack of *sox17* expression in both lineages (Dickmeis *et al.*, 2001), yet DFC formation was never discussed in the context of *sox32* overexpression. Our results suggest that *sox32* OE differential accessible regions (DARs) are proximal to genes that are not only associated with endoderm but also DFCs relative to control. In relation to this, *sox32* OE transcripts were highly concordant with endoderm-enriched genes at early stages of development and also DFCs but not with mesoderm and ectoderm-enriched genes from published single-cell RNA-seq data (Wagner *et al.*, 2018). This is in agreement with previous literature which showed that *sox32* mRNA injection of a single marginal blastomere at 16-cell stage, which would normally contribute to prechordal plate and notochord, instead accumulated the entire endoderm and adopted an endoderm fate (Kikuchi *et al.*, 2001). Consistent with this, fate mapping experiments revealed that in the absence of *sox32*, labelled marginal cells adopt a mesodermal fate (Dickmeis *et al.*, 2001). Thus our data not only suggests that *sox32* OE induces ectopic endoderm formation but also DFCs, a phenomenon

not previously mentioned in literature and that this is therefore in accordance with the function of *sox32* in zebrafish during development (Alexander *et al.*, 1999; Dickmeis *et al.*, 2001; Kikuchi *et al.*, 2001; Essner *et al.*, 2005).

Though *sox32* plays a central role in endoderm and DFC formation (Alexander *et al.*, 1999), it is specific to teleosts and appears to be absent in other species (Voldoire *et al.*, 2017). In other vertebrates, *sox17* is required for both endoderm formation (Séguin *et al.*, 2008; Niakan *et al.*, 2010) and is also expressed in the laterality organ, known as the node in mice (Hassoun, Püschel and Viebahn, 2010) and Henson's node in chick (Chapman *et al.*, 2007). In zebrafish, *sox17* acts downstream of *sox32* and analogous to *sox32* mutants, *Sox17*-null mice exhibit endoderm and laterality defects (Kanai-Azuma *et al.*, 2002; Saund *et al.*, 2012) and thus, *sox32* appears to be functionally equivalent to *Sox17* in other species. Our data suggests that overexpression of *sox32* does not drive artificial accessibility at loci that are not accessible in control embryos. Since *sox32* is functionally equivalent to *Sox17* in other vertebrates, *sox32* OE can be used to study CRMs that are not only important for endoderm formation but also for left/right asymmetry establishment.

6.2 *Sox32* OE upregulates expression of Nodal ligands

Endoderm formation is dependent on a number of TFs that act downstream of Nodal signalling but upstream of *sox17*, and these include *mixl1*, *gata5* and *sox32* (Kikuchi *et al.*, 2000, 2001; Reiter, Kikuchi and Stainier, 2001). Embryos that lack Nodal ligands *ndr1* and *ndr2* or maternal-zygotic mutants of *tdgf1* (*MZtdgf1*), an essential coreceptor for Nodal signalling, completely lack expression of *sox32* and fail to induce endoderm and DFCs (Feldman *et al.*, 1998; Gritsman *et al.*, 1999; Warga and Kane, 2018). Unlike *sox32*, which is completely absent or severely reduced in *MZtdgf1* or *Ztdgf1* mutants (Aoki *et al.*, 2002), DFCs are unaffected in *Ztdgf1* but are completely lacking in *MZtdgf1* (Warga and Kane, 2018). This means that higher Nodal signalling is likely required to induce endoderm formation compared to DFCs. Though Nodal signalling is required to induce expression of *sox32* and therefore endoderm formation, *sox32*-expressing *MZtdgf1* cells injected at the margin of control embryos were able to induce a population of *sox17*⁺ cells (Kikuchi *et al.*, 2001), that occasionally populated the anterior endoderm, albeit at low levels (Aoki *et al.*, 2002). Thus, endoderm formation is completely dependent on *sox32* expression but likely also requires Nodal signalling at early stages of development.

Previous data have shown that *sox32* OE upregulates expression of *ndr1* and *ndr2* in zebrafish at 5.3-6 h.p.f (Kikuchi *et al.*, 2001). Contrary to this, our data suggests that *sox32* OE induces

significant upregulation in expression of *ndr1* at high stage (~3.3 h.p.f.) but not shield stage (6 h.p.f.) while no significant difference in expression was observed for *ndr2* at all stages tested between *sox32* OE and control. Though this contradicts the data in Kikuchi *et al.* (2001), our results are in agreement with previous studies which show that *sox32* OE increases the expression of *sox17* but decreases the expression of *ndr1* and *ndr2* at marginal blastomeres at 50% epiboly (Sakaguchi, Kuroiwa and Takeda, 2001). However, in accordance with Kikuchi *et al.* (2001), which indicated that *sox32* OE in MZ*tdgf1* cannot induce expression of *gata5* or *mixl1*, our results also suggest that *sox32* OE requires Nodal signalling to induce expression of *gata5* in the embryo while induction of *sox17* is independent of Nodal. This is again inconsistent with *in situ* hybridisation results from Sakaguchi, Kuroiwa and Takeda, (2001) which show that *sox32* MO abolished *gata5* expression at 80% epiboly despite the fact that *sox32* is downstream of *gata5* in the endoderm specification pathway. However, this is likely because beyond early gastrula stages, endodermal cells from *sox32* MO likely adopt a mesodermal fate (Dickmeis *et al.*, 2001) and no longer express endodermal regulators.

These differences in our data and results from previous studies are likely due to a number of reasons; unlike Kikuchi *et al.* (2001), which observed ectopic expression of Nodal ligands by *in situ* hybridisation, our results were demonstrated quantitatively by qRT-PCR. Thus, though it may appear that *sox32* OE induces high expression of Nodal ligands, our results suggest that *ndr1* and *ndr2* expression levels are likely to be low across the embryo. However, our data are in line with studies from Sakaguchi, Kuroiwa and Takeda, (2001), where despite using *in situ* hybridisation, showed that *sox32* OE does not upregulate the expression of *ndr1* and *ndr2*. Another reason for the discrepancies in *ndr2* expression is that perhaps high levels of *sox32* OE and subsequently *ndr1* are required to induce expression of *ndr2*. Thus, it is difficult to compare our datasets to data from Kikuchi *et al.* (2001) and Sakaguchi, Kuroiwa and Takeda, (2001) since different experiments were performed. Nevertheless, our results show that *ndr1* was upregulated at an earlier stage in *sox32* OE embryos relative to control and this is perhaps required to induce ectopic endoderm and DFC formation.

Notably, our results also demonstrate that overexpression of *sox32* significantly upregulated expression of *vgl14l*, and other DFC markers based on qRT-PCR and RNA-seq analyses. Both endoderm and DFC formation are dependent on Nodal signalling but unlike endoderm, DFC formation is unaffected by loss of *mixl1* or *gata5* (Kikuchi *et al.*, 2001) and requires maternal and not zygotic *tdgf1* (Warga and Kane, 2018), and thus, it is likely that different factors downstream of maternal *tdgf1* are required to regulate expression of *sox32* in DFCs compared to endoderm. In addition, our qRT-PCR analysis likely suggests that ectopic DFC formation occurs at an earlier stage than endoderm since endoderm is dependent on the expression of

crucial endoderm inducers including *gata5* and *mixl1* (Kikuchi *et al.*, 2000; Reiter, Kikuchi and Stainier, 2001).

6.3 Endoderm factors likely bind enhancers downstream of *sox32*

Previously, *Tbx16* and *Tbx16* were demonstrated to regulate expression of endodermal regulators *mixl1* and *gata5* as well as regulators of endoderm proliferation, *cxcl12a/b* (Nelson *et al.*, 2017). In relation to this, *Tbx16* and *Tbx16* in combination with other endoderm regulators including *Smad2* and *Eomesa*, were shown to bind to CRMs proximal to *mixl1* and *foxa2* (Nelson *et al.*, 2014, 2017). *Tbx16* and *Tbx16* also have established cooperative roles in normal KV formation and left/right establishment (Melby, Warga and Kimmel, 1996; Essner *et al.*, 2005; Amack, Wang and Yost, 2007). However, to our knowledge, *Tbx16* and *Tbx16* have not been previously shown to induce gene expression downstream of *sox32*. To test this, *sox32* OE DARs were compared to *Tbx16* and *Tbx16* ChIP-seq coordinates. Our analysis revealed that *sox32* OE DARs were highly concordant with *Tbx16* and *Tbx16* binding regions and that these regions were proximal to known regulators of endoderm formation, including, *gata5* and *foxa2* but also *cxcr4a*, which is involved in endoderm proliferation (Stückemann *et al.*, 2012) and KV formation (Liu *et al.*, 2019). Thus, our data suggests that *Tbx16* and *Tbx16* bind *sox32*-regulated regions perhaps to induce endoderm and KV formation. It is noteworthy however that *sox32* OE induces expression of Nodal ligands, *ndr1* and *ndr2*, which in turn induce expression of *tbx16* and *tbx16* (Rodaway *et al.*, 1999). Thus, an alternative hypothesis is that some of the regions that overlap with *Tbx16* and *Tbx16* ChIP-seq peaks are actually peaks that become accessible downstream of Nodal and not *sox32*. However, there isn't enough data to support this theory without conducting further ATAC-seq experiments on *sox32* OE *MZtdgf1* embryos to see whether those same regions are open and thus, this experiment will help clarify if those regions are open as a result of Nodal induction by *sox32* OE.

Moreover, maternally-contributed factor, *Nanog* has been shown to be an essential endoderm regulator since it induces the formation of the YSL, from which Nodal ligands arise (Xu *et al.*, 2012). Interestingly, previous studies have shown that a minority of *Nanog* ChIP-seq peaks overlap with that for *Smad2*, *Mixl1*, *Pou5f3*, *Mxtx2* and *Eomesa* and that these common regions were proximal to key endoderm factors such as *sox32*, *tbx16*, *gata5* and *ndr1* (Lunde, Belting and Driever, 2004; Reim *et al.*, 2004; Xu *et al.*, 2012; Nelson *et al.*, 2014, 2017). Previously, *Nanog* was also shown to interact with *Sox32*, which disrupts *Pou5f3*-*Nanog* complexes at the dorsal endoderm (Perez-Camps *et al.*, 2016). However, the role of *Nanog* downstream of *sox32* has never been explored. Despite a minority of common regions

between Nanog ChIP-seq peaks and *sox32* OE DARs, these regions were proximal to endoderm and DFC genes such as *tbx16* and *vgl4l*. Therefore our data suggests that Nanog likely binds to enhancers downstream of *sox32* and that this interaction may be important for inducing endoderm and DFC fate. This therefore contradicts previous studies which showed that MZ*nanog* results in a substantial loss of endoderm markers, while DFCs appeared less affected (Veil *et al.*, 2018). This study presents a new role for Nanog in DFC induction and that Nanog may be sufficient but not necessarily required to induce DFC formation in zebrafish.

6.4 Separating *sox17*+ endoderm from contaminating mesoderm at early organogenesis

Although *sox17* marks endodermal progenitor cells during gastrulation, beyond gastrulation, *sox17* is no longer expressed in the endoderm. Despite this, the study of the endoderm beyond gastrulation is facilitated by GFP expression from the *Tg(sox17:EGFP)* pan-endodermal transgenic line, which marks all endoderm progenitors after the endogenous *sox17* is silenced. However, by organ bud formation, the GFP expression from *Tg(sox17:EGFP)* not only marks the endoderm but also haemopoietic lineages and endothelial cells (Chung *et al.*, 2011). Thus, this makes it difficult to separate *sox17*+ endoderm from contaminating mesoderm since GFP from *Tg(sox17:EGFP)* marks both lineages.

Multiple lines of evidence led us to believe that *sox17*+ endoderm can be separated from *sox17*+ mesoderm based on GFP intensity. For instance, *sox17* expression appears to be strong during gastrulation but negligible beyond this stage (White *et al.*, 2017). Also, GFP immunostaining in Nelson *et al.* (2017) and fluorescence images of *Tg(sox17:EGFP)* at 24 h.p.f. show strong signal for GFP in the endoderm but no GFP can be detected in the LPM. Thus, we conducted a trial experiment where we sorted cells based on GFP intensity, where cells that express high levels of GFP were hypothesised to be endoderm while cells that express low levels of GFP were believed to be mesoderm. We then sought to investigate whether endoderm can be successfully isolated from mesoderm by qRT-PCR. However our results show that endoderm marker *foxa3* was enriched in GFP^{low} while *sox17* showed strong expression in GFP^{high} population and *gata1a* showed no difference in expression between GFP^{high} and GFP^{low}. Thus, this study concludes that GFP intensity cannot be used to isolate *sox17*+ endoderm from *sox17*+ mesoderm beyond gastrulation. Notably an alternative approach to this method is to use FACS on *Tg(sox17:EGFP)* embryos injected with *sox32* morpholino to set gating parameters against *sox17*+ mesoderm. However, based on the results obtained in this study, it is likely that some mesoderm and endoderm lineages share

similar GFP expression levels and thus, isolation of endoderm from mesoderm lineages may not be completely achieved with this approach. In addition to this, a drawback to both of these methods is that only a handful of markers are selected for qPCR, where some markers are likely based on *in situ* hybridisation results, which show location of gene expression patterns but not at single-cell resolution. Therefore, it is often difficult to interpret whether these genes are localised more to a specific cell population, i.e.: endoderm vs mesoderm. Furthermore, comparison of *in situ* hybridisation results to RNA-seq data have shown that some markers may not be appropriate. For instance, *in situ* hybridisation show strong *hhex* expression in the endoderm, while whole embryo single-cell RNA-seq data show high expression of *hhex* in the mesoderm compared to the endoderm (Wagner *et al.*, 2018). Thus, the study in this report highlights that it is likely near impossible to isolate endoderm from mesoderm based on GFP intensity alone and that other approaches are required.

6.5 Sorted GFP+ cells comprise both endoderm and KV transdifferentiated lineages at 28 h.p.f.

To characterise GFP expression in the endoderm and mesoderm, we generated *sox32*-deficient *Tg(sox17:EGFP)* embryos by using *sox32*-MO. Similar to *sox32* mutants (Dickmeis *et al.*, 2001; Kikuchi *et al.*, 2001), *sox32* morphants completely lack endoderm but also show a lack of GFP+ cells in the posterior notochord. Previous fate mapping experiments have shown that KV cells later contribute to caudal tail notochord (Melby, Warga and Kimmel, 1996). Therefore, we sought to investigate the developmental fate of the KV by injecting the red fluorescent dye Rhodamine-dextran (RD) into the yolk sac of mid-blastula staged (~3 h.p.f.) *Tg(sox17:EGFP)* embryos, since at this stage, cytoplasmic bridges are only open between the yolk sac and Wilson cells, which give rise to DFCs (Amack and Yost, 2004). At 35 h.p.f. we show that the posterior notochord was double labelled with GFP and RD and thus, these *sox32*-dependent cells are likely derivatives of the KV in accordance with Melby, Warga and Kimmel, (1996). However, there were clear drawbacks to this experiment. For instance, much of the RD was trapped in the yolk sac of a large number of injected embryos and thus failed to transfer to the blastomeres. This is likely due to the large size of dextran-tagged dyes which work much more efficiently when injected directly into cells (Paredes *et al.*, 2008). Nevertheless, our results demonstrate that GFP+ cells in the posterior notochord are likely derivatives of the KV, in accordance with previous data.

To test that *sox17*+ endoderm can be sufficiently separated from hemopoietic and endothelial lineages, a triple transgenic line was created which expresses different fluorescent proteins in *sox17*+ lineages (GFP), endothelial cells (mCherry) and blood lineages (dsRed). The use of

the triple transgenic line was first validated by qRT-PCR and showed that markers of endoderm were highly expressed in *sox17E* while mesoderm markers were enriched in *sox17M*, and thus, the results demonstrate that *sox17+* endoderm can be sufficiently isolated from *sox17+* mesoderm lineages by using the triple transgenic line. In relation to this, *sox17E* DARs were proximal to genes that are expressed and play a role in the endoderm compared to *sox17M* and *sox17N*, which were associated with genes of the central nervous system (CNS) and hemopoietic cell lineages, respectively. Though some endodermal terms including digestive system and pharynx were also associated with *sox17N*, this is likely because many endoderm-expressed genes are not unique to endoderm and play a role in the CNS. In relation to this, *sox17N* exhibit higher accessibility for *sox17E* DARs compared to *sox17E*, while *sox17E* was also accessible for *sox17N* DARs. This likely suggests that regulatory regions are also not unique to *sox17+* endoderm since they also likely play a role in non-*sox17+* cell populations.

Also motif analysis revealed enrichment of binding motifs of factors in *sox17E* DARs that play a role in the endoderm. For example in zebrafish, interactions between Tead factors and transcriptional regulator Yap1, are crucial for maintenance and formation of endoderm progenitors, and Yap1 morphants exhibit defects in anterior endoderm at 26 h.p.f. (Fukui *et al.*, 2014). Similarly, in humans TEAD/YAP1 signalling occupied the same CRMs as other pancreatic regulators including PDX1 and HNF1B and was crucial in activating expression of key pancreatic factors including *GATA4*, *GATA6* and *FGFR2* (Cebola *et al.*, 2015). However, Tead4 binding motif was also enriched in *sox17E* DARs which is expressed in other species (Nishioka *et al.*, 2008), but appears to be missing in zebrafish. Since Tead proteins bind the same consensus sequence (Anbanandam *et al.*, 2006), it is likely that they play redundant roles in other species (Nishioka *et al.*, 2008). In line with this, previous experiments have used mRNA of mouse *Tead2* fused to the transcriptional repressor domain Engrailed to test the effects of Tead repression on pancreas development in zebrafish and showed that the pancreas was reduced in size at 48 h.p.f., despite the fact that zebrafish do not express *tead2* (Cebola *et al.*, 2015). In addition to their role in the endoderm, Tead proteins are also crucial for KV formation since ablation of Tead1a and Tead3a results in laterality defects within zebrafish (Fillatre *et al.*, 2019). Interestingly, Tead proteins appear to be expressed in the pharyngeal endoderm at 24 h.p.f., based on single-cell RNA-seq data (Wagner *et al.*, 2018). Thus, Tead factors play important roles during early stages of the zebrafish embryo, particularly during endoderm development but also during the morphogenesis of a number of crucial endodermal-derived organs. Our results therefore uncover a number of putative enhancers with Tead binding sites that are likely important during early organogenesis.

Binding motifs of Six factors were also enriched in *sox17E* DARs. In mice, Six1 and Six4 act synergistically to control the morphogenesis of the thymus and parathyroid (Zou *et al.*, 2006). On the other hand, human SIX2 is required for maturation of β -cells into secreting cells in the pancreas (Velazco-Cruz *et al.*, 2020). However there are currently no information regarding the role of Six TFs in the endoderm in zebrafish, though they appear to be expressed in the endoderm based on single-cell RNA-seq data (Wagner *et al.*, 2018). Thus, our data suggests the presence of binding sites of factors with less well established roles in the endoderm in zebrafish compared to other species. Hence, functional studies of these regions will likely decipher a novel role for Six factors in relation to the endoderm in zebrafish.

To further test the relationship between *sox17E* sorted cell population and endoderm and KV, *sox17E* upregulated transcripts were compared to genes enriched in the endoderm at stages 6, 8, 10, 14, 18 and 24 h.p.f. and DFCs at 8 h.p.f., attained from published RNA-seq data (Wagner *et al.*, 2018). Our results reveal a significant correlation between *sox17E* transcripts and endoderm-enriched genes at later stages of development, particularly, 14, 18, and 24 h.p.f. but also DFCs at 8 h.p.f. This analysis further confirms the presence of genes likely required for the maintenance of endoderm fate but also that *sox17E* likely comprises KV transdifferentiated cells even at 28 h.p.f.. It is therefore likely that some of the genes expressed in the KV continue to be expressed at later stages of development despite KV cells transdifferentiating to the posterior notochord.

Though our data supports with the use of the triple transgenic line in studying the endoderm at early organogenesis, we also observe some drawbacks. For instance, the number of reads that map to each consensus region is different between the two *sox17E* replicate samples. This difference in read numbers may be the underlying reason behind the small number of DARs observed for *sox17N* relative to the other cell populations. Alternatively, this difference in read numbers may also be due to the differences in chromatin accessibility across the different heterogeneous endoderm populations, which means that many regions are likely lost due to ensemble averaging. Moreover, whole-mount *in situ* hybridisation for *sox17* shows that at post-gastrula stages, *sox17* is expressed in neural tube amongst other tissues including LPM, ICM and KV (Chung *et al.*, 2011). Whilst the triple transgenic line distinguishes between haemopoietic and cell lineages from the endoderm, it does not take into account the expression of *sox17* in the neural tube. However, GFP staining was never observed in the neural tube based on GFP-immunostaining (Nelson *et al.*, 2017) and fluorescent images of *Tg(sox17:EGFP)* at 24 h.p.f, and thus, it is unlikely that this population would be captured by FACS when sorting for *sox17E* cell populations. All in all, our results suggest that the triple

transgenic line can be effectively used to separate *sox17+* endoderm from mesoderm and as a result, can be used to look for endoderm-specific CRMs during development.

6.6 The chromatin landscape changes in the endoderm during development

The endoderm makes major contributions to the respiratory and gastrointestinal tracts and all associated organs (Wlizla and Zorn, 2014). A number of studies have already identified developmentally important CRMs that drive the differentiation of embryonic stem cells into endoderm-specific lineages. For example, studies on human embryonic stem cells have identified more than 100,000 enhancers across five different stages of pancreatic development but also identified a number of poised chromatin states linked to genes of other gut-tube-derived tissues, such as the liver and lung (Wang *et al.*, 2015). In relation to this, studies in mice reveal that early midgut enhancers harbour the potential to express midgut and foregut genes and that foregut enhancers are later decommissioned upon differentiation of the intestine (Banerjee *et al.*, 2018). Thus, these studies demonstrate that the chromatin landscape changes between endoderm specification and differentiation of the foregut and midgut lineages, however, a list of developmentally important CRMs required to direct the differentiation of embryonic stem cells into other endodermal organs, including the thymus and parathyroid, is still missing. To find developmentally important CRMs during two different stages of embryonic development, ATAC-seq data from *sox32* OE (6 h.p.f.) were compared to data from *sox17E* (28 h.p.f), and it was evident that the accessibility profile differs between the two sample groups, consistent with Wang *et al.* (2015). The data also shows that ~80% of DARs exhibit higher accessibility in *sox32* OE than in *sox17E*, and this is because at 6 h.p.f. prior to organ budding, the accessibility profile is likely to be similar across all cells in the embryo while at 28 h.p.f., the endoderm is much more heterogeneous, with each cell lineage adopting different accessibility profiles. Thus, it is likely that as a result of this, many of the *sox17E* DARs are lost due to ensemble averaging.

Our results also highlight that *sox32* OE DARs were proximal to genes that play a role during endoderm and DFC specification while *sox17E* DARs were proximal to genes that are crucial during early organogenesis and KV formation. However, it is apparent that many endoderm-expressed genes also play roles in other cell identities. For instance, based on Gene Ontology (GO) analysis using PANTHER (Thomas, Campbell, *et al.*, 2003; Thomas, Kejariwal, *et al.*, 2003), genes required for endoderm formation including *tbxta* and *bmp4* were also associated with epiblast and brain, while *yap1*, *gata5* and *fscn1a* were associated with hypoblast and mesoderm. Interestingly, GO analysis revealed that genes such as *tbxta* and *tbx16* which were associated with mesoderm, but are also important for KV formation. This is in line with previous

data which demonstrated that both *Tbx16* and *Tbx16* are required for intermediate mesoderm formation and somitogenesis (Mueller, Huang and Ho, 2010; Warga *et al.*, 2013; Kimelman, 2016) but also cooperatively work together to establish left/right asymmetry in zebrafish (Amack, Wang and Yost, 2007). Thus, this means that a large number of genes proximal to *sox32* OE DARs also play a role in other cell identities.

Similarly, genes proximal to *sox17E* DARs were associated with endodermal organ morphogenesis. For instance, terms including the digestive system, pharynx, respiratory system, and gills were over-represented in *sox17E* compared to *sox32* OE DARs. However, genes proximal to *sox17E* DARs also play roles in other cell identities. Based on GO analysis, genes required for liver and pancreas development, including *jag1b*, *fgfr2* and *rargb*, were also associated with the CNS (Yeo and Chitnis, 2007; Thisse and Thisse, 2008; Blanco, Bertucci and Unniappan, 2020). Thus, many endoderm-expressed genes are not unique to endoderm since they play roles in other cell identities.

TF binding motifs also appear to be different between *sox32* OE and *sox17E* DARs. For example, the most over-represented TF binding site in *sox32* OE DARs was for CTCF, which is known to regulate endoderm formation at the expense of ectodermal fate in mammals (Kaspi *et al.*, 2013). *Sox32* OE DARs were also enriched for Brn1 binding motifs, which is expressed and plays crucial roles in the kidneys in mammals (Nakai *et al.*, 2003). This is in line with previous studies which used *in situ* hybridisation and microarray expression data and showed that in mice, the endodermal marker *Sox17* is expressed at several key stages of kidney and urinary development and mutations of *SOX17* in humans are associated with congenital anomalies of the kidney and the urinary tract (Gimelli *et al.*, 2010). Thus our data may provide a likely link between endoderm specification and kidney function in zebrafish.

On the other hand, *sox17E* DARs were enriched for binding sites of factors that play a role in the endoderm at later stages of development. For instance, binding sites for Tead factors were amongst the most enriched motifs in *sox17E* DARs, where Tead/Yap1 signalling is not only crucial for correct migration and maintenance of endodermal progenitors, but also essential for regulating key pancreatic genes during development in both humans and zebrafish (Fukui *et al.*, 2014; Cebola *et al.*, 2015). Interestingly, binding motifs for a number of Sox factors were also over-represented in *sox17E* DARs. The most over-represented TF binding motif in *sox17E* DARs was for Sox21 which acts as a transcriptional repressor within the developing nervous system (Ohba *et al.*, 2004; Sandberg, Källström and Muhr, 2005; Whittington *et al.*, 2015). In addition to this, binding motifs for other Sox factors including Sox2, 3, 6, 10 and 17 were also enriched in *sox17E* DARs. However, Sox factors are known to bind the same

consensus region and thus it is likely that these motifs are bound by Sox factors that play a role in the endoderm. One example of an important Sox factor is Sox4b, which is required for the differentiation of alpha cells within the pancreas (Mavropoulos *et al.*, 2005). Since the position weight matrix of Sox4b is missing in HOMER, there is a high chance that these Sox motifs correspond to binding sites for other Sox members, including Sox4b, that play essential roles in the endoderm. Thus, our data suggests the presence of binding sites of factors that are likely important during early organ formation, and that further functional studies including the use of transient reporter assays will further decipher their roles within the endoderm during development.

Our study has helped identify a list of putative enhancers with essential TF binding sites that are likely important during two essential time points of early endoderm formation. Functional studies on these CRMs will undoubtedly give insight into the transcriptional control that govern early endoderm formation. Our work has also helped pave the way to finding endoderm-specific CRMs at early organ formation. Further studies on the endoderm using ATAC-seq at different developmental stages will likely provide a more dynamic insight into how the endoderm is regulated differently during development. Such studies will likely be invaluable for understanding the genetic basis of many congenital diseases in humans.

6.7 Future directions and experiments

The questions raised from this study are as follows:

Is Nodal signalling required to induce ectopic endoderm and DFC formation?

In this study, *sox32* OE was shown to upregulate expression of Nodal ligand, *ndr1*, which is required for endoderm formation but also *sox17*, which marks the endoderm and DFCs. Our data also shows that *sox32* is sufficient but not required for the enhanced accessibility observed in *sox32* OE vs control embryos. Thus, it is likely that other factors are required to induce ectopic formation of endoderm and DFCs. Since our data suggests that *sox32* OE upregulates expression of *ndr1*, which in turn upregulates expression of *gata5* and endogenous *sox32*, it is likely that ectopic endoderm and DFC formation relies on upregulation of Nodal signalling. This is in accordance with published data which shows that *sox32* OE upregulates expression of both *ndr1* and *ndr2* (Kikuchi *et al.*, 2001).

One way to test the effects of Nodal signalling on ectopic endoderm and DFC formation is by performing ATAC-seq on *MZtdgf1* mutants, which lack the essential coreceptor of Nodal

signalling in addition to *sox32* OE MZ*tdgf1* embryos. By looking at read densities across DARs from both mutants and comparing the results to data obtained from this study using *sox32* OE embryos, this will likely help determine whether Nodal signalling is required for ectopic endoderm and DFC induction.

Do cells within *sox32* OE represent two cell entities?

Our results demonstrate that *sox32* OE induces ectopic formation of endoderm and DFCs, however it is not known whether cells in *sox32* OE express genes of both entities or whether they represent two separate lineages. One way to test this is by performing single-cell RNA-seq and ATAC-seq on control and *sox32* OE embryos and looking at upregulated transcripts at the single-cell level. This will therefore help identify whether all cells have the same accessible CRMs and upregulated transcripts or whether they represent two separate entities and therefore express different genes.

Are DFCs induced earlier than endoderm in *sox32* OE embryos?

Overexpression of *sox32* was shown to upregulate expression of *vgl14l* at high stage relative to control, while expression of *gata5* and endogenous *sox32*, which are required to specify endoderm, were induced at later stages of development likely through activation of Nodal signalling. Thus, this suggests that *sox32* OE induces ectopic DFCs prior to endoderm. However, not enough markers were used for this experiment to confirm this. One way this can be explored is by using a wide variety of endoderm and DFC-markers for qPCR while also extending the number of developmental stages investigated between 1-cell stage to shield stage (6 h.p.f.). This will likely help identify the specific developmental stage at which both DFCs and endoderm are induced but also help determine whether DFCs are induced prior to endoderm fate.

Are CRMs identified in this study evolutionary conserved among zebrafish and humans?

Our results have identified a number of putative CRMs that are likely important during the different stages of endoderm formation: endoderm specification and at early onset of organogenesis. Regions of accessible chromatin identified by ATAC-seq include promoters and enhancers. Thus, to ensure that these CRMs are associated with enhancers, the regions from this study should be compared to known enhancer histone hallmarks including H3K27ac identified in whole-embryo ChIP-seq, which is used to distinguish between active and inactive

enhancers but also H3K4me1 which is present in either active or poised enhancers (Heintzman *et al.*, 2009). This will help eliminate any CRMs that are associated with either promoters or silencers, or any regions that are not proximal to endodermal genes. Following this, to further test whether these are active enhancers within the endoderm, tol2-based transient reporter assays should be performed. This entails designing a construct which will encompass an enhancer of interest upstream of a basal promoter that drives GFP expression. This construct will also harbour a *crystallin, alpha α* promoter, upstream of the *mCherry* reporter gene, which drives lens expression and provides a constitutive marker to control for injection. This method has already been used to find active enhancers within endothelial cells, which comprise less than 5% of the whole embryo (Quillien *et al.*, 2017). In addition to this, mutagenesis studies should also be performed to further dissect the functionality of enhancers. This can be performed by using transcription factor binding site (TFBS) analysis to identify TF binding sites within all CRMs from this study and this be achieved by computational programmes such as Meme suite FIMO (Grant, Bailey and Noble, 2011).

Finally, to find CRMs that are conserved between humans and zebrafish, a computational programme called Ancora should be used. Ancora provides a set of highly conserved non-coding elements between multiple species which may facilitate the discovery of developmentally important CRMs in humans. Previous studies have shown that approximately 12% of pancreatic enhancers from zebrafish aligned to the human genome. However, despite an overall lack of sequence conservation, a large number of TFBS were shared between human and zebrafish pancreas datasets (Bordeira-Carriço *et al.*, 2020). Our results may not only aid in identifying conserved sites but may also aid in the discovery of important TFBS that operate at different stages of endoderm development, particularly during endoderm specification and early organogenesis.

Our results uncover a number of putative CRMs that either act downstream of *sox32* or during organ budding stage. The more we understand the regulatory network that controls endoderm speciation and organ morphogenesis, the better our understanding of how enhancer-related mutations result in developmental defects as a result of aberrant gene expression. By validating these putative sites through the proposed experiments outlined above, a number of potentially important CRMs can be identified which may underlie a wide variety of diseases in the endoderm during development.

References

- Aamar, E. and Dawid, I. B. (2010) 'Sox17 and chordin are required for formation of Kupffer's vesicle and left-right asymmetry determination in zebrafish', *Developmental Dynamics*. NIH Public Access, 239(11), pp. 2980–2988. doi: 10.1002/dvdy.22431.
- Abdulghani, M., Jain, A. and Tuteja, G. (2019) 'Genome-wide identification of enhancer elements in the placenta', *Placenta*. W.B. Saunders Ltd, pp. 72–77. doi: 10.1016/j.placenta.2018.09.003.
- Ackermann, A. M. *et al.* (2016) 'Integration of ATAC-seq and RNA-seq identifies human alpha cell and beta cell signature genes', *Molecular Metabolism*. Elsevier, 5(3), pp. 233–244. doi: 10.1016/j.molmet.2016.01.002.
- Aday, A. W. *et al.* (2011) 'Identification of cis regulatory features in the embryonic zebrafish genome through large-scale profiling of H3K4me1 and H3K4me3 binding sites', *Developmental Biology*, 357(2), pp. 450–462. doi: 10.1016/j.ydbio.2011.03.007.
- Adli, M. and Bernstein, B. E. (2011) 'Whole-genome chromatin profiling from limited numbers of cells using nano-ChIP-seq', *Nature Protocols*. Nature Publishing Group, 6(10), pp. 1656–1668. doi: 10.1038/nprot.2011.402.
- Ahn, D., You, K. H. and Kim, C. H. (2012) *Evolution of the Tbx6/16 subfamily genes in vertebrates: Insights from zebrafish*, *Molecular Biology and Evolution*. doi: 10.1093/molbev/mss199.
- Aksoy, I. *et al.* (2013) 'Oct4 switches partnering from Sox2 to Sox17 to reinterpret the enhancer code and specify endoderm', *EMBO Journal*. John Wiley & Sons, Ltd, 32(7), pp. 938–953. doi: 10.1038/emboj.2013.31.
- Alexander, J. *et al.* (1999) *casanova Plays an early and essential role in endoderm formation in zebrafish*, *Developmental Biology*. doi: 10.1006/dbio.1999.9441.
- Alexander, J. and Stainier, D. Y. R. (1999) *A molecular pathway leading to endoderm formation in zebrafish*, *Current Biology*. doi: 10.1016/S0960-9822(00)80016-0.
- Amacher, S. L. *et al.* (2002) *The zebrafish T-box genes no tail and spadetail are required for development of trunk and tail mesoderm and medial floor plate*, *Development*. doi: 10.1242/dev.129.14.3311.
- Amack, J. D., Wang, X. and Yost, H. J. (2007) 'Two T-box genes play independent and cooperative roles to regulate morphogenesis of ciliated Kupffer's vesicle in zebrafish', *Developmental Biology*. Academic Press Inc., 310(2), pp. 196–210. doi: 10.1016/j.ydbio.2007.05.039.
- Amack, J. D. and Yost, H. J. (2004) 'The T box transcription factor no tail in ciliated cells

controls zebrafish left-right asymmetry', *Current Biology*. Elsevier, 14(8), pp. 685–690.

Anbanandam, A. *et al.* (2006) 'Insights into transcription enhancer factor 1 (TEF-1) activity from the solution structure of the TEA domain', *Proceedings of the National Academy of Sciences of the United States of America*. National Academy of Sciences, 103(46), pp. 17225–17230. doi: 10.1073/pnas.0607171103.

Anders, S. (2010) *Babraham Bioinformatics - FastQC A Quality Control tool for High Throughput Sequence Data*, *Soil*. Available at: <https://www.bioinformatics.babraham.ac.uk/projects/fastqc/> (Accessed: 19 October 2021).

Anders, S., Pyl, P. T. and Huber, W. (2015) 'HTSeq-A Python framework to work with high-throughput sequencing data', *Bioinformatics*. Oxford University Press, 31(2), pp. 166–169. doi: 10.1093/bioinformatics/btu638.

Andersson, R. (2015) 'Promoter or enhancer, what's the difference? Deconstruction of established distinctions and presentation of a unifying model', *BioEssays*. John Wiley and Sons Inc., 37(3), pp. 314–323. doi: 10.1002/bies.201400162.

Ang, S. L. *et al.* (1993) *The formation and maintenance of the definitive endoderm lineage in the mouse: Involvement of HNF3/forkhead proteins*, *Development*. doi: 10.1242/dev.119.4.1301.

Ang, S. L. and Rossant, J. (1994) 'HNF-3 β is essential for node and notochord formation in mouse development', *Cell*. Elsevier, 78(4), pp. 561–574. doi: 10.1016/0092-8674(94)90522-3.

Ángel, V.-V. *et al.* (2017) 'HMGB Proteins from Yeast to Human. Gene Regulation, DNA Repair and Beyond', in *Old Yeasts - New Questions*. InTech. doi: 10.5772/intechopen.70126.

Antoshechkin, I. and Sternberg, P. W. (2007) 'The versatile worm: Genetic and genomic resources for *Caenorhabditis elegans* research', *Nature Reviews Genetics*. Nature Publishing Group, pp. 518–532. doi: 10.1038/nrg2105.

Aoki, T. O. *et al.* (2002) 'Molecular integration of casanova in the Nodal signalling pathway controlling endoderm formation', *Development*, 129(2), pp. 275–286. Available at: <https://dev.biologists.org/content/129/2/275.short> (Accessed: 12 January 2021).

Aranguren, X. L. *et al.* (2011) 'Transcription factor COUP-TFII is indispensable for venous and lymphatic development in zebrafish and *Xenopus laevis*', *Biochemical and Biophysical Research Communications*. Academic Press, 410(1), pp. 121–126. doi: 10.1016/j.bbrc.2011.05.117.

Argenton, F. *et al.* (2004) 'Ectopic expression and knockdown of a zebrafish sox21 reveal its role as a transcriptional repressor in early development', *Mechanisms of Development*, 121(2), pp. 131–142. doi: 10.1016/j.mod.2004.01.001.

Arnold, S. J. *et al.* (2008) 'Pivotal roles for eomesodermin during axis formation, epithelium-to-mesenchyme transition and endoderm specification in the mouse', *Development*. The

Company of Biologists Ltd, 135(3), pp. 501–511. doi: 10.1242/dev.014357.

Aronson, B. E., Stapleton, K. A. and Krasinski, S. D. (2014) 'Role of GATA factors in development, differentiation, and homeostasis of the small intestinal epithelium', *American Journal of Physiology - Gastrointestinal and Liver Physiology*. Am J Physiol Gastrointest Liver Physiol, 306(6). doi: 10.1152/ajpgi.00119.2013.

Babu, D. A., Deering, T. G. and Mirmira, R. G. (2007) 'A feat of metabolic proportions: Pdx1 orchestrates islet development and function in the maintenance of glucose homeostasis', *Molecular Genetics and Metabolism*, 92(1–2), pp. 43–55. doi: 10.1016/j.ymgme.2007.06.008.

Bahar Halpern, K., Vana, T. and Walker, M. D. (2014) 'Paradoxical role of DNA methylation in activation of FoxA2 gene expression during endoderm development', *Journal of Biological Chemistry*. Elsevier, 289(34), pp. 23882–23892. doi: 10.1074/jbc.M114.573469.

Bai, L. and Morozov, A. V. (2010) 'Gene regulation by nucleosome positioning', *Trends in Genetics*, 26(11), pp. 476–483. doi: 10.1016/j.tig.2010.08.003.

Banerjee, K. K. *et al.* (2018) 'Enhancer, transcriptional, and cell fate plasticity precedes intestinal determination during endoderm development', *Genes and Development*. Cold Spring Harbor Laboratory Press, 32(21–22), pp. 1430–1442. doi: 10.1101/gad.318832.118.

Bellipanni, G. *et al.* (2006) 'Essential and opposing roles of zebrafish β -catenins in the formation of dorsal axial structures and neurectoderm', *Development*. The Company of Biologists Ltd, 133(7), pp. 1299–1309. doi: 10.1242/dev.02295.

Bennett, J. T. *et al.* (2007) 'Nodal signaling activates differentiation genes during zebrafish gastrulation', *Developmental Biology*, 304(2), pp. 525–540. doi: 10.1016/j.ydbio.2007.01.012.

Binot, A. C. *et al.* (2010) 'Nkx6.1 and nkx6.2 regulate α - and β -cell formation in zebrafish by acting on pancreatic endocrine progenitor cells', *Developmental Biology*. Academic Press Inc., 340(2), pp. 397–407. doi: 10.1016/j.ydbio.2010.01.025.

Birnhuber, A. *et al.* (2019) 'Transcription factor Fra-2 and its emerging role in matrix deposition, proliferation and inflammation in chronic lung diseases', *Cellular Signalling*. Pergamon, 64, p. 109408. doi: 10.1016/j.cellsig.2019.109408.

Bjornson *et al.* (2005) 'Eomesodermin is a localized maternal determinant required for endoderm induction in zebrafish', *Developmental Cell*, 9(4), pp. 523–533. doi: 10.1016/j.devcel.2005.08.010.

Blanco, A. M., Bertucci, J. I. and Unniappan, S. (2020) 'FGF21 Mimics a Fasting-Induced Metabolic State and Increases Appetite in Zebrafish', *Scientific Reports*. Nature Research, 10(1). doi: 10.1038/s41598-020-63726-w.

Bogdanović, O. *et al.* (2012) 'Dynamics of enhancer chromatin signatures mark the transition from pluripotency to cell specification during embryogenesis', *Genome Research*. Cold Spring Harbor Laboratory Press, 22(10), pp. 2043–2053.

Bolger, A. M., Lohse, M. and Usadel, B. (2014) 'Trimmomatic: A flexible trimmer for Illumina

sequence data', *Bioinformatics*. Oxford University Press, 30(15), pp. 2114–2120. doi: 10.1093/bioinformatics/btu170.

Bordeira-Carriço, R. *et al.* (2020) 'Cis-regulatory similarities in the zebrafish and human pancreas uncover potential disease-related enhancers', *bioRxiv*. Cold Spring Harbor Laboratory, p. 2020.04.27.064220. doi: 10.1101/2020.04.27.064220.

van Boxtel, A. L. *et al.* (2015) 'A Temporal Window for Signal Activation Dictates the Dimensions of a Nodal Signaling Domain', *Developmental Cell*. Cell Press, 35(2), pp. 175–185. doi: 10.1016/j.devcel.2015.09.014.

Boyle, A. P. *et al.* (2008) 'High-Resolution Mapping and Characterization of Open Chromatin across the Genome', *Cell*, 132(2), pp. 311–322. doi: 10.1016/j.cell.2007.12.014.

Bozek, M. and Gompel, N. (2020) 'Developmental Transcriptional Enhancers: A Subtle Interplay between Accessibility and Activity: Considering Quantitative Accessibility Changes between Different Regulatory States of an Enhancer Deconvolutes the Complex Relationship between Accessibility and Activity', *BioEssays*. John Wiley & Sons, Ltd, 42(4), p. 1900188. doi: 10.1002/bies.201900188.

Brasset, E. and Vaury, C. (2005) 'Insulators are fundamental components of the eukaryotic genomes', *Heredity*. Nature Publishing Group, 94(6), pp. 571–576. doi: 10.1038/sj.hdy.6800669.

Brennan, J. *et al.* (2001) 'Nodal signalling in the epiblast patterns the early mouse embryo', *Nature*. Nature Publishing Group, 411(6840), pp. 965–969. doi: 10.1038/35082103.

Bruce, A. E. E. *et al.* (2003) 'The maternally expressed zebrafish T-box gene eomesodermin regulates organizer formation', *Development*, 130(22), pp. 5503–5517. doi: 10.1242/dev.00763.

Bruce, A. E. E. *et al.* (2005) 'T-box gene eomesodermin and the homeobox-containing mix/bix gene *mtx2* regulate epiboly movements in the zebrafish', *Developmental Dynamics*. Wiley-Liss Inc., 233(1), pp. 105–114. doi: 10.1002/dvdy.20305.

Buenrostro, J. D. *et al.* (2013) 'Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position', *Nature Methods*, 10(12), pp. 1213–1218. doi: 10.1038/nmeth.2688.

Buenrostro, J. D. *et al.* (2015) 'ATAC-seq: A method for assaying chromatin accessibility genome-wide', *Current Protocols in Molecular Biology*, 2015, pp. 21.29.1–21.29.9.

Burtscher, I. and Lickert, H. (2009) 'Foxa2 regulates polarity and epithelialization in the endoderm germ layer of the mouse embryo', *Development*, 136(6), pp. 1029–1038. doi: 10.1242/dev.028415.

Butler, J. E. F. and Kadonaga, J. T. (2002) 'The RNA polymerase II core promoter: A key component in the regulation of gene expression', *Genes and Development*. Cold Spring Harbor Laboratory Press, pp. 2583–2592. doi: 10.1101/gad.1026202.

Cai, Y. *et al.* (2021) 'H3K27me3-rich genomic regions can function as silencers to repress gene expression via chromatin interactions', *Nature Communications*. Nature Publishing Group, 12(1), pp. 1–22. doi: 10.1038/s41467-021-20940-y.

Calo, E. and Wysocka, J. (2013) 'Modification of Enhancer Chromatin: What, How, and Why?', *Molecular Cell*, 49(5), pp. 825–837. doi: 10.1016/j.molcel.2013.01.038.

Carvalho, L. and Heisenberg, C. P. (2010) 'The yolk syncytial layer in early zebrafish development', *Trends in Cell Biology*, pp. 586–592. doi: 10.1016/j.tcb.2010.06.009.

Cebola, I. *et al.* (2015) 'TEAD and YAP regulate the enhancer network of human embryonic pancreatic progenitors', *Nature Cell Biology*. Nature Publishing Group, 17(5), pp. 615–626. doi: 10.1038/ncb3160.

Chan, T. M. *et al.* (2009) 'Functional analysis of the evolutionarily conserved cis-regulatory elements on the sox17 gene in zebrafish', *Developmental Biology*. Academic Press, 326(2), pp. 456–470. doi: 10.1016/j.ydbio.2008.11.010.

Chapman, S. C. *et al.* (2007) 'Specification of germ layer identity in the chick gastrula', *BMC Developmental Biology*. BMC Dev Biol, 7. doi: 10.1186/1471-213X-7-91.

Chatterjee, S., Khunti, K. and Davies, M. J. (2017) 'Type 2 diabetes', *The Lancet*, 389(10085), pp. 2239–2251. doi: 10.1016/S0140-6736(17)30058-2.

Chen *et al.* (1996) 'Mutations affecting the cardiovascular system and other internal organs in zebrafish.', *Development (Cambridge, England)*, 123(1), pp. 293–302.

Chen, S. and Kimelman, D. (2000) 'The role of the yolk syncytial layer in germ layer patterning in zebrafish.', *Development*, 127(21), pp. 4681–9. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/11023870> (Accessed: 25 August 2017).

Chen, X. *et al.* (1997) 'Smad4 and FAST-1 in the assembly of activin-responsive factor', *Nature*. Nature Publishing Group, 389(6646), pp. 85–89. doi: 10.1038/38008.

Chen, X., Rubock, M. J. and Whitman, M. (1996) 'A transcriptional partner for mad proteins in TGF- β signalling', *Nature*, 383(6602), pp. 691–696. doi: 10.1038/383691a0.

Chen, Y. and Schler, A. F. (2001) 'The zebrafish nodal signal Squint functions as a morphogen', *Nature*. Nature Publishing Group, 411(6837), pp. 607–610. doi: 10.1038/35079121.

Chereji, R. V., Bryson, T. D. and Henikoff, S. (2019) 'Quantitative MNase-seq accurately maps nucleosome occupancy levels', *Genome biology*. NLM (Medline), 20(1), p. 198. doi: 10.1186/s13059-019-1815-z.

Chereji, R. V. and Morozov, A. V. (2015) 'Functional roles of nucleosome stability and dynamics', *Briefings in Functional Genomics*. Oxford University Press, 14(1), pp. 50–60. doi: 10.1093/bfgp/elu038.

Choudhry, P. and Trede, N. S. (2013) 'DiGeorge Syndrome Gene *tbx1* Functions through *wnt11r* to Regulate Heart Looping and Differentiation', *PLoS ONE*. Edited by S.-P. L. Hwang.

Public Library of Science, 8(3), p. e58145. doi: 10.1371/journal.pone.0058145.

Chung, M. I. S. *et al.* (2011) 'Characterization of sry-related hmg box group f genes in zebrafish hematopoiesis', *Experimental Hematology*, 39(10), pp. 986–998.e5. doi: 10.1016/j.exphem.2011.06.010.

Chung, W. S., Shin, C. H. and Stainier, D. Y. R. (2008) 'Bmp2 Signaling Regulates the Hepatic versus Pancreatic Fate Decision', *Developmental Cell*, 15(5), pp. 738–748. doi: 10.1016/j.devcel.2008.08.019.

Cirillo, L. A. *et al.* (1998) 'Binding of the winged-helix transcription factor HNF3 to a linker histone site on the nucleosome', *EMBO Journal*, 17(1), pp. 244–254. doi: 10.1093/emboj/17.1.244.

Clotman, F. *et al.* (2002) *The onecut transcription factor HNF6 is required for normal development of the biliary tract*, *Development*. Available at: <https://dev.biologists.org/content/develop/129/8/1819.full.pdf> (Accessed: 23 January 2021).

Cole, M. F. *et al.* (2008) 'Tcf3 is an integral component of the core regulatory circuitry of embryonic stem cells', *Genes and Development*. Cold Spring Harbor Laboratory Press, 22(6), pp. 746–755. doi: 10.1101/gad.1642408.

Cooper, M. S. and D'Amico, L. A. (1996) 'A cluster of noninvoluting endocytic cells at the margin of the zebrafish blastoderm marks the site of embryonic shield formation', *Developmental Biology*, 180(1), pp. 184–198. doi: 10.1006/dbio.1996.0294.

Coppola, C. J., Ramaker, R. C. and Mendenhall, E. M. (2016) 'Identification and function of enhancers in the human genome', *Human Molecular Genetics*. Oxford University Press, pp. R190–R197. doi: 10.1093/hmg/ddw216.

Corces, M. R. *et al.* (2017) 'An improved ATAC-seq protocol reduces background and enables interrogation of frozen tissues', *Nature Methods*. Nature Publishing Group, 14(10), pp. 959–962. doi: 10.1038/nmeth.4396.

Cordero, D. R. *et al.* (2011) 'Cranial neural crest cells on the move: Their roles in craniofacial development', *American Journal of Medical Genetics, Part A*. NIH Public Access, 155(2), pp. 270–279. doi: 10.1002/ajmg.a.33702.

Crawford, G. E. *et al.* (2006) 'Genome-wide mapping of DNase hypersensitive sites using massively parallel signature sequencing (MPSS)', *Genome Research*, 16(1), pp. 123–131. doi: 10.1101/gr.4074106.

Creyghton, M. P. *et al.* (2010) 'Histone H3K27ac separates active from poised enhancers and predicts developmental state', *Proceedings of the National Academy of Sciences of the United States of America*, 107(50), pp. 21931–21936. doi: 10.1073/pnas.1016071107.

D'Amour, K. A. *et al.* (2005) 'Efficient differentiation of human embryonic stem cells to definitive endoderm', *Nature Biotechnology*, 23(12), pp. 1534–1541. doi: 10.1038/nbt1163.

Dal-Pra, S., Thisse, C. and Thisse, B. (2011) 'FoxA transcription factors are essential for the

development of dorsal axial structures', *Developmental Biology*. Academic Press, 350(2), pp. 484–495. doi: 10.1016/j.ydbio.2010.12.018.

Dalgin, G. and Prince, V. E. (2021) 'Midline morphogenesis of zebrafish foregut endoderm is dependent on Hoxb5b', *Developmental Biology*. Elsevier Inc., 471, pp. 1–9. doi: 10.1016/j.ydbio.2020.12.001.

David, Nicolas B *et al.* (2002) 'Requirement for endoderm and FGF3 in ventral head skeleton formation.', *Development (Cambridge, England)*, 129(19), pp. 4457–4468.

David, Nicolas B. *et al.* (2002) *Requirement for endoderm and FGF3 in ventral head skeleton formation*, *Development*. doi: 10.1242/dev.129.19.4457.

David, N. B. and Rosa, F. M. (2001) *Cell autonomous commitment to an endodermal fate and behaviour by activation of Nodal signalling*, *Development*. doi: 10.1242/dev.128.20.3937.

Dawid, I. B. (2004) 'Developmental biology of zebrafish', *Annals of the New York Academy of Sciences*, 1038(1), pp. 88–93. doi: 10.1196/annals.1315.015.

Dean, A., Larson, D. R. and Sartorelli, V. (2021) 'Enhancers, gene regulation, and genome organization', *Genes and Development*. Cold Spring Harbor Laboratory Press, 35(7), pp. 427–433. doi: 10.1101/gad.348372.121.

Decker, K. *et al.* (2006) 'Gata6 is an important regulator of mouse pancreas development', *Developmental Biology*. Academic Press Inc., 298(2), pp. 415–429. doi: 10.1016/j.ydbio.2006.06.046.

Dee, C. T. *et al.* (2008) 'Sox3 regulates both neural fate and differentiation in the zebrafish ectoderm', *Developmental Biology*. Academic Press, 320(1), pp. 289–301. doi: 10.1016/j.ydbio.2008.05.542.

Defossez, P. A. *et al.* (2005) 'The human enhancer blocker CTC-binding factor interacts with the transcription factor Kaiso', *Journal of Biological Chemistry*, 280(52), pp. 43017–43023. doi: 10.1074/jbc.M510802200.

Defronzo, R. A. (2009) 'From the triumvirate to the ominous octet: A new paradigm for the treatment of type 2 diabetes mellitus', in *Diabetes*. American Diabetes Association, pp. 773–795. doi: 10.2337/db09-9028.

DeLaForest, A. *et al.* (2011) 'HNF4A is essential for specification of hepatic progenitors from human pluripotent stem cells', *Development*. Oxford University Press for The Company of Biologists Limited, 138(19), pp. 4143–4153. doi: 10.1242/dev.062547.

Delgado-Olguín, P. *et al.* (2011) 'CTCF promotes muscle differentiation by modulating the activity of myogenic regulatory factors', *Journal of Biological Chemistry*. Elsevier, 286(14), pp. 12483–12494. doi: 10.1074/jbc.M110.164574.

Despang, A. *et al.* (2019) 'Functional dissection of the Sox9–Kcnj2 locus identifies nonessential and instructive roles of TAD architecture', *Nature Genetics*, 51(8), pp. 1263–1271. doi: 10.1038/s41588-019-0466-z.

- Devakanmalai, G. S., Zumrut, H. E. and Özbudak, E. M. (2013) 'Cited3 activates Mef2c to control muscle cell differentiation and survival', *Biology Open*. The Company of Biologists, 2(5), pp. 505–514. doi: 10.1242/bio.20132550.
- Devoy, A. *et al.* (2012) 'Genomically humanized mice: Technologies and promises', *Nature Reviews Genetics*, 13(1), pp. 14–20. doi: 10.1038/nrg3116.
- Dickel, D. E. *et al.* (2018) 'Ultraconserved Enhancers Are Required for Normal Development', *Cell*. Cell Press, 172(3), pp. 491-499.e15. doi: 10.1016/j.cell.2017.12.017.
- Dickmeis, T. *et al.* (2001) 'A crucial component of the endoderm formation pathway, casanova, is encoded by a novel sox-related gene', *Genes and Development*. Cold Spring Harbor Laboratory Press, 15(12), pp. 1487–1492.
- Dirksen, M. -L and Jamrich, M. (1995) 'Differential expression of fork head genes during early Xenopus and zebrafish development', *Developmental Genetics*, 17(2), pp. 107–116. doi: 10.1002/dvg.1020170203.
- Dobin, A. *et al.* (2013) 'STAR: Ultrafast universal RNA-seq aligner', *Bioinformatics*. Oxford Academic, 29(1), pp. 15–21. doi: 10.1093/bioinformatics/bts635.
- Dobrovolskaia-Zavadskaia, N. (1927) 'Sur la mortification spontanee de la queue chez la souris nouveau-nee et sur l'existence d'un caractere hereditaire "non viable"', *CR Soc Biol*, 97, pp. 114–116. Available at: <https://ci.nii.ac.jp/naid/10009426192/> (Accessed: 11 September 2021).
- Dokmanovic-Chouinard, M. *et al.* (2008) 'Positional cloning of "Lisch-like", a candidate modifier of susceptibility to type 2 diabetes in mice', *PLoS Genetics*. Public Library of Science, 4(7), p. e1000137. doi: 10.1371/journal.pgen.1000137.
- Dong, C., Wilhelm, D. and Koopman, P. (2004) 'Sox genes and cancer', *Cytogenetic and Genome Research*. IntechOpen, 105(2–4), pp. 442–447. doi: 10.1159/000078217.
- Dooley, K. A., Davidson, A. J. and Zon, L. I. (2005) 'Zebrafish scl functions independently in hematopoietic and endothelial development', *Developmental Biology*. Academic Press Inc., 277(2), pp. 522–536. doi: 10.1016/j.ydbio.2004.09.004.
- Dougan, S. T. *et al.* (2003) 'The role of the zebrafish nodal-related genes squint and cyclops in patterning of mesendoderm.', *Development (Cambridge, England)*. Development, 130(9), pp. 1837–51. doi: 10.1242/dev.00400.
- Downen, J. M. *et al.* (2014) 'Control of cell identity genes occurs in insulated neighborhoods in mammalian chromosomes', *Cell*. Cell Press, 159(2), pp. 374–387. doi: 10.1016/j.cell.2014.09.030.
- Driever, W. *et al.* (1996) *A genetic screen for mutations affecting embryogenesis in zebrafish*, *Development*. doi: 10.5167/uzh-215.
- Du, S. *et al.* (2012) 'Differential regulation of epiboly initiation and progression by zebrafish Eomesodermin A', *Developmental Biology*. Academic Press, 362(1), pp. 11–23. doi:

10.1016/j.ydbio.2011.10.036.

Dufort, D. *et al.* (1998) *The transcription factor HNF3 β is required in visceral endoderm for normal primitive streak morphogenesis*, *Development*. doi: 10.1242/dev.125.16.3015.

Erokhin, M. *et al.* (2015) 'Eukaryotic enhancers: Common features, regulation, and participation in diseases', *Cellular and Molecular Life Sciences*. *Cell Mol Life Sci*, 72(12), pp. 2361–2375. doi: 10.1007/s00018-015-1871-9.

Erter, C. E., Solnica-Krezel, L. and Wright, C. V. E. (1998) 'Zebrafish nodal-related 2 encodes an early mesendodermal inducer signaling from the extraembryonic yolk syncytial layer', *Developmental Biology*, 204(2), pp. 361–372. doi: 10.1006/dbio.1998.9097.

Essner, J. J. *et al.* (2005) 'Kupffer's vesicle is a ciliated organ of asymmetry in the zebrafish embryo that initiates left-right development of the brain, heart and gut', *Development*, 132(6), pp. 1247–1260. doi: 10.1242/dev.01663.

Eufrásio, A. *et al.* (2020) 'In vivo reporter assays uncover changes in enhancer activity caused by type 2 diabetes-associated single nucleotide polymorphisms', *Diabetes*. American Diabetes Association Inc., 69(12), pp. 2794–2805. doi: 10.2337/db19-1049.

Fabian, P. *et al.* (2020) 'Lineage analysis reveals an endodermal contribution to the vertebrate pituitary', *Science*. American Association for the Advancement of Science, 370(6515), pp. 463–467. doi: 10.1126/science.aba4767.

Faial, T. *et al.* (2015) 'Brachyury and SMAD signalling collaboratively orchestrate distinct mesoderm and endoderm gene regulatory networks in differentiating human embryonic stem cells', *Development (Cambridge)*, 142(12), pp. 2121–2135. doi: 10.1242/dev.117838.

Fan, X. *et al.* (2007) 'Nodal signals mediate interactions between the extra-embryonic and embryonic tissues in zebrafish', *Developmental Biology*. Academic Press Inc., 310(2), pp. 363–378. doi: 10.1016/j.ydbio.2007.08.008.

Farooq, M. *et al.* (2008) 'Histone deacetylase 3 (hdac3) is specifically required for liver development in zebrafish', *Developmental Biology*, 317(1), pp. 336–353. doi: 10.1016/j.ydbio.2008.02.034.

Feldman, B. *et al.* (1998) 'Zebrafish organizer development and germ-layer formation require nodal-related signals', *Nature*. Nature Publishing Group, 395(6698), pp. 181–185. doi: 10.1038/26013.

Ferg, M. *et al.* (2014) 'Gene transcription in the zebrafish embryo: Regulators and networks', *Briefings in Functional Genomics*. Oxford University Press, 13(2), pp. 131–143.

Figiel, D. M., Elsayed, R. and Nelson, A. C. (2021) 'Investigating the molecular guts of endoderm formation using zebrafish', *Briefings in Functional Genomics*. doi: 10.1093/bfpg/elab013.

Fillatre, J. *et al.* (2019) 'TEADS, Yap, Taz, vgl14s transcription factors control the establishment of left-right asymmetry in zebrafish', *eLife*, 8. doi: 10.7554/eLife.45241.

Fish, J. E. *et al.* (2008) 'miR-126 Regulates Angiogenic Signaling and Vascular Integrity', *Developmental Cell*. Elsevier, 15(2), pp. 272–284.

Flannick, J. *et al.* (2019) 'Exome sequencing of 20,791 cases of type 2 diabetes and 24,440 controls', *Nature*. Nature, 570(7759), pp. 71–76. doi: 10.1038/s41586-019-1231-2.

Foley, A. C. and Stern, C. D. (2001) 'Evolution of vertebrate forebrain development: how many different mechanisms?', *Journal of Anatomy*. Wiley-Blackwell, 199(1), pp. 35–52. doi: 10.1046/j.1469-7580.2001.19910035.x.

Friedman, J. R. and Kaestner, K. H. (2006) 'The Foxa family of transcription factors in development and metabolism', *Cellular and Molecular Life Sciences*. Cell Mol Life Sci, 63(19–20), pp. 2317–2328. doi: 10.1007/s00018-006-6095-6.

Friedman, R. Z. *et al.* (2021) 'Information content differentiates enhancers from silencers in mouse photoreceptors', *eLife*. eLife Sciences Publications Ltd, 10. doi: 10.7554/eLife.67403.

Fujimoto, K. *et al.* (2009) 'Autophagy regulates pancreatic beta cell death in response to Pdx1 deficiency and nutrient deprivation', *Journal of Biological Chemistry*, 284(40), pp. 27664–27673. doi: 10.1074/jbc.M109.041616.

Fukui, H. *et al.* (2014) 'S1P-Yap1 signaling regulates endoderm formation required for cardiac precursor cell migration in zebrafish', *Developmental Cell*, 31(1), pp. 128–136. doi: 10.1016/j.devcel.2014.08.014.

Furey, T. S. (2012) 'ChIP-seq and beyond: New and improved methodologies to detect and characterize protein-DNA interactions', *Nature Reviews Genetics*. Nature Publishing Group, pp. 840–852. doi: 10.1038/nrg3306.

Gagnon, J. A., Obbad, K. and Schier, A. F. (2018) 'The primary role of Zebrafish nanog is in extra-embryonic tissue', *Development (Cambridge)*. Oxford University Press for The Company of Biologists Limited, 145(1). doi: 10.1242/dev.147793.

Gao, C. *et al.* (2019) 'Zebrafish Hhex-null mutant develops an intrahepatic intestinal tube due to de-repression of cdx1b and pdx1', *Journal of Molecular Cell Biology*. Oxford Academic, 11(6), pp. 448–462. doi: 10.1093/jmcb/mjy068.

Gao, N. *et al.* (2008) 'Dynamic regulation of Pdx1 enhancers by Foxa1 and Foxa2 is essential for pancreas development', *Genes and Development*. Cold Spring Harbor Laboratory Press, 22(24), pp. 3435–3448. doi: 10.1101/gad.1752608.

Garnett, A. T. *et al.* (2009) 'Identification of direct T-box target genes in the developing zebrafish mesoderm', *Development*, 136(5), pp. 749–760. doi: 10.1242/dev.024703.

Gentsch, G. E. *et al.* (2013) 'InVivo T-Box Transcription Factor Profiling Reveals Joint Regulation of Embryonic Neuromesodermal Bipotency', *Cell Reports*, 4(6), pp. 1185–1196. doi: 10.1016/j.celrep.2013.08.012.

Germain *et al.* (2000) 'Homeodomain and winged-helix transcription factors recruit activated Smads to distinct promoter elements via a common Smad interaction motif', *Genes and*

Development. Cold Spring Harbor Laboratory Press, 14(4), pp. 435–451. doi: 10.1101/gad.14.4.435.

Gerstein, M. B. *et al.* (2010) 'Integrative analysis of the *Caenorhabditis elegans* genome by the modENCODE project', *Science*. American Association for the Advancement of Science, 330(6012), pp. 1775–1787. doi: 10.1126/science.1196914.

Gerstein, M. B. *et al.* (2012) 'Architecture of the human regulatory network derived from ENCODE data', *Nature*. Nature Publishing Group, 489(7414), pp. 91–100. doi: 10.1038/nature11245.

Ghiasvand, N. M. *et al.* (2011) 'Deletion of a remote enhancer near ATOH7 disrupts retinal neurogenesis, causing NCRNA disease', *Nature Neuroscience*, 14(5), pp. 578–588. doi: 10.1038/nn.2798.

Ghosh, T. K., Brook, J. D. and Wilsdon, A. (2017) 'T-Box Genes in Human Development and Disease', *Current Topics in Developmental Biology*. Academic Press, 122, pp. 383–415. doi: 10.1016/bs.ctdb.2016.08.006.

Gilfillan, G. D. *et al.* (2012) 'Limitations and possibilities of low cell number ChIP-seq', *BMC Genomics*. BioMed Central, 13(1), p. 645.

Gimelli, S. *et al.* (2010) 'Mutations in SOX17 are associated with congenital anomalies of the kidney and the urinary tract', *Human Mutation*. Wiley-Blackwell, 31(12), pp. 1352–1359. doi: 10.1002/humu.21378.

Giraldez, A. J. *et al.* (2006) 'Zebrafish MiR-430 promotes deadenylation and clearance of maternal mRNAs', *Science*. American Association for the Advancement of Science, 312(5770), pp. 75–79. doi: 10.1126/science.1122689.

Goering, L. M. *et al.* (2003) 'An interacting network of T-box genes directs gene expression and fate in the zebrafish mesoderm', *Proceedings of the National Academy of Sciences of the United States of America*, 100(16), pp. 9410–9415. doi: 10.1073/pnas.1633548100.

Goessling, W. *et al.* (2008) 'APC mutant zebrafish uncover a changing temporal requirement for wnt signaling in liver development', *Developmental Biology*. Dev Biol, 320(1), pp. 161–174. doi: 10.1016/j.ydbio.2008.05.526.

Gokey, J. J. *et al.* (2016) 'Kupffer's vesicle size threshold for robust left-right patterning of the zebrafish embryo', *Developmental Dynamics*. John Wiley and Sons Inc., 245(1), pp. 22–33. doi: 10.1002/dvdy.24355.

Golson, M. L. and Kaestner, K. H. (2016) 'Fox transcription factors: From development to disease', *Development (Cambridge)*. Company of Biologists, 143(24), pp. 4558–4570. doi: 10.1242/dev.112672.

Le Good, J. A. *et al.* (2005) 'Nodal stability determines signaling range', *Current Biology*. Cell Press, 15(1), pp. 31–36. doi: 10.1016/j.cub.2004.12.062.

Gore, A. V. *et al.* (2005) 'The zebrafish dorsal axis is apparent at the four-cell stage', *Nature*.

Nature Publishing Group, 438(7070), pp. 1030–1035. doi: 10.1038/nature04184.

Gore, A. V. and Sampath, K. (2002) 'Localization of transcripts of the Zebrafish morphogen Squint is dependent on egg activation and the microtubule cytoskeleton', *Mechanisms of Development*, 112(1–2), pp. 153–156. doi: 10.1016/S0925-4773(01)00622-0.

Gorkin, D. U., Leung, D. and Ren, B. (2014) 'The 3D genome in transcriptional regulation and pluripotency', *Cell Stem Cell*. Cell Press, pp. 762–775. doi: 10.1016/j.stem.2014.05.017.

Göttgens, B. *et al.* (2002) 'Transcriptional regulation of the stem cell leukemia gene (SCL) - Comparative analysis of five vertebrate SCL loci', *Genome Research*. Cold Spring Harbor Laboratory Press, 12(5), pp. 749–759. doi: 10.1101/gr.45502.

Grant, C. E., Bailey, T. L. and Noble, W. S. (2011) 'FIMO: Scanning for occurrences of a given motif', *Bioinformatics*. Bioinformatics, 27(7), pp. 1017–1018. doi: 10.1093/bioinformatics/btr064.

Griffin, K. J. P. *et al.* (1998) *Molecular identification of spadetail: Regulation of zebrafish trunk and tail mesoderm formation by T-box genes*, *Development*. Available at: <https://dev.biologists.org/content/125/17/3379.short> (Accessed: 8 March 2021).

Gritsman, K. *et al.* (1999) 'The EGF-CFC protein one-eyed pinhead is essential for nodal signaling', *Cell*. Elsevier, 97(1), pp. 121–132. doi: 10.1016/S0092-8674(00)80720-5.

Hadzhiev, Y. *et al.* (2019) 'A cell cycle-coordinated Polymerase II transcription compartment encompasses gene expression before global genome activation', *Nature Communications*. Nature Publishing Group, 10(1). doi: 10.1038/s41467-019-08487-5.

Haffter, P. *et al.* (1996) 'The identification of genes with unique and essential functions in the development of the zebrafish, *Danio rerio*', *Development*, 123(1), pp. 1–36.

Hagos, E. G. and Dougan, S. T. (2007) 'Time-dependent patterning of the mesoderm and endoderm by Nodal signals in zebrafish', *BMC Developmental Biology*. BioMed Central, 7(1), p. 22. doi: 10.1186/1471-213X-7-22.

Hagos, E. G., Fan, X. and Dougan, S. T. (2007) 'The role of maternal Activin-like signals in zebrafish embryos', *Developmental Biology*, 309(2), pp. 245–258. doi: 10.1016/j.ydbio.2007.07.010.

Hall, C. *et al.* (2006) 'An essential role for zebrafish Fgfr1 during gill cartilage development', *Mechanisms of Development*. Elsevier, 123(12), pp. 925–940. doi: 10.1016/j.mod.2006.08.006.

Halpern, M. E. *et al.* (1993) 'Induction of muscle pioneers and floor plate is distinguished by the zebrafish no tail mutation', *Cell*. Elsevier, 75(1), pp. 99–111. doi: 10.1016/S0092-8674(05)80087-X.

Haramis, A. P. G. *et al.* (2006) 'Adenomatous polyposis coli-deficient zebrafish are susceptible to digestive tract neoplasia', *EMBO Reports*, 7(4), pp. 444–449. doi: 10.1038/sj.embor.7400638.

Harris, M. B., Mostecky, J. and Rothman, P. B. (2005) 'Repression of an interleukin-4-responsive promoter requires cooperative BCL-6 function', *Journal of Biological Chemistry*, 280(13), pp. 13114–13121. doi: 10.1074/jbc.M412649200.

Harvey, K. F., Pflieger, C. M. and Hariharan, I. K. (2003) 'The *Drosophila* Mst ortholog, hippo, restricts growth and cell proliferation and promotes apoptosis', *Cell*. Cell, 114(4), pp. 457–467. doi: 10.1016/S0092-8674(03)00557-9.

Hassoun, R., Püschel, B. and Viebahn, C. (2010) 'Sox17 expression patterns during gastrulation and early neurulation in the rabbit suggest two sources of endoderm formation', *Cells Tissues Organs*. Cells Tissues Organs, 191(2), pp. 68–83. doi: 10.1159/000236044.

Hauptmann, G. and Gerster, T. (1996) 'Complex expression of the *zp-50 pou* gene in the embryonic zebrafish brain is altered by overexpression of sonic hedgehog', *Development*. The Company of Biologists, 122(6), pp. 1769–1780. doi: 10.1242/dev.122.6.1769.

Heintzman, N. D. *et al.* (2009) 'Histone modifications at human enhancers reflect global cell-type-specific gene expression', *Nature*. NIH Public Access, 459(7243), pp. 108–112.

Heinz, S. *et al.* (2010) 'Simple Combinations of Lineage-Determining Transcription Factors Prime cis-Regulatory Elements Required for Macrophage and B Cell Identities', *Molecular Cell*. Cell Press, 38(4), pp. 576–589. doi: 10.1016/j.molcel.2010.05.004.

Heinz, S. *et al.* (2015) 'The selection and function of cell type-specific enhancers', *Nature Reviews Molecular Cell Biology*. Nature Publishing Group, pp. 144–154. doi: 10.1038/nrm3949.

Henry, G. L. and Melton, D. A. (1998) 'Mixer, a homeobox gene required for endoderm development', *Science*. American Association for the Advancement of Science, 281(5373), pp. 91–96. doi: 10.1126/science.281.5373.91.

Heyn, P. *et al.* (2014) 'The earliest transcribed zygotic genes are short, newly evolved, and different across species', *Cell Reports*, 6(2), pp. 285–292. doi: 10.1016/j.celrep.2013.12.030.

Ho, R. K. and Kane, D. A. (1990) *Cell-autonomous action of zebrafish spt-1 mutation in specific mesodermal precursors*, *Nature*. Academic. doi: 10.1038/348728a0.

Holtzinger, A. (2005) 'Gata4 regulates the formation of multiple organs', *Development*. The Company of Biologists Ltd, 132(17), pp. 4005–4014. doi: 10.1242/dev.01978.

Holtzinger, A. and Evans, T. (2005) 'Gata4 regulates the formation of multiple organs', *Development*. The Company of Biologists Ltd, 132(17), pp. 4005–4014. doi: 10.1242/dev.01978.

Hong, S. K. *et al.* (2011) 'Embryonic mesoderm and endoderm induction requires the actions of non-embryonic Nodal-related ligands and Mxtx2', *Development*. Oxford University Press for The Company of Biologists Limited, 138(4), pp. 787–795. doi: 10.1242/dev.058974.

Hore, T. A., Deakin, J. E. and Marshall Graves, J. A. (2008) 'The evolution of epigenetic regulators CTCF and BORIS/CTCF in amniotes', *PLoS Genetics*. Public Library of Science,

4(8), p. 1000169. doi: 10.1371/journal.pgen.1000169.

Hou, L., Srivastava, Y. and Jauch, R. (2017) 'Molecular basis for the genome engagement by Sox proteins', *Seminars in Cell and Developmental Biology*. Academic Press, 63, pp. 2–12. doi: 10.1016/j.semcdb.2016.08.005.

Howe, K. *et al.* (2013) 'The zebrafish reference genome sequence and its relationship to the human genome', *Nature*. Nature Publishing Group, 496(7446), pp. 498–503.

Hozumi, S., Aoki, S. and Kikuchi, Y. (2017) 'Nuclear movement regulated by Non-Smad nodal signaling via JNK is associated with smad signaling during zebrafish endoderm specification', *Development (Cambridge)*. Company of Biologists Ltd, 144(21), pp. 4015–4025. doi: 10.1242/dev.151746.

Hsu, D. K. W. *et al.* (1996) 'Identification of a murine TEF-1-related gene expressed after mitogenic stimulation of quiescent fibroblasts and during myogenic differentiation', *Journal of Biological Chemistry*. Elsevier, 271(23), pp. 13786–13795. doi: 10.1074/jbc.271.23.13786.

Huang, D. W., Sherman, B. T. and Lempicki, R. A. (2009a) 'Bioinformatics enrichment tools: Paths toward the comprehensive functional analysis of large gene lists', *Nucleic Acids Research*. Nucleic Acids Res, 37(1), pp. 1–13. doi: 10.1093/nar/gkn923.

Huang, D. W., Sherman, B. T. and Lempicki, R. A. (2009b) 'Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources', *Nature Protocols*. Nat Protoc, 4(1), pp. 44–57. doi: 10.1038/nprot.2008.211.

Huang, J. H. *et al.* (2018) 'Borders of Cis-Regulatory DNA Sequences Preferentially Harbor the Divergent Transcription Factor Binding Motifs in the Human Genome', *Frontiers in Genetics*. Frontiers, 9, p. 571. doi: 10.3389/fgene.2018.00571.

Hudson, C. *et al.* (1997) 'Xsox17 α and - β mediate endoderm formation in xenopus', *Cell*. Cell Press, 91(3), pp. 397–405. doi: 10.1016/S0092-8674(00)80423-7.

Ibragimov, A. N., Bylino, O. V. and Shidlovskii, Y. V. (2020) 'Molecular Basis of the Function of Transcriptional Enhancers', *Cells*. NLM (Medline), 9(7). doi: 10.3390/cells9071620.

Ishikawa, T. o. *et al.* (2007) 'The zebrafish genome contains two inducible, functional cyclooxygenase-2 genes', *Biochemical and Biophysical Research Communications*. Academic Press, 352(1), pp. 181–187. doi: 10.1016/j.bbrc.2006.11.007.

Iwafuchi-Doi, M. (2019) 'The mechanistic basis for chromatin regulation by pioneer transcription factors', *Wiley interdisciplinary reviews. Systems biology and medicine*. John Wiley & Sons, Ltd, 11(1), p. e1427. doi: 10.1002/wsbm.1427.

Jackson, H. E. *et al.* (2015) 'The role of Sox6 in zebrafish muscle fiber type specification', *Skeletal Muscle*. BioMed Central Ltd., 5(1), pp. 1–17. doi: 10.1186/s13395-014-0026-2.

Jahangiri, L. *et al.* (2016) 'The AP-1 transcription factor component Fosl2 potentiates the rate of myocardial differentiation from the zebrafish second heart field', *Development (Cambridge)*. Company of Biologists Ltd, 143(1), pp. 113–122. doi: 10.1242/dev.126136.

- Jiang, Z. *et al.* (2008) 'Exdpf is a key regulator of exocrine pancreas development controlled by retinoic acid and ptf1a in zebrafish', *PLoS Biology*. Edited by D. Stemple. Public Library of Science, 6(11), pp. 2450–2464. doi: 10.1371/journal.pbio.0060293.
- Jonsson, J. *et al.* (1994) 'Insulin-promoter-factor 1 is required for pancreas development in mice', *Nature*. Nature Publishing Group, 371(6498), pp. 606–609. doi: 10.1038/371606a0.
- Jung, H. M. *et al.* (2017) 'Development of the larval lymphatic system in zebrafish', *Development (Cambridge)*. Company of Biologists Ltd, 144(11), pp. 2070–2081. doi: 10.1242/dev.145755.
- Kaestner, K. H., Hiemisch, H. and Schütz, G. (1998) *Targeted Disruption of the Gene Encoding Hepatocyte Nuclear Factor 3 γ Results in Reduced Transcription of Hepatocyte-Specific Genes*, *Molecular and Cellular Biology*. doi: 10.1128/mcb.18.7.4245.
- Kamachi, Y. and Kondoh, H. (2013) 'Sox proteins: Regulators of cell fate specification and differentiation', *Development (Cambridge)*, 140(20), pp. 4129–4144. doi: 10.1242/dev.091793.
- Kan, N. G., Junghans, D. and Belmonte, J. C. I. (2009) 'Compensatory growth mechanisms regulated by BMP and FGF signaling mediate liver regeneration in zebrafish after partial hepatectomy', *The FASEB Journal*. Wiley, 23(10), pp. 3516–3525. doi: 10.1096/fj.09-131730.
- Kanai-Azuma *et al.* (2002) *Depletion of definitive gut endoderm in Sox17-null mutant mice.*, *Development (Cambridge, England)*.
- Kane, D. A. and Kimmel, C. B. (1993) 'The zebrafish midblastula transition', *Development*, 119(2), pp. 447–456.
- Kane, D. A., Warga, R. M. and Kimmel, C. B. (1992) 'Mitotic domains in the early embryo of the zebrafish', *Nature*, 360(6406), pp. 735–737. doi: 10.1038/360735a0.
- Karnuta, J. M. and Scacheri, P. C. (2018) 'Enhancers: bridging the gap between gene control and human disease', *Human molecular genetics*. Oxford Academic, 27(R2), pp. R219–R227. doi: 10.1093/hmg/ddy167.
- Kaspi, H. *et al.* (2013) 'Brief Report: MiR-290-295 regulate embryonic stem cell differentiation propensities by repressing Pax6', *Stem Cells*. John Wiley & Sons, Ltd, 31(10), pp. 2266–2272. doi: 10.1002/stem.1465.
- Kelly, G. M. and Drysdale, T. A. (2015) 'Retinoic acid and the development of the endoderm', *Journal of Developmental Biology*. Multidisciplinary Digital Publishing Institute, 3(2), pp. 25–56. doi: 10.3390/jdb3020025.
- Kelsh, R. N. (2006) 'Sorting out Sox10 functions in neural crest development', *BioEssays*, 28(8), pp. 788–798. doi: 10.1002/bies.20445.
- Kiecker, C., Bates, T. and Bell, E. (2016) 'Molecular specification of germ layers in vertebrate embryos', *Cellular and Molecular Life Sciences*. Springer, pp. 923–947.
- Kiesow, K. *et al.* (2015) 'Junb controls lymphatic vascular development in zebrafish via MIR-

182', *Scientific Reports*. Nature Publishing Group, 5. doi: 10.1038/srep15007.

Kikuchi, Y. *et al.* (2000) 'The zebrafish bonnie and clyde gene encodes a Mix family homeodomain protein that regulates the generation of endodermal precursors', *Genes and Development*. Cold Spring Harbor Laboratory Press, 14(10), pp. 1279–1289. doi: 10.1101/gad.14.10.1279.

Kikuchi, Y. *et al.* (2001) 'casanova encodes a novel Sox-related protein necessary and sufficient for early endoderm formation in zebrafish', *Genes and Development*, 15(12), pp. 1493–1505. doi: 10.1101/gad.892301.

Kikuchi, Y. *et al.* (2004) 'Notch Signaling Can Regulate Endoderm Formation in Zebrafish', *Developmental Dynamics*. John Wiley & Sons, Ltd, 229(4), pp. 756–762. doi: 10.1002/dvdy.10483.

Kim, B. M. *et al.* (2005) 'The stomach mesenchymal transcription factor barx1 specifies gastric epithelial identity through inhibition of transient Wnt signaling', *Developmental Cell*. Elsevier, 8(4), pp. 611–622. doi: 10.1016/j.devcel.2005.01.015.

Kim, I., Saunders, T. L. and Morrison, S. J. (2007) 'Sox17 Dependence Distinguishes the Transcriptional Regulation of Fetal from Adult Hematopoietic Stem Cells', *Cell*. NIH Public Access, 130(3), pp. 470–483. doi: 10.1016/j.cell.2007.06.011.

Kim, J. *et al.* (2008) 'An Extended Transcriptional Network for Pluripotency of Embryonic Stem Cells', *Cell*. Howard Hughes Medical Institute, 132(6), pp. 1049–1061. doi: 10.1016/j.cell.2008.02.039.

Kim, S., Yu, N. K. and Kaang, B. K. (2015) 'CTCF as a multifunctional protein in genome regulation and gene expression', *Experimental & molecular medicine*. Nature Publishing Group, p. e166. doi: 10.1038/emm.2015.33.

Kimelman, D. (2016) 'Tales of Tails (and Trunks). Forming the Posterior Body in Vertebrate Embryos', *Current Topics in Developmental Biology*. Academic Press, 116, pp. 517–536. doi: 10.1016/bs.ctdb.2015.12.008.

Kimelman, D. *et al.* (2017) 'Regulation of posterior body and epidermal morphogenesis in zebrafish by localized Yap1 and Wwtr1', *eLife*. eLife Sciences Publications Ltd, 6. doi: 10.7554/eLife.31065.

Kimmel, C. B. *et al.* (1989) 'A mutation that changes cell movement and cell fate in the zebrafish embryo', *Nature*, 337(6205), pp. 358–362. doi: 10.1038/337358a0.

Kimmel, C. B. *et al.* (1995) 'Stages of embryonic development of the zebrafish', *Developmental Dynamics*, 203(3), pp. 253–310. doi: 10.1002/aja.1002030302.

Kimmel, C. B., Warga, R. M. and Schilling, T. F. (1990) 'Origin and organization of the zebrafish fate map', *Development*, 108(4), pp. 581–594. doi: 10.1242/dev.108.4.581.

Kimmel, R. A. *et al.* (2011) 'Requirement for Pdx1 in specification of latent endocrine progenitors in zebrafish', *BMC Biology*. BioMed Central, 9(1), pp. 1–15. doi: 10.1186/1741-

7007-9-75.

Kimmel, R. A. *et al.* (2015) 'Diabetic pdx1-mutant zebrafish show conserved responses to nutrient overload and anti-glycemic treatment', *Scientific Reports*. Nature Publishing Group, 5. doi: 10.1038/srep14241.

Kinder, S. J. *et al.* (2001) *The organizer of the mouse gastrula is composed of a dynamic population of progenitor cells for the axial mesoderm*, *Development*.

Kinkel, M. D. and Prince, V. E. (2009) 'On the diabetic menu: Zebrafish as a model for pancreas development and function', *BioEssays*. John Wiley & Sons, Ltd, 31(2), pp. 139–152. doi: 10.1002/bies.200800123.

Kispert, A. and Herrmann, B. G. (1993) 'The Brachyury gene encodes a novel DNA binding protein', *EMBO Journal*, 12(8), pp. 3211–3220. doi: 10.1002/j.1460-2075.1993.tb05990.x.

Klein, C. *et al.* (2011) 'Neuron navigator 3a regulates liver organogenesis during zebrafish embryogenesis', *Development*. Company of Biologists, 138(10), pp. 1935–1945. doi: 10.1242/dev.056861.

Klein, D. C. and Hainer, S. J. (2020) 'Genomic methods in profiling DNA accessibility and factor localization', *Chromosome research: an international journal on the molecular, supramolecular and evolutionary aspects of chromosome biology*. NLM (Medline), 28(1), pp. 69–85. doi: 10.1007/s10577-019-09619-9.

Klemm, S. L., Shipony, Z. and Greenleaf, W. J. (2019) 'Chromatin accessibility and the regulatory epigenome', *Nature Reviews Genetics*, 20(4), pp. 207–220. doi: 10.1038/s41576-018-0089-8.

Ko, L. J. and Engel, J. D. (1993) *DNA-binding specificities of the GATA transcription factor family.*, *Molecular and Cellular Biology*. doi: 10.1128/mcb.13.7.4011.

Komisarczuk, A. Z., Kawakami, K. and Becker, T. S. (2009) 'Cis-regulation and chromosomal rearrangement of the *fgf8* locus after the teleost/tetrapod split', *Developmental Biology*. Academic Press, 336(2), pp. 301–312.

Kosaki, K. and Casey, B. (1998) 'Genetics of human left-right axis malformations', *Seminars in Cell and Developmental Biology*. Elsevier Ltd, 9(1), pp. 89–99. doi: 10.1006/scdb.1997.0187.

Koshikawa, S. (2015) 'Enhancer modularity and the evolution of new traits', *Fly*. Taylor & Francis, 9(4), pp. 155–159. doi: 10.1080/19336934.2016.1151129.

Koutsourakis, M. *et al.* (2001) 'Branching and differentiation defects in pulmonary epithelium with elevated Gata6 expression', *Mechanisms of Development*. Mech Dev, 105(1–2), pp. 105–114. doi: 10.1016/S0925-4773(01)00386-0.

Kramer-Zucker, A. G. *et al.* (2005) 'Cilia-driven fluid flow in the zebrafish pronephros, brain and Kupffer's vesicle is required for normal organogenesis', *Development*. The Company of Biologists Ltd, 132(8), pp. 1907–1921. doi: 10.1242/dev.01772.

Krebs, A. R. *et al.* (2010) 'seqMINER: an integrated ChIP-seq data interpretation platform', *Nucleic Acids Research*, 39(6), pp. e35–e35. doi: 10.1093/nar/gkq1287.

Krefting, J., Andrade-Navarro, M. A. and Ibn-Salem, J. (2018) 'Evolutionary stability of topologically associating domains is associated with conserved gene regulation', *BMC Biology*. BioMed Central Ltd., 16(1), pp. 1–12. doi: 10.1186/s12915-018-0556-x.

Krijger, P. H. L. and De Laat, W. (2016) 'Regulation of disease-associated gene expression in the 3D genome', *Nature Reviews Molecular Cell Biology*. Nature Publishing Group, 17(12), pp. 771–782. doi: 10.1038/nrm.2016.138.

Kulkeaw, K. and Sugiyama, D. (2012) 'Zebrafish erythropoiesis and the utility of fish as models of anemia', *Stem Cell Research and Therapy*. BioMed Central, p. 55. doi: 10.1186/srct146.

Kunwar, P. S. *et al.* (2003) 'Mixer/Bon and FoxH1/Sur have overlapping and divergent roles in nodal signaling and mesendoderm induction', *Development*. The Company of Biologists Ltd, pp. 5589–5599. doi: 10.1242/dev.00803.

Kupffer, C. (1868) 'Beobachtungen über die Entwicklung der Knochenfische', *Archiv für mikroskopische Anatomie*. Springer, 4(1), pp. 209–272. doi: 10.1007/BF02955363.

Kwan, K. Y. *et al.* (2008) 'SOX5 postmitotically regulates migration, postmigratory differentiation, and projections of subplate and deep-layer neocortical neurons', *Proceedings of the National Academy of Sciences of the United States of America*. National Academy of Sciences, 105(41), pp. 16021–16026. doi: 10.1073/pnas.0806791105.

Kyrchanova, O. *et al.* (2008) 'Orientation-dependent interaction between Drosophila insulators is a property of this class of regulatory elements', *Nucleic Acids Research*, 36(22), pp. 7019–7028. doi: 10.1093/nar/gkn781.

De Laat, W. and Duboule, D. (2013) 'Topology of mammalian developmental enhancers and their regulatory landscapes', *Nature*. Nature Publishing Group, pp. 499–506. doi: 10.1038/nature12753.

Lancman, J. J. *et al.* (2013) 'Specification of hepatopancreas progenitors in zebrafish by hnf1ba and wnt2bb', *Development (Cambridge)*. Company of Biologists, 140(13), pp. 2669–2679. doi: 10.1242/dev.090993.

Lander, E. S. *et al.* (2001) 'Initial sequencing and analysis of the human genome', *Nature*, 409(6822), pp. 860–921. doi: 10.1038/35057062.

Langheinrich, U. (2003) 'Zebrafish: A new model on the pharmaceutical catwalk', *BioEssays*, 25(9), pp. 904–912. doi: 10.1002/bies.10326.

Langmead, B. and Salzberg, S. L. (2012) 'Fast gapped-read alignment with Bowtie 2', *Nature Methods*. NIH Public Access, 9(4), pp. 357–359. doi: 10.1038/nmeth.1923.

Lee, C. T. *et al.* (2004) 'The Nuclear Orphan Receptor COUP-TFII Is Required for Limb and Skeletal Muscle Development', *Molecular and Cellular Biology*. American Society for Microbiology, 24(24), pp. 10835–10843. doi: 10.1128/mcb.24.24.10835-10843.2004.

- Lee, M. T. *et al.* (2013) 'Nanog, Pou5f1 and SoxB1 activate zygotic gene expression during the maternal-to-zygotic transition', *Nature*, 503(7476), pp. 360–364. doi: 10.1038/nature12632.
- Lee, M. T., Bonneau, A. R. and Giraldez, A. J. (2014) 'Zygotic genome activation during the maternal-to-zygotic transition', *Annual review of cell and developmental biology*. NIH Public Access, pp. 581–613. doi: 10.1146/annurev-cellbio-100913-013027.
- Leichsenring, M. *et al.* (2013) 'Pou5f1 transcription factor controls zygotic gene activation in vertebrates', *Science*. American Association for the Advancement of Science, 341(6149), pp. 1005–1009. doi: 10.1126/science.1242527.
- Li, G. *et al.* (2012) 'Extensive promoter-centered chromatin interactions provide a topological basis for transcription regulation', *Cell*. Cell, 148(1–2), pp. 84–98. doi: 10.1016/j.cell.2011.12.014.
- Li, H. *et al.* (2009) 'The Sequence Alignment/Map format and SAMtools', *Bioinformatics*. Oxford University Press, 25(16), pp. 2078–2079. doi: 10.1093/bioinformatics/btp352.
- Li, L. *et al.* (2004) 'Gene regulation by Sp1 and Sp3', *Biochemistry and Cell Biology*. NRC Research Press Ottawa, Canada, 82(4), pp. 460–471. doi: 10.1139/o04-045.
- Li, Z. *et al.* (2019) 'Identification of transcription factor binding sites using ATAC-seq', *Genome Biology*. BioMed Central Ltd., 20(1), p. 45. doi: 10.1186/s13059-019-1642-2.
- Lim, C. Y. *et al.* (2004) 'The MTE, a new core promoter element for transcription by RNA polymerase II', *Genes and Development*. Cold Spring Harbor Laboratory Press, 18(13), pp. 1606–1617. doi: 10.1101/gad.1193404.
- Lin, S. J. *et al.* (2014) 'Minireview: Pathophysiological roles of the TR4 nuclear receptor: Lessons learned from mice lacking TR4', *Molecular Endocrinology*. The Endocrine Society, 28(6), pp. 805–821. doi: 10.1210/me.2013-1422.
- Liu, C., Morrissey, E. E. and Whitsett, J. A. (2002) 'GATA-6 is required for maturation of the lung in late gestation', *American Journal of Physiology - Lung Cellular and Molecular Physiology*. American Physiological Society, 283(2 27-2). doi: 10.1152/ajplung.00044.2002.
- Liu, D. W. *et al.* (2011) 'A variant of Fibroblast growth factor receptor 2 (Fgfr2) regulates left-right asymmetry in Zebrafish', *PLoS ONE*. Edited by H. Escriva. Public Library of Science, 6(7), p. e21793. doi: 10.1371/journal.pone.0021793.
- Liu, G. *et al.* (2018) 'Inherited DNA methylation primes the establishment of accessible chromatin during genome activation', *Genome Research*. Cold Spring Harbor Laboratory Press, 28(7), pp. 998–1007. doi: 10.1101/gr.228833.117.
- Liu, J. *et al.* (2019) 'Chemokine signaling links cell-cycle progression and cilia formation for left-right symmetry breaking', *PLoS Biology*. Public Library of Science, 17(8). doi: 10.1371/journal.pbio.3000203.
- Liu, J. K., Bergman, Y. and Zaret, K. S. (1988) *The mouse albumin promoter and a distal*

upstream site are simultaneously DNase I hypersensitive in liver chromatin and bind similar liver-abundant factors in vitro., *Genes & development*. doi: 10.1101/gad.2.5.528.

Liu, Z. *et al.* (2016) 'Fscn1 is required for the trafficking of TGF- β family type I receptors during endoderm formation', *Nature Communications*. Nature Publishing Group, 7, p. 12603. doi: 10.1038/ncomms12603.

Livak, K. J. and Schmittgen, T. D. (2001) 'Analysis of relative gene expression data using real-time quantitative PCR and the 2- $\Delta\Delta$ CT method', *Methods*. Academic Press Inc., 25(4), pp. 402–408. doi: 10.1006/meth.2001.1262.

Loh, Y. H. *et al.* (2006) 'The Oct4 and Nanog transcription network regulates pluripotency in mouse embryonic stem cells', *Nature Genetics*. Nature Publishing Group, 38(4), pp. 431–440. doi: 10.1038/ng1760.

Lokmane, L. *et al.* (2008) 'Crucial role of vHNF1 in vertebrate hepatic specification', *Development*. The Company of Biologists, 135(16), pp. 2777–2786. doi: 10.1242/dev.023010.

Lolas, M. *et al.* (2014) 'Charting Brachyury-mediated developmental pathways during early mouse embryogenesis', *Proceedings of the National Academy of Sciences of the United States of America*, 111(12), pp. 4478–4483. doi: 10.1073/pnas.1402612111.

Long, S., Ahmad, N. and Rebagliati, M. (2003) 'The zebrafish nodal-related gene southpaw is required for visceral and diencephalic left-right asymmetry', *Development*, 130(11), pp. 2303–2316. doi: 10.1242/dev.00436.

Lorent, K. *et al.* (2004) 'Inhibition of Jagged-mediated notch signaling disrupts zebrafish biliary development and generates multi-organ defects compatible with an Alagille syndrome phenocopy', *Development*. The Company of Biologists Ltd, 131(22), pp. 5753–5766. doi: 10.1242/dev.01411.

Love, M. I., Huber, W. and Anders, S. (2014) 'Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2', *Genome Biology*. BioMed Central, 15(12), p. 550. doi: 10.1186/s13059-014-0550-8.

Lu, H. *et al.* (2013) 'EpCAM Is an Endoderm-Specific Wnt Derepressor that Licenses Hepatic Development', *Developmental Cell*. Cell Press, 24(5), pp. 543–553. doi: 10.1016/j.devcel.2013.01.021.

Lunde, K., Belting, H. G. and Driever, W. (2004) 'Zebrafish pou5f1/pou2, Homolog of Mammalian Oct4, Functions in the Endoderm Specification Cascade', *Current Biology*. Cell Press, 14(1), pp. 48–55. doi: 10.1016/j.cub.2003.11.022.

Lupiáñez, D. G. *et al.* (2015) 'Disruptions of topological chromatin domains cause pathogenic rewiring of gene-enhancer interactions', *Cell*, 161(5), pp. 1012–1025. doi: 10.1016/j.cell.2015.04.004.

Maier, E. C. and Whitfield, T. T. (2014) 'RA and FGF Signalling Are Required in the Zebrafish Otic Vesicle to Pattern and Maintain Ventral Otic Identities', *PLoS Genetics*. Edited by M. C.

Mullins. Public Library of Science, 10(12), p. e1004858. doi: 10.1371/journal.pgen.1004858.

Manfroid, I. *et al.* (2007) 'Reciprocal endoderm-mesoderm interactions mediated by fgf24 and fgf10 govern pancreas development', *Development*. The Company of Biologists Ltd, 134(22), pp. 4011–4021. doi: 10.1242/dev.007823.

Martin, B. L. and Kimelman, D. (2008) 'Regulation of Canonical Wnt Signaling by Brachyury Is Essential for Posterior Mesoderm Formation', *Developmental Cell*. Cell Press, 15(1), pp. 121–133. doi: 10.1016/j.devcel.2008.04.013.

Martinez-Barbera, J. P. and Beddington, R. S. P. (2001) *Getting your head around Hex and Hesx1: Forebrain formation in mouse*, *International Journal of Developmental Biology*. doi: 10.1387/ijdb.11291863.

Marty-Santos, L. and Cleaver, O. (2016) 'Pdx1 regulates pancreas tubulogenesis and E-cadherin expression', *Development (Cambridge)*. Company of Biologists Ltd, 143(1), pp. 101–112. doi: 10.1242/dev.126755.

Maston, G. A., Evans, S. K. and Green, M. R. (2006) 'Transcriptional regulatory elements in the human genome', *Annual Review of Genomics and Human Genetics*. Annual Reviews, 7, pp. 29–59. doi: 10.1146/annurev.genom.7.080505.115623.

Masui, T. *et al.* (2008) 'Transcriptional Autoregulation Controls Pancreatic Ptf1a Expression during Development and Adulthood', *Molecular and Cellular Biology*. American Society for Microbiology, 28(17), pp. 5458–5468. doi: 10.1128/mcb.00549-08.

Matthews, R. P. *et al.* (2004) 'The zebrafish oncut gene hnf-6 functions in an evolutionarily conserved genetic pathway that regulates vertebrate biliary development', *Developmental Biology*. Academic Press Inc., 274(2), pp. 245–259. doi: 10.1016/j.ydbio.2004.06.016.

Mavropoulos, A. *et al.* (2005) 'Sox4B Is a Key Player of Pancreatic A Cell Differentiation in Zebrafish', *Developmental Biology*. Dev Biol, 285(1), pp. 211–223. doi: 10.1016/j.ydbio.2005.06.024.

McCord, R. P. *et al.* (2011) 'Distant cis-regulatory elements in human skeletal muscle differentiation', *Genomics*. Genomics, 98(6), pp. 401–411. doi: 10.1016/j.ygeno.2011.08.003.

McGrath, J. *et al.* (2003) 'Two populations of node monocilia initiate left-right asymmetry in the mouse', *Cell*. Cell Press, 114(1), pp. 61–73. doi: 10.1016/S0092-8674(03)00511-7.

McLean, C. Y. *et al.* (2010) 'GREAT improves functional interpretation of cis-regulatory regions', *Nature Biotechnology*. NIH Public Access, 28(5), pp. 495–501. doi: 10.1038/nbt.1630.

McLin, V. A., Rankin, S. A. and Zorn, A. M. (2007) 'Repression of Wnt/ β -catenin signaling in the anterior endoderm is essential for liver and pancreas development', *Development*. Development, 134(12), pp. 2207–2217. doi: 10.1242/dev.001230.

McRae, J. F. *et al.* (2017) 'Prevalence and architecture of de novo mutations in developmental disorders', *Nature*. Nature Publishing Group, 542(7642), pp. 433–438. doi:

10.1038/nature21062.

Melby, A. E., Warga, R. M. and Kimmel, C. B. (1996) 'Specification of cell fates at the dorsal margin of the zebrafish gastrula', *Development*, 122(7), pp. 2225–2237.

Meno, C. *et al.* (2001) 'Diffusion of Nodal Signaling Activity in the Absence of the Feedback Inhibitor Lefty2', *Developmental Cell*. Elsevier, 1(1), pp. 127–138. doi: 10.1016/S1534-5807(01)00006-5.

Merika, M. and Orkin, S. H. (1993) 'DNA-binding specificity of GATA family transcription factors.', *Molecular and Cellular Biology*. American Society for Microbiology, 13(7), pp. 3999–4010. doi: 10.1128/mcb.13.7.3999.

Miguel-Escalada, I., Pasquali, L. and Ferrer, J. (2015) 'Transcriptional enhancers: Functional insights and role in human disease', *Current Opinion in Genetics and Development*. Elsevier Current Trends, 33, pp. 71–76. doi: 10.1016/j.gde.2015.08.009.

Milani, P. *et al.* (2016) 'Cell freezing protocol suitable for ATAC-Seq on motor neurons derived from human induced pluripotent stem cells', *Scientific Reports*. Nature Publishing Group, 6. doi: 10.1038/srep25474.

Mitra, S. *et al.* (2010) 'Identification and characterization of the transcription factors involved in T-cell development, t-bet, stat6 and foxp3, within the zebrafish, *Danio rerio*', *FEBS Journal*. John Wiley & Sons, Ltd, 277(1), pp. 128–147. doi: 10.1111/j.1742-4658.2009.07460.x.

Mitsui, K. *et al.* (2003) 'The homeoprotein nanog is required for maintenance of pluripotency in mouse epiblast and ES cells', *Cell*. Cell Press, 113(5), pp. 631–642. doi: 10.1016/S0092-8674(03)00393-3.

Mizoguchi, T. *et al.* (2008) 'Sdf1/Cxcr4 signaling controls the dorsal migration of endodermal cells during zebrafish gastrulation', *Development*, 135(15), pp. 2521–2529. doi: 10.1242/dev.020107.

Mizuno, T. *et al.* (1996) *Mesoderm induction in zebrafish [3]*, *Nature*. doi: 10.1038/383131a0.

Molkentin, J. D. *et al.* (1997) *Requirement of the transcription factor GATA4 for heart tube formation and ventral morphogenesis*, *Genes and Development*. doi: 10.1101/gad.11.8.1061.

Molkentin, J. D. (2000) 'The zinc finger-containing transcription factors GATA-4, -5, and -6: Ubiquitously expressed regulators of tissue-specific gene expression', *Journal of Biological Chemistry*, 275(50), pp. 38949–38952. doi: 10.1074/jbc.R000029200.

Monaghan, A. P. *et al.* (1993) *Postimplantation expression patterns indicate a role for the mouse forkhead/HNF-3 α , β and γ genes in determination of the definitive endoderm, chordamesoderm and neuroectoderm*, *Development*. doi: 10.1242/dev.119.3.567.

Montague, T. G. and Schier, A. F. (2017) 'Vg1-nodal heterodimers are the endogenous inducers of mesendoderm', *eLife*. eLife Sciences Publications Ltd, 6. doi: 10.7554/eLife.28183.

Morley, R. H. *et al.* (2009) *A gene regulatory network directed by zebrafish No tail accounts*

for its roles in mesoderm formation, *Proceedings of the National Academy of Sciences of the United States of America*. doi: 10.1073/pnas.0808382106.

Mueller, R. L., Huang, C. and Ho, R. K. (2010) 'Spatio-temporal regulation of Wnt and retinoic acid signaling by *tbx16*/*spadetail* during zebrafish mesoderm differentiation', *BMC Genomics*. *BMC Genomics*, 11(1). doi: 10.1186/1471-2164-11-492.

Mullins, M. C. *et al.* (1994) 'Large-scale mutagenesis in the zebrafish: in search of genes controlling development in a vertebrate', *Current Biology*, 4(3), pp. 189–202. doi: 10.1016/S0960-9822(00)00048-8.

Nakai, S. *et al.* (2003) 'Crucial roles fo *Brn1* in distal tubule formation and function in mouse kidney', *Development*. Oxford University Press for The Company of Biologists Limited, 130(19), pp. 4751–4759. doi: 10.1242/dev.00666.

Narlikar, L. and Ovcharenko, I. (2009) 'Identifying regulatory elements in eukaryotic genomes', *Briefings in Functional Genomics and Proteomics*. Oxford University Press, 8(4), pp. 215–230. doi: 10.1093/bfgp/elp014.

Nasevicius, A. and Ekker, S. C. (2000) 'Effective targeted gene "knockdown" in zebrafish', *Nature Genetics*, 26(2), pp. 216–220. doi: 10.1038/79951.

Navis, A., Marjoram, L. and Bagnat, M. (2013) 'Cftr controls lumen expansion and function of Kupffer's vesicle in zebrafish', *Development*, 140(8), pp. 1703–1712. doi: 10.1242/dev.091819.

Nelson, A. C. *et al.* (2012) 'An integrated functional genomics approach identifies the regulatory network directed by brachyury (T) in chordoma', *Journal of Pathology*, 228(3), pp. 274–285. doi: 10.1002/path.4082.

Nelson, A. C. *et al.* (2014) 'Global identification of *smad2* and *eomesodermin* targets in zebrafish identifies a conserved transcriptional network in mesendoderm and a novel role for *eomesodermin* in repression of ectodermal gene expression', *BMC Biology*. BioMed Central, 12(1), p. 81.

Nelson, A. C. *et al.* (2017) 'In Vivo Regulation of the Zebrafish Endoderm Progenitor Niche by T-Box Transcription Factors', *Cell Reports*. Elsevier, 19(13), pp. 2782–2795.

Nelson, A. C. and Wardle, F. C. (2013) 'Conserved non-coding elements and cis regulation: actions speak louder than words', *Development*. Oxford University Press for The Company of Biologists Limited, 140(7), pp. 1385–1395.

Niakan, K. K. *et al.* (2010) '*Sox17* promotes differentiation in mouse embryonic stem cells by directly regulating extraembryonic gene expression and indirectly antagonizing self-renewal', *Genes and Development*. Cold Spring Harbor Laboratory Press, 24(3), pp. 312–326. doi: 10.1101/gad.1833510.

Nishioka, N. *et al.* (2008) 'Tead4 is required for specification of trophectoderm in pre-implantation mouse embryos', *Mechanisms of Development*. Elsevier, 125(3–4), pp. 270–283.

doi: 10.1016/j.mod.2007.11.002.

Nissim, S. *et al.* (2014) 'Prostaglandin E2 regulates liver versus pancreas cell-fate decisions and endodermal outgrowth', *Developmental Cell*. Elsevier, 28(4), pp. 423–437. doi: 10.1016/j.devcel.2014.01.006.

Noll, M. and Kornberg, R. D. (1977) 'Action of micrococcal nuclease on chromatin and the location of histone H1', *Journal of Molecular Biology*. J Mol Biol, 109(3), pp. 393–404. doi: 10.1016/S0022-2836(77)80019-3.

Nonaka, S. *et al.* (1998) 'Randomization of left-right asymmetry due to loss of nodal cilia generating leftward flow of extraembryonic fluid in mice lacking KIF3B motor protein', *Cell*. Cell Press, 95(6), pp. 829–837. doi: 10.1016/S0092-8674(00)81705-5.

Nord, H. *et al.* (2014) 'Differential regulation of myosin heavy chains defines new muscle domains in zebrafish', *Molecular Biology of the Cell*. American Society for Cell Biology, 25(8), pp. 1384–1395. doi: 10.1091/mbc.E13-08-0486.

Norton, W. H. *et al.* (2005) 'Monorail/Foxa2 regulates floorplate differentiation and specification of oligodendrocytes, serotonergic raphé neurones and cranial motoneurones', *Development*, 132(4), pp. 645–658. doi: 10.1242/dev.01611.

Nowotschin, S. *et al.* (2019) 'The emergent landscape of the mouse gut endoderm at single-cell resolution', *Nature*. Nature Publishing Group, 569(7756), pp. 361–367. doi: 10.1038/s41586-019-1127-1.

Nowotschin, S., Hadjantonakis, A. K. and Campbell, K. (2019) 'The endoderm: a divergent cell lineage with many commonalities', *Development (Cambridge, England)*. Oxford University Press for The Company of Biologists Limited, 146(11). doi: 10.1242/dev.150920.

Ober, E. A., Field, H. A. and Stainier, D. Y. R. (2003) 'From endoderm formation to liver and pancreas development in zebrafish', *Mechanisms of Development*, pp. 5–18. doi: 10.1016/S0925-4773(02)00327-1.

Odenthal, J. and Nüsslein-Volhard, C. (1998) 'Fork Head Domain Genes in Zebrafish', *Development Genes and Evolution*. Springer, 208(5), pp. 245–258. doi: 10.1007/s004270050179.

Ohba, H. *et al.* (2004) 'Sox21 is a repressor of neuronal differentiation and is antagonized by YB-1', *Neuroscience Letters*, 358(3), pp. 157–160. doi: 10.1016/j.neulet.2004.01.026.

Okada, Y. *et al.* (2005) 'Mechanism of nodal flow: A conserved symmetry breaking event in left-right axis determination', *Cell*, 121(4), pp. 633–644. doi: 10.1016/j.cell.2005.04.008.

Onichtchouk, D. and Driever, W. (2016) 'Zygotic Genome Activators, Developmental Timing, and Pluripotency', in *Current Topics in Developmental Biology*. Academic Press Inc., pp. 273–297. doi: 10.1016/bs.ctdb.2015.12.004.

Pálffy, M. *et al.* (2020) 'Chromatin accessibility established by Pou5f3, Sox19b and Nanog primes genes for activity during zebrafish genome activation', *PLoS Genetics*. Edited by A. A.

Aboobaker. Public Library of Science, 16(1), p. e1008546. doi: 10.1371/journal.pgen.1008546.

Pander, C. (1817) 'Beyträge zur Entwicklungsgeschichte des Hühnchens (Würzburg)', in.

Panigrahi, A. and O'Malley, B. W. (2021) 'Mechanisms of enhancer action: the known and the unknown', *Genome Biology*. BioMed Central, 22(1), pp. 1–30. doi: 10.1186/s13059-021-02322-1.

Paredes, R. M. *et al.* (2008) 'Chemical calcium indicators', *Methods*. NIH Public Access, 46(3), pp. 143–151. doi: 10.1016/j.ymeth.2008.09.025.

Pashos, E. *et al.* (2013) 'Distinct enhancers of ptf1a mediate specification and expansion of ventral pancreas in zebrafish', *Developmental Biology*. Academic Press Inc., 381(2), pp. 471–481. doi: 10.1016/j.ydbio.2013.07.011.

Passegué, E. *et al.* (2002) 'JunB can substitute for Jun in mouse development and cell proliferation', *Nature Genetics*. Nature Publishing Group, 30(2), pp. 158–166. doi: 10.1038/ng790.

Pawlak, M. *et al.* (2019) 'Dynamics of cardiomyocyte transcriptome and chromatin landscape demarcates key events of heart development', *Genome Research*. Cold Spring Harbor Laboratory Press, 29(3), pp. 506–519. doi: 10.1101/gr.244491.118.

Pennacchio, L. A. *et al.* (2006) 'In vivo enhancer analysis of human conserved non-coding sequences', *Nature*. Nature Publishing Group, 444(7118), pp. 499–502. doi: 10.1038/nature05295.

Pennacchio, L. A. *et al.* (2013) 'Enhancers: Five essential questions', *Nature Reviews Genetics*. NIH Public Access, pp. 288–295. doi: 10.1038/nrg3458.

Pereira, F. A. *et al.* (1999) *The orphan nuclear receptor COUP-TFII is required for angiogenesis and heart development*, *Genes and Development*. doi: 10.1101/gad.13.8.1037.

Perez-Camps, M. *et al.* (2016) 'Quantitative imaging reveals real-time Pou5f3-nanog complexes driving dorsoventral mesendoderm patterning in zebrafish', *eLife*. eLife Sciences Publications Ltd, 5(September2016). doi: 10.7554/eLife.11475.

Pérez-Rico, Y. A., Barillot, E. and Shkumatava, A. (2020) 'Demarcation of Topologically Associating Domains Is Uncoupled from Enriched CTCF Binding in Developing Zebrafish', *iScience*. Elsevier, 23(5), p. 101046. doi: 10.1016/j.isci.2020.101046.

Pevny, L. H. and Lovell-Badge, R. (1997) 'Sox genes find their feet', *Current Opinion in Genetics and Development*. Elsevier Current Trends, pp. 338–344.

Peyriéras, N., Strähle, U. and Rosa, F. (1998) 'Conversion of zebrafish blastomeres to an endodermal fate by TGF- β -related signalling', *Current Biology*. Cell Press, 8(13), pp. 783–788. doi: 10.1016/s0960-9822(98)70303-3.

Pézeron, G. *et al.* (2008) 'Ras11b knock down in zebrafish suppresses one-eyed-pinhead mutant phenotype', *PLoS ONE*. Edited by T. Zwaka. Public Library of Science, 3(1), p. e1434.

doi: 10.1371/journal.pone.0001434.

Picozzi, P. *et al.* (2009) 'Eomesodermin requires transforming growth factor- β /activin signaling and binds Smad2 to activate mesodermal genes', *Journal of Biological Chemistry*. J Biol Chem, 284(4), pp. 2397–2408. doi: 10.1074/jbc.M808704200.

Piotrowski, T. and Nüsslein-Volhard, C. (2000) 'The endoderm plays an important role in patterning the segmented pharyngeal region in zebrafish (*Danio rerio*)', *Developmental Biology*, 225(2), pp. 339–356. doi: 10.1006/dbio.2000.9842.

Piper, K. *et al.* (2002) 'Novel SOX9 expression during human pancreas development correlates to abnormalities in Campomelic dysplasia', *Mechanisms of Development*. Elsevier, 116(1–2), pp. 223–226. doi: 10.1016/S0925-4773(02)00145-4.

Pogoda, H. M. *et al.* (2000) 'The zebrafish forkhead transcription factor FoxH1/Fast1 is a modulator of Nodal signaling required for organizer formation', *Current Biology*. Current Biology Ltd, 10(17), pp. 1041–1049. doi: 10.1016/S0960-9822(00)00669-2.

Porazzi, P. *et al.* (2012) 'Disruptions of global and Jagged1-mediated notch signaling affect thyroid morphogenesis in the zebrafish', *Endocrinology*. Oxford Academic, 153(11), pp. 5645–5658. doi: 10.1210/en.2011-1888.

Poulain, M. and Lepage, T. (2002) 'Mezzo, a paired-like homeobox protein is an immediate target of nodal signalling and regulates endoderm specification in zebrafish', *Development*, 129(21), pp. 4901–4914. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/12397099> (Accessed: 21 March 2020).

Pradeepa, M. M. *et al.* (2016) 'Histone H3 globular domain acetylation identifies a new class of enhancers', *Nature Genetics*. Nature Publishing Group, 48(6), pp. 681–686. doi: 10.1038/ng.3550.

Prelich, G. (2012) 'Gene overexpression: Uses, mechanisms, and interpretation', *Genetics*. Genetics Society of America, pp. 841–854. doi: 10.1534/genetics.111.136911.

Di Prinzio, C. M. *et al.* (2010) 'Growth hormone receptors in zebrafish (*Danio rerio*): Adult and embryonic expression patterns', *Gene Expression Patterns*, 10(4–5), pp. 214–225. doi: 10.1016/j.gep.2010.03.001.

Proulx, K., Lu, A. and Sumanas, S. (2010) 'Cranial vasculature in zebrafish forms by angioblast cluster-derived angiogenesis', *Developmental Biology*. Academic Press, 348(1), pp. 34–46.

Pugacheva, E. M. *et al.* (2006) 'Cloning and characterization of zebrafish CTCF: Developmental expression patterns, regulation of the promoter region, and evolutionary aspects of gene organization', *Gene*. Elsevier, 375(1–2), pp. 26–36. doi: 10.1016/j.gene.2006.01.036.

Quillien, A. *et al.* (2017) 'Robust Identification of Developmentally Active Endothelial Enhancers in Zebrafish Using FANS-Assisted ATAC-Seq', *Cell Reports*. Elsevier Company.,

20(3), pp. 709–720.

Quinlan, A. R. and Hall, I. M. (2010) 'BEDTools: A flexible suite of utilities for comparing genomic features', *Bioinformatics*, 26(6), pp. 841–842. doi: 10.1093/bioinformatics/btq033.

Randall, R. A. *et al.* (2002) 'Different Smad2 partners bind a common hydrophobic pocket in Smad2 via a defined proline-rich motif', *EMBO Journal*. European Molecular Biology Organization, 21(1–2), pp. 145–156. doi: 10.1093/emboj/21.1.145.

Rankin, S. A. *et al.* (2018) 'Timing is everything: Reiterative Wnt, BMP and RA signaling regulate developmental competence during endoderm organogenesis', *Developmental Biology*. NIH Public Access, 434(1), pp. 121–132. doi: 10.1016/j.ydbio.2017.11.018.

Rebagliati, M. R., Toyama, R., Haffter, P., *et al.* (1998) *Cyclops Encodes a Nodal-Related Factor Involved in Midline Signaling*, *Proceedings of the National Academy of Sciences of the United States of America*. doi: 10.1073/pnas.95.17.9932.

Rebagliati, M. R., Toyama, R., Fricke, C., *et al.* (1998) 'Zebrafish nodal-related genes are implicated in axial patterning and establishing left-right asymmetry', *Developmental Biology*. Academic Press, 199(2), pp. 261–272. doi: 10.1006/dbio.1998.8935.

Reim, G. *et al.* (2004) 'The POU domain protein Spg (Pou2/Oct4) is essential for endoderm formation in cooperation with the HMG domain protein casanova', *Developmental Cell*. University of Oregon Press, Oregon, 6(1), pp. 91–101. doi: 10.1016/S1534-5807(03)00396-4.

Reiter *et al.* (1999) 'Gata5 is required for the development of the heart and endoderm in zebrafish', *Genes and Development*. Cold Spring Harbor Laboratory Press, 13(22), pp. 2983–2995. doi: 10.1101/gad.13.22.2983.

Reiter, Kikuchi and Stainier (2001a) 'Multiple roles for Gata5 in zebrafish endoderm formation', *Development*, 128(1), pp. 125–135. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/11092818> (Accessed: 21 March 2020).

Reiter, Kikuchi and Stainier (2001b) *Zebrafish endoderm formation requires Gata5*.

Robert, K. H. O. (1992) 'Cell movements and cell fate during zebrafish gastrulation', *Development*, 116(4 SUPPL.), pp. 65–73.

Rodaway, A. *et al.* (1999) *Induction of the mesendoderm in the zebrafish germ ring by yolk cell-derived TGF- β family signals and discrimination of mesoderm and endoderm by FGF*, *Development*. doi: 10.1242/dev.126.14.3067.

Roy, S. *et al.* (2010) 'Identification of functional elements and regulatory circuits by *Drosophila* modENCODE', *Science*. American Association for the Advancement of Science, 330(6012), pp. 1787–1797. doi: 10.1126/science.1198374.

Sadler, K. C. *et al.* (2007) 'Liver growth in the embryo and during liver regeneration in zebrafish requires the cell cycle regulator, uhrf1', *Proceedings of the National Academy of Sciences of the United States of America*, 104(5), pp. 1570–1575. doi: 10.1073/pnas.0610774104.

Sakaguchi, T., Kuroiwa, A. and Takeda, H. (2001) 'A novel sox gene, 226D7, acts downstream

of Nodal signaling to specify endoderm precursors in zebrafish', *Mechanisms of Development*. Elsevier, 107(1–2), pp. 25–38. doi: 10.1016/S0925-4773(01)00453-1.

Sampath, K. *et al.* (1998) 'Induction of the zebrafish ventral brain and floorplate requires cyclops/nodal signalling', *Nature*. Nature Publishing Group, 395(6698), pp. 185–189. doi: 10.1038/26020.

Sandberg, M., Källström, M. and Muhr, J. (2005) 'Sox21 promotes the progression of vertebrate neurogenesis', *Nature Neuroscience*, 8(8), pp. 995–1001. doi: 10.1038/nn1493.

Santoriello, C. and Zon, L. I. (2012) 'Hooked! modeling human disease in zebrafish', *Journal of Clinical Investigation*, pp. 2337–2343. doi: 10.1172/JCI60434.

Saund, R. S. *et al.* (2012) 'Gut endoderm is involved in the transfer of left-right asymmetry from the node to the lateral plate mesoderm in the mouse embryo', *Development (Cambridge)*. Oxford University Press for The Company of Biologists Limited, 139(13), pp. 2426–2435. doi: 10.1242/dev.079921.

Sawada, A. *et al.* (2008) 'Redundant Roles of Tead1 and Tead2 in Notochord Development and the Regulation of Cell Proliferation and Survival', *Molecular and Cellular Biology*. American Society for Microbiology, 28(10), pp. 3177–3189. doi: 10.1128/mcb.01759-07.

Saxena, R. *et al.* (2007) 'Genome-wide association analysis identifies loci for type 2 diabetes and triglyceride levels', *Science*. American Association for the Advancement of Science, 316(5829), pp. 1331–1336. doi: 10.1126/science.1142358.

Schaffner, W. (2015) 'Enhancers, enhancers - From their discovery to today's universe of transcription enhancers', *Biological Chemistry*. De Gruyter, 396(4), pp. 311–327. doi: 10.1515/hsz-2014-0303.

Schier, A. F. (2003) 'Nodal Signaling in Vertebrate Development', *Annual Review of Cell and Developmental Biology*, 19(1), pp. 589–621. doi: 10.1146/annurev.cellbio.19.041603.094522.

Schier, A. F. (2005) 'Axis formation: Squint comes into focus', *Current Biology*, 15(24). doi: 10.1016/j.cub.2005.11.051.

Schier, A. F. (2009) 'Nodal morphogens.', *Cold Spring Harbor perspectives in biology*. Cold Spring Harbor Laboratory Press, 1(5), p. a003459. doi: 10.1101/cshperspect.a003459.

Schier, A. F. and Talbot, W. S. (2005) 'Molecular genetics of axis formation in zebrafish', *Annual Review of Genetics*. Annual Reviews, 39(1), pp. 561–613. doi: 10.1146/annurev.genet.37.110801.143752.

Schindelin, J. *et al.* (2012) 'Fiji: An open-source platform for biological-image analysis', *Nature Methods*. Nature Publishing Group, pp. 676–682. doi: 10.1038/nmeth.2019.

Schulte-Merker, S. *et al.* (1994) 'no tail (ntl) is the zebrafish homologue of the mouse T (Brachyury) gene', *Development*, 120(4), pp. 1009–1015.

Schulz, K. N. and Harrison, M. M. (2019) 'Mechanisms regulating zygotic genome activation', *Nature Reviews Genetics*, 20(4), pp. 221–234. doi: 10.1038/s41576-018-0087-x.

- Schweickert, A. *et al.* (2007) 'Cilia-Driven Leftward Flow Determines Laterality in *Xenopus*', *Current Biology*, 17(1), pp. 60–66. doi: 10.1016/j.cub.2006.10.067.
- Sedykh, I. *et al.* (2017) 'Zebrafish *zic2* controls formation of periocular neural crest and choroid fissure morphogenesis', *Developmental Biology*. NIH Public Access, 429(1), pp. 92–104. doi: 10.1016/j.ydbio.2017.07.003.
- Segert, J. A., Gisselbrecht, S. S. and Bulyk, M. L. (2021) 'Transcriptional Silencers: Driving Gene Expression with the Brakes On', *Trends in Genetics*. Elsevier. doi: 10.1016/j.tig.2021.02.002.
- Séguin, C. A. *et al.* (2008) 'Establishment of Endoderm Progenitors by SOX Transcription Factor Expression in Human Embryonic Stem Cells', *Cell Stem Cell*. Elsevier, 3(2), pp. 182–195. doi: 10.1016/j.stem.2008.06.018.
- Selleck, M. A. J. and Stern, C. D. (1991) *Fate mapping and cell lineage analysis of Hensen's node in the chick embryo*, *Development*.
- Sellick, G. S. *et al.* (2004) 'Mutations in PTF1A cause pancreatic and cerebellar agenesis', *Nature Genetics*. Nat Genet, 36(12), pp. 1301–1305. doi: 10.1038/ng1475.
- Seth, A., Stemple, D. L. and Barroso, I. (2013) 'The emerging use of zebrafish to model metabolic disease', *DMM Disease Models and Mechanisms*. Dis Model Mech, pp. 1080–1088. doi: 10.1242/dmm.011346.
- Shen, M. M. (2007) 'Nodal signaling: Development roles and regulation', *Development*, 134(6), pp. 1023–1034. doi: 10.1242/dev.000166.
- Shen, W. *et al.* (2001) 'Foxa3 (Hepatocyte Nuclear Factor 3 γ) is Required for the Regulation of Hepatic GLUT2 Expression and the Maintenance of Glucose Homeostasis during a Prolonged Fast', *Journal of Biological Chemistry*. Elsevier, 276(46), pp. 42812–42817. doi: 10.1074/jbc.M106344200.
- Shin, C. H. *et al.* (2008) 'Multiple roles for Med12 in vertebrate endoderm development', *Developmental Biology*. Academic Press, 317(2), pp. 467–479. doi: 10.1016/j.ydbio.2008.02.031.
- Shlyueva, D., Stampfel, G. and Stark, A. (2014) 'Transcriptional enhancers: from properties to genome-wide predictions', *Nature Reviews Genetics*, 15(4), pp. 272–286. doi: 10.1038/nrg3682.
- Short, P. J. *et al.* (2017) 'De novo mutations in regulatory elements cause neurodevelopmental disorders', *bioRxiv*. doi: 10.1101/112896.
- Siew, H. L. *et al.* (2006) 'Transcriptome kinetics of arsenic-induced adaptive response in zebrafish liver', *Physiological Genomics*. American Physiological Society, 27(3), pp. 351–361. doi: 10.1152/physiolgenomics.00201.2005.
- Simon, J. M. *et al.* (2012) 'Using formaldehyde-assisted isolation of regulatory elements (FAIRE) to isolate active regulatory DNA', *Nature Protocols*. Nature Publishing Group, 7(2),

pp. 256–267.

Sladek, R. *et al.* (2007) 'A genome-wide association study identifies novel risk loci for type 2 diabetes', *Nature*. *Nature*, 445(7130), pp. 881–885. doi: 10.1038/nature05616.

Slagle, C. E., Aoki, T. and Burdine, R. D. (2011) 'Nodal-dependent mesendoderm specification requires the combinatorial activities of FoxH1 and eomesodermin', *PLoS Genetics*. Edited by M. C. Mullins. Public Library of Science, 7(5), p. e1002072.

Solnica-Krezel, L., Schier, A. F. and Driever, W. (1994) 'Efficient recovery of ENU-induced mutations from the zebrafish germline', *Genetics*, 136(4), pp. 1401–1420.

Song, L. and Crawford, G. E. (2010) 'DNase-seq: A high-resolution technique for mapping active gene regulatory elements across the genome from mammalian cells', *Cold Spring Harbor Protocols*. Cold Spring Harbor Laboratory Press, 5(2), p. pdb.prot5384.

Spence, J. R. and Wells, J. M. (2007) 'Translational embryology: Using embryonic principles to generate pancreatic endocrine cells from embryonic stem cells', *Developmental Dynamics*. Wiley-Blackwell, 236(12), pp. 3218–3227.

Spitz, F. and Furlong, E. E. M. (2012) 'Transcription factors: From enhancer binding to developmental control', *Nature Reviews Genetics*. Nature Publishing Group, pp. 613–626. doi: 10.1038/nrg3207.

Srinivasan, L. and Atchison, M. L. (2004) 'YY1 DNA binding and PcG recruitment requires CtBP', *Genes and Development*. Cold Spring Harbor Laboratory Press, 18(21), pp. 2596–2601. doi: 10.1101/gad.1228204.

Stafford, D. (2006) 'Retinoids signal directly to zebrafish endoderm to specify insulin-expressing β -cells', *Development*, 133(5), pp. 949–956. doi: 10.1242/dev.02263.

Stafford, D. *et al.* (2006) 'Retinoids signal directly to zebrafish endoderm to specify insulin-expressing β -cells', *Development*. The Company of Biologists Ltd, 133(5), pp. 949–956. doi: 10.1242/dev.02263.

Stafford, D. and Prince, V. E. (2002) 'Retinoic acid signaling is required for a critical early step in zebrafish pancreatic development', *Current Biology*. *Curr Biol*, 12(14), pp. 1215–1220. doi: 10.1016/S0960-9822(02)00929-6.

Stainier, D. Y. R. (2002) 'A glimpse into the molecular entrails of endoderm formation', *Genes and Development*. Cold Spring Harbor Laboratory Press, 16(8), pp. 893–907. doi: 10.1101/gad.974902.

Stark, R and Brown, G. (2016) 'DiffBind: differential binding analysis of ChIP-Seq peak data', pp. 1–29. Available at: <http://bioconductor.org/packages/release/bioc/vignettes/DiffBind/inst/doc/DiffBind.pdf> (Accessed: 12 September 2018).

Stark, R. and Brown, G. (2011) 'DiffBind : differential binding analysis of ChIP-Seq peak data', *Bioconductor*, pp. 1–27. Available at:

<http://bioconductor.org/packages/release/bioc/vignettes/DiffBind/inst/doc/DiffBind.pdf>

(Accessed: 29 September 2019).

Stoffers, D. A. *et al.* (1997) 'Pancreatic agenesis attributable to a single nucleotide deletion in the human IPF1 gene coding sequence', *Nature Genetics*. Nature Publishing Group, 15(1), pp. 106–110. doi: 10.1038/ng0197-106.

Strahle, U. *et al.* (1993) 'Axial, a zebrafish gene expressed along the developing body axis, shows altered expression in cyclops mutant embryos', *Genes and Development*, 7(7 B), pp. 1436–1446. doi: 10.1101/gad.7.7b.1436.

Stubbs, J. L. *et al.* (2008) 'The forkhead protein Foxj1 specifies node-like cilia in *Xenopus* and zebrafish embryos', *Nature Genetics*, 40(12), pp. 1454–1460. doi: 10.1038/ng.267.

Stückemann, T. *et al.* (2012) 'Zebrafish *Cxcr4a* determines the proliferative response to Hedgehog signalling.', *Development (Cambridge, England)*, 139(15), pp. 2711–2720. doi: 10.1242/dev.074930.

Subramanian, A. *et al.* (2005) 'Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles', *Proceedings of the National Academy of Sciences of the United States of America*. National Academy of Sciences, 102(43), pp. 15545–15550. doi: 10.1073/pnas.0506580102.

Sulik, K. *et al.* (1994) 'Morphogenesis of the murine node and notochordal plate', *Developmental Dynamics*, 201(3), pp. 260–278. doi: 10.1002/aja.1002010309.

Sun, Y. L. *et al.* (1986) *Digestion of the chicken beta-globin gene chromatin with micrococcal nuclease reveals the presence of an altered nucleosomal array characterized by an atypical ladder of DNA fragments.*, *The EMBO journal*. doi: 10.1002/j.1460-2075.1986.tb04212.x.

Sun, Y., Miao, N. and Sun, T. (2019) 'Detect accessible chromatin using ATAC-seq, from principle to applications', *Hereditas*. NLM (Medline), p. 29. doi: 10.1186/s41065-019-0105-9.

Suryamohan, K. and Halfon, M. S. (2015) 'Identifying transcriptional cis-regulatory modules in animal genomes', *Wiley Interdisciplinary Reviews: Developmental Biology*, 4(2), pp. 59–84.

Tada, S. *et al.* (2005) 'Characterization of mesendoderm: A diverging point of the definitive endoderm and mesoderm in embryonic stem cell differentiation culture', *Development*. The Company of Biologists Ltd, 132(19), pp. 4363–4374. doi: 10.1242/dev.02005.

Tadros, W. and Lipshitz, H. D. (2009) 'The maternal-to-zygotic transition: A play in two acts', *Development*. Development, 136(18), pp. 3033–3042. doi: 10.1242/dev.033183.

Takizawa, F. *et al.* (2007) 'Expression analysis of two Eomesodermin homologues in zebrafish lymphoid tissues and cells', *Molecular Immunology*, 44(9), pp. 2324–2331. doi: 10.1016/j.molimm.2006.11.018.

Talbot, C. D. *et al.* (2020) 'Eomesodermin is functionally conserved between zebrafish and mouse in spite of different mutant phenotypic severities, and controls left/right organiser

formation via interlocking feedforward loops', *bioRxiv*. bioRxiv, p. 2020.10.02.324244. doi: 10.1101/2020.10.02.324244.

Talbot, J. C. *et al.* (2012) 'Fras1 Shapes Endodermal Pouch 1 and Stabilizes Zebrafish Pharyngeal Skeletal Development.', *Development (Cambridge, England)*, 139(15), pp. 2804–2813. doi: 10.1242/dev.074906.

Talbot, J. C. *et al.* (2019) 'Muscle precursor cell movements in zebrafish are dynamic and require six family genes', *Development (Cambridge)*. Company of Biologists Ltd, 146(10). doi: 10.1242/dev.171421.

Tantin, D. (2013) 'Oct transcription factors in development and stem cells: Insights and mechanisms', *Development (Cambridge)*. Company of Biologists, 140(14), pp. 2857–2866. doi: 10.1242/dev.095927.

Tavares, B. *et al.* (2017) 'Notch/Her12 signalling modulates, motile/immotile cilia ratio downstream of Foxj1a in zebrafish left-right organizer', *eLife*. eLife Sciences Publications Ltd, 6. doi: 10.7554/eLife.25165.

Teo, A. K. K. *et al.* (2011) 'Pluripotency factors regulate definitive endoderm specification through eomesodermin', *Genes and Development*. Cold Spring Harbor Laboratory Press, 25(3), pp. 238–250. doi: 10.1101/gad.607311.

Thisse, C., and Thisse, B. (2008) 'Expression from: Unexpected Novel Relational Links Uncovered by Extensive Developmental Profiling of Nuclear Receptor Expression', *ZFIN Direct Data Submission*. Available at: <http://zfin.org/ZDB-PUB-010810-1> (Accessed: 20 January 2022).

Thisse, B. *et al.* (2001) 'Expression of the zebrafish genome during embryogenesis (NIH R01 RR15402-01)', *ZFIN Direct Data Submission*. Available at: <https://zfin.org/ZDB-PUB-010810-1> (Accessed: 29 April 2020).

Thisse and Thisse (2005) 'High Throughput Expression Analysis of ZF-Models Consortium Clones.', *ZFIN Direct Data Submission (<http://zfin.org>)*.

Thomas, I. H. *et al.* (2009) 'Neonatal diabetes mellitus with pancreatic agenesis in an infant with homozygous IPF-1 Pro63fsX60 mutation', *Pediatric Diabetes*. John Wiley & Sons, Ltd (10.1111), 10(7), pp. 492–496. doi: 10.1111/j.1399-5448.2009.00526.x.

Thomas, P. D., Kejariwal, A., *et al.* (2003) 'PANTHER: A browsable database of gene products organized by biological function, using curated protein family and subfamily classification', *Nucleic Acids Research*. Nucleic Acids Res, pp. 334–341. doi: 10.1093/nar/gkg115.

Thomas, P. D., Campbell, M. J., *et al.* (2003) 'PANTHER: A library of protein families and subfamilies indexed by function', *Genome Research*. Genome Res, 13(9), pp. 2129–2141. doi: 10.1101/gr.772403.

Thorvaldsdóttir, H., Robinson, J. T. and Mesirov, J. P. (2013) 'Integrative Genomics Viewer (IGV): High-performance genomics data visualization and exploration', *Briefings in*

Bioinformatics. Oxford University Press, 14(2), pp. 178–192.

Thurman, R. E. *et al.* (2012) 'The accessible chromatin landscape of the human genome', *Nature*. Nature Publishing Group, 489(7414), pp. 75–82. doi: 10.1038/nature11232.

Tiso, N. *et al.* (2002) 'BMP signalling regulates anteroposterior endoderm patterning in zebrafish', *Mechanisms of Development*. Mech Dev, 118(1–2), pp. 29–37. doi: 10.1016/S0925-4773(02)00252-6.

Tobia, C. *et al.* (2019) 'Atypical chemokine receptor 3 generates guidance cues for CXCL12-mediated endothelial cell migration', *Frontiers in Immunology*. Frontiers Media S.A., 10(MAY). doi: 10.3389/fimmu.2019.01092.

Ton, C. *et al.* (2000) 'Identification, characterization, and mapping of expressed sequence tags from an embryonic zebrafish heart cDNA library', *Genome Research*, 10(12), pp. 1915–1927. doi: 10.1101/gr.10.12.1915.

Tong Ihn Lee and Young, R. A. (2000) 'Transcription of eukaryotic protein-coding genes', *Annual Review of Genetics*, 34, pp. 77–137. doi: 10.1146/annurev.genet.34.1.77.

Tremblay, K. D. (2010) 'Formation of the murine endoderm: Lessons from the mouse, frog, fish, and chick', in *Progress in Molecular Biology and Translational Science*. Elsevier B.V., pp. 1–34. doi: 10.1016/B978-0-12-381280-3.00001-4.

Tremblay, M., Sanchez-Ferras, O. and Bouchard, M. (2018) 'Gata transcription factors in development and disease', *Development (Cambridge)*, 145(20). doi: 10.1242/dev.164384.

Trinkaus, J. P. (1993) 'The yolk syncytial layer of *Fundulus*: Its origin and history and its significance for early embryogenesis', *Journal of Experimental Zoology*, 265(3), pp. 258–284. doi: 10.1002/jez.1402650308.

Tsankov, A. M. *et al.* (2015) 'Transcription factor binding dynamics during human ES cell differentiation', *Nature*, 518(7539), pp. 344–349. doi: 10.1038/nature14233.

Tseng, W. F. *et al.* (2011) 'An evolutionarily conserved kernel of *gata5*, *gata6*, *otx2* and *prdm1a* operates in the formation of endoderm in zebrafish', *Developmental Biology*. Academic Press Inc., 357(2), pp. 541–557. doi: 10.1016/j.ydbio.2011.06.040.

Tsompana, M. and Buck, M. J. (2014) 'Chromatin accessibility: A window into the genome', *Epigenetics and Chromatin*. BioMed Central Ltd. doi: 10.1186/1756-8935-7-33.

Vandermeer, J. E. *et al.* (2014) 'Genome-wide identification of signaling center enhancers in the developing limb', *Development (Cambridge)*, 141(21), pp. 4194–4198. doi: 10.1242/dev.110965.

Vannier, C. *et al.* (2013) 'Zeb1 regulates E-cadherin and Epcam (epithelial cell adhesion molecule) expression to control cell behavior in early zebrafish development', *Journal of Biological Chemistry*. Elsevier, 288(26), pp. 18643–18659. doi: 10.1074/jbc.M113.467787.

Varlet, I., Collignon, J. and Robertson, E. J. (1997) *Nodal expression in the primitive endoderm is required for specification of the anterior axis during mouse gastrulation*, *Development*. doi:

10.1242/dev.124.5.1033.

Veil, M. *et al.* (2018) 'Maternal nanog is required for zebrafish embryo architecture and for cell viability during gastrulation', *Development (Cambridge)*. Company of Biologists Ltd, 145(1). doi: 10.1242/dev.155366.

Velazco-Cruz, L. *et al.* (2020) 'SIX2 Regulates Human β Cell Differentiation from Stem Cells and Functional Maturation In Vitro', *Cell Reports*. Cell Press, 31(8), p. 107687. doi: 10.1016/j.celrep.2020.107687.

Viotti, M. *et al.* (2012) 'Role of the gut endoderm in relaying left-right patterning in mice', *PLoS Biology*. Edited by B. L. M. Hogan. Public Library of Science, 10(3), p. e1001276. doi: 10.1371/journal.pbio.1001276.

Viotti, M., Foley, A. C. and Hadjantonakis, A.-K. (2014) 'Gutsy moves in mice: cellular and molecular dynamics of endoderm morphogenesis', *Philosophical Transactions of the Royal Society B: Biological Sciences*. The Royal Society, 369(1657), pp. 20130547–20130547. doi: 10.1098/rstb.2013.0547.

Visel, A. *et al.* (2008) 'Ultraconservation identifies a small subset of extremely constrained developmental enhancers', *Nature Genetics*. Nat Genet, 40(2), pp. 158–160. doi: 10.1038/ng.2007.55.

Visel, A. *et al.* (2009) 'ChIP-seq accurately predicts tissue-specific activity of enhancers', *Nature*. Nature Publishing Group, 457(7231), pp. 854–858. doi: 10.1038/nature07730.

Visel, A. *et al.* (2013) 'A high-resolution enhancer atlas of the developing telencephalon', *Cell*. Cell Press, 152(4), pp. 895–908. doi: 10.1016/j.cell.2012.12.041.

Voldoire, E. *et al.* (2017) 'Expansion by whole genome duplication and evolution of the sox gene family in teleost fish', *PLoS ONE*. Public Library of Science, 12(7), p. e0180936. doi: 10.1371/journal.pone.0180936.

Voong, L. N. *et al.* (2017) 'Genome-wide Mapping of the Nucleosome Landscape by Micrococcal Nuclease and Chemical Mapping', *Trends in Genetics*, 33(8), pp. 495–507. doi: 10.1016/j.tig.2017.05.007.

Wagner, D. E. *et al.* (2018) 'Single-cell mapping of gene expression landscapes and lineage in the zebrafish embryo', *Science*. American Association for the Advancement of Science, 360(6392), pp. 981–987. doi: 10.1126/science.aar4362.

Wallace, K. N. and Pack, M. (2003) 'Unique and conserved aspects of gut development in zebrafish', *Developmental Biology*. Academic Press, 255(1), pp. 12–29. doi: 10.1016/S0012-1606(02)00034-9.

Wang, A. *et al.* (2015) 'Epigenetic priming of enhancers predicts developmental competence of hESC-derived endodermal lineage intermediates', *Cell Stem Cell*. Cell Press, 16(4), pp. 386–399. doi: 10.1016/j.stem.2015.02.013.

Wang, G. *et al.* (2011) 'The Rho kinase rock2b establishes anteroposterior asymmetry of the

ciliated Kupffer's vesicle in zebrafish', *Development*. Oxford University Press for The Company of Biologists Limited, 138(1), pp. 45–54. doi: 10.1242/dev.052985.

Wang, G., Manning, M. L. and Amack, J. D. (2012) 'Regional cell shape changes control form and function of Kupffer's vesicle in the zebrafish embryo', *Developmental Biology*, 370(1), pp. 52–62. doi: 10.1016/j.ydbio.2012.07.019.

Wangler, M. F., Yamamoto, S. and Bellen, H. J. (2015) 'Fruit flies in biomedical research', *Genetics*. *Genetics*, 199(3), pp. 639–653. doi: 10.1534/genetics.114.171785.

Warga and Kane (2018) 'Wilson cell origin for kupffer's vesicle in the zebrafish', *Developmental Dynamics*. John Wiley and Sons Inc., 247(9), pp. 1057–1069. doi: 10.1002/dvdy.24657.

Warga and Nüsslein-Volhard, C. (1999) 'Origin and development of the zebrafish endoderm', *Development*, 126(4), pp. 827–838. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/9895329> (Accessed: 23 November 2019).

Warga, R. M. *et al.* (2013) 'Zebrafish Tbx16 regulates intermediate mesoderm cell fate by attenuating Fgf activity', *Developmental Biology*. Academic Press, 383(1), pp. 75–89. doi: 10.1016/j.ydbio.2013.08.018.

Warga, R. M. and Kimmel, C. B. (1990) *Cell movements during epiboly and gastrulation in zebrafish*, *Development*. Available at: <https://dev.biologists.org/content/develop/108/4/569.full.pdf> (Accessed: 24 November 2019).

Warga, R. M. and Nüsslein-Volhard, C. (1999) *Zebrafish endoderm development*, *Development*. Available at: <https://dev.biologists.org/content/develop/126/4/827.full.pdf> (Accessed: 24 November 2019).

Wassef, M. and Margueron, R. (2017) 'Regulation of Cellular Identity by Polycomb and Trithorax Proteins', in *Chromatin Regulation and Dynamics*. Elsevier Inc., pp. 165–189. doi: 10.1016/B978-0-12-803395-1.00007-1.

Wasserman, W. W. *et al.* (2000) 'Human-mouse genome comparisons to locate regulatory sites', *Nature Genetics*. Nature Publishing Group, 26(2), pp. 225–228. doi: 10.1038/79965.

Waterston, R. H. *et al.* (2002) 'Initial sequencing and comparative analysis of the mouse genome', *Nature*, 420(6915), pp. 520–562. doi: 10.1038/nature01262.

Weedon, M. N. *et al.* (2014) 'Recessive mutations in a distal PTF1A enhancer cause isolated pancreatic agenesis', *Nature Genetics*. Nature Publishing Group, 46(1), pp. 61–64. doi: 10.1038/ng.2826.

Wei, Z. *et al.* (2018) 'esATAC: an easy-to-use systematic pipeline for ATAC-seq data analysis', *Bioinformatics*. Edited by B. Berger. Oxford University Press, 34(15), pp. 2664–2665.

Weigel, D. *et al.* (1989) 'The homeotic gene fork head encodes a nuclear protein and is expressed in the terminal regions of the Drosophila embryo', *Cell*, 57(4), pp. 645–658. doi: 10.1016/0092-8674(89)90133-5.

Weinstein *et al.* (1994) 'The winged-helix transcription factor HNF-3 β is required for notochord development in the mouse embryo', *Cell*, 78(4), pp. 575–588. doi: 10.1016/0092-8674(94)90523-1.

Westerfield, M. (1995) 'The zebrafish book: a guide for the laboratory use of zebrafish', *University of Oregon Press, Eugene, OR*, 40(2), pp. 269–274. doi: 10.2527/jas1975.402269x.

White, R. J. *et al.* (2017) 'A high-resolution mRNA expression time course of embryonic development in zebrafish', *eLife*, 6. doi: 10.7554/eLife.30860.

Whittington, N. *et al.* (2015) 'Sox21 regulates the progression of neuronal differentiation in a dose-dependent manner', *Developmental Biology*. NIH Public Access, 397(2), pp. 237–247. doi: 10.1016/j.ydbio.2014.11.012.

Wiebe, P. O. *et al.* (2007) 'Ptf1a Binds to and Activates Area III, a Highly Conserved Region of the Pdx1 Promoter That Mediates Early Pancreas-Wide Pdx1 Expression', *Molecular and Cellular Biology*, 27(11), pp. 4093–4104. doi: 10.1128/mcb.01978-06.

Williams, D. W. *et al.* (1996) 'High transgene activity in the yolk syncytial layer affects quantitative transient expression assays in zebrafish (*Danio rerio*) embryos', *Transgenic Research*. Kluwer Academic Publishers, 5(6), pp. 433–442. doi: 10.1007/BF01980208.

Wills, A. *et al.* (2008) 'Bmp signaling is necessary and sufficient for ventrolateral endoderm specification in *Xenopus*', *Developmental Dynamics*. Dev Dyn, 237(8), pp. 2177–2186. doi: 10.1002/dvdy.21631.

Wilson, V. and Conlon, F. L. (2002) 'The T-box family', *Genome Biology*. BioMed Central, p. reviews3008.1. doi: 10.1186/gb-2002-3-6-reviews3008.

Winata, C. L. *et al.* (2013) 'Genome Wide Analysis Reveals Zic3 Interaction with Distal Regulatory Elements of Stage Specific Developmental Genes in Zebrafish', *PLoS Genetics*. Public Library of Science, 9(10), p. e1003852. doi: 10.1371/journal.pgen.1003852.

Windner, S. E. *et al.* (2015) 'Tbx6, Mesp-b and rippy1 regulate the onset of skeletal myogenesis in zebrafish', *Development (Cambridge)*, 142(6), pp. 1159–1168. doi: 10.1242/dev.113431.

Wlizla, M. and Zorn, A. M. (2014) 'Vertebrate Endoderm Formation', in *Principles of Developmental Genetics: Second Edition*. Elsevier, pp. 237–253. doi: 10.1016/B978-0-12-405945-0.00013-2.

Woolfe, A. *et al.* (2005) 'Highly conserved non-coding sequences are associated with vertebrate development', *PLoS Biology*. PLoS Biol, 3(1). doi: 10.1371/journal.pbio.0030007.

Wu, C. (1980) 'The 5' ends of drosophila heat shock genes in chromatin are hypersensitive to DNase I', *Nature*, 286(5776), pp. 854–860. doi: 10.1038/286854a0.

Xiong, N., Kang, C. and Raulet, D. H. (2002) 'Redundant and unique roles of two enhancer elements in the TCR γ locus in gene regulation and $\gamma\delta$ T cell development', *Immunity*, 16(3), pp. 453–463. doi: 10.1016/S1074-7613(02)00285-6.

- Xu, C. *et al.* (2012) 'Nanog-like Regulates Endoderm Formation through the Mxtx2-Nodal Pathway', *Developmental Cell*, 22(3), pp. 625–638. doi: 10.1016/j.devcel.2012.01.003.
- Xu, P. *et al.* (2014) 'Maternal eomesodermin regulates zygotic nodal gene expression for mesendoderm induction in zebrafish embryos', *Journal of Molecular Cell Biology*, 6(4), pp. 272–285. doi: 10.1093/jmcb/mju028.
- Xue, C. *et al.* (2018) 'The expression patterns of vestigial like family member 4 genes in zebrafish embryogenesis', *Gene Expression Patterns*. Elsevier, 28, pp. 34–41. doi: 10.1016/j.gep.2018.02.001.
- Yan, F. *et al.* (2020) 'From reads to insight: A hitchhiker's guide to ATAC-seq data analysis', *Genome Biology*. BioMed Central, pp. 1–16. doi: 10.1186/s13059-020-1929-3.
- Yang, H. *et al.* (2002) *GATA6 regulates differentiation of distal lung epithelium*, *Development*. Available at: <https://pubmed.ncbi.nlm.nih.gov/11959831/> (Accessed: 6 February 2021).
- Yang, H. *et al.* (2020) 'A map of cis-regulatory elements and 3D genome structures in zebrafish', *Nature*. NIH Public Access, 588(7837), pp. 337–343. doi: 10.1038/s41586-020-2962-9.
- Yeh, L. K. *et al.* (2008) 'Molecular analysis and characterization of zebrafish Keratocan (zKera) gene', *Journal of Biological Chemistry*, 283(1), pp. 506–517. doi: 10.1074/jbc.M707656200.
- Yeo, S. Y. and Chitnis, A. B. (2007) 'Jagged-mediated Notch signaling maintains proliferating neural progenitors and regulates cell diversity in the ventral spinal cord', *Proceedings of the National Academy of Sciences of the United States of America*, 104(14), pp. 5913–5918. doi: 10.1073/pnas.0607062104.
- Yiangou, L. *et al.* (2018) 'Human Pluripotent Stem Cell-Derived Endoderm for Modeling Development and Clinical Applications', *Cell Stem Cell*, pp. 485–499. doi: 10.1016/j.stem.2018.03.016.
- Yu, G., Wang, L. G. and He, Q. Y. (2015) 'ChIP seeker: An R/Bioconductor package for ChIP peak annotation, comparison and visualization', *Bioinformatics*, 31(14), pp. 2382–2383. doi: 10.1093/bioinformatics/btv145.
- Yu, X. *et al.* (2008) 'Foxj1 transcription factors are master regulators of the motile ciliogenic program', *Nature Genetics*, 40(12), pp. 1445–1453. doi: 10.1038/ng.263.
- Yuh, C. H., Dorman, E. R. and Davidson, E. H. (2005) 'Brn1/2/4, the predicted midgut regulator of the endo16 gene of the sea urchin embryo', *Developmental Biology*. Academic Press Inc., 281(2), pp. 286–298. doi: 10.1016/j.ydbio.2005.02.034.
- Zaret, K. S. *et al.* (2008) 'Pioneer factors, genetic competence, and inductive signaling: Programming liver and pancreas progenitors from the endoderm', *Cold Spring Harbor Symposia on Quantitative Biology*. Cold Spring Harbor Laboratory Press, 73, pp. 119–126. doi: 10.1101/sqb.2008.73.040.
- Zhang, P. *et al.* (2002) 'Expression of COUP-TFII in metabolic tissues during development',

Mechanisms of Development. Elsevier, 119(1), pp. 109–114. doi: 10.1016/S0925-4773(02)00286-1.

Zhang, Y. *et al.* (2008) 'Model-based analysis of ChIP-Seq (MACS)', *Genome Biology*. BioMed Central, 9(9), p. R137.

Zhang, Y. and Hou, L. (2021) 'Alternate roles of sox transcription factors beyond transcription initiation', *International Journal of Molecular Sciences*. Multidisciplinary Digital Publishing Institute (MDPI), 22(11). doi: 10.3390/ijms22115949.

Zhao, B. *et al.* (2007) 'Inactivation of YAP oncoprotein by the Hippo pathway is involved in cell contact inhibition and tissue growth control', *Genes and Development*. Cold Spring Harbor Laboratory Press, 21(21), pp. 2747–2761. doi: 10.1101/gad.1602907.

Zhao, B. *et al.* (2008) 'TEAD mediates YAP-dependent gene induction and growth control', *Genes and Development*. Cold Spring Harbor Laboratory Press, 22(14), pp. 1962–1971. doi: 10.1101/gad.1664408.

Zhao, F. Q. (2013) 'Octamer-binding transcription factors: Genomics and functions', *Frontiers in Bioscience*. NIH Public Access, 18(3), pp. 1051–1071. doi: 10.2741/4162.

Zhao, J. *et al.* (2013) 'The transcription factor Vox represses endoderm development by interacting with casanova and Pou2', *Development (Cambridge)*, 140(5), pp. 1090–1099. doi: 10.1242/dev.082008.

Zhou, V. W., Goren, A. and Bernstein, B. E. (2011) 'Charting histone modifications and the functional organization of mammalian genomes', *NATurE rEvIEWS | Genetics*, 12(7). doi: 10.1038/nrg2905.

Zhu, Q. *et al.* (2014) 'The transcription factor Pou3f1 promotes neural fate commitment via activation of neural lineage genes and inhibition of external signaling pathways', *eLife*. eLife Sciences Publications Ltd, 2014(3). doi: 10.7554/eLife.02224.

Zon, L. I. (1999) 'Zebrafish: A new model for human disease', *Genome Research*, 9(2), pp. 99–100. doi: 10.1101/gr.9.2.99.

Zorn, A. M. and Wells, J. M. (2007) 'Molecular Basis of Vertebrate Endoderm Development', *International Review of Cytology*, pp. 49–111. doi: 10.1016/S0074-7696(06)59002-3.

Zorn, A. M. and Wells, J. M. (2009) 'Vertebrate Endoderm Development and Organ Formation', *Annual Review of Cell and Developmental Biology*. NIH Public Access, 25(1), pp. 221–251. doi: 10.1146/annurev.cellbio.042308.113344.

Zou, D. *et al.* (2006) 'Patterning of the third pharyngeal pouch into thymus/parathyroid by Six and Eya1', *Developmental Biology*. NIH Public Access, 293(2), pp. 499–512. doi: 10.1016/j.ydbio.2005.12.015.

Chapter 7 Supplementary

Table 7.1: List of DARs between *sox32* KD and control ranked by FDR from Diffbind. Fold enrichment score > 0 represents control DARs while *sox32* KD DARs have a fold enrichment score < 0. FDR = False discovery rate

Chromosome	Chromosome start	Chromosome end	Fold enrichment	FDR
chr22	8964228	8965208	7.3	8.29E-07
chr7	50456381	50456919	5.62	0.00056
chr4	29894411	29895067	4.77	0.00189
chr4	40573379	40573850	-1.45	0.00523
chr7	2295075	2295640	1.14	0.0274
chr6	47515877	47516884	1.34	0.034

Table 7.2: List of top 200 DARs between *sox32* OE and control ranked by FDR from Diffbind. Fold enrichment score > 0 represents control DARs while *sox32* OE DARs have a fold enrichment score < 0. FDR = False discovery rate

Chromosome	Chromosome start	Chromosome end	Fold enrichment	FDR
chr1	57856552	57857972	-4.75	2.81E-68
chr1	15984509	15986246	-4.4	6.18E-58
chr7	47085558	47086697	-4.13	6.18E-56
chr23	30158932	30159819	-3.88	1.59E-55
chr5	3371277	3372032	-3.58	1.09E-48
chr20	11153454	11154577	-3.67	9.09E-47
chr16	5523029	5523870	-3.78	2.62E-46
chr7	46696288	46698694	-3.69	5.35E-46
chr23	36383051	36384179	-3.41	1.76E-44
chr12	26174554	26176255	-3.51	1.61E-43
chr4	18767736	18769742	-3.59	6.10E-43
chr9	5111669	5112876	-3.74	6.76E-43
chr20	11163913	11164405	-3.41	6.89E-43
chr9	14017037	14017603	-3.29	3.78E-42
chr24	28361171	28362235	-3.25	7.58E-42
chr20	28565855	28566819	-3.06	1.05E-41
chr19	37659459	37660141	-3.33	3.78E-41
chr3	29478796	29479678	-3.43	4.02E-41

chr20	22709025	22709939	-3.37	7.33E-41
chr10	20276616	20277702	-3.35	1.51E-40
chr16	41657071	41657976	-2.95	2.75E-39
chr10	20978013	20978651	-3.41	3.68E-39
chr22	27081231	27081955	-3.84	4.81E-39
chr6	8974336	8975696	-3.25	7.76E-39
chr3	34059698	34060741	-3.1	8.53E-39
chr24	4293023	4294131	-3.32	1.19E-38
chr2	22188169	22189044	-3.35	2.07E-38
chr6	8941348	8942186	-3.18	1.40E-37
chr5	43842200	43842804	-3.24	2.10E-37
chr6	41298007	41298707	-3.37	4.97E-37
chr17	19505893	19507041	-3.29	4.97E-37
chr12	6273593	6274624	-2.96	6.05E-37
chr19	3884642	3885077	-3.49	7.26E-37
chr20	29662794	29663794	-2.81	1.21E-36
chr11	44170324	44171465	-3.32	2.27E-36
chr10	17008549	17009395	-3.05	3.40E-36
chr12	20382569	20383160	-3.08	5.41E-36
chr8	29477862	29478679	-3.4	6.33E-36
chr13	6555907	6556917	-3.14	1.40E-35
chr3	43077327	43078228	-2.88	1.40E-35
chr11	36260510	36261503	-3.16	2.79E-35
chr6	19328081	19329055	-3.46	5.05E-35
chr13	7981392	7982670	-3	1.12E-34
chr2	28996805	28997622	-2.84	1.13E-34
chr1	51153321	51154105	-2.81	1.41E-34
chr4	19247511	19248181	-2.98	2.69E-34
chr20	33695146	33696276	-3.43	3.90E-34
chr12	21703731	21704586	-3.33	5.40E-34
chr25	14492841	14494448	-3.2	1.41E-33
chr10	21221574	21223065	-3.06	2.29E-33
chr7	27313533	27314018	-3.01	3.55E-33
chr18	40283722	40284544	-3.02	7.96E-33
chr23	30150056	30150497	-2.84	1.13E-32

chr9	16201898	16202442	-3.24	1.16E-32
chr16	41144607	41144858	-3.62	1.23E-32
chr10	4987001	4988397	-3.02	2.80E-32
chr4	14381886	14382365	-3.01	3.79E-32
chr14	22652263	22654363	-2.68	3.79E-32
chr6	14065306	14066829	-3.19	5.04E-32
chr19	8450691	8452040	-3.06	7.85E-32
chr16	42618563	42619244	-2.92	9.02E-32
chr20	20999858	21001548	-2.98	1.21E-31
chr8	17753875	17754710	-3.07	1.39E-31
chr18	44463170	44464408	-3.45	1.63E-31
chr21	38042360	38043207	-2.75	1.99E-31
chr2	29234371	29235345	-2.69	2.61E-31
chr4	11552377	11554364	-2.74	2.92E-31
chr18	2618812	2619302	-3	5.00E-31
chr24	29311032	29312051	-2.8	8.59E-31
chr18	49261119	49262419	-2.67	9.92E-31
chr13	42844902	42845632	-3.24	9.92E-31
chr9	14060909	14062218	-2.81	1.64E-30
chr25	32173817	32174740	-2.98	1.69E-30
chr13	23423046	23424030	-3.05	2.29E-30
chr11	19695482	19696042	-2.59	2.29E-30
chr20	14263267	14264236	-2.88	2.80E-30
chr12	4087320	4088286	-2.59	4.05E-30
chr5	8256189	8257022	-4.92	5.28E-30
chr2	28995364	28996461	-2.8	6.96E-30
chr16	9040591	9041297	-2.66	6.99E-30
chr14	22009904	22010916	-2.58	7.40E-30
chr10	20954625	20955993	-3.15	7.52E-30
chr15	36653609	36654053	-3.03	1.40E-29
chr20	33683489	33684175	-3.45	2.01E-29
chr11	26806051	26806755	-2.82	2.27E-29
chr7	26905834	26907002	-2.92	2.45E-29
chr7	58770339	58775013	-3.03	2.81E-29
chr14	19375948	19376924	-2.92	3.04E-29

chr13	26572869	26573710	-2.55	3.52E-29
chr9	10451315	10451710	-2.83	6.93E-29
chr4	6318266	6319516	-2.74	7.68E-29
chr15	19295707	19296676	-2.68	8.00E-29
chr7	40358116	40358646	-2.55	8.15E-29
chr21	15695126	15696132	-2.65	8.75E-29
chr8	32204529	32205926	-2.97	8.78E-29
chr7	26415663	26417108	-2.63	9.32E-29
chr13	36505009	36505610	-2.97	1.15E-28
chr16	22930225	22931717	-2.78	1.15E-28
chr2	39448502	39449431	-2.75	1.37E-28
chr24	40522346	40523075	-2.71	1.37E-28
chr8	9891047	9891644	-2.91	2.42E-28
chr13	40270760	40271390	-3.25	2.47E-28
chr4	24348609	24349177	-3.06	2.70E-28
chr2	29549950	29550680	-2.67	3.17E-28
chr23	8007395	8008427	-2.62	3.99E-28
chr6	19324490	19325472	-2.83	5.09E-28
chr6	42573421	42574079	-2.52	5.71E-28
chr10	28673409	28676728	-2.78	7.74E-28
chr14	20575898	20576649	-2.81	8.49E-28
chr15	17182318	17183700	-2.39	1.07E-27
chr18	41758993	41760049	-2.65	1.74E-27
chr13	39185158	39186656	-2.75	1.80E-27
chr6	41163492	41164282	-2.87	1.85E-27
chr20	15463911	15464683	-2.88	3.60E-27
chr10	23029457	23030312	-2.71	3.86E-27
chr24	32600743	32601761	-2.77	3.99E-27
chr1	26991338	26992483	-3	5.36E-27
chr1	9423456	9424693	-2.51	6.14E-27
chr15	36364932	36365837	-2.4	6.50E-27
chr20	34076077	34077221	-2.69	1.14E-26
chr20	23343353	23344559	-2.61	1.14E-26
chr3	34053219	34055524	-2.46	1.18E-26
chr1	17258151	17258699	-2.86	1.43E-26

chr7	58790590	58791288	-2.66	1.52E-26
chr7	47056547	47057401	-2.76	1.66E-26
chr10	17862545	17863307	-3.11	1.86E-26
chr23	21072955	21075896	-2.58	1.89E-26
chr16	46689150	46689911	-2.33	2.05E-26
chr13	23369232	23369905	-2.79	2.09E-26
chr8	30174857	30175265	-2.52	2.12E-26
chr7	37061961	37062503	-2.64	2.27E-26
chr21	42692847	42693879	-2.59	3.28E-26
chr10	43923480	43924001	-4.63	3.72E-26
chr4	20975306	20976806	-2.94	4.34E-26
chr6	50297650	50298586	-2.75	4.40E-26
chr9	17621800	17622446	-2.45	4.60E-26
chr19	2432857	2433299	-2.93	6.64E-26
chr6	11134488	11136495	-2.88	7.47E-26
chr23	36384702	36385621	-2.34	9.78E-26
chr9	5059791	5060451	-3.09	1.11E-25
chr20	16817381	16817933	-2.61	1.14E-25
chr18	10358434	10358921	-3.01	1.49E-25
chr17	3305358	3305748	-2.45	1.65E-25
chr25	3391426	3392580	-2.27	1.78E-25
chr9	16170945	16171911	-2.46	2.15E-25
chr15	46361580	46362484	-2.48	2.22E-25
chr14	17952755	17953240	-2.91	3.43E-25
chr13	26825570	26825904	-2.72	3.64E-25
chr7	19654657	19655263	-2.71	5.06E-25
chr16	22098019	22099435	-3.18	5.40E-25
chr15	64238	64574	-2.85	6.56E-25
chr7	71706010	71707506	-2.45	6.69E-25
chr25	6032976	6034199	-2.54	6.89E-25
chr8	27383288	27384587	-2.75	6.97E-25
chr7	40421307	40422215	-2.63	7.17E-25
chr12	40909313	40910432	-2.32	7.22E-25
chr6	28304969	28305739	-2.72	8.20E-25
chr25	33117268	33118467	-2.63	1.08E-24

chr1	19277757	19278475	-2.45	1.33E-24
chr5	5517012	5518050	-2.29	1.39E-24
chr2	29548585	29549614	-2.59	2.05E-24
chr11	35780678	35781236	-2.44	2.12E-24
chr5	21738747	21739330	-2.98	2.16E-24
chr1	16332743	16333613	-2.51	2.45E-24
chr15	15900631	15901743	-2.66	2.50E-24
chr6	26600951	26602250	-2.26	3.01E-24
chr6	42605255	42605966	-3.94	4.81E-24
chr3	55569374	55570788	-2.33	4.94E-24
chr14	25178296	25179042	-2.34	4.98E-24
chr18	26530769	26531664	-2.45	6.23E-24
chr17	1567573	1568143	-2.56	6.24E-24
chr6	42614324	42615970	-3.11	6.36E-24
chr17	22340129	22341205	-2.56	8.16E-24
chr23	7349508	7350757	-2.55	8.33E-24
chr24	36490147	36490834	-2.63	9.49E-24
chr20	32524693	32525632	-2.79	1.02E-23
chr10	17393495	17394283	-2.48	1.07E-23
chr25	33753465	33754187	-2.81	1.13E-23
chr15	15462080	15462714	-2.4	1.25E-23
chr15	20671448	20671953	-2.57	1.48E-23
chr11	32011969	32012829	-2.35	1.56E-23
chr3	38055860	38056507	-2.41	1.60E-23
chr23	34512557	34513184	-2.42	2.25E-23
chr21	19635479	19635790	-3.5	2.35E-23
chr10	27489339	27490301	-2.58	2.45E-23
chr21	41361215	41361897	-2.82	2.70E-23
chr6	46581333	46581953	-2.63	2.99E-23
chr1	33034320	33035134	-2.55	3.18E-23
chr6	57492345	57492810	-3.05	5.15E-23
chr13	40267732	40268338	-2.44	6.12E-23
chr12	34371509	34372137	-2.48	6.23E-23
chr5	18001854	18002733	-2.95	6.99E-23
chr18	20695448	20696241	-2.52	7.43E-23

chr2	48060337	48061452	-2.19	1.19E-22
chr11	1279273	1280679	-2.65	1.32E-22
chr8	9297092	9297971	-2.38	1.35E-22
chr7	46757157	46757514	-2.33	1.58E-22
chr16	41748436	41748915	-2.3	1.60E-22
chr7	28211459	28212556	-2.38	1.79E-22

Table 7.3: A list of 25 genes that are found proximal to *sox32* OE vs control DARs. These genes were attained from GREAT by first converting *sox32* OE DARs to danRer7 genome build using LiftOver. TSS = transcriptional start site























Chromosome	Chromosome start	Chromosome end	Gene (distance to TSS)
chr1	524088	525560	appa (-16770)
chr1	69706	71745	grtp1a (-4239)
chr1	109456	110407	tfdp1b (+7959)
chr1	131949	132742	cenpe (-2384)
chr1	268516	269148	rasa3 (+17190)
Zv9_NA908	1819	2573	NONE
chr1	848108	848774	creg1 (+72190)
chr1	1188085	1188807	urp2 (-111398)
chr1	1192653	1193099	urp2 (-106968)
chr1	1201942	1202869	urp2 (-97438)
chr1	1204003	1204589	urp2 (-95548)
chr1	1775770	1776603	si:ch211-132g1.5 (-5684)
chr1	2062216	2062860	farp1 (-15632)
chr1	2064553	2065428	farp1 (-13179)
chr1	2069530	2070134	farp1 (-8338)
chr1	2148893	2149248	ggact.3 (+57110)
chr1	4141736	4143004	spry2 (-28549)
chr1	4183462	4184116	spry2 (+12870)
chr1	4202076	4202655	spry2 (+31447)
chr1	4238612	4239127	tuba8l2 (-52760)
chr1	4354719	4355555	mnx2b (-10481)
chr1	4357305	4358025	mnx2b (-13009)
chr1	4600032	4600970	nrp2a (-34737)
chr1	4620717	4622088	nrp2a (-55639)

chr1	4639172	4640023	nrp2a (-73834)
------	---------	---------	----------------

Table 7.4: A list of 25 genes that are found proximal to control vs *sox32* OE DARs. These genes were attained from GREAT by first converting control DARs to danRer7 genome build using LiftOver. TSS = transcriptional start site

Chromosome	Chromosome start	Chromosome end	Gene (distance to TSS)
chr1	473130	473974	jam2a (+4119)
chr1	1326118	1326724	si:ch73-272o12.2 (+5245)
chr1	4465369	4465768	cpo (+674)
chr1	5372109	5372366	klf7a (-14646)
chr1	5942776	5943339	erbb4a (+248103)
chr1	7787615	7788058	zgc:77849 (+17557)
chr1	8693301	8693806	ugt5b3 (+24375)
chr1	9626994	9627356	tnrc5 (+22558)
chr1	10123030	10123563	dmd (-54407)
chr1	11044844	11045065	ttyh3b (-13479)
chr1	11896642	11897221	zgc:77739 (+206322)
chr2	25063969	25064356	nkeh1a (-154552)
chr2	25080341	25081102	si:dkey-5n7.2 (-149616)
chr2	25119050	25119674	si:dkey-5n7.2 (-110976)
chr2	25140544	25140885	si:dkey-5n7.2 (-89623)
chr2	25262737	25263635	si:dkey-5n7.2 (+32848)
chr2	25308278	25308997	si:dkey-5n7.2 (+78300)
chr2	25311851	25312866	ppp2r3a (-81472)
chr1	12162327	12162901	pcdh18a (-6742)
chr1	12173536	12174116	pcdh18a (-17954)
chr1	12241968	12242686	pcdh18a (-86455)
chr1	15261370	15262327	mtmr7b (-12170)
chr1	15313292	15313710	pcm1 (-54072)
chr1	15490382	15490879	zgc:112426 (-59862)
chr1	17123671	17123988	mtnr1aa (-45057)



























Table 7.5: TF binding motifs found over-represented in *sox32* OE vs control DARs. Motifs are ranked by statistical significance.

Rank	Motif	Name	P-value	log P-value	q-value (Benjamini)	# Sw
1		Tbx21(T-box)/GM12878-TBX21-ChIP-Seq(Encode)/Homer	1e-252	-5.805e+02	0.0000	2
2		Tbx6(T-box)/ESC-Tbx6-ChIP-Seq(GSE93524)/Homer	1e-241	-5.557e+02	0.0000	2
3		Tbet(T-box)/CD8-Tbet-ChIP-Seq(GSE33802)/Homer	1e-237	-5.457e+02	0.0000	2
4		Eomes(T-box)/H9-Eomes-ChIP-Seq(GSE26097)/Homer	1e-189	-4.356e+02	0.0000	4
5		Tbr1(T-box)/Cortex-Tbr1-ChIP-Seq(GSE71384)/Homer	1e-178	-4.114e+02	0.0000	3
6		Tbx5(T-box)/HL1-Tbx5.biotin-ChIP-Seq(GSE21529)/Homer	1e-140	-3.244e+02	0.0000	4
7		FOXA1(Forkhead)/MCF7-FOXA1-ChIP-Seq(GSE26831)/Homer	1e-114	-2.635e+02	0.0000	2
8		FOXA1(Forkhead)/LNCAP-FOXA1-ChIP-Seq(GSE27824)/Homer	1e-106	-2.452e+02	0.0000	3
9		FOXM1(Forkhead)/MCF7-FOXM1-ChIP-Seq(GSE72977)/Homer	1e-98	-2.268e+02	0.0000	2
10		Brn1(POU,Homeobox)/NPC-Brn1-ChIP-Seq(GSE35496)/Homer	1e-97	-2.244e+02	0.0000	9
11		Sox10(HMG)/SciaticNerve-Sox3-ChIP-Seq(GSE35132)/Homer	1e-94	-2.182e+02	0.0000	2
12		Sox3(HMG)/NPC-Sox3-ChIP-Seq(GSE33059)/Homer	1e-91	-2.108e+02	0.0000	2
13		Oct6(POU,Homeobox)/NPC-Pou3f1-ChIP-Seq(GSE35496)/Homer	1e-89	-2.066e+02	0.0000	1
14		Sox17(HMG)/Endoderm-Sox17-ChIP-Seq(GSE61475)/Homer	1e-87	-2.017e+02	0.0000	1
15		Foxa2(Forkhead)/Liver-Foxa2-ChIP-Seq(GSE25694)/Homer	1e-85	-1.971e+02	0.0000	2
16		Oct11(POU,Homeobox)/NCIH1048-POU2F3-ChIP-seq(GSE115123)/Homer	1e-78	-1.812e+02	0.0000	8
17		Gata2(Zf)/K562-GATA2-ChIP-Seq(GSE18829)/Homer	1e-76	-1.759e+02	0.0000	1
18		Gata4(Zf)/Heart-Gata4-ChIP-Seq(GSE35151)/Homer	1e-72	-1.679e+02	0.0000	1
19		Sox21(HMG)/ESC-SOX21-ChIP-Seq(GSE110505)/Homer	1e-72	-1.672e+02	0.0000	2
20		GATA3(Zf)/iTreg-Gata3-ChIP-Seq(GSE20898)/Homer	1e-71	-1.643e+02	0.0000	2
21		Fox:Ebox(Forkhead,bHLH)/Panc1-Foxa2-ChIP-Seq(GSE47459)/Homer	1e-71	-1.636e+02	0.0000	2
22		Gata1(Zf)/K562-GATA1-ChIP-Seq(GSE18829)/Homer	1e-70	-1.627e+02	0.0000	1

23		Oct4:Sox17(POU,Homeobox,HMG)/F9-Sox17-ChIP-Seq(GSE44553)/Homer	1e-68	-1.569e+02	0.0000	4
24		Sox15(HMG)/CPA-Sox15-ChIP-Seq(GSE62909)/Homer	1e-66	-1.540e+02	0.0000	1
25		TRPS1(Zf)/MCF7-TRPS1-ChIP-Seq(GSE107013)/Homer	1e-66	-1.537e+02	0.0000	3
26		Gata6(Zf)/HUG1N-GATA6-ChIP-Seq(GSE51936)/Homer	1e-66	-1.534e+02	0.0000	1
27		Sox2(HMG)/mES-Sox2-ChIP-Seq(GSE11431)/Homer	1e-58	-1.346e+02	0.0000	
28		Sox6(HMG)/Myotubes-Sox6-ChIP-Seq(GSE32627)/Homer	1e-57	-1.324e+02	0.0000	2
29		Foxa3(Forkhead)/Liver-Foxa3-ChIP-Seq(GSE77670)/Homer	1e-55	-1.288e+02	0.0000	9
30		Foxo3(Forkhead)/U2OS-Foxo3-ChIP-Seq(E-MTAB-2701)/Homer	1e-55	-1.284e+02	0.0000	2
31		Oct2(POU,Homeobox)/Bcell-Oct2-ChIP-Seq(GSE21512)/Homer	1e-51	-1.195e+02	0.0000	6
32		Sox4(HMG)/proB-Sox4-ChIP-Seq(GSE50066)/Homer	1e-51	-1.192e+02	0.0000	1
33		FOXK1(Forkhead)/HEK293-FOXK1-ChIP-Seq(GSE51673)/Homer	1e-45	-1.036e+02	0.0000	2
34		Sox7(HMG)/ESC-Sox7-ChIP-Seq(GSE133899)/Homer	1e-44	-1.015e+02	0.0000	5
35		Sox9(HMG)/Limb-SOX9-ChIP-Seq(GSE73225)/Homer	1e-43	-1.011e+02	0.0000	1
36		Foxf1(Forkhead)/Lung-Foxf1-ChIP-Seq(GSE77951)/Homer	1e-42	-9.758e+01	0.0000	2
37		FoxL2(Forkhead)/Ovary-FoxL2-ChIP-Seq(GSE60858)/Homer	1e-41	-9.508e+01	0.0000	2
38		FOXP1(Forkhead)/H9-FOXP1-ChIP-Seq(GSE31006)/Homer	1e-38	-8.796e+01	0.0000	1
39		FoxD3(forkhead)/ZebrafishEmbryo-Foxd3.biotin-ChIP-seq(GSE106676)/Homer	1e-37	-8.545e+01	0.0000	1
40		Otx2(Homeobox)/EpiLC-Otx2-ChIP-Seq(GSE56098)/Homer	1e-32	-7.440e+01	0.0000	1
41		Foxo1(Forkhead)/RAW-Foxo1-ChIP-Seq(Fan_et_al.)/Homer	1e-31	-7.173e+01	0.0000	3
42		Oct4(POU,Homeobox)/mES-Oct4-ChIP-Seq(GSE11431)/Homer	1e-30	-7.052e+01	0.0000	1
43		GSC(Homeobox)/FrogEmbryos-GSC-ChIP-Seq(DRA000576)/Homer	1e-27	-6.334e+01	0.0000	1
44		FOXK2(Forkhead)/U2OS-FOXK2-ChIP-Seq(E-MTAB-2204)/Homer	1e-26	-6.197e+01	0.0000	1
45		Zic2(Zf)/ESC-Zic2-ChIP-Seq(SRP197560)/Homer	1e-26	-5.991e+01	0.0000	5
46		KLF14(Zf)/HEK293-KLF14.GFP-ChIP-Seq(GSE58341)/Homer	1e-24	-5.578e+01	0.0000	1
47		Lhx3(Homeobox)/Neuron-Lhx3-ChIP-Seq(GSE31456)/Homer	1e-23	-5.422e+01	0.0000	3
48		Isl1(Homeobox)/Neuron-Isl1-ChIP-Seq(GSE31456)/Homer	1e-21	-5.038e+01	0.0000	2

49		Zic(Zf)/Cerebellum-ZIC1.2-ChIP-Seq(GSE60731)/Homer	1e-19	-4.477e+01	0.0000	1
50		Nkx6.1(Homeobox)/Islet-Nkx6.1-ChIP-Seq(GSE40975)/Homer	1e-19	-4.446e+01	0.0000	4
51		En1(Homeobox)/SUM149-EN1-ChIP-Seq(GSE120957)/Homer	1e-17	-4.140e+01	0.0000	3
52		GATA(Zf),IR3/iTreg-Gata3-ChIP-Seq(GSE20898)/Homer	1e-17	-4.029e+01	0.0000	3
53		Brachyury(T-box)/Mesoendoderm-Brachyury-ChIP-exo(GSE54963)/Homer	1e-17	-3.943e+01	0.0000	6
54		Zic3(Zf)/mES-Zic3-ChIP-Seq(GSE37889)/Homer	1e-16	-3.798e+01	0.0000	5
55		CRX(Homeobox)/Retina-Crx-ChIP-Seq(GSE20012)/Homer	1e-16	-3.778e+01	0.0000	3
56		Foxh1(Forkhead)/hESC-FOXH1-ChIP-Seq(GSE29422)/Homer	1e-16	-3.750e+01	0.0000	1
57		Bapx1(Homeobox)/VertebralCol-Bapx1-ChIP-Seq(GSE36672)/Homer	1e-16	-3.701e+01	0.0000	2
58		DLX2(Homeobox)/BasalGanglia-Dlx2-ChIP-seq(GSE124936)/Homer	1e-16	-3.698e+01	0.0000	2
59		Nkx2.2(Homeobox)/NPC-Nkx2.2-ChIP-Seq(GSE61673)/Homer	1e-15	-3.543e+01	0.0000	2
60		MafA(bZIP)/Islet-MafA-ChIP-Seq(GSE30298)/Homer	1e-14	-3.372e+01	0.0000	1
61		c-Jun-CRE(bZIP)/K562-cJun-ChIP-Seq(GSE31477)/Homer	1e-14	-3.346e+01	0.0000	5
62		Zac1(Zf)/Neuro2A-Plagl1-ChIP-Seq(GSE75942)/Homer	1e-14	-3.328e+01	0.0000	1
63		Tgif1(Homeobox)/mES-Tgif1-ChIP-Seq(GSE55404)/Homer	1e-14	-3.263e+01	0.0000	4
64		PPARE(NR),DR1/3T3L1-Pparg-ChIP-Seq(GSE13511)/Homer	1e-13	-3.138e+01	0.0000	8
65		Meis1(Homeobox)/MastCells-Meis1-ChIP-Seq(GSE48085)/Homer	1e-13	-3.135e+01	0.0000	2
66		Pitx1(Homeobox)/Chicken-Pitx1-ChIP-Seq(GSE38910)/Homer	1e-13	-3.095e+01	0.0000	5
67		Hoxa10(Homeobox)/ChickenMSG-Hoxa10.Flag-ChIP-Seq(GSE86088)/Homer	1e-12	-2.874e+01	0.0000	9
68		Nur77(NR)/K562-NR4A1-ChIP-Seq(GSE31363)/Homer	1e-12	-2.819e+01	0.0000	3
69		Nkx2.5(Homeobox)/HL1-Nkx2.5.biotin-ChIP-Seq(GSE21529)/Homer	1e-12	-2.805e+01	0.0000	2
70		Tbx20(T-box)/Heart-Tbx20-ChIP-Seq(GSE29636)/Homer	1e-12	-2.804e+01	0.0000	3
71		CREB5(bZIP)/LNCaP-CREB5.V5-ChIP-Seq(GSE137775)/Homer	1e-12	-2.801e+01	0.0000	6
72		PPARa(NR),DR1/Liver-Ppara-ChIP-Seq(GSE47954)/Homer	1e-12	-2.779e+01	0.0000	9
73		Atf2(bZIP)/3T3L1-Atf2-ChIP-Seq(GSE56872)/Homer	1e-12	-2.771e+01	0.0000	5
74		Atf4(bZIP)/MEF-Atf4-ChIP-Seq(GSE35681)/Homer	1e-11	-2.729e+01	0.0000	5






















75		WT1(Zf)/Kidney-WT1-ChIP-Seq(GSE90016)/Homer	1e-11	-2.728e+01	0.0000	4
76		DLX1(Homeobox)/BasalGanglia-Dlx1-ChIP-seq(GSE124936)/Homer	1e-11	-2.707e+01	0.0000	2
77		Atf7(bZIP)/3T3L1-Atf7-ChIP-Seq(GSE56872)/Homer	1e-11	-2.657e+01	0.0000	7
78		Tgif2(Homeobox)/mES-Tgif2-ChIP-Seq(GSE55404)/Homer	1e-10	-2.530e+01	0.0000	4
79		Lhx2(Homeobox)/HFSC-Lhx2-ChIP-Seq(GSE48068)/Homer	1e-10	-2.508e+01	0.0000	1
80		Sp5(Zf)/mES-Sp5.Flag-ChIP-Seq(GSE72989)/Homer	1e-10	-2.484e+01	0.0000	6
81		Nanog(Homeobox)/mES-Nanog-ChIP-Seq(GSE11724)/Homer	1e-10	-2.464e+01	0.0000	6
82		JunD(bZIP)/K562-JunD-ChIP-Seq/Homer	1e-10	-2.372e+01	0.0000	2
83		Lhx1(Homeobox)/EmbryoCarcinoma-Lhx1-ChIP-Seq(GSE70957)/Homer	1e-10	-2.336e+01	0.0000	2
84		PU.1(ETS)/ThioMac-PU.1-ChIP-Seq(GSE21512)/Homer	1e-10	-2.323e+01	0.0000	4
85		LHX9(Homeobox)/Hct116-LHX9.V5-ChIP-Seq(GSE116822)/Homer	1e-10	-2.323e+01	0.0000	2
86		Nkx2.1(Homeobox)/LungAC-Nkx2.1-ChIP-Seq(GSE43252)/Homer	1e-9	-2.216e+01	0.0000	3
87		GLIS3(Zf)/Thyroid-Glis3.GFP-ChIP-Seq(GSE103297)/Homer	1e-9	-2.158e+01	0.0000	7
88		PU.1-IRF(ETS:IRF)/Bcell-PU.1-ChIP-Seq(GSE21512)/Homer	1e-8	-2.025e+01	0.0000	1
89		HNF6(Homeobox)/Liver-Hnf6-ChIP-Seq(ERP000394)/Homer	1e-8	-1.996e+01	0.0000	1
90		RXR(NR),DR1/3T3L1-RXR-ChIP-Seq(GSE13511)/Homer	1e-8	-1.975e+01	0.0000	8
91		Cux2(Homeobox)/Liver-Cux2-ChIP-Seq(GSE35985)/Homer	1e-8	-1.968e+01	0.0000	9
92		Tbox:Smad(T-box,MAD)/ESCd5-Smad2_3-ChIP-Seq(GSE29422)/Homer	1e-8	-1.948e+01	0.0000	2
93		Atf1(bZIP)/K562-ATF1-ChIP-Seq(GSE31477)/Homer	1e-7	-1.810e+01	0.0000	1
94		OCT4-SOX2-TCF-NANOG(POU,Homeobox,HMG)/mES-Oct4-ChIP-Seq(GSE11431)/Homer	1e-7	-1.761e+01	0.0000	3
95		E2F3(E2F)/MEF-E2F3-ChIP-Seq(GSE71376)/Homer	1e-7	-1.697e+01	0.0000	3
96		ETS1(ETS)/Jurkat-ETS1-ChIP-Seq(GSE17954)/Homer	1e-7	-1.671e+01	0.0000	9
97		Hoxa9(Homeobox)/ChickenMSG-Hoxa9.Flag-ChIP-Seq(GSE86088)/Homer	1e-7	-1.661e+01	0.0000	4
98		Maz(Zf)/HepG2-Maz-ChIP-Seq(GSE31477)/Homer	1e-7	-1.657e+01	0.0000	6
99		ZNF711(Zf)/SHSY5Y-ZNF711-ChIP-Seq(GSE20673)/Homer	1e-6	-1.586e+01	0.0000	9
100		ZFX(Zf)/mES-Zfx-ChIP-Seq(GSE11431)/Homer	1e-6	-1.551e+01	0.0000	6

						
101		Chop(bZIP)/MEF-Chop-ChIP-Seq(GSE35681)/Homer	1e-6	-1.530e+01	0.0000	4
102		GATA(Zf),IR4/iTreg-Gata3-ChIP-Seq(GSE20898)/Homer	1e-6	-1.517e+01	0.0000	1
103		NFIL3(bZIP)/HepG2-NFIL3-ChIP-Seq(Encode)/Homer	1e-6	-1.468e+01	0.0000	1
104		Prop1(Homeobox)/GHFT1-PROP1.biotin-ChIP-Seq(GSE77302)/Homer	1e-6	-1.427e+01	0.0000	1
105		SpiB(ETS)/OCILY3-SPIB-ChIP-Seq(GSE56857)/Homer	1e-6	-1.427e+01	0.0000	2
106		LEF1(HMG)/H1-LEF1-ChIP-Seq(GSE64758)/Homer	1e-6	-1.415e+01	0.0000	8
107		Sp2(Zf)/HEK293-Sp2.eGFP-ChIP-Seq(Encode)/Homer	1e-6	-1.400e+01	0.0000	9
108		Rbpj1(?)/Panc1-Rbpj1-ChIP-Seq(GSE47459)/Homer	1e-5	-1.379e+01	0.0000	1
109		Hoxb4(Homeobox)/ES-Hoxb4-ChIP-Seq(GSE34014)/Homer	1e-5	-1.375e+01	0.0000	3
110		p53(p53)/mES-cMyc-ChIP-Seq(GSE11431)/Homer	1e-5	-1.351e+01	0.0000	1
111		PR(NR)/T47D-PR-ChIP-Seq(GSE31130)/Homer	1e-5	-1.340e+01	0.0000	2
112		Hoxd10(Homeobox)/ChickenMSG-Hoxd10.Flag-ChIP-Seq(GSE86088)/Homer	1e-5	-1.333e+01	0.0000	1
113		Myf5(bHLH)/GM-Myf5-ChIP-Seq(GSE24852)/Homer	1e-5	-1.327e+01	0.0000	8
114		COUP-TFII(NR)/K562-NR2F1-ChIP-Seq(Encode)/Homer	1e-5	-1.319e+01	0.0000	1
115		TR4(NR),DR1/Hela-TR4-ChIP-Seq(GSE24685)/Homer	1e-5	-1.276e+01	0.0000	1
116		Phox2a(Homeobox)/Neuron-Phox2a-ChIP-Seq(GSE31456)/Homer	1e-5	-1.250e+01	0.0000	7
117		RAR:RXR(NR),DR5/ES-RAR-ChIP-Seq(GSE56893)/Homer	1e-5	-1.229e+01	0.0000	3
118		Erra(NR)/HepG2-Erra-ChIP-Seq(GSE31477)/Homer	1e-5	-1.219e+01	0.0000	1
119		HNF4a(NR),DR1/HepG2-HNF4a-ChIP-Seq(GSE25021)/Homer	1e-5	-1.212e+01	0.0000	4
120		Phox2b(Homeobox)/CLBGA-PHOX2B-ChIP-Seq(GSE90683)/Homer	1e-5	-1.174e+01	0.0000	4
121		Hoxd12(Homeobox)/ChickenMSG-Hoxd12.Flag-ChIP-Seq(GSE86088)/Homer	1e-4	-1.089e+01	0.0001	2
122		CRE(bZIP)/Promoter/Homer	1e-4	-1.059e+01	0.0001	3
123		Egr1(Zf)/K562-Egr1-ChIP-Seq(GSE32465)/Homer	1e-4	-9.966e+00	0.0002	6
124		Tcf7(HMG)/GM12878-TCF7-ChIP-Seq(Encode)/Homer	1e-4	-9.756e+00	0.0002	4
125		DLX5(Homeobox)/BasalGanglia-Dlx5-ChIP-seq(GSE124936)/Homer	1e-4	-9.588e+00	0.0002	1

126		Nkx3.1(Homeobox)/LNCaP-Nkx3.1-ChIP-Seq(GSE28264)/Homer	1e-4	-9.541e+00	0.0003	3
127		ZNF467(Zf)/HEK293-ZNF467.GFP-ChIP-Seq(GSE58341)/Homer	1e-4	-9.525e+00	0.0003	4
128		HOXA1(Homeobox)/mES-Hoxa1-ChIP-Seq(SRP084292)/Homer	1e-4	-9.327e+00	0.0003	5
129		TATA-Box(TBP)/Promoter/Homer	1e-4	-9.216e+00	0.0003	1
130		Gfi1b(Zf)/HPC7-Gfi1b-ChIP-Seq(GSE22178)/Homer	1e-3	-9.108e+00	0.0004	9
131		ZNF652/HepG2-ZNF652.Flag-ChIP-Seq(Encode)/Homer	1e-3	-9.092e+00	0.0004	3
132		E2F6(E2F)/HeLa-E2F6-ChIP-Seq(GSE31477)/Homer	1e-3	-8.998e+00	0.0004	2
133		DUX4(Homeobox)/Myoblasts-DUX4.V5-ChIP-Seq(GSE75791)/Homer	1e-3	-8.820e+00	0.0005	1
134		KLF5(Zf)/LoVo-KLF5-ChIP-Seq(GSE49402)/Homer	1e-3	-8.412e+00	0.0007	8
135		Hnf6b(Homeobox)/LNCaP-Hnf6b-ChIP-Seq(GSE106305)/Homer	1e-3	-8.254e+00	0.0008	1
136		EAR2(NR)/K562-NR2F6-ChIP-Seq(Encode)/Homer	1e-3	-8.220e+00	0.0009	1
137		Unknown(Homeobox)/Limb-p300-ChIP-Seq/Homer	1e-3	-8.182e+00	0.0009	1
138		GATA:SCL(Zf,bHLH)/Ter119-SCL-ChIP-Seq(GSE18720)/Homer	1e-3	-8.122e+00	0.0009	1
139		Bach2(bZIP)/OCILy7-Bach2-ChIP-Seq(GSE44420)/Homer	1e-3	-8.021e+00	0.0010	1
140		EBF1(EBF)/Near-E2A-ChIP-Seq(GSE21512)/Homer	1e-3	-7.904e+00	0.0012	3
141		Duxbl(Homeobox)/NIH3T3-Duxbl.HA-ChIP-Seq(GSE119782)/Homer	1e-3	-7.824e+00	0.0012	1
142		Dlx3(Homeobox)/Kerainocytes-Dlx3-ChIP-Seq(GSE89884)/Homer	1e-3	-7.735e+00	0.0014	1
143		Barx1(Homeobox)/Stomach-Barx1.3xFlag-ChIP-Seq(GSE69483)/Homer	1e-3	-7.622e+00	0.0015	8
144		Hoxa11(Homeobox)/ChickenMSG-Hoxa11.Flag-ChIP-Seq(GSE86088)/Homer	1e-3	-7.573e+00	0.0016	3
145		GATA3(Zf),DR8/iTreg-Gata3-ChIP-Seq(GSE20898)/Homer	1e-3	-7.569e+00	0.0016	1
146		Ets1-distal(ETS)/CD4+-PolIII-ChIP-Seq(Barski_et_al.)/Homer	1e-3	-7.311e+00	0.0020	3
147		Unknown-ESC-element(?)/mES-Nanog-ChIP-Seq(GSE11724)/Homer	1e-3	-6.991e+00	0.0028	7
148		ETV4(ETS)/HepG2-ETV4-ChIP-Seq(ENCODE)/Homer	1e-3	-6.921e+00	0.0029	8
149		NF1:FOXA1(CTF,Forkhead)/LNCAP-FOXA1-ChIP-Seq(GSE27824)/Homer	1e-2	-6.499e+00	0.0044	8
150		ERG(ETS)/VCaP-ERG-ChIP-Seq(GSE14097)/Homer	1e-2	-6.403e+00	0.0049	1
151		KLF6(Zf)/PDAC-KLF6-ChIP-Seq(GSE64557)/Homer	1e-2	-6.219e+00	0.0058	6

152		CEBP(bZIP)/ThioMac-CEBPb-ChIP-Seq(GSE21512)/Homer	1e-2	-6.016e+00	0.0071	1
153		p73(p53)/Trachea-p73-ChIP-Seq(PRJNA310161)/Homer	1e-2	-5.819e+00	0.0085	5
154		Zfp281(Zf)/ES-Zfp281-ChIP-Seq(GSE81042)/Homer	1e-2	-5.681e+00	0.0097	6
155		COUP-TFII(NR)/Artia-Nr2f2-ChIP-Seq(GSE46497)/Homer	1e-2	-5.655e+00	0.0099	1
156		Tcf3(HMG)/mES-Tcf3-ChIP-Seq(GSE11724)/Homer	1e-2	-5.632e+00	0.0101	2
157		Hoxa13(Homeobox)/ChickenMSG-Hoxa13.Flag-ChIP-Seq(GSE86088)/Homer	1e-2	-5.624e+00	0.0101	3
158		p53(p53)/Saos-p53-ChIP-Seq(GSE15780)/Homer	1e-2	-5.423e+00	0.0122	1
159		p53(p53)/Saos-p53-ChIP-Seq/Homer	1e-2	-5.423e+00	0.0122	1
160		p63(p53)/Keratinocyte-p63-ChIP-Seq(GSE17611)/Homer	1e-2	-5.263e+00	0.0142	3
161		ZNF675(Zf)/HEK293-ZNF675.GFP-ChIP-Seq(GSE58341)/Homer	1e-2	-5.194e+00	0.0152	1
162		Atf3(bZIP)/GBM-ATF3-ChIP-Seq(GSE33912)/Homer	1e-2	-5.186e+00	0.0152	7
163		KLF10(Zf)/HEK293-KLF10.GFP-ChIP-Seq(GSE58341)/Homer	1e-2	-5.068e+00	0.0170	1
164		Srebp1a(bHLH)/HepG2-Srebp1a-ChIP-Seq(GSE31477)/Homer	1e-2	-5.040e+00	0.0174	2
165		PBX2(Homeobox)/K562-PBX2-ChIP-Seq(Encode)/Homer	1e-2	-5.009e+00	0.0178	1
166		PAX3:FKHR-fusion(Paired,Homeobox)/Rh4-PAX3:FKHR-ChIP-Seq(GSE19063)/Homer	1e-2	-4.911e+00	0.0195	3
167		Bach1(bZIP)/K562-Bach1-ChIP-Seq(GSE31477)/Homer	1e-2	-4.899e+00	0.0196	6
168		PU.1:IRF8(ETS:IRF)/pDC-Irf8-ChIP-Seq(GSE66899)/Homer	1e-2	-4.883e+00	0.0198	1
169		Pax7(Paired,Homeobox),long/Myoblast-Pax7-ChIP-Seq(GSE25064)/Homer	1e-2	-4.877e+00	0.0198	6
170		AP-1(bZIP)/ThioMac-PU.1-ChIP-Seq(GSE21512)/Homer	1e-2	-4.855e+00	0.0202	8
171		FOXA1:AR(Forkhead,NR)/LNCAP-AR-ChIP-Seq(GSE27824)/Homer	1e-2	-4.813e+00	0.0209	1
172		Fos(bZIP)/TSC-Fos-ChIP-Seq(GSE110950)/Homer	1e-2	-4.796e+00	0.0211	7
173		LRF(Zf)/Erythroblasts-ZBTB7A-ChIP-Seq(GSE74977)/Homer	1e-2	-4.781e+00	0.0213	8
174		PAX5(Paired,Homeobox),condensed/GM12878-PAX5-ChIP-Seq(GSE32465)/Homer	1e-2	-4.759e+00	0.0217	8
175		HLF(bZIP)/HSC-HLF.Flag-ChIP-Seq(GSE69817)/Homer	1e-2	-4.751e+00	0.0217	1
176		Klf9(Zf)/GBM-Klf9-ChIP-Seq(GSE62211)/Homer	1e-2	-4.682e+00	0.0232	2
177		KLF1(Zf)/HUDEP2-KLF1-CutnRun(GSE136251)/Homer	1e-2	-4.616e+00	0.0246	5

Table 7.6: TF binding motifs found over-represented in control vs *sox32* OE DARs. Motifs are ranked by statistical significance.

Rank	Motif	Name	P-value	log P-value	q-value (Benjamini)	# Target Sequences with Motif	% Target Sequences with Motif
1		Brn1(POU,Homeobox)/NPC-Brn1-ChIP-Seq(GSE35496)/Homer	1e-58	-1.339e+02	0.0000	219.0	15.0
2		Oct6(POU,Homeobox)/NPC-Pou3f1-ChIP-Seq(GSE35496)/Homer	1e-56	-1.290e+02	0.0000	258.0	18.0
3		OCT4-SOX2-TCF-NANOG(POU,Homeobox,HMG)/mES-Oct4-ChIP-Seq(GSE11431)/Homer	1e-52	-1.201e+02	0.0000	156.0	11.0
4		Sox21(HMG)/ESC-SOX21-ChIP-Seq(GSE110505)/Homer	1e-43	-1.003e+02	0.0000	494.0	35.0
5		Oct11(POU,Homeobox)/NCIH1048-POU2F3-ChIP-seq(GSE115123)/Homer	1e-41	-9.446e+01	0.0000	179.0	12.0
6		Sox3(HMG)/NPC-Sox3-ChIP-Seq(GSE33059)/Homer	1e-40	-9.421e+01	0.0000	470.0	33.0
7		Sox2(HMG)/mES-Sox2-ChIP-Seq(GSE11431)/Homer	1e-36	-8.488e+01	0.0000	278.0	19.0
8		Sox10(HMG)/SciaticNerve-Sox3-ChIP-Seq(GSE35132)/Homer	1e-30	-7.007e+01	0.0000	411.0	29.0
9		Oct4(POU,Homeobox)/mES-Oct4-ChIP-Seq(GSE11431)/Homer	1e-30	-6.981e+01	0.0000	269.0	19.0
10		Oct2(POU,Homeobox)/Bcell-Oct2-ChIP-Seq(GSE21512)/Homer	1e-29	-6.777e+01	0.0000	147.0	10.0
11		Sox6(HMG)/Myotubes-Sox6-ChIP-Seq(GSE32627)/Homer	1e-27	-6.422e+01	0.0000	427.0	30.0
12		Sox15(HMG)/CPA-Sox15-ChIP-Seq(GSE62909)/Homer	1e-23	-5.428e+01	0.0000	297.0	21.0
13		Sox17(HMG)/Endoderm-Sox17-ChIP-Seq(GSE61475)/Homer	1e-23	-5.305e+01	0.0000	219.0	15.0
14		Sox9(HMG)/Limb-SOX9-ChIP-Seq(GSE73225)/Homer	1e-17	-4.093e+01	0.0000	182.0	13.0
15		Nkx6.1(Homeobox)/Islet-Nkx6.1-ChIP-Seq(GSE40975)/Homer	1e-17	-4.026e+01	0.0000	687.0	49.0
16		Lhx2(Homeobox)/HFSC-Lhx2-ChIP-Seq(GSE48068)/Homer	1e-16	-3.840e+01	0.0000	337.0	24.0
17		Sox4(HMG)/proB-Sox4-ChIP-Seq(GSE50066)/Homer	1e-14	-3.384e+01	0.0000	200.0	14.0
18		LHX9(Homeobox)/Hct116-LHX9.V5-ChIP-Seq(GSE116822)/Homer	1e-13	-3.190e+01	0.0000	406.0	29.0
19		DLX5(Homeobox)/BasalGanglia-Dlx5-ChIP-seq(GSE124936)/Homer	1e-13	-3.006e+01	0.0000	247.0	17.0
20		Lhx1(Homeobox)/EmbryoCarcinoma-Lhx1-ChIP-Seq(GSE70957)/Homer	1e-12	-2.938e+01	0.0000	348.0	24.0
21		Isl1(Homeobox)/Neuron-Isl1-ChIP-Seq(GSE31456)/Homer	1e-12	-2.938e+01	0.0000	446.0	31.0
22		En1(Homeobox)/SUM149-EN1-ChIP-Seq(GSE120957)/Homer	1e-12	-2.932e+01	0.0000	522.0	37.0

23		DLX1(Homeobox)/BasalGanglia-Dlx1-ChIP-seq(GSE124936)/Homer	1e-12	-2.869e+01	0.0000	389.0	27.0
24		DLX2(Homeobox)/BasalGanglia-Dlx2-ChIP-seq(GSE124936)/Homer	1e-12	-2.785e+01	0.0000	434.0	31.0
25		Dlx3(Homeobox)/Kerainocytes-Dlx3-ChIP-Seq(GSE89884)/Homer	1e-11	-2.673e+01	0.0000	216.0	15.0
26		Sox7(HMG)/ESC-Sox7-ChIP-Seq(GSE133899)/Homer	1e-11	-2.636e+01	0.0000	87.0	6.2
27		Lhx3(Homeobox)/Neuron-Lhx3-ChIP-Seq(GSE31456)/Homer	1e-10	-2.520e+01	0.0000	471.0	33.0
28		Pit1(Homeobox)/GCrat-Pit1-ChIP-Seq(GSE58009)/Homer	1e-9	-2.156e+01	0.0000	292.0	20.0
29		Nanog(Homeobox)/mES-Nanog-ChIP-Seq(GSE11724)/Homer	1e-9	-2.102e+01	0.0000	865.0	61.0
30		Zic3(Zf)/mES-Zic3-ChIP-Seq(GSE37889)/Homer	1e-6	-1.560e+01	0.0000	93.0	6.6
31		LEF1(HMG)/H1-LEF1-ChIP-Seq(GSE64758)/Homer	1e-6	-1.399e+01	0.0000	146.0	10.0
32		Zac1(Zf)/Neuro2A-Plagl1-ChIP-Seq(GSE75942)/Homer	1e-5	-1.348e+01	0.0000	299.0	21.0
33		Zic2(Zf)/ESC-Zic2-ChIP-Seq(SRP197560)/Homer	1e-5	-1.270e+01	0.0000	72.0	5.1
34		Tcf3(HMG)/mES-Tcf3-ChIP-Seq(GSE11724)/Homer	1e-4	-1.104e+01	0.0002	61.0	4.3
35		AP-2gamma(AP2)/MCF7-TFAP2C-ChIP-Seq(GSE21234)/Homer	1e-4	-1.076e+01	0.0003	93.0	6.6
36		AP-2alpha(AP2)/Hela-AP2alpha-ChIP-Seq(GSE31477)/Homer	1e-4	-1.063e+01	0.0003	68.0	4.8
37		Foxh1(Forkhead)/hESC-FOXH1-ChIP-Seq(GSE29422)/Homer	1e-4	-1.022e+01	0.0004	152.0	10.0
38		Zic(Zf)/Cerebellum-ZIC1.2-ChIP-Seq(GSE60731)/Homer	1e-4	-1.002e+01	0.0005	143.0	10.0
39		OCT:OCT(POU,Homeobox)/NPC-OCT6-ChIP-Seq(GSE43916)/Homer	1e-4	-9.559e+00	0.0008	37.0	2.6
40		ZFX(Zf)/mES-Zfx-ChIP-Seq(GSE11431)/Homer	1e-3	-8.822e+00	0.0016	115.0	8.2
41		Hoxa9(Homeobox)/ChickenMSG-Hoxa9.Flag-ChIP-Seq(GSE86088)/Homer	1e-3	-8.752e+00	0.0017	547.0	39.0
42		PPARa(NR),DR1/Liver-Ppara-ChIP-Seq(GSE47954)/Homer	1e-3	-7.580e+00	0.0053	139.0	9.9
43		GRHL2(CP2)/HBE-GRHL2-ChIP-Seq(GSE46194)/Homer	1e-2	-6.889e+00	0.0104	96.0	6.8
44		TCFL2(HMG)/K562-TCF7L2-ChIP-Seq(GSE29196)/Homer	1e-2	-6.781e+00	0.0114	22.0	1.5
45		Tcf7(HMG)/GM12878-TCF7-ChIP-Seq(Encode)/Homer	1e-2	-6.522e+00	0.0144	70.0	5.0
46		CTCF-SatelliteElement(Zf?)/CD4+-CTCF-ChIP-Seq(Barski_et_al.)/Homer	1e-2	-6.425e+00	0.0155	6.0	0.4
47		GATA(Zf),IR4/iTreg-Gata3-ChIP-Seq(GSE20898)/Homer	1e-2	-6.267e+00	0.0178	21.0	1.5

48		KLF6(Zf)/PDAC-KLF6-ChIP-Seq(GSE64557)/Homer	1e-2	-6.129e+00	0.0200	118.0	8.4
49		Atf3(bZIP)/GBM-ATF3-ChIP-Seq(GSE33912)/Homer	1e-2	-6.114e+00	0.0200	103.0	7.3
50		TR4(NR),DR1/Hela-TR4-ChIP-Seq(GSE24685)/Homer	1e-2	-6.084e+00	0.0201	26.0	1.8
51		NFE2L2(bZIP)/HepG2-NFE2L2-ChIP-Seq(Encode)/Homer	1e-2	-5.917e+00	0.0232	13.0	0.9
52		TEAD(TEA)/Fibroblast-PU.1-ChIP-Seq(Unpublished)/Homer	1e-2	-5.716e+00	0.0279	81.0	5.7
53		RUNX(Runt)/HPC7-Runx1-ChIP-Seq(GSE22178)/Homer	1e-2	-5.494e+00	0.0341	134.0	9.5
54		Nur77(NR)/K562-NR4A1-ChIP-Seq(GSE31363)/Homer	1e-2	-5.425e+00	0.0359	39.0	2.7
55		Fos(bZIP)/TSC-Fos-ChIP-Seq(GSE110950)/Homer	1e-2	-5.304e+00	0.0398	96.0	6.8
56		MYNN(Zf)/HEK293-MYNN.eGFP-ChIP-Seq(Encode)/Homer	1e-2	-5.281e+00	0.0400	56.0	4.0
57		BATF(bZIP)/Th17-BATF-ChIP-Seq(GSE39756)/Homer	1e-2	-5.161e+00	0.0443	97.0	6.9
58		AP-1(bZIP)/ThioMac-PU.1-ChIP-Seq(GSE21512)/Homer	1e-2	-5.117e+00	0.0455	116.0	8.2
59		TEAD2(TEA)/Py2T-Tead2-ChIP-Seq(GSE55709)/Homer	1e-2	-5.048e+00	0.0479	61.0	4.3
60		Hoxd11(Homeobox)/ChickenMSG-Hoxd11.Flag-ChIP-Seq(GSE86088)/Homer	1e-2	-5.046e+00	0.0479	493.0	35.0
61		EAR2(NR)/K562-NR2F6-ChIP-Seq(Encode)/Homer	1e-2	-5.044e+00	0.0479	184.0	13.0
62		Hoxd13(Homeobox)/ChickenMSG-Hoxd13.Flag-ChIP-Seq(GSE86088)/Homer	1e-2	-5.039e+00	0.0479	345.0	24.0
63		JunB(bZIP)/DendriticCells-Junb-ChIP-Seq(GSE36099)/Homer	1e-2	-4.870e+00	0.0536	82.0	5.8
64		Maz(Zf)/HepG2-Maz-ChIP-Seq(GSE31477)/Homer	1e-2	-4.858e+00	0.0536	111.0	7.9
65		EBF1(EBF)/Near-E2A-ChIP-Seq(GSE21512)/Homer	1e-2	-4.849e+00	0.0536	71.0	5.0
66		OCT:OCT(POU,Homeobox)/NPC-Brn1-ChIP-Seq(GSE35496)/Homer	1e-2	-4.811e+00	0.0543	6.0	0.4
67		COUP-TFII(NR)/Artia-Nr2f2-ChIP-Seq(GSE46497)/Homer	1e-2	-4.805e+00	0.0543	245.0	17.0
68		Zfp281(Zf)/ES-Zfp281-ChIP-Seq(GSE81042)/Homer	1e-2	-4.780e+00	0.0543	14.0	1.0
69		Zfp57(Zf)/H1-ZFP57.HA-ChIP-Seq(GSE115387)/Homer	1e-2	-4.778e+00	0.0543	57.0	4.0
70		RXR(NR),DR1/3T3L1-RXR-ChIP-Seq(GSE13511)/Homer	1e-2	-4.720e+00	0.0561	121.0	8.6
71		ZNF519(Zf)/HEK293-ZNF519.GFP-ChIP-Seq(GSE58341)/Homer	1e-2	-4.710e+00	0.0561	25.0	1.7
72		KLF5(Zf)/LoVo-KLF5-ChIP-Seq(GSE49402)/Homer	1e-2	-4.675e+00	0.0570	124.0	8.8
73		OCT:OCT-short(POU,Homeobox)/NPC-OCT6-ChIP-Seq(GSE43916)/Homer	1e-2	-4.641e+00	0.0581	177.0	12.0


74		Hoxa13(Homeobox)/ChickenMSG- Hoxa13.Flag-ChIP- Seq(GSE86088)/Homer	1e-2	-4.632e+00	0.0581	516.0	36.
----	---	--	------	------------	--------	-------	-----

Table 7.7: A list of genes found proximal to sox32 OE DARs and Nanog ChIP-seq common regions according to GREAT. TSS = transcriptional start site

Chromosome	Chromosome start	Chromosome end	Gene (distance to TSS)
chr1	4354719	4355555	mnx2b (-10481)
chr1	4799722	4799939	abcb6a (-199005)
chr1	10456357	10456610	knstrn (+165245)
chr1	12437424	12437715	pcdh18a (-281698)
chr1	18156879	18157532	lap3 (+16199)
chr1	21446039	21446264	fgfbp1 (+6174)
chr1	23007389	23007943	rps3a (-122706)
chr1	23790604	23791096	mnd1 (-244656)
chr1	25878867	25879169	sh3gl2 (-130376)
chr1	30875293	30875611	si:dkey-179o14.1 (+83664)
chr1	35219516	35219976	gab1 (+16794)
chr1	39933923	39934863	tenm3 (+183989)
chr1	40234924	40236023	irf2a (+7193)
chr1	41222074	41222375	cpz (+39570)
chr1	41650380	41651001	rgs12b (+18195)
chr1	41889929	41890567	zgc:163052 (+10792)
chr1	51403770	51404555	otx1a (-26425)
chr1	51692638	51693260	jag1a (-25585)
chr1	55262320	55262694	pgam1b (-4234)
chr10	10306848	10307216	cercam (-10549)
chr10	13707873	13709585	cntfr (+27908)
chr10	18018756	18019215	prkab1a (+52284)
chr10	20166558	20166880	dmtn (+4519)
chr10	21186203	21186519	pcdh1a (-58340)
chr10	24008947	24009335	si:dkey-181c13.1 (+247894)
chr10	24617532	24618301	inpp5k (+130427)
chr10	26237491	26238040	tiam1a (+22385)
chr10	26279091	26279437	tiam1a (-19113)
chr10	27017149	27017565	dchs1b (+28393)
chr10	28994920	28995461	ptrh2 (+6661)
chr10	29842645	29843211	picalm (-36248)
chr10	34904511	34905536	kl (-124702)
chr11	6724228	6724747	pde4cb (-122386)
chr11	8244326	8245030	ttl7 (+27603)
chr11	10368469	10368802	ect2 (+151746)
chr11	17619938	17620602	suclg2 (-260604)
chr11	20224726	20224994	zgc:55764 (-29528)

chr11	20368353	20368946	fezf2 (-108426)
chr11	20440135	20440652	fezf2 (-36682)
chr11	25264993	25265489	rbm39a (-7703)
chr11	26164517	26165543	myh7ba (-17820)
chr11	27547826	27548266	map1lc3a (+5378)
chr11	28458896	28459448	barx1 (+39239)
chr11	32906005	32906614	diaph3 (-34828)
chr11	35627832	35628791	fam212aa (-1311)
chr11	36453895	36454442	sema3fb (-43669)
chr11	38143769	38144414	egfl7 (-326950)
chr12	1406702	1407305	pemt (-83993)
chr12	6859994	6860586	a1cf (-4640)
chr12	7781936	7782203	si:dkey-187i8.2 (+47368)
chr12	20036912	20037486	ep300a (+20574)
chr12	21735062	21735653	aqp8a.1 (-9807)
chr12	21740253	21741036	aqp8a.1 (-15094)
chr12	25178487	25178899	svila (-56278)
chr12	27180573	27181360	prkcea (+12835)
chr12	28847051	28847406	arl4d (+33139)
chr12	30567088	30567573	valopb (+23228)
chr12	33147848	33148579	acsl5 (-3760)
chr12	33374985	33375327	habp2 (-25184)
chr12	33491945	33492478	lrrc59 (+22045)
chr12	34628520	34629108	c1qtnf1 (+10420)
chr12	40374454	40374913	abca5 (+23241)
chr12	48905801	48906497	fh (-42103)
chr13	9075580	9075950	sema4ga (-14919)
chr13	10029647	10030053	prepl (-156603)
chr13	11012774	11013577	zbtb18 (-142706)
chr13	14806005	14806296	cdc25 (-15902)
chr13	16746309	16746599	kcnma1a (+174873)
chr13	17305463	17306004	c10orf11 (+207601)
chr13	19818472	19818901	fam204a (+109802)
chr13	21977412	21978046	zswim8 (+11421)
chr13	22047408	22047914	zswim8 (+81353)
chr13	23615846	23616301	khdrbs2 (+109635)
chr13	24575176	24575916	si:ch211-202m22.1 (-19161)
chr13	24698537	24699248	slc1a4 (-27671)
chr13	24772168	24772659	lgalslb (+861)
chr13	26259074	26259944	bcl11aa (+182574)
chr13	30017153	30017518	pax2a (+22897)
chr13	33703761	33704179	btbd6a (-81280)

chr13	34006261	34006713	flrt3 (+45147)
chr13	34462049	34462485	macrod2 (-80404)
chr13	35068894	35069436	sptlc3 (+10296)
chr13	36403238	36403751	si:ch211-67f13.8 (-17017)
chr13	39763475	39763845	fam135a (+21552)
chr13	39766430	39766792	fam135a (+24503)
chr13	41338560	41338982	st3gal5l (+152)
chr13	41683135	41683490	dkk1a (+48753)
chr13	41870570	41871141	dkk1a (+236296)
chr13	44047346	44048147	dact2 (+18244)
chr13	44097203	44097709	dact2 (+67953)
chr13	50440325	50440775	irf2bp2a (-17630)
chr14	5056320	5056696	lbx2 (-335463)
chr14	11573442	11573806	zgc:112315 (-20388)
chr14	18710648	18711367	slbp (-30661)
chr14	18785447	18785717	slbp (+43913)
chr14	19103894	19104486	si:dkey-203a12.2 (+12923)
chr14	14970509	14970986	slc43a3b (+1368)
chr14	20774700	20775133	fmr1 (-383029)
chr14	21604185	21604565	ids (+288637)
chr14	21632422	21632761	ids (+260420)
chr14	24449691	24450165	med12 (-119138)
chr14	26479317	26479677	slc26a2 (-17441)
chr14	26767315	26767818	glra1 (-59614)
chr14	30229696	30230128	zgc:153146 (-415119)
chr14	30840474	30840775	cyp4v7 (-4431)
chr14	31686939	31687409	cf11 (+21986)
chr14	36239512	36239955	ebf1a (+56855)
chr14	36508312	36508980	clint1a (-18687)
chr14	37426687	37427018	pitx2 (-19749)
chr14	40889846	40890318	diaph2 (+245420)
chr14	41767752	41768198	pcdh19 (-74352)
chr15	6974326	6974917	foxl2 (-10062)
chr15	9535463	9536196	zgc:110155 (-32570)
chr15	14083473	14083997	zgc:114130 (-12513)
chr15	15198871	15199194	glod4 (+3565)
chr15	16308800	16309143	cltcb (+3703)
chr15	16322113	16322520	cltcb (+17048)
chr15	17235439	17235845	or131-2 (-51838)
chr15	20692433	20692922	ppp2r1b (-13644)
chr15	24639891	24640377	pafah1b1a (+10622)
chr15	25945582	25945883	serpinf2b (-56999)

chr15	26587582	26588065	tbx2b (-125458)
chr15	26621218	26621624	tbx2b (-91861)
chr15	26707009	26707604	tbx2b (-5975)
chr15	28967295	28967534	nrip1a (-44080)
chr15	31320238	31320552	brca2 (+162037)
chr15	32835252	32835895	mab21l1 (-31996)
chr15	34287900	34288227	zgc:55621 (+44349)
chr15	34718138	34718579	mecom (-44021)
chr15	34751869	34752319	mecom (-77756)
chr15	35076420	35077828	irs1 (+120786)
chr16	700096	700469	irx1a (-120358)
chr16	3488577	3488943	cdcp1a (-14299)
chr16	6594712	6595204	plecb (+87347)
chr16	8394258	8394993	nt5c3a (+7894)
chr16	8575607	8576215	si:dkey-11f5.5 (+5913)
chr16	14265168	14265698	cox6b2 (-2422)
chr16	18239077	18240019	stmn2a (-6443)
chr16	19060414	19060885	rpl18 (+21164)
chr16	22663626	22664005	cpvl (+77141)
chr16	25848218	25849165	kcnn3 (+10491)
chr16	32075196	32075833	tnfaip8l2b (-115712)
chr16	32969393	32970049	fam210a (+61776)
Zv9_scaffold3563	128781	129115	capn7 (-33424)
chr16	40982061	40982834	cdc42se1 (+16696)
chr16	41034288	41034725	zgc:113232 (+9337)
chr16	42392725	42393014	tgfbr2 (-155520)
chr16	42813484	42813988	trmt11 (+103545)
chr4	51650514	51650872	si:ch211-286l10.1 (-129098)
chr16	47810925	47811384	sult2st3 (+25775)
chr16	47877901	47878481	tmem147 (-867)
chr16	50394337	50394644	mios (+117845)
chr16	54352047	54352344	zgc:110372 (+86629)
chr17	150097	150606	dio3a (+77756)
chr17	1375004	1375701	spint1a (+44396)
chr17	8157924	8158822	lft2 (-5402)
chr17	10234785	10235105	foxa1 (+48342)
chr17	10224901	10225259	foxa1 (+58207)
chr17	17293745	17294393	nrxn3a (-22475)
chr17	18467879	18468353	vrk1 (+159282)
chr17	19393630	19394092	cyp26c1 (-28380)
chr17	20323517	20323750	neurlb (-6805)

chr17	21546753	21547194	hmx2 (+67193)
chr17	21794162	21794707	htra1a (+45289)
chr17	22640060	22640298	birc6 (+222620)
chr17	23182257	23182595	ppp1r3ca (+41222)
chr17	23215243	23215958	pcgf5a (+60912)
chr17	26740914	26741449	rca3 (+27085)
chr17	27127390	27128635	id3 (-13027)
chr17	31733121	31733666	vsx2 (-14614)
chr17	35382553	35383419	itgb1bp1 (+22503)
chr17	37386050	37387181	slc30a1b (+3278)
chr17	37391374	37392065	hadhab (-3734)
chr17	40379634	40380314	ubr1 (+422035)
chr17	45573279	45573941	tagapa (+1336)
chr17	46024975	46025341	aspg (+119544)
chr18	4202567	4202996	zbbx (-47151)
chr18	4807344	4808409	fgf7 (-27858)
chr18	8630338	8630587	sema3d (-5918)
chr18	9821132	9821682	hgfb (+298446)
chr18	12684112	12684375	cotl1 (+3737)
chr18	17670844	17671312	si:ch211-212o1.2 (+11149)
chr18	23325939	23326680	nr2f2 (+332768)
chr18	23445387	23446114	nr2f2 (+213327)
chr18	23667778	23668202	nr2f2 (-8912)
chr18	27663176	27663513	tspan18b (-17848)
chr18	28183297	28183546	syt9a (+85049)
chr18	36659326	36659838	si:dkey-10o6.2 (-5709)
chr18	37159908	37160433	onecut1 (-32141)
chr18	37527794	37528519	bcl2l10 (-15852)
chr18	38989816	38990133	si:ch211-132b12.8 (+15653)
chr18	42867066	42867571	aplp2 (+76276)
chr18	48227987	48228402	yif1b (+128344)
chr19	3746733	3747137	stm (-16528)
chr19	11718637	11719002	ssr2 (+10151)
chr19	22695984	22696458	nfatc1 (-4854)
chr19	28579505	28579827	irx1b (+36657)
chr19	33130022	33130634	zbtb10 (+15445)
chr19	33280836	33281708	zgc:103761 (-121618)
chr19	33806186	33806540	atxn1a (-26075)
chr19	34934742	34935330	triqk (+105674)
chr2	4110376	4110928	pter (-2352)
chr2	9979227	9980312	dmbx1a (+1614)
chr2	10053554	10054060	gadd45aa (-18143)

chr2	18719543	18720021	lppr5a (+117416)
chr2	20961386	20962025	bmi1b (+4811)
chr2	26519414	26519918	efna2 (+67869)
chr2	26904400	26905343	si:dkey-181m9.8 (-35931)
chr2	28052378	28052851	rnf152 (+14754)
chr2	755368	756205	foxc1a (+114602)
chr2	27932566	27932817	dhcr7 (+45271)
chr2	28706151	28707175	si:dkeyp-68h1.1 (-3945)
chr2	30224607	30224904	march6 (+19848)
chr2	34642757	34643182	pappa2 (+180932)
chr2	38709706	38710141	copb2 (+9402)
chr2	41955898	41956603	keap1a (+5907)
chr2	41957030	41957655	keap1a (+4815)
chr20	7294846	7295087	si:ch211-121a2.3 (+35961)
chr20	8546553	8546950	prkaa2 (+340055)
chr20	14985852	14986211	faslg (+12342)
chr20	21738222	21738608	syne2a (+6715)
chr20	23445775	23446254	slc10a4 (-19491)
chr20	24006803	24007517	cx52.6 (+64536)
chr20	24017590	24018285	bach2b (+63442)
chr20	24292595	24292827	pdgfra (-17195)
chr20	26462294	26462768	akap12b (-11985)
chr20	28641906	28642452	rgs6 (+25759)
chr20	30129231	30130031	sox11b (-68397)
chr20	32175532	32175957	cep5711 (+12156)
chr20	32460159	32460695	snx3 (+15535)
chr20	32563574	32564513	sec63 (+23415)
chr20	32821383	32821672	fam84a (+25884)
chr20	33412182	33412441	mycn (+79003)
chr20	35600814	35601445	si:dkey-30j22.2 (+1913)
chr20	35978804	35979176	opn5 (-55695)
chr20	39308366	39308681	clu (-9828)
chr20	39841936	39842402	rnf217 (-27424)
chr20	45127341	45127829	atad2b (+339366)
chr20	47071503	47071834	esrrgb (-3449)
chr20	54390618	54391168	si:dkey-24117.2 (+1230)
chr21	2582706	2583387	si:ch211-241b2.1 (-5448)
chr7	29242437	29243446	lmo1 (+241621)
chr21	9234302	9234920	lman1 (-2739)
chr21	9626296	9626579	prlra (-13888)
chr21	9760711	9761217	zgc:101635 (-9327)
chr21	14419064	14419530	tnfrsfa (+3171)

chr21	14948761	14949176	unc5da (-260682)
chr21	15671984	15672420	bcr (+50019)
chr21	16493677	16494146	zgc:158666 (+25471)
chr21	20256657	20256970	rtn2b (+54662)
Zv9_scaffold3533	115573	115758	NONE
chr21	30377072	30377528	si:ch211-166i24.1 (+179938)
chr21	33744575	33745233	wdr55 (-35844)
chr21	34162855	34163333	ubtd2 (-18533)
chr21	36502892	36503878	fgfr4 (+16771)
chr21	36504685	36505175	fgfr4 (+15226)
chr21	36534952	36535705	fgfr4 (-15173)
chr22	12813942	12814559	adarb1a (+18879)
chr22	14015424	14015718	sh3bp4a (-15436)
chr22	19051777	19052087	atp5d (-48857)
chr22	22252214	22252467	gna15.4 (-20271)
chr22	32152590	32152861	adra2a (+123557)
chr23	655004	655325	nadl1.1 (-1602)
chr23	6879889	6880760	zgc:158254 (+15951)
chr23	6881855	6882558	zgc:158254 (+17833)
chr23	9843714	9844559	lama5 (+31452)
chr23	10316746	10317486	krt5 (-32769)
chr23	18273735	18274158	mfsd4b (+149)
chr23	21189546	21190093	pax7b (+37843)
chr23	21372750	21373096	zgc:171453 (+11392)
chr23	23525690	23526557	her9 (+169323)
chr23	23714947	23715607	her9 (-19830)
chr23	29898827	29900567	clstn1 (+10917)
chr23	32471922	32472508	grasp (+2486)
chr23	34751822	34752447	chst13 (-240783)
chr23	35451251	35451527	ftsjd2 (-3024)
chr23	35975932	35976702	rarga (+6243)
chr23	37690036	37690702	tor1l3 (-17229)
chr23	37740964	37741253	tor1l3 (-67969)
chr23	37773250	37774137	tor1l3 (-100554)
chr23	37791536	37792382	irx7 (-116911)
chr23	38937010	38937585	zfp64 (+26519)
chr23	44473357	44473975	ephb4b (-2718)
chr24	4358764	4359263	ccny (-92061)
chr24	4362774	4363403	ccny (-87986)
chr24	6007465	6008073	acbd5a (+8692)
chr24	10342062	10343211	myca (+128423)

chr24	14418345	14419211	ngdn (-26368)
chr24	21862680	21863464	boc (+4325)
chr24	23343964	23344684	zc2hc1a (+20132)
chr24	24117655	24118182	hnf4g (+161412)
chr24	27433094	27433519	tnikb (-4534)
chr24	31884536	31885024	ptbp2a (+205914)
chr25	11396390	11397002	mespab (-13370)
chr25	14720329	14721021	exoc3l1 (-26495)
chr25	21483977	21484312	ergic2 (+121178)
chr25	22935814	22936436	hexa (-3937)
chr25	23436120	23436419	abcc8 (-7772)
Zv9_NA210	18848	19643	NONE
chr3	24374869	24375764	nfe2l1a (+9068)
chr3	25556175	25556806	mchr1b (-53023)
chr3	26556849	26557789	xylt1 (+64302)
chr3	26053312	26053554	nubp1 (+14383)
chr3	31419019	31420063	apnl (-9629)
chr3	33595935	33596897	gtpbp1 (+21683)
chr3	35389708	35390086	cacng3b (+6523)
chr3	38418308	38418705	asic2 (+69328)
chr3	38921672	38922447	si:dkey-106c17.3 (+133779)
chr3	40588770	40589298	fscn1a (+17805)
chr3	43689493	43690259	tmem184a (+2344)
chr3	44357427	44358008	rab11fip4a (-65048)
chr3	46312909	46313524	wu:f104e06 (-29315)
chr3	50019593	50020172	hs3st3b1a (+16256)
chr3	55607573	55608419	si:ch211-74m13.1 (-5245)
chr3	61279934	61280254	foxj1a (-1220)
chr3	63093985	63094298	proza (-38424)
chr4	1016880	1017792	chr2a (+5978)
chr4	4078579	4078949	cd9b (+52812)
chr4	7717563	7718024	fbxl14b (+169823)
chr4	11705674	11706228	lmo3 (+24215)
chr4	12981238	12982141	yaf2 (+4040)
chr4	13040180	13041815	prickle1b (-3388)
chr4	13151337	13152153	pus7l (+105008)
chr4	13525769	13526325	plxnb2a (+26167)
chr4	13670803	13671200	plxnb2a (+18871)
chr4	14391126	14391638	plxna4 (+94462)
chr4	16113993	16114950	etnk1 (+160434)
chr4	16233712	16234099	lrmp (+113206)
chr4	20266248	20266949	syt1a (+124626)

chr4	22032850	22033138	slc35e3 (-20253)
chr4	22432882	22433160	si:ch211-198g19.1 (+145941)
chr4	24187821	24188378	gata3 (-36457)
chr4	24606716	24607952	rbm17 (-4610)
chr4	47159999	47160675	si:dkey-43f9.1 (+24769)
chr4	32139221	32139603	zgc:174654 (-91218)
chr4	49608423	49608921	zgc:174180 (-230516)
chr4	33138119	33138397	si:dkey-122c11.9 (+24306)
Zv9_scaffold3503	952014	952379	NONE
chr4	56869685	56870416	si:dkey-82i20.3 (+151239)
chr4	42795865	42796190	zgc:173702 (-116587)
chr4	42777128	42777423	zgc:173702 (-135339)
chr4	30585327	30585608	si:dkey-25i10.1 (-40173)
chr5	12976552	12977055	fbxw8 (+17496)
chr5	16133071	16133571	kat6a (+13718)
chr5	20925542	20925999	unc13ba (-19694)
chr5	28250703	28251317	tango2 (-80178)
chr5	28341516	28341966	tango2 (+10553)
chr5	28930261	28930574	unc5db (+18614)
chr5	34259123	34259954	ntmt1 (-136365)
chr5	34264079	34264373	ntmt1 (-141052)
chr5	36755215	36755562	fcho2 (+63192)
chr5	41140387	41140714	fgf5 (-2599)
chr5	44835532	44836184	wdr54 (+144325)
chr5	44901988	44902307	rhobtb2a (-186978)
chr5	45803933	45804389	si:ch73-337i15.2 (+101199)
chr5	46418697	46419172	ctsla (-4438)
chr5	46837703	46838244	smarca2 (+80458)
chr5	47426360	47427033	slc4a4a (+61259)
chr5	49443849	49444096	rasa1a (-201454)
chr5	51292965	51293559	nr2f1a (-234242)
chr5	52016107	52016745	fam172a (-194878)
chr5	52258226	52258551	zgc:194908 (+209445)
chr5	53544061	53544715	lhfp12b (+63546)
chr5	58792915	58793450	cdc26 (+10162)
chr5	59934865	59935437	slc37a2 (+13048)
chr5	60858046	60858688	nup88 (-48237)
chr5	60938659	60939254	prkrip1 (+6372)
chr5	62221720	62222298	lig3 (-306084)
chr5	64682252	64683072	si:dkey-261j4.5 (-3076)
chr5	66119492	66120145	specc1 (-118602)

chr5	71182251	71182961	eif4e1b (+12308)
chr5	73176110	73176526	rgs3a (-16552)
chr5	73174169	73174387	rgs3a (-14512)
chr5	73167426	73168187	rgs3a (-8041)
chr6	8830680	8831015	zgc:162584 (-142150)
chr6	12820466	12821309	ical1 (-8814)
chr6	15683095	15683756	iqca1 (+20948)
chr6	16009332	16009646	gbx2 (+60498)
chr6	16037897	16038262	gbx2 (+89089)
chr6	18073574	18074081	ntn1a (+74882)
chr6	22290500	22290877	aanat1 (-22207)
chr6	31491588	31491938	ror1 (+188347)
chr6	31722524	31722841	pgm1 (-5288)
chr6	31856196	31857295	foxd3 (-70746)
chr6	32358177	32358761	mafa (+42493)
chr6	33192263	33192584	pik3r3b (-42488)
chr6	33359557	33360053	pik3r3b (+124893)
chr6	33773395	33773747	ap1m3 (+36728)
chr6	34372572	34373227	rgs4 (+11745)
chr6	37966140	37966958	oca2 (+152550)
chr6	37998482	37999094	oca2 (+184789)
chr6	40031969	40032301	zgc:92046 (-2750)
chr6	41135310	41135652	opn1mw4 (+4117)
chr6	42652883	42653898	sema3fa (-97356)
chr6	43819698	43819990	foxp1b (-5848)
chr6	44002933	44003345	rybpb (-27600)
chr6	45847861	45848600	epha2 (-5301)
chr6	48670503	48671078	si:ch211-105j21.7 (-248011)
chr6	54837389	54837894	csrp1b (-18321)
chr6	57442419	57442884	snrpb (-7167)
chr6	59837769	59838167	slc32a1 (-26175)
chr7	8605408	8605693	zgc:153341 (+11022)
chr7	10487230	10487600	aldh1a3 (+137479)
chr7	13497484	13498217	zgc:158785 (+143716)
chr7	17625226	17625554	dbx1a (+177859)
chr7	19650749	19651367	rgs12a (+109029)
chr7	20812253	20812859	zgc:171731 (+54698)
chr7	21034084	21034499	mus81 (+25771)
chr7	24617489	24617809	zgc:109889 (-45445)
chr7	29423633	29423957	lmo1 (+60768)
chr7	32134250	32134603	mcee (+18134)
chr7	32333062	32333451	tjp1a (+28335)

chr7	33545957	33546484	kif18a (-190853)
chr7	34441209	34441619	tph1b (+7980)
chr7	36396808	36398347	ddx28 (-20857)
chr7	37062911	37063450	irx6a (-92250)
chr7	39534233	39534773	abcc12 (+40930)
chr7	42035165	42036231	mnx1 (+4683)
chr7	47402706	47403222	pop4 (-170326)
chr7	50128470	50128943	trpm5 (-3674)
chr7	55166936	55167453	rps17 (-58336)
chr7	54690065	54690696	fgf3 (-41287)
chr7	54635332	54635909	fgf4 (+2037)
chr7	54626265	54626792	fgf4 (+11129)
chr7	58157686	58158110	myo9a1 (+4756)
chr7	61027237	61027518	ppp1r14bb (+192233)
chr7	61900367	61900668	rwdd (-140442)
chr7	68455169	68455553	parvab (+12771)
chr8	3002963	3003291	crb2b (+11240)
chr8	4921211	4921986	cldn5a (+15294)
chr8	5405646	5405972	bnip3la (+31549)
chr8	10874129	10874471	tbc1d22b (+5201)
chr8	14339942	14340191	cabp4 (-20739)
chr8	15998165	15998536	spata6 (-175922)
chr8	16798232	16798863	dmrta2 (-20085)
chr8	16841681	16842208	dmrta2 (-63482)
chr8	16862082	16862467	faf1 (+63565)
chr8	20309323	20309809	foxd2 (+97179)
chr8	22120718	22120974	si:dkey-24f17.5 (+199475)
chr8	24558376	24558915	pparbd (-6193)
chr8	26037738	26038092	gnai3 (-5051)
chr8	27414121	27414742	sema3bl (+39476)
chr8	31297664	31298303	foxd5 (-3003)
chr8	31459037	31459555	specc1la (-8665)
chr8	31959617	31959947	vgll4l (-4203)
chr8	32786171	32786578	si:dkey-250l23.4 (+39)
chr8	34472561	34472984	lmx1bb (-242818)
chr8	53047032	53047284	egr3 (+54472)
chr8	53049007	53049283	egr3 (+56459)
chr8	54042686	54043096	tbx16 (-11175)
chr8	55336986	55337452	zfyve20 (-2198)
chr9	1999802	2000352	evx2 (+6577)
chr9	2444403	2444916	scrn3 (-19281)
chr9	7140192	7140840	coa5 (-9697)

chr9	10061525	10061968	fstl1b (+1814)
chr9	12656718	12657332	zgc:162707 (+44001)
chr9	13674651	13675212	fzd7a (-43663)
chr9	23539993	23540219	si:dkey-189g17.3 (+59177)
chr9	23601589	23602104	etv5a (-63285)
chr9	26588976	26589512	acvr2aa (-112805)
chr9	27409508	27410114	efnb2a (+272652)
chr9	28357852	28358535	slc10a2 (-14869)
chr9	29262334	29262907	fzd5 (-3457)
chr9	29355140	29355655	creb1b (-47039)
chr9	31406530	31407045	asb11 (-2233)

Table 7.8: A list of genes found proximal to control DARs and Nanog ChIP-seq common regions according to GREAT. TSS = transcriptional start site

Chromosome	Chromosome start	Chromosome end	Gene (distance to TSS)
chr1	5942776	5943339	erbb4a (+248103)
chr1	9626994	9627356	tnrc5 (+22558)
chr1	11044844	11045065	ttyh3b (-13479)
chr1	17125433	17125805	mtnr1aa (-46846)
chr1	20074204	20074680	npy1r (+5952)
chr1	28554581	28555074	si:dkey-1h24.2 (+43884)
chr1	36215364	36215710	pou4f2 (+70486)
chr1	49405037	49405226	cxcl-c1c (+327157)
chr1	49821218	49821726	msxb (-37827)
chr10	9091681	9092070	fstb (+33167)
chr10	9245233	9245670	rasgef1bb (-4312)
chr10	10181113	10181538	sh2d3ca (+37181)
chr10	13518045	13518622	il11ra (-210713)
chr10	13518950	13520146	il11ra (-211927)
chr10	14012380	14012633	cntfr (-275870)
chr10	14931518	14931904	smad2 (+17849)
chr10	32301931	32302497	robo4 (+45284)
chr10	32433537	32434226	robo3 (+84568)
chr10	35731946	35732734	dclk1a (+161748)
chr10	39341102	39341721	si:ch211-55a7.1 (-15163)
chr11	4980972	4981612	ptprga (-57025)
chr11	5542079	5542617	cse1l (-10639)
chr12	6135549	6135939	itga3b (+12651)
chr12	11178814	11179393	slc4a1b (+46166)
chr12	31138789	31139483	gfra1b (-172571)
chr12	32690849	32691189	tcf7l2 (+99269)
chr12	40005373	40005812	rpl38 (-94142)
chr12	45634318	45634847	bccip (+296128)
chr13	6364785	6365283	h2afy2 (-20064)
chr13	17466989	17467241	c10orf11 (+46220)
chr13	23001347	23001775	zgc:113291 (+784)
chr13	30524485	30525271	rps24 (+73193)
chr13	33138991	33139284	rock2a (-16214)
chr13	34771047	34771507	kif16bb (+134791)
chr14	7033642	7034199	hbegfa (+18686)
chr14	26988737	26989064	slc36a1 (-13880)
chr14	27041224	27041792	slc36a1 (+38727)
chr14	27924595	27925175	zgc:162298 (+110138)

chr14	35362416	35362858	gria1a (+2411)
chr14	40678729	40679998	diaph2 (+34702)
chr14	40692552	40693014	diaph2 (+48121)
chr14	41560552	41561067	pcdh19 (+132813)
chr15	5894299	5894730	sh3bgr (+76852)
chr15	10752908	10753461	odz4 (+102805)
chr15	32554946	32555338	mab211l (+248436)
chr15	32528193	32528573	mab211l (+275195)
chr15	31752939	31753642	arl4aa (+51489)
chr16	21368304	21369487	dnah11 (-38490)
chr16	26279827	26280142	apoc2 (+55937)
chr16	32344947	32345288	sema6e (-181864)
chr16	55720773	55721052	pdcl (-105607)
chr17	11601457	11601905	efcab2 (+237625)
chr17	18453238	18453934	vrk1 (+173812)
chr17	27444305	27444825	pacrg (+40461)
chr17	29200941	29201482	foxg1a (-22765)
chr17	29922972	29923721	esrrga (+37753)
chr17	30055012	30055365	esrrga (-94089)
chr17	31043038	31043851	evla (-8357)
chr17	33243848	33244131	snx9a (-120118)
chr17	36146368	36147055	colec11 (+549822)
chr17	39796093	39796732	eprs (-68822)
chr17	41644552	41645048	insm1b (+65142)
chr17	41655244	41656303	insm1b (+76116)
chr17	48056081	48056449	zgc:103755 (-85453)
chr18	2310451	2310968	si:ch211-248g20.5 (-55825)
chr18	10782553	10783161	tspan9a (-273928)
chr18	14155517	14156082	si:dkey-246g23.4 (+2825)
chr18	17157722	17157987	foxl1 (-13268)
chr18	23066153	23066598	mef2aa (+91845)
chr18	25973915	25974627	znf710a (+142448)
chr18	36425769	36426129	luzp2 (+99439)
chr18	36438337	36438800	luzp2 (+86819)
chr18	37909986	37910678	scg3 (+12)
chr19	5811778	5812215	si:dkey-89b17.4 (+44666)
chr19	8086289	8086660	gabpb2a (-23427)
chr19	22175372	22175725	znf516 (+73209)
chr19	22244411	22244755	znf516 (+4175)
chr19	23054802	23055132	pleca (+51141)
chr2	47027435	47027919	si:ch211-226f6.1 (+155117)
chr19	17837960	17838389	klf2b (-9313)

chr20	17484699	17485487	cdh2 (-300932)
chr20	18822923	18823908	fam167ab (+10738)
chr20	33369125	33370246	mycn (+36377)
chr20	38955947	38956362	si:dkey-221h15.1 (+21232)
chr21	15634770	15635415	bcr (+12910)
chr21	21370925	21371177	capslb (-4106)
chr21	23874914	23875773	im:7150060 (+35155)
Zv9_scaffold3460	16154	16766	pcyt1bb (-32831)
chr22	19426074	19426831	fgf22 (-3320)
chr23	26962893	26963539	zgc:158263 (-10653)
chr24	8279027	8279604	slc35b3 (+28516)
chr24	8410256	8410546	slc35b3 (-102569)
chr24	13905188	13905675	eya1 (+76170)
chr24	18356246	18356807	cul1b (-407279)
chr25	15013353	15013785	mpped2 (+58016)
chr25	17179303	17179645	dyrk4 (-4223)
chr25	19604924	19605177	rlbp1b (+9175)
chr25	20370153	20370473	kcna6 (-67057)
chr25	23303411	23303762	stra6 (-7058)
chr3	45174233	45174554	kctd5 (+145069)
chr3	46853133	46853594	elavl3 (+2587)
chr3	50244683	50244925	zgc:112146 (+10627)
chr4	2182465	2183113	phlda1 (+28719)
chr4	9735938	9736848	wnt16 (-34960)
chr4	11588744	11589295	arhgef39 (-47272)
chr4	21035890	21036330	zgc:123180 (-197419)
chr5	54649611	54649953	mtx3 (+72008)
Zv9_scaffold3503	372902	373509	NONE
chr4	55288050	55288431	si:ch211-241n15.3 (-2643)
chr5	14022864	14023299	p2rx2 (+3140)
chr5	42247921	42248581	isl1 (+110924)
chr5	43169140	43169598	dnajb5 (+10087)
chr5	45117554	45117979	rhobtb2a (+28641)
chr5	46612853	46613074	dmrt3a (-6215)
chr5	61782192	61782777	zgc:194800 (+530642)
chr5	64337156	64337726	ncf1 (+170875)
chr5	67281959	67282810	pappab (+89672)
chr6	1610332	1610662	chchd6b (+7136)
chr6	6368354	6368797	cep170b (+30711)
chr6	11678435	11679077	zgc:165651 (+7060)
chr6	19213360	19213987	rhbdl3 (+1330)
chr6	25465224	25466002	lmo4b (+221991)

chr6	26456422	26456860	hdlbp (+144045)
chr6	33452690	33453342	tmem69 (+122988)
chr7	15182316	15183105	rhcga (+14019)
chr7	33508893	33509148	kif18a (-153653)
chr7	44878219	44878773	cdh11 (-117741)
chr7	44948909	44949886	cdh11 (-188643)
chr7	50846753	50847361	hrasb (+261279)
chr7	66441370	66441635	tpcn2 (+110821)
chr7	57854139	57854589	hp (+13579)
chr8	24804100	24804470	lhfp15b (-67030)
chr8	25143797	25144323	mmp9 (+9605)
chr8	32714168	32714631	c7a (-6634)
chr8	47534996	47535630	ywhabb (-19820)
chr8	55045369	55045721	si:dkey-9018.3 (+3247)
chr9	6742677	6742990	pou3f3a (-40109)
chr9	16866992	16867512	epha3l (+39021)
chr9	20646283	20646836	fgf9 (-2779)

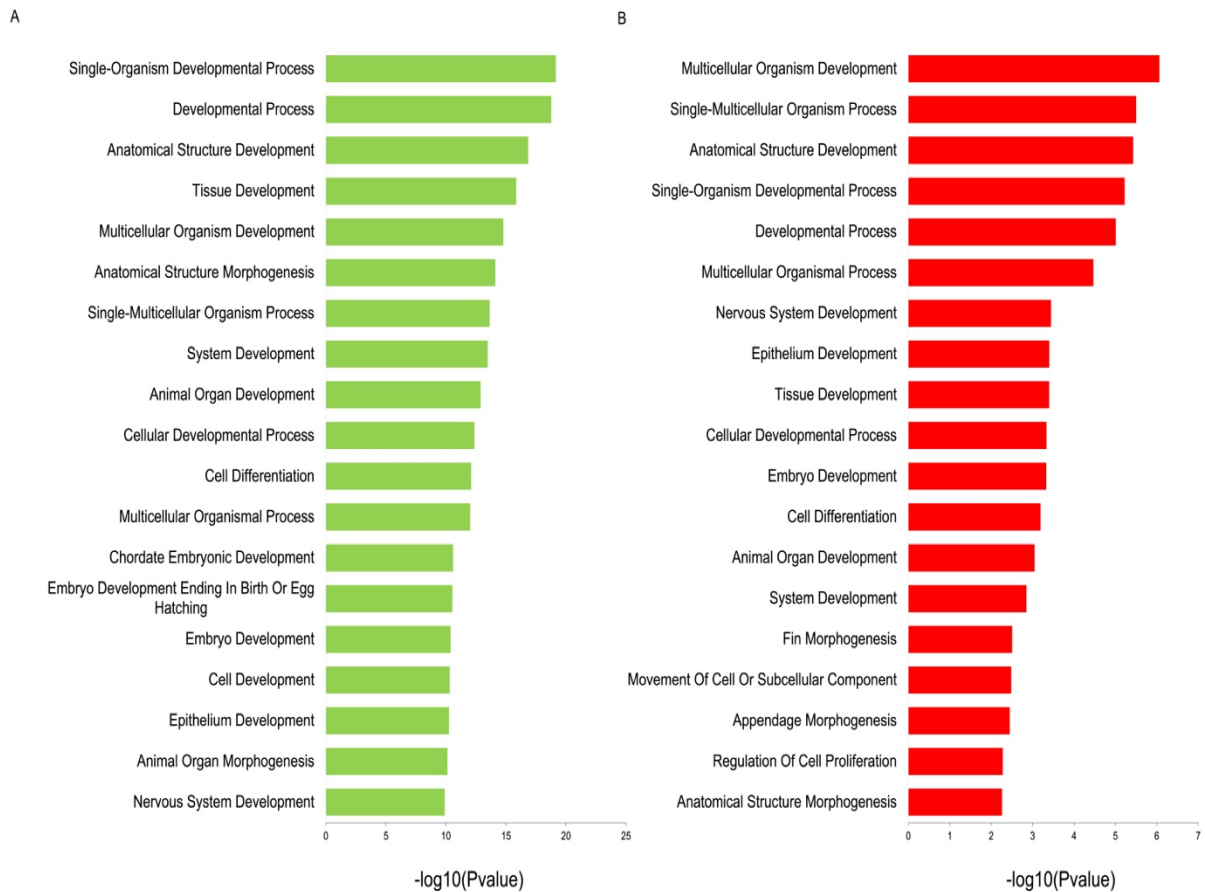


Figure 7-1: Top GO terms associated with genes found proximal to common regions between (A) *sox32* OE or (B) control DARs and Nanog ChIP-seq peaks according to GO analysis using DAVID.

Table 7.9: A list of genes found proximal to *sox32* OE DARs and *Tbx16* ChIP-seq common regions according to GREAT. TSS = transcriptional start site

Chromosome	Chromosome start	Chromosome end	Gene (distance to TSS)
chr1	8450993	8451553	<i>uncx4.1</i> (+85086)
chr1	15085408	15086241	<i>fgf20a</i> (+50543)
chr1	15179993	15181387	<i>vps37a</i> (-24284)
chr1	24193094	24193831	<i>trim2a</i> (+34218)
chr1	25878867	25879169	<i>sh3gl2</i> (-130376)
chr1	25887836	25889152	<i>sh3gl2</i> (-139852)
chr1	35495650	35495945	<i>hhip</i> (-2399)
chr1	39055343	39056154	<i>gpm6ab</i> (+8236)
chr1	39900983	39901622	<i>tenm3</i> (+150899)
chr1	39923267	39923585	<i>tenm3</i> (+173022)
chr1	57662790	57663409	<i>b3gntl1</i> (+7382)
chr1	59463790	59464317	<i>sid4</i> (+64834)
chr10	1145923	1146751	<i>bmpr1bb</i> (-2706)
chr10	2652481	2652806	<i>wu:fb59d01</i> (+38548)

chr10	5381799	5382324	nfil3 (-35564)
chr10	13690174	13690630	cntfr (+46235)
chr10	23330500	23331175	slc25a1a (+11192)
chr10	24959597	24960380	slc7a1 (+83095)
chr10	24977806	24978285	slc46a3 (-82434)
chr10	26282927	26283832	tiam1a (-23229)
chr10	28041542	28041838	auts2 (-209155)
chr10	28788440	28788998	med13a (+10802)
chr10	29478502	29478893	alcama (-35372)
chr10	32446472	32447683	robo3 (+71372)
chr11	6694796	6695462	pde4cb (-151745)
chr11	7981614	7982625	ttl17 (+290161)
chr11	24616306	24617358	pik3c2b (-8842)
chr11	28458896	28459448	barx1 (+39239)
chr11	32916624	32917484	diaph3 (-24084)
chr12	9652506	9653710	cyp26a1 (+59447)
chr12	19101500	19102146	trrap (+143607)
chr12	20796558	20797147	shisa9 (+165014)
chr12	21202608	21204150	mgrn1a (-15374)
chr12	33082157	33082802	vti1a (+27157)
chr12	36748131	36748894	fscn2b (-115374)
chr12	38289794	38290343	cox10 (-139258)
chr12	43490812	43491428	epc1 (-70574)
chr12	44014149	44014785	ebf3 (+268171)
chr12	50552201	50552654	rgrb (+29498)
chr13	3483601	3484004	ncoa4 (+107336)
chr13	4325956	4326740	pla2g12b (-74)
chr13	4599291	4599855	micu1 (+51051)
chr13	5206762	5207414	meis1 (+39110)
chr13	9844385	9845556	six3a (-18392)
chr13	11270762	11271732	desi2 (+5300)
chr13	18251801	18252707	tet1 (+578)
chr13	18331846	18332300	eif4e1c (-28310)
chr13	26329813	26330374	bcl11aa (+253159)
chr13	26402159	26402466	bcl11aa (+325378)
chr13	33703761	33704179	btbd6a (-81280)
chr13	34462049	34462485	macrod2 (-80404)
chr13	47068600	47069033	fgfr2 (-28374)
Zv9 NA48	126177	126763	ate1 (+110912)
chr14	359942	360884	hs3st1 (+116007)
chr14	434769	435279	zgc:172215 (+63429)
chr20	6024180	6024907	cep128 (+63185)

chr14	6433979	6434572	abca1b (+38637)
chr14	9812772	9813122	tmem129 (-19256)
chr14	9812772	9813122	tmem129 (-19256)
chr14	27314377	27314776	emx3 (+71565)
chr14	30169767	30170645	zgc:153146 (-474825)
chr14	32951972	32953434	zic6 (+60552)
chr15	6697297	6697784	sirt2 (+25168)
chr15	14563665	14564114	traf4a (+28263)
chr15	15028801	15029588	hnf1ba (+2758)
chr15	17006313	17007529	arcn1b (+52308)
chr15	25719550	25719901	rtn4rl1b (-15265)
chr15	26195135	26195769	pitpm3 (-15734)
chr15	26621218	26621624	tbx2b (-91861)
chr15	26937448	26938160	brip1 (-66159)
chr16	700096	700469	irx1a (-120358)
chr16	31485293	31486399	rhbg (+19506)
chr16	34031310	34032895	wnt4b (+7972)
chr16	39053894	39054574	ptprub (+95446)
chr16	49306034	49306370	bola1 (+36006)
chr17	4600261	4600898	cdc5l (-228528)
chr17	12790531	12791275	psma6a (+39416)
chr17	13819173	13819856	jkamp (+507170)
chr17	14704580	14706043	ptger2a (-22755)
chr17	16812270	16813036	dio2 (+97997)
chr17	17109677	17110292	nrxn3a (+161609)
chr17	19390080	19390964	cyp26c1 (-25041)
chr17	22573878	22574450	birc6 (+156605)
chr17	26246096	26246988	zgc:158651 (+199697)
chr17	28070732	28071376	im:7158001 (-37714)
chr17	28071595	28072443	im:7158001 (-36749)
chr17	33486318	33486568	dnajc27 (-16229)
chr17	34222124	34222782	mpp5a (-160632)
chr17	46046243	46046771	aspg (+140893)
chr17	49074131	49075132	gnmt (+17105)
chr17	53180089	53180772	spred1 (+6821)
chr17	53268580	53268842	spred1 (-81459)
chr18	23445387	23446114	nr2f2 (+213327)
chr18	25874642	25875392	znf710a (+43194)
chr18	36524735	36525262	luzp2 (+389)
Zv9_NA278	46113	46602	NONE
chr18	39438249	39439039	zbtb38 (+103903)
chr18	42624337	42624632	si:ch211-151h10.1 (+62228)

chr19	9406288	9406900	cers2a (+28898)
chr19	10677290	10677589	grin2da (+26143)
chr19	31867713	31868355	fam49a (-10817)
chr19	43276960	43277379	tinagl1 (-27778)
chr19	48173449	48174185	EIF3HB (-83137)
chr2	12132003	12132445	ARHGAP21B (+32424)
chr2	15804513	15804956	VAV3B (+175051)
chr2	19140120	19140432	ZFYVE9A (+8656)
chr2	26519414	26519918	EFNA2 (+67869)
chr2	27463607	27464443	DSel (-417)
chr2	28295453	28296863	CDH6 (-42395)
chr2	33477172	33478017	si:dkey-31m5.7 (+8628)
chr2	42734619	42735634	MARCH11 (+41437)
chr20	6699024	6699859	si:ch211-191a24.4 (+59287)
chr20	21501862	21502511	JAG2 (+38831)
chr20	24349581	24349844	PDGFRA (+39807)
chr20	22518563	22519094	GSX2 (-47912)
chr20	23324813	23325532	Ociad1 (-84030)
chr20	26468845	26469516	AKAP12B (-18635)
chr20	36588230	36589232	ENAH (-4167)
chr20	37653085	37653846	AIG1 (-32530)
chr20	39370631	39371563	ESCO2 (+1499)
chr21	530720	531362	RAD23B (-43646)
chr21	9283714	9284235	RX3 (-7498)
chr21	9312636	9312972	GRP (-4141)
chr21	10569592	10570138	CELF4 (-26145)
chr21	14266252	14266805	MZT2B (+24396)
chr21	18476296	18477131	FGF10A (+25)
chr21	19399389	19400111	EFNA5A (+83213)
chr21	23562856	23563566	SC5D (+250020)
chr21	23638666	23639450	SC5D (+174173)
chr21	25744307	25745082	TMSB (-38183)
chr21	29095134	29095646	FOXI3A (+73)
chr21	29902644	29903152	ZGC:175145 (+4059)
chr21	36960815	36961570	Cldn2 (-4942)
chr22	11599819	11600622	TSPAN7B (+3165)
chr22	32980406	32981336	Mxi1 (-39378)
chr23	3057970	3058659	ZGC:158828 (+130625)
chr23	13807900	13808605	ZBTB46 (-25563)
chr23	23068374	23069193	Rerea (-45329)
chr23	24620715	24621313	FAM131C (-14563)
chr23	26036206	26036694	si:dkey-21c19.3 (+30884)

chr23	41639698	41640243	scube3 (+62187)
chr24	90362	91234	stk17a (-138258)
chr24	5381759	5382357	plod2 (+64585)
chr24	14298465	14299215	prdm14 (-56697)
chr24	21756423	21757402	zdhhc23b (-16854)
chr24	24754227	24755007	map7d2b (+16261)
chr24	30440498	30440804	aglb (-121896)
chr24	42450343	42451284	ceb1 (-39031)
chr25	900111	900590	irak4 (-12823)
chr25	11400650	11401039	mespab (-17519)
chr25	11699060	11699630	mespab (-316019)
chr25	14366591	14366873	zgc:86725 (-3571)
chr25	18794033	18795193	dusp6 (+23180)
chr25	32256182	32256773	tnnt3a (+8008)
chr25	37796055	37796737	wwox (+301579)
chr3	6844800	6845143	cygb1 (-19313)
chr3	26320099	26320426	rab40c (+89850)
chr3	44279397	44280154	rab11fip4a (+12894)
chr3	40137840	40138410	myo15aa (-6600)
chr3	41110500	41111305	cyp3c1 (-139539)
chr3	41811290	41811673	grifin (-25522)
chr3	53823297	53824257	brd4 (-6754)
chr3	63003152	63004150	zgc:194423 (-12264)
chr4	11127900	11129141	si:dkey-222f8.3 (-23652)
chr4	11792419	11792942	zgc:92357 (-2644)
chr4	13464583	13465567	si:dkey-261e22.1 (+39311)
chr4	16123290	16124627	etnk1 (+169921)
chr4	16306615	16307273	lrmp (+40168)
chr4	20266248	20266949	syt1a (+124626)
chr5	37918159	37919394	efnb1 (+10170)
chr5	43521514	43522602	ncor1 (+93039)
chr5	45250114	45250481	si:dkey-40c11.1 (+51214)
chr5	46632836	46633337	dmrt2a (+6543)
chr5	47061734	47062219	npffr2.1 (+214198)
chr5	58011603	58012128	tbx2a (-88857)
chr5	66119492	66120145	specc1 (-118602)
chr5	75256197	75257587	abl1 (-2929)
chr6	12824127	12825029	cxcr4a (-9344)
chr6	15725234	15725597	iqca1 (+62938)
chr6	16037897	16038262	gbx2 (+89089)
chr6	18070956	18071768	ntn1a (+72416)
chr6	18065052	18065846	ntn1a (+66503)

chr6	22906113	22906615	prkca (-5846)
chr6	29861886	29862869	lrrc40 (+34832)
chr6	30583208	30583851	insl5a (+187532)
chr6	37588039	37588857	cyfip1 (-14978)
chr6	45603257	45603786	cntn4 (+125089)
chr6	45847861	45848600	epha2 (-5301)
chr6	50476799	50477603	mych (+74098)
chr6	50739700	50740196	pigu (-57896)
chr7	22201709	22202540	efnb3b (+67778)
chr7	31293022	31293972	tpma (+4740)
chr7	35031519	35032062	gro1 (-12649)
chr7	35039923	35041263	gro1 (-21451)
chr7	35878987	35879842	irx3a (-64984)
chr7	39330679	39331043	heatr3 (-19284)
chr7	40013907	40014424	c1qtnf4 (-9823)
chr7	40337544	40338144	creb3l1 (+47534)
chr7	42501961	42502498	shcbp1 (-247643)
chr7	47366371	47367014	pop4 (-206597)
chr7	45789979	45790823	ca7 (+101)
chr7	48014317	48014788	znf536 (+32562)
chr7	48513542	48513888	hel_dr4 (+319932)
chr7	66552815	66553520	tpcn2 (+222486)
chr7	57279868	57280332	cdh15 (-81685)
chr7	68426252	68426818	parvab (+41597)
chr7	68426252	68426818	parvab (+41597)
chr8	17622172	17622842	plk2b (+24415)
chr8	18388392	18388944	slc44a5b (+30157)
chr8	20376267	20376693	si:ch211-120j21.1 (+65662)
chr8	20469831	20470570	si:ch211-120j21.1 (-28059)
chr8	26620801	26621673	cacna1sb (+684)
chr8	28156940	28157493	rhead (-251098)
chr8	30464209	30464864	atoh1a (+22509)
chr8	33363409	33363845	ncs1a (-575)
chr8	34190355	34191243	lmx1bb (+39156)
chr8	45156058	45156464	ccdc92 (+42656)
chr8	53049007	53049283	egr3 (+56459)
chr8	53811782	53812873	fgfr1a (-27037)
chr8	55597036	55597873	fgd (+1664)
chr9	6377343	6377755	ecrg4a (+43923)
chr9	15122302	15123082	npr2b (-341626)
chr9	18528130	18528516	tnfsf11 (-11328)
chr9	21956494	21958158	gdap2 (-70542)

chr9	22616248	22616593	arhgef7a (+46954)
chr9	24449817	24450332	arl4cb (-58249)
chr9	27218609	27219307	efnb2a (+81799)
chr9	27317634	27318439	efnb2a (+180878)

Table 7.10: A list of genes found proximal to control DARs and Tbx16 ChIP-seq common regions according to GREAT. TSS = transcriptional start site

Chromosome	Chromosome start	Chromosome end	Gene (distance to TSS)
chr10	26818995	26819390	zgc:153126 (+18003)
chr13	30845527	30846056	zcchc24 (+23775)
chr14	7033642	7034199	hbegfa (+18686)
chr17	15424792	15425350	zgc:92375 (+27843)
chr20	4339539	4340574	zgc:154018 (+83063)
chr21	6672096	6672565	olfm1b (-35731)

Table 7.11: A list of genes found proximal to sox32 OE DARs and Tbx16 ChIP-seq common regions according to GREAT. TSS = transcriptional start site

Chromosome	Chromosome start	Chromosome end	Gene (distance to TSS)
chr1	4183462	4184116	spry2 (+12870)
chr1	4357305	4358025	mnx2b (-13009)
chr1	8483284	8483883	micall2b (+78580)
chr1	15187019	15187656	vps37a (-17636)
chr1	16058459	16059478	fat1 (-59848)
chr1	16064718	16065902	fat1 (-53507)
chr1	23778434	23779869	mnd1 (-256354)
chr1	23925969	23926456	mnd1 (-109293)
chr1	23961260	23962311	mnd1 (-73720)
chr1	23967020	23967791	mnd1 (-68100)
chr1	29170757	29171797	vps8 (+100245)
chr1	29736553	29737930	dachc (+8019)
chr1	33023030	33024073	gyg2 (-4522)
chr1	35509940	35511036	hhip (+12291)
chr1	36064483	36065597	pou4f2 (-80011)
chr1	39795598	39796936	tenm3 (+45863)
chr1	40234924	40236023	irf2a (+7193)
chr1	40250647	40252467	tmem134 (+6889)
chr1	40623013	40623497	si:dkey-44g23.9 (-1297)
chr1	42169832	42170526	rnf24 (+30125)
chr1	43872746	43873555	suclg1 (+193953)

chr1	47106281	47107176	cab39l (+10858)
chr10	5032212	5033169	enoph1 (+3593)
chr10	5650980	5651994	psd3l (+23690)
chr10	19061008	19061360	zgc:101123 (+13818)
chr10	27651169	27652304	si:dkey-88p24.9 (-6831)
chr10	28754784	28755697	med13a (+44280)
chr10	32911375	32912207	lhfp (+5662)
chr11	1505758	1506295	ift52 (+7658)
chr11	19152260	19152839	id1 (-26915)
chr11	19231131	19232283	zgc:175264 (+7227)
chr11	19444037	19445696	rho1 (+73223)
chr11	21172067	21173381	cdh4 (+161208)
chr11	23871193	23871800	mdm4 (-20489)
chr11	24306108	24306519	kiss1 (-57957)
chr11	31291788	31292719	pecr (-51953)
chr11	33219316	33220013	diaph3 (+278527)
chr11	33233931	33235027	diaph3 (+293341)
chr11	35627832	35628791	fam212aa (-1311)
chr11	36390954	36391833	sema3fb (+19106)
chr11	36453895	36454442	sema3fb (-43669)
chr11	45526552	45527512	tomm20b (-33741)
chr12	4764725	4765844	aldoab (-14714)
chr12	7083216	7084331	dkk1b (-65549)
chr12	18613027	18613418	cyth3 (+41115)
chr12	27206841	27207702	prkcea (+39140)
chr12	29110317	29111723	etv4 (+3894)
chr12	32687089	32687873	tcf7l2 (+102807)
chr12	36649237	36649914	slc38a10 (-89883)
chr12	37819740	37821473	st6galnac (-9122)
chr12	38455346	38455859	cox10 (+26276)
chr13	5612448	5614381	spred2b (+684)
chr13	5835729	5836664	ppm1g (-30077)
chr13	10421405	10422182	abcg8 (+88690)
chr13	13561003	13561780	cfp (+18698)
chr13	17313949	17314743	c10orf11 (+198989)
chr13	21858651	21860010	sprn (-30943)
chr13	22184772	22185712	ndst2a (-20179)
chr13	28679717	28680571	fgf8a (-38877)
chr13	28748844	28749649	fgf8a (+30226)
chr13	28822937	28823736	cyp17a1 (+11230)
chr13	34078918	34080118	flrt3 (-27884)
chr13	34115461	34116343	flrt3 (-64268)

chr13	35563551	35564786	jag1b (+3395)
chr13	35765794	35766397	wdr27 (-8150)
chr13	36298962	36299741	tacc3 (-5510)
chr13	50342655	50343090	irf2bp2a (+80047)
chr14	6420795	6422000	abca1b (+25759)
chr14	7724948	7725652	ppp3cb (-45221)
chr14	8049408	8051094	zgc:92242 (+424)
chr14	21572338	21572900	ids (+320393)
chr14	24671973	24672803	fgf1a (-30483)
chr14	27458901	27461395	gpr137 (+11823)
chr14	27494214	27495341	smad5 (+12516)
chr14	31563219	31563946	efemp2a (+10121)
chr14	31735190	31736123	cfl1 (+70469)
chr14	33769021	33769654	lnx2b (-4233)
chr14	40265005	40266614	csnk1a1 (+14261)
chr14	40281995	40283331	csnk1a1 (-2592)
chr14	42129462	42130311	ef1 (-11349)
chr14	46460868	46461309	grhpra (+551252)
chr15	4356249	4357261	tfdp2 (-5885)
chr15	14697553	14698597	tada2a (-107150)
chr15	16971121	16972699	arcn1b (+87319)
chr15	18196161	18197130	vps26b (-28505)
chr15	20785398	20786103	sik2b (+12614)
chr15	22251792	22252421	grik4 (+31385)
chr15	22266089	22266774	grik4 (+45710)
chr15	27016027	27016689	lhx1a (+43039)
chr15	27279689	27280335	dhrs13a.3 (+154352)
chr15	27946388	27948145	slc6a4a (+20119)
chr15	30068613	30069398	msi2b (+45020)
chr15	37594406	37595381	robo1 (+89977)
Zv9_NA810	23825	24891	NONE
chr15	46181778	46183227	chd (-9219)
chr15	47214940	47216114	zgc:153039 (-91469)
chr16	12690699	12691388	erfl3 (-43091)
chr16	23064894	23065383	skap2 (+46356)
chr16	25104799	25105501	flad1 (+58914)
chr16	32867165	32868010	lmna (+289)
chr16	32971609	32972470	fam210a (+64095)
chr16	39483166	39484670	cited4b (-24255)
chr16	42380360	42380919	tgfbr2 (-143290)
chr16	42967459	42968351	rspo3 (+331)
chr16	55390066	55390702	col27a1a (+16974)

chr17	10479206	10479942	sec23a (+31146)
chr17	10218077	10218751	foxa1 (+64873)
chr17	17293745	17294393	nrxn3a (-22475)
chr17	19951976	19952694	actn2 (+21742)
chr17	23878037	23879616	b3gnt2b (-13448)
chr17	26698460	26699470	phf17 (+29248)
chr17	26740914	26741449	rcan3 (+27085)
chr17	27807278	27809083	qkia (+25581)
chr17	28255439	28257951	htr1bd (-29040)
chr17	29486513	29487083	kctd3 (+82082)
chr17	30447642	30448674	greb1 (+19652)
chr17	34482366	34482969	eif2s1a (+71193)
chr17	37386050	37387181	slc30a1b (+3278)
chr17	38420673	38423608	nkx2.9 (-482)
chr17	42345536	42346397	foxa2 (+48385)
chr17	43633154	43633811	zfyve28 (+6168)
chr17	51198412	51199606	fermt2 (-37873)
chr18	2013162	2013970	si:ch211-248g20.1 (-47153)
chr18	8878409	8879287	sema3ab (+42167)
chr18	14972626	14973839	rfx4 (-13543)
chr18	19923071	19924187	gro2 (-83043)
chr18	20146781	20147259	kif23 (+92103)
chr18	21451562	21452397	mbtps1 (+13786)
chr18	21471188	21471878	mbtps1 (-5767)
chr18	23674871	23675716	nr2f2 (-16216)
chr18	32754002	32754856	ssr3 (+55036)
chr18	32999510	33000355	tiparp (+58180)
chr18	35665737	35667405	yap1 (+28931)
chr18	35678034	35679390	yap1 (+16790)
chr18	37231822	37232507	onecut1 (+39853)
chr18	38580813	38581635	sema6dl (-80599)
chr18	42676789	42677662	si:ch211-151h10.1 (+9487)
chr18	42891938	42894197	aplp2 (+50527)
chr18	44755336	44755927	foxa3 (+7090)
chr19	1948621	1949391	rplp1 (+29767)
chr19	3054409	3055020	fam126a (+28663)
chr19	6819034	6819983	erf (-4249)
chr19	9112793	9113974	efna1a (+9046)
chr19	9130728	9131405	efna1a (+26729)
chr19	12852242	12853193	grhl2b (-4383)
chr19	30640106	30640633	ntla (-2461)

chr19	30673803	30674483	ntla (-36234)
chr19	14850586	14851160	serinc2 (+1879)
chr19	15407586	15408316	ctps1a (-42049)
chr19	26475660	26478454	jarid2b (-81232)
chr19	35487251	35488854	si:dkey-184p18.2 (+44292)
chr19	43553898	43554916	si:dkey-158b13.2 (-27490)
Zv9_NA898	3913	4630	NONE
chr2	10237494	10238758	scinlb (+3963)
chr2	17575731	17577464	ptprfb (+23183)
chr2	17581469	17583267	ptprfb (+17413)
chr2	21185873	21186755	ctdsplb (+33814)
chr2	24803817	24804925	slc35g2a (-38341)
chr2	25562335	25562950	fndc3ba (+36016)
chr2	31851475	31852777	mycb (+117995)
chr2	31874013	31874927	fam49ba (+105692)
chr2	36616906	36617659	zgc:153654 (+21976)
chr2	39292554	39293136	zgc:136870 (-30302)
chr20	13783650	13784329	nek2 (-13915)
chr20	20857453	20858194	ckbb (+36121)
chr20	22004295	22005301	daam1b (-22595)
chr20	22006397	22007595	daam1b (-24793)
chr20	23024262	23025172	rasl11b (-10981)
chr20	28230328	28230932	dll4 (-44141)
chr20	28402880	28403890	rhov (+9036)
chr20	51502999	51503810	rbm25 (+17830)
chr21	5096522	5098024	notch1a (+40034)
chr21	8009493	8010016	crb2a (+3347)
chr21	9234302	9234920	lman1 (-2739)
chr21	13376694	13377470	sfswap (+108438)
chr21	13925831	13927533	p2rx4a (+542)
chr21	14208400	14209164	lhx5 (-76673)
chr21	14429168	14429840	tnfrsfa (-7036)
chr21	20176150	20176992	kcnk12l (-8624)
chr21	31938958	31940188	arl3l1 (-68)
chr21	32956002	32956734	ebf1b (+4633)
chr22	2470248	2471378	si:ch73-138e16.5 (-28)
chr22	8672779	8673766	rnh1 (-269512)
chr22	12146674	12147666	mgat5 (+64234)
chr22	19088875	19089978	bsg (-64087)
chr22	19953050	19954790	rgmd (+4698)
chr22	21305735	21308008	fkbp8 (-36078)
chr22	36031288	36031754	dag1 (+16177)

chr23	241572	242133	taf11 (-83462)
chr23	6630968	6631806	rbm38 (-5893)
chr23	6634332	6635290	rbm38 (-2469)
chr23	6879889	6880760	zgc:158254 (+15951)
chr23	7473172	7473803	gata5 (+20195)
chr23	7501331	7502741	gata5 (+48743)
chr23	7759544	7760745	pofut1 (-5462)
chr23	9843714	9844559	lama5 (+31452)
chr23	17278274	17279905	commd7 (+5412)
chr23	19278912	19279835	bcl2l1 (-8325)
chr23	21744640	21745616	her4.1 (+2628)
chr23	23770175	23771105	agr1 (-9133)
chr23	24687306	24688358	fam131c (-81381)
chr23	29907688	29908095	clstn1 (+2722)
chr23	29918156	29919028	clstn1 (-7978)
chr23	29935854	29936575	clstn1 (-25601)
chr23	29972865	29974523	ctnnbip1 (+3135)
chr23	32454024	32454976	grasp (+20201)
chr23	35993060	35994556	rarga (+23734)
chr23	35998648	35999528	rarga (+29014)
chr23	37740964	37741253	tor1l3 (-67969)
chr23	37770385	37771411	tor1l3 (-97758)
chr23	38918463	38919283	zfp64 (+8094)
chr23	44893228	44893929	zgc:100911 (+18014)
chr24	4545128	4545879	fzd8a (-17548)
chr24	14384203	14384937	prdm14 (+29033)
chr24	20274725	20275124	csrnp1b (-7421)
chr24	25753588	25754435	si:dkey-11p10.9 (+40831)
chr24	36502292	36503351	st18 (-112880)
chr24	39466396	39468012	metrn (-8197)
chr25	669219	669997	zgc:162740 (-35397)
chr25	11750827	11752017	mespab (-368096)
chr25	13067434	13068191	sept7b (+23721)
chr25	17734698	17736654	cyp2x7 (-131)
chr25	18374812	18375595	pth1a (+27116)
Zv9_NA210	18848	19643	NONE
chr3	15499748	15500730	hn1l (+2601)
chr3	17934798	17935854	tob1a (-17103)
chr3	33595935	33596897	gtpbp1 (+21683)
chr3	35358718	35359866	cacng3b (-24082)
chr3	36204642	36205566	csnk1da (+4358)
chr3	46020089	46020906	zgc:194346 (+132633)

chr3	47862202	47863235	cdc42ep4a (-105948)
chr3	53704035	53705311	lpar2a (+25858)
chr3	55533661	55534825	si:ch211-74m13.3 (-61774)
chr4	14252	14682	msgn1 (-2091)
chr4	4461975	4463014	zgc:194209 (-54595)
Zv9_scaffold3539	85501	87277	arl1 (-39433)
Zv9_scaffold3539	92411	93120	arl1 (-45810)
chr4	10650514	10651321	net1 (+7808)
chr4	10990572	10991032	cry1a (-86765)
chr4	12860749	12861940	cntn1b (+62983)
chr4	16082230	16084075	etnk1 (+129115)
chr4	16116563	16119224	etnk1 (+163856)
chr4	19770625	19771054	nav3 (+22250)
chr4	22358077	22358975	si:ch211-149p5.1 (+198955)
chr4	23947514	23948613	gata3 (+203579)
chr5	16298698	16299968	sc:d411 (-11189)
chr5	21528581	21529848	fam222a (+33876)
chr5	24544272	24545042	npas2 (-30357)
chr5	24545517	24546856	npas2 (-28827)
chr5	29327128	29328149	zmat4a (+99645)
chr5	38938670	38940080	zgc:163098 (+48669)
chr5	42097463	42098307	ndufs4 (-12748)
chr5	42356822	42357625	isl1 (+1951)
chr5	45336788	45337656	si:dkey-40c11.1 (-35710)
chr5	47467365	47468117	slc4a4a (+20215)
chr5	62020578	62021087	lig3 (-507260)
chr5	70790172	70790999	ephb4a (-11392)
chr6	10939668	10940302	abcc6a (+2684)
chr6	15645493	15646964	cxcr7b (+11466)
chr6	18073574	18074081	ntn1a (+74882)
chr6	30434475	30435718	insl5a (+39099)
chr6	33359557	33360053	pik3r3b (+124893)
chr6	35303955	35304752	lrp8 (+7123)
chr6	42681330	42682345	sema3fa (-68909)
chr6	43222097	43223557	arl6ip5a (+57314)
chr6	43988490	43989279	rybpb (-13346)
chr6	45829723	45830612	epha2 (+12762)
chr6	45852969	45853401	h6pd (-2980)
chr6	50778519	50778942	pigu (-96679)
chr6	52998996	53000061	fam212ab (+1428)

chr7	10112088	10113567	chsy1 (+5972)
chr7	16219095	16219485	igf1rb (+3504)
chr7	17853962	17855122	prmt3 (-26950)
chr7	21681224	21683721	slc3a2a (+813)
chr7	22898320	22900157	tmem88b (-56745)
chr7	29943048	29944071	scube2 (+3840)
chr7	32134250	32134603	mcee (+18134)
chr7	33274387	33276768	bdnf (-8345)
chr7	34716202	34717607	anp32a (-32313)
chr7	36396808	36398347	ddx28 (-20857)
chr7	38689825	38690593	sall1a (-79228)
chr7	40018858	40019707	c6ast4 (+11628)
chr7	44796502	44796998	cdh11 (-35995)
chr7	46052134	46052673	fam96b (-93463)
chr7	48004818	48005872	znf536 (+23354)
chr7	48287556	48288083	znf536 (+305829)
chr7	49626772	49627445	cpeb1b (+81114)
chr7	50058588	50059462	kcnq1 (+46195)
chr7	50083398	50084050	cdkn1c (+40320)
chr7	54559220	54559653	ccnd1 (-18154)
chr7	54532011	54533330	ccnd1 (-44920)
chr7	61265381	61265997	pc (+40984)
chr7	62745217	62746278	rbpjb (+22652)
chr7	74111173	74111889	ccdc149a (+176864)
chr8	11909182	11910017	mapk13 (-25840)
chr8	16302724	16303819	agbl4 (+207654)
chr8	16798232	16798863	dmrta2 (-20085)
chr8	20652258	20653936	rfx2 (+23172)
chr8	23130104	23132265	si:ch211-261n11.3 (+6521)
chr8	26463093	26464240	slc38a5a (-5057)
chr8	31197400	31199469	zgc:162939 (-33957)
chr8	31285579	31286285	cbwd (+12190)
chr8	34147989	34148751	lmx1bb (+81585)
chr8	39156614	39157386	rab11fip1a (-43224)
chr8	45572799	45574511	ncor2 (-8679)
chr8	45601408	45602490	ncor2 (-36973)
chr8	51584455	51585350	prickle3 (-19611)
chr8	54046363	54047266	tbx16 (-15099)
chr9	3134284	3135262	pdk1 (+13631)
chr9	3671740	3672381	sp5 (-3700)
chr9	10061525	10061968	fstl1b (+1814)
chr9	10303079	10304869	spopla (-11247)

chr9	11530782	11531540	chpfa (+8299)
chr9	11544537	11545264	chpfa (-5441)
chr9	14097543	14098405	raph1 (+50193)
chr9	14612922	14613411	abcb6b (-63452)
chr9	19411278	19412769	gtf2f2b (-35382)
chr9	19564713	19565854	kctd4 (-69646)
chr9	19567366	19568216	kctd4 (-72153)
chr9	21438300	21438942	si:dkey-174k18.1 (-78979)
chr9	27063976	27065201	arhgap15 (-37395)
chr9	27105105	27105770	arglu1a (-14042)
chr9	27283193	27283679	efnb2a (+146277)
chr9	28308754	28309660	slc10a2 (-63856)
chr9	28357852	28358535	slc10a2 (-14869)

Table 7.12: A list of genes found proximal to control DARs and Tbxta ChIP-seq common regions according to GREAT. TSS = transcriptional start site

Chromosome	Chromosome start	Chromosome end	Gene (distance to TSS)
chr2	25080341	25081102	si:dkey-5n7.2 (-149616)
chr16	21368304	21369487	dnah11 (-38490)
chr17	41651757	41653169	insm1b (+72805)
chr18	41553802	41554841	pvr1b (-1828)
chr20	4361554	4362913	zgc:154018 (+105240)
chr20	33360974	33361837	mycn (+28097)
chr22	33176235	33177012	zgc:171679 (-15669)
chr23	26932578	26933324	hspg2 (-40041)
chr24	13897515	13898268	eya1 (+68630)
chr24	29710238	29710817	olfm3a (-76827)
chr3	7459647	7460708	zgc:173577 (-365043)
chr3	11235354	11236826	zgc:165627 (+61902)
Zv9_scaffold353 0	1076723	1077910	zgc:174311 (+697202)
chr5	26353345	26354243	si:ch211-106a19.1 (+27546)
chr5	42247921	42248581	isl1 (+110924)
chr5	65161799	65162493	akap10 (+90608)
chr6	57883198	57884935	cbfa2t2 (-278956)

chr7	15182316	15183105	rhcga (+14019)
chr8	35284857	35285483	zgc:77614 (+7141)

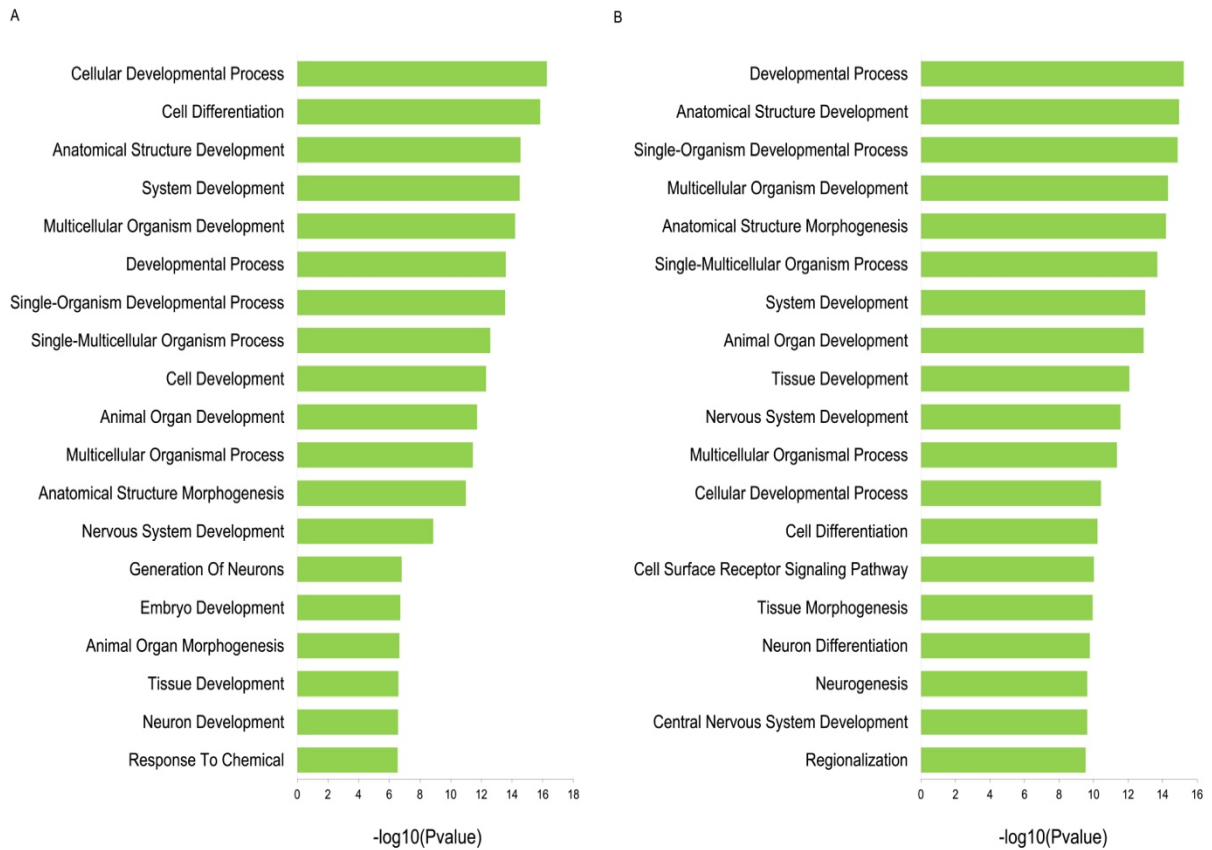


Figure 7-2: Top GO terms associated with genes found proximal to common regions between (A) Tbx16 or (B) Tbx1a ChIP-seq peaks and *sox32* OE DARs according to GO analysis using DAVID.

Table 7.13: A list of upregulated transcripts in *sox32* OE vs control embryos. Genes are ordered based on adjusted Pvalues.

Gene_ID	Log2FoldChange	Stat	Pvalue	Padj	Symbol
ENSDARG00000101717	-4.5948819	62.04157 9	0	0	<i>sox17</i>
ENSDARG00000057633	-2.9978738	46.46383 5	0	0	<i>cxcr4a</i>
ENSDARG00000068409	-3.6206945	43.20144 1	0	0	<i>vgl1l1</i>

ENSDARG0000005 9534	-6.9294629	- 41.63632 5	0	0	<i>si:ch211- 146m13.3</i>
ENSDARG0000009 0170	-3.4886483	- 38.98643 7	0	0	<i>rab11fip4a</i>
ENSDARG0000010 0753	-2.8966365	- 37.89207 1	0	0	<i>scarb2b</i>
ENSDARG0000004 0100	-4.4886226	- 37.69749 8	0	0	<i>si:dkey-95h12.1</i>
ENSDARG0000008 6996	-3.8721268	- 35.73425 7	1.16E- 279	3.03E- 276	<i>si:ch73- 234b20.5</i>
ENSDARG0000011 2644	-3.3522903	- 34.52317 2	3.60E- 261	8.35E- 258	<i>st6galnac</i>
ENSDARG0000001 7388	-3.121359	- 34.30477 8	6.66E- 258	1.39E- 254	<i>gstt1b</i>
ENSDARG0000003 6791	-2.4221538	- 33.96493 6	7.34E- 253	1.39E- 249	<i>dnmt3bb.1</i>
ENSDARG0000010 1919	-4.1267967	- 33.91976 7	3.41E- 252	5.92E- 249	<i>foxj1a</i>
ENSDARG0000005 7708	-2.4089148	- 32.46295 6	3.56E- 231	5.70E- 228	<i>zgc:175264</i>
ENSDARG0000005 8597	-2.9034968	- 30.41583 3	3.39E- 203	5.05E- 200	<i>nt5c3a</i>
ENSDARG0000010 0591	-2.6627759	- 30.28740 6	1.68E- 201	2.33E- 198	<i>sox32</i>

ENSDARG0000007 8256	-3.2570725	- 30.23101 6	9.27E- 201	1.21E- 197	<i>ovo11b</i>
ENSDARG0000006 2593	-2.2021903	- 29.48162 4	4.95E- 191	6.07E- 188	<i>stox1</i>
ENSDARG0000006 0368	-3.7748823	- 29.26744 2	2.69E- 188	3.12E- 185	<i>syn1</i>
ENSDARG0000005 5373	-2.6436336	- 28.04449 6	4.66E- 173	5.12E- 170	<i>sema3fb</i>
ENSDARG0000002 9443	-2.08617	- 27.84193 9	1.35E- 170	1.41E- 167	<i>zgc:92242</i>
ENSDARG0000001 3853	-1.8855357	- 27.78464 8	6.65E- 170	6.60E- 167	<i>lmo4a</i>
ENSDARG0000005 5945	-1.6399151	- 27.67206 6	1.51E- 168	1.44E- 165	<i>asph</i>
ENSDARG0000005 9231	-4.6708669	- 27.61614 8	7.12E- 168	6.45E- 165	<i>heph11a</i>
ENSDARG0000010 4495	-2.9700626	- 27.58469 3	1.70E- 167	1.48E- 164	<i>pltp</i>
ENSDARG0000006 9431	-3.0824949	- 27.32473 6	2.16E- 164	1.80E- 161	<i>slc26a4</i>
ENSDARG0000005 5064	-2.1691707	- 27.11674 8	6.25E- 162	5.01E- 159	<i>prdx5</i>
ENSDARG0000005 8179	-2.5277286	- 25.98306 1	7.70E- 149	5.73E- 146	<i>ackr3b</i>

ENSDARG0000005 7477	-2.5469905	- 25.77529 3	1.68E- 146	1.21E- 143	<i>lrrc28</i>
ENSDARG0000000 4363	-2.9898263	- 25.56499 8	3.74E- 144	2.60E- 141	<i>lhfp16</i>
ENSDARG0000000 0690	-2.4437712	- 24.88025 9	1.22E- 136	8.18E- 134	<i>syp12b</i>
ENSDARG0000001 3128	-2.2268619	- 24.46843 1	3.20E- 132	2.02E- 129	<i>pdk1</i>
ENSDARG0000004 0503	-1.6860372	- 24.07752 4	4.30E- 128	2.64E- 125	<i>LOC100334443</i>
ENSDARG0000008 7373	-3.8335987	- 23.95122 1	8.97E- 127	5.34E- 124	<i>dnah5</i>
ENSDARG0000010 4540	-1.6382774	- 23.90786	2.54E- 126	1.47E- 123	<i>cyp2aa8</i>
ENSDARG0000003 9082	-2.0661269	- 23.58144 6	5.98E- 123	3.37E- 120	<i>zgc:123010</i>
ENSDARG0000005 9933	-2.1717824	- 23.27599 1	7.76E- 120	4.15E- 117	<i>plpp3</i>
ENSDARG0000009 9430	-1.7575903	- 23.23563 9	1.99E- 119	1.04E- 116	<i>atg16l1</i>
ENSDARG0000006 0106	-3.1054898	- 23.07150 9	8.95E- 118	4.55E- 115	<i>crb2a</i>
ENSDARG0000010 0513	-2.2449992	- 23.01893 5	3.01E- 117	1.50E- 114	<i>rps27l</i>

ENSDARG0000007 8640	-2.8197986	- 22.95689	1.26E- 116	5.96E- 114	<i>fam117aa</i>
ENSDARG0000009 3760	-1.9268311	- 22.72752 1	2.39E- 114	1.11E- 111	<i>si:ch211- 197h24.9</i>
ENSDARG0000009 2035	-2.2804157	- 22.51857 9	2.73E- 112	1.24E- 109	<i>si:ch211- 156j16.1</i>
ENSDARG0000006 2045	-5.8601773	- 22.47112	7.96E- 112	3.53E- 109	<i>il1rapl1a</i>
ENSDARG0000011 4589	-3.8664665	- 22.01728 8	1.97E- 107	8.20E- 105	<i>jam3a</i>
ENSDARG0000010 5127	-1.3387852	- 22.01326 6	2.15E- 107	8.79E- 105	<i>si:ch211- 155e24.3</i>
ENSDARG0000000 1879	-1.7375749	- 21.99519 5	3.20E- 107	1.28E- 104	<i>nav2b</i>
ENSDARG0000003 5559	-1.3201451	- 21.93295 3	1.26E- 106	4.96E- 104	<i>tp53</i>
ENSDARG0000010 1735	-2.0981219	- 21.90535 3	2.31E- 106	8.92E- 104	<i>chn1</i>
ENSDARG0000001 1533	-1.6495795	- 21.88871 3	3.33E- 106	1.26E- 103	<i>sema6dl</i>
ENSDARG0000000 4621	-2.7866373	- 21.81426 6	1.70E- 105	6.32E- 103	<i>gpm6ab</i>
ENSDARG0000003 2831	-3.9589981	- 21.74530 5	7.65E- 105	2.80E- 102	<i>htra1a</i>

ENSDARG0000004 1107	-3.3693649	- 21.69419 8	2.33E- 104	8.37E- 102	<i>cftr</i>
ENSDARG0000002 7088	-1.7681707	- 21.54111 4	6.41E- 103	2.27E- 100	<i>ptgdsb.1</i>
ENSDARG0000007 9403	-3.3022351	- 21.37050 7	2.51E- 101	8.73E- 99	<i>si:dkey- 204111.1</i>
ENSDARG0000010 3980	-1.6456901	- 21.30624	9.94E- 101	3.40E- 98	<i>ets2</i>
ENSDARG0000002 4694	-1.9594262	- 21.29167 9	1.36E- 100	4.56E- 98	<i>myo1b</i>
ENSDARG0000002 6322	-2.6326273	- 21.24079 5	4.01E- 100	1.33E- 97	<i>dhrs13a.1</i>
ENSDARG0000010 9678	-1.6401933	- 21.23366 2	4.67E- 100	1.52E- 97	<i>LOC101885164</i>
ENSDARG0000004 5694	-1.5152661	- 21.20323	8.92E- 100	2.86E- 97	<i>prickle1b</i>
ENSDARG0000007 6557	-3.5901014	- 21.13649 9	3.67E- 99	1.14E- 96	<i>bicc2</i>
ENSDARG0000009 3659	-4.1447849	- 21.04006 3	2.82E- 98	8.65E- 96	<i>entpd5b</i>
ENSDARG0000007 0047	-2.1665852	- 20.83439 6	2.11E- 96	6.29E- 94	<i>rgs4</i>
ENSDARG0000000 1976	-5.1749727	- 20.74686 4	1.31E- 95	3.84E- 93	<i>fkbp16</i>

ENSDARG0000002 0000	-1.8172728	- 20.67845 5	5.41E- 95	1.57E- 92	<i>sh3bp4a</i>
ENSDARG0000003 0289	-2.6108705	- 20.66223	7.58E- 95	2.16E- 92	<i>jag1a</i>
ENSDARG0000006 3292	-1.8584874	- 20.54013 9	9.43E- 94	2.66E- 91	<i>ssuh2rs1</i>
ENSDARG0000011 1625	-1.6675745	- 20.51189 5	1.69E- 93	4.69E- 91	<i>wu:fc75a09</i>
ENSDARG0000008 6842	-1.5996981	- 20.28332 1	1.81E- 91	4.89E- 89	<i>dap1b</i>
ENSDARG0000005 6561	-1.4190332	- 20.22099 7	6.40E- 91	1.71E- 88	<i>asb11</i>
ENSDARG0000005 9510	-2.6734798	- 19.92819 4	2.32E- 88	6.04E- 86	<i>coq2</i>
ENSDARG0000010 1936	-2.0786075	- 19.84553 1	1.20E- 87	3.06E- 85	<i>tmem106ba</i>
ENSDARG0000006 0871	-5.9291138	- 19.70839 7	1.83E- 86	4.48E- 84	<i>mctp1b</i>
ENSDARG0000007 0669	-2.0173344	- 19.67468 4	3.55E- 86	8.61E- 84	<i>cxcr3.3</i>
ENSDARG0000002 0953	-2.3948785	- 19.51107 2	8.84E- 85	2.12E- 82	<i>dnajb6b</i>
ENSDARG0000010 3277	-2.4253069	- 19.46282 6	2.27E- 84	5.38E- 82	<i>cyp24a1</i>

ENSDARG0000003 7804	-2.5163016	- 19.45038 8	2.89E- 84	6.77E- 82	<i>phlda3</i>
ENSDARG0000005 8325	-2.1673103	- 19.32886	3.07E- 83	7.11E- 81	<i>casp8</i>
ENSDARG0000003 7997	-1.624115	- 19.17074 1	6.50E- 82	1.47E- 79	<i>tubb5</i>
ENSDARG0000002 7582	-5.9200339	- 18.99947 6	1.72E- 80	3.82E- 78	<i>angptl7</i>
ENSDARG0000007 6895	-1.3790709	- 18.98782 6	2.15E- 80	4.72E- 78	<i>flrt3</i>
ENSDARG0000001 8312	-1.3748433	- 18.84070 3	3.50E- 79	7.53E- 77	<i>rell1</i>
ENSDARG0000007 0453	-3.4665338	- 18.61891 5	2.26E- 77	4.71E- 75	<i>gch1</i>
ENSDARG0000006 9494	-2.1734088	- 18.60264 6	3.06E- 77	6.31E- 75	<i>tmem223</i>
ENSDARG0000008 6374	-2.688365	- 18.58157 5	4.53E- 77	9.26E- 75	<i>isg15</i>
ENSDARG0000010 2176	-1.8956064	- 18.32909 5	4.85E- 75	9.81E- 73	<i>pla1a</i>
ENSDARG0000007 0426	-1.4559278	- 18.29312 4	9.39E- 75	1.88E- 72	<i>chac1</i>
ENSDARG0000007 1871	-3.3232223	- 18.23593 8	2.68E- 74	5.31E- 72	<i>glod5</i>

ENSDARG0000010 3667	-2.4722017	- 18.23494 9	2.73E- 74	5.36E- 72	<i>gga3a</i>
ENSDARG0000001 5887	-2.0520336	- 18.16736	9.36E- 74	1.82E- 71	<i>b2ml</i>
ENSDARG0000009 7753	-1.7172053	- 18.07022 2	5.47E- 73	1.06E- 70	<i>slirp</i>
ENSDARG0000009 5914	-3.3746348	- 18.01627 1	1.45E- 72	2.75E- 70	<i>LOC110439871</i>
ENSDARG0000003 5332	-1.5532901	- 17.99760 6	2.03E- 72	3.82E- 70	<i>cfap298</i>
ENSDARG0000001 0878	-1.2639627	- 17.90882	1.01E- 71	1.87E- 69	<i>cdkn1ca</i>
ENSDARG0000008 6786	-2.0956085	- 17.79064 3	8.35E- 71	1.53E- 68	<i>mfsd10</i>
ENSDARG0000005 9843	-1.4732741	- 17.64476 3	1.12E- 69	2.02E- 67	<i>fam135a</i>
ENSDARG0000003 5557	-1.7267802	- 17.51210 3	1.16E- 68	2.08E- 66	<i>gabarapa</i>
ENSDARG0000010 1752	-1.6888553	- 17.48699 5	1.80E- 68	3.21E- 66	<i>plekhg5b</i>
ENSDARG0000009 4160	-1.3360106	- 17.46756 4	2.53E- 68	4.47E- 66	<i>si:ch73- 160h15.3</i>
ENSDARG0000000 4880	-2.3315274	- 17.30003 2	4.70E- 67	8.17E- 65	<i>acsm3</i>

ENSDARG0000007 5954	-2.5020388	- 17.17891 1	3.82E- 66	6.53E- 64	<i>serpinh1a</i>
ENSDARG0000006 8965	-1.4424455	- 17.07341 3	2.34E- 65	3.97E- 63	<i>nrip1a</i>
ENSDARG0000003 8012	-3.4553087	- 16.96600 1	1.47E- 64	2.43E- 62	<i>fam83fa</i>
ENSDARG0000007 9611	-1.7143024	- 16.95902	1.65E- 64	2.71E- 62	<i>sema4c</i>
ENSDARG0000003 9997	-1.8138427	- 16.94112 1	2.24E- 64	3.65E- 62	<i>ptp4a3a</i>
ENSDARG0000000 1807	-3.1188481	- 16.93539 7	2.47E- 64	3.99E- 62	<i>tnfrsf21</i>
ENSDARG0000004 0764	-0.9242393	- 16.86056	8.78E- 64	1.41E- 61	<i>id1</i>
ENSDARG0000006 3297	-1.7744101	- 16.82141 2	1.70E- 63	2.69E- 61	<i>abcb6a</i>
ENSDARG0000003 9269	-1.8108864	- 16.79452 3	2.68E- 63	4.20E- 61	<i>arg2</i>
ENSDARG0000005 7683	-1.0815321	- 16.77564 8	3.68E- 63	5.72E- 61	<i>mcm6</i>
ENSDARG0000005 9792	-1.4673497	- 16.76691 7	4.26E- 63	6.58E- 61	<i>trpm5</i>
ENSDARG0000006 0862	-3.5494051	- 16.63404 3	3.95E- 62	6.05E- 60	<i>atxn1b</i>

ENSDARG0000007 4143	-2.0646498	- 16.52150 6	2.57E- 61	3.88E- 59	<i>myo10l3</i>
ENSDARG0000004 3716	-2.1211945	- 16.47027 9	6.00E- 61	9.00E- 59	<i>cldn5a</i>
ENSDARG0000001 9856	-2.9197795	- 16.44896 9	8.53E- 61	1.27E- 58	<i>atp1a1b</i>
ENSDARG0000001 9208	-3.8204995	- 16.42927 6	1.18E- 60	1.75E- 58	<i>camsap1a</i>
ENSDARG0000003 4518	-2.2748468	- 16.40741	1.69E- 60	2.48E- 58	<i>trdmt1</i>
ENSDARG0000007 9027	-1.7589308	- 16.36720 2	3.28E- 60	4.78E- 58	<i>chsy1</i>
ENSDARG0000007 9373	-6.3911749	- 16.26060 5	1.88E- 59	2.70E- 57	<i>fosl1b</i>
ENSDARG0000000 3411	-1.3745512	- 16.25275 9	2.14E- 59	3.05E- 57	<i>foxa2</i>
ENSDARG0000007 9378	-1.3729478	- 16.19394 9	5.56E- 59	7.84E- 57	<i>phldb1b</i>
ENSDARG0000003 0297	-1.5670135	- 16.18174 7	6.78E- 59	9.49E- 57	<i>sox13</i>
ENSDARG0000010 0702	-2.5677275	- 16.10177	2.48E- 58	3.45E- 56	<i>zgc:174680</i>
ENSDARG0000000 7077	-2.1404516	- 16.06874 2	4.23E- 58	5.83E- 56	<i>ankrd50l</i>

ENSDARG0000003 9117	-1.5648684	- 15.94625 5	3.03E- 57	4.12E- 55	<i>tefa</i>
ENSDARG0000004 5294	-1.2601843	- 15.92802 6	4.05E- 57	5.48E- 55	<i>sh3glb1a</i>
ENSDARG0000001 4098	-1.6885444	- 15.89929 4	6.41E- 57	8.62E- 55	<i>pkd2</i>
ENSDARG0000006 1817	-4.7553091	-15.7976	3.23E- 56	4.29E- 54	<i>kif1aa</i>
ENSDARG0000002 7112	-1.3527615	- 15.79589 1	3.32E- 56	4.38E- 54	<i>ephb4b</i>
ENSDARG0000001 6837	-1.8227111	- 15.79496 4	3.37E- 56	4.42E- 54	<i>glipr2l</i>
ENSDARG0000007 5048	-1.1461057	- 15.67537 3	2.23E- 55	2.85E- 53	<i>lonrf1</i>
ENSDARG0000009 4300	-1.6070197	- 15.52291 2	2.43E- 54	3.03E- 52	<i>nupr1a</i>
ENSDARG0000000 4877	-1.3619834	- 15.49322 3	3.85E- 54	4.75E- 52	<i>rock2b</i>
ENSDARG0000003 5808	-2.0348903	- 15.36760 9	2.70E- 53	3.25E- 51	<i>clcn4</i>
ENSDARG0000000 7329	-1.0787012	- 15.32984	4.83E- 53	5.69E- 51	<i>tbx16</i>
ENSDARG0000001 8693	-1.0380466	- 15.29291 9	8.52E- 53	9.98E- 51	<i>cdh2</i>

ENSDARG0000010 2300	-1.0133703	- 15.26017 2	1.41E- 52	1.62E- 50	<i>ca9</i>
ENSDARG0000003 5018	-1.5608841	- 15.14746 1	7.88E- 52	9.02E- 50	<i>thy1</i>
ENSDARG0000009 6603	-1.4705631	- 14.95732	1.40E- 50	1.57E- 48	<i>bmb</i>
ENSDARG0000001 8968	-1.0282933	- 14.93920 5	1.83E- 50	2.05E- 48	<i>acvr1ba</i>
ENSDARG0000001 1582	-1.9846064	- 14.91306 1	2.71E- 50	3.01E- 48	<i>sox5</i>
ENSDARG0000000 9544	-1.4431963	- 14.90816 3	2.92E- 50	3.22E- 48	<i>cldnb</i>
ENSDARG0000003 5519	-1.1180286	- 14.88935 8	3.86E- 50	4.24E- 48	<i>hsth1l</i>
ENSDARG0000003 5198	-1.5768686	- 14.85713 4	6.25E- 50	6.83E- 48	<i>gcnt4a</i>
ENSDARG0000007 6371	-3.0077846	- 14.83467 6	8.74E- 50	9.44E- 48	<i>shroom2b</i>
ENSDARG0000004 2900	-1.8590576	- 14.77308	2.18E- 49	2.32E- 47	<i>gtpbp1l</i>
ENSDARG0000003 5719	-0.9540029	- 14.73625 5	3.77E- 49	3.99E- 47	<i>arl5c</i>
ENSDARG0000006 3665	-1.5538808	- 14.72287 5	4.60E- 49	4.84E- 47	<i>mat2al</i>

ENSDARG0000004 2727	-1.7512189	- 14.56526 2	4.67E- 48	4.80E- 46	<i>exo5</i>
ENSDARG0000003 1126	-1.33444	- 14.56178 4	4.92E- 48	5.02E- 46	<i>notum1a</i>
ENSDARG0000007 8567	-1.2616927	- 14.54527 5	6.26E- 48	6.36E- 46	<i>lonrf1l</i>
ENSDARG0000007 5733	-1.2100009	- 14.49988 1	1.21E- 47	1.22E- 45	<i>zyx</i>
ENSDARG0000006 9922	-1.8463723	- 14.47965 4	1.63E- 47	1.62E- 45	<i>popdc2</i>
ENSDARG0000001 9747	-0.9013845	- 14.47076 2	1.85E- 47	1.84E- 45	<i>hsd3b2</i>
ENSDARG0000001 0008	-4.2104805	- 14.43698	3.03E- 47	2.96E- 45	<i>vim</i>
ENSDARG0000010 9443	-1.3446024	- 14.40231 5	5.00E- 47	4.87E- 45	<i>eml4</i>
ENSDARG0000001 2818	-1.3281256	- 14.37977 5	6.93E- 47	6.69E- 45	<i>csnk2a2a</i>
ENSDARG0000001 7121	-4.0150396	- 14.35207 6	1.03E- 46	9.93E- 45	<i>mafba</i>
ENSDARG0000004 3323	-1.6014098	- 14.29224 2	2.45E- 46	2.34E- 44	<i>lnx1</i>
ENSDARG0000004 4511	-0.950132	- 14.25398 1	4.23E- 46	4.03E- 44	<i>etv5b</i>

ENSDARG0000009 0369	-3.5773082	- 14.23523 5	5.54E- 46	5.25E- 44	<i>zgc:86896</i>
ENSDARG0000000 4402	-1.2947537	- 14.17864 1	1.24E- 45	1.17E- 43	<i>elovl6</i>
ENSDARG0000001 0962	-1.3414694	- 14.16882 1	1.43E- 45	1.34E- 43	<i>fkbp7</i>
ENSDARG0000006 9961	-3.5288068	- 14.15411 6	1.76E- 45	1.65E- 43	<i>il21r.1</i>
ENSDARG0000005 7688	-1.9669403	- 13.96057 5	2.71E- 44	2.50E- 42	<i>ical1</i>
ENSDARG0000006 2785	-1.5063961	- 13.95856 6	2.79E- 44	2.56E- 42	<i>adpgk2</i>
ENSDARG0000001 5293	-1.5216769	- 13.91680 4	5.01E- 44	4.56E- 42	<i>fam110a</i>
ENSDARG0000006 8176	-1.5404871	- 13.89761 1	6.55E- 44	5.94E- 42	<i>uqcc1</i>
ENSDARG0000007 4023	-0.8565186	- 13.84736 4	1.32E- 43	1.19E- 41	<i>rbms1a</i>
ENSDARG0000009 7746	-5.9778191	- 13.83652 3	1.53E- 43	1.37E- 41	<i>si:rp71-7711.1</i>
ENSDARG0000001 0563	-0.9764008	- 13.80340 5	2.43E- 43	2.17E- 41	<i>spopla</i>
ENSDARG0000000 3961	-2.4149227	- 13.77298 9	3.71E- 43	3.26E- 41	<i>parp3</i>

ENSDARG0000005 2387	-2.4392806	- 13.66084 9	1.74E- 42	1.51E- 40	<i>pou2f3</i>
ENSDARG0000008 7452	-2.2398895	- 13.54449 8	8.54E- 42	7.30E- 40	<i>pel3</i>
ENSDARG0000003 1911	-0.9375552	- 13.51409	1.29E- 41	1.10E- 39	<i>si:ch211- 191a24.3</i>
ENSDARG0000001 7004	-0.904399	- 13.46854 1	2.40E- 41	2.02E- 39	<i>myo10</i>
ENSDARG0000003 5556	-1.1152828	- 13.41203 1	5.14E- 41	4.32E- 39	<i>rps6ka3a</i>
ENSDARG0000007 0914	-1.0053541	- 13.40791 6	5.43E- 41	4.55E- 39	<i>dusp6</i>
ENSDARG0000003 4773	-0.8770728	- 13.37487 9	8.48E- 41	7.01E- 39	<i>ncapd3</i>
ENSDARG0000010 1127	-0.8684172	- 13.36672 5	9.46E- 41	7.80E- 39	<i>map1lc3b</i>
ENSDARG0000003 6965	-1.8279238	- 13.21393 1	7.29E- 40	5.94E- 38	<i>rnf24</i>
ENSDARG0000003 4717	-2.167075	- 13.20892 7	7.79E- 40	6.32E- 38	<i>def6c</i>
ENSDARG0000002 3217	-5.2584447	- 13.15800 4	1.53E- 39	1.23E- 37	<i>crema</i>
ENSDARG0000009 1111	-1.3896676	- 13.10023 9	3.28E- 39	2.59E- 37	<i>si:ch211- 15b10.6</i>

ENSDARG0000002 0623	-0.977094	- 13.08568 7	3.98E- 39	3.13E- 37	<i>baxa</i>
ENSDARG0000001 9564	-1.0966858	- 13.03712	7.52E- 39	5.87E- 37	<i>asap2b</i>
ENSDARG0000001 2199	-0.8624117	- 13.01524 6	1.00E- 38	7.80E- 37	<i>gpt2</i>
ENSDARG0000010 3522	-0.9747205	- 13.00513	1.14E- 38	8.87E- 37	<i>si:zfos- 943e10.1</i>
ENSDARG0000003 2242	-1.8582353	- 12.99987 4	1.23E- 38	9.46E- 37	<i>tnnt2c</i>
ENSDARG0000001 3250	-0.88943	- 12.97835 4	1.62E- 38	1.25E- 36	<i>tars1</i>
ENSDARG0000004 1904	-1.4986559	- 12.96210 8	2.01E- 38	1.53E- 36	<i>ankzf1</i>
ENSDARG0000006 8369	-4.1890622	- 12.95939 9	2.08E- 38	1.58E- 36	<i>angptl2b</i>
ENSDARG0000004 3663	-1.1265944	- 12.94794 4	2.41E- 38	1.83E- 36	<i>faub</i>
ENSDARG0000008 6826	-0.9324846	- 12.90182	4.40E- 38	3.30E- 36	<i>sult6b1</i>
ENSDARG0000006 1844	-0.8592162	- 12.89294 8	4.93E- 38	3.69E- 36	<i>si:dkey-38p12.3</i>
ENSDARG0000006 2972	-0.8060115	- 12.87769 8	6.01E- 38	4.47E- 36	<i>hip1</i>
ENSDARG0000007 9712	-1.8942238	- 12.87019 7	6.62E- 38	4.90E- 36	<i>gal3st4</i>

ENSDARG0000006 1233	-1.8101001	- 12.80129	1.61E- 37	1.18E- 35	<i>abcc5</i>
ENSDARG0000004 2641	-1.1722483	- 12.73844 8	3.62E- 37	2.62E- 35	<i>cyp51</i>
ENSDARG0000000 8372	-1.3295816	- 12.67978 3	7.65E- 37	5.48E- 35	<i>spred2b</i>
ENSDARG0000004 3531	-1.723299	- 12.56612 6	3.24E- 36	2.25E- 34	<i>jun</i>
ENSDARG0000007 7053	-2.865614	- 12.54495	4.24E- 36	2.93E- 34	<i>ppfia3</i>
ENSDARG0000001 8508	-0.9942554	- 12.53765 8	4.65E- 36	3.20E- 34	<i>zdhhc8b</i>
ENSDARG0000003 5715	-0.6734863	- 12.51338 8	6.31E- 36	4.33E- 34	<i>marcksl1b</i>
ENSDARG0000009 0982	-1.802358	- 12.48043 5	9.55E- 36	6.48E- 34	<i>slc44a1b</i>
ENSDARG0000003 7760	-2.9251013	- 12.45529 5	1.31E- 35	8.80E- 34	<i>uncx4.1</i>
ENSDARG0000007 3684	-2.4727114	- 12.42362 3	1.95E- 35	1.30E- 33	<i>quo</i>
ENSDARG0000003 7140	-1.1124666	- 12.29767 3	9.32E- 35	6.09E- 33	<i>pfkfb1</i>
ENSDARG0000001 4646	-1.0298794	- 12.23367 3	2.05E- 34	1.32E- 32	<i>aoc2</i>

ENSDARG0000003 5694	-0.7737215	- 12.20959 1	2.76E- 34	1.76E- 32	<i>stm</i>
ENSDARG0000007 6867	-2.2979727	- 12.19009 7	3.51E- 34	2.22E- 32	<i>mrtfbb</i>
ENSDARG0000004 5815	-0.8899216	- 12.13363 2	7.01E- 34	4.41E- 32	<i>mcm10</i>
ENSDARG0000002 1241	-3.8730293	- 12.10870 3	9.50E- 34	5.96E- 32	<i>zgc:165604</i>
ENSDARG0000007 8411	-4.9849347	- 12.09824 6	1.08E- 33	6.75E- 32	<i>hsqb15</i>
ENSDARG0000004 2904	-1.508535	- 12.08906 6	1.21E- 33	7.53E- 32	<i>foxo3b</i>
ENSDARG0000011 3961	-2.0290819	- 12.06396 9	1.64E- 33	1.02E- 31	<i>nudt15</i>
ENSDARG0000010 4825	-1.0487153	- 12.03676 5	2.28E- 33	1.41E- 31	<i>trim35-7</i>
ENSDARG0000006 0169	-1.0079379	- 12.01443 8	2.98E- 33	1.85E- 31	<i>mns1</i>
ENSDARG0000003 9697	-3.4557827	- 12.00542 3	3.33E- 33	2.05E- 31	<i>shn1</i>
ENSDARG0000003 6893	-1.9232368	- 11.98824 1	4.10E- 33	2.52E- 31	<i>f13a1b</i>
ENSDARG0000000 7377	-0.7717017	- 11.96525 3	5.40E- 33	3.30E- 31	<i>odc1</i>

ENSDARG0000002 0235	-0.8763362	- 11.94093 2	7.24E- 33	4.40E- 31	<i>sept9a</i>
ENSDARG0000004 0362	-2.8540119	- 11.91889 3	9.44E- 33	5.72E- 31	<i>ehd2b</i>
ENSDARG0000009 3452	-3.1107799	- 11.91585 5	9.79E- 33	5.91E- 31	<i>LOC100535312</i>
ENSDARG0000007 8552	-0.8170947	- 11.91021 9	1.05E- 32	6.31E- 31	<i>grhl3</i>
ENSDARG0000000 2787	-1.1706048	- 11.90388 2	1.13E- 32	6.79E- 31	<i>tle3a</i>
ENSDARG0000004 4753	-0.976337	- 11.88069 5	1.49E- 32	8.91E- 31	<i>idh3b</i>
ENSDARG0000002 8396	-2.3205038	-11.8758	1.58E- 32	9.39E- 31	<i>fkbp5</i>
ENSDARG0000007 1626	-1.0622776	- 11.86256 7	1.85E- 32	1.10E- 30	<i>ptgdsb.2</i>
ENSDARG0000009 2310	-1.126448	- 11.84908 9	2.18E- 32	1.28E- 30	<i>si:ch211- 199g17.2</i>
ENSDARG0000000 4105	-2.8917382	- 11.84892 8	2.18E- 32	1.28E- 30	<i>tie1</i>
ENSDARG0000008 8202	-0.9175199	- 11.83591 1	2.55E- 32	1.49E- 30	<i>sh3d19</i>
ENSDARG0000001 0655	-1.1554397	- 11.81915 4	3.11E- 32	1.81E- 30	<i>ppm1k</i>

ENSDARG0000007 8882	-1.3983232	- 11.79433 8	4.17E- 32	2.43E- 30	<i>slc22a31</i>
ENSDARG0000005 4864	-1.1947905	- 11.75810 2	6.42E- 32	3.70E- 30	<i>aplp2</i>
ENSDARG0000009 2825	-3.164912	-11.7353	8.40E- 32	4.83E- 30	<i>si:dkey-42p14.3</i>
ENSDARG0000009 4894	-5.3517676	- 11.71625 2	1.05E- 31	5.99E- 30	<i>tyrobp</i>
ENSDARG0000005 4616	-3.0425556	- 11.71138 7	1.11E- 31	6.33E- 30	<i>cldni</i>
ENSDARG0000007 8481	-2.4194681	- 11.56611 5	6.12E- 31	3.38E- 29	<i>uts2r2</i>
ENSDARG0000002 9955	-1.6142811	- 11.55437 6	7.02E- 31	3.86E- 29	<i>glb1l</i>
ENSDARG0000009 8810	-4.9954239	- 11.50552 4	1.24E- 30	6.75E- 29	<i>zgc:194678</i>
ENSDARG0000004 1137	-1.161104	- 11.36846 7	6.00E- 30	3.20E- 28	<i>dhrs13a.3</i>
ENSDARG0000007 0432	-0.7764991	- 11.31934 2	1.05E- 29	5.60E- 28	<i>ino80</i>
ENSDARG0000001 5224	-0.7912064	- 11.26389 6	1.98E- 29	1.04E- 27	<i>cd2ap</i>
ENSDARG0000006 1294	-0.7649768	- 11.25485 6	2.19E- 29	1.15E- 27	<i>arhgap5</i>

ENSDARG0000010 3546	-1.0017636	- 11.22501 1	3.07E- 29	1.61E- 27	<i>purbb</i>
ENSDARG0000000 7289	-2.2986696	- 11.21430 1	3.47E- 29	1.81E- 27	<i>metrnlb</i>
ENSDARG0000006 2489	-1.5732607	- 11.20874 9	3.69E- 29	1.92E- 27	<i>Mar-08</i>
ENSDARG0000011 7312	-0.7556669	- 11.20390 9	3.90E- 29	2.02E- 27	<i>sb:cb288</i>
ENSDARG0000010 4068	-1.3065706	- 11.11798 7	1.03E- 28	5.18E- 27	<i>gstp1</i>
ENSDARG0000011 0967	-1.1549228	- 11.10521 2	1.18E- 28	5.96E- 27	<i>LOC101884738</i>
ENSDARG0000010 1576	-0.6677803	- 11.08957 5	1.41E- 28	7.08E- 27	<i>tbxta</i>
ENSDARG0000007 6667	-0.9055876	- 11.08887 8	1.42E- 28	7.10E- 27	<i>ccng1</i>
ENSDARG0000009 6797	-1.323799	- 11.07941	1.58E- 28	7.88E- 27	<i>si:ch73- 127m5.2</i>
ENSDARG0000001 1027	-0.729268	- 11.07093	1.74E- 28	8.62E- 27	<i>fgfr1a</i>
ENSDARG0000007 0448	-0.8644788	- 11.03813 1	2.50E- 28	1.24E- 26	<i>grk4</i>
ENSDARG0000006 1223	-4.3501559	- 11.03351 9	2.63E- 28	1.30E- 26	<i>ogfr11</i>

ENSDARG0000001 8643	-1.2629518	- 11.02318 4	2.95E- 28	1.45E- 26	<i>igf2a</i>
ENSDARG0000001 3575	-0.8438448	- 11.01507 8	3.23E- 28	1.58E- 26	<i>rfx2</i>
ENSDARG0000008 6317	-1.6132535	- 11.00326 4	3.69E- 28	1.79E- 26	<i>si:ch211- 67e16.4</i>
ENSDARG0000005 2435	-0.9902625	- 10.99511 2	4.03E- 28	1.95E- 26	<i>spred2a</i>
ENSDARG0000005 3113	-1.5996633	- 10.96996 5	5.33E- 28	2.57E- 26	<i>ly75</i>
ENSDARG0000004 2825	-0.8504232	- 10.95519 9	6.27E- 28	3.02E- 26	<i>rnf169</i>
ENSDARG0000004 3662	-3.0921978	- 10.94627 1	6.92E- 28	3.33E- 26	<i>cnih2</i>
ENSDARG0000001 7821	-0.7897182	- 10.93079 5	8.21E- 28	3.92E- 26	<i>gata5</i>
ENSDARG0000005 6480	-0.9268571	- 10.86387 6	1.71E- 27	8.10E- 26	<i>cpvl</i>
ENSDARG0000005 3263	-0.9215712	- 10.82618 1	2.59E- 27	1.22E- 25	<i>zgc:113372</i>
ENSDARG0000001 0487	-0.6112189	- 10.81498 6	2.92E- 27	1.38E- 25	<i>sae1</i>
ENSDARG0000003 6080	-0.7337566	- 10.79634 8	3.58E- 27	1.67E- 25	<i>cd81a</i>

ENSDARG0000009 8228	-0.7924285	- 10.79212 3	3.75E- 27	1.74E- 25	<i>thrap3b</i>
ENSDARG0000003 8694	-0.7332462	- 10.78779 8	3.93E- 27	1.82E- 25	<i>zgc:101744</i>
ENSDARG0000000 4023	-4.8569881	- 10.76941 8	4.80E- 27	2.21E- 25	<i>isl1</i>
ENSDARG0000007 9355	-5.2006135	- 10.75912 6	5.37E- 27	2.45E- 25	<i>flrt2</i>
ENSDARG0000003 5676	-0.6289164	- 10.74214 9	6.45E- 27	2.93E- 25	<i>ptp4a2b</i>
ENSDARG0000006 2983	-0.6663456	- 10.74040 2	6.58E- 27	2.98E- 25	<i>fbxo18</i>
ENSDARG0000010 2394	-0.8288962	- 10.73347 6	7.09E- 27	3.19E- 25	<i>ap1g1</i>
ENSDARG0000005 2690	-0.9299326	- 10.72306 6	7.93E- 27	3.56E- 25	<i>arrdc3a</i>
ENSDARG0000000 2285	-1.3318125	- 10.70325	9.83E- 27	4.40E- 25	<i>mcoln1a</i>
ENSDARG0000004 0526	-0.8786095	- 10.64741 4	1.79E- 26	7.90E- 25	<i>s1pr5a</i>
ENSDARG0000008 8247	-2.6812918	- 10.63109 3	2.14E- 26	9.39E- 25	<i>si:ch1073- 303k11.2</i>
ENSDARG0000010 3383	-4.4884331	- 10.62099 2	2.38E- 26	1.04E- 24	<i>dnah9</i>

ENSDARG0000005 3455	-1.6652693	- 10.59329 1	3.20E- 26	1.40E- 24	<i>ccdc103</i>
ENSDARG0000002 6655	-1.379217	- 10.57693 2	3.81E- 26	1.65E- 24	<i>tspo</i>
ENSDARG0000011 2943	-1.7752099	- 10.57672 3	3.82E- 26	1.65E- 24	<i>LOC100536657</i>
ENSDARG0000000 7720	-0.7988827	- 10.57252 7	4.00E- 26	1.72E- 24	<i>sub1b</i>
ENSDARG0000000 9499	-2.6530859	- 10.56849 5	4.17E- 26	1.80E- 24	<i>syne1a</i>
ENSDARG0000000 4714	-1.4936021	- 10.55619 1	4.76E- 26	2.03E- 24	<i>tcf12</i>
ENSDARG0000002 0086	-2.5167011	- 10.54925 2	5.12E- 26	2.18E- 24	<i>nuak1a</i>
ENSDARG0000001 9438	-1.0602719	- 10.54205 2	5.53E- 26	2.34E- 24	<i>rnf13</i>
ENSDARG0000006 3407	-4.5567624	- 10.52677 5	6.50E- 26	2.74E- 24	<i>adamtsl4</i>
ENSDARG0000006 8401	-0.6448925	- 10.51813 4	7.13E- 26	3.00E- 24	<i>yap1</i>
ENSDARG0000009 4458	-4.4454785	- 10.49624 8	8.99E- 26	3.77E- 24	<i>si:dkey- 65b12.12</i>
ENSDARG0000004 2418	-2.9615203	- 10.47930 5	1.08E- 25	4.50E- 24	<i>galnt17</i>

ENSDARG0000009 8946	-0.7966025	- 10.44394 7	1.56E- 25	6.51E- 24	<i>pald1a</i>
ENSDARG0000008 9307	-1.1890048	- 10.41102 7	2.21E- 25	9.17E- 24	<i>pmaip1</i>
ENSDARG0000009 9161	-2.7938941	- 10.38626 1	2.86E- 25	1.19E- 23	<i>dnaaf4</i>
ENSDARG0000010 2211	-4.4843903	- 10.35886 9	3.81E- 25	1.57E- 23	<i>csf3a</i>
ENSDARG0000002 5567	-0.7459696	- 10.30738 4	6.53E- 25	2.65E- 23	<i>nf2b</i>
ENSDARG0000005 8366	-0.8540714	- 10.30161 5	6.93E- 25	2.79E- 23	<i>si:dkey-222f8.3</i>
ENSDARG0000003 8910	-1.18164	- 10.27148	9.47E- 25	3.80E- 23	<i>pccb</i>

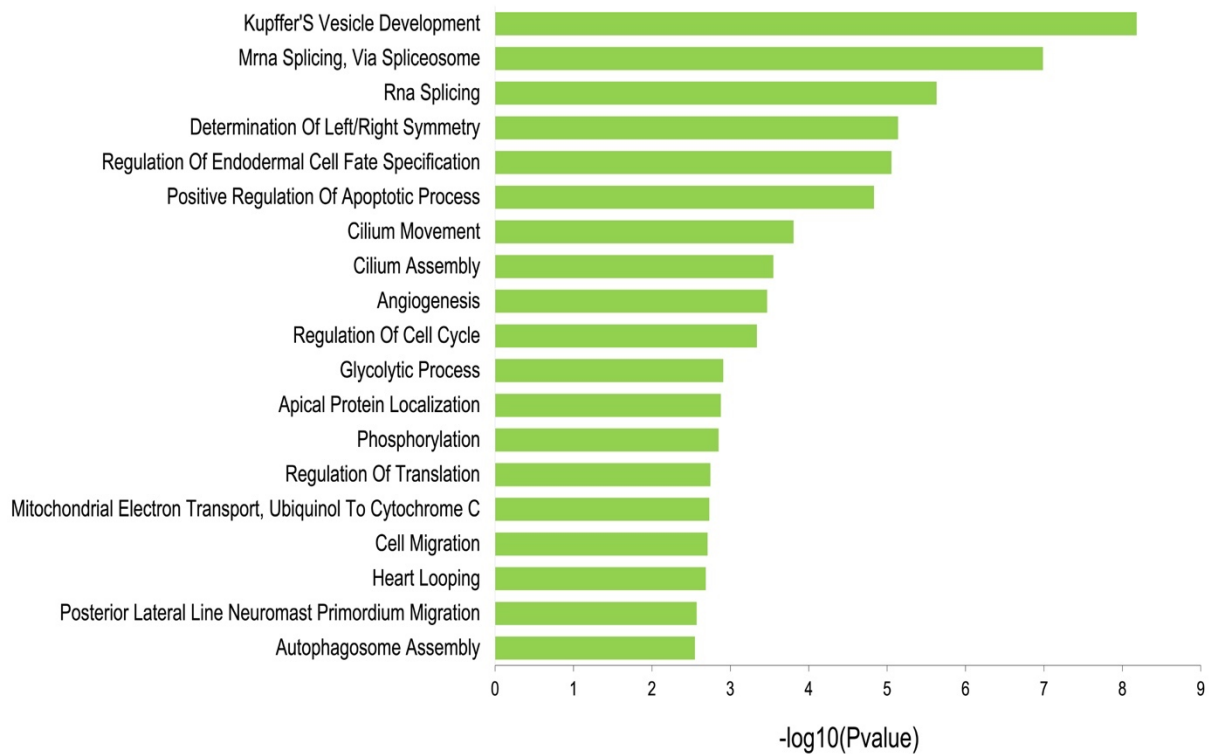


Figure 7-3: Top GO terms associated with transcripts that were upregulated in sox32 OE vs control according GO analysis using DAVID.

Table 7.14: List of top 200 DARs between sox17E vs sox17N ranked by FDR from Diffbind. Fold enrichment score > 0 represents sox17N DARs while sox17E DARs have a fold enrichment score < 0. FDR = False discovery rate

Chromosome	Chromosome start	Chromosome end	Fold enrichment	FDR
chr11	42035126	42039448	1.92	2.50E-29
chr18	12147070	12148621	1.51	9.80E-23
chr13	39899825	39901164	1.71	1.07E-17
chr11	42040979	42045977	1.87	1.55E-17
chr13	44497989	44498907	1.79	1.55E-17
chr19	41565268	41566618	1.64	1.55E-17
chr8	14511036	14512742	1.28	3.17E-17
chr7	46384248	46386527	1.39	3.17E-17
chr24	10636497	10637722	1.57	6.76E-16
chr24	8542991	8545047	1.36	7.04E-16
chr23	24067581	24073452	-1.59	1.83E-15
chr3	55837985	55840824	1.24	2.37E-15
chr11	39226793	39229071	1.43	2.43E-15
chr24	27680892	27682154	1.42	3.16E-15

chr3	53683695	53685521	1.41	3.16E-15
chr3	53541034	53544858	1.15	3.16E-15
chr11	42029085	42031697	1.48	3.29E-15
chr13	5290630	5291750	1.87	4.56E-15
chr14	40821190	40822186	-2.16	5.72E-15
chr11	42066583	42071458	1.79	6.11E-15
chr21	20703507	20705308	1.52	8.42E-15
chr24	4584878	4586351	1.17	1.04E-14
chr2	26428553	26429870	1.37	1.52E-14
chr16	32517491	32518843	1.55	1.78E-14
chr2	18280628	18281987	1.54	1.98E-14
chr10	37092730	37093804	1.33	2.28E-14
chr13	33793098	33794614	1.59	3.39E-14
chr10	31548591	31549634	1.85	3.60E-14
chr14	39906595	39907979	1.78	4.64E-14
chr23	7838278	7839580	1.26	4.64E-14
chr16	449465	451279	1.7	5.50E-14
chr19	28217587	28219141	1.52	5.55E-14
chr7	49778932	49780269	1.6	7.23E-14
chr6	14777683	14779606	1.64	8.57E-14
chr5	17321815	17323323	1.32	8.57E-14
chr11	42056074	42058464	1.68	8.57E-14
chr6	45910459	45911408	1.67	1.26E-13
chr20	39958403	39959765	1.54	1.99E-13
chr19	28157391	28158619	1.83	2.22E-13
chr21	23608279	23609225	1.2	2.22E-13
chr16	33248584	33249613	1.34	2.43E-13
chr10	31580933	31582850	1.46	2.89E-13
chr6	14891618	14893424	1.71	2.89E-13
chr6	43115685	43116837	1.89	3.44E-13
chr21	23288142	23291009	1.24	6.06E-13
chr14	40201768	40203306	1.71	9.47E-13
chr2	4083994	4085365	1.31	9.54E-13
chr13	28385449	28387099	1.6	1.20E-12
chr18	15372667	15374901	1.61	1.24E-12

chr15	21368255	21369829	1.52	1.29E-12
chr19	37619826	37624307	1.16	1.48E-12
chr24	8090361	8091479	1.61	1.54E-12
chr20	33340460	33341730	1.31	1.58E-12
chr14	36372713	36375292	1.22	1.67E-12
chr16	16837693	16838911	1.61	1.80E-12
chr3	46886692	46888837	1.23	1.83E-12
chr7	35884324	35886258	1.6	1.85E-12
chr21	20712838	20717753	1.45	1.94E-12
chr24	21245538	21247808	1.06	1.95E-12
chr16	42903610	42905333	1.27	2.09E-12
chr5	272866	274295	1.6	2.09E-12
chr18	6093357	6095390	1.37	2.09E-12
chr11	25822784	25825163	1.48	2.09E-12
chr18	23757970	23758944	1.77	2.16E-12
chr21	7433325	7435116	1.62	2.24E-12
chr7	14334164	14335400	1.33	2.24E-12
chr7	68753694	68754952	1.27	2.24E-12
chr8	10323173	10324165	-2.16	2.24E-12
chr7	36047528	36050384	1.66	2.24E-12
chr19	41553125	41554119	1.79	2.24E-12
chr12	22291956	22293222	1.5	2.41E-12
chr10	42271956	42273249	1.05	2.41E-12
chr14	32660694	32661022	-2.66	2.75E-12
chr17	1485183	1486607	1.32	2.77E-12
chr20	10104801	10107687	1.26	3.45E-12
chr2	59088251	59089649	1.87	3.54E-12
chr2	35692402	35693204	1.24	3.75E-12
chr10	3162251	3163746	1.26	4.02E-12
chr7	12069101	12070860	1.44	4.12E-12
chr7	11799826	11801579	1.15	4.12E-12
chr15	29768794	29770912	1.36	4.12E-12
chr11	28339539	28341962	1.36	4.58E-12
chr2	26986934	26988947	1.45	6.86E-12
chr17	33085529	33087008	1.92	7.66E-12

chr1	19387583	19388355	-2.59	9.46E-12
chr7	28580662	28582527	1.47	9.46E-12
chr14	32669804	32670943	1.38	9.57E-12
chr18	47884590	47886407	1.18	1.02E-11
chr15	3629713	3630781	1.7	1.03E-11
chr2	5941909	5942912	1.84	1.20E-11
chr25	2859177	2860485	1.25	1.20E-11
chr14	49695733	49697204	1.23	1.48E-11
chr7	35859378	35860676	1.46	1.48E-11
chr7	53781526	53782958	1.47	1.49E-11
chr19	29761120	29762860	0.97	1.51E-11
chr6	19815222	19818455	1.41	1.51E-11
chr5	18517227	18518138	1.45	1.54E-11
chr3	43319144	43320721	1.56	1.68E-11
chr12	31003705	31007052	1.33	1.68E-11
chr10	38131831	38132928	1.59	1.90E-11
chr25	427242	429408	1.29	1.90E-11
chr13	13250003	13252083	1.07	1.91E-11
chr16	35491253	35493105	1.23	1.91E-11
chr7	28606574	28607951	1.92	1.99E-11
chr17	24349647	24351422	1.38	2.00E-11
chr2	33105835	33110406	1.2	2.00E-11
chr11	42023081	42027065	1.72	2.00E-11
chr5	186361	188920	1.39	2.02E-11
chr24	5104615	5107052	1.96	2.29E-11
chr16	46228672	46230253	1.74	2.57E-11
chr13	52139582	52141572	1.62	2.60E-11
chr15	8516433	8518361	1.52	2.81E-11
chr11	5426182	5427262	1.7	2.87E-11
chr3	28580837	28582243	1.27	3.02E-11
chr5	50572009	50574220	1.49	3.04E-11
chr5	16087402	16088254	1.07	3.13E-11
chr12	45744983	45746197	1.25	3.13E-11
chr22	10797318	10799487	1.39	3.26E-11
chr25	412766	414662	1.64	3.33E-11

chr7	54834087	54835239	1.28	3.65E-11
chr13	46432263	46433810	1.2	3.79E-11
chr10	21851284	21852644	1.88	3.93E-11
chr16	19482663	19485134	1.24	4.07E-11
chr23	18869203	18872132	1.04	4.07E-11
chr17	36029625	36033076	1.21	4.59E-11
chr20	21411992	21414608	1.02	4.60E-11
chr10	38117134	38118590	1.69	4.66E-11
chr21	37336596	37337938	1.27	4.66E-11
chr6	55884227	55885254	-2.7	4.74E-11
chr20	21207703	21208750	1.47	5.85E-11
chr22	17677517	17679041	-2.07	5.94E-11
chr14	35062948	35064350	1.32	6.01E-11
chr24	41200540	41202015	1.86	6.09E-11
chr10	22147713	22150709	-1.61	6.54E-11
chr19	1056511	1057601	1.53	6.98E-11
chr17	1483578	1484513	1.56	7.11E-11
chr14	26637730	26640057	1.39	7.11E-11
chr6	19941710	19942704	1.89	7.30E-11
chr1	42743580	42744448	2.07	7.56E-11
chr19	28913397	28914153	1.16	7.90E-11
chr17	42032531	42034036	1.58	8.21E-11
chr17	18645921	18646765	1.32	8.28E-11
chr7	15980491	15981413	1.27	8.40E-11
chr17	7188717	7189603	1.54	8.40E-11
chr16	33487703	33489667	1.52	9.13E-11
chr1	51179371	51180478	1.96	9.45E-11
chr16	33494022	33496141	1.43	9.96E-11
chr3	48528684	48529609	1.6	9.96E-11
chr11	37746663	37748256	1.48	1.05E-10
chr10	23637946	23638855	1.58	1.11E-10
chr18	16639637	16640517	1.8	1.12E-10
chr16	16712079	16714870	1.57	1.15E-10
chr19	37605617	37607115	1.28	1.16E-10
chr1	54012099	54016053	0.97	1.20E-10

chr5	36610370	36617079	1.23	1.22E-10
chr25	36512008	36514305	1.35	1.22E-10
chr15	33058796	33060245	1.41	1.25E-10
chr10	10843680	10844946	0.87	1.26E-10
chr24	4486645	4487870	1.64	1.38E-10
chr15	12167111	12168479	1.59	1.38E-10
chr12	24828357	24829273	1.57	1.42E-10
chr18	49305888	49308225	1.52	1.45E-10
chr8	53365758	53368894	1.54	1.49E-10
chr20	40301016	40302236	1.56	1.64E-10
chr16	1520011	1520631	1.52	1.66E-10
chr22	28787963	28792849	0.85	1.86E-10
chr11	27910653	27911597	1.46	1.87E-10
chr14	42156969	42158368	1.1	1.91E-10
chr7	65856097	65858308	1.19	2.03E-10
chr7	33624507	33625828	1.55	2.04E-10
chr1	11495547	11497306	1.51	2.04E-10
chr5	36025099	36026959	1.12	2.04E-10
chr17	35996711	35998160	1.56	2.04E-10
chr1	48574163	48574646	2.31	2.06E-10
chr23	10050111	10051593	1.34	2.08E-10
chr23	27223027	27224378	1.13	2.18E-10
chr20	6708936	6711414	1.14	2.25E-10
chr20	43856152	43857224	1.31	2.25E-10
chr7	62058382	62060774	1.19	2.36E-10
chr15	21346441	21347792	1.72	2.48E-10
chr17	41883438	41886046	1.2	2.65E-10
chr23	21452855	21455761	1.63	2.68E-10
chr2	31270585	31271384	1.28	2.73E-10
chr7	25256350	25257316	2.26	2.83E-10
chr3	25721573	25722743	1.25	2.85E-10
chr21	38153447	38154064	-2.07	2.91E-10
chr16	19442000	19447244	1.38	3.16E-10
chr13	45974532	45975738	1.29	3.35E-10
chr16	33484342	33485641	1.37	3.68E-10

chr11	16353722	16356492	1.26	3.92E-10
chr20	39014751	39015535	1.01	3.92E-10
chr22	18264977	18266307	1.43	4.08E-10
chr2	3114663	3115761	1.65	4.15E-10
chr7	68097705	68099100	1.19	4.26E-10
chr1	21762042	21763598	1.23	4.41E-10
chr3	46814436	46820200	1.19	4.97E-10
chr12	36118421	36119537	1.37	4.99E-10
chr14	32638408	32640381	1.26	4.99E-10
chr6	14495169	14496531	1.68	5.13E-10

Table 7.15: List of top 200 DARs between *sox17N* vs *sox17M* ranked by FDR from Diffbind. Fold enrichment score > 0 represents *sox17N* DARs while *sox17M* DARs have a fold enrichment score < 0. FDR = False discovery rate

Chromosome	Chromosome start	Chromosome end	Fold enrichment	FDR
chr11	42035126	42039448	4.49	3.66E-95
chr16	32517491	32518843	4.94	1.38E-81
chr3	18108243	18110666	4.26	1.88E-78
chr8	14511036	14512742	3.87	2.22E-77
chr3	53683695	53685521	4.14	1.51E-75
chr14	36372713	36375292	3.99	5.01E-75
chr11	42029085	42031697	4.72	1.60E-74
chr3	55143618	55149276	-4.08	1.60E-74
chr11	42040979	42045977	4.49	1.34E-69
chr7	46384248	46386527	4.22	5.49E-68
chr17	34795668	34796717	3.9	2.81E-66
chr24	27680892	27682154	4.96	3.98E-66
chr20	39958403	39959765	4.36	9.47E-66
chr24	21245538	21247808	3.83	6.35E-65
chr17	36029625	36033076	4.16	1.34E-64
chr2	15146628	15148046	3.74	2.25E-63
chr20	21458391	21459523	4.34	1.96E-62
chr7	14334164	14335400	4.26	5.91E-62
chr12	42546963	42549262	3.2	3.51E-61
chr13	13250003	13252083	4.41	5.55E-61
chr20	33340460	33341730	4.14	5.79E-60

chr3	55112842	55124356	-3.95	4.91E-59
chr18	38436537	38437692	4.03	6.57E-59
chr19	6258071	6259074	4.25	9.55E-59
chr5	17674494	17676439	3.81	1.20E-58
chr2	18280628	18281987	4.57	1.45E-58
chr23	17720683	17721729	4.45	1.87E-58
chr17	24349647	24351422	4.16	2.97E-58
chr14	42168986	42175882	3.28	1.68E-57
chr23	7838278	7839580	4.47	3.74E-56
chr22	10797318	10799487	4.15	1.24E-55
chr21	23608279	23609225	4.57	1.39E-54
chr12	31073252	31076121	3.36	1.41E-54
chr13	39899825	39901164	4.19	2.03E-54
chr23	27223027	27224378	3.96	7.06E-54
chr16	33494022	33496141	4.25	1.58E-53
chr2	4083994	4085365	4.67	2.08E-53
chr7	15980491	15981413	3.93	3.57E-53
chr18	20278642	20280286	3.91	5.77E-53
chr12	24828357	24829273	4.1	6.53E-53
chr5	17710233	17711700	3.16	9.23E-53
chr7	68753694	68754952	3.91	9.64E-53
chr22	16532742	16538407	-2.94	1.16E-52
chr20	30148352	30149528	4.92	3.12E-52
chr21	20712838	20717753	4.22	4.03E-52
chr18	40326265	40327465	4.93	6.97E-52
chr14	42156969	42158368	4.22	1.10E-51
chr18	15372667	15374901	4.54	1.34E-51
chr16	19442000	19447244	4.42	3.02E-51
chr10	38131831	38132928	4.61	7.33E-51
chr12	47663175	47666007	-2.91	1.52E-50
chr11	44714113	44715707	-3.49	1.54E-50
chr19	28217587	28219141	4.68	2.46E-50
chr14	49695733	49697204	3.77	2.71E-50
chr11	25822784	25825163	4.25	4.53E-50
chr17	26559447	26560491	4.61	8.45E-50

chr15	28714855	28717143	4.1	8.77E-50
chr18	47884590	47886407	3.39	1.17E-49
chr6	6561807	6563179	4.43	2.55E-49
chr12	45744983	45746197	4.28	3.50E-49
chr13	11259351	11260257	4.01	4.29E-49
chr2	35692402	35693204	4.37	4.48E-49
chr2	31270585	31271384	4.05	6.36E-49
chr20	43856152	43857224	4.24	7.77E-49
chr5	52105107	52106756	3.81	1.62E-48
chr14	39906595	39907979	4.62	1.71E-48
chr16	33248584	33249613	4.87	1.83E-48
chr17	18645921	18646765	4.44	3.17E-48
chr2	37322206	37323304	-3.18	5.30E-48
chr14	31109627	31111443	3.91	5.52E-48
chr24	38830913	38831933	3.81	5.79E-48
chr22	28787963	28792849	3.4	7.30E-48
chr6	45910459	45911408	4.78	9.67E-48
chr3	55124636	55128714	-4.11	1.01E-47
chr16	29239971	29240883	4.66	3.71E-47
chr13	28385449	28387099	4.06	5.26E-47
chr16	16837693	16838911	4.24	5.59E-47
chr19	37605617	37607115	3.71	7.37E-47
chr3	46886692	46888837	3.66	8.40E-47
chr3	43319144	43320721	4.4	1.14E-46
chr1	7763269	7764426	4.46	1.31E-46
chr3	55137158	55140344	-4.04	1.97E-46
chr16	449465	451279	4.32	4.42E-46
chr14	35098980	35100096	3.85	4.75E-46
chr1	47060306	47063745	-3.29	5.23E-46
chr24	8542991	8545047	4.68	7.92E-46
chr20	21207703	21208750	4.51	8.15E-46
chr15	21368255	21369829	4.24	8.53E-46
chr7	62058382	62060774	4.42	8.81E-46
chr11	21736951	21737897	3.99	9.33E-46
chr25	22346697	22347640	-4.84	9.48E-46

chr6	56301978	56302751	3.92	1.35E-45
chr24	2406996	2409017	-4.2	1.46E-45
chr7	33595724	33597222	4.13	1.80E-45
chr18	25616193	25617422	3.22	1.80E-45
chr7	35884324	35886258	4.36	1.94E-45
chr19	41565268	41566618	4.53	3.83E-45
chr6	14891618	14893424	4.11	4.86E-45
chr3	25027332	25028325	3.44	9.54E-45
chr10	31580933	31582850	4.09	1.09E-44
chr16	33415974	33418933	3.83	1.44E-44
chr13	30445591	30446915	4.03	1.73E-44
chr20	39014751	39015535	3.74	3.01E-44
chr5	19679687	19680786	3.92	7.18E-44
chr10	10843680	10844946	3.77	8.39E-44
chr7	36979114	36980412	4.53	8.69E-44
chr12	48674278	48675613	-3.78	9.43E-44
chr7	54834087	54835239	3.66	1.01E-43
chr11	16353722	16356492	3.86	1.01E-43
chr11	5426182	5427262	4.94	1.03E-43
chr15	33058796	33060245	4.55	1.57E-43
chr13	42393869	42396723	-3.58	2.22E-43
chr6	19815222	19818455	3.96	2.47E-43
chr6	44909443	44911071	3.9	2.53E-43
chr11	42056074	42058464	4.18	2.99E-43
chr3	53541034	53544858	3.85	3.28E-43
chr2	50365814	50366886	4.45	3.32E-43
chr6	33539504	33544686	2.5	4.63E-43
chr20	35823769	35825345	-3.27	5.82E-43
chr24	17194761	17195689	3.84	7.40E-43
chr16	29246457	29248787	3.41	7.40E-43
chr4	17478409	17480906	3.66	8.05E-43
chr12	25563688	25566098	3.35	8.67E-43
chr24	27633975	27636138	3.64	8.97E-43
chr21	2254894	2258844	-3.91	9.73E-43
chr5	67615814	67617196	3.63	1.14E-42

chr7	20991710	20992714	3.89	1.20E-42
chr2	29482722	29486411	-3.44	1.21E-42
chr2	40030466	40031575	3.69	1.21E-42
chr3	55837985	55840824	3.8	1.43E-42
chr19	40267728	40269288	3.88	2.10E-42
chr2	27516120	27518053	3	2.20E-42
chr25	8527946	8529022	3.92	2.39E-42
chr21	8508283	8510412	4.01	3.70E-42
chr3	14084817	14086253	4.26	4.40E-42
chr14	39546066	39548400	3.33	4.64E-42
chr18	49305888	49308225	4.07	4.83E-42
chr8	4712651	4713742	3.69	5.30E-42
chr5	19411505	19413296	3.65	6.04E-42
chr18	48199435	48200682	-4.26	7.08E-42
chr19	41553125	41554119	4.23	7.18E-42
chr13	46432263	46433810	4.24	7.66E-42
chr2	33105835	33110406	3.01	1.46E-41
chr8	10078400	10079386	4.24	1.47E-41
chr4	20831346	20833048	-4.46	1.50E-41
chr20	47570108	47571030	3.83	2.00E-41
chr21	37336596	37337938	3.6	2.12E-41
chr14	26637730	26640057	4.73	3.08E-41
chr5	65774356	65775499	4.55	3.50E-41
chr4	12295100	12295813	4.18	4.20E-41
chr13	30487158	30488485	2.9	4.32E-41
chr19	33708884	33710172	-3.7	4.85E-41
chr7	49778932	49780269	4.4	5.47E-41
chr5	17363640	17365256	4.63	6.93E-41
chr4	7750149	7750959	-3.71	7.58E-41
chr13	43250910	43251849	4.19	7.71E-41
chr16	33484342	33485641	4.01	7.79E-41
chr6	36697412	36698450	4.04	1.12E-40
chr19	1056511	1057601	4.18	1.34E-40
chr10	27171309	27172030	-3.88	1.61E-40
chr13	28844799	28846579	-3.73	1.76E-40

chr17	34841247	34842535	-4.47	2.36E-40
chr10	32135314	32136517	4.74	3.13E-40
chr10	38117134	38118590	4.63	3.77E-40
chr2	54463163	54464575	3.45	4.23E-40
chr5	66577077	66578539	4.3	4.42E-40
chr21	24587778	24590172	3.29	4.52E-40
chr7	53822466	53824024	3.96	5.69E-40
chr7	54805267	54806263	3.25	5.82E-40
chr1	40384011	40386303	-3.3	5.82E-40
chr10	31407231	31408133	4.34	7.14E-40
chr6	36513350	36514998	4.38	7.82E-40
chr24	10636497	10637722	4.23	8.60E-40
chr24	4584878	4586351	4.18	1.07E-39
chr14	32714323	32715899	3.54	1.23E-39
chr3	51036609	51040194	3.31	1.39E-39
chr11	14285822	14286443	-4.84	1.46E-39
chr17	37960736	37962045	3.72	1.54E-39
chr12	22291956	22293222	3.85	1.68E-39
chr16	42903610	42905333	3.83	2.34E-39
chr6	6156507	6156991	-4.26	2.61E-39
chr15	40668489	40669576	3.31	2.67E-39
chr18	30321934	30323028	3.55	2.82E-39
chr25	16665911	16668519	-4.35	2.84E-39
chr14	32725055	32725805	5.02	3.22E-39
chr13	46618743	46620090	3.99	3.56E-39
chr11	44985683	44987168	-3.41	3.81E-39
chr6	40582586	40583848	-2.84	4.05E-39
chr13	26214666	26215770	3.54	4.08E-39
chr23	21452855	21455761	4.24	4.20E-39
chr17	35996711	35998160	4.33	4.60E-39
chr22	16946380	16948948	2.83	5.31E-39
chr19	33730753	33732527	2.79	5.64E-39
chr7	28580662	28582527	3.93	5.92E-39
chr15	8452680	8453597	4.09	6.53E-39
chr7	60375947	60377579	-3.78	7.50E-39

chr23	22860175	22862867	2.86	7.69E-39
chr20	33431130	33431923	3.85	8.67E-39
chr5	48002158	48005836	3.27	8.73E-39

Table 7.16: List of top 200 DARs between *sox17M* vs *sox17E* ranked by FDR from Diffbind. Fold enrichment score > 0 represents *sox17M* DARs while *sox17E* DARs have a fold enrichment score < 0. FDR = False discovery rate

Chromosome	Chromosome start	Chromosome end	Fold enrichment	FDR
chr6	158029	162475	-3.13	2.66E-52
chr15	242882	245186	-3.14	3.42E-50
chr13	51995625	51999550	-2.78	2.43E-47
chr11	44697333	44698951	-2.7	1.54E-42
chr3	32094042	32104258	-2.21	3.65E-41
chr6	59191350	59194677	-3.45	2.43E-40
chr22	1244778	1245476	-2.7	4.86E-37
chr15	6825548	6826410	-3.22	1.06E-36
chr2	33324974	33328776	-2.71	1.43E-36
chr18	46150685	46157851	-2.23	4.44E-36
chr2	112120	125619	-2.64	6.36E-35
chr5	32056272	32058218	-4.33	1.42E-33
chr17	53417361	53420274	-3.17	1.42E-33
chr5	22274035	22276112	-3.59	3.90E-33
chr22	9918076	9928048	-3.43	4.18E-33
chr16	19442000	19447244	3.04	6.79E-33
chr17	36029625	36033076	2.94	7.22E-33
chr3	55124636	55128714	-3.83	9.50E-33
chr22	17442821	17443772	-3.67	9.50E-33
chr20	30148352	30149528	3.77	1.36E-32
chr3	55143618	55149276	-3.68	1.70E-32
chr18	17378407	17379728	-4.24	2.21E-32
chr7	38001756	38002738	-3.29	2.29E-32
chr16	45416287	45417923	-2.47	3.09E-32
chr12	48674278	48675613	-3.46	3.56E-32
chr12	31073252	31076121	2.88	1.50E-31
chr6	40582586	40583848	-2.69	2.40E-31
chr20	21458391	21459523	3.34	4.43E-31

chr11	290175	292913	-3.1	1.86E-30
chr7	58775752	58781510	-2.1	2.35E-30
chr25	11208940	11211694	-3.08	5.83E-30
chr3	55137158	55140344	-3.81	7.08E-30
chr21	247123	252602	-1.83	1.03E-29
chr21	39548345	39550298	-3.65	5.49E-29
chr14	166373	168652	-2.53	5.95E-29
chr2	53668544	53670119	-3.67	8.45E-29
chr24	21245538	21247808	2.77	9.48E-29
chr14	40824550	40825938	-3.95	1.19E-28
chr20	46674694	46675457	-3.82	1.22E-28
chr11	44964456	44966174	-3.48	2.05E-28
chr5	17710233	17711700	2.75	2.12E-28
chr1	47060306	47063745	-3.05	2.41E-28
chr18	30428976	30430779	-2.51	2.58E-28
chr5	24705689	24707350	-3.08	2.65E-28
chr13	13250003	13252083	3.34	2.80E-28
chr16	32517491	32518843	3.39	2.80E-28
chr16	31011604	31012996	-3.44	3.43E-28
chr1	51721748	51726015	-2.69	3.89E-28
chr11	44714113	44715707	-2.9	3.89E-28
chr19	33697239	33699024	-3.44	4.71E-28
chr16	27588785	27590628	-3.7	6.00E-28
chr3	15000769	15001909	-3.42	6.10E-28
chr17	38569651	38571459	-3.15	1.36E-27
chr2	49857196	49861202	-2.83	2.38E-27
chr7	60375947	60377579	-3.33	4.02E-27
chr2	37322206	37323304	-2.73	5.09E-27
chr18	28766	31437	-2.58	6.57E-27
chr5	63059721	63061606	-3.57	6.72E-27
chr3	9712847	9714651	-3.05	8.25E-27
chr13	4361755	4364081	-3.12	9.91E-27
chr18	30425648	30427535	-2.78	1.30E-26
chr2	29482722	29486411	-3.54	1.67E-26
chr11	43756863	43758431	-3.3	3.55E-26

chr10	27171309	27172030	-3.56	3.91E-26
chr3	22109288	22110534	-3.25	5.75E-26
chr21	2261596	2265045	-3.68	5.81E-26
chr5	27403123	27405512	-2.79	7.41E-26
chr20	23291035	23295006	-2.98	8.98E-26
chr2	33329285	33333517	-2.75	1.14E-25
chr19	46868076	46868927	-2.76	1.40E-25
chr3	59691	61149	-4.02	1.48E-25
chr20	53148223	53150122	-4.01	1.55E-25
chr13	42393869	42396723	-3.14	1.58E-25
chr8	23567461	23569265	-3.6	1.65E-25
chr25	7714994	7716565	-3.33	1.92E-25
chr5	30408166	30410732	-3.3	1.97E-25
chr22	16532742	16538407	-2.79	2.14E-25
chr1	46485601	46485854	-4.32	3.74E-25
chr12	42546963	42549262	2.89	9.11E-25
chr19	46043463	46045342	-3.85	9.45E-25
chr12	20349636	20351329	-3.72	9.86E-25
chr11	25422809	25424151	-2.31	9.90E-25
chr5	17674494	17676439	2.95	1.07E-24
chr21	3930838	3933547	-3.08	1.25E-24
chr12	45232576	45233356	-4.02	1.47E-24
chr21	20712838	20717753	2.77	1.50E-24
chr5	72006246	72009630	-3.11	1.51E-24
chr18	45864753	45865870	-3.42	1.75E-24
chr5	9560494	9561702	-3.65	1.95E-24
chr16	84808	88103	-2.92	2.15E-24
chr3	55130834	55133989	-3.69	2.15E-24
chr11	31563546	31565925	-2.76	2.36E-24
chr6	6156507	6156991	-4.23	2.88E-24
chr5	52105107	52106756	2.72	3.85E-24
chr18	40326265	40327465	3.83	3.85E-24
chr18	15372667	15374901	2.93	4.36E-24
chr11	3851700	3854010	-3.02	4.50E-24
chr3	17622270	17624203	-3.21	5.36E-24

chr25	6050621	6054029	-3.42	5.69E-24
chr21	12106590	12107699	-4.07	5.97E-24
chr21	2254894	2258844	-3.65	9.22E-24
chr23	17720683	17721729	3.45	1.10E-23
chr13	21647325	21650647	-3.06	1.16E-23
chr24	28418907	28420223	-2.93	1.28E-23
chr7	8314500	8316312	-3.34	1.50E-23
chr7	55740764	55742253	-3.25	1.56E-23
chr17	29205258	29207124	-3.6	1.56E-23
chr3	21049120	21050131	-3.9	1.62E-23
chr18	38436537	38437692	3	2.21E-23
chr18	20271556	20274223	2.46	2.67E-23
chr3	32600227	32603813	-3.19	2.95E-23
chr17	34795668	34796717	3	3.16E-23
chr21	75933	77632	-2.34	3.40E-23
chr19	25942835	25945426	-2.93	4.14E-23
chr3	18108243	18110666	3.08	5.14E-23
chr7	23776582	23777961	-3.36	5.36E-23
chr10	26535002	26536743	-4.28	5.38E-23
chr1	35541602	35543828	-3.2	5.56E-23
chr10	10843680	10844946	2.9	7.92E-23
chr17	41744757	41746516	-3.4	8.04E-23
chr17	24277015	24278039	-4.04	1.07E-22
chr15	30467766	30469057	-3.77	1.20E-22
chr16	36987382	36988729	-2.92	1.35E-22
chr18	3701752	3702403	-3.89	1.51E-22
chr18	50598636	50600730	-3.09	1.52E-22
chr8	26058388	26061638	-2.4	1.72E-22
chr3	17714780	17716720	-3.23	1.72E-22
chr2	4083994	4085365	3.36	1.76E-22
chr2	15146628	15148046	2.89	1.82E-22
chr8	516266	517158	-3.83	1.89E-22
chr2	37319149	37321209	-3.29	2.01E-22
chr6	8603462	8605035	-3.37	2.16E-22
chr21	42802316	42804549	-3.09	2.30E-22

chr10	31156392	31157027	3.7	2.81E-22
chr12	47663175	47666007	-2.76	2.84E-22
chr25	28118248	28119385	-3.23	2.90E-22
chr7	33595724	33597222	3.01	4.11E-22
chr4	18666008	18666619	3	4.33E-22
chr22	5686955	5688645	-2.94	5.55E-22
chr11	4995083	4995877	-3.92	7.39E-22
chr7	74152565	74154271	-2.96	7.89E-22
chr5	471813	473346	-2.86	7.99E-22
chr8	25338163	25341806	-2.99	8.10E-22
chr3	53683695	53685521	2.73	8.13E-22
chr7	32680348	32681127	-5.29	8.90E-22
chr18	39791375	39792656	-3.29	9.65E-22
chr23	11268421	11268763	-4.06	1.04E-21
chr13	28844799	28846579	-2.9	1.15E-21
chr14	26637730	26640057	3.35	1.21E-21
chr22	5197773	5199628	-2.77	1.22E-21
chr7	41025519	41027897	-3.28	1.35E-21
chr5	65364175	65364475	-3.53	1.38E-21
chr13	30445591	30446915	3.29	1.42E-21
chr5	9634013	9635756	-3.22	1.53E-21
chr17	24349647	24351422	2.78	1.54E-21
chr8	25609742	25611455	-3.42	2.00E-21
chr14	42156969	42158368	3.12	2.50E-21
chr19	35264955	35266628	-3.73	2.77E-21
chr1	40442551	40444250	-2.89	2.84E-21
chr23	21470841	21472197	4.47	3.41E-21
chr18	49300005	49301074	-3.43	3.69E-21
chr16	9618548	9619388	-4.28	4.15E-21
chr5	65774356	65775499	3.5	4.21E-21
chr14	35098980	35100096	3.1	5.45E-21
chr24	39857766	39859447	-2.62	7.04E-21
chr24	28281410	28282087	-3.66	7.04E-21
chr25	19802142	19802984	-3.53	7.48E-21
chr21	23608279	23609225	3.37	7.51E-21

chr16	38100903	38102091	3.12	8.15E-21
chr17	31599531	31600935	-3.17	8.31E-21
chr18	23984104	23985441	3.17	8.53E-21
chr22	12208328	12209430	-4.12	9.02E-21
chr14	31109627	31111443	3	9.59E-21
chr11	40798360	40799927	-3.4	9.59E-21
chr7	67729009	67735291	-1.84	9.70E-21
chr20	21185711	21186695	-3.49	9.70E-21
chr3	11595210	11595994	-3.48	1.10E-20
chr2	43468785	43471388	-2.38	1.15E-20
chr11	42029085	42031697	3.25	1.21E-20
chr6	40085078	40086723	-2.6	1.21E-20
chr13	36403568	36404676	-3.06	1.27E-20
chr3	25027332	25028325	3.01	1.35E-20
chr3	29888162	29891722	-2.65	1.37E-20
chr1	50269445	50270478	-4.5	1.54E-20
chr22	28787963	28792849	2.55	1.57E-20
chr12	49042331	49043056	-3.16	1.63E-20
chr20	30905540	30906537	-3.17	1.63E-20
chr11	25822784	25825163	2.77	1.69E-20
chr1	40384011	40386303	-2.55	1.77E-20
chr15	28714855	28717143	3.05	1.80E-20
chr3	53541034	53544858	2.7	1.87E-20
chr3	31860479	31862340	-3.44	2.28E-20
chr18	16794405	16797705	-3.16	2.37E-20
chr19	25410399	25413194	-3.53	2.55E-20
chr3	9800469	9801340	-3.42	2.62E-20
chr19	28217587	28219141	3.16	2.63E-20
chr2	24059891	24062252	-3.25	2.73E-20
chr5	63050287	63051558	-2.18	2.73E-20
chr5	7302202	7302908	-3.47	2.89E-20

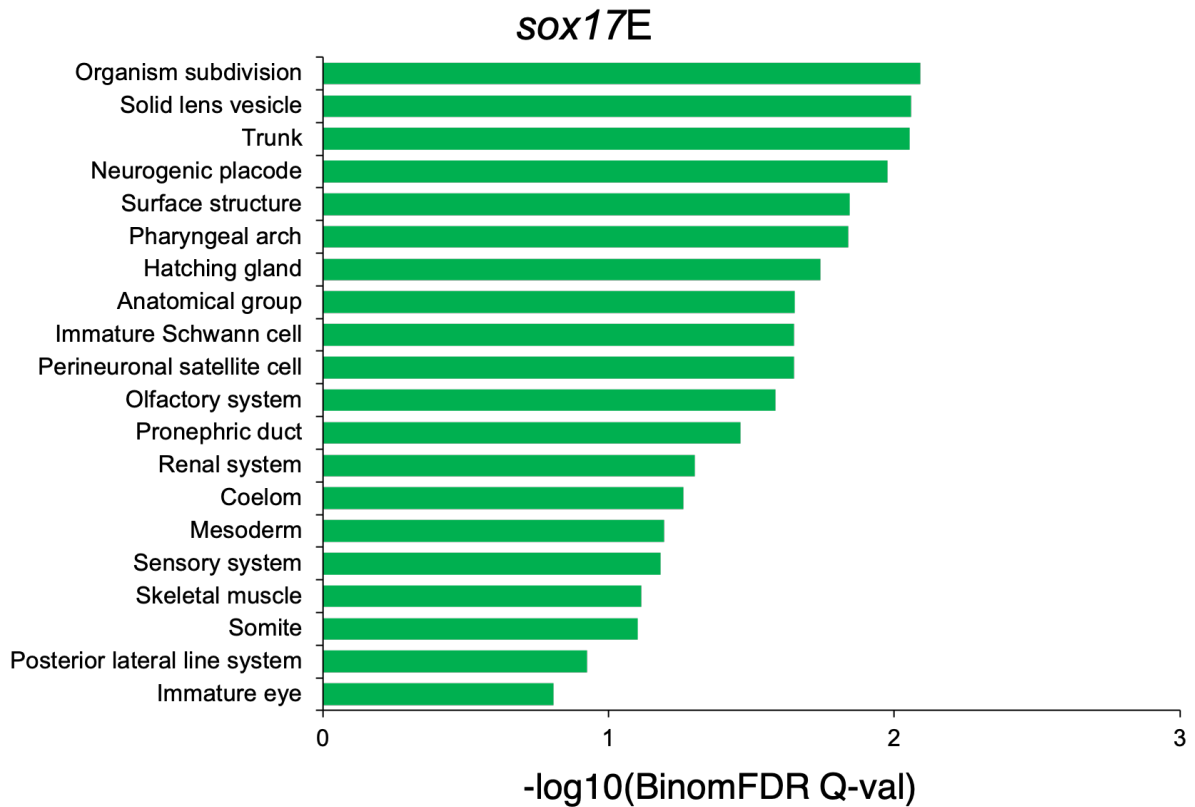


Figure 7-4: Bar graph displaying over-represented anatomical terms associated with *sox17E* at 28 h.p.f.. FDR = False discovery rate

Table 7.17: A list of 25 genes that are found proximal to *sox17E* vs *sox17N* DARs. These genes were attained from GREAT by first converting *sox17E* DARs to danRer7 genome build using LiftOver. TSS = transcriptional start site

Chromosome	Chromosome start	Chromosome end	Gene (distance to TSS)
chr1	3404019	3404271	rcan1b (+711337)
chr1	4604651	4606245	nrp2a (-39684)
chr1	5265215	5265587	klf7a (+92191)
chr1	6515759	6516762	mylz3 (-3998)
chr1	17436951	17437443	tbc1d1 (+76939)
chr1	17660047	17661409	limch1a (-47137)
chr1	17713711	17714719	limch1a (+6350)
chr1	18398208	18399513	tmem192 (+7421)
chr1	19374872	19375526	zgc:136652 (-126)
chr1	36524032	36524579	prmt10 (-44044)
chr1	38811798	38812264	cep44 (+40052)
chr1	46990569	46991299	mcf2la (-117)
chr1	59805804	59807256	zgc:173915 (-36801)

chr10	167449	169576	rippy3 (+5569)
chr10	264151	265759	psmg1 (+10853)
chr10	266720	267503	psmg1 (+8696)
chr10	2969580	2970627	oclna (-12)
chr10	11087404	11087975	ptbp3 (+81319)
chr10	15519724	15521412	erbb2ip (+37099)
chr10	22347823	22351261	cldn7b (+1087)
chr10	22729692	22730354	efnb3a (-6880)
chr10	28873988	28877968	rps6kb1a (-555)
chr10	29307477	29308260	cblb (+37687)
chr10	29437292	29437746	alcama (+5807)
chr10	32314742	32315566	robo4 (+58224)

Table 7.18: A list of 25 genes that are found proximal to *sox17M* vs *sox17N* DARs. These genes were attained from GREAT by first converting *sox17M* DARs to danRer7 genome build using LiftOver. TSS = transcriptional start site

Chromosome	Chromosome start	Chromosome end	Gene (distance to TSS)
chr1	3149	4057	zgc:163025 (-197)
chr1	66310	68441	grtp1a (-889)
chr1	133888	135831	cenpe (+130)
chr1	488306	489561	atp5j (+346)
chr1	875770	876993	creg1 (+100131)
chr1	1165651	1166225	runx1 (+102561)
chr1	2016765	2019372	rap2ab (+38920)
chr1	2125690	2126761	farp1 (+48056)
chr1	4290898	4292769	tuba8l2 (+204)
chr1	4299439	4300258	tuba8l2 (+8219)
chr1	4583841	4585715	nrp2a (-19014)
chr1	4849477	4850234	abcb6a (-148980)
chr1	4997916	4999263	abcb6a (-246)
chr1	5003359	5003950	abcb6a (+4819)
chr1	5005252	5007530	abcb6a (+7555)
chr1	5024637	5026693	cnppd1 (+75)
chr1	5474187	5475390	si:ch211-93g23.2 (-28824)
chr1	7597170	7598368	smcr7b (-2532)

















chr1	7628246	7629714	zgc:112980 (+399)
chr1	7722213	7732222	actb1 (+1598)
chr1	7769456	7772356	zgc:77849 (+626)
chr1	8053692	8054590	tlr2 (-6011)
chr1	8186839	8188300	si:ch211-14k19.8 (+3333)
chr1	8190090	8193022	vap (-438)
chr1	8261947	8262531	em2 (+5916)

Table 7.19: A list of 25 genes that are found proximal to *sox17N* vs *sox17E* DARs. These genes were attained from GREAT by first converting *sox17N* DARs to danRer7 genome build using LiftOver. TSS = transcriptional start site

Chromosome	Chromosome start	Chromosome end	Gene (distance to TSS)
chr1	2797488	2798741	rca1b (+105307)
chr1	3293692	3294865	rca1b (+601471)
chr1	3288889	3289635	rca1b (+596454)
chr1	4104976	4106331	spry2 (-65265)
chr1	4165707	4166788	spry2 (-4671)
chr1	4796843	4797472	abcb6a (-201678)
chr1	4807131	4807705	abcb6a (-191418)
chr1	4872609	4873400	abcb6a (-125831)
chr1	4927815	4928696	abcb6a (-70580)
chr1	5323386	5325068	klf7a (+33365)
chr1	6403708	6405094	efnb2b (+5749)
chr1	6669344	6670585	eng1b (+84516)
chr1	10676661	10678420	si:dkey-26i13.7 (+100411)
chr1	11597717	11598963	gne (+6711)
chr2	25104402	25105957	si:dkey-5n7.2 (-125158)
chr2	25118572	25119683	si:dkey-5n7.2 (-111210)
chr1	12157322	12159753	pcdh18a (-2666)
chr1	12180090	12181738	pcdh18a (-25042)
chr1	15312771	15314206	pcm1 (-54084)
chr1	16376848	16377769	si:ch211-197n1.2 (+39405)
chr1	19053330	19054673	ndst3 (+529)
chr1	19059889	19061392	ndst3 (-6110)
chr1	19948186	19948854	si:dkey-253i9.4 (+12540)

chr1	20439150	20439765	adamtsl7 (-57587)
chr1	20532610	20533666	zcchc7 (+15854)

Table 7.20: TF binding motifs found over-represented in *sox17E* vs *sox17N* DARs. Motifs are ranked by statistical significance.

Rank	Motif	Name	P-value	log P-value	q-value (Benjamini)	# Target Sequences with Motif	% Target Sequences with Motif
1		Fosl2(bZIP)/3T3L1-Fosl2-ChIP-Seq(GSE56872)/Homer	1e-19	-4.398e+01	0.0000	47.0	12.0
2		Fra2(bZIP)/Striatum-Fra2-ChIP-Seq(GSE43429)/Homer	1e-18	-4.306e+01	0.0000	60.0	15.0
3		Jun-AP1(bZIP)/K562-cJun-ChIP-Seq(GSE31477)/Homer	1e-17	-4.035e+01	0.0000	37.0	9.0
4		JunB(bZIP)/DendriticCells-Junb-ChIP-Seq(GSE36099)/Homer	1e-16	-3.721e+01	0.0000	62.0	16.0
5		TEAD1(TEAD)/HepG2-TEAD1-ChIP-Seq(Encode)/Homer	1e-14	-3.451e+01	0.0000	63.0	16.0
6		Fra1(bZIP)/BT549-Fra1-ChIP-Seq(GSE46166)/Homer	1e-14	-3.384e+01	0.0000	61.0	16.0
7		Fos(bZIP)/TSC-Fos-ChIP-Seq(GSE110950)/Homer	1e-14	-3.269e+01	0.0000	64.0	16.0
8		BATF(bZIP)/Th17-BATF-ChIP-Seq(GSE39756)/Homer	1e-14	-3.231e+01	0.0000	65.0	17.0
9		Atf3(bZIP)/GBM-ATF3-ChIP-Seq(GSE33912)/Homer	1e-13	-3.192e+01	0.0000	67.0	17.0
10		AP-1(bZIP)/ThioMac-PU.1-ChIP-Seq(GSE21512)/Homer	1e-13	-3.154e+01	0.0000	71.0	18.0
11		TEAD4(TEA)/Tropoblast-Tead4-ChIP-Seq(GSE37350)/Homer	1e-12	-2.920e+01	0.0000	54.0	14.0
12		TEAD3(TEA)/HepG2-TEAD3-ChIP-Seq(Encode)/Homer	1e-12	-2.891e+01	0.0000	68.0	17.0
13		Six4(Homeobox)/MCF7-SIX4-ChIP-Seq(Encode)/Homer	1e-10	-2.394e+01	0.0000	16.0	4.0
14		Six1(Homeobox)/Myoblast-Six1-ChIP-Chip(GSE20150)/Homer	1e-9	-2.224e+01	0.0000	33.0	8.0
15		Six2(Homeobox)/NephronProgenitor-Six2-ChIP-Seq(GSE39837)/Homer	1e-9	-2.108e+01	0.0000	83.0	21.0
16		TEAD2(TEA)/Py2T-Tead2-ChIP-Seq(GSE55709)/Homer	1e-9	-2.102e+01	0.0000	34.0	8.0
17		TEAD(TEA)/Fibroblast-PU.1-ChIP-Seq(Unpublished)/Homer	1e-8	-2.057e+01	0.0000	40.0	10.0
18		Bach2(bZIP)/OCILy7-Bach2-ChIP-Seq(GSE44420)/Homer	1e-7	-1.785e+01	0.0000	22.0	5.0
19		p63(p53)/Keratinocyte-p63-ChIP-Seq(GSE17611)/Homer	1e-6	-1.522e+01	0.0000	29.0	7.0
20		p53(p53)/Saos-p53-ChIP-Seq(GSE15780)/Homer	1e-6	-1.404e+01	0.0000	14.0	3.0
21		p53(p53)/Saos-p53-ChIP-Seq/Homer	1e-6	-1.404e+01	0.0000	14.0	3.0
22		FOXO1(Forkhead)/MCF7-FOXO1-ChIP-Seq(GSE72977)/Homer	1e-5	-1.381e+01	0.0000	89.0	23.0

23		FOXA1(Forkhead)/LNCAP-FOXA1-ChIP-Seq(GSE27824)/Homer	1e-5	-1.303e+01	0.0000	106.0	27	
24		Fox:Ebox(Forkhead,bHLH)/Panc1-Foxa2-ChIP-Seq(GSE47459)/Homer	1e-5	-1.289e+01	0.0000	92.0	24	
25		FOXA1(Forkhead)/MCF7-FOXA1-ChIP-Seq(GSE26831)/Homer	1e-5	-1.252e+01	0.0001	89.0	23	
26		Snail1(Zf)/LS174T-SNAI1.HA-ChIP-Seq(GSE127183)/Homer	1e-5	-1.197e+01	0.0001	56.0	14	
27		ZEB1(Zf)/PDAC-ZEB1-ChIP-Seq(GSE64557)/Homer	1e-4	-1.138e+01	0.0002	100.0	26	
28		HOXB13(Homeobox)/ProstateTumor-HOXB13-ChIP-Seq(GSE56288)/Homer	1e-4	-1.087e+01	0.0003	82.0	21	
29		AP-2alpha(AP2)/Hela-AP2alpha-ChIP-Seq(GSE31477)/Homer	1e-4	-1.082e+01	0.0003	25.0	6.5	
30		Myf5(bHLH)/GM-Myf5-ChIP-Seq(GSE24852)/Homer	1e-4	-1.080e+01	0.0003	49.0	12	
31		AP-2gamma(AP2)/MCF7-TFAP2C-ChIP-Seq(GSE21234)/Homer	1e-4	-1.065e+01	0.0003	31.0	8.3	
32		ZEB2(Zf)/SNU398-ZEB2-ChIP-Seq(GSE103048)/Homer	1e-4	-1.059e+01	0.0003	66.0	17	
33		Gata1(Zf)/K562-GATA1-ChIP-Seq(GSE18829)/Homer	1e-4	-9.873e+00	0.0007	39.0	10	
34		E2A(bHLH),near_PU.1/Bcell-PU.1-ChIP-Seq(GSE21512)/Homer	1e-4	-9.468e+00	0.0010	80.0	21	
35		FoxD3(forkhead)/ZebrafishEmbryo-Foxd3.biotin-ChIP-seq(GSE106676)/Homer	1e-4	-9.329e+00	0.0011	75.0	19	
36		E2A(bHLH)/proBcell-E2A-ChIP-Seq(GSE21978)/Homer	1e-3	-9.119e+00	0.0013	76.0	20	
37		CDX4(Homeobox)/ZebrafishEmbryos-Cdx4.Myc-ChIP-Seq(GSE48254)/Homer	1e-3	-9.082e+00	0.0014	71.0	18	
38		Smad4(MAD)/ESC-SMAD4-ChIP-Seq(GSE29422)/Homer	1e-3	-8.937e+00	0.0015	101.0	26	
39		Foxo3(Forkhead)/U2OS-Foxo3-ChIP-Seq(E-MTAB-2701)/Homer	1e-3	-8.789e+00	0.0017	77.0	20	
40		Gata4(Zf)/Heart-Gata4-ChIP-Seq(GSE35151)/Homer	1e-3	-8.593e+00	0.0020	61.0	16	
41		MyoG(bHLH)/C2C12-MyoG-ChIP-Seq(GSE36024)/Homer	1e-3	-8.305e+00	0.0027	60.0	15	
42		Foxa3(Forkhead)/Liver-Foxa3-ChIP-Seq(GSE77670)/Homer	1e-3	-7.852e+00	0.0041	36.0	9.4	
43		Foxa2(Forkhead)/Liver-Foxa2-ChIP-Seq(GSE25694)/Homer	1e-3	-7.733e+00	0.0045	70.0	18	
44		FoxL2(Forkhead)/Ovary-FoxL2-ChIP-Seq(GSE60858)/Homer	1e-3	-7.693e+00	0.0046	77.0	20	
45		GATA3(Zf)/iTreg-Gata3-ChIP-Seq(GSE20898)/Homer	1e-3	-7.617e+00	0.0048	80.0	21	
46		Slug(Zf)/Mesoderm-Snai2-ChIP-Seq(GSE61475)/Homer	1e-3	-7.597e+00	0.0048	33.0	8.3	
47		Cdx2(Homeobox)/mES-Cdx2-ChIP-Seq(GSE14586)/Homer	1e-3	-7.444e+00	0.0055	53.0	13	

48		RXR(NR),DR1/3T3L1-RXR-ChIP-Seq(GSE13511)/Homer	1e-3	-7.225e+00	0.0067	43.0	11
49		COUP-TFII(NR)/Artia-Nr2f2-ChIP-Seq(GSE46497)/Homer	1e-3	-7.224e+00	0.0067	82.0	21
50		KLF10(Zf)/HEK293-KLF10.GFP-ChIP-Seq(GSE58341)/Homer	1e-3	-7.201e+00	0.0067	79.0	20
51		EAR2(NR)/K562-NR2F6-ChIP-Seq(Encode)/Homer	1e-3	-7.166e+00	0.0067	65.0	17
52		Hoxd11(Homeobox)/ChickenMSG-Hoxd11.Flag-ChIP-Seq(GSE86088)/Homer	1e-3	-6.986e+00	0.0078	143.0	37
53		NFIL3(bZIP)/HepG2-NFIL3-ChIP-Seq(Encode)/Homer	1e-3	-6.943e+00	0.0080	51.0	13
54		Ptf1a(bHLH)/Panc1-Ptf1a-ChIP-Seq(GSE47459)/Homer	1e-2	-6.897e+00	0.0082	131.0	34
55		p73(p53)/Trachea-p73-ChIP-Seq(PRJNA310161)/Homer	1e-2	-6.782e+00	0.0091	6.0	1.5
56		HEB(bHLH)/mES-Heb-ChIP-Seq(GSE53233)/Homer	1e-2	-6.766e+00	0.0091	88.0	23
57		FOXK1(Forkhead)/HEK293-FOXK1-ChIP-Seq(GSE51673)/Homer	1e-2	-6.526e+00	0.0113	89.0	23
58		Hoxd13(Homeobox)/ChickenMSG-Hoxd13.Flag-ChIP-Seq(GSE86088)/Homer	1e-2	-6.311e+00	0.0138	100.0	26
59		Ascl1(bHLH)/NeuralTubes-Ascl1-ChIP-Seq(GSE55840)/Homer	1e-2	-6.296e+00	0.0138	67.0	17
60		Gata6(Zf)/HUG1N-GATA6-ChIP-Seq(GSE51936)/Homer	1e-2	-6.078e+00	0.0168	51.0	13
61		Gata2(Zf)/K562-GATA2-ChIP-Seq(GSE18829)/Homer	1e-2	-5.874e+00	0.0203	37.0	9.7
62		TRPS1(Zf)/MCF7-TRPS1-ChIP-Seq(GSE107013)/Homer	1e-2	-5.874e+00	0.0203	100.0	26
63		Hoxa11(Homeobox)/ChickenMSG-Hoxa11.Flag-ChIP-Seq(GSE86088)/Homer	1e-2	-5.793e+00	0.0213	140.0	36
64		KLF1(Zf)/HUDEP2-KLF1-CutnRun(GSE136251)/Homer	1e-2	-5.722e+00	0.0225	29.0	7.6
65		KLF14(Zf)/HEK293-KLF14.GFP-ChIP-Seq(GSE58341)/Homer	1e-2	-5.408e+00	0.0303	47.0	12
66		Hoxc9(Homeobox)/Ainv15-Hoxc9-ChIP-Seq(GSE21812)/Homer	1e-2	-5.393e+00	0.0303	37.0	9.7
67		EKLF(Zf)/Erythrocyte-Klf1-ChIP-Seq(GSE20478)/Homer	1e-2	-5.379e+00	0.0303	10.0	2.6
68		Pdx1(Homeobox)/Islet-Pdx1-ChIP-Seq(SRA008281)/Homer	1e-2	-5.350e+00	0.0307	69.0	18
69		SCRT1(Zf)/HEK293-SCRT1.eGFP-ChIP-Seq(Encode)/Homer	1e-2	-4.991e+00	0.0434	22.0	5.7
70		PRDM15(Zf)/ESC-Prdm15-ChIP-Seq(GSE73694)/Homer	1e-2	-4.985e+00	0.0434	40.0	10
71		MafA(bZIP)/Islet-MafA-ChIP-Seq(GSE30298)/Homer	1e-2	-4.958e+00	0.0436	43.0	11
72		FOXP1(Forkhead)/H9-FOXP1-ChIP-Seq(GSE31006)/Homer	1e-2	-4.904e+00	0.0453	44.0	11
73		Sp2(Zf)/HEK293-Sp2.eGFP-ChIP-Seq(Encode)/Homer	1e-2	-4.862e+00	0.0466	40.0	10

Table 7.21: A list of upregulated transcripts in *sox17E* vs *sox17N* sample populations. Genes are ordered based on adjusted Pvalues.

Gene_ID	Log2FoldChange	Stat	Pvalue	Padj	Symbol
ENSDARG00000012944	3.397588873	8.3168938	9.03E-17	1.49E-12	<i>myhz2</i>
ENSDARG000000071095	3.927160611	7.31833506	2.51E-13	2.07E-09	<i>abi3bpb</i>
ENSDARG000000037889	4.321446583	7.23416311	4.68E-13	2.57E-09	<i>wnt9b</i>
ENSDARG000000039133	3.429199021	7.17202106	7.39E-13	3.04E-09	<i>lamb4</i>
ENSDARG000000004539	3.606312941	7.02730328	2.11E-12	6.93E-09	<i>ptgs2a</i>
ENSDARG000000077084	5.968189161	6.94326884	3.83E-12	1.05E-08	<i>col28a1a</i>
ENSDARG000000075521	3.681751524	6.90861586	4.89E-12	1.15E-08	<i>cpz</i>
ENSDARG000000021720	3.256177569	6.78790929	1.14E-11	2.34E-08	<i>col7a1</i>
ENSDARG000000102332	3.167175595	6.67484688	2.47E-11	4.53E-08	<i>spint1a</i>
ENSDARG000000099371	2.644743677	6.64012279	3.13E-11	5.16E-08	<i>cygb1</i>
ENSDARG000000058381	3.195807987	6.60546853	3.96E-11	5.93E-08	<i>zgc:171775</i>
ENSDARG000000014047	3.06797251	6.530257	6.57E-11	9.00E-08	<i>cldn7b</i>
ENSDARG000000018923	2.939208582	6.50025127	8.02E-11	9.89E-08	<i>fat2</i>
ENSDARG000000079372	3.06706453	6.49295026	8.42E-11	9.89E-08	<i>si:ch211-264f5.6</i>
ENSDARG000000061173	2.791620301	6.33617646	2.36E-10	2.42E-07	<i>st14a</i>
ENSDARG000000076856	3.951469977	6.1970405	5.75E-10	5.57E-07	<i>frem2a</i>
ENSDARG000000054616	2.991159905	6.09434695	1.10E-09	9.72E-07	<i>cldni</i>

ENSDARG00000113 572	3.179908923	6.091023 43	1.12E-09	9.72E-07	<i>si:dkey-102c8.3</i>
ENSDARG00000027 867	3.977879859	6.065975 81	1.31E-09	1.08E-06	<i>paplna</i>
ENSDARG00000052 652	3.773878925	6.057336 61	1.38E-09	1.08E-06	<i>fermt1</i>
ENSDARG00000075 441	2.833695744	6.003622 56	1.93E-09	1.44E-06	<i>vwa2</i>
ENSDARG00000069 946	2.786795051	5.929461 87	3.04E-09	2.17E-06	<i>itga6b</i>
ENSDARG00000009 133	2.821991559	5.833690 87	5.42E-09	3.60E-06	<i>myo1eb</i>
ENSDARG00000012 824	2.665926992	5.826291 12	5.67E-09	3.60E-06	<i>itga3b</i>
ENSDARG00000058 543	2.39243814	5.825803 48	5.68E-09	3.60E-06	<i>lama5</i>
ENSDARG00000102 750	2.399407141	5.765161 89	8.16E-09	4.79E-06	<i>cdh1</i>
ENSDARG00000088 717	3.650917883	5.745163 64	9.18E-09	5.21E-06	<i>ecrg4b</i>
ENSDARG00000004 658	3.386496412	5.718244 36	1.08E-08	5.90E-06	<i>zgc:101810</i>
ENSDARG00000063 631	3.227702918	5.709856 28	1.13E-08	5.96E-06	<i>si:ch1073- 291c23.1</i>
ENSDARG00000040 534	2.797513105	5.701439 08	1.19E-08	5.96E-06	<i>epcam</i>
ENSDARG00000044 356	2.639740678	5.700312 01	1.20E-08	5.96E-06	<i>tp63</i>
ENSDARG00000057 590	2.933366895	5.669284 2	1.43E-08	6.94E-06	<i>si:ch1073- 396h14.1</i>
ENSDARG00000094 324	3.111970356	5.644643 67	1.66E-08	7.78E-06	<i>efemp2a</i>
ENSDARG00000044 001	2.997339522	5.633971 15	1.76E-08	7.91E-06	<i>lgals3b</i>
ENSDARG00000054 619	2.323483187	5.632273 39	1.78E-08	7.91E-06	<i>fras1</i>
ENSDARG00000053 232	2.970065873	5.613089 89	1.99E-08	8.43E-06	<i>itgb1b.1</i>

ENSDARG00000090 268	3.432051999	5.612120 46	2.00E-08	8.43E-06	<i>krtt1c19e</i>
ENSDARG00000069 692	2.40178775	5.575663 37	2.47E-08	1.01E-05	<i>col7a1l</i>
ENSDARG00000056 938	2.974134751	5.567958 8	2.58E-08	1.03E-05	<i>ker</i>
ENSDARG00000038 151	3.329848692	5.557389 82	2.74E-08	1.07E-05	<i>zgc:92360</i>
ENSDARG00000056 767	3.082615611	5.536121 94	3.09E-08	1.18E-05	<i>itgb3a</i>
ENSDARG00000100 223	2.822384191	5.512734 64	3.53E-08	1.32E-05	<i>gpa33a</i>
ENSDARG00000002 847	3.544958503	5.507245 16	3.64E-08	1.33E-05	<i>fndc1</i>
ENSDARG00000063 295	2.848474774	5.472841 11	4.43E-08	1.58E-05	<i>myh9a</i>
ENSDARG00000014 190	8.956705956	5.438589 59	5.37E-08	1.88E-05	<i>sst2</i>
ENSDARG00000077 410	2.89770368	5.428193 6	5.69E-08	1.93E-05	<i>myo9b</i>
ENSDARG00000062 974	2.989112191	5.426594 85	5.74E-08	1.93E-05	<i>itga2.2</i>
ENSDARG00000006 526	2.382753053	5.422365 24	5.88E-08	1.94E-05	<i>fn1b</i>
ENSDARG00000095 930	3.007209191	5.398292 13	6.73E-08	2.17E-05	<i>myha</i>
ENSDARG00000062 579	2.486890891	5.378156 5	7.53E-08	2.38E-05	<i>kremen1</i>
ENSDARG00000068 409	3.137813464	5.326687 07	1.00E-07	3.05E-05	<i>vgl14l</i>
ENSDARG00000012 467	2.970847552	5.313797 31	1.07E-07	3.19E-05	<i>spint1b</i>
ENSDARG00000101 816	2.949007574	5.311662 67	1.09E-07	3.19E-05	<i>col5a3b</i>
ENSDARG00000045 517	5.151967981	5.273235 72	1.34E-07	3.87E-05	<i>itih5</i>
ENSDARG00000057 113	2.430675356	5.255010 5	1.48E-07	4.20E-05	<i>c6</i>

ENSDARG00000101 986	3.117266542	5.248570 4	1.53E-07	4.28E-05	<i>irf6</i>
ENSDARG00000017 320	2.394227922	5.244908 48	1.56E-07	4.29E-05	<i>f11r.1</i>
ENSDARG00000090 369	2.988902351	5.231787 5	1.68E-07	4.53E-05	<i>zgc:86896</i>
ENSDARG00000062 750	2.959349666	5.206734 13	1.92E-07	5.10E-05	<i>si:ch73-74h11.1</i>
ENSDARG00000102 626	2.201135298	5.201081 33	1.98E-07	5.12E-05	<i>frem2b</i>
ENSDARG00000006 112	3.018084245	5.199973 93	1.99E-07	5.12E-05	<i>myof</i>
ENSDARG00000051 975	2.596862954	5.184684 59	2.16E-07	5.48E-05	<i>cd99</i>
ENSDARG00000032 623	2.813674583	5.175867 27	2.27E-07	5.57E-05	<i>rca3</i>
ENSDARG00000019 063	3.433259661	5.157701 05	2.50E-07	5.97E-05	<i>fat1b</i>
ENSDARG00000031 164	2.383655246	5.155100 93	2.53E-07	5.97E-05	<i>tuba8l2</i>
ENSDARG00000020 811	3.43534698	5.152910 41	2.56E-07	5.97E-05	<i>fibpb</i>
ENSDARG00000020 811	3.43534698	5.152910 41	2.56E-07	5.97E-05	<i>efemp2b</i>
ENSDARG00000044 167	2.959224135	5.151993 58	2.58E-07	5.97E-05	<i>padi2</i>
ENSDARG00000016 188	2.356298082	5.126680 03	2.95E-07	6.74E-05	<i>si:ch73-63e15.2</i>
ENSDARG00000042 021	3.157274633	5.122813 8	3.01E-07	6.79E-05	<i>mapk12a</i>
ENSDARG00000012 972	2.950517811	5.119822 57	3.06E-07	6.80E-05	<i>cfl1</i>
ENSDARG00000001 968	2.681985874	5.076222 52	3.85E-07	8.17E-05	<i>dock5</i>
ENSDARG00000098 239	3.327236068	5.076101 97	3.85E-07	8.17E-05	<i>zgc:85932</i>
ENSDARG00000037 517	2.397338309	5.073660 67	3.90E-07	8.17E-05	<i>slc35c2</i>

ENSDARG00000037 517	2.397338309	5.073660 67	3.90E-07	8.17E-05	<i>LOC101884357</i>
ENSDARG00000045 262	2.557211389	5.072634 85	3.92E-07	8.17E-05	<i>gsnb</i>
ENSDARG00000092 567	3.333308202	4.992793 12	5.95E-07	0.000120 91	<i>si:dkey-6111.4</i>
ENSDARG00000074 094	2.20915566	4.967722 34	6.77E-07	0.000134 31	<i>tgm2b</i>
ENSDARG00000057 992	2.618387068	4.952164 33	7.34E-07	0.000143 78	<i>fstb</i>
ENSDARG00000043 010	2.575923853	4.938334 16	7.88E-07	0.000152 54	<i>camk2d1</i>
ENSDARG00000040 295	2.954165065	4.927276 3	8.34E-07	0.000159 55	<i>apoeb</i>
ENSDARG00000003 808	2.482080314	4.914262 72	8.91E-07	0.000168 57	<i>aqp3a</i>
ENSDARG00000018 721	2.571272998	4.911109 29	9.06E-07	0.000169 35	<i>npnta</i>
ENSDARG00000002 494	3.0280468	4.892382 6	9.96E-07	0.000184 2	<i>itgb6</i>
ENSDARG00000013 310	2.672978611	4.863808 63	1.15E-06	0.000208 23	<i>map3k15</i>
ENSDARG00000045 887	4.741261007	4.852007 25	1.22E-06	0.000216 78	<i>mmp30</i>
ENSDARG00000017 128	3.545903571	4.849411 73	1.24E-06	0.000216 78	<i>myofl</i>
ENSDARG00000005 108	2.838747384	4.825520 1	1.40E-06	0.000241 88	<i>oclna</i>
ENSDARG00000099 424	5.8940327	4.749586 4	2.04E-06	0.000338 82	<i>lcat</i>
ENSDARG00000002 909	2.4984211	4.743910 4	2.10E-06	0.000344 97	<i>tjp3</i>
ENSDARG00000042 172	2.280805606	4.733673 97	2.20E-06	0.000355 73	<i>c7a</i>
ENSDARG00000035 350	5.848632186	4.726463 33	2.28E-06	0.000365 01	<i>ins</i>
ENSDARG00000011 672	2.382247522	4.707330 77	2.51E-06	0.000396 27	<i>sema3b</i>

ENSDARG00000036 840	3.719990145	4.678330 01	2.89E-06	0.000444 81	<i>krt15</i>
ENSDARG00000052 556	2.484045299	4.673729 23	2.96E-06	0.000450 68	<i>fgfr11b</i>
ENSDARG00000092 881	2.500587034	4.669487 31	3.02E-06	0.000455 87	<i>rbm20</i>
ENSDARG00000015 955	4.8268637	4.651605 47	3.29E-06	0.000486 67	<i>cldnc</i>
ENSDARG00000012 013	2.555360193	4.650439 16	3.31E-06	0.000486 67	<i>cpa6</i>
ENSDARG00000040 237	1.795728621	4.638426 99	3.51E-06	0.000506 9	<i>mef2d</i>
ENSDARG00000104 181	2.54224518	4.638375 17	3.51E-06	0.000506 9	<i>tmsb1</i>
ENSDARG00000101 980	2.262364634	4.626702 01	3.72E-06	0.000527 07	<i>bmper</i>
ENSDARG00000045 129	2.169171703	4.603259 05	4.16E-06	0.000573 79	<i>fkbp10b</i>
ENSDARG00000039 577	2.918044709	4.602017 36	4.18E-06	0.000573 79	<i>ptk2bb</i>
ENSDARG00000068 191	2.354636479	4.598360 65	4.26E-06	0.000577 07	<i>cdh15</i>
ENSDARG00000032 932	2.566930006	4.597386 69	4.28E-06	0.000577 07	<i>cnksr1</i>
ENSDARG00000105 630	2.762686745	4.590533 66	4.42E-06	0.000591 5	<i>alpk3b</i>
ENSDARG00000021 242	2.341903667	4.581260 97	4.62E-06	0.000608 45	<i>mvp</i>
ENSDARG00000071 215	2.771000795	4.579592 46	4.66E-06	0.000608 46	<i>palm3</i>
ENSDARG00000116 617	2.295097651	4.561066 98	5.09E-06	0.000650 45	<i>spaca4l</i>
ENSDARG00000052 470	2.925309219	4.553799 33	5.27E-06	0.000659 68	<i>igfbp2a</i>
ENSDARG00000030 722	2.789374935	4.534582 43	5.77E-06	0.000708 81	<i>xirp1</i>
ENSDARG00000069 888	3.678836517	4.530259 5	5.89E-06	0.000714 68	<i>cldna</i>

ENSDARG00000018 757	2.940671781	4.529712 16	5.91E-06	0.000714 68	<i>klf5l</i>
ENSDARG00000035 018	2.993693225	4.521349 41	6.14E-06	0.000735 21	<i>thy1</i>
ENSDARG00000040 362	2.652598053	4.520634 9	6.17E-06	0.000735 21	<i>ehd2b</i>
ENSDARG00000055 784	2.636324037	4.518635 33	6.22E-06	0.000736 84	<i>ptpn3</i>
ENSDARG00000103 590	2.704411553	4.508801 6	6.52E-06	0.000766 32	<i>cyp2aa6</i>
ENSDARG00000061 970	3.222092606	4.505533 86	6.62E-06	0.000772 69	<i>lingo3a</i>
ENSDARG00000097 153	1.995746623	4.497222 54	6.88E-06	0.000797 85	<i>LOC108182821</i>
ENSDARG00000039 099	2.498626747	4.475452 64	7.62E-06	0.000861 92	<i>aep1</i>
ENSDARG00000075 555	2.896545591	4.474071 12	7.67E-06	0.000861 92	<i>adamts17</i>
ENSDARG00000018 820	2.404119469	4.466015 68	7.97E-06	0.000886 06	<i>flncb</i>
ENSDARG00000097 118	5.043490002	4.453985 8	8.43E-06	0.000913 49	<i>si:dkey-33c14.3</i>
ENSDARG00000074 677	2.456286319	4.453766 99	8.44E-06	0.000913 49	<i>frem3</i>
ENSDARG00000076 945	2.21030602	4.448050 38	8.67E-06	0.000932	<i>dsg2.1</i>
ENSDARG00000025 218	2.30681359	4.424847 39	9.65E-06	0.001018 06	<i>myo5ab</i>
ENSDARG00000005 476	1.799350779	4.423021 04	9.73E-06	0.001020 17	<i>nav3</i>
ENSDARG00000006 615	2.763621326	4.418233 51	9.95E-06	0.001029 91	<i>hnf1ba</i>
ENSDARG00000019 564	2.440065949	4.407316 07	1.05E-05	0.001063 13	<i>asap2b</i>
ENSDARG00000078 362	3.173392943	4.393168 4	1.12E-05	0.001120 92	<i>zmp:000000084</i> <i>6</i>
ENSDARG00000079 327	2.410412633	4.382083 35	1.18E-05	0.001158 32	<i>hmcn2</i>

ENSDARG00000005 762	1.850439283	4.379177 17	1.19E-05	0.001165 72	<i>col14a1a</i>
ENSDARG000000075 463	2.69974156	4.378101 59	1.20E-05	0.001165 72	<i>mss51</i>
ENSDARG000000010 758	3.440950731	4.375616 7	1.21E-05	0.001172 15	<i>capn12</i>
ENSDARG000000009 544	2.342694355	4.373553 52	1.22E-05	0.001173 51	<i>cldnb</i>
ENSDARG000000014 259	2.010489216	4.363577	1.28E-05	0.001203 2	<i>eya1</i>
ENSDARG000000099 291	2.216130925	4.355655 89	1.33E-05	0.001214 01	<i>lsr</i>
ENSDARG000000063 435	2.809643432	4.355638 29	1.33E-05	0.001214 01	<i>trpm4b.2</i>
ENSDARG000000054 723	2.11070502	4.354297 99	1.33E-05	0.001214 01	<i>si:ch211- 242b18.1</i>
ENSDARG000000076 309	3.306584974	4.353470 67	1.34E-05	0.001214 01	<i>mxra5b</i>
ENSDARG000000016 718	2.86926011	4.353033 18	1.34E-05	0.001214 01	<i>mmp11b</i>
ENSDARG000000074 508	2.345356315	4.331383 67	1.48E-05	0.001325 2	<i>si:dkey-28e7.3</i>
ENSDARG000000051 861	2.44072722	4.327898 99	1.51E-05	0.001339 06	<i>pkp3a</i>
ENSDARG000000103 454	2.175913146	4.316862 6	1.58E-05	0.001400 2	<i>trpm6</i>
ENSDARG000000029 168	2.553320218	4.304211 58	1.68E-05	0.001456 85	<i>ppfibp2b</i>
ENSDARG000000029 168	2.553320218	4.304211 58	1.68E-05	0.001456 85	<i>olfml1</i>
ENSDARG000000036 558	1.968414111	4.303454 68	1.68E-05	0.001456 85	<i>col18a1a</i>
ENSDARG000000013 415	2.556181684	4.303386 11	1.68E-05	0.001456 85	<i>lmna</i>
ENSDARG000000034 080	2.264082037	4.285841 65	1.82E-05	0.001552 22	<i>plcd1b</i>
ENSDARG000000091 699	2.245465411	4.280634 29	1.86E-05	0.001580 81	<i>capn2a</i>

ENSDARG00000077 540	2.878394261	4.262037 15	2.03E-05	0.001700 78	<i>f2r1.2</i>
ENSDARG00000025 254	2.672967484	4.260229 48	2.04E-05	0.001705 89	<i>s100a10b</i>
ENSDARG00000011 245	2.466683023	4.253102 41	2.11E-05	0.001729 84	<i>esrp1</i>
ENSDARG00000031 658	2.087901127	4.252614 87	2.11E-05	0.001729 84	<i>si:ch211-207d6.2</i>
ENSDARG00000009 208	2.581157548	4.240360 39	2.23E-05	0.001783 08	<i>prkcda</i>
ENSDARG00000079 324	3.218791784	4.236637 41	2.27E-05	0.001786 44	<i>ano9b</i>
ENSDARG00000024 775	2.654930494	4.235569 52	2.28E-05	0.001786 44	<i>si:ch211-129c21.1</i>
ENSDARG00000055 093	2.95602635	4.232335 36	2.31E-05	0.001803 74	<i>cdh27</i>
ENSDARG00000025 348	1.818393185	4.222407 07	2.42E-05	0.001850 02	<i>igfbp5b</i>
ENSDARG00000092 170	3.690296533	4.220772 73	2.43E-05	0.001854 85	<i>apoc1</i>
ENSDARG00000074 646	2.677451697	4.207832 21	2.58E-05	0.001933 65	<i>LOC568792</i>
ENSDARG00000015 552	1.825701426	4.196750 71	2.71E-05	0.002016 21	<i>phactr4a</i>
ENSDARG00000100 312	4.33028723	4.195545 25	2.72E-05	0.002017 83	<i>hoxa13a</i>
ENSDARG00000060 980	2.038539304	4.188297 4	2.81E-05	0.002074 01	<i>atp8b1</i>
ENSDARG00000096 533	7.24650976	4.187146 71	2.82E-05	0.002075 24	<i>rtlgr</i>
ENSDARG00000008 732	2.295217075	4.184797 74	2.85E-05	0.002087 5	<i>zgc:66479</i>
ENSDARG00000020 265	4.268540781	4.173185 06	3.00E-05	0.002177 49	<i>angptl6</i>
ENSDARG00000011 824	2.364263784	4.171697 19	3.02E-05	0.002182 15	<i>pbxip1b</i>
ENSDARG00000040 606	2.216536768	4.168012 85	3.07E-05	0.002208 03	<i>tfap2c</i>

ENSDARG00000098 420	1.971792123	4.166377 12	3.09E-05	0.002214 25	<i>snta1</i>
ENSDARG00000088 590	1.990519435	4.162251 09	3.15E-05	0.002235 22	<i>pxna</i>
ENSDARG00000055 100	1.769270698	4.157021 14	3.22E-05	0.002267 44	<i>cxcl12b</i>
ENSDARG00000104 267	1.842314763	4.153845 54	3.27E-05	0.002283 29	<i>postnb</i>
ENSDARG00000044 132	2.495468934	4.153482 8	3.27E-05	0.002283 29	<i>ogna</i>
ENSDARG00000077 641	1.95885359	4.146782 34	3.37E-05	0.002323 47	<i>thbs3a</i>
ENSDARG00000078 853	2.684957104	4.146598 32	3.37E-05	0.002323 47	<i>arhgef19</i>
ENSDARG00000078 964	2.660127908	4.142851 23	3.43E-05	0.002351 94	<i>esyt2b</i>
ENSDARG00000009 262	2.320799189	4.140324 1	3.47E-05	0.002355 46	<i>si:dkey-97m3.1</i>
ENSDARG00000004 695	2.031448384	4.139658 59	3.48E-05	0.002355 46	<i>six4a</i>
ENSDARG00000022 309	1.982932022	4.129720 22	3.63E-05	0.002439 55	<i>dspa</i>
ENSDARG00000068 219	3.015386248	4.124286 8	3.72E-05	0.002487 7	<i>b3gnt2l</i>
ENSDARG00000098 465	2.231020464	4.119689 61	3.79E-05	0.002527 58	<i>si:ch73-366l1.5</i>
ENSDARG00000103 515	3.010250549	4.107353 5	4.00E-05	0.002634 4	<i>vcana</i>
ENSDARG00000018 814	2.436005133	4.103058 21	4.08E-05	0.002673 11	<i>esrp2</i>
ENSDARG00000039 806	4.815357171	4.096323 01	4.20E-05	0.002712 18	<i>tbxtb</i>
ENSDARG00000060 626	2.156075623	4.095504 62	4.21E-05	0.002712 18	<i>dgkaa</i>
ENSDARG00000086 539	2.839947467	4.095029 33	4.22E-05	0.002712 18	<i>sh2d3a</i>
ENSDARG00000044 588	1.693974526	4.094877 71	4.22E-05	0.002712 18	<i>emp2</i>

ENSDARG00000039 486	2.72273396	4.094423 68	4.23E-05	0.002712 18	<i>bag3</i>
ENSDARG00000040 907	6.615110417	4.087774 06	4.36E-05	0.002767 23	<i>gcgb</i>
ENSDARG00000089 235	4.929913515	4.079591 73	4.51E-05	0.002855 42	<i>tspearb</i>
ENSDARG00000102 825	3.237162584	4.078314 28	4.54E-05	0.002860 16	<i>olfm2b</i>
ENSDARG00000077 982	3.423348012	4.073306 56	4.64E-05	0.002903 35	<i>elf3</i>
ENSDARG00000055 416	3.442443501	4.072109 64	4.66E-05	0.002903 35	<i>serpinb1</i>
ENSDARG00000040 944	2.815187744	4.071823 91	4.66E-05	0.002903 35	<i>ntd5</i>
ENSDARG00000061 383	3.189847345	4.071286 29	4.68E-05	0.002903 35	<i>serpinf2b</i>
ENSDARG00000027 611	2.16947128	4.068014 81	4.74E-05	0.002918 05	<i>cavin2a</i>
ENSDARG00000014 670	3.375878491	4.067048 23	4.76E-05	0.002918 05	<i>colgalt2</i>
ENSDARG00000019 442	1.960482113	4.048544 88	5.15E-05	0.003129 5	<i>amotl1</i>
ENSDARG00000041 339	2.615774082	4.044225 06	5.25E-05	0.003176 04	<i>zgc:92380</i>
ENSDARG00000022 807	2.673435273	4.042315 32	5.29E-05	0.003190 29	<i>pitpnc1b</i>
ENSDARG00000018 566	2.901208582	4.036553 04	5.42E-05	0.003245 86	<i>flnca</i>
ENSDARG00000011 170	1.488325167	4.029453 97	5.59E-05	0.003321 29	<i>ndrg2</i>
ENSDARG00000017 624	2.237651597	4.028182 9	5.62E-05	0.003327 29	<i>krt4</i>
ENSDARG00000070 961	1.932766935	4.026440 56	5.66E-05	0.003340 01	<i>lepr</i>
ENSDARG00000091 902	2.009884531	4.025357 37	5.69E-05	0.003343 44	<i>b3gnt2b</i>
ENSDARG00000098 766	1.748718913	4.020250 04	5.81E-05	0.003404 6	<i>pcxa</i>

ENSDARG00000037 278	3.746238102	4.017189 08	5.89E-05	0.003436 88	<i>lrata</i>
ENSDARG00000103 747	2.151703914	4.011464 49	6.03E-05	0.003495 75	<i>cav1</i>
ENSDARG00000045 958	2.272811864	4.010844 06	6.05E-05	0.003495 75	<i>egfl6</i>
ENSDARG00000076 332	1.611247791	4.010686 23	6.05E-05	0.003495 75	<i>si:ch211-159i8.4</i>
ENSDARG00000007 219	2.082196497	4.005073 53	6.20E-05	0.003554 85	<i>actn1</i>
ENSDARG00000056 210	3.639647156	4.001239 39	6.30E-05	0.003586 09	<i>hspa1b</i>
ENSDARG00000076 192	3.49132523	4.001187 47	6.30E-05	0.003586 09	<i>ankrd1b</i>
ENSDARG00000089 162	2.478176328	4.000544 84	6.32E-05	0.003586 09	<i>afap1l1a</i>
ENSDARG00000018 971	1.919090003	3.999168 31	6.36E-05	0.003594 62	<i>b3gnt5a</i>
ENSDARG00000079 296	4.31819718	3.997618 83	6.40E-05	0.003605 83	<i>gcga</i>
ENSDARG00000033 161	10.59758471	3.996416 97	6.43E-05	0.003611 81	<i>sst1.2</i>
ENSDARG00000008 772	2.424856471	3.993898 52	6.50E-05	0.003634 77	<i>cacng1a</i>
ENSDARG00000062 590	1.694492053	3.993303 87	6.52E-05	0.003634 77	<i>pleca</i>
ENSDARG00000114 086	2.106146516	3.988475 44	6.65E-05	0.003684 55	<i>zmp:000000120</i> <i>0</i>
ENSDARG00000035 514	2.497162016	3.986161 7	6.72E-05	0.003708 16	<i>adam28</i>
ENSDARG00000014 465	2.604193415	3.981193 99	6.86E-05	0.003761 29	<i>arhgef25b</i>
ENSDARG00000089 645	3.100265136	3.979051 24	6.92E-05	0.003773 61	<i>si:ch1073-</i> <i>406i10.2</i>
ENSDARG00000044 261	2.820129592	3.978837 41	6.93E-05	0.003773 61	<i>si:ch211-</i> <i>243g18.2</i>
ENSDARG00000045 064	1.600533318	3.974538 4	7.05E-05	0.003817 14	<i>ablim1b</i>

ENSDARG00000076 280	2.879329788	3.970315 54	7.18E-05	0.003834 97	<i>LOC100333163</i>
ENSDARG00000076 280	2.879329788	3.970315 54	7.18E-05	0.003834 97	<i>ppp1r1b</i>
ENSDARG00000089 441	2.788363957	3.965032 31	7.34E-05	0.003883 09	<i>si:ch211-105c13.3</i>
ENSDARG00000002 509	4.11951212	3.963071 77	7.40E-05	0.003901 88	<i>zgc:153911</i>
ENSDARG00000088 091	2.228522986	3.962134 93	7.43E-05	0.003901 88	<i>pfn1</i>
ENSDARG00000102 777	3.1782457	3.959061 19	7.52E-05	0.003930 89	<i>thbs4a</i>
ENSDARG00000009 401	2.929424354	3.957355 63	7.58E-05	0.003946 53	<i>vcanb</i>
ENSDARG00000091 253	2.350709041	3.948947 89	7.85E-05	0.004049 11	<i>smyd1b</i>
ENSDARG00000038 703	4.421535976	3.948208 57	7.87E-05	0.004049 11	<i>hkdc1</i>
ENSDARG00000042 934	1.7437747	3.939739 97	8.16E-05	0.004168 68	<i>ccn2a</i>
ENSDARG00000016 505	2.8432416	3.938644 22	8.19E-05	0.004174 8	<i>apbb1ip</i>
ENSDARG00000056 248	4.449180663	3.935486 59	8.30E-05	0.004217 02	<i>si:dkey-183i3.5</i>
ENSDARG00000076 870	1.72832769	3.933338 05	8.38E-05	0.004241 81	<i>piezo1</i>
ENSDARG00000041 257	2.313586474	3.928402 77	8.55E-05	0.004315 63	<i>smtnl1</i>
ENSDARG00000056 499	1.896975917	3.927715 2	8.58E-05	0.004315 63	<i>ca6</i>
ENSDARG00000099 880	7.417180964	3.923303 05	8.73E-05	0.004368 75	<i>sp6</i>
ENSDARG00000100 088	2.148017951	3.909689 51	9.24E-05	0.004608 42	<i>zmp:0000001114</i>
ENSDARG00000035 562	2.026989783	3.907980 23	9.31E-05	0.004627 12	<i>mpdu1a</i>
ENSDARG00000078 605	3.549620716	3.901390 92	9.56E-05	0.004726 36	<i>ptger1a</i>

ENSDARG00000036 279	1.503001832	3.892206 18	9.93E-05	0.004836 35	<i>lamc1</i>
ENSDARG00000008 697	1.789906553	3.886716 12	0.000101 61	0.004919 11	<i>epas1a</i>
ENSDARG00000071 196	1.602581374	3.886656 28	0.000101 64	0.004919 11	<i>cavin2b</i>
ENSDARG00000014 246	2.83679717	3.878148 14	0.000105 26	0.005079 38	<i>jag2a</i>
ENSDARG00000022 689	2.625395135	3.865600 11	0.000110 82	0.005301 15	<i>itgb1b.2</i>
ENSDARG00000076 781	1.80981955	3.864290 57	0.000111 41	0.005314 22	<i>trim45</i>
ENSDARG00000014 358	1.919175768	3.859242 1	0.000113 74	0.005409 52	<i>optc</i>
ENSDARG00000024 030	1.952834596	3.853887 18	0.000116 26	0.005513 34	<i>angptl2a</i>
ENSDARG00000043 211	2.056466839	3.851563 14	0.000117 37	0.005549 94	<i>ripk4</i>
ENSDARG00000020 239	2.015073806	3.849847 25	0.000118 19	0.005572 95	<i>lpin1</i>
ENSDARG00000101 423	2.267249466	3.844951 43	0.000120 58	0.005643 29	<i>cyp2y3</i>
ENSDARG00000043 242	2.37529548	3.835652 7	0.000125 23	0.005805 09	<i>si:dkey-222f2.1</i>
ENSDARG00000079 745	1.90141165	3.833661 87	0.000126 25	0.005819 85	<i>si:ch211-166a6.5</i>
ENSDARG00000060 235	2.619761127	3.833647 39	0.000126 26	0.005819 85	<i>mst1ra</i>
ENSDARG00000096 887	1.964659492	3.826587 23	0.000129 93	0.005955 89	<i>znf469</i>
ENSDARG00000040 009	2.106639814	3.825026 91	0.000130 76	0.005958 32	<i>palld</i>
ENSDARG00000103 311	2.887723517	3.817499 52	0.000134 81	0.006077 95	<i>fbln1</i>
ENSDARG00000105 230	2.38023864	3.812581 64	0.000137 52	0.006183 27	<i>fgf4</i>
ENSDARG00000074 212	2.514057097	3.807502 54	0.000140 38	0.006294 41	<i>si:dkeyp-9d4.5</i>

ENSDARG00000098 641	2.026387465	3.802261 95	0.000143 38	0.006411 63	<i>nbeal2</i>
ENSDARG00000005 216	2.118058604	3.794347 18	0.000148 03	0.006566 1	<i>zgc:158328</i>
ENSDARG00000069 473	2.3273916	3.788481 36	0.000151 57	0.006687	<i>frem1a</i>
ENSDARG00000068 148	3.286446182	3.785352 16	0.000153 49	0.006742 87	<i>arrdc1a</i>
ENSDARG00000103 854	1.810056527	3.785083 93	0.000153 66	0.006742 87	<i>pdlim7</i>
ENSDARG00000007 077	1.984291538	3.781708 74	0.000155 76	0.006816 79	<i>ankrd50l</i>
ENSDARG00000040 284	2.295444411	3.779724 98	0.000157	0.006853 1	<i>si:dkey-79d12.5</i>
ENSDARG00000096 932	2.293941781	3.778904 12	0.000157 52	0.006857 54	<i>si:ch211-37e10.2</i>
ENSDARG00000037 655	1.806131606	3.773220 66	0.000161 15	0.006974 19	<i>pls3</i>
ENSDARG00000061 764	1.633414793	3.772730 05	0.000161 47	0.006974 19	<i>ahnak</i>
ENSDARG00000079 281	3.98324661	3.768332 14	0.000164 34	0.007079 61	<i>si:dkeyp-75h12.2</i>
ENSDARG00000094 752	2.069860698	3.767391 08	0.000164 96	0.007087 79	<i>rpe65b</i>
ENSDARG00000091 127	2.078328133	3.763775 13	0.000167 37	0.007153 76	<i>klf15</i>
ENSDARG00000103 347	2.465995133	3.761127 84	0.000169 15	0.007185 19	<i>cyp2aa3</i>
ENSDARG00000005 690	1.758196708	3.760738 26	0.000169 41	0.007185 19	<i>slc25a23a</i>
ENSDARG00000019 235	1.449586428	3.757802 79	0.000171 41	0.007251 3	<i>sema3aa</i>
ENSDARG00000075 752	2.132093139	3.756621 99	0.000172 22	0.007266 9	<i>myo18aa</i>
ENSDARG00000074 201	1.682528091	3.748972 83	0.000177 56	0.007439 18	<i>flna</i>
ENSDARG00000016 311	2.045935017	3.748391 16	0.000177 97	0.007439 18	<i>dock9b</i>

ENSDARG00000026 473	1.99692807	3.748192 34	0.000178 11	0.007439 18	<i>six1b</i>
ENSDARG00000061 579	1.66265004	3.744395 23	0.000180 83	0.007533 47	<i>myo1cb</i>
ENSDARG00000068 397	2.35862334	3.743062 67	0.000181 79	0.007538 05	<i>tns2b</i>
ENSDARG00000041 065	2.313600122	3.742974 03	0.000181 86	0.007538 05	<i>hspb1</i>
ENSDARG00000087 303	2.338097767	3.736311 01	0.000186 74	0.007686 92	<i>cebpd</i>
ENSDARG00000103 612	2.330428493	3.736097 84	0.000186 9	0.007686 92	<i>akt2l</i>
ENSDARG00000006 314	1.891628567	3.735537 03	0.000187 32	0.007686 92	<i>itgav</i>
ENSDARG00000102 106	2.247040968	3.731644 57	0.000190 23	0.007787 28	<i>tgm1l1</i>
ENSDARG00000070 787	1.592640551	3.719398 36	0.000199 7	0.008074 27	<i>jupa</i>
ENSDARG00000030 006	2.524682887	3.718254 91	0.000200 6	0.008091 02	<i>slc6a7</i>
ENSDARG00000034 210	2.155755609	3.715790 24	0.000202 57	0.008150 33	<i>si:dkey-240h12.4</i>
ENSDARG00000016 691	2.125046804	3.712089 86	0.000205 56	0.008250 28	<i>cd9b</i>
ENSDARG00000093 931	2.91546858	3.710484 02	0.000206 86	0.008282 59	<i>rflnb</i>
ENSDARG00000000 796	2.29971847	3.708445 64	0.000208 54	0.008329 27	<i>nr4a1</i>
ENSDARG00000110 186	2.259712814	3.706058 54	0.000210 51	0.008367 51	<i>cdcp1b</i>
ENSDARG00000015 947	1.815048106	3.702749 95	0.000213 28	0.008457 <i>matn4</i>	
ENSDARG00000039 987	1.541903576	3.701660 21	0.000214 19	0.008472 99	<i>hivep2a</i>
ENSDARG00000015 793	2.499960737	3.693064 32	0.000221 57	0.008722 78	<i>creb3l1</i>
ENSDARG00000002 172	2.107952233	3.689745	0.000224 48	0.008816 29	<i>aplnra</i>

ENSDARG00000103 720	2.018821403	3.688846 91	0.000225 27	0.008826 4	<i>zgc:162730</i>
ENSDARG00000010 478	2.506104296	3.683087 22	0.000230 43	0.009006 88	<i>hsp90aa1.1</i>
ENSDARG00000009 026	1.918033559	3.677775 98	0.000235 28	0.009152 98	<i>ank2a</i>
ENSDARG00000010 085	2.110637513	3.677079 32	0.000235 92	0.009156 36	<i>p4ha2</i>
ENSDARG00000104 022	2.429491951	3.674889 09	0.000237 95	0.009198 18	<i>gcgra</i>
ENSDARG00000069 463	1.611107622	3.674442 15	0.000238 37	0.009198 18	<i>alox12</i>
ENSDARG00000017 036	2.523353582	3.674116 21	0.000238 67	0.009198 18	<i>svilc</i>
ENSDARG00000070 353	3.774484232	3.672334 77	0.000240 35	0.009234 16	<i>hoxc13a</i>
ENSDARG00000100 518	1.91550702	3.671925 26	0.000240 73	0.009234 16	<i>zmp:000000066 0</i>
ENSDARG00000057 606	2.434884055	3.670625 01	0.000241 96	0.009238 2	<i>arl4ca</i>
ENSDARG00000088 813	2.398679073	3.669944 42	0.000242 6	0.009241 39	<i>ppp1r3ab</i>
ENSDARG00000053 535	2.341824557	3.668453 22	0.000244 02	0.009273 98	<i>lmo7b</i>
ENSDARG00000090 072	2.503302118	3.664164 02	0.000248 15	0.009409	<i>krtcap3</i>
ENSDARG00000101 576	2.497288572	3.662482 63	0.000249 78	0.009427 58	<i>tbxta</i>
ENSDARG00000078 114	1.782678175	3.660943 69	0.000251 29	0.009462 69	<i>si:ch73-237c6.1</i>
ENSDARG00000008 548	2.129210456	3.656777 07	0.000255 41	0.009595 81	<i>arhgap12a</i>
ENSDARG00000030 896	5.126819064	3.652657 75	0.000259 54	0.009684 79	<i>foxq1a</i>
ENSDARG00000089 563	3.859078192	3.652068 33	0.000260 14	0.009685 09	<i>sfrp2l</i>
ENSDARG00000041 747	2.828847876	3.651124 9	0.000261 09	0.009698 8	<i>sspn</i>

ENSDARG00000039 677	1.652178666	3.644619 86	0.000267 79	0.009884 39	<i>dsc2l</i>
ENSDARG00000060 871	2.070875437	3.638584 4	0.000274 14	0.010042 07	<i>mctp1b</i>
ENSDARG00000054 060	2.129766856	3.638146 8	0.000274 61	0.010042 07	<i>pof1b</i>
ENSDARG00000020 656	1.971453744	3.636767 53	0.000276 08	0.010064 88	<i>prkg1a</i>
ENSDARG00000012 796	2.191594053	3.636419 47	0.000276 45	0.010064 88	<i>hnmt</i>
ENSDARG00000024 546	1.894736411	3.634281 13	0.000278 76	0.010115 24	<i>pla2g4aa</i>
ENSDARG00000034 588	2.53115357	3.633779 6	0.000279 3	0.010115 24	<i>scn4ab</i>
ENSDARG00000033 466	2.230343381	3.632799 44	0.000280 36	0.010117 66	<i>tagln2</i>
ENSDARG00000098 954	1.970253946	3.632169 63	0.000281 05	0.010120 2	<i>stard13b</i>
ENSDARG00000040 695	1.964355451	3.630597 04	0.000282 77	0.010159 83	<i>fkbp14</i>
ENSDARG00000090 190	1.59529118	3.626705 19	0.000287 06	0.010269 29	<i>bcam</i>
ENSDARG00000061 941	2.419376123	3.625135 86	0.000288 81	0.010298 69	<i>trpv4</i>
ENSDARG00000053 617	2.318789164	3.624786 58	0.000289 2	0.010298 69	<i>camk2a</i>
ENSDARG00000026 736	1.776290059	3.619823 45	0.000294 8	0.010421 4	<i>il11ra</i>
ENSDARG00000026 736	1.776290059	3.619823 45	0.000294 8	0.010421 4	<i>LOC556326</i>
ENSDARG00000100 697	2.004786647	3.615953 5	0.000299 24	0.010508 99	<i>LOC103908986</i>
ENSDARG00000055 439	4.088364754	3.615724 49	0.000299 51	0.010508 99	<i>adamtsl7</i>
ENSDARG00000034 207	3.176053624	3.605065 92	0.000312 07	0.010903 73	<i>cx44.2</i>
ENSDARG00000078 989	2.654871205	3.605057 2	0.000312 08	0.010903 73	<i>alpk3a</i>

ENSDARG00000055 759	1.748607587	3.598148 23	0.000320 49	0.011126 58	<i>efhd2</i>
ENSDARG00000092 947	2.116369447	3.594818 67	0.000324 62	0.011207 74	<i>cyt1</i>
ENSDARG00000006 422	2.802960077	3.591820 61	0.000328 38	0.011257 82	<i>esyt3</i>
ENSDARG00000021 184	2.278020382	3.591204 75	0.000329 15	0.011261	<i>rbfox1l</i>
ENSDARG00000054 122	2.396025284	3.586098 81	0.000335 66	0.011436 13	<i>tmem30b</i>
ENSDARG00000015 536	1.395997576	3.582277 8	0.000340 61	0.011580 78	<i>sox6</i>
ENSDARG00000042 409	1.788184076	3.573131 7	0.000352 74	0.011919 19	<i>si:ch211- 251j10.3</i>
ENSDARG00000101 023	2.867850599	3.570051 32	0.000356 91	0.012035 52	<i>msx2b</i>
ENSDARG00000052 341	2.020949863	3.569197 68	0.000358 08	0.012050 11	<i>sgcb</i>
ENSDARG00000020 759	1.665168973	3.562278 43	0.000367 65	0.012321 9	<i>elf1</i>
ENSDARG00000056 922	2.222603787	3.558892 73	0.000372 42	0.012456 44	<i>ltbp1</i>
ENSDARG00000005 629	2.350970122	3.558114 53	0.000373 53	0.012468 06	<i>smyd2b</i>
ENSDARG00000045 929	1.669796329	3.557279 08	0.000374 72	0.012482 44	<i>oaz2a</i>
ENSDARG00000098 764	2.350307502	3.555014 63	0.000377 96	0.012539 67	<i>musk</i>
ENSDARG00000023 062	2.069600316	3.553612 39	0.000379 98	0.012581 35	<i>ccn1</i>
ENSDARG00000074 148	1.838969607	3.552348 38	0.000381 81	0.012616 56	<i>rbpj1</i>
ENSDARG00000076 673	1.66871237	3.539845 06	0.000400 36	0.013059 35	<i>wu:fi04e12</i>
ENSDARG00000076 321	1.841774585	3.539637 67	0.000400 68	0.013059 35	<i>col28a2a</i>
ENSDARG00000012 788	2.94442388	3.539580 14	0.000400 76	0.013059 35	<i>foxa3</i>

ENSDARG00000031 782	1.682252832	3.535338 13	0.000407 25	0.013192 45	<i>crybg1a</i>
ENSDARG00000062 430	1.758974261	3.534512 67	0.000408 53	0.013207 73	<i>gpd2</i>
ENSDARG00000012 405	2.5101946	3.533021 73	0.000410 84	0.013208 17	<i>col1a1a</i>
ENSDARG00000079 834	2.415263027	3.532307 58	0.000411 95	0.013208 17	<i>kdf1a</i>
ENSDARG00000000 212	2.843570488	3.532078 58	0.000412 31	0.013208 17	<i>krt97</i>
ENSDARG00000019 601	1.508764065	3.526228 45	0.000421 52	0.013443 02	<i>col12a1b</i>
ENSDARG00000003 193	2.12011812	3.525168 71	0.000423 21	0.013470 79	<i>rassf7b</i>
ENSDARG00000070 352	3.166555971	3.523982 65	0.000425 11	0.013505 11	<i>hoxc12a</i>
ENSDARG00000036 830	2.390114109	3.519001 77	0.000433 17	0.013734 7	<i>krt91</i>
ENSDARG00000074 843	1.59106609	3.510387 59	0.000447 45	0.014133 02	<i>phldb2b</i>
ENSDARG00000103 479	1.595473114	3.506827 54	0.000453 48	0.014270 24	<i>ptprfa</i>
ENSDARG00000056 795	2.899501122	3.506798 38	0.000453 53	0.014270 24	<i>serpine1</i>
ENSDARG00000078 419	2.675805909	3.502327 17	0.000461 21	0.014456 61	<i>LOC101884302</i>
ENSDARG00000078 552	4.79531851	3.501682 8	0.000462 33	0.014464 06	<i>grhl3</i>
ENSDARG00000060 102	1.516528131	3.500775 31	0.000463 91	0.014479 99	<i>kank1a</i>
ENSDARG00000038 694	2.178875005	3.498665 56	0.000467 59	0.014545 76	<i>zgc:101744</i>
ENSDARG00000002 968	4.886873173	3.494765 96	0.000474 48	0.014704 34	<i>a1cf</i>
ENSDARG00000054 137	2.835426385	3.491016 77	0.000481 19	0.014857 65	<i>adgrg6</i>
ENSDARG00000078 468	1.600138234	3.490667 57	0.000481 82	0.014857 65	<i>fap</i>

ENSDARG00000045 070	3.143322452	3.490491 28	0.000482 13	0.014857 65	<i>itgb3b</i>
ENSDARG00000009 438	2.420498047	3.489636 11	0.000483 68	0.014877 41	<i>myog</i>
ENSDARG00000103 117	2.375977881	3.483096 73	0.000495 65	0.015203 5	<i>tmprss4a</i>
ENSDARG00000032 619	1.584636728	3.479610 7	0.000502 14	0.015330 73	<i>tob1a</i>
ENSDARG00000115 657	1.734878306	3.477163 99	0.000506 75	0.015442 67	<i>prkcaa</i>
ENSDARG00000020 929	2.051290633	3.475832 11	0.000509 27	0.015490 88	<i>fam49ba</i>
ENSDARG00000027 088	1.860648729	3.474909 39	0.000511 03	0.015515 58	<i>ptgdsb.1</i>
ENSDARG00000006 546	2.173368127	3.470311 5	0.000519 86	0.015754 58	<i>ak4</i>
ENSDARG00000068 177	1.459986406	3.469060 78	0.000522 28	0.015780 86	<i>pak2a</i>
ENSDARG00000078 318	2.130009203	3.468548 43	0.000523 28	0.015780 86	<i>tmod1</i>
ENSDARG00000093 899	2.720531999	3.468383 7	0.000523 6	0.015780 86	<i>si:ch211- 142k18.1</i>
ENSDARG00000078 037	1.948830445	3.466361 85	0.000527 55	0.015870 95	<i>chpfb</i>
ENSDARG00000095 675	2.283144784	3.463031 4	0.000534 13	0.016029 28	<i>ccdc141</i>
ENSDARG00000037 943	2.405368429	3.462710 25	0.000534 76	0.016029 28	<i>cpt1cb</i>
ENSDARG00000027 609	2.565592177	3.460207 59	0.000539 76	0.016131 32	<i>hpn</i>
ENSDARG00000021 833	1.454066468	3.460023 27	0.000540 13	0.016131 32	<i>ahr2</i>
ENSDARG00000016 364	2.521883418	3.453320 95	0.000553 73	0.016446 91	<i>gna15.1</i>
ENSDARG00000040 046	1.840597218	3.452834 91	0.000554 73	0.016446 91	<i>snai2</i>
ENSDARG00000023 578	1.549891768	3.451964 98	0.000556 52	0.016446 91	<i>lpp</i>

ENSDARG00000044 655	2.597643841	3.451457 46	0.000557 57	0.016446 91	<i>st14b</i>
ENSDARG00000100 973	1.606294077	3.450765 01	0.000559	0.016446 91	<i>arhgap24</i>
ENSDARG00000019 845	1.721095709	3.450407 04	0.000559 74	0.016446 91	<i>pdlim1</i>
ENSDARG00000025 299	1.96368341	3.450138 53	0.000560 3	0.016446 91	<i>tspan9a</i>
ENSDARG00000003 395	2.46301961	3.449950 14	0.000560 69	0.016446 91	<i>col4a3</i>
ENSDARG00000068 182	3.285157486	3.447127 48	0.000566 58	0.016590 14	<i>crb3b</i>
ENSDARG00000056 530	1.561388198	3.445340 46	0.000570 34	0.016670 56	<i>cpamd8</i>
ENSDARG00000091 234	2.677604958	3.436674 55	0.000588 9	0.017153	<i>si:ch73-335/21.4</i>
ENSDARG00000052 063	1.512729439	3.436649 24	0.000588 96	0.017153	<i>col4a5</i>
ENSDARG00000099 889	2.104613473	3.436182 71	0.000589 97	0.017153	<i>dusp27</i>
ENSDARG00000109 267	2.070269156	3.435692 4	0.000591 04	0.017153 76	<i>bcam</i>
ENSDARG00000037 961	2.223309244	3.432433 5	0.000598 19	0.017270 73	<i>rcn3</i>
ENSDARG00000016 233	2.083634805	3.431987 62	0.000599 18	0.017270 73	<i>sdr16c5a</i>
ENSDARG00000016 260	1.799444324	3.431635 23	0.000599 95	0.017270 73	<i>fxr2</i>
ENSDARG00000074 575	2.027911652	3.430996 22	0.000601 37	0.017270 73	<i>ppp1r9alb</i>
ENSDARG00000086 261	5.757739482	3.428720 52	0.000606 43	0.017385 84	<i>fhdc4</i>
ENSDARG00000115 540	2.657528258	3.425094 93	0.000614 59	0.017554 72	<i>cacng4a</i>
ENSDARG00000071 150	1.462652875	3.424679 73	0.000615 53	0.017554 72	<i>cica</i>
ENSDARG00000042 652	1.314942117	3.422698 96	0.000620 03	0.017652 54	<i>rreb1b</i>

ENSDARG00000019 949	2.131648386	3.421109 66	0.000623 66	0.017694 79	<i>serpinh1b</i>
ENSDARG00000078 430	1.322330271	3.420160 24	0.000625 84	0.017726 1	<i>tiam1a</i>
ENSDARG00000061 798	2.088099736	3.417994 46	0.000630 84	0.017837 06	<i>arhgap27l</i>
ENSDARG00000090 585	1.753031084	3.416586 33	0.000634 12	0.017898 81	<i>gpc1b</i>
ENSDARG00000012 066	1.554169597	3.408950 27	0.000652 13	0.018313 16	<i>dcn</i>
ENSDARG00000030 494	2.24811277	3.407504 92	0.000655 6	0.018379 06	<i>hvj</i>
ENSDARG00000102 867	1.769981113	3.404617 44	0.000662 57	0.018485 02	<i>srfb</i>
ENSDARG00000007 398	1.853589678	3.404505 7	0.000662 84	0.018485 02	<i>lrrk1</i>
ENSDARG00000039 455	2.310777859	3.404183 54	0.000663 62	0.018485 02	<i>tspan15</i>
ENSDARG00000037 748	2.445596313	3.404081 21	0.000663 87	0.018485 02	<i>slc43a1b</i>
ENSDARG00000045 946	2.050171841	3.402982 64	0.000666 55	0.018528 15	<i>sec24d</i>
ENSDARG00000063 572	1.42106128	3.400381 71	0.000672 92	0.018673 77	<i>perp</i>
ENSDARG00000092 638	1.457509117	3.396770 65	0.000681 86	0.018890 07	<i>ptprua</i>
ENSDARG00000019 753	2.119746164	3.393780 36	0.000689 35	0.019058 2	<i>kcnn3</i>
ENSDARG00000062 059	1.989988222	3.391055 12	0.000696 24	0.019191 53	<i>si:ch211- 236l14.4</i>
ENSDARG00000073 935	2.881312513	3.389225 41	0.000700 9	0.019243 37	<i>wnt6b</i>
ENSDARG00000008 487	1.963121483	3.387395 54	0.000705 6	0.019287 85	<i>dmd</i>
ENSDARG00000079 543	3.431800568	3.386901 97	0.000706 87	0.019290 54	<i>dpys</i>
ENSDARG00000055 715	2.221512395	3.385546 29	0.000710 37	0.019353 98	<i>capn8</i>

ENSDARG00000026 165	1.734160129	3.383542 28	0.000715 57	0.019431 44	<i>col11a1a</i>
ENSDARG00000043 066	2.039768727	3.381424 4	0.000721 11	0.019549 59	<i>nog2</i>
ENSDARG00000002 445	1.634027519	3.378901 4	0.000727 76	0.019697 43	<i>prdm1a</i>
ENSDARG00000052 039	2.558170741	3.375246 35	0.000737 5	0.019928 15	<i>caspb</i>
ENSDARG00000039 652	1.850609269	3.374773 46	0.000738 77	0.019929 69	<i>si:ch211- 221n20.8</i>
ENSDARG00000067 829	1.512176034	3.373752 03	0.000741 51	0.019971 05	<i>ppargc1a</i>
ENSDARG00000078 108	2.986394337	3.372616 24	0.000744 58	0.020020 84	<i>dok1a</i>
ENSDARG00000099 088	2.773251727	3.370750 05	0.000749 64	0.020077 38	<i>edar</i>
ENSDARG00000063 563	1.809587944	3.370539 16	0.000750 21	0.020077 38	<i>creb3l2</i>
ENSDARG00000019 874	2.462876004	3.370492 42	0.000750 34	0.020077 38	<i>hsph1</i>
ENSDARG00000063 050	2.316802463	3.369550 14	0.000752 91	0.020113 45	<i>rc3h2</i>
ENSDARG00000091 783	4.066141738	3.366464 03	0.000761 39	0.020274 04	<i>LOC100001051</i>
ENSDARG00000077 295	1.437829199	3.365353 62	0.000764 46	0.020302 02	<i>akap6</i>
ENSDARG00000008 433	2.291963748	3.362675 95	0.000771 91	0.020420 91	<i>unc45b</i>
ENSDARG00000046 019	1.622407426	3.362248 26	0.000773 11	0.020420 91	<i>snai1b</i>
ENSDARG00000040 432	1.793014363	3.360441 61	0.000778 18	0.020522	<i>klf2b</i>
ENSDARG00000080 020	2.707063542	3.358287 94	0.000784 27	0.020649 48	<i>il13ra1</i>
ENSDARG00000060 435	2.523646576	3.352968 18	0.000799 5	0.020983 34	<i>bnip1</i>
ENSDARG00000058 358	1.630628908	3.352074 34	0.000802 09	0.021017 69	<i>krt8</i>

ENSDARG00000071 562	2.205495626	3.351332 57	0.000804 24	0.021040 57	<i>mtus1a</i>
ENSDARG00000017 105	1.486218079	3.342661 92	0.000829 79	0.021640 27	<i>vaspb</i>
ENSDARG00000057 633	1.945532278	3.341092 14	0.000834 5	0.021727 18	<i>cxcr4a</i>
ENSDARG00000076 962	1.618396865	3.340670 86	0.000835 76	0.021727 18	<i>gdpd5b</i>
ENSDARG00000007 671	1.730166367	3.338406 16	0.000842 61	0.021836 07	<i>ghrb</i>
ENSDARG00000005 749	1.872179307	3.334844 45	0.000853 47	0.022013 67	<i>cand2</i>
ENSDARG00000055 592	2.054947704	3.334100 07	0.000855 76	0.022038 13	<i>capn2b</i>
ENSDARG00000052 405	2.64397301	3.326935 16	0.000878 07	0.022490 22	<i>pak6b</i>
ENSDARG00000088 975	3.096856238	3.326827 22	0.000878 41	0.022490 22	<i>si:ch211-14k19.8</i>
ENSDARG00000028 664	2.226411711	3.326709 2	0.000878 78	0.022490 22	<i>ahsa1a</i>
ENSDARG00000007 009	1.881080491	3.325791 79	0.000881 68	0.022501 14	<i>hoxa11b</i>
ENSDARG00000026 052	2.410440901	3.325708 25	0.000881 94	0.022501 14	<i>ccdc3a</i>
ENSDARG00000054 003	2.428887282	3.323656 73	0.000888 46	0.022597 24	<i>rargb</i>
ENSDARG00000039 943	1.778043266	3.323124 47	0.000890 15	0.022605 46	<i>tent5ba</i>
ENSDARG00000029 764	2.108858759	3.317777 7	0.000907 37	0.022971 73	<i>mef2ca</i>
ENSDARG00000062 446	2.46228236	3.315364 73	0.000915 24	0.023064 51	<i>neurl2</i>
ENSDARG00000086 778	2.332027411	3.313764 5	0.000920 49	0.023161 44	<i>si:ch211-79m20.1</i>
ENSDARG00000017 058	1.393462632	3.313121 97	0.000922 61	0.023179 28	<i>large2</i>
ENSDARG00000025 465	1.60762991	3.306300 5	0.000945 37	0.023535 48	<i>spon2a</i>

ENSDARG00000098 639	1.716288129	3.305809 68	0.000947 02	0.023541 13	<i>pycr1b</i>
ENSDARG00000096 454	2.425826824	3.301566	0.000961 47	0.023817 96	<i>ap1m2</i>
ENSDARG00000006 060	2.000250093	3.297465 8	0.000975 62	0.024106 2	<i>crybg1b</i>
ENSDARG00000059 993	2.053060723	3.295598 84	0.000982 12	0.024162 15	<i>trpm4a</i>
ENSDARG00000076 094	2.901035988	3.295484 52	0.000982 52	0.024162 15	<i>klhl30</i>
ENSDARG00000032 856	1.655162866	3.295343 72	0.000983 01	0.024162 15	<i>ccdc120</i>
ENSDARG00000007 682	1.973795345	3.295132 48	0.000983 75	0.024162 15	<i>pdpfa</i>
ENSDARG00000031 983	1.842483471	3.291013 56	0.000998 27	0.024373 22	<i>six4b</i>
ENSDARG00000073 859	1.690958694	3.290112 08	0.001001 48	0.024415 21	<i>kif13bb</i>
ENSDARG00000039 066	2.33752657	3.288395 85	0.001007 6	0.024528 22	<i>klhl31</i>
ENSDARG00000002 411	2.19581767	3.286712 34	0.001013 64	0.024622 06	<i>pde4cb</i>
ENSDARG00000010 515	1.937188049	3.286488 83	0.001014 45	0.024622 06	<i>dvl1a</i>
ENSDARG00000062 518	1.766876242	3.285856 56	0.001016 73	0.024641 05	<i>ulk1a</i>
ENSDARG00000039 093	1.271324174	3.284076 28	0.001023 17	0.024758 01	<i>got1</i>
ENSDARG00000012 790	1.950744228	3.283693 57	0.001024 56	0.024758 01	<i>tmem161a</i>
ENSDARG00000005 526	2.439375776	3.282950 62	0.001027 27	0.024786 95	<i>igfn1.1</i>
ENSDARG00000021 948	1.453836062	3.280335 49	0.001036 84	0.024944 73	<i>tnc</i>
ENSDARG00000003 216	2.049104361	3.278507 18	0.001043 58	0.025027 45	<i>anxa2a</i>
ENSDARG00000043 818	1.905677989	3.277755 74	0.001046 36	0.025027 45	<i>si:ch211- 119c20.2</i>

ENSDARG00000013 921	1.727856969	3.276544 99	0.001050 86	0.025064 97	<i>frya</i>
ENSDARG00000044 011	1.719839743	3.275177 2	0.001055 96	0.025094 76	<i>xkrx</i>
ENSDARG00000004 724	2.230064312	3.274889 95	0.001057 03	0.025094 76	<i>tcea3</i>
ENSDARG00000041 502	1.855667696	3.274593 3	0.001058 14	0.025094 76	<i>tgfb1a</i>
ENSDARG00000105 422	3.640963464	3.274544 94	0.001058 32	0.025094 76	<i>si:dkey-1k23.3</i>
ENSDARG00000099 448	2.742245052	3.272505 33	0.001065 99	0.025240 16	<i>sh3d21</i>
ENSDARG00000041 706	2.229545399	3.271024 39	0.001071 59	0.025336 25	<i>vgl2a</i>
ENSDARG00000011 407	1.670096646	3.269738 53	0.001076 47	0.025389 05	<i>col2a1b</i>
ENSDARG00000004 034	1.733491455	3.269517 41	0.001077 31	0.025389 05	<i>arhgdig</i>
ENSDARG00000069 833	2.148653395	3.269206 86	0.001078 49	0.025389 05	<i>rhbdl2</i>
ENSDARG00000044 155	1.707019019	3.267429 2	0.001085 29	0.025450 99	<i>mafaa</i>
ENSDARG00000017 676	1.768749352	3.267317 33	0.001085 72	0.025450 99	<i>mmp2</i>
ENSDARG00000037 097	2.317358361	3.265057 92	0.001094 42	0.025582	<i>slc7a2</i>
ENSDARG00000098 294	1.265409857	3.263485 18	0.001100 51	0.025651 53	<i>col5a3a</i>
ENSDARG00000030 967	1.784094093	3.262945 93	0.001102 61	0.025664 04	<i>si:ch211-51a6.2</i>
ENSDARG00000074 808	1.751314921	3.262133 71	0.001105 77	0.025701 34	<i>megf6b</i>
ENSDARG00000061 231	1.480109019	3.261512 05	0.001108 2	0.025721 43	<i>tinagl1</i>
ENSDARG00000006 757	2.17349714	3.258357 76	0.001120 59	0.025935 91	<i>klhl41b</i>
ENSDARG00000103 991	2.033232411	3.255468 42	0.001132 06	0.026127 76	<i>sec141</i>

ENSDARG00000045 676	1.688019517	3.254681 28	0.001135 2	0.026163 59	<i>calua</i>
ENSDARG00000077 588	1.571196555	3.252377 31	0.001144 44	0.026266 25	<i>pdgfc</i>
ENSDARG00000069 476	1.799559108	3.249569 13	0.001155 8	0.026490 03	<i>spint2</i>
ENSDARG00000070 857	3.144819289	3.248356 67	0.001160 74	0.026566 18	<i>si:dkey-32e6.6</i>
ENSDARG00000075 759	1.350380659	3.247213 7	0.001165 41	0.026636 06	<i>samd4a</i>
ENSDARG00000067 729	1.615262411	3.245490 76	0.001172 48	0.026723 53	<i>arid3c</i>
ENSDARG00000076 729	1.319056226	3.244618 47	0.001176 08	0.026768 45	<i>myo9aa</i>
ENSDARG00000070 000	2.105261449	3.243196 61	0.001181 97	0.026842 17	<i>txnipb</i>
ENSDARG00000078 217	2.860786447	3.243048 15	0.001182 58	0.026842 17	<i>cblc</i>
ENSDARG00000077 301	2.173881556	3.241575 05	0.001188 71	0.026944 12	<i>mlf1</i>
ENSDARG00000098 431	1.209305138	3.241007 33	0.001191 08	0.026960 71	<i>comp</i>
ENSDARG00000041 505	1.784776749	3.240142 63	0.001194 7	0.027005 45	<i>itm2bb</i>
ENSDARG00000035 556	1.576087829	3.239010 19	0.001199 45	0.027020 74	<i>rps6ka3a</i>
ENSDARG00000002 084	1.664092635	3.238808 51	0.001200 3	0.027020 74	<i>lamb2</i>
ENSDARG00000058 605	2.480172416	3.238063 15	0.001203 44	0.027054 42	<i>vsig10</i>
ENSDARG00000011 797	2.325388515	3.233051 89	0.001224 75	0.027495 96	<i>tent5bb</i>
ENSDARG00000037 455	1.920064874	3.232602 09	0.001226 68	0.027501 77	<i>ugt8</i>
ENSDARG00000075 556	1.368587692	3.231643 34	0.001230 81	0.027522 02	<i>luzp1</i>
ENSDARG00000011 809	2.435480818	3.226423 51	0.001253 48	0.027865 66	<i>mical1</i>

ENSDARG00000019 392	1.680693617	3.225508 7	0.001257 49	0.027865 66	<i>stat5a</i>
ENSDARG00000028 096	2.826587831	3.225357 61	0.001258 15	0.027865 66	<i>cldn23a</i>
ENSDARG00000058 551	2.51525208	3.223703 05	0.001265 45	0.027951 91	<i>popdc3</i>
ENSDARG00000078 404	1.791309023	3.223049 12	0.001268 34	0.027978 25	<i>cdh26.1</i>
ENSDARG00000077 897	1.817529664	3.219439 82	0.001284 41	0.028221 98	<i>znf296</i>
ENSDARG00000116 806	2.529274725	3.218351 98	0.001289 3	0.028288 85	<i>ppp1r27a</i>
ENSDARG00000101 209	1.37821943	3.217800 9	0.001291 78	0.028305 52	<i>lamb1a</i>
ENSDARG00000076 312	2.245374284	3.216441 7	0.001297 91	0.028402 12	<i>myot</i>
ENSDARG00000010 565	2.344273172	3.214835 77	0.001305 19	0.028523 55	<i>aqp4</i>
ENSDARG00000071 478	2.071232009	3.212948 25	0.001313 8	0.028597 73	<i>lnx2b</i>
ENSDARG00000078 522	2.652871618	3.208776 37	0.001333 01	0.028825 28	<i>si:ch211-80h18.1</i>
ENSDARG00000101 175	1.64402962	3.207108 75	0.001340 76	0.028954 85	<i>mcu</i>
ENSDARG00000026 348	1.663872091	3.206610 86	0.001343 09	0.028957 91	<i>csad</i>
ENSDARG00000014 626	2.093203984	3.206324 22	0.001344 42	0.028957 91	<i>dlx3b</i>
ENSDARG00000002 037	2.530706399	3.204957 09	0.001350 83	0.029019 84	<i>pkfb2b</i>
ENSDARG00000079 198	2.210009388	3.202718 16	0.001361 37	0.029170 22	<i>usp13</i>
ENSDARG00000036 168	1.994059215	3.200782 61	0.001370 55	0.029328 68	<i>nfatc1</i>
ENSDARG00000041 098	3.071414666	3.199952 82	0.001374 5	0.029375 05	<i>barx2</i>
ENSDARG00000099 960	2.712121115	3.198993 85	0.001379 08	0.029434 71	<i>elovl1a</i>

ENSDARG00000038 964	3.459191066	3.195669 86	0.001395 07	0.029698 85	<i>traf4b</i>
ENSDARG00000076 735	1.407157315	3.194862 68	0.001398 97	0.029707 57	<i>si:dkey-193c22.2</i>
ENSDARG00000078 145	1.496315367	3.194839 41	0.001399 09	0.029707 57	<i>si:ch211-218g4.2</i>
ENSDARG00000015 293	2.46343392	3.193761 74	0.001404 32	0.029780 26	<i>fam110a</i>
ENSDARG00000020 594	2.256819393	3.188806 66	0.001428 61	0.030256 46	<i>txlnba</i>
ENSDARG00000103 546	1.365382287	3.187222 89	0.001436 46	0.030383 54	<i>purbb</i>
ENSDARG00000057 138	2.470775635	3.186477 03	0.001440 17	0.030422 88	<i>zgc:174164</i>
ENSDARG00000079 199	3.041821215	3.185537 65	0.001444 85	0.030458 51	<i>megf6a</i>
ENSDARG00000101 569	1.228910414	3.185396 62	0.001445 56	0.030458 51	<i>ppargc1b</i>
ENSDARG00000010 673	2.502868551	3.183117 48	0.001456 99	0.030659 06	<i>trim16</i>
ENSDARG00000062 222	2.913610486	3.182756 65	0.001458 8	0.030659 06	<i>mxra8a</i>
ENSDARG00000052 881	2.19933676	3.180594 7	0.001469 73	0.030770 86	<i>dnajb5</i>
ENSDARG00000089 691	1.560754069	3.179914 69	0.001473 18	0.030803 96	<i>ism2a</i>
ENSDARG00000021 987	1.946321985	3.179388 51	0.001475 86	0.030820 56	<i>plecb</i>
ENSDARG00000052 997	1.845788754	3.178474 73	0.001480 52	0.030820 56	<i>sema4e</i>
ENSDARG00000059 746	1.808972241	3.177948 76	0.001483 21	0.030820 56	<i>plod1a</i>
ENSDARG00000016 570	3.459346898	3.177436 36	0.001485 83	0.030820 56	<i>prlra</i>
ENSDARG00000036 074	1.739035839	3.177274 17	0.001486 66	0.030820 56	<i>cebpa</i>
ENSDARG00000017 901	1.643114584	3.176964 73	0.001488 25	0.030820 56	<i>tln2a</i>

ENSDARG00000068 732	1.419105874	3.175474 14	0.001495 92	0.030886 87	<i>spry4</i>
ENSDARG00000079 462	2.462205003	3.172080 5	0.001513 51	0.031174 01	<i>si:ch211-131k2.3</i>
ENSDARG00000051 809	2.549414617	3.171697 4	0.001515 51	0.031174 01	<i>btr12</i>
ENSDARG00000071 082	1.707908424	3.168201 26	0.001533 85	0.031361 91	<i>p4ha1b</i>
ENSDARG00000052 148	1.964572248	3.168192 63	0.001533 9	0.031361 91	<i>ptgs1</i>
ENSDARG00000045 636	1.552996702	3.167819 51	0.001535 87	0.031361 91	<i>rbl2</i>
ENSDARG00000062 263	1.772933058	3.167310 27	0.001538 56	0.031361 91	<i>arhgap17b</i>
ENSDARG00000100 797	3.368739078	3.166352 27	0.001543 64	0.031399 38	<i>LOC101884488</i>
ENSDARG00000075 600	2.348612037	3.163726 32	0.001557 63	0.031616 33	<i>si:dkeyp-41f9.3</i>
ENSDARG00000100 698	2.531565881	3.162971 98	0.001561 67	0.031639 89	<i>eva1bb</i>
ENSDARG00000040 306	2.872670677	3.159806 8	0.001578 74	0.031800 33	<i>otomp</i>
ENSDARG00000036 987	1.834869636	3.159793 84	0.001578 81	0.031800 33	<i>has2</i>
ENSDARG00000009 194	1.623484703	3.158151 14	0.001587 73	0.031905 74	<i>col16a1</i>
ENSDARG00000020 007	1.615071563	3.158012 59	0.001588 49	0.031905 74	<i>col1a2</i>
ENSDARG00000031 796	1.547558021	3.156922 47	0.001594 44	0.031958 66	<i>ache</i>
ENSDARG00000101 592	1.882903539	3.155838 63	0.001600 37	0.031992 75	<i>col22a1</i>
ENSDARG00000020 107	1.847419993	3.155547 79	0.001601 97	0.031992 75	<i>usp2a</i>
ENSDARG00000101 452	1.888447005	3.154048 05	0.001610 23	0.032076 16	<i>im:7156467</i>
ENSDARG00000059 029	1.806943399	3.153727 64	0.001612	0.032076 16	<i>mmp28</i>

ENSDARG00000075 930	2.221125632	3.152067 48	0.001621 19	0.032181 27	<i>adamts2</i>
ENSDARG00000046 079	2.192787316	3.151210 7	0.001625 95	0.032203 15	<i>cacng6b</i>
ENSDARG00000043 182	1.927788722	3.151128 7	0.001626 41	0.032203 15	<i>hlla2a.1</i>
ENSDARG00000026 611	1.739518955	3.150814 02	0.001628 16	0.032203 15	<i>socs3b</i>
ENSDARG00000020 072	2.078881271	3.145414 11	0.001658 52	0.032643 41	<i>thbs4b</i>
ENSDARG00000055 897	1.503260689	3.144989 62	0.001660 93	0.032643 41	<i>zgc:154093</i>
ENSDARG00000002 204	2.504085938	3.144743 99	0.001662 32	0.032643 41	<i>hspb11</i>
ENSDARG00000061 509	2.976524244	3.144011 67	0.001666 49	0.032686 2	<i>tbx3b</i>
ENSDARG00000006 353	1.396086329	3.1428	0.001673 4	0.032782 73	<i>itga5</i>
ENSDARG00000052 061	1.326137268	3.141262 76	0.001682 21	0.032899 93	<i>col4a6</i>
ENSDARG00000054 864	1.425903733	3.140711 11	0.001685 38	0.032899 93	<i>aplp2</i>
ENSDARG00000058 365	2.258844607	3.139921 13	0.001689 93	0.032913 15	<i>hspb8</i>
ENSDARG00000075 707	1.875271076	3.139899 31	0.001690 06	0.032913 15	<i>nid2a</i>
ENSDARG00000102 948	2.766233673	3.135174 57	0.001717 52	0.033401 42	<i>crb3a</i>
ENSDARG00000089 413	1.633270609	3.134889 11	0.001719 19	0.033401 42	<i>mbpb</i>
ENSDARG00000004 196	1.324403615	3.131416 68	0.001739 65	0.033719 32	<i>tnfsf10l</i>
ENSDARG00000078 227	1.540120695	3.129992 2	0.001748 11	0.033763 95	<i>cspg4</i>
ENSDARG00000071 259	1.475072757	3.129260 89	0.001752 47	0.033805 07	<i>fbxo34</i>
ENSDARG00000098 237	1.36348156	3.128945 77	0.001754 35	0.033805 07	<i>fbn2b</i>

ENSDARG00000077 722	1.611062306	3.127648 11	0.001762 11	0.033875 33	<i>ppp2r3a</i>
ENSDARG00000025 513	1.513376397	3.123942 55	0.001784 45	0.034251 23	<i>dip2cb</i>
ENSDARG00000117 419	2.964538801	3.123716 16	0.001785 83	0.034251 23	<i>LOC110437884</i>
ENSDARG00000102 415	2.004150763	3.122013 72	0.001796 19	0.034369 8	<i>scinla</i>
ENSDARG00000036 371	2.111213544	3.121110 89	0.001801 7	0.034435 31	<i>acta1a</i>
ENSDARG00000038 826	2.111039965	3.120057 2	0.001808 16	0.034478 64	<i>jph1b</i>
ENSDARG00000013 414	2.012712666	3.119357	0.001812 46	0.034498 3	<i>lin7a</i>
ENSDARG00000029 457	2.305938786	3.119207 33	0.001813 38	0.034498 3	<i>cacna1sa</i>
ENSDARG00000004 877	1.611841521	3.116253 54	0.001831 65	0.034748 16	<i>rock2b</i>
ENSDARG00000038 585	1.863231171	3.115265 64	0.001837 79	0.034748 16	<i>EIF4E2</i>
ENSDARG00000055 075	1.612424331	3.114563 29	0.001842 17	0.034748 16	<i>svila</i>
ENSDARG00000073 870	1.502423331	3.114428 77	0.001843 02	0.034748 16	<i>gdpd2</i>
ENSDARG00000068 246	1.265199061	3.112362 77	0.001855 96	0.034881 84	<i>plcb3</i>
ENSDARG00000091 161	1.819260119	3.110702 45	0.001866 43	0.034933 5	<i>rpz</i>
ENSDARG00000028 816	1.80144553	3.110335 92	0.001868 75	0.034933 5	<i>tmed3</i>
ENSDARG00000059 259	2.044133413	3.109514 61	0.001873 95	0.034933 5	<i>pabpc4</i>
ENSDARG00000013 335	1.939896221	3.109346 28	0.001875 02	0.034933 5	<i>anxa6</i>
ENSDARG00000014 804	1.998268026	3.109098 28	0.001876 59	0.034933 5	<i>cacna2d1a</i>
ENSDARG00000058 548	2.135154858	3.106478 16	0.001893 3	0.035161 35	<i>bves</i>

ENSDARG00000079 009	1.463274053	3.106214 82	0.001894 99	0.035161 35	<i>abca1b</i>
ENSDARG00000104 949	1.975855398	3.106175 35	0.001895 24	0.035161 35	<i>slc27a1b</i>
ENSDARG00000041 546	1.893925008	3.103196 72	0.001914 42	0.035468 14	<i>gypc</i>
ENSDARG00000054 031	1.539988523	3.102562 69	0.001918 53	0.035468 14	<i>mxtd4</i>
ENSDARG00000088 232	1.926858455	3.101772 43	0.001923 66	0.035488 47	<i>shisa2</i>
ENSDARG00000016 470	2.292338595	3.099779 03	0.001936 65	0.035688 16	<i>anxa5b</i>
ENSDARG00000103 519	1.559887557	3.098910 63	0.001942 34	0.035752 89	<i>adamts1</i>
ENSDARG00000058 943	2.144443575	3.098220 06	0.001946 87	0.035756 32	<i>cdcp1a</i>
ENSDARG00000102 191	1.442107005	3.097642 45	0.001950 67	0.035786 14	<i>LOC101886030</i>
ENSDARG00000027 017	2.465269728	3.096520 3	0.001958 07	0.035821 38	<i>ppp2r5a</i>
ENSDARG00000061 758	2.132184884	3.095996 35	0.001961 53	0.035821 38	<i>sh3pxd2ab</i>
ENSDARG00000088 515	2.590762776	3.095944 64	0.001961 87	0.035821 38	<i>st6gal2b</i>
ENSDARG00000053 104	1.828681246	3.095865 95	0.001962 39	0.035821 38	<i>fam102ba</i>
ENSDARG00000007 709	2.239623469	3.095702 64	0.001963 47	0.035821 38	<i>adamts8a</i>
ENSDARG00000044 808	1.677123493	3.093140 35	0.001980 5	0.036062 18	<i>slc4a4b</i>
ENSDARG00000053 362	1.203550884	3.093058 15	0.001981 05	0.036062 18	<i>cilp</i>
ENSDARG00000076 564	1.314210087	3.088487 68	0.002011 78	0.036540 68	<i>hspg2</i>
ENSDARG00000063 255	1.599869021	3.084595 6	0.002038 29	0.036859 47	<i>btbd11a</i>
ENSDARG00000027 500	1.368735832	3.083131 39	0.002048 35	0.037000 66	<i>oxsr1b</i>

ENSDARG00000071 113	2.215085985	3.081828 04	0.002057 34	0.037122 29	<i>xirp2a</i>
ENSDARG00000086 881	1.839836328	3.081105	0.002062 34	0.037171 8	<i>ier2b</i>
ENSDARG00000019 815	1.437520408	3.079657 94	0.002072 39	0.037230 52	<i>fn1a</i>
ENSDARG00000051 729	1.731947245	3.078863 71	0.002077 92	0.037289 21	<i>nfatc3b</i>
ENSDARG00000036 426	2.652580821	3.077186 83	0.002089 64	0.037458 78	<i>zgc:103586</i>
ENSDARG00000043 531	1.69174413	3.074758 01	0.002106 73	0.037698 79	<i>jun</i>
ENSDARG00000030 012	1.377640245	3.074321 46	0.002109 82	0.037698 79	<i>lrrfp1a</i>
ENSDARG00000062 082	1.222799712	3.073952 61	0.002112 43	0.037698 79	<i>hipk3b</i>
ENSDARG00000018 881	2.035489186	3.073662 31	0.002114 49	0.037698 79	<i>apobec2a</i>
ENSDARG00000076 229	1.256470239	3.072805 19	0.002120 57	0.037766 32	<i>mrtfab</i>
ENSDARG00000098 191	1.735157838	3.070522 86	0.002136 84	0.037898 98	<i>si:dkey-16l2.20</i>
ENSDARG00000061 391	2.615624714	3.070232 21	0.002138 92	0.037898 98	<i>grhl1</i>
ENSDARG00000074 415	2.464824352	3.070147 21	0.002139 53	0.037898 98	<i>zgc:171482</i>
ENSDARG00000103 322	2.915213124	3.066973 61	0.002162 38	0.038180 38	<i>si:ch73-347e22.8</i>
ENSDARG00000010 059	1.901375972	3.064636 34	0.002179 35	0.038317 14	<i>itpkb</i>
ENSDARG00000059 202	1.720501219	3.064623 96	0.002179 44	0.038317 14	<i>tspan2b</i>
ENSDARG00000030 465	1.678232445	3.062401 17	0.002195 69	0.038561 65	<i>tmem263</i>
ENSDARG00000037 030	1.342936131	3.059352 53	0.002218 16	0.038831 95	<i>casz1</i>
ENSDARG00000104 484	1.246572487	3.058148 44	0.002227 09	0.038946 9	<i>itgb1b</i>

ENSDARG00000006		3.056162	0.002241	0.039081	
301	1.225417582	31	9	23	<i>raph1a</i>
ENSDARG000000019		3.053554	0.002261	0.039339	
342	2.002218375	69	48	15	<i>chrnd</i>
ENSDARG000000069		3.052110	0.002272	0.039445	
415	2.453700122	93	38	47	<i>col17a1a</i>
ENSDARG000000100		3.050979	0.002280	0.039552	
940	1.636233751	68	96	67	<i>nr5a2</i>
ENSDARG000000098		3.049776	0.002290	0.039669	
777	1.581686421	9	11	6	<i>crybg2</i>
ENSDARG000000031		3.047117	0.002310	0.039980	
809	1.894755891	66	47	16	<i>rbm24b</i>

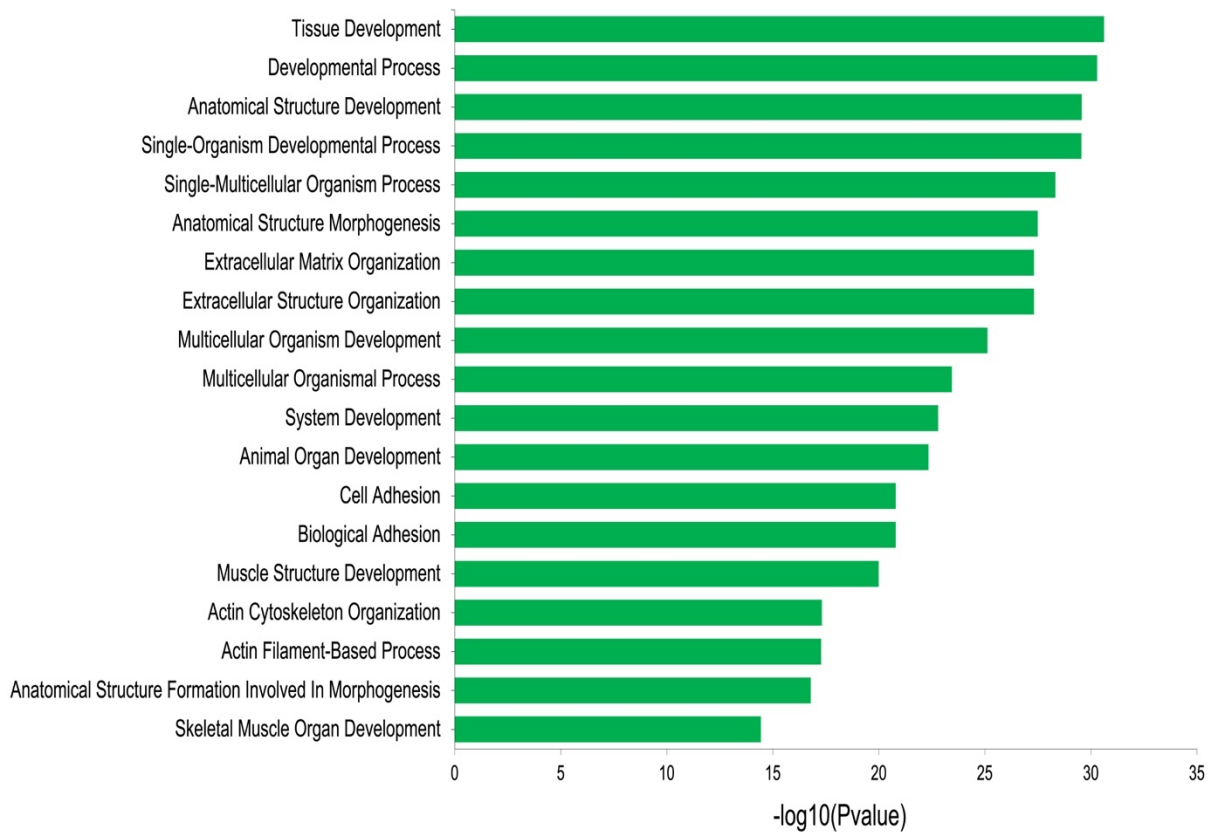


Figure 7-5: Top GO terms associated with transcripts that were upregulated in *sox17E* vs *sox17N* according to GO analysis using DAVID.

Table 7.22: List of top 200 DARs between *sox32* OE and *sox17E* ranked by FDR from Diffbind. Fold enrichment score > 0 represents *sox32* OE DARs while *sox17E* DARs have a fold enrichment score < 0. FDR = False discovery rate

Chromosome	Chromosome start	Chromosome end	Fold enrichment	FDR
chr10	10843687	10844703	-4.27	7.06E-150
chr23	17107556	17109187	5.71	9.58E-109
chr22	28787961	28792815	-3.06	1.59E-103
chr7	58829716	58833449	5.08	1.78E-96
chr11	10875156	10876474	-4.29	3.29E-94
chr1	24872898	24873957	4.56	3.70E-94
chr12	34587943	34589315	-3.63	1.34E-93
chr13	50614419	50615700	4.51	1.53E-89
chr13	49480422	49482221	-3.15	2.67E-89
chr24	35609029	35611029	-3.01	5.17E-88
chr16	26739950	26741259	5.21	5.17E-88
chr21	38459851	38463506	-2.79	4.38E-86
chr11	29821007	29822884	4.67	8.57E-86
chr17	25883012	25885121	5.24	1.80E-84
chr3	5230075	5231303	-4.28	2.93E-84
chr20	30857559	30859674	4.4	1.24E-83
chr11	18740669	18741644	6.17	1.35E-83
chr12	9470844	9472152	4.59	1.80E-83
chr16	11088859	11090718	-3.1	2.76E-82
chr11	35597658	35598995	4.27	4.70E-82
chr23	29656028	29657659	4.87	2.02E-81
chr11	6452180	6453517	4.37	6.56E-81
chr6	36581607	36583127	-3.39	3.65E-80
chr7	58822612	58828705	3.64	3.09E-78
chr23	3760625	3761844	5.04	7.50E-78
chr16	35456890	35458331	-2.9	9.14E-78
chr22	10034539	10036717	4.54	4.48E-77

chr5	35428327	35431025	4.09	5.77E-77
chr11	18810787	18811939	5.46	8.07E-77
chr10	8619340	8620391	5.26	8.42E-77
chr6	15745665	15747136	4.17	8.42E-77
chr5	61492762	61494430	4.9	1.14E-76
chr23	28021223	28024733	4.77	3.50E-76
chr13	38946513	38947990	4.67	4.61E-76
chr20	27746838	27747903	4.65	7.31E-76
chr10	6456437	6458574	4.22	7.31E-76
chr5	17710136	17711268	-3.5	9.96E-76
chr23	33545757	33547033	4.46	1.10E-75
chr5	9023496	9025731	4.7	2.24E-75
chr5	65971035	65972393	5.54	2.78E-75
chr23	28025040	28027491	4.99	4.36E-75
chr2	15146642	15148046	-3.46	6.54E-75
chr13	5049412	5051190	5.08	1.74E-74
chr10	5272348	5273733	4.69	5.07E-74
chr4	9864804	9865897	5.56	5.90E-74
chr25	7387863	7389721	-3.58	6.28E-74
chr21	23608451	23609197	-3.58	1.54E-73
chr5	14325997	14328368	-3.33	8.56E-73
chr15	23047042	23048972	4.27	1.66E-72
chr1	17322823	17324007	4.41	2.15E-72
chr7	9104376	9105784	4.53	2.84E-72
chr13	44838132	44841809	3.47	3.90E-72
chr19	43773869	43774804	5.67	5.80E-72
chr5	30548245	30550056	4.96	9.54E-72
chr25	18332814	18334236	4.46	3.82E-71
chr6	32164026	32165110	4.87	4.24E-71
chr23	37700602	37701377	4.53	6.13E-71
chr21	25816978	25820912	3.94	7.06E-71
chr24	36139428	36142025	4.12	1.06E-70
chr12	16303563	16304917	4.34	1.67E-70
chr11	44170324	44171465	5.24	1.67E-70
chr7	10801821	10802818	5.19	1.07E-69

chr15	25568253	25569447	5.23	1.09E-69
chr1	2295861	2297462	4.06	1.66E-69
chr2	42121446	42123654	5.62	1.90E-69
chr20	28366075	28370252	3.77	2.12E-69
chr2	30695071	30696948	4.14	4.37E-69
chr14	26637890	26640054	-2.82	4.94E-69
chr2	37099608	37101807	3.63	3.81E-68
chr6	39990352	39995719	3.21	4.88E-68
chr22	38453618	38454825	4.53	5.78E-68
chr11	19023693	19025352	4.56	9.66E-68
chr19	27390453	27392095	4.67	1.31E-67
chr25	12501299	12502464	4.77	1.31E-67
chr13	603702	606288	-2.81	1.57E-67
chr20	27750690	27757813	3.2	1.77E-67
chr13	50620009	50626116	3.75	1.78E-67
chr21	13142574	13148282	3.53	3.52E-67
chr5	68803125	68809823	3.27	7.37E-67
chr3	45787337	45788796	4.96	8.23E-67
chr25	34139130	34140538	3.86	8.71E-67
chr6	42742516	42744123	3.87	1.75E-66
chr18	30306526	30308522	-3.4	2.28E-66
chr17	27257528	27260230	4.05	2.54E-66
chr23	28028152	28029129	4.53	3.23E-66
chr3	32563978	32565593	3.58	3.89E-66
chr13	9362502	9363950	5	1.29E-65
chr20	45890498	45891642	4.38	2.04E-65
chr23	42394861	42395876	5.05	2.07E-65
chr14	31167467	31168702	3.96	2.28E-65
chr23	29717376	29720359	4.26	4.07E-65
chr23	34338917	34339749	-4.07	4.35E-65
chr23	7399607	7400238	4.53	4.99E-65
chr7	54710285	54711399	4.45	4.99E-65
chr13	23386117	23387497	4.46	5.02E-65
chr13	14969472	14971239	5.4	5.90E-65
chr15	22339721	22341134	4.87	6.24E-65

chr14	35099003	35099851	-4.08	7.25E-65
chr4	21660562	21661425	5.46	1.14E-64
chr18	49333203	49334333	4.01	1.93E-64
chr21	15722763	15724465	4.15	2.12E-64
chr19	43115178	43117325	3.49	2.65E-64
chr3	45379747	45381314	4.73	2.65E-64
chr8	15216192	15217868	4.53	3.33E-64
chr5	16087406	16088187	-2.71	5.93E-64
chr2	58003304	58004633	4.52	7.90E-64
chr7	54969177	54970639	-4.42	8.69E-64
chr15	45604493	45606086	5.05	9.69E-64
chr25	29386900	29390158	4.21	1.22E-63
chr19	3169616	3171102	4.94	1.71E-63
chr21	23195741	23198210	-2.56	2.40E-63
chr22	18548380	18549624	4.04	2.55E-63
chr23	17769621	17770777	3.95	2.67E-63
chr1	10900556	10902149	3.94	2.87E-63
chr11	6615568	6616511	4.83	3.52E-63
chr5	3575729	3577127	3.39	4.24E-63
chr11	41980547	41980999	-2.87	4.38E-63
chr2	48229742	48231621	4.61	5.96E-63
chr20	27742308	27744090	3.82	6.80E-63
chr25	29051607	29055151	4.24	7.79E-63
chr7	33199461	33201555	5.01	9.11E-63
chr13	28459201	28460422	4.45	1.23E-62
chr15	23724919	23728065	3.58	1.40E-62
chr4	13796225	13797416	4.4	1.62E-62
chr12	3857640	3858686	4.06	3.07E-62
chr13	30506026	30507168	-3.26	6.64E-62
chr7	20457323	20458398	4.09	6.81E-62
chr16	22930225	22931717	3.78	1.06E-61
chr5	58555536	58556454	3.96	1.06E-61
chr19	14227486	14228166	4.51	1.07E-61
chr6	32154478	32155572	4.56	1.14E-61
chr25	31323075	31324226	-3.43	3.72E-61

chr1	6142464	6143755	4.68	1.01E-60
chr4	12809440	12810893	-3.11	1.48E-60
chr12	13727959	13730589	3.8	1.85E-60
chr17	16564989	16569490	-2.35	1.93E-60
chr17	7720713	7723149	3.96	2.45E-60
chr4	13978670	13980325	4.42	2.89E-60
chr25	11022336	11023270	4.08	3.23E-60
chr5	22069210	22073795	3.6	4.43E-60
chr23	44128363	44129830	4.63	5.06E-60
chr13	49334096	49335421	-3.29	5.37E-60
chr17	29410487	29412005	4.48	5.42E-60
chr17	18320859	18321796	-3.75	9.58E-60
chr3	61521359	61525148	3.79	9.62E-60
chr1	15984509	15986246	5.04	1.71E-59
chr22	8642996	8645043	4.08	3.18E-59
chr14	28567495	28568998	4.31	5.27E-59
chr17	52866257	52868257	-2.71	6.23E-59
chr16	12248708	12251363	4.35	8.30E-59
chr1	50922033	50924962	2.97	9.30E-59
chr7	9331466	9332945	3.06	2.57E-58
chr1	29893279	29894535	4.14	2.85E-58
chr6	6988098	6989681	4.01	3.89E-58
chr7	46696288	46698694	3.67	4.78E-58
chr23	6768398	6769269	4.34	6.47E-58
chr24	25225256	25226646	4.38	1.52E-57
chr18	45129299	45132034	3.02	2.02E-57
chr12	9013251	9014083	4.59	3.31E-57
chr10	43884238	43885657	-3.04	4.20E-57
chr15	46837615	46838845	4.52	4.41E-57
chr23	23310478	23311700	4.83	5.22E-57
chr3	23640986	23644464	3.14	5.43E-57
chr23	29723859	29725379	4.15	8.51E-57
chr1	7616516	7617635	4.04	8.56E-57
chr13	14975611	14977025	3.8	1.06E-56
chr16	36065145	36066724	3.77	1.21E-56

chr4	21923180	21925426	4.2	2.29E-56
chr17	42591554	42592670	3.88	2.74E-56
chr2	26987030	26988546	-3.02	2.98E-56
chr6	55341962	55343897	4.09	3.00E-56
chr4	8764174	8767098	4.12	4.46E-56
chr17	20283536	20285856	4.07	7.08E-56
chr5	27521858	27523332	3.86	7.50E-56
chr16	26671546	26672772	4.16	8.34E-56
chr7	56750880	56753648	2.91	9.60E-56
chr17	37898283	37900864	-2.49	1.28E-55
chr16	12252776	12253919	4.27	1.72E-55
chr14	34498803	34501206	3.25	2.51E-55
chr13	50616453	50618536	3.32	2.66E-55
chr24	41694761	41697401	3.71	3.43E-55
chr21	15913205	15914248	4.32	3.53E-55
chr13	4432558	4434355	4.23	3.61E-55
chr24	14514231	14515073	4.39	4.47E-55
chr12	26174554	26176255	4.52	5.20E-55
chr13	43419789	43421964	3.68	7.16E-55
chr6	12920757	12925333	3.44	7.87E-55
chr8	9381950	9383230	3.74	9.57E-55
chr13	28494182	28496079	3.24	1.54E-54
chr3	35459934	35462247	3.42	1.60E-54
chr1	46828066	46833272	2.96	2.47E-54
chr2	14675229	14676433	3.85	3.44E-54
chr7	47085558	47086697	4.39	3.65E-54
chr7	51437725	51440726	3.59	4.48E-54
chr25	14459652	14460344	4.22	4.66E-54
chr19	14039243	14039932	4.2	4.82E-54
chr6	6548385	6549642	-3	5.38E-54
chr19	28131462	28132282	3.31	7.13E-54
chr11	28608768	28609885	4.53	7.54E-54

Table 7.23: A list of 25 genes that are found proximal to *sox32* OE vs *sox17E* DARs. These genes were attained from GREAT by first converting *sox32* OE DARs to danRer7 genome build using LiftOver. TSS = transcriptional start site

Chromosome	Chromosome start	Chromosome end	Gene (distance to TSS)
chr1	609224	611499	cep97 (-6204)
chr1	600714	607144	rpl24 (-24)
chr1	576937	580681	zgc:110091 (+619)
chr1	52123	55117	pcid2 (-55)
chr1	69706	71745	grtp1a (-4239)
chr1	100881	102869	tfdp1b (-98)
chr1	109456	110407	tfdp1b (+7959)
chr1	131949	132742	cenpe (-2384)
chr1	133883	135939	cenpe (+181)
chr1	139297	140414	cenpe (+5126)
chr1	142013	142670	cenpe (+7612)
chr1	145908	146354	cenpe (+11401)
chr1	157804	158575	cenpe (+23460)
chr1	170334	171959	cenpe (+36417)
chr1	227939	228990	pros1 (+392)
chr1	251182	251542	pros1 (-22505)
chr1	291288	292007	si:ch73-244f7.4 (-1912)
chr1	298787	299103	si:ch73-244f7.4 (+5385)
chr1	315370	316497	si:ch73-244f7.4 (+22374)
chr1	348903	349540	zgc:92518 (-3729)
chr1	368610	369983	ercc5 (-320)
chr1	371474	371957	ercc5 (-2739)
chr1	375736	376756	ercc5 (-7269)
chr1	387964	389158	txnl4b (-122)
chr1	396855	397646	mrpl16 (-1834)

Table 7.24: A list of 25 genes that are found proximal to *sox17E* vs *sox32* OE DARs. These genes were attained from GREAT by first converting *sox17E* DARs to danRer7 genome build using LiftOver. TSS = transcriptional start site

Chromosome	Chromosome start	Chromosome end	Gene (distance to TSS)
chr1	184086	185396	gas6 (+28395)
chr1	195554	196011	gas6 (+17353)
chr1	1418254	1420344	atp1a1a.4 (-27262)
chr1	1868972	1871060	mbnl2 (-15945)
chr1	1898162	1898671	mbnl2 (+12456)
chr1	3293917	3294559	rca1b (+601430)
chr1	4187081	4187912	spry2 (+16578)

chr1	4218281	4219504	spry2 (+47974)
chr1	4231830	4233040	tuba8l2 (-59195)
chr1	4413081	4413913	fn1b (+18722)
chr1	4610387	4611211	nrp2a (-45035)
chr1	4604651	4606245	nrp2a (-39684)
chr1	4620717	4622270	nrp2a (-55730)
chr1	4906927	4907541	abcb6a (-91602)
chr1	5362689	5363184	klf7a (-5345)
chr1	5369207	5370188	klf7a (-12106)
chr1	6515759	6516762	mylz3 (-3998)
chr1	6669479	6670403	eng1b (+84540)
chr1	10687301	10690586	si:dkey-26i13.7 (+89008)
chr1	11003392	11004212	ttyh3b (+27674)
chr1	11025195	11026153	ttyh3b (+5802)
chr2	25047502	25048130	nceh1a (-138205)
chr1	12180222	12181519	pcdh18a (-24999)
chr1	12184876	12185192	pcdh18a (-29162)
chr1	12187631	12188850	pcdh18a (-32369)

Table 7.25: TF binding motifs found over-represented in *sox32* OE vs *sox17E* DARs. Motifs are ranked by statistical significance.

Rank	Motif	Name	P-value	log P-value	q-value (Benjamini)	# Target Sequences with Motif
1		CTCF(Zf)/CD4+-CTCF-ChIP-Seq(Barski_et_al.)/Homer	1e-219	-5.044e+02	0.0000	1414.0
2		Brn1(POU,Homeobox)/NPC-Brn1-ChIP-Seq(GSE35496)/Homer	1e-216	-4.991e+02	0.0000	2927.0
3		Tbx6(T-box)/ESC-Tbx6-ChIP-Seq(GSE93524)/Homer	1e-208	-4.791e+02	0.0000	5590.0
4		Oct6(POU,Homeobox)/NPC-Pou3f1-ChIP-Seq(GSE35496)/Homer	1e-191	-4.402e+02	0.0000	3523.0
5		BORIS(Zf)/K562-CTCF-ChIP-Seq(GSE32465)/Homer	1e-189	-4.371e+02	0.0000	1396.0
6		Tbx21(T-box)/GM12878-TBX21-ChIP-Seq(Encode)/Homer	1e-184	-4.256e+02	0.0000	6051.0
7		Eomes(T-box)/H9-Eomes-ChIP-Seq(GSE26097)/Homer	1e-178	-4.117e+02	0.0000	11541.0
8		Tbet(T-box)/CD8-Tbet-ChIP-Seq(GSE33802)/Homer	1e-171	-3.942e+02	0.0000	6262.0
9		Sox10(HMG)/SciaticNerve-Sox3-ChIP-Seq(GSE35132)/Homer	1e-149	-3.453e+02	0.0000	7253.0
10		Sox21(HMG)/ESC-SOX21-ChIP-Seq(GSE110505)/Homer	1e-147	-3.395e+02	0.0000	8382.0
11		Sox3(HMG)/NPC-Sox3-ChIP-Seq(GSE33059)/Homer	1e-140	-3.234e+02	0.0000	8078.0
12		Tbr1(T-box)/Cortex-Tbr1-ChIP-Seq(GSE71384)/Homer	1e-138	-3.200e+02	0.0000	8470.0
13		Sox17(HMG)/Endoderm-Sox17-ChIP-Seq(GSE61475)/Homer	1e-129	-2.972e+02	0.0000	3754.0
14		Sox6(HMG)/Myotubes-Sox6-ChIP-Seq(GSE32627)/Homer	1e-124	-2.856e+02	0.0000	8125.0
15		Oct11(POU,Homeobox)/NCIH1048-POU2F3-ChIP-seq(GSE115123)/Homer	1e-119	-2.745e+02	0.0000	2403.0
16		Sox2(HMG)/mES-Sox2-ChIP-Seq(GSE11431)/Homer	1e-118	-2.730e+02	0.0000	4128.0
17		Tbx5(T-box)/HL1-Tbx5.biotin-ChIP-Seq(GSE21529)/Homer	1e-107	-2.477e+02	0.0000	12587.0
18		Zic2(Zf)/ESC-Zic2-ChIP-Seq(SRP197560)/Homer	1e-99	-2.282e+02	0.0000	1601.0
19		Zic3(Zf)/mES-Zic3-ChIP-Seq(GSE37889)/Homer	1e-95	-2.194e+02	0.0000	1928.0
20		Sox15(HMG)/CPA-Sox15-ChIP-Seq(GSE62909)/Homer	1e-95	-2.194e+02	0.0000	5154.0
21		FOXA1(Forkhead)/MCF7-FOXA1-ChIP-Seq(GSE26831)/Homer	1e-93	-2.153e+02	0.0000	7260.0
22		Oct2(POU,Homeobox)/Bcell-Oct2-ChIP-Seq(GSE21512)/Homer	1e-92	-2.140e+02	0.0000	2109.0

23		FOXA1(Forkhead)/LNCAP-FOXA1-ChIP-Seq(GSE27824)/Homer	1e-91	-2.115e+02	0.0000	8573.0
24		Gata2(Zf)/K562-GATA2-ChIP-Seq(GSE18829)/Homer	1e-84	-1.939e+02	0.0000	3009.0
25		GATA3(Zf)/iTreg-Gata3-ChIP-Seq(GSE20898)/Homer	1e-81	-1.884e+02	0.0000	6850.0
26		Sox4(HMG)/proB-Sox4-ChIP-Seq(GSE50066)/Homer	1e-81	-1.883e+02	0.0000	3556.0
27		OCT4-SOX2-TCF-NANOG(POU,Homeobox,HMG)/mES-Oct4-ChIP-Seq(GSE11431)/Homer	1e-78	-1.809e+02	0.0000	1534.0
28		Gata6(Zf)/HUG1N-GATA6-ChIP-Seq(GSE51936)/Homer	1e-73	-1.694e+02	0.0000	4295.0
29		Sox9(HMG)/Limb-SOX9-ChIP-Seq(GSE73225)/Homer	1e-72	-1.661e+02	0.0000	3138.0
30		Gata4(Zf)/Heart-Gata4-ChIP-Seq(GSE35151)/Homer	1e-66	-1.541e+02	0.0000	4819.0
31		Sox7(HMG)/ESC-Sox7-ChIP-Seq(GSE133899)/Homer	1e-65	-1.507e+02	0.0000	1484.0
32		TRPS1(Zf)/MCF7-TRPS1-ChIP-Seq(GSE107013)/Homer	1e-64	-1.496e+02	0.0000	9167.0
33		Oct4(POU,Homeobox)/mES-Oct4-ChIP-Seq(GSE11431)/Homer	1e-64	-1.478e+02	0.0000	4093.0
34		Foxo3(Forkhead)/U2OS-Foxo3-ChIP-Seq(E-MTAB-2701)/Homer	1e-61	-1.413e+02	0.0000	6172.0
35		Gata1(Zf)/K562-GATA1-ChIP-Seq(GSE18829)/Homer	1e-60	-1.400e+02	0.0000	2531.0
36		FOXM1(Forkhead)/MCF7-FOXM1-ChIP-Seq(GSE72977)/Homer	1e-60	-1.383e+02	0.0000	6711.0
37		Foxf1(Forkhead)/Lung-Foxf1-ChIP-Seq(GSE77951)/Homer	1e-59	-1.369e+02	0.0000	7407.0
38		Foxa2(Forkhead)/Liver-Foxa2-ChIP-Seq(GSE25694)/Homer	1e-59	-1.359e+02	0.0000	5596.0
39		Fox:Ebox(Forkhead,bHLH)/Panc1-Foxa2-ChIP-Seq(GSE47459)/Homer	1e-57	-1.325e+02	0.0000	5897.0
40		FoxL2(Forkhead)/Ovary-FoxL2-ChIP-Seq(GSE60858)/Homer	1e-55	-1.270e+02	0.0000	6529.0
41		Oct4:Sox17(POU,Homeobox,HMG)/F9-Sox17-ChIP-Seq(GSE44553)/Homer	1e-53	-1.239e+02	0.0000	990.0
42		FOXK1(Forkhead)/HEK293-FOXK1-ChIP-Seq(GSE51673)/Homer	1e-49	-1.140e+02	0.0000	7638.0
43		FoxD3(forkhead)/ZebrafishEmbryo-Foxd3.biotin-ChIP-seq(GSE106676)/Homer	1e-43	-9.903e+01	0.0000	5713.0
44		Nkx6.1(Homeobox)/Islet-Nkx6.1-ChIP-Seq(GSE40975)/Homer	1e-41	-9.632e+01	0.0000	15111.0
45		FOXP1(Forkhead)/H9-FOXP1-ChIP-Seq(GSE31006)/Homer	1e-41	-9.518e+01	0.0000	3817.0
46		Zic(Zf)/Cerebellum-ZIC1.2-ChIP-Seq(GSE60731)/Homer	1e-38	-8.763e+01	0.0000	3187.0
47		Foxo1(Forkhead)/RAW-Foxo1-ChIP-Seq(Fan_et_al.)/Homer	1e-37	-8.598e+01	0.0000	9271.0
48		Lhx3(Homeobox)/Neuron-Lhx3-ChIP-Seq(GSE31456)/Homer	1e-35	-8.282e+01	0.0000	10493.0

49		KLF14(Zf)/HEK293-KLF14.GFP-ChIP-Seq(GSE58341)/Homer	1e-35	-8.089e+01	0.0000	3481.0
50		Nanog(Homeobox)/mES-Nanog-ChIP-Seq(GSE11724)/Homer	1e-32	-7.384e+01	0.0000	19077.0
51		Hoxa9(Homeobox)/ChickenMSG-Hoxa9.Flag-ChIP-Seq(GSE86088)/Homer	1e-31	-7.248e+01	0.0000	13208.0
52		Foxh1(Forkhead)/hESC-FOXH1-ChIP-Seq(GSE29422)/Homer	1e-31	-7.242e+01	0.0000	3451.0
53		Foxa3(Forkhead)/Liver-Foxa3-ChIP-Seq(GSE77670)/Homer	1e-30	-7.100e+01	0.0000	2466.0
54		Isl1(Homeobox)/Neuron-Isl1-ChIP-Seq(GSE31456)/Homer	1e-30	-7.087e+01	0.0000	9093.0
55		DLX2(Homeobox)/BasalGanglia-Dlx2-ChIP-seq(GSE124936)/Homer	1e-29	-6.891e+01	0.0000	9071.0
56		GSC(Homeobox)/FrogEmbryos-GSC-ChIP-Seq(DRA000576)/Homer	1e-28	-6.495e+01	0.0000	5396.0
57		Zfp281(Zf)/ES-Zfp281-ChIP-Seq(GSE81042)/Homer	1e-27	-6.318e+01	0.0000	271.0
58		FOXK2(Forkhead)/U2OS-FOXK2-ChIP-Seq(E-MTAB-2204)/Homer	1e-26	-6.060e+01	0.0000	4401.0
59		Sp5(Zf)/mES-Sp5.Flag-ChIP-Seq(GSE72989)/Homer	1e-25	-5.939e+01	0.0000	2148.0
60		En1(Homeobox)/SUM149-EN1-ChIP-Seq(GSE120957)/Homer	1e-24	-5.687e+01	0.0000	10925.0
61		Lhx2(Homeobox)/HFSC-Lhx2-ChIP-Seq(GSE48068)/Homer	1e-22	-5.254e+01	0.0000	6294.0
62		Atf4(bZIP)/MEF-Atf4-ChIP-Seq(GSE35681)/Homer	1e-22	-5.145e+01	0.0000	1700.0
63		RUNX-AML(Runt)/CD4+-PolII-ChIP-Seq(Barski_et_al.)/Homer	1e-22	-5.139e+01	0.0000	2855.0
64		ETS:RUNX(ETS,Runt)/Jurkat-RUNX1-ChIP-Seq(GSE17954)/Homer	1e-22	-5.088e+01	0.0000	198.0
65		Unknown-ESC-element(?)/mES-Nanog-ChIP-Seq(GSE11724)/Homer	1e-22	-5.071e+01	0.0000	2269.0
66		LHX9(Homeobox)/Hct116-LHX9.V5-ChIP-Seq(GSE116822)/Homer	1e-21	-5.033e+01	0.0000	8039.0
67		GATA(Zf),IR3/iTreg-Gata3-ChIP-Seq(GSE20898)/Homer	1e-21	-5.026e+01	0.0000	823.0
68		GLIS3(Zf)/Thyroid-Glis3.GFP-ChIP-Seq(GSE103297)/Homer	1e-21	-5.003e+01	0.0000	2277.0
69		Chop(bZIP)/MEF-Chop-ChIP-Seq(GSE35681)/Homer	1e-20	-4.745e+01	0.0000	1335.0
70		Hoxa10(Homeobox)/ChickenMSG-Hoxa10.Flag-ChIP-Seq(GSE86088)/Homer	1e-20	-4.670e+01	0.0000	2711.0
71		Zac1(Zf)/Neuro2A-Plagl1-ChIP-Seq(GSE75942)/Homer	1e-20	-4.659e+01	0.0000	5088.0
72		LEF1(HMG)/H1-LEF1-ChIP-Seq(GSE64758)/Homer	1e-18	-4.289e+01	0.0000	2532.0
73		DLX1(Homeobox)/BasalGanglia-Dlx1-ChIP-seq(GSE124936)/Homer	1e-18	-4.238e+01	0.0000	7938.0
74		RUNX2(Runt)/PCa-RUNX2-ChIP-Seq(GSE33889)/Homer	1e-17	-4.080e+01	0.0000	3357.0

75		Atf1(bZIP)/K562-ATF1-ChIP-Seq(GSE31477)/Homer	1e-17	-4.066e+01	0.0000	3458.0
76		Barx1(Homeobox)/Stomach-Barx1.3xFlag-ChIP-Seq(GSE69483)/Homer	1e-17	-4.062e+01	0.0000	2854.0
77		Hoxa11(Homeobox)/ChickenMSG-Hoxa11.Flag-ChIP-Seq(GSE86088)/Homer	1e-17	-4.016e+01	0.0000	12634.0
78		Otx2(Homeobox)/EpiLC-Otx2-ChIP-Seq(GSE56098)/Homer	1e-16	-3.890e+01	0.0000	3209.0
79		Maz(Zf)/HepG2-Maz-ChIP-Seq(GSE31477)/Homer	1e-16	-3.883e+01	0.0000	1880.0
80		ZNF467(Zf)/HEK293-ZNF467.GFP-ChIP-Seq(GSE58341)/Homer	1e-16	-3.868e+01	0.0000	1269.0
81		JunD(bZIP)/K562-JunD-ChIP-Seq/Homer	1e-16	-3.734e+01	0.0000	691.0
82		Nkx2.2(Homeobox)/NPC-Nkx2.2-ChIP-Seq(GSE61673)/Homer	1e-15	-3.572e+01	0.0000	7070.0
83		WT1(Zf)/Kidney-WT1-ChIP-Seq(GSE90016)/Homer	1e-15	-3.541e+01	0.0000	1257.0
84		c-Myc(bHLH)/mES-cMyc-ChIP-Seq(GSE11431)/Homer	1e-15	-3.511e+01	0.0000	1231.0
85		RUNX(Runt)/HPC7-Runx1-ChIP-Seq(GSE22178)/Homer	1e-14	-3.452e+01	0.0000	3070.0
86		Unknown(Homeobox)/Limb-p300-ChIP-Seq/Homer	1e-14	-3.417e+01	0.0000	3651.0
87		RUNX1(Runt)/Jurkat-RUNX1-ChIP-Seq(GSE29180)/Homer	1e-14	-3.411e+01	0.0000	3726.0
88		NFIL3(bZIP)/HepG2-NFIL3-ChIP-Seq(Encode)/Homer	1e-14	-3.250e+01	0.0000	4051.0
89		Cux2(Homeobox)/Liver-Cux2-ChIP-Seq(GSE35985)/Homer	1e-13	-3.217e+01	0.0000	2801.0
90		ETS1(ETS)/Jurkat-ETS1-ChIP-Seq(GSE17954)/Homer	1e-13	-3.147e+01	0.0000	2865.0
91		ZFX(Zf)/mES-Zfx-ChIP-Seq(GSE11431)/Homer	1e-13	-3.146e+01	0.0000	1890.0
92		Tcf3(HMG)/mES-Tcf3-ChIP-Seq(GSE11724)/Homer	1e-13	-3.136e+01	0.0000	899.0
93		Pitx1(Homeobox)/Chicken-Pitx1-ChIP-Seq(GSE38910)/Homer	1e-13	-3.134e+01	0.0000	16698.0
94		Sp2(Zf)/HEK293-Sp2.eGFP-ChIP-Seq(Encode)/Homer	1e-13	-3.084e+01	0.0000	2940.0
95		Brachyury(T-box)/Mesoendoderm-Brachyury-ChIP-exo(GSE54963)/Homer	1e-12	-2.921e+01	0.0000	1608.0
96		Nkx2.1(Homeobox)/LungAC-Nkx2.1-ChIP-Seq(GSE43252)/Homer	1e-12	-2.892e+01	0.0000	10163.0
97		MafA(bZIP)/Islet-MafA-ChIP-Seq(GSE30298)/Homer	1e-12	-2.875e+01	0.0000	2787.0
98		Lhx1(Homeobox)/EmbryoCarcinoma-Lhx1-ChIP-Seq(GSE70957)/Homer	1e-12	-2.851e+01	0.0000	6943.0
99		DLX5(Homeobox)/BasalGanglia-Dlx5-ChIP-seq(GSE124936)/Homer	1e-12	-2.837e+01	0.0000	4543.0
100		Hoxd10(Homeobox)/ChickenMSG-	1e-	-2.732e+01	0.0000	5134.0

		Hoxd10.Flag-ChIP-Seq(GSE86088)/Homer	1e-11			
101		E2F4(E2F)/K562-E2F4-ChIP-Seq(GSE31477)/Homer	1e-11	-2.715e+01	0.0000	1253.0
102		Hoxa13(Homeobox)/ChickenMSG-Hoxa13.Flag-ChIP-Seq(GSE86088)/Homer	1e-11	-2.684e+01	0.0000	13313.0
103		Dlx3(Homeobox)/Kerainocytes-Dlx3-ChIP-Seq(GSE89884)/Homer	1e-11	-2.654e+01	0.0000	4050.0
104		CRE(bZIP)/Promoter/Homer	1e-11	-2.561e+01	0.0000	1271.0
105		HNF6(Homeobox)/Liver-Hnf6-ChIP-Seq(ERP000394)/Homer	1e-11	-2.561e+01	0.0000	3344.0
106		Atf2(bZIP)/3T3L1-Atf2-ChIP-Seq(GSE56872)/Homer	1e-10	-2.513e+01	0.0000	1621.0
107		Nkx2.5(Homeobox)/HL1-Nkx2.5.biotin-ChIP-Seq(GSE21529)/Homer	1e-10	-2.491e+01	0.0000	8074.0
108		Eif4(ETS)/BMDM-Eif4-ChIP-Seq(GSE88699)/Homer	1e-10	-2.440e+01	0.0000	2584.0
109		PU.1-IRF(ETS:IRF)/Bcell-PU.1-ChIP-Seq(GSE21512)/Homer	1e-10	-2.396e+01	0.0000	3491.0
110		Hoxd11(Homeobox)/ChickenMSG-Hoxd11.Flag-ChIP-Seq(GSE86088)/Homer	1e-10	-2.368e+01	0.0000	12310.0
111		ERG(ETS)/VCaP-ERG-ChIP-Seq(GSE14097)/Homer	1e-10	-2.349e+01	0.0000	3732.0
112		Prop1(Homeobox)/GHFT1-PROP1.biotin-ChIP-Seq(GSE77302)/Homer	1e-9	-2.299e+01	0.0000	4106.0
113		Atf7(bZIP)/3T3L1-Atf7-ChIP-Seq(GSE56872)/Homer	1e-9	-2.288e+01	0.0000	2264.0
114		Gfi1b(Zf)/HPC7-Gfi1b-ChIP-Seq(GSE22178)/Homer	1e-9	-2.247e+01	0.0000	2675.0
115		Nkx3.1(Homeobox)/LNCaP-Nkx3.1-ChIP-Seq(GSE28264)/Homer	1e-9	-2.221e+01	0.0000	8876.0
116		ZNF711(Zf)/SHSY5Y-ZNF711-ChIP-Seq(GSE20673)/Homer	1e-9	-2.175e+01	0.0000	2750.0
117		ETV4(ETS)/HepG2-ETV4-ChIP-Seq(ENCODE)/Homer	1e-9	-2.131e+01	0.0000	2885.0
118		c-Jun-CRE(bZIP)/K562-cJun-ChIP-Seq(GSE31477)/Homer	1e-8	-2.063e+01	0.0000	1442.0
119		Bapx1(Homeobox)/VertebralCol-Bapx1-ChIP-Seq(GSE36672)/Homer	1e-8	-2.061e+01	0.0000	8062.0
120		KLF5(Zf)/LoVo-KLF5-ChIP-Seq(GSE49402)/Homer	1e-8	-2.049e+01	0.0000	2523.0
121		Nur77(NR)/K562-NR4A1-ChIP-Seq(GSE31363)/Homer	1e-8	-2.033e+01	0.0000	723.0
122		GATA3(Zf),DR8/iTreg-Gata3-ChIP-Seq(GSE20898)/Homer	1e-8	-1.975e+01	0.0000	370.0
123		Pax7(Paired,Homeobox),long/Myoblast-Pax7-ChIP-Seq(GSE25064)/Homer	1e-8	-1.969e+01	0.0000	235.0
124		E2F7(E2F)/Hela-E2F7-ChIP-Seq(GSE32673)/Homer	1e-8	-1.930e+01	0.0000	239.0
125		ETS(ETS)/Promoter/Homer	1e-7	-1.790e+01	0.0000	1062.0























126		E2F6(E2F)/Hela-E2F6-ChIP-Seq(GSE31477)/Homer	1e-7	-1.771e+01	0.0000	869.0
127		Hoxd12(Homeobox)/ChickenMSG-Hoxd12.Flag-ChIP-Seq(GSE86088)/Homer	1e-7	-1.768e+01	0.0000	9564.0
128		YY1(Zf)/Promoter/Homer	1e-7	-1.739e+01	0.0000	331.0
129		PU.1(ETS)/ThioMac-PU.1-ChIP-Seq(GSE21512)/Homer	1e-7	-1.719e+01	0.0000	1247.0
130		Hoxd13(Homeobox)/ChickenMSG-Hoxd13.Flag-ChIP-Seq(GSE86088)/Homer	1e-7	-1.710e+01	0.0000	8712.0
131		Rbpj1(?)/Panc1-Rbpj1-ChIP-Seq(GSE47459)/Homer	1e-7	-1.675e+01	0.0000	3433.0
132		ELF1(ETS)/Jurkat-ELF1-ChIP-Seq(SRA014231)/Homer	1e-7	-1.612e+01	0.0000	1418.0
133		TCFL2(HMG)/K562-TCF7L2-ChIP-Seq(GSE29196)/Homer	1e-6	-1.610e+01	0.0000	287.0
134		Etv2(ETS)/ES-ER71-ChIP-Seq(GSE59402)/Homer	1e-6	-1.604e+01	0.0000	2229.0
135		E2F3(E2F)/MEF-E2F3-ChIP-Seq(GSE71376)/Homer	1e-6	-1.600e+01	0.0000	1060.0
136		ZNF652/HepG2-ZNF652.Flag-ChIP-Seq(Encode)/Homer	1e-6	-1.595e+01	0.0000	1051.0
137		NFE2L2(bZIP)/HepG2-NFE2L2-ChIP-Seq(Encode)/Homer	1e-6	-1.573e+01	0.0000	146.0
138		HLF(bZIP)/HSC-HLF.Flag-ChIP-Seq(GSE69817)/Homer	1e-6	-1.514e+01	0.0000	4530.0
139		MYNN(Zf)/HEK293-MYNN.eGFP-ChIP-Seq(Encode)/Homer	1e-6	-1.471e+01	0.0000	1198.0
140		CRX(Homeobox)/Retina-Crx-ChIP-Seq(GSE20012)/Homer	1e-6	-1.470e+01	0.0000	9512.0
141		CEBP:AP1(bZIP)/ThioMac-CEBPb-ChIP-Seq(GSE21512)/Homer	1e-6	-1.447e+01	0.0000	3583.0
142		OCT:OCT(POU,Homeobox)/NPC-Brn1-ChIP-Seq(GSE35496)/Homer	1e-6	-1.440e+01	0.0000	86.0
143		Tcf7(HMG)/GM12878-TCF7-ChIP-Seq(Encode)/Homer	1e-5	-1.376e+01	0.0000	1240.0
144		RXR(NR),DR1/3T3L1-RXR-ChIP-Seq(GSE13511)/Homer	1e-5	-1.361e+01	0.0000	2192.0
145		KLF1(Zf)/HUDEP2-KLF1-CutnRun(GSE136251)/Homer	1e-5	-1.324e+01	0.0000	1897.0
146		GATA(Zf),IR4/iTreg-Gata3-ChIP-Seq(GSE20898)/Homer	1e-5	-1.272e+01	0.0000	359.0
147		GABPA(ETS)/Jurkat-GABPa-ChIP-Seq(GSE17954)/Homer	1e-5	-1.236e+01	0.0000	2370.0
148		Phox2a(Homeobox)/Neuron-Phox2a-ChIP-Seq(GSE31456)/Homer	1e-5	-1.223e+01	0.0000	2631.0
149		Tgif1(Homeobox)/mES-Tgif1-ChIP-Seq(GSE55404)/Homer	1e-5	-1.221e+01	0.0000	13776.0
150		PPARa(NR),DR1/Liver-Ppara-ChIP-Seq(GSE47954)/Homer	1e-5	-1.210e+01	0.0000	2450.0
151		PPARE(NR),DR1/3T3L1-Pparg-ChIP-Seq(GSE13511)/Homer	1e-5	-1.208e+01	0.0000	1994.0

152		Phox2b(Homeobox)/CLBGA-PHOX2B-ChIP-Seq(GSE90683)/Homer	1e-5	-1.172e+01	0.0000	1528.0
153		CREB5(bZIP)/LNCaP-CREB5.V5-ChIP-Seq(GSE137775)/Homer	1e-5	-1.171e+01	0.0000	1958.0
154		Fli1(ETS)/CD8-FLI-ChIP-Seq(GSE20898)/Homer	1e-5	-1.167e+01	0.0000	3341.0
155		RAR:RXR(NR),DR5/ES-RAR-ChIP-Seq(GSE56893)/Homer	1e-4	-1.132e+01	0.0000	74.0
156		Hoxb4(Homeobox)/ES-Hoxb4-ChIP-Seq(GSE34014)/Homer	1e-4	-1.122e+01	0.0000	947.0
157		SpiB(ETS)/OCILY3-SPIB-ChIP-Seq(GSE56857)/Homer	1e-4	-1.116e+01	0.0000	671.0
158		EHF(ETS)/LoVo-EHF-ChIP-Seq(GSE49402)/Homer	1e-4	-1.072e+01	0.0001	3289.0
159		ELF5(ETS)/T47D-ELF5-ChIP-Seq(GSE30407)/Homer	1e-4	-1.019e+01	0.0001	1849.0
160		ETV1(ETS)/GIST48-ETV1-ChIP-Seq(GSE22441)/Homer	1e-4	-9.973e+00	0.0001	3485.0
161		EWS:FLI1-fusion(ETS)/SK_N_MC-EWS:FLI1-ChIP-Seq(SRA014231)/Homer	1e-4	-9.805e+00	0.0002	1524.0
162		KLF6(Zf)/PDAC-KLF6-ChIP-Seq(GSE64557)/Homer	1e-4	-9.395e+00	0.0002	2127.0
163		SCRT1(Zf)/HEK293-SCRT1.eGFP-ChIP-Seq(Encode)/Homer	1e-3	-8.902e+00	0.0004	1059.0
164		Duxbl(Homeobox)/NIH3T3-Duxbl.HA-ChIP-Seq(GSE119782)/Homer	1e-3	-8.884e+00	0.0004	522.0
165		Elk1(ETS)/Hela-Elk1-ChIP-Seq(GSE31477)/Homer	1e-3	-8.689e+00	0.0004	1750.0
166		FOXA1:AR(Forkhead,NR)/LNCAP-AR-ChIP-Seq(GSE27824)/Homer	1e-3	-8.648e+00	0.0005	385.0
167		ZNF528(Zf)/HEK293-ZNF528.GFP-ChIP-Seq(GSE58341)/Homer	1e-3	-8.361e+00	0.0006	15.0
168		EKLF(Zf)/Erythrocyte-Klf1-ChIP-Seq(GSE20478)/Homer	1e-3	-8.124e+00	0.0008	424.0
169		EWS:ERG-fusion(ETS)/CADO_ES1-EWS:ERG-ChIP-Seq(SRA014231)/Homer	1e-3	-7.910e+00	0.0010	1902.0
170		AP-2alpha(AP2)/Hela-AP2alpha-ChIP-Seq(GSE31477)/Homer	1e-3	-7.677e+00	0.0012	820.0
171		Tbox:Smad(T-box,MAD)/ESCd5-Smad2_3-ChIP-Seq(GSE29422)/Homer	1e-3	-7.625e+00	0.0013	605.0
172		Bach1(bZIP)/K562-Bach1-ChIP-Seq(GSE31477)/Homer	1e-3	-7.608e+00	0.0013	162.0
173		MITF(bHLH)/MastCells-MITF-ChIP-Seq(GSE48085)/Homer	1e-3	-7.524e+00	0.0014	2866.0
174		Npas4(bHLH)/Neuron-Npas4-ChIP-Seq(GSE127793)/Homer	1e-3	-7.510e+00	0.0014	3661.0
175		Meis1(Homeobox)/MastCells-Meis1-ChIP-Seq(GSE48085)/Homer	1e-3	-7.402e+00	0.0015	6486.0
176		THRb(NR)/HepG2-THRb.Flag-ChIP-Seq(Encode)/Homer	1e-3	-7.146e+00	0.0020	1127.0
177		Elk4(ETS)/Hela-Elk4-ChIP-Seq(GSE31477)/Homer	1e-3	-7.079e+00	0.0021	1916.0

178		MNT(bHLH)/HepG2-MNT-ChIP-Seq(Encode)/Homer	1e-3	-7.008e+00	0.0022	3662.0
179		CTCF-SatelliteElement(Zf?)/CD4+-CTCF-ChIP-Seq(Barski_et_al.)/Homer	1e-3	-7.008e+00	0.0022	58.0
180		PR(NR)/T47D-PR-ChIP-Seq(GSE31130)/Homer	1e-2	-6.692e+00	0.0030	7708.0
181		NF-E2(bZIP)/K562-NFE2-ChIP-Seq(GSE31477)/Homer	1e-2	-6.688e+00	0.0030	137.0
182		GFY(?)/Promoter/Homer	1e-2	-6.664e+00	0.0031	130.0
183		BMAL1(bHLH)/Liver-Bmal1-ChIP-Seq(GSE39860)/Homer	1e-2	-6.601e+00	0.0033	6784.0
184		HOXA1(Homeobox)/mES-Hoxa1-ChIP-Seq(SRP084292)/Homer	1e-2	-6.426e+00	0.0039	1353.0
185		ELF3(ETS)/PDAC-ELF3-ChIP-Seq(GSE64557)/Homer	1e-2	-6.351e+00	0.0041	1798.0
186		NPAS(bHLH)/Liver-NPAS-ChIP-Seq(GSE39860)/Homer	1e-2	-6.243e+00	0.0046	6220.0
187		KLF3(Zf)/MEF-Klf3-ChIP-Seq(GSE44748)/Homer	1e-2	-5.928e+00	0.0063	1027.0
188		Hnf6b(Homeobox)/LNCaP-Hnf6b-ChIP-Seq(GSE106305)/Homer	1e-2	-5.829e+00	0.0069	5008.0
189		Ets1-distal(ETS)/CD4+-PolII-ChIP-Seq(Barski_et_al.)/Homer	1e-2	-5.751e+00	0.0074	827.0
190		TEAD1(TEAD)/HepG2-TEAD1-ChIP-Seq(Encode)/Homer	1e-2	-5.675e+00	0.0079	1973.0
191		TEAD4(TEA)/Tropoblast-Tead4-ChIP-Seq(GSE37350)/Homer	1e-2	-5.598e+00	0.0085	1725.0
192		Max(bHLH)/K562-Max-ChIP-Seq(GSE31477)/Homer	1e-2	-5.548e+00	0.0089	2064.0
193		TATA-Box(TBP)/Promoter/Homer	1e-2	-5.300e+00	0.0114	6074.0
194		AP-2gamma(AP2)/MCF7-TFAP2C-ChIP-Seq(GSE21234)/Homer	1e-2	-5.262e+00	0.0118	1142.0
195		ZNF675(Zf)/HEK293-ZNF675.GFP-ChIP-Seq(GSE58341)/Homer	1e-2	-5.231e+00	0.0121	383.0
196		NRF1(NRF)/MCF7-NRF1-ChIP-Seq(Unpublished)/Homer	1e-2	-5.141e+00	0.0131	242.0
197		Pax8(Paired,Homeobox)/Thyroid-Pax8-ChIP-Seq(GSE26938)/Homer	1e-2	-5.081e+00	0.0139	478.0
198		n-Myc(bHLH)/mES-nMyc-ChIP-Seq(GSE11431)/Homer	1e-2	-5.065e+00	0.0140	1827.0
199		E2F1(E2F)/Hela-E2F1-ChIP-Seq(GSE22478)/Homer	1e-2	-5.041e+00	0.0143	417.0
200		Erra(NR)/HepG2-Erra-ChIP-Seq(GSE31477)/Homer	1e-2	-4.927e+00	0.0160	4998.0
201		BMYB(HTH)/Hela-BMYB-ChIP-Seq(GSE27030)/Homer	1e-2	-4.880e+00	0.0166	7754.0
202		Bach2(bZIP)/OCIly7-Bach2-ChIP-Seq(GSE44420)/Homer	1e-2	-4.834e+00	0.0173	457.0
203		NF1:FOXA1(CTF,Forkhead)/LNCAP-	1e-2	-4.831e+00	0.0173	224.0

ACTGTTTATTGGCA
CAGGCTTGGCA

Table 7.26: TF binding motifs found over-represented in *sox17E* vs *sox32* OE DARs. Motifs are ranked by statistical significance.

Rank	Motif	Name	P-value	log P-value	q-value (Benjamini)	# Target Sequences with Motif	% Ta Se wi
1		Sox21(HMG)/ESC-SOX21-ChIP-Seq(GSE110505)/Homer	1e-76	-1.766e+02	0.0000	1855.0	27
2		TR4(NR),DR1/Hela-TR4-ChIP-Seq(GSE24685)/Homer	1e-62	-1.441e+02	0.0000	239.0	3.5
3		Sox3(HMG)/NPC-Sox3-ChIP-Seq(GSE33059)/Homer	1e-61	-1.409e+02	0.0000	1652.0	24
4		Sox10(HMG)/SciaticNerve-Sox3-ChIP-Seq(GSE35132)/Homer	1e-57	-1.323e+02	0.0000	1526.0	22
5		EAR2(NR)/K562-NR2F6-ChIP-Seq(Encode)/Homer	1e-51	-1.196e+02	0.0000	1120.0	16
6		COUP-TFII(NR)/Artia-Nr2f2-ChIP-Seq(GSE46497)/Homer	1e-46	-1.072e+02	0.0000	1423.0	20
7		TEAD3(TEA)/HepG2-TEAD3-ChIP-Seq(Encode)/Homer	1e-45	-1.057e+02	0.0000	823.0	12
8		Sox2(HMG)/mES-Sox2-ChIP-Seq(GSE11431)/Homer	1e-42	-9.726e+01	0.0000	846.0	12
9		TEAD1(TEAD)/HepG2-TEAD1-ChIP-Seq(Encode)/Homer	1e-41	-9.564e+01	0.0000	679.0	9.5
10		Maz(Zf)/HepG2-Maz-ChIP-Seq(GSE31477)/Homer	1e-40	-9.397e+01	0.0000	738.0	10
11		COUP-TFII(NR)/K562-NR2F1-ChIP-Seq(Encode)/Homer	1e-40	-9.219e+01	0.0000	1143.0	16
12		Sox6(HMG)/Myotubes-Sox6-ChIP-Seq(GSE32627)/Homer	1e-36	-8.491e+01	0.0000	1519.0	22
13		Sox17(HMG)/Endoderm-Sox17-ChIP-Seq(GSE61475)/Homer	1e-34	-8.037e+01	0.0000	713.0	10
14		TEAD4(TEA)/Tropoblast-Tead4-ChIP-Seq(GSE37350)/Homer	1e-32	-7.584e+01	0.0000	596.0	8.5
15		Sox15(HMG)/CPA-Sox15-ChIP-Seq(GSE62909)/Homer	1e-32	-7.516e+01	0.0000	1019.0	14
16		Sox4(HMG)/proB-Sox4-ChIP-Seq(GSE50066)/Homer	1e-32	-7.403e+01	0.0000	772.0	11
17		PPARE(NR),DR1/3T3L1-Pparg-ChIP-Seq(GSE13511)/Homer	1e-31	-7.366e+01	0.0000	663.0	9.5
18		TEAD(TEA)/Fibroblast-PU.1-ChIP-Seq(Unpublished)/Homer	1e-30	-7.021e+01	0.0000	450.0	6.5
19		PPARa(NR),DR1/Liver-Ppara-ChIP-Seq(GSE47954)/Homer	1e-27	-6.335e+01	0.0000	717.0	10
20		RXR(NR),DR1/3T3L1-RXR-ChIP-Seq(GSE13511)/Homer	1e-27	-6.229e+01	0.0000	710.0	10
21		KLF10(Zf)/HEK293-KLF10.GFP-ChIP-Seq(GSE58341)/Homer	1e-25	-5.794e+01	0.0000	1245.0	18
22		Tgif1(Homeobox)/mES-Tgif1-ChIP-Seq(GSE55404)/Homer	1e-23	-5.500e+01	0.0000	2888.0	42

23		Erra(NR)/HepG2-Erra-ChIP-Seq(GSE31477)/Homer	1e-23	-5.391e+01	0.0000	1334.0	19.0
24		TEAD2(TEA)/Py2T-Tead2-ChIP-Seq(GSE55709)/Homer	1e-22	-5.098e+01	0.0000	361.0	5.0
25		Sox7(HMG)/ESC-Sox7-ChIP-Seq(GSE133899)/Homer	1e-20	-4.741e+01	0.0000	306.0	4.0
26		Nkx6.1(Homeobox)/Islet-Nkx6.1-ChIP-Seq(GSE40975)/Homer	1e-20	-4.735e+01	0.0000	2606.0	38.0
27		Zfp281(Zf)/ES-Zfp281-ChIP-Seq(GSE81042)/Homer	1e-20	-4.681e+01	0.0000	124.0	1.0
28		ZNF467(Zf)/HEK293-ZNF467.GFP-ChIP-Seq(GSE58341)/Homer	1e-20	-4.646e+01	0.0000	444.0	6.0
29		Sox9(HMG)/Limb-SOX9-ChIP-Seq(GSE73225)/Homer	1e-19	-4.392e+01	0.0000	663.0	9.0
30		Nur77(NR)/K562-NR4A1-ChIP-Seq(GSE31363)/Homer	1e-18	-4.311e+01	0.0000	210.0	3.0
31		Hoxc9(Homeobox)/Ainv15-Hoxc9-ChIP-Seq(GSE21812)/Homer	1e-18	-4.299e+01	0.0000	583.0	8.0
32		KLF14(Zf)/HEK293-KLF14.GFP-ChIP-Seq(GSE58341)/Homer	1e-18	-4.166e+01	0.0000	952.0	13.0
33		Atoh1(bHLH)/Cerebellum-Atoh1-ChIP-Seq(GSE22111)/Homer	1e-17	-4.042e+01	0.0000	903.0	13.0
34		HNF4a(NR),DR1/HepG2-HNF4a-ChIP-Seq(GSE25021)/Homer	1e-17	-4.032e+01	0.0000	335.0	4.0
35		BHLHA15(bHLH)/NIH3T3-BHLHB8.HA-ChIP-Seq(GSE119782)/Homer	1e-17	-3.942e+01	0.0000	1155.0	16.0
36		Olig2(bHLH)/Neuron-Olig2-ChIP-Seq(GSE30882)/Homer	1e-16	-3.881e+01	0.0000	1576.0	23.0
37		NeuroG2(bHLH)/Fibroblast-NeuroG2-ChIP-Seq(GSE75910)/Homer	1e-14	-3.425e+01	0.0000	1268.0	18.0
38		Fra1(bZIP)/BT549-Fra1-ChIP-Seq(GSE46166)/Homer	1e-14	-3.369e+01	0.0000	440.0	6.0
39		Meis1(Homeobox)/MastCells-Meis1-ChIP-Seq(GSE48085)/Homer	1e-14	-3.361e+01	0.0000	1553.0	22.0
40		LHX9(Homeobox)/Hct116-LHX9.V5-ChIP-Seq(GSE116822)/Homer	1e-14	-3.360e+01	0.0000	1437.0	21.0
41		FOXM1(Forkhead)/MCF7-FOXM1-ChIP-Seq(GSE72977)/Homer	1e-14	-3.311e+01	0.0000	1124.0	16.0
42		Zac1(Zf)/Neuro2A-Plagl1-ChIP-Seq(GSE75942)/Homer	1e-14	-3.279e+01	0.0000	1317.0	19.0
43		Six1(Homeobox)/Myoblast-Six1-ChIP-Chip(GSE20150)/Homer	1e-14	-3.255e+01	0.0000	264.0	3.0
44		Lhx2(Homeobox)/HFSC-Lhx2-ChIP-Seq(GSE48068)/Homer	1e-13	-3.214e+01	0.0000	1117.0	16.0
45		Fox:Ebox(Forkhead,bHLH)/Panc1-Foxa2-ChIP-Seq(GSE47459)/Homer	1e-13	-3.202e+01	0.0000	1137.0	16.0
46		THRb(NR)/Liver-NR1A2-ChIP-Seq(GSE52613)/Homer	1e-13	-3.151e+01	0.0000	2580.0	37.0
47		Foxa2(Forkhead)/Liver-Foxa2-ChIP-Seq(GSE25694)/Homer	1e-13	-3.130e+01	0.0000	934.0	13.0
48		Atf3(bZIP)/GBM-ATF3-ChIP-Seq(GSE33912)/Homer	1e-13	-3.064e+01	0.0000	518.0	7.0

49		Isl1(Homeobox)/Neuron-Isl1-ChIP-Seq(GSE31456)/Homer	1e-13	-3.035e+01	0.0000	1626.0	23
50		Hoxd11(Homeobox)/ChickenMSG-Hoxd11.Flag-ChIP-Seq(GSE86088)/Homer	1e-12	-2.942e+01	0.0000	2225.0	32
51		Hoxa9(Homeobox)/ChickenMSG-Hoxa9.Flag-ChIP-Seq(GSE86088)/Homer	1e-12	-2.934e+01	0.0000	2390.0	35
52		Fos(bZIP)/TSC-Fos-ChIP-Seq(GSE110950)/Homer	1e-12	-2.925e+01	0.0000	478.0	7.0
53		TCF4(bHLH)/SHSY5Y-TCF4-ChIP-Seq(GSE96915)/Homer	1e-12	-2.908e+01	0.0000	1227.0	17
54		Hoxa11(Homeobox)/ChickenMSG-Hoxa11.Flag-ChIP-Seq(GSE86088)/Homer	1e-12	-2.892e+01	0.0000	2206.0	32
55		Myf5(bHLH)/GM-Myf5-ChIP-Seq(GSE24852)/Homer	1e-12	-2.863e+01	0.0000	632.0	9.3
56		Hoxd10(Homeobox)/ChickenMSG-Hoxd10.Flag-ChIP-Seq(GSE86088)/Homer	1e-12	-2.824e+01	0.0000	993.0	14
57		Ascl1(bHLH)/NeuralTubes-Ascl1-ChIP-Seq(GSE55840)/Homer	1e-11	-2.761e+01	0.0000	1030.0	15
58		Esrrb(NR)/mES-Esrrb-ChIP-Seq(GSE11431)/Homer	1e-11	-2.757e+01	0.0000	456.0	6.0
59		Twist2(bHLH)/Myoblast-Twist2.Ty1-ChIP-Seq(GSE127998)/Homer	1e-11	-2.721e+01	0.0000	1444.0	21
60		Fra2(bZIP)/Striatum-Fra2-ChIP-Seq(GSE43429)/Homer	1e-11	-2.691e+01	0.0000	362.0	5.3
61		WT1(Zf)/Kidney-WT1-ChIP-Seq(GSE90016)/Homer	1e-11	-2.667e+01	0.0000	420.0	6.3
62		BATF(bZIP)/Th17-BATF-ChIP-Seq(GSE39756)/Homer	1e-11	-2.638e+01	0.0000	472.0	6.9
63		NeuroD1(bHLH)/Islet-NeuroD1-ChIP-Seq(GSE30298)/Homer	1e-11	-2.636e+01	0.0000	610.0	8.9
64		ZFX(Zf)/mES-Zfx-ChIP-Seq(GSE11431)/Homer	1e-11	-2.627e+01	0.0000	518.0	7.0
65		Pdx1(Homeobox)/Islet-Pdx1-ChIP-Seq(SRA008281)/Homer	1e-11	-2.616e+01	0.0000	1080.0	15
66		En1(Homeobox)/SUM149-EN1-ChIP-Seq(GSE120957)/Homer	1e-11	-2.588e+01	0.0000	1887.0	27
67		Atf1(bZIP)/K562-ATF1-ChIP-Seq(GSE31477)/Homer	1e-11	-2.585e+01	0.0000	697.0	10
68		FOXA1(Forkhead)/LNCAP-FOXA1-ChIP-Seq(GSE27824)/Homer	1e-11	-2.574e+01	0.0000	1359.0	19
69		Atf7(bZIP)/3T3L1-Atf7-ChIP-Seq(GSE56872)/Homer	1e-10	-2.495e+01	0.0000	487.0	7.3
70		Hoxa13(Homeobox)/ChickenMSG-Hoxa13.Flag-ChIP-Seq(GSE86088)/Homer	1e-10	-2.490e+01	0.0000	2302.0	33
71		RAR:RXR(NR),DR5/ES-RAR-ChIP-Seq(GSE56893)/Homer	1e-10	-2.428e+01	0.0000	35.0	0.5
72		AP-2alpha(AP2)/Hela-AP2alpha-ChIP-Seq(GSE31477)/Homer	1e-10	-2.414e+01	0.0000	288.0	4.2
73		Atf2(bZIP)/3T3L1-Atf2-ChIP-Seq(GSE56872)/Homer	1e-10	-2.369e+01	0.0000	364.0	5.3
74		MyoD(bHLH)/Myotube-MyoD-ChIP-Seq(GSE21614)/Homer	1e-10	-2.359e+01	0.0000	616.0	9.0

75		FOXA1(Forkhead)/MCF7-FOXA1-ChIP-Seq(GSE26831)/Homer	1e-10	-2.341e+01	0.0000	1121.0	16
76		JunB(bZIP)/DendriticCells-Junb-ChIP-Seq(GSE36099)/Homer	1e-10	-2.313e+01	0.0000	398.0	5.8
77		Lhx3(Homeobox)/Neuron-Lhx3-ChIP-Seq(GSE31456)/Homer	1e-9	-2.294e+01	0.0000	1706.0	25
78		GATA3(Zf)/iTreg-Gata3-ChIP-Seq(GSE20898)/Homer	1e-9	-2.277e+01	0.0000	1107.0	16
79		PBX2(Homeobox)/K562-PBX2-ChIP-Seq(Encode)/Homer	1e-9	-2.260e+01	0.0000	959.0	14
80		FosI2(bZIP)/3T3L1-FosI2-ChIP-Seq(GSE56872)/Homer	1e-9	-2.248e+01	0.0000	228.0	3.3
81		Tgif2(Homeobox)/mES-Tgif2-ChIP-Seq(GSE55404)/Homer	1e-9	-2.230e+01	0.0000	2880.0	42
82		AP-1(bZIP)/ThioMac-PU.1-ChIP-Seq(GSE21512)/Homer	1e-9	-2.217e+01	0.0000	555.0	8.3
83		FOXK1(Forkhead)/HEK293-FOXK1-ChIP-Seq(GSE51673)/Homer	1e-9	-2.193e+01	0.0000	1274.0	18
84		FoxL2(Forkhead)/Ovary-FoxL2-ChIP-Seq(GSE60858)/Homer	1e-9	-2.162e+01	0.0000	1030.0	15
85		Foxo1(Forkhead)/RAW-Foxo1-ChIP-Seq(Fan_et_al.)/Homer	1e-9	-2.154e+01	0.0000	1733.0	25
86		ZNF711(Zf)/SHSY5Y-ZNF711-ChIP-Seq(GSE20673)/Homer	1e-9	-2.090e+01	0.0000	726.0	10
87		CDX4(Homeobox)/ZebrafishEmbryos-Cdx4.Myc-ChIP-Seq(GSE48254)/Homer	1e-8	-2.049e+01	0.0000	927.0	13
88		c-Jun-CRE(bZIP)/K562-cJun-ChIP-Seq(GSE31477)/Homer	1e-8	-2.044e+01	0.0000	313.0	4.5
89		Hoxd13(Homeobox)/ChickenMSG-Hoxd13.Flag-ChIP-Seq(GSE86088)/Homer	1e-8	-2.030e+01	0.0000	1499.0	21
90		MafA(bZIP)/Islet-MafA-ChIP-Seq(GSE30298)/Homer	1e-8	-1.997e+01	0.0000	642.0	9.4
91		AP-2gamma(AP2)/MCF7-TFAP2C-ChIP-Seq(GSE21234)/Homer	1e-8	-1.960e+01	0.0000	374.0	5.4
92		Lhx1(Homeobox)/EmbryoCarcinoma-Lhx1-ChIP-Seq(GSE70957)/Homer	1e-8	-1.943e+01	0.0000	1142.0	16
93		TRPS1(Zf)/MCF7-TRPS1-ChIP-Seq(GSE107013)/Homer	1e-8	-1.931e+01	0.0000	1489.0	21
94		VDR(NR),DR3/GM10855-VDR+vitD-ChIP-Seq(GSE22484)/Homer	1e-7	-1.832e+01	0.0000	141.0	2.0
95		BMAL1(bHLH)/Liver-Bmal1-ChIP-Seq(GSE39860)/Homer	1e-7	-1.789e+01	0.0000	1493.0	21
96		PU.1-IRF(ETS:IRF)/Bcell-PU.1-ChIP-Seq(GSE21512)/Homer	1e-7	-1.783e+01	0.0000	817.0	11
97		RARa(NR)/K562-RARa-ChIP-Seq(Encode)/Homer	1e-7	-1.776e+01	0.0000	1983.0	29
98		Rfx2(HTH)/LoVo-RFX2-ChIP-Seq(GSE49402)/Homer	1e-7	-1.764e+01	0.0000	92.0	1.3
99		Foxo3(Forkhead)/U2OS-Foxo3-ChIP-Seq(E-MTAB-2701)/Homer	1e-7	-1.756e+01	0.0000	983.0	14
100		DLX1(Homeobox)/BasalGanglia-Dlx1-	1e-7	-1.749e+01	0.0000	1336.0	19

		ChIP-seq(GSE124936)/Homer					
101		CREB5(bZIP)/LNCaP-CREB5.V5-ChIP-Seq(GSE137775)/Homer	1e-7	-1.730e+01	0.0000	386.0	5.0
102		Foxf1(Forkhead)/Lung-Foxf1-ChIP-Seq(GSE77951)/Homer	1e-7	-1.716e+01	0.0000	1153.0	16.0
103		Gata1(Zf)/K562-GATA1-ChIP-Seq(GSE18829)/Homer	1e-7	-1.714e+01	0.0000	421.0	6.0
104		Hoxb4(Homeobox)/ES-Hoxb4-ChIP-Seq(GSE34014)/Homer	1e-7	-1.714e+01	0.0000	218.0	3.0
105		HOXB13(Homeobox)/ProstateTumor-HOXB13-ChIP-Seq(GSE56288)/Homer	1e-7	-1.667e+01	0.0000	990.0	14.0
106		Jun-AP1(bZIP)/K562-cJun-ChIP-Seq(GSE31477)/Homer	1e-7	-1.653e+01	0.0000	154.0	2.0
107		JunD(bZIP)/K562-JunD-ChIP-Seq/Homer	1e-7	-1.629e+01	0.0000	152.0	2.0
108		Foxa3(Forkhead)/Liver-Foxa3-ChIP-Seq(GSE77670)/Homer	1e-6	-1.596e+01	0.0000	398.0	5.0
109		PRDM1(Zf)/Hela-PRDM1-ChIP-Seq(GSE31477)/Homer	1e-6	-1.586e+01	0.0000	484.0	7.0
110		Tcf21(bHLH)/ArterySmoothMuscle-Tcf21-ChIP-Seq(GSE61369)/Homer	1e-6	-1.576e+01	0.0000	758.0	11.0
111		MyoG(bHLH)/C2C12-MyoG-ChIP-Seq(GSE36024)/Homer	1e-6	-1.564e+01	0.0000	832.0	12.0
112		FOXP1(Forkhead)/H9-FOXP1-ChIP-Seq(GSE31006)/Homer	1e-6	-1.561e+01	0.0000	601.0	8.0
113		Gata2(Zf)/K562-GATA2-ChIP-Seq(GSE18829)/Homer	1e-6	-1.547e+01	0.0000	478.0	7.0
114		Pbx3(Homeobox)/GM12878-PBX3-ChIP-Seq(GSE32465)/Homer	1e-6	-1.530e+01	0.0000	262.0	3.0
115		RFX(HTH)/K562-RFX3-ChIP-Seq(SRA012198)/Homer	1e-6	-1.502e+01	0.0000	89.0	1.0
116		ERRg(NR)/Kidney-ESRRG-ChIP-Seq(GSE104905)/Homer	1e-6	-1.498e+01	0.0000	539.0	7.0
117		Six4(Homeobox)/MCF7-SIX4-ChIP-Seq(Encode)/Homer	1e-6	-1.479e+01	0.0000	61.0	0.0
118		LEF1(HMG)/H1-LEF1-ChIP-Seq(GSE64758)/Homer	1e-6	-1.479e+01	0.0000	483.0	7.0
119		Rfx1(HTH)/NPC-H3K4me1-ChIP-Seq(GSE16256)/Homer	1e-6	-1.471e+01	0.0000	140.0	2.0
120		NPAS(bHLH)/Liver-NPAS-ChIP-Seq(GSE39860)/Homer	1e-6	-1.443e+01	0.0000	1356.0	19.0
121		OCT:OCT(POU,Homeobox)/NPC-Brn1-ChIP-Seq(GSE35496)/Homer	1e-6	-1.408e+01	0.0000	30.0	0.0
122		FOXK2(Forkhead)/U2OS-FOXK2-ChIP-Seq(E-MTAB-2204)/Homer	1e-5	-1.381e+01	0.0000	731.0	10.0
123		LXRE(NR),DR4/RAW-LXRb.biotin-ChIP-Seq(GSE21512)/Homer	1e-5	-1.355e+01	0.0000	50.0	0.0
124		Gata6(Zf)/HUG1N-GATA6-ChIP-Seq(GSE51936)/Homer	1e-5	-1.351e+01	0.0000	674.0	9.0
125		Mef2b(MADS)/HEK293-Mef2b.V5-ChIP-Seq(GSE67450)/Homer	1e-5	-1.349e+01	0.0000	646.0	9.0

126		Tcf12(bHLH)/GM12878-Tcf12-ChIP-Seq(GSE32465)/Homer	1e-5	-1.283e+01	0.0000	735.0	10
127		Tbx20(T-box)/Heart-Tbx20-ChIP-Seq(GSE29636)/Homer	1e-5	-1.278e+01	0.0000	199.0	2.9
128		HOXA1(Homeobox)/mES-Hoxa1-ChIP-Seq(SRP084292)/Homer	1e-5	-1.231e+01	0.0000	301.0	4.4
129		Egr1(Zf)/K562-Egr1-ChIP-Seq(GSE32465)/Homer	1e-5	-1.225e+01	0.0000	518.0	7.6
130		Mef2d(MADS)/Retina-Mef2d-ChIP-Seq(GSE61391)/Homer	1e-5	-1.218e+01	0.0000	177.0	2.6
131		DLX2(Homeobox)/BasalGanglia-Dlx2-ChIP-seq(GSE124936)/Homer	1e-5	-1.198e+01	0.0000	1464.0	21
132		PAX5(Paired,Homeobox)/GM12878-PAX5-ChIP-Seq(GSE32465)/Homer	1e-5	-1.195e+01	0.0000	237.0	3.4
133		Gata4(Zf)/Heart-Gata4-ChIP-Seq(GSE35151)/Homer	1e-5	-1.190e+01	0.0000	752.0	11
134		p53(p53)/Saos-p53-ChIP-Seq(GSE15780)/Homer	1e-5	-1.164e+01	0.0000	82.0	1.2
135		p53(p53)/Saos-p53-ChIP-Seq/Homer	1e-5	-1.164e+01	0.0000	82.0	1.2
136		Cdx2(Homeobox)/mES-Cdx2-ChIP-Seq(GSE14586)/Homer	1e-5	-1.162e+01	0.0000	668.0	9.8
137		Mef2c(MADS)/GM12878-Mef2c-ChIP-Seq(GSE32465)/Homer	1e-5	-1.157e+01	0.0000	366.0	5.3
138		GLIS3(Zf)/Thyroid-Glis3.GFP-ChIP-Seq(GSE103297)/Homer	1e-5	-1.156e+01	0.0000	558.0	8.3
139		Arnt(Ahr(bHLH)/MCF7-Arnt-ChIP-Seq(Lo_et_al.)/Homer	1e-4	-1.140e+01	0.0000	821.0	12
140		Dlx3(Homeobox)/Kerainocytes-Dlx3-ChIP-Seq(GSE89884)/Homer	1e-4	-1.140e+01	0.0000	642.0	9.4
141		ZNF652/HepG2-ZNF652.Flag-ChIP-Seq(Encode)/Homer	1e-4	-1.140e+01	0.0000	198.0	2.9
142		Cux2(Homeobox)/Liver-Cux2-ChIP-Seq(GSE35985)/Homer	1e-4	-1.120e+01	0.0000	483.0	7.0
143		PAX3:FKHR-fusion(Paired,Homeobox)/Rh4-PAX3:FKHR-ChIP-Seq(GSE19063)/Homer	1e-4	-1.100e+01	0.0001	187.0	2.7
144		HNF6(Homeobox)/Liver-Hnf6-ChIP-Seq(ERP000394)/Homer	1e-4	-1.075e+01	0.0001	584.0	8.5
145		GSC(Homeobox)/FrogEmbryos-GSC-ChIP-Seq(DRA000576)/Homer	1e-4	-1.062e+01	0.0001	921.0	13
146		Pitx1(Homeobox)/Chicken-Pitx1-ChIP-Seq(GSE38910)/Homer	1e-4	-1.048e+01	0.0001	3031.0	44
147		X-box(HTH)/NPC-H3K4me1-ChIP-Seq(GSE16256)/Homer	1e-4	-1.039e+01	0.0001	88.0	1.2
148		FoxD3(forkhead)/ZebrafishEmbryo-Foxd3.biotin-ChIP-seq(GSE106676)/Homer	1e-4	-1.028e+01	0.0001	917.0	13
149		Hnf6b(Homeobox)/LNCaP-Hnf6b-ChIP-Seq(GSE106305)/Homer	1e-4	-1.025e+01	0.0001	906.0	13
150		Brn1(POU,Homeobox)/NPC-Brn1-ChIP-Seq(GSE35496)/Homer	1e-4	-1.007e+01	0.0001	311.0	4.5
151		NFIL3(bZIP)/HepG2-NFIL3-ChIP-Seq(Encode)/Homer	1e-4	-9.983e+00	0.0001	642.0	9.4

152		Barx1(Homeobox)/Stomach-Barx1.3xFlag-ChIP-Seq(GSE69483)/Homer	1e-4	-9.968e+00	0.0001	445.0	6.3	
153		SCL(bHLH)/HPC7-Sci-ChIP-Seq(GSE13511)/Homer	1e-4	-9.854e+00	0.0002	3544.0	51	
154		PAX6(Paired,Homeobox)/Forebrain-Pax6-ChIP-Seq(GSE66961)/Homer	1e-4	-9.758e+00	0.0002	85.0	1.2	
155		p73(p53)/Trachea-p73-ChIP-Seq(PRJNA310161)/Homer	1e-4	-9.678e+00	0.0002	39.0	0.5	
156		DLX5(Homeobox)/BasalGanglia-Dlx5-ChIP-seq(GSE124936)/Homer	1e-4	-9.678e+00	0.0002	719.0	10	
157		Etv2(ETS)/ES-ER71-ChIP-Seq(GSE59402)/Homer	1e-4	-9.573e+00	0.0002	478.0	7.0	
158		ZNF143 STAF(Zf)/CUTLL-ZNF143-ChIP-Seq(GSE29600)/Homer	1e-4	-9.566e+00	0.0002	184.0	2.7	
159		Pax8(Paired,Homeobox)/Thyroid-Pax8-ChIP-Seq(GSE26938)/Homer	1e-4	-9.435e+00	0.0002	137.0	2.0	
160		Bapx1(Homeobox)/VertebralCol-Bapx1-ChIP-Seq(GSE36672)/Homer	1e-4	-9.301e+00	0.0003	1528.0	22	
161		Nkx2.2(Homeobox)/NPC-Nkx2.2-ChIP-Seq(GSE61673)/Homer	1e-4	-9.280e+00	0.0003	1457.0	21	
162		Otx2(Homeobox)/EpiLC-Otx2-ChIP-Seq(GSE56098)/Homer	1e-4	-9.265e+00	0.0003	577.0	8.4	
163		ETS1(ETS)/Jurkat-ETS1-ChIP-Seq(GSE17954)/Homer	1e-4	-9.263e+00	0.0003	606.0	8.8	
164		RBPJ:Ebox(?,bHLH)/Panc1-Rbpj1-ChIP-Seq(GSE47459)/Homer	1e-3	-8.969e+00	0.0003	165.0	2.4	
165		PU.1(ETS)/ThioMac-PU.1-ChIP-Seq(GSE21512)/Homer	1e-3	-8.858e+00	0.0004	278.0	4.0	
166		Smad3(MAD)/NPC-Smad3-ChIP-Seq(GSE36673)/Homer	1e-3	-8.797e+00	0.0004	2084.0	30	
167		RORgt(NR)/EL4-RORgt.Flag-ChIP-Seq(GSE56019)/Homer	1e-3	-8.752e+00	0.0004	93.0	1.3	
168		RORgt(NR)/EL4-RORgt.Flag-ChIP-Seq(GSE56019)/Homer	1e-3	-8.752e+00	0.0004	93.0	1.3	
169		Pknox1(Homeobox)/ES-Prep1-ChIP-Seq(GSE63282)/Homer	1e-3	-8.741e+00	0.0004	224.0	3.2	
170		Tcf3(HMG)/mES-Tcf3-ChIP-Seq(GSE11724)/Homer	1e-3	-8.729e+00	0.0004	189.0	2.7	
171		SF1(NR)/H295R-Nr5a1-ChIP-Seq(GSE44220)/Homer	1e-3	-8.651e+00	0.0005	192.0	2.8	
172		Ap4(bHLH)/AML-Tfap4-ChIP-Seq(GSE45738)/Homer	1e-3	-8.576e+00	0.0005	1023.0	15	
173		TCFL2(HMG)/K562-TCF7L2-ChIP-Seq(GSE29196)/Homer	1e-3	-8.518e+00	0.0005	71.0	1.0	
174		ZBTB18(Zf)/HEK293-ZBTB18.GFP-ChIP-Seq(GSE58341)/Homer	1e-3	-8.517e+00	0.0005	353.0	5.3	
175		Smad4(MAD)/ESC-SMAD4-ChIP-Seq(GSE29422)/Homer	1e-3	-8.288e+00	0.0006	1460.0	21	
176		Reverb(NR),DR2/RAW-Reverba.biotin-ChIP-Seq(GSE45914)/Homer	1e-3	-8.234e+00	0.0007	93.0	1.3	

177		Unknown(Homeobox)/Limb-p300-ChIP-Seq/Homer	1e-3	-8.080e+00	0.0008	610.0	8.9
178		Mef2a(MADS)/HL1-Mef2a.biotin-ChIP-Seq(GSE21529)/Homer	1e-3	-8.053e+00	0.0008	291.0	4.2
179		EBF1(EBF)/Near-E2A-ChIP-Seq(GSE21512)/Homer	1e-3	-8.015e+00	0.0008	316.0	4.6
180		REST-NRSF(Zf)/Jurkat-NRSF-ChIP-Seq/Homer	1e-3	-7.935e+00	0.0009	8.0	0.3
181		Ets1-distal(ETS)/CD4+-PolIII-ChIP-Seq(Barski_et_al.)/Homer	1e-3	-7.932e+00	0.0009	199.0	2.9
182		MITF(bHLH)/MastCells-MITF-ChIP-Seq(GSE48085)/Homer	1e-3	-7.865e+00	0.0009	609.0	8.9
183		Sp5(Zf)/mES-Sp5.Flag-ChIP-Seq(GSE72989)/Homer	1e-3	-7.805e+00	0.0010	512.0	7.5
184		NFAT(RHD)/Jurkat-NFATC1-ChIP-Seq(Jolma_et_al.)/Homer	1e-3	-7.746e+00	0.0010	580.0	8.5
185		LRF(Zf)/Erythroblasts-ZBTB7A-ChIP-Seq(GSE74977)/Homer	1e-3	-7.659e+00	0.0011	659.0	9.6
186		ERG(ETS)/VCaP-ERG-ChIP-Seq(GSE14097)/Homer	1e-3	-7.419e+00	0.0014	784.0	11.4
187		Hoxd12(Homeobox)/ChickenMSG-Hoxd12.Flag-ChIP-Seq(GSE86088)/Homer	1e-3	-7.236e+00	0.0017	1550.0	22.2
188		EWS:ERG-fusion(ETS)/CADO_ES1-EWS:ERG-ChIP-Seq(SRA014231)/Homer	1e-3	-7.214e+00	0.0017	403.0	5.9
189		THRb(NR)/HepG2-THRb.Flag-ChIP-Seq(Encode)/Homer	1e-3	-7.199e+00	0.0017	284.0	4.2
190		Foxh1(Forkhead)/hESC-FOXH1-ChIP-Seq(GSE29422)/Homer	1e-3	-7.117e+00	0.0019	569.0	8.3
191		Pit1(Homeobox)/GCrat-Pit1-ChIP-Seq(GSE58009)/Homer	1e-3	-7.098e+00	0.0019	980.0	14.2
192		PR(NR)/T47D-PR-ChIP-Seq(GSE31130)/Homer	1e-3	-6.984e+00	0.0021	1594.0	23.1
193		NF-E2(bZIP)/K562-NFE2-ChIP-Seq(GSE31477)/Homer	1e-2	-6.795e+00	0.0026	37.0	0.5
194		ZNF264(Zf)/HEK293-ZNF264.GFP-ChIP-Seq(GSE58341)/Homer	1e-2	-6.768e+00	0.0026	373.0	5.4
195		E2F6(E2F)/Hela-E2F6-ChIP-Seq(GSE31477)/Homer	1e-2	-6.513e+00	0.0033	228.0	3.3
196		THRa(NR)/C17.2-THRa-ChIP-Seq(GSE38347)/Homer	1e-2	-6.376e+00	0.0038	201.0	2.9
197		MNT(bHLH)/HepG2-MNT-ChIP-Seq(Encode)/Homer	1e-2	-6.223e+00	0.0044	810.0	11.4
198		Oct6(POU,Homeobox)/NPC-Pou3f1-ChIP-Seq(GSE35496)/Homer	1e-2	-6.181e+00	0.0046	387.0	5.6
199		Six2(Homeobox)/NephronProgenitor-Six2-ChIP-Seq(GSE39837)/Homer	1e-2	-6.084e+00	0.0050	819.0	12.0
200		ZSCAN22(Zf)/HEK293-ZSCAN22.GFP-ChIP-Seq(GSE58341)/Homer	1e-2	-6.055e+00	0.0052	53.0	0.7
201		Zic(Zf)/Cerebellum-ZIC1.2-ChIP-Seq(GSE60731)/Homer	1e-2	-6.014e+00	0.0053	631.0	9.2
202		Bach2(bZIP)/OCILy7-Bach2-ChIP-Seq(GSE44420)/Homer	1e-2	-6.002e+00	0.0054	116.0	1.7

203		Tlx7(NR)/NPC-H3K4me1-ChIP-Seq(GSE16256)/Homer	1e-2	-5.935e+00	0.0057	148.0	2.1
204		Nrf2(bZIP)/Lymphoblast-Nrf2-ChIP-Seq(GSE37589)/Homer	1e-2	-5.872e+00	0.0061	33.0	0.4
205		Nkx2.5(Homeobox)/HL1-Nkx2.5.biotin-ChIP-Seq(GSE21529)/Homer	1e-2	-5.850e+00	0.0062	1599.0	23.0
206		Zic3(Zf)/mES-Zic3-ChIP-Seq(GSE37889)/Homer	1e-2	-5.711e+00	0.0071	359.0	5.2
207		p63(p53)/Keratinocyte-p63-ChIP-Seq(GSE17611)/Homer	1e-2	-5.676e+00	0.0073	205.0	3.0
208		CRX(Homeobox)/Retina-Crx-ChIP-Seq(GSE20012)/Homer	1e-2	-5.603e+00	0.0078	1679.0	24.0
209		Hoxa10(Homeobox)/ChickenMSG-Hoxa10.Flag-ChIP-Seq(GSE86088)/Homer	1e-2	-5.602e+00	0.0078	440.0	6.4
210		CRE(bZIP)/Promoter/Homer	1e-2	-5.410e+00	0.0094	239.0	3.5
211		NF1(CTF)/LNCAP-NF1-ChIP-Seq(Unpublished)/Homer	1e-2	-5.369e+00	0.0097	130.0	1.9
212		Oct11(POU,Homeobox)/NCIH1048-POU2F3-ChIP-seq(GSE115123)/Homer	1e-2	-5.348e+00	0.0099	272.0	3.9
213		Bach1(bZIP)/K562-Bach1-ChIP-Seq(GSE31477)/Homer	1e-2	-5.275e+00	0.0106	43.0	0.6
214		HLF(bZIP)/HSC-HLF.Flag-ChIP-Seq(GSE69817)/Homer	1e-2	-5.253e+00	0.0108	687.0	10.0
215		ZNF7(Zf)/HepG2-ZNF7.Flag-ChIP-Seq(Encode)/Homer	1e-2	-5.218e+00	0.0111	426.0	6.2
216		Hnf1(Homeobox)/Liver-Foxa2-Chip-Seq(GSE25694)/Homer	1e-2	-5.196e+00	0.0113	135.0	1.9
217		Srebp1a(bHLH)/HepG2-Srebp1a-ChIP-Seq(GSE31477)/Homer	1e-2	-5.024e+00	0.0133	127.0	1.8
218		PBX1(Homeobox)/MCF7-PBX1-ChIP-Seq(GSE28007)/Homer	1e-2	-4.954e+00	0.0142	86.0	1.2
219		HNF1b(Homeobox)/PDAC-HNF1B-ChIP-Seq(GSE64557)/Homer	1e-2	-4.930e+00	0.0145	168.0	2.4
220		Nr5a2(NR)/Pancreas-LRH1-ChIP-Seq(GSE34295)/Homer	1e-2	-4.906e+00	0.0148	271.0	3.9
221		Zic2(Zf)/ESC-Zic2-ChIP-Seq(SRP197560)/Homer	1e-2	-4.871e+00	0.0153	280.0	4.1
222		SpiB(ETS)/OCILY3-SPIB-ChIP-Seq(GSE56857)/Homer	1e-2	-4.825e+00	0.0159	141.0	2.0
223		Rfx5(HTH)/GM12878-Rfx5-ChIP-Seq(GSE31477)/Homer	1e-2	-4.801e+00	0.0162	170.0	2.4
224		TATA-Box(TBP)/Promoter/Homer	1e-2	-4.731e+00	0.0173	1082.0	15.0
225		Srebp2(bHLH)/HepG2-Srebp2-ChIP-Seq(GSE31477)/Homer	1e-2	-4.728e+00	0.0173	82.0	1.2
226		OCT:OCT(POU,Homeobox)/NPC-OCT6-ChIP-Seq(GSE43916)/Homer	1e-2	-4.726e+00	0.0173	135.0	1.9
227		Phox2a(Homeobox)/Neuron-Phox2a-ChIP-Seq(GSE31456)/Homer	1e-2	-4.656e+00	0.0184	378.0	5.5
228		RORg(NR)/Liver-Rorc-ChIP-Seq(GSE101115)/Homer	1e-2	-4.616e+00	0.0191	71.0	1.0

