THE UNIVERSITY
of EDINBURGH

This thesis has been submitted in fulfilment of the requirements for a postgraduate degree (e.g. PhD, MPhil, DClinPsychol) at the University of Edinburgh. Please note the following terms and conditions of use:

- This work is protected by copyright and other intellectual property rights, which are retained by the thesis author, unless otherwise stated.
- A copy can be downloaded for personal non-commercial research or study, without prior permission or charge.
- This thesis cannot be reproduced or quoted extensively from without first obtaining permission in writing from the author.
- The content must not be changed in any way or sold commercially in any format or medium without the formal permission of the author.
- When referring to this work, full bibliographic details including the author, title, awarding institution and date of the thesis must be given.

INTERROGATION OF TRANSCRIPTIONAL, REGULATORY AND SIGNALLING NETWORKS IN FETAL THYMIC EPITHELIAL CELL DEVELOPMENT VIA *IN SILICO* ANALYSES

Anastasia Kousa

Thesis submitted for the degree of Doctor of Philosophy
The University of Edinburgh
August 2018

Declaration

I declare that this thesis was composed by myself, that the work contained herein is my own except where explicitly stated otherwise in the text, and that this work has not been submitted for any other degree or professional qualification except as specified.

Anastasia Kousa

Edinburgh, 2018

Acknowledgements

Firstly, I would like to express my sincere gratitude to my supervisor Dr Simon Tomlinson and co-supervisor Prof Clare Blackburn for their constant support, advice and motivation. Their meticulous suggestions and critical feedback has guided me through my PhD project (even in times of self-doubt) over the last four years. Further thanks go to my PhD committee member Professor Ian Chambers for his valuable help, which kept my project on track.

My sincere appreciation goes to my undergraduate supervisor Assoc Prof Nicholas M. Glykos -who will probably never read this thesis-, for his continuous guidance and astute criticism over the years; I will always consider him as my mentor because he gave me a zeal for science and replied “Google it” to every single question of mine.

I am deeply grateful to Dr Florian Halbritter and Dr Duncan Godwin for the many enlightening discussions we had over the years; discussions that simplified many complex concepts in the bioinformatics and the programming world. Moreover, I give particular thanks to Dr Harsh Vaidya and Dr Dong Liu because without their data contribution, there would be no project for me to work on. Also, to my flatmate, Dr Stamatina Fragkogianni for the many late-night bioinformatics discussions and for bearing with me during the four beautiful years that we shared a flat. To Dr Filip Wymeersch for countless moments of laughter, weird conversations over lunch breaks and for still calling me his “gym buddy” when I kept standing him up for abs class. To Dr Alison McGarvey for her friendship, countless coffee breaks (even if she was drinking tea) and moral support. And to Dr Alastair Kilpatrick for proofreading my thesis when no one else would.

Being a member of two research groups, I have been lucky to share multiple science trips and group meetings with several people throughout my project, namely, Dr Dominique Meunier, Dr Svetlana Ulyanchenko, Dr Timothy Henderson, Paul Rouse, Dr Kathy O’Neil, Dr Karen Jent, Dr Alberto Briones Leon, Joanna Sweetman, James Ashmore, Dr Jonathan Manning, Dr Aidan McGlinchey and Will Bowring.

Also, to Riccardo, Theo, Ira, Theodora, Vaso, Marina-Elena, Marina, Ioanna, Andrea and Sebastián and many other important people that have been part of my life in Edinburgh, with whom I have shared a lot of laughter, fun, travels and comradery in the ups and downs of the PhD life and without them grey Edinburgh weather would feel much more miserable.

Finally, I would like to thank my family, especially my mom, who has given me strength and moral support in every step that I took, even if she still does not exactly understand what I am doing.

“If you can't explain it simply, you don't understand it well enough.”

Albert Einstein

Abstract

The thymus is the primary lymphoid organ responsible for the development and maturation of T lymphocytes (aka T-cells) in vertebrates. The complex architecture of the thymic microenvironment orchestrates the formation of a diverse and self-tolerant T-cell repertoire capable of supporting the development and maintenance of a functional immune system. The main component of this microenvironment, the thymic epithelium, is crucially required to direct thymus organogenesis and homeostasis, and to mediate T-cell repertoire development and selection. The thymic epithelial progenitor cells (TEPCs) from which the mature thymus develops originate from the endoderm of the 3rd pharyngeal pouch by embryonic day 9 in mouse development (or early week 6 in human embryos). Expression of the transcription factor FOXN1 is required to drive TEPCs differentiation in each thymic epithelial lineage (TEC), while the absence of functional FOXN1 causes athymia. Moreover, forced expression of *Foxn1* in mouse embryonic fibroblasts (MEFs) converts these MEFs into TECs that can support the development of a normal thymic system. Despite the great therapeutic potential that TEPCs present in regenerative medicine, there is currently no detailed model describing regulation of the TEPC state and its differentiation into cortical (c) and medullary (m) TECs, or explaining the dominant role of FOXN1 in the thymic epithelial system. Comparative transcriptomics analysis in conjunction with pathway enrichment analysis of the developing TEPCs could reveal the signalling pathways that regulate the early TEPC state and progression into differentiation. Additionally, integrative bioinformatics analysis of transcriptomics and genomics datasets could identify the functional networks that are directly regulated by FOXN1 during early TEC progression. In this thesis I provide, for the first time, an *in silico* model explaining fetal TEPC differentiation into the functionally distinct TEC lineages, in the cellular, molecular and signalling contexts of thymus early development. Furthermore, I present evidence which suggests that FOXN1 could be a pioneer factor, capable of fully establishing the transcriptional programme that underpins thymic epithelial cell identity and function. Finally, in this thesis, I introduce the development of an interactive thymic-specific database that provides a platform for easy access, analysis and integration of curated bioinformatics datasets.

Lay summary

The thymus is the organ majorly responsible for the normal function of the immune system. Thymocytes (a.k.a. T-cells), specialised defence cells that attack foreign invaders in the body (known as antigens), are developed and trained in the thymus. Within the thymus, T-cell maturation takes place via another group of cells, the thymic epithelial cells or TECs. TECs are of two kinds, cortical thymic epithelial cells (cTECs) which reside in the outer region of the thymus called the cortex, and medullary thymic epithelial cells (mTECs) that exist in several regions inside the organ, called thymic medullary regions. During thymus development, these TEC populations seem to form from an earlier, more immature cell type, the thymic epithelial progenitor cell (TEPC) population. This transition from TEPCs towards cTECs and mTECs cannot occur in the absence of a factor named FOXN1. A block of FOXN1 expression in TEPCs causes athymia, while its forced expression in an unrelated cell type is able to support from scratch the development of a functional thymus. This thesis investigates the transition model of the earlier TEPC population towards the more mature and specialised cTECs and mTECs, based on representative datasets that measure all of the genes present in the above populations. Furthermore, in this thesis, I explore the signalling cues that act in the very early TEPC state, to identify ways to improve TEPC culture conditions. Additionally, by comparing FOXN1-dependent genes in TEPCs with candidate direct FOXN1 target genes in cTECs, I also try to demystify how FOXN1 applies its leading role in TEC fate. Lastly, I introduce the development of an online interface specific for the thymic system that provides an easy way to store, access and share data within experimental groups, and offers further analysis and visualisation options for non-bioinformatics specialists.

Abbreviations

3rd PP – Third pharyngeal pouch

ATAC-seq – Assay for transposase-accessible chromatin using sequencing

BMP – Bone morphogenetic protein

bp – base pair

ChIP-seq – Chromatin immunoprecipitation coupled with sequencing

cKO – Conditional knockouts

CMJ – Corticomedullary junction

cTECs – Cortical thymic epithelial cells

cTEPCs – Cortical thymic epithelial progenitor cells

DamID-seq – DNA adenine methyltransferase identification coupled with sequencing

DAPI – 4',6-Diamidino-2-phenylindole

DAPT – γ-secretase inhibitor

DC – Dendritic cells

DN – Double negative

DP – Double positive

E – Embryonic day

ECM – Extracellular matrix

FACS – Fluorescence-activated cell sorting

FGF – Fibroblast growth factor

FTOC – Fetal thymus organ culture

GOF – Gain-of-function

LOF – Loss-of-function

MHC – Major histocompatibility complex

mTECs – Medullary thymic epithelial cells

mTEPCs – Medullary thymic epithelial progenitor cells

NICD – NOTCH intracellular domain

PC – Principal component

PCA – Principal component analysis

RNA-seq – Ribonucleic acid sequencing

RT-qPCR – Real-time quantitative polymerase chain reaction

scRNA-seq – single cell RNA-seq

TCR – T-cell receptor

TECs – Thymic epithelial cells xii

TEPCs – Thymic epithelial progenitor cells

MHC –Major histocompatibility complex

RTOC – Reaggregate thymic organ culture

Table of Contents

Declaration……………………………………………………………………...i

Acknowledgements………………………………………………………………ii

Abstract……………………………………………………………………………v

Lay Summary……………………………………………………………………vi

Abbreviations……………………………………………………………………vii

Table of Contents………………………………………………………………...ix

List of Figures…………………………………………………………………...xvi

List of Tables……………………………………………………………………xix

CHAPTER 1 – INTRODUCTION 1

1.0 THE THYMUS (PRÉCIS I) 2

1.1 THE THYMIC MICROENVIRONMENT 4

1.1.1 Thymic epithelial cells and other stromal cells 4

1.1.2 T-cell repertoire selection 5

1.2 THYMUS ORGANOGENESIS 6

1.2.1 Origin of thymic epithelial cells 8

1.2.2 Specification and differentiation of thymic epithelial cells in the embryonic thymus 8

1.2.2.1 A genetic regulatory network acting prior to TEC specification 9

1.2.2.2 Bipotent and lineage-fated thymic epithelial progenitor cells............................ 11

1.2.2.3 A serial progression model into the embryonic cortical and medullary thymic epithelial cell lineages 13

1.2.2.4 Cell-crosstalk dependence for TEC progression... 16

1.2.3 Signalling mechanisms involved in early thymus progression 16

1.2.3.1 Sonic Hedgehog (SHH)... 16

1.2.3.2 Bone Morphogenic Proteins (BMPs) 17

1.2.3.3 Fibroblast Growth Factors (FGFs) 18

1.2.3.4 Wingless-type MMTV integration site family members (WNTs) 19

1.2.4 The role of FOXN1 in thymus development 20

1.2.4.1 FOXN1-independent TEC specification 20

1.2.4.2 FOXN1-dependent TEC differentiation 21

1.2.4.3 FOXN1 binding motif 22

1.2.4.4 FOXN1 target genes 23

1.2.5 The role of FOXN1 in skin 24

1.3 CLASSICAL MODELS FOR STEM CELL DIFFERENTIATION AND MAINTENANCE 27

1.4 MODELS BY WHICH NOTCH OPERATES IN OTHER CELL TYPES 30

1.5 MECHANISMS OF GENE REGULATION 32

1.5.1 Mechanisms of transcriptional activation and repression 33

1.5.2 Chromatin structure controlled by activators and repressors 34

1.5.3 Chromatin modifications as indicators of promoter/enhancer activity 36

1.5.4 The role of pioneer factors 36

1.6 NEXT GENERATION SEQUENCING TECHNOLOGIES AND DATA ANALYSIS (PRÉCIS II) 39

1.6.1 A short history of DNA sequencing technologies 39

1.6.2 RNA-seq data processing for novel biological insight 42

1.6.3 ChIP-seq data processing for novel biological insight 44

1.6.4 NGS and data storing 47

1.6.5 Understanding gene regulatory mechanisms through RNA-seq and ChIP-seq data integration 48

1.7 PROJECT AIMS AND THESIS STRUCTURE 50

CHAPTER 2 – MATERIALS AND METHODS 52

2.1 PRÉCIS 53

2.2 DATA REPOSITORY 54

2.2.1 Collection of transcriptomic datasets in mouse (RNA-seq) 54

2.2.1.1 A Developmental series of fetal TEPC data 54

2.2.1.2 A *Foxn1* Allelic series of fetal TEPC data 54

2.2.1.3 An RBPJ-mutant series of fetal TEPC data 55
2.2.1.4 TEC subpopulations from newborn and adult mice 56
2.2.2 Collection of regulatory datasets in mouse (ChIP-seq) 57
2.2.2.1 A FOXN1-flagged peptide protein dataset in adult TECs 57
2.2.3 Collection of accessibility datasets in mouse (ChIP-seq) 59
2.2.3.1 Histone modification data from fetal TEPCs 59
2.3 BIOINFORMATICS ANALYSIS AND WORKFLOWS 60
2.3.1 A common pre-analysis step in the RNA-seq and ChIP-seq analysis pipelines 60
2.3.1.1 Raw reads quality control 60
2.3.2 RNA-seq pre-analysis pipeline 61
2.3.2.1 Mapping reads to genome 61
2.3.2.2 Quantification of gene counts 61
2.3.3 RNA-seq core analysis pipeline 63
2.3.3.1 Low gene counts 63
2.3.3.2 Data normalisation and differential expression analysis 65
2.3.3.3 Data inspection and visualisation 65
2.3.3.4 Batch effect correction 66
2.3.3.5 Differential expression analysis 66
2.3.3.6 Pathway enrichment analysis 67
2.3.4 ChIP-seq pre-analysis pipeline 70
2.3.4.1 Mapping reads to genome 70
2.3.4.2 Peak calling 70
2.3.5 ChIP-seq core analysis pipeline 73
2.3.5.1 Significant peaks 73
2.3.5.2 Data inspection 73
2.3.5.3 Peaks to genes 73
2.3.5.4 Motif analysis 74

CHAPTER 3 – MODELLING THE DEVELOPMENTAL PROGRESSION OF EARLY TEPCS DURING THYMUS DEVELOPMENT 76

3.1 PRÉCIS 77

3.2 RESULTS ... 79

3.2.1 Pseudo-timing of a *Foxn1* Allelic series dataset represents normal developmental progression of TEPCs ... 79

3.2.2 Projection of differentiating TEPC genes onto representative populations of the TEC sublineages indicates TEPC heterogeneity among the E12.5 TEPCs ... 92

3.2.3 *Foxn1* heterogeneity among E12.5 bipotent TEPCs ... 100

3.2.4 A well-defined TEP/SC gene signature from scRNA-seq data ... 103

3.2.5 Alternative markers to use in combination with PLET1 and EpCAM to identify a homogeneous population of bipotent TEP/SCs ... 103

3.3 DISCUSSION ... 106

3.3.1 A potential TEP/SC differentiation model ... 108

3.4 SUMMARY ... 112

CHAPTER 4 – AN *IN SILICO* MODEL OF SIGNALLING REQUIREMENTS FOR THE SURVIVAL AND EXPANSION OF BIPOTENT FETAL TEPCS ... 113

4.1 PRÉCIS ... 114

4.2 RESULTS ... 116

4.2.1 Pathways reshaped at the onset of *Foxn1* expression in TEP/SCs: pathways switching off ... 116

4.2.1.1 FGF signalling pathway ... 120

4.2.1.2 NOTCH, BMP and canonical WNT signalling pathways ... 122

4.2.1.3 SHH and WNT (non-canonical) signalling pathways ... 125

4.2.1.4 IGF signalling pathway ... 127

4.2.2 A testable model of the signalling pathways required to maintain and expand TEP/SCs *in vivo* ... 129

4.2.3 Exploring NOTCH signalling in the thymic epithelium: regulation of the medullary thymic epithelial cell lineage via specification of medullary progenitor cells. ... 134

4.2.3.1 Genetic ablation of NOTCH signalling in early TEPCs under the control of *Foxn1* demonstrates an effect only in the mTEC compartment 134
4.2.3.2 Time-dependence of mTECs ablation during early thymus development 140
4.2.3.3 Genetic reinforcement of NOTCH signalling in early TEPCs perturbs the exit from the early TEPC state 142
4.3 DISCUSSION 150
4.3.1 A clear separation of TEC and parathyroid fate 150
4.3.2 Changes in TEPC identity could be reflecting changes in cell morphology 151
4.3.3 Switching to a neighbouring cell-dependent fate 154
4.3.4 Establishing the mTEC lineage from a common TEPC progenitor 156
4.4 SUMMARY 159

CHAPTER 5 – FOXN1: A MASTER REGULATOR OF TEC DIFFERENTIATION 160

5.1 PRÉCIS 161
5.2 RESULTS 162
5.2.1 Prediction of candidate direct FOXN1 targets in mouse fetal TEPCs 162
5.2.1.1 FOXN1-dependent genes in fetal TEPCs 162
5.2.1.2 Direct binding events of FOXN1 in newborn TECs 166
5.2.1.3 Inferring candidate direct FOXN1 targets in fetal TEPCs from FOXN1 binding events in newborn TECs 166
5.2.1.4 Insights in the TEPC chromatin accessibility landscape based on promoter and enhancer histone modification marks in E12.5 TEPCs 170
5.2.1.5 Prediction of high confidence candidate direct FOXN1 targets in fetal TEPCs from overlapping FOXN1 binding sites in newborn TECs with accessibility histone marks in fetal TEPCs 174
5.2.2 Comparative analysis and insights on the highly confident direct candidate FOXN1 targets in fetal TEPCs 177
5.2.3 Comparative analysis and insight of the highly confident direct candidate FOXN1 targets in newborn cTECs 181

5.2.4 Differential FOXN1 binding profiles in fetal TEPCs and newborn TECs 182

5.2.5 Motif discovery in regions defined by FOXN1 peaks in newborn TECs and histone modification marks in fetal TEPCs 186

5.2.6 Nodal points that FOXN1 regulates in fetal TEPCs during thymus organogenesis 189

5.3 DISCUSSION 193

5.3.1 FOXN1: a potential pioneer factor 193

5.3.2 A first-described high-confidence list of FOXN1 direct targets in fetal TEPCs and dynamic regulation of pathways 194

5.3.3 FOXN1 drives and maintains the antigen processing and presentation programme via multiple mechanisms 195

5.3.3.1 General mechanism of interferon gamma signalling 195

5.3.3.2 MHC class II antigen presentation pathway and IFN-γ 197

5.3.3.3 The MHC class II transactivator (CIITA) 197

5.3.3.4 CIITA and TECs 197

5.3.3.5 FOXN1 consists part of the unknown mechanism that maintains CIITA's expression and the MHC class II programme in the absence of IFN-γ signalling 198

5.3.4 The FOXN1-NOTCH interplay in determining TEC specification 199

5.3.5 A potential role for *Tfap2a* in the early TEPCs 201

5.4 SUMMARY ..... ERROR! BOOKMARK NOT DEFINED.

CHAPTER 6 – THYMIBASE 204

6.1 PRÉCIS 205

6.2 METHODS 207

6.3 RESULTS / DISCUSSION 208

6.3.1 A curated collection of thymic datasets 208

6.3.2 An interface to support easy reanalysis of data 211

6.4 SUMMARY 216

CHAPTER 7 – CONCLUDING REMARKS 217

7.1 CONCLUSIONS 218
7.1.1 A revised serial progression model of TEPCs into cortical and medullary TECs 218
7.1.2 An *in silico* model of the intrinsic and extrinsic cues acting in the early undifferentiated TEPC state: NOTCH signalling in mTEC specification 220
7.1.3 FOXN1: A master regulator of the thymic system and a potential pioneer factor 222
7.1.4 Role of the FOXN1-NOTCH interplay in the fetal TEPCs and potential mechanisms of how it deciphers fate specification during TEPC progression 223
7.1.5 ThymiBase 227
7.2 FUTURE WORK 228
7.2.1 Single cell RNA-seq analysis of E12.5 TEPCs 228
7.2.2 Validation of FOXN1's pioneer activity in TEPCs 228
7.2.3 ThymiBase: additional components and ChIP-seq data....................... 229

References.. 230

List of Figures

Figure 1.1: An overview of early thymus organogenesis. 7

Figure 1.2: A serial progression model of TEC early progression in development.. 15

Figure 2.1: Relative levels of expression for the *Foxn1*^{WT} and *Foxn1*^{R} variants and schematic representation of the *Foxn1*^{R} design... 58

Figure 2.2: Schematic representation of the FOXN1-tagged protein design construct. 58

Figure 2.3: Overview plot of the RNA-seq pre-analysis pipeline............................ 62

Figure 2.4: Density plots of the log2-transformed CPM values from example datasets. 64

Figure 2.5: Overview plot of the RNA-seq core analysis pipeline. 69

Figure 2.6: Overview plot of the ChIP-seq pre-analysis pipeline............................ 72

Figure 2.7: Overview plot of the ChIP-seq core analysis pipeline. 75

Figure 3.1: The *Foxn1*^{R} hypomorphic allele. 81

Figure 3.2: Graphical representation of the TEPC Developmental and the *Foxn1* Allelic series 82

Figure 3.3: Cluster dendrogram of the E10.5, E11.5 and E12.5 sample triplicates from the TEPC Developmental series dataset.. 85

Figure 3.4: Cluster dendrogram of the *Foxn1* variant samples from the *Foxn1* Allelic series dataset... 85

Figure 3.5: PCA/variable associations among the TEPC Developmental series and *Foxn1* Allelic series integrated dataset before batch effect correction 87

Figure 3.6: PCA/variable associations among the TEPC Developmental series and *Foxn1* Allelic series integrated dataset after batch effect correction 87

Figure 3.7: Principal component analysis of the TEPC Developmental series and *Foxn1* Allelic series integrated dataset. 89

Figure 3.8: Lineage-specific gene heatmaps for the TEPC Developmental series and the *Foxn1* Allelic series datasets before data integration... 91

Figure 3.9: Representative levels of *Foxn1* in the TEPC Developmental series dataset... 93

Figure 3.10: Projection of TEPC differentially expressed genes on E14.5 and 1w TEC subpopulations. 95
Figure 3.11: Projection of FOXN1-dependent, differentially expressed TEPC genes on E14.5 and 1w TEC subpopulations 96
Figure 3.12: Expression levels of the FOXN1-dependent, down-regulated genes (E12.5 versus E10.5) in E10.5 TEPCs, E12.5 TEPCs, 1w old cTECs and 1w old mTECs 99
Figure 3.13: Scatterplot of *Plet1* versus *Foxn1* among various TEC samples 102
Figure 3.14: Candidate marker list to be used in combination with PLET1 for TEP/SC isolation. 105
Figure 3.15: Schematic representation of a potential TEPC differentiation model towards the TEC sublineages. 110
Figure 4.1: Grouped overview of down-regulated signalling pathways based on core enrichment genes overlap. 119
Figure 4.2: GSEA Enrichment plot for the SHC-MEDIATED CASCADE pathway and gene list of core enrichment. 121
Figure 4.3: GSEA Enrichment plot for the NEURAL CREST DIFFERENTIATION pathway and gene list of core enrichment. 124
Figure 4.4: GSEA Enrichment plot for the HEDGEHOG SIGNALING PATHWAY and gene list of core enrichment. 126
Figure 4.5: GSEA Enrichment plot for the IGF1R SIGNALING CASCADE and gene list of core enrichment. 128
Figure 4.6: Schematic representation of the down-regulated signalling pathways activity model in the early TEPCs to assist TEPCs expansion *in vitro*. 132
Figure 4.7: Expression profiles of *Wnt5a* receptors and interactors in TEPCs.......133
Figure 4.8: PCA plot of the E14.5 RBPJ WT and LOF (RBPJ cKO) samples. 138
Figure 4.9: Differences among RBPJ cKO, NICD and WT groups. 144
Figure 4.10: GSEA Enrichment plot for NOTCH1 INTRACELLULAR DOMAIN REGULATES TRANSCRIPTION and gene list of core enrichment. 145
Figure 4.11: Heatmap of TEPC and TEC-specific lineage markers for NOTCH mutant and control samples. 148

Figure 5.1: Overview of the integrative analysis for identification of FOXN1 target candidates in TEPCs. 165
Figure 5.2: FOXN1 ChIP-seq peaks distribution in the mouse reference genome assembly gene regions.......... 169
Figure 5.3: Genomic region profile plots and heatmaps over the FOXN1 IDR peak regions in newborn TECs for the FOXN1 and histone modification ChIP-seq datasets. 172
Figure 5.4: Volcano plots of the FOXN1 candidate targets in fetal TEPCs 176
Figure 5.5: Heatmap of the high confidence FOXN1 target candidates in fetal TEPCs 180
Figure 5.6: Genomic region profile plots and heatmap clusters over the FOXN1-H3K4me3, FOXN1-H3K27ac, FOXN1-H3K4me3-H3K27ac and FOXN1 IDR overlapping peaks for the FOXN1 and histone modification ChIP-seq datasets..... 184
Figure 5.7: Heatmap of the high confidence FOXN1 target candidates that regulate enriched signalling pathways in fetal TEPCs.......... 192
Figure 5.8: Regulation the MHCII programme through direct binding of the interferon gamma activated STAT1 and IRF1 on the pIV promoter of *Ciita*.......... 196
Figure 6.1: ThymiBase database panel. 210
Figure 6.2: ThymiBase visualisation panel.......... 213
Figure 6.3: ThymiBase panel for differential expression analysis. 214
Figure 6.4: ThymiBase panel for pathway enrichment analysis. 215
Figure 7.1: Model of the FOXN1-NOTCH interplay in TEPC progression.......... 224
Figure 7.2: Potential fate decision mechanisms for the E10.5 TEP/SCs…………226

List of Tables

Table 4.1: List of enriched signalling pathways with decreased activity in TEPCs between E10.5 and E12.5 timepoints 117

Table 4.2: Expression values of NOTCH targets for the E14.5 RBPK cKO (LOF) and WT samples. 139

Table 4.3: Up-regulated signalling pathways in the E14.5 PLET1$^+$ GOF versus E14.5 PLET1$^+$ WT comparison. 145

Table 5.1: Unique FOXN1 targets in newborn cTECs compared to targets in fetal TEPCs. 182

Table 5.2: *De novo* identification of canonical and extended FOXN1 binding motifs in the regions under the IDR FOXN1 peaks in newborn TECs that overlap (or not) with histone modification marks in fetal TEPCs. 188

Table 5.3: List of signalling pathways whose activity increases during the developmental progression from E10.5 undifferentiated TEPCs to E12.5 TEPCs.. 191

Chapter 1

Introduction

Chapter 1

Introduction

1.0 THE THYMUS (PRÉCIS I)

Since antiquity, a halo of mystery has surrounded the thymus organ. Originating from the Greek word *θύμος*, meaning “warty excrescence” (like a thyme bud), the thymus was also misinterpreted by the ancient Greeks to mean “house of the human soul” (*θυμός*), possibly due to the location of the thymus just above the heart in the human body (Fanu, 1963; Fischbein, 1984; Jacobs et al., 1999). Galen of Pergamum (130-200 AC) was the first to note that this organ grew to its full size during childhood (Coleman, 1969), however it was not until 1961 that Jacqes Miller finally determined thymus true function in host immunity (Miller, 2002). Miller used thymectomised young mice to demonstrate the importance of the thymus in the generation of key defence cells for their immune system, the nowadays well-known T lymphocytes (also known as thymocytes or T-cells).

Today, the thymus is known as the central lymphoid organ responsible for the development and maturation of T-cells in vertebrates. Circulating hematopoietic lymphoid precursors are attracted by the thymus and migrate inside the organ, where they undergo maturation to engender self-tolerant and naïve thymocytes equipped with a broad repertoire of antigen specific T-cell receptors (TCRs), before they are released back into the bloodstream (Boyd et al., 1993; van Ewijk et al., 1999; Jenkinson, 1992). The complex architecture of the thymic microenvironment encompasses the cellular crosstalk between the non-hematopoietic stromal cells residing in the thymus and the incoming immature hematopoietic progenitors, directing them in a patterned fashion through the compartments of the thymus that orchestrate thymocyte stepwise differentiation and selection (Benjamini and Hochberg, 1995; Takahama, 2006).

These processes subsequently contribute to the development and maintenance of a functional immune system.

Thymocyte differentiation and specificity occurs at its highest efficacy during the very early stages of the thymus development. It decreases after puberty and is almost (if not fully) absent during the elderly years (age-related thymus atrophy), highlighting the importance of the thymus structure in the early stages of life (discussed in Boehm and Swann, 2013). This progressive shrinking of the thymus that advances with age is called thymic involution and is characterised by an acute decline in naïve T-cells. This sequentially impacts on the host's immune system faculty in producing effective responses against pathogens (Gruver et al., 2007; Linton and Dorshkind, 2004). Consequently, thymic atrophy may have considerable consequences in the outcomes of chemotherapy and/or other cytoablative treatments on ageing patients, since these are heavily dependent on thymic efficacy for reconstitution of an adequate repertoire of T-cell receptors. Regenerative strategies designed to prevent or invert thymus waning continue to be examined for clinical use (Chidgey et al., 2008).

Thymus regeneration can be achieved through reactivation of the endogenous tissue (tissue regeneration) by chemically stimulating residual thymic epithelial progenitor cells (TEPCs) or by creating *de novo* thymic tissue (tissue replacement). Administration of cytokines and growth factors (GFs), such as Interleukin 7 (IL7) and Interleukin 7 (IL22) or the Fibroblast Growth Factor 7 (FGF7) and the keratinocyte growth factor (KGF), as well as enforced Forkhead Box Protein N1 (FOXN1) expression have been experimentally shown to promote thymus regeneration (Alpdogan, 2006; Bredenkamp et al., 2014a; Dudakov et al., 2012; Mackall et al., 2001). Thymic epithelial cells (TECs) can be generated through transdifferentiation from unrelated tissues into TECs or alternatively TEC-like cells can be generated through directed differentiation of pluripotent stem cells (PSCs) (Bredenkamp et al., 2014b; Parent et al., 2013; Sun et al., 2013). TECs constitute the main component of the thymic stroma (which is essential in thymus development and maintenance) and the above findings underline the possibility of tissue replacement. Significantly, long-term self-renewing TEPCs and adequately compartmentalised TEC sublineage regions need to be contained in such tissue (Bredenkamp et al., 2015).

1.1 THE THYMIC MICROENVIRONMENT

The mature thymic microenvironment comprises a complex network of various cell types that together establish and maintain a functional immune system. More than 95% of the thymic cellularity is made up by developing thymocytes (T-cells), whose maturation is mediated by the thymic stroma. Within the thymic stroma, specialised thymic epithelial cells (TECs), neural crest (NC) – derived mesenchymal cells and bone marrow (BM) – derived lymphoid cells form a complex three-dimensional network which orchestrates T-cell development and differentiation (Boyd et al., 1993). Two main compartments are histologically apparent in the thymus: the cortex and medulla. The thymic medulla is located in the innermost layer of the organ and is surrounded by the thymic cortex, forming a thin corticomedullary junction (CMJ) in the area of contact. In turn, the cortex is surrounded by a thin layer of simple epithelium, the subcapsule, which separates it from the capsule, a thick outer layer that lies around the organ (Boyd et al., 1993).

1.1.1 Thymic epithelial cells and other stromal cells

Thymic epithelial cells (TECs) can be split in three main categories: cortical TECs (cTECs), medullary TECs (mTECs) and subcapsular epithelial cells. cTECs reside in the thymus cortex where they form three-dimensional networks with densely packed developing thymocytes. mTECs reside in the thymic medulla forming similar 3D networks with the available thymocytes, with the mTEC/thymocyte ratio much higher compared to the one in the cortical compartment (Boyd et al., 1993). In addition to TECs, bone marrow (BM) – derived cells reside in the thymic stroma: these are principally dendritic cells and macrophages. Dendritic cells spread across the thymus, but mostly accumulate in the CMJ, while they express strongly MHC class II molecules. Macrophages can be found in both the cortex and the CMJ compartments but express MHC class II molecules in varying levels contrary to the high MHC class II expression of dendritic cells (Duijvestijn and Hoefsmit, 1981; Milićević et al., 1987). Finally, mesenchymal cells consist a vital part of the thymic stroma (Boyd et al., 1993). Extracellular matrix (ECM) components are secreted by the thymic mesenchyme and are potentially involved in the regulation of TECs (Schreiber et al., 1991; Watt et al., 1991) and T-cells (Cardarelli and Pierschbacher, 1986; Cardarelli et al., 1988) via ECM receptors that the latter two populations have been found to express.

1.1.2 T-cell repertoire selection

The highly organised inner thymic architecture supports thymocyte development through its complex network of stromal cells. Hematopoietic progenitors enter the thymic structure (early thymic progenitors; ETPs) through the corticomedullary junction (CMJ) and move towards the subcapsule (Wu et al., 1991). At this early stage, ETPs do not express the T-cell receptor (TCR) or either of the cluster of differentiation 4 (CD4) or cluster of differentiation 8 (CD8) co-receptors and are characterised as "double negative" (DN) cells. Throughout four DN stages (DN1, DN2, DN3 and DN4), T-cells gradually acquire expression of the above molecules and become ready to undergo positive selection at the cortex. Only cells expressing TCR, and the CD4 and CD8 glycoproteins are destined to go through the DP stage, where they can receive a survival signal if they come in contact with MHC class I or MHC class II molecules (Jameson et al., 1995). Cells which bind MHC class I molecules maintain expression of CD8 and stop expressing CD4, whereas cells that bind MHC class II molecules maintain expression of CD4 and stop expressing CD8. The developing T-cells will be located in distinct anatomical positions based on their developmental stage. Absence of MHC–DP cell interaction leads to death by neglect. T-cells that received enough signal move inside the medulla, where negative selection takes place to eliminate cells that over-express both CD4 and CD8 molecules which could attack self-peptides (Palmer, 2003).

1.2 THYMUS ORGANOGENESIS

During early organogenesis, the thymic gland originates from the endoderm of the 3rd pharyngeal pouch (3rd PP), from the same primordium that gives rise to the parathyroid gland (shown in **Figure 1.1**). Around embryonic day 10 (E10), cells from the 3rd PP start proliferating, leading to the formation of the bilateral primordia. Mesenchymal cells in the area, derived from a transient neural crest (NC) population that has migrated into the pharyngeal region at E9.5 and populated the pharyngeal arches, surround both primordia and expand similarly to the epithelial component until approximately E12.5. These cells will eventually comprise the capsules that surround the thymus (reviewed in Rodewald, 2008). At E12.5, both organ structures separate from the pharynx and migrate to their final anatomical locations; the midline and laterally of the thyroid respectively. Critical to this early separation of the thymus and the parathyroid structures is the expression of the glial cells missing-2 (*Gcm2*) transcription factor (TF), whose expression is apparent as early as E9.5 in mouse development, setting the barrier between the parathyroid and the thymic cell fate (Gordon et al., 2001).

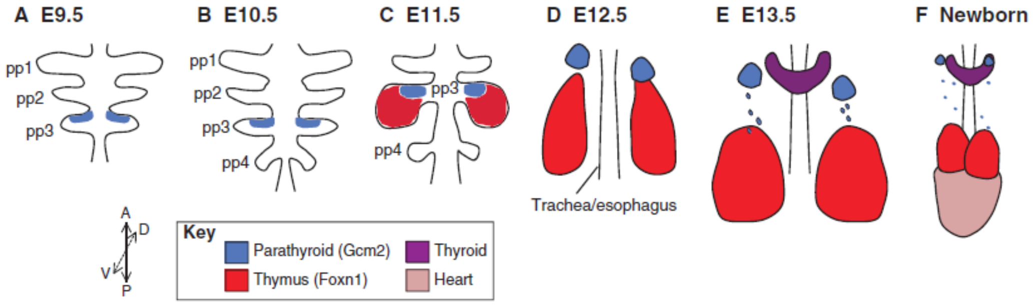

A E9.5
B E10.5
C E11.5
D E12.5
E E13.5
F Newborn
pp1
pp2
pp3
pp1
pp2
pp3
pp4
pp3
pp4
Trachea/esophagus
A
D
V
P
Key
Parathyroid (Gcm2)
Thymus (Foxn1)
Thyroid
Heart

Figure 1.1: An overview of early thymus organogenesis. A) Representation of the pharyngeal pouches (pp) structures that are aligned sequentially on the flanking structures of the foregut, with the 3rd pair of pouches (pp3) to appear by E9.5. *Gcm2* expression that marks the parathyroid region is already evident from E9.5 (shown in blue). **B)** Appearance of the 4th pouch by E10.5. **C)** Development of the pouches into thymus (red colour) and parathyroid (blue colour) primordia by E11.5; organs are identified by expression of *Foxn1* and *Gcm2* respectively. **D)** Detachment of the primordia from the pharynx by E12.5. **E)** Total separation of the parathyroid buds from the thymus, although they stay adjacent to the thyroid (purple); few parathyroid cells remain attached to the thymus. **F)** Organs in final positions, in relevance to the heart (shown in pink). Double-headed arrows indicate orientation: Dorsal (D), Ventral (V), Anterior (A) and Posterior (P). Image reproduced from Gordon and Manley, 2011.

1.2.1 Origin of thymic epithelial cells

Despite early controversy over the origin of TECs suggesting both endoderm and ectoderm as potential germ layers from which TECs may derive (Cordier and Haumont, 1980; Cordier and Heremans, 1975), cell transplantation experiments (in both chicken and mouse) verified the unique endodermal origin of TECs (Gordon et al., 2004; Le Douarin and Jotereau, 1975). Transplantation of quail pharyngeal endoderm into chick hosts generated a chimeric thymus, where TECs were of donor origin, while mesenchymal tissue and T-cells were of host origin (Le Douarin and Jotereau, 1975). Additionally, to study the potency of thymic epithelial cells in murine, pharyngeal tissue or reaggregated thymic organ culture (RTOC) was transplanted under the kidney capsule (Anderson et al., 1993; Blackburn and Manley, 2004; Rodewald et al., 2001; Sheridan et al., 2009). When the E9.0 3rd PP endoderm was dissected and grafted, it gave rise to an ectopic thymus characterised by normal corticomedullary regions, capable of supporting thymocyte development. The latter experiment established a sole endodermic origin of TECs (Gordon et al., 2004). Importantly, this finding suggested that a subset of cells in the 3rd PP endoderm had inherited the thymic fate, even though thymic definitive markers (for instance, *Foxn1*) were not yet switched on.

1.2.2 Specification and differentiation of thymic epithelial cells in the embryonic thymus

Although the molecular mechanisms that establish TEPC fate in the 3rd PP endoderm are not yet fully defined, previous studies have suggested that early patterning of the thymic epithelium can be influenced by Paired Box 3 (PAX3) and Sonic Hedgehog (SHH) (Griffith et al., 2009; Moore-Scott and Manley, 2005). Formation of a thymic rudiment in mice lacking the *nude* product demonstrated that TEC specification is independent from expression of the *nude* gene, nevertheless, subsequent differentiation of TECs, including assembly of the complex TEC architecture and T-cell precursor recruitment and differentiation, are essentially dependent on its expression (Bleul et al., 2006; Corbeaux et al., 2010; Vaidya et al., 2016). The *nude* mutant phenotype (Flanagan, 1966) is characterised by a developmental arrest in early thymus organogenesis and absence of thymopoiesis (Cordier and Haumont, 1980; Pantelouris, 1968; Pantelouris and Hair, 1970). Further experiments, including positional cloning and targeted gene disruption identified FOXN1 as the protein encoded by the *nude* gene product and the *nu* mutation to be a single nucleotide deletion located in the third exon (exon3) of the *Foxn1* locus (Nehls et al., 1994, 1996). Moreover, thymus – bone marrow cross-transplantation studies between *nude* and *wild type* mice confirmed that stromal rather than haematopoietic defects caused the thymic phenotype (Wortis et al., 1971). Further, analysis of *nude – wild type* chimeras identified *nude* TECs incorporated into the thymic epithelium of normal mice to be still expressing immature markers, indicating that FOXN1 promotes TEC differentiation in a cell-autonomous way (Blackburn et al., 1996). Significantly, TECs that lack FOXN1 expression maintain a progenitor potency that allows them to form again a patterned and functional thymus, as demonstrated by quasi-clonal activation of *Foxn1* in these cells postnatally (Bleul et al., 2006).

A number of other genes, namely Homeobox A3 (*Hoxa3*), Eyes Absent (EYA) Transcriptional Coactivator And Phosphatase 1 (*Eya1*), SIX Homeobox 1 (*Six1*), Paired Box 1 (*Pax1*), Paired Box 9 (*Pax9*) and T-Box 1 (*Tbx1*), which act prior to FOXN1 expression, have also been found to cause abnormalities in thymus development (for example, hyperplasia, athymia, faulty lobe migration). However, these defective phenotypes may also be due to the absence of these genes from the other germ layers or earlier developmental points (Conway et al., 1997; Dietrich and Gruss, 1995; Hetzer-Egger et al., 2002; Jerome and Papaioannou, 2001; Laclef et al., 2003; Peters et al., 1998; Su and Manley, 2000; Su et al., 2001; Wallin et al., 1996; Xu et al., 2002; Zou et al., 2006). This group of genes collectively suggests that there is a genetic regulatory network that acts upstream of TEC specification; their role is extensively described in the next section.

1.2.2.1 A genetic regulatory network acting prior to TEC specification

To better understand this early-acting genetic regulatory network, the role of the genes expressed prior to *Foxn1* is discussed further here. *Hoxa3* is evident in both the endoderm and surrounding mesenchyme of the pharyngeal region – foregut, while newborn mice lacking this gene are missing the thymus, parathyroid and part of the thyroid gland (Manley and Capecchi, 1995). *Foxn1* expression was delayed (but still initiated) in *Hoxa3*$^{-/-}$ mice, suggesting that *Hoxa3* alone cannot lead specification of the thymic epithelium (Chojnowski et al., 2014). Nonetheless, the lack of thymus could be explained by increased apoptosis in those mice. Specific deletion of *Hoxa3* in the endoderm caused a more severe phenotype compared to deletion in the mesenchyme due to *Bmp4* endoderm-specific dependence on *Hoxa3* (Chojnowski et al., 2014).

PAX1 and PAX9 are highly homologous transcription factors (Blake and Ziman, 2014) that are commonly expressed in the 3rd PP at E10.5, however, their knockout phenotypes differ substantially (Peters et al., 1998; Wallin et al., 1996). *Pax9*$^{-/-}$ mice display severe 3rd – 4th PP growth defects at E11.5 and their thymic tissue remains stuck in the neck region (absence of migration), however they still express *Foxn1* and initiate thymocyte colonization that lasts only up to E16.5 due to pervasive apoptosis (Hetzer-Egger et al., 2002). *Pax1*$^{-/-}$ mice display normal migration, with a mild delay in growth and thymocyte transition (Wallin et al., 1996). The maintenance of *Pax1* expression is dependent on *Hoxa3* expression (Manley and Capecchi, 1995); however, *Hoxa3* expression is not affected in *Pax1 Pax9* double knockouts (Zou et al., 2006), which suggests an upstream regulatory role for Hoxa3 compared to the above PAX genes. Additionally, loss of one *Hoxa3* allele (*Hoxa3*$^{+/-}$) in combination with a *Pax1* mutant (*Pax1*$^{-/-}$) leads to more severe aberrations, indicating a HOXA3-PAX1 partially regulating thymus organogenesis (Su and Manley, 2000; Su et al., 2001).

Expression of *Eya1* and *Six1* is evident in the pharyngeal pouch and afterwards in the fetal thymus (Xu et al., 2002; Zou et al., 2006). *Eya1*$^{-/-}$ mice demonstrate a complete lack of the thymus-parathyroid primordium, while *Six1*$^{-/-}$ mice (or *Six1*$^{-/-}$ *Six4*$^{-/-}$ double knockout) initially form a 3rd PP rudiment, which rapidly degenerates via apoptosis (Zou et al., 2006). *Eya1*$^{-/-}$ *Six1*$^{-/-}$ double mutants also exhibit a more severe phenotype than *Eya1*$^{-/-}$ alone (Xu et al., 2002; Zou et al., 2006). However, since EYA1 cannot bind DNA (Rebay et al., 2005), it is possible that EYA1 regulation of SIX1 may be through an alternative mechanism (for instance via EYA1's phosphatase activity).

DiGeorge syndrome (DGS) is caused by the deletion of the 22q11.2 chromosomal region and among other symptoms, the immune system of the patients is compromised because of thymus hypoplasia or aplasia (Holländer et al., 2006). In mice, knockout of the *Tbx1* locus causes a very similar phenotype to DGS, verifying that deletion of *Tbx1* is largely responsible for aberrations (Jerome and Papaioannou, 2001; Lindsay et al., 2001). Expression of *Tbx1* is apparent in the pharyngeal region (endoderm of the PP and arch mesenchyme; Lindsay et al., 2001), and dynamic regulation of TBX1 is necessary to succeed normal thymus organogenesis since both *Tbx1* knockout (Arnold et al., 2006) and gain-of-function mutants (Vitelli et al., 2009) led to aplasia and hypoplasia respectively. TBX1 was shown to specifically repress *Foxn1*, as demonstrated from constitutive expression of *Tbx1* under the control of the *Foxn1* promoter (*Foxn1*^{Cre}), thus TEC differentiation can only proceed by *Tbx1* down-regulation (Reeh et al., 2014). Overall, the above studies suggest a temporary role for *Tbx1* in PP patterning and morphogenesis regulation and that happens partially through *Fgf8* and *Pax9* activation (Arnold et al., 2006; Okubo et al., 2011).

It is likely that a genetic regulatory network orchestrates patterning and morphogenesis of the thymic anlage and this happens through cross-interaction of this network's components. Unlike the phenotype in *nude* mice, in which thymic rudiments are formed and survive into adulthood, most of these genes knockouts (except for *Pax1*) cause degeneration of the thymic primordia through apoptosis. Therefore, two phases are evident during early progression of TECs: a) an early FOXN1-independent phase, in which a fine-tuned network of factors drives the cells into a stable and apoptosis-resistant state ensuring TEC specification and survival and b) a later FOXN1-dependent phase, in which FOXN1 activation mediates TEC differentiation into the functionally distinct cTEC and mTEC populations and allows further proliferation of these cells.

1.2.2.2 Bipotent and lineage-fated thymic epithelial progenitor cells

The early thymic primordium is populated by thymic epithelial progenitor cells (TEPCs) that can differentiate into the functionally distinct cortical and medullary TEC lineages and support the development of an adequate immune system (Bennett et al., 2002). Determination of the TEPC phenotype and isolation of these cells are therefore of major importance. Evidence for the existence of such a progenitor cell type initially emerged from studying human thymic epithelial tumours where few cells were able to express both cTEC and mTEC markers (Schluep et al., 1988). A more definitive TEPC phenotype was determined by a study addressing the role of FOXN1 in mice. As described above, this study demonstrated that cell-autonomous FOXN1 expression is required for TEPCs to differentiate towards the cortical and medullary TEC lineages (Blackburn et al., 1996). In more detail, analysis of adult aggregation chimeras from *wild type* and *nude* cells identified few *nude*-like cells marked by MTS20 and MTS24 monoclonal antibodies (Abs), and absence of any differentiation markers. MTS20 and MTS24 Abs (Depreter et al., 2008) target the Placenta-expressed Transcript 1 (PLET1) antibody, which has been identified by two independent laboratories as a fetal TEC marker capable of reconstituting a full thymus organ in RTOC grafts (Bennett et al., 2002; Gill et al., 2002). EpCAM$^+$PLET1$^+$ TEPCs at E12.5 of embryonic development were all identified to be capable of generating both cTECs and mTECs (designated bipotent) through a single cell transplantation assay (Rossi et al., 2006). Additionally, improved flow cytometry methods identified all E12.5 TEPCs to be PLET1$^+$, with the PLET1$^+$ expression to subsequently decrease in organogenesis (Cook, 2010).

A later block in differentiation due to differentiation arrest of thymocytes generated cells that co-expressed K8 and K5 markers which normally reside in the CMJ, while transplantation of these cells to mice with a later block showed their conversion into K8$^+$K5$^-$ cells, suggesting their role as cTEC progenitors (Klug et al., 1998). Also recently, a self-renewing subpopulation of embryonic TECs that expresses high levels of Claudin 3 (*Cldn3*), *Cldn4* and stage-specific embryonic antigen (SSEA1) was able to generate mTECs in the long-term and this subset of TECs has been characterised as mTEC stem cells (Sekai et al., 2014). Importantly, cells identified in the *nude* primordium express a similar set of markers (Baik et al., 2016), which suggests that FOXN1 is indispensable for TEC fate selection (Nowell et al., 2011). Another embryonic progenitor mTEC population that is able to give rise to AIRE$^+$ mTECs was shown to express RANK, a key regulator in the formation of the thymus medulla (Baik et al., 2016).

Despite current progress, the bipotent phenotype of specified TEC progenitors that populate the thymic anlage before the onset of FOXN1 has not yet been described in full to allow confident identification and purification of this bipotent TEPC population.

1.2.2.3 A serial progression model into the embryonic cortical and medullary thymic epithelial cell lineages

In the adult thymus, cTEC and mTEC populations are easily distinguishable based on the differential expression of a panel of markers (intracellular and cell surface ones) (Lucas et al., 2016; Ohigashi et al., 2013). However, this is not the case in the embryonic thymus, where current TEC markers fail to clearly define the cTEC-mTEC early populations, setting a barrier in the analysis of the early cTEC and mTEC lineage emergence (Klug et al., 2000, 2002). In a recent study, the cTEC-specific marker CD205 and the mTEC-specific marker CD40, which normally identify these populations in the adult thymus, were used to assess early cTEC and mTEC development (Shakib et al., 2009). This study identified a $CD205^{+}CD40^{-}$ population that was apparent around E12.0-13.0, while over time some cells progressively became $CD205^{+}$ $CD40^{+}$ and eventually $CD205^{-}CD40^{+}$. Interestingly, a purified $CD205^{+}CD40^{-}$ population was capable of giving rise to both cTEC and mTEC lineages through an RTOC transplantation assay, suggesting that mTEC lineages experience an early phase of cTEC markers expression (Baik et al., 2013). In that direction, fate mapping data clearly demonstrated that all TECs experience expression of the cTEC marker *β5t* (Ohigashi et al., 2013).

In particular for the mTEC lineage progression, $SSEA1^{+}$ mTEC stem cells ontogenic appearance was examined against the key regulator of mTEC formation, RANK (Akiyama et al., 2008; Hikosaka et al., 2008; Rossi et al., 2007a). The $SSEA1^{+}$ mTEC stem cells was found to express RANK uniformly and also emerge earlier than the $RANK^{-}$ mTEC progenitor population (Baik et al., 2016). Interestingly, $SSEA1^{+}$ mTEC stem cells are evident in the embryonic thymus of mice which exhibit an mTEC developmental block (*Relb*$^{-/-}$ mice), whereas no $RANK^{+}$ mTEC progenitors were detected (Baik et al., 2016). Collectively, these studies agree on mTEC specification having already occurred by E12.0. In brief, mTEC lineage progression generates

SSEA1$^+$ mTEC stem cells in an RELB-independent way, however, RELB is required downstream of this stage to produce RANK$^+$ mTEC progenitors which can also act as a subsequent source for mature AIRE$^+$ mTECs generation (Akiyama et al., 2008). The relationship between the mTEC stem cells and the mTEC progenitors is not yet known, however it seems that multiple members of the Tumour Necrosis Factor family (TNFR) take part in mTEC serial differentiation process.

Collectively, the above findings are significant, as they propose that TEPCs are initially cTEC-fated and then progressively acquire their mTEC lineage in a stepwise manner, with mTEC sublineage differentiation to also proceed gradually through premature mTEC populations towards the mature mTEC phenotypes (see TEC differentiation diagram in **Figure 1.2**). Nonetheless, the transition from the bipotent progenitor stage towards these phenotypes has not yet been deciphered.

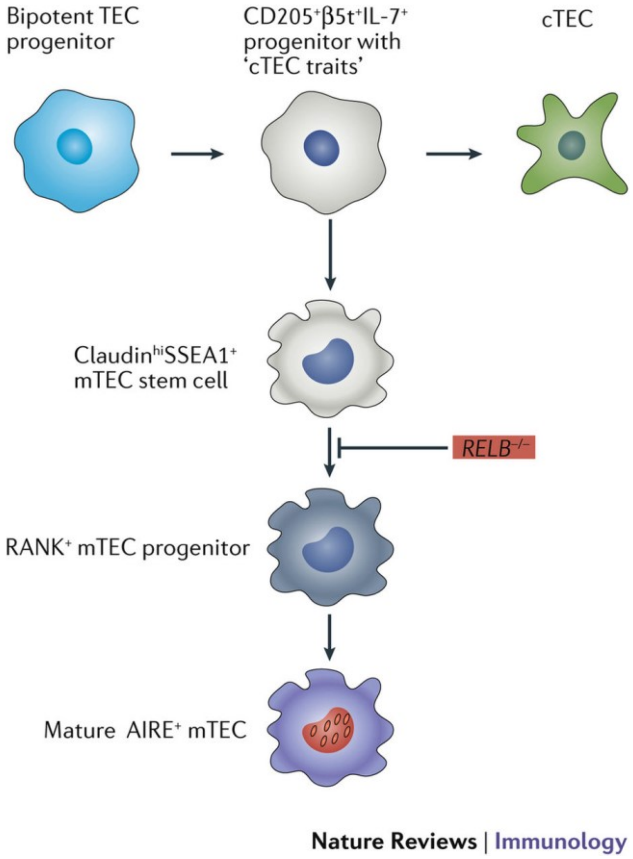

Bipotent TEC progenitor
CD205⁺β5t⁺IL-7⁺ progenitor with 'cTEC traits'
cTEC
Claudin^hi SSEA1⁺ mTEC stem cell
RELB−/−
RANK⁺ mTEC progenitor
Mature AIRE⁺ mTEC
Nature Reviews | Immunology

Figure 1.2: A serial progression model of TEC early progression in development. The diagram depicts a summarised overview of early TEC differentiation from the bipotent progenitor stage towards the cortical (c) and medullary (m) TEC lineages based on current literature analysing the embryonic thymus. Early undifferentiated bipotent progenitors (TEPCs) undergo a stage where they express cTEC-like genes before proceeding towards the cTEC or mTEC distinct subpopulations. An unknown mechanism determining mTEC specification results in the generation of SSEA1$^+$ mTEC stem cells, which are subsequently dependent on RELB for further differentiation and seem to also involve the mTEC regulator RANK. Finally, crosstalk among the RANK$^+$ mTEC precursors sets off differentiation towards AIRE$^+$ mTECs. Image reproduced from Takahama et al., 2017.

1.2.2.4 Cell-crosstalk dependence for TEC progression

TEC differentiation and maturation depends on extracellular signalling from the neural crest (NCC)-derived mesenchyme and the developing thymocytes which migrate in the thymic rudiment during organogenesis (Auerbach, 1960; van Ewijk et al., 1994). The thymic lobes are surrounded by a NCC-derived mesenchymal population that stimulates proliferation of immature TECs by secretion of FGF factors (Jenkinson et al., 2003; Revest et al., 2001). Around E11.5, haematopoietic progenitors begin to colonise the thymus (Luis et al., 2016) coinciding with the onset of FOXN1 and the morphological shift of thymic epithelial cells. In transgenic mice where thymopoiesis is blocked (hCD3ε; block at DN1 stage), TECs maintain a two-dimensional (2D) configuration, with the predominant TEC population to be $K5^+$ $K8^+$, while thymi were hypoplastic with large cysts (van Ewijk et al., 2000; Klug et al., 1998). It is possible that the presence of haematopoietic progenitors is required to maintain TEC organisation, since T-cell progenitors are not necessary for the initial stages of fetal epithelium patterning (Klug et al., 2002). A subsequent block in T-cell development using *Rag1*$^{-/-}$ mutants (block at DN3 stage) gave rise to thymi with reticular cTEC organisation and a detectable K5-K8+ population, although bare detection of mTECs was apparent (Klug et al., 1998). Indeed, presence of more mature T-cells is essential for mTEC sublineage elaboration (Hikosaka et al., 2008). Together, these results demonstrate that TEC differentiation after the initial specification stage encompasses several checkpoints and necessitates cell non-autonomous interactions to progress.

1.2.3 Signalling mechanisms involved in early thymus progression

1.2.3.1 Sonic Hedgehog (SHH)

The role of SHH signalling in early development has been investigated through *Shh*$^{-/-}$ embryos, in which disturbed separation of the thymus-parathyroid structures was observed. In the *Shh*$^{-/-}$ mutants, the overall pouch domain was diminished due to cell death, while subsequent loss of the *Gcm2* expression resulted in an expanded thymus domain marked by expanded expression of *Bmp4* and *Foxn1* (Moore-Scott and

Manley, 2005). Additionally, in experiments where *Shh* was constitutively expressed in the 3rd PP, cells did not acquire a parathyroid fate, but instead TBX1 was further induced and in turn repressed *Foxn1* expression in the thymic domain (Bain et al., 2016; Reeh et al., 2014), suggesting that SHH expression acts as a control gate of the thymic-parathyroid fate, rather than imposing it.

Following the initial patterning of the 3rd PP, expression of *Shh* is also evident during TEC differentiation. Subsequent investigation of the *Shh*$^{-/-}$ mutants has been difficult since development for these embryos fails around E16.5 (Shah and Zuniga-Pflucker, 2014). Analysis of fetal culture of *Shh*$^{-/-}$ thymi or of chemically blocked SHH signalling demonstrated lower number of cTECs and mTECs, increased levels of MHC class II expression and reduced numbers of AIRE^{+} mTECs (Saldaña et al., 2016). Collectively, the above findings propose a role for SHH in TEC differentiation, although it is not clear if these effects are the outcome of cells acting autonomously or if they have been partially caused indirectly through lymphocyte crosstalk, since *Shh* also regulates thymocyte development (Barbarulo et al., 2016).

1.2.3.2 Bone Morphogenic Proteins (BMPs)

During early organogenesis of the thymus *Bmp4* shows a dynamic spatiotemporal expression pattern, by which *Bmp4* expression is restricted in the few mesenchymal cells of the 3rd pharyngeal arch at E9.5. Its expression it then continues to expand at the ventral aspect of the 3rd PP at E10.5-E11.5, and finally its expression is evident throughout the primordium and the mesenchymal capsule (Patel et al., 2006). Therefore, it is possible that *Bmp4* is involved in the specification of the thymic epithelium starting from the ventral posterior aspect of the 3rd pouch. Some *in vitro* evidence has suggested that *Foxn1* expression in TECs may be induced by *Bmp4*, *Foxn1* expression in the posterior area of the 3rd pouch at E11.25 (Tsai, 2003). In agreement with this, expression of the BMP inhibitor NOGGIN (normally expressed in a complementary pattern to that of *Bmp4*) under the control of *Foxn1* disrupts thymus migration, while the existent thymus is heavily reduced (Bleul and Boehm, 2005). Additionally, sustained expression of NOGGIN caused partial loss of *Foxn1* in the thymic primordium which resulted in domain reversion to a cystic structure, indicatory of a more immature state (Soza-Ried et al., 2008), while specific deletion of *Bmp4* as early as E9.5 (*Foxg1^{Cre}*) also resulted in a similar phenotype (Gordon et al., 2010). Finally, specific deletion of *Bmp4* in committed TECs and the NC-derived surrounding mesenchyme (using *Foxn1^{Cre}* and *Wnt1^{Cre}*, respectively) demonstrated normal migration of the structure but again reduced size, suggesting a continuous requirement of BMP signalling after *Foxn1* initiation, a requirement no longer needed though for the later stages of the mesenchyme.

In general, BMP signalling seems to up-regulate *Foxn1* expression in TECs as shown by addition of BMP4 in FTOC medium (Tsai, 2003). In our hands, only combinatorial addition of FGF8 and BMP4 in 3rd PP culture resulted in *Foxn1* up-regulation (Poppis, Blackburn lab, unpublished). However, neither *Foxg1^{Cre}* induced *Bmp4* nor deletion or forced *Noggin* expression completely abolished *Foxn1* expression, although in the latter case *Foxn1* expression was lost from part of the thymic rudiment and the rudiment itself exhibited a cystic structure (Bleul and Boehm, 2005; Gordon et al., 2010; Soza-Ried et al., 2008). Interestingly, suppression of BMP signalling by NOGGIN activates *Bmp2* and *Bmp4* in a negative feedback fashion (Bleul and Boehm, 2005). This feedback loop, as well as potential redundancy among BMP factors, could explain the small impact on *Foxn1* expression in the *Bmp4* deletion model using the *Foxg1^{Cre}* (Soza-Ried et al., 2008). In summary, these data suggest a potential role for the BMP signalling in *Foxn1* up-regulation, although this is possibly dependent also on other signalling pathways acting in the region and TECs maturation status itself.

1.2.3.3 Fibroblast Growth Factors (FGFs)

FGF signalling is involved in the development of several structures in the pharyngeal region, with mutations disrupting the normal FGF signalling to affect normal organogenesis in the thymus and/or the parathyroid system. High expression of *Fgf8* is apparent in the distal posterior presumptive thymus region around E9.5. FGFR2-IIIb expression is limited in the thymic epithelium and its expression is first evident at E13.5, while its ligands (FGF3, FGF7 and FGF10) are expressed in the mesenchymal tissue surrounding the developing thymus. Aplasia or hypoplasia is observed in *Fgf8* mutants because they fail to form normal 3rd PPs (Abu-Issa et al., 2002; Frank et al., 2002), while reduced TEC proliferation occurs after E12.5 in FGFR2-IIIb and FGF10 mutants (Ohuchi et al., 2000; Revest et al., 2001). Importantly, continuous localised inhibition of the FGF signalling by the sprouty genes (*Spry1* and *Spry2*) is also necessary for the formation of a normal-sized parathyroid (via *Gcm2* induction), thymus-parathyroid separation (via apoptosis) and expansion of the thymic primordium (via cell proliferation) (Gardiner et al., 2012). Overall, FGF signalling regulates early thymus organogenesis in two stages: a) rudiment patterning and morphogenesis through an *Fgf8* negative loop and b) TEC proliferation through a mesenchymal-epithelial crosstalk.

1.2.3.4 Wingless-type MMTV integration site family members (WNTs)

Several WNT ligands are expressed in the thymic anlage during early organogenesis (around E10.5), however only WNT4 seems to persist in adult development (Balciunaite et al., 2002). A critical role for WNT4 as a *Foxn1* inducer has been identified by using cTEC-derived lines in which both co-culture with WNT4-transfected T-cells and WNT activation by lithium chloride led to an increase in *Foxn1* expression. In support to this finding, a later study co-cultured E15.5 thymi with supernatant from a WNT4-secreting cell line; analysis of isolated TECs demonstrated *Foxn1* up-regulation as well as up-regulation of the TEC differentiation markers, *Il7* and *MHC II* (Kvell et al., 2014). Furthermore, analysis of *Wnt4*$^{-/-}$ thymi by Heinonen et al. (2011) demonstrated decreased TEC cellularity, an effect that was stronger if the deletion occurred in neonatal, rather than adult mice. In agreement with this result, specific knockout of *Grp16* (transporter of WNT ligands to the cell surface) in TECs reduced thymocyte and TEC cellularity (Brunk et al., 2015). Interestingly, elevated activity of WNT signalling by loss of the WNT antagonist *Kremen1*, gradually disrupted the corticomedullary compartmentalisation (Osada et al., 2006). A more dramatic disruption of the thymic structure and loss of *Foxn1* resulted from two independent studies where the WNT effector β-catenin (Swann et al., 2017; Zuklys et al., 2009). In addition, up-regulation of WNT4 alone demonstrated milder effects, suggesting that a negative loop may be regulating WNT4 expression (Swann et al., 2017). Collectively, the above results highlight a potential role of WNT4 in *Foxn1* activation/up-regulation and dynamic expression of WNT signalling is essential for normal thymus development.

Understanding the genetic network and signalling mechanisms that orchestrate TEPC early specification and expansion in early organogenesis will enable improved culturing conditions for survival and proliferation of TEPCs *in vitro* with significant effects in regeneration.

1.2.4 The role of FOXN1 in thymus development

1.2.4.1 FOXN1-independent TEC specification

Foxn1 expression is evident in low levels during the initial outpocketing of the 3rd PP (E9.5) (Nehls et al., 1994), while high levels of *Foxn1* become evident from E11.25 onwards, with this strong expression being apparent in the most ventral part of the 3rd PP and gradually expanding to cover the entire thymus organ (Gordon et al., 2001, reviewed in Vaidya et al., 2016). As discussed above, formation of the thymic primordium does not require FOXN1 (established by histological analysis, Cordier and Haumont, 1980; Cordier and Heremans, 1975). In agreement with this, several studies strongly support that FOXN1 does not determine the thymic epithelial lineage specification, and their findings are summarised in this paragraph. E9.0 3rd PPs ectopic transplantation (prior to *Foxn1* expression) is capable of giving rise to a functional thymus organ with evident cTEC and mTEC compartments, denoting that 3rd PP cells have already been specified towards the TEC fate (Gordon et al., 2004). In addition, expression of genes that are specific markers of the thymic domain (such as the interleukin 7; *Il7*, and the Forkhead transcription factor g1; *Foxg1*) occurs independently of FOXN1 (Wei and Condie, 2011; Zamisch et al., 2005). Finally, the developmental arrest that occurs to TEPCs with *Foxn1* revertible *null* or subfunctional hypomorphic alleles can be reversed at later stages, denoting that fetal TEPCs can exist in a stable state *in vivo* (Blackburn et al., 1996; Bleul et al., 2006; Jin et al., 2014) and therefore suggesting the existence of an active transcriptional network acting upstream of FOXN1 that establishes TEPC identity and mediates TEC specification. Taken together, the above evidence indicates that TEC lineage commitment is not dependent on FOXN1, but potentially relies on a yet unexplored genetic network.

1.2.4.2 FOXN1-dependent TEC differentiation

Previous literature has demonstrated that TEC differentiation (and proliferation), including formation of the three-dimensional (3D) TEC microarchitecture are critically dependent on FOXN1 expression (reviewed in Vaidya et al., 2016). In one of the studies, homozygous mice for a hypomorphic *Foxn1* allele (*Foxn1*$^{\Delta}$; lacking the FOXN1 N-terminal domain) were generated and analysis of the resulting thymi demonstrated thymus hypoplasia; the cTEC and mTEC regions failed to separate, while additionally these abnormal thymi could only support aberrant thymopoiesis in adult mice (Su et al., 2003). The above observed phenotypes indicated that TEC differentiation was initiated but a subsequent block occurred leading to defects in the adult thymus. With the use of a different *Foxn1* hypomorphic allele (*Foxn1*^{R}), Nowell et al. (2011) investigated the impact of variable FOXN1 levels on TEC differentiation. *Foxn1*^{R} expresses around 15% of normal *Foxn1* transcript levels compared to the wild type (WT) *Foxn1* allele and mice homozygous for *Foxn1*^{R} (*Foxn1*$^{R/R}$) were characterised by hypoplastic thymi and suboptimal T-cell development. Examination of a *Foxn1* Allelic series which was generated based on the *Foxn1*^{WT}, *Foxn1*^{R} and *Foxn1*$^{-}$ alleles revealed *Foxn1* dose-dependent TEC differentiation, in summary showing that increasing levels of *Foxn1* control both cTEC and mTEC exit from the TEPC stage and progression through intermediate stages to terminal differentiation in the fetal and adult thymus (Nowell et al., 2011). Of note is that another study using a revertible *Foxn1* allele (*Foxn1*^{SA2}), in which *Foxn1* could be reactivated in single cells by tamoxifen induction (Cre-ERT2 system), generated miniature thymi with well-defined cortical and medullary compartments, containing both Cytokeratin 5-high, Claudin 4-high (K5^{hi}Cldn4^{hi}), and K5^{-}Cldn4$^{\mathrm{lo}/-}$ populations (Bleul et al., 2006). The outcome of this analysis, in combination with published data showing that mTEC-restricted progenitors (mTEPCs) during thymus organogenesis are *Cldn4*^{hi} (Hamazaki et al., 2007; Sekai et al., 2014), suggests that the appearance of the mTEC lineage may be independent from *Foxn1* expression and therefore that the divergence of the cTEC and mTEC populations may happen earlier than indicated by Nowell et al. in the lines above.

In summary, several lines of evidence described above have proved the importance of FOXN1 in TEC progression. Further literature has also revealed a role for FOXN1 in managing the attraction and differentiation of haematopoietic progenitors in the thymic rudiment, as well as in the regulation of neural crest cell (NCC) migration and maturation that will shape the thymic mesenchyme. Altogether, these data prove FOXN1 high significance in the establishment of a functional thymic system.

1.2.4.3 FOXN1 binding motif

Despites FOXN1's (product of the *nude* locus) indisputable role in the thymic system, the molecular function and the genetic network around this transcription factor have not yet been determined in completeness, while only recently direct targets of FOXN1 and the FOXN1 binding motif were identified in TECs by a chromatin immunoprecipitation assay combined with high throughput DNA sequencing (ChIP-seq) (Žuklys et al., 2016). The very first prediction for a FOXN1 binding motif came from an *in vitro* selection of binding sites (SELEX), in which the oligo-bound sequences contained an identical tetranucleotide core subsequence, 5'-ACGC-3' (Schlake et al., 1997). Following this, a study investigating the evolution of FOX factors used a protein-binding microarray (PBM; Berger and Bulyk, 2009; Berger et al., 2006) assay to identify preferred bound sequences by the different FOX family members (Nakagawa et al., 2013). This competitive study identified an almost identical sequence, 5'-GACGC-3', as the strongest bound sequence by FOXN1, while it also suggested that FOXN1 had lost over time the ability to bind the canonical FOX motif (5'-AAACAA-3'). Even though two separate studies had predicted the same core nucleotide sequence as the preferred FOXN1 binding motif, both of them relied on experiments using short oligonucleotide binding in an artificial environment. Only recently, the FOXN1 binding motif was confirmed by mapping the FOXN1-DNA interactions *in vivo* on a genome scale (Žuklys et al., 2016).

Žuklys et al. expressed a FOXN1 protein, tagged with an octapeptide, under the *Foxn1* promoter in TECs of *nude* mice using a bacterial artificial chromosome (BAC). Mice homozygous for the BAC (designated *Foxn1*$^{wt*/wt*}$) demonstrated a normal coat and thymus architecture, as well as normal *Foxn1* expression levels with a mildly reduced cellularity which was still capable of supporting thymocyte development. *De novo* motif analysis of the consistently enriched FOXN1 peaks around the proximal transcription start site of all expressed genes (5 kilobases (kb) upstream and 3kb downstream of genes TSSs) predicted 5'-GACGC-3' to be the FOXN1 binding motif, in full agreement with the above *in vitro* studies, further validating this consensus as the preferential FOXN1 binding sequence.

1.2.4.4 FOXN1 target genes

In terms of gene regulation in TECs, *Foxn1* has been shown to be generally required for expression of proteins that play a vital role in promoting T-cell development (Bredenkamp et al., 2014a; Calderón and Boehm, 2012; Chen et al., 2009; Nowell et al., 2011). For instance, the C-C Motif Chemokine Ligand 25, C-X-C Motif chemokine ligand 12 (CXCL12) chemokine is necessary for the attraction of thymic seeding progenitors in both fetal and adult thymus (Liu, 2005; Liu et al., 2006; Plotkin et al., 2003), while Delta like canonical NOTCH ligand 4 (DLL4) allows commitment of migrating hematopoietic precursors to the T-cell lineage (Hozumi et al., 2008; Koch et al., 2008) and KIT Ligand (KITL) is required for T-cell survival and proliferation (Buono et al., 2016). Furthermore, Proteasome subunit beta 11 (PSMB11; also known as β5t) and Cathepsin L (CTSL) regulate peptide production in TECs to positively select CD8^{+} and CD4^{+} T-cells respectively (Honey et al., 2002; Murata et al., 2007; Nakagawa et al., 1998; Sasaki et al., 2015), while MHC Class II expression contributes in peptide selection specifically for the CD4^{+} T-cells. Additionally, FOXN1 is also regulating genes involved in differentiation, proliferation and function of TECs, namely, the Paired Box 1 (*Pax1*), Tumor Protein P63 (*Trp63*), Cyclin D1 (*Ccnd1)*, Cluster of differentiation 40 (*CD40*), Cluster of differentiation 80 (*CD80*), Fibroblast growth factor receptor 2 isoform IIIB (*Fgfr2IIIb)*, Autoimmune regulator (*Aire*), as well as some WNT regulators (Bredenkamp et al., 2014a; Nowell et al., 2011). Finally, the FOXN1-tagged ChIP-seq analysis by Žuklys et al. (2016) has identified 450 highly confident FOXN1 direct target candidates, while it has also experimentally verified for the first time in neonatal mice that Cluster of differentiation 38 (*Cd38*) and *Psmb11* are directly regulated by FOXN1 in cTECs. Suggested from previous literature (Nowell et al., 2011), the usual FOXN1 candidate targets (*Ccl25*, *Dll4*, *Cxcl12*) are also included in this list, strongly suggesting that they are directly regulated by FOXN1.

FOXN1 is now regarded as the master regulator in the thymic system, since its continuous expression is required for differentiation and maintenance of the cortical and medullary TEC lineages as well as subsequent thymocyte maturation, processes indispensably linked to a functional immune system. In 2016, a FOXN1-tagged ChIP-seq analysis finally provided more insights into FOXN1 direct binding potency in TECs *in vivo*, further identifying highly confident FOXN1 direct target candidates, specifically in the thymic cortex of 1 week old mice. The same study also proposed a critical role for FOXN1 in the maintenance of the antigen processing and presentation program in newborn cTECs based on these data. Nevertheless, how this single transcription factor establishes from scratch a robust transcriptional network that orchestrates thymus organogenesis and function in the early residing TEPCs or (as recently demonstrated) in an unrelated cell type (Bredenkamp et al., 2014b) still remains to be elucidated. It also remains unclear how FOXN1 initiation is regulated by the early genetic regulatory network that is acting in the undifferentiated thymic anlage.

1.2.5 The role of FOXN1 in skin

Expression of *Foxn1* has been identified in the skin, submatrix region of the nails, tongue, oral cavity and nasal placode (Lee et al., 1999; Meier et al., 1999). In the skin, *Foxn1* expression is observed in the epidermis and hair follicles but it is limited in the epithelium of these cell types (Baxter and Brissette, 2002; Weiner et al., 2007). In the epidermis, *Foxn1* participates in the transition from the proliferative state (basal layer) of the keratinocytes to the differentiated state (suprabasal and rest keratinocyte layers) (Lee et al., 1999; Prowse et al., 1999). FOXN1 is most prominent in the keratinocytes suprabasal layer when cells seem to exit their cell cycle and begin to differentiate, while almost all of the proliferative basal cells of the epidermis are FOXN1 negative

(Lee et al., 1999; Meier et al., 1999). This intrinsic supplement of source of cells from stem cells dispersed in the basal layer is required to generate defensive, terminally differentiated layers of skin (Morasso and Tomic-Canic, 2005). Cells positive for FOXN1 also express early differentiation markers (Keratin 1 and 10; K1 and K10), indicating that FOXN1 acts as a prodifferentiation transcription factor (reviewed in Grabowska and Wilanowski, 2017).

FOXN1 expression is also evident in the hair follicles, which exhibit synchronised dynamic cycles of active growth, regression and quiescence (anagen, catagen and telogen respectively) with FOXN1 levels peaking during anagen, when multipotent stem cells get activated and exit their niche, while FOXN1 is absent from the telogen stage (reviewed in Grabowska and Wilanowski, 2017)

The molecular mechanisms that orchestrate FOXN1 function in skin are still to be better understood. Only a few genes have been identified to be FOXN1 targets and the exact mechanism of their regulation is poorly understood. A microarray analysis in which human epidermal keratinocytes with induced human FOXN1 were compared to a human epidermal keratinocyte control population were compared revealed up-regulation of gene groups linked to signalling, cytokines, growth arrest, extracellular matrix and metabolism, including K10 suggesting that FOXN1 induction may be promoting keratinocyte differentiation (Janes, 2004). In another study, FOXN1 was shown to impose pleiotropic effects, with WNT signalling family genes and metalloproteinases to be enriched in the keratinocytes of nude mice, while BMP and NOTCH signalling members demonstrated the opposite effect (Kur-Piotrowska et al., 2017).

In hair follicles, the regulatory network surrounding FOXN1 is not clearly defined. In some reports, FOXN1 has been positioned downstream of BMP and upstream of NOTCH canonical pathway (Cai et al., 2009), whereas in other reports it seems to exist downstream of the NOTCH pathway (Hu et al., 2010). As will be discussed later (**section 1.4**), the NOTCH pathway has been involved in the fate maintenance inside the hair matrix. In an initial study (Cai et al., 2009), *Foxn1* and *Msx2* expression has been shown to be required for normal expression of *Notch1* in the hair follicle matrix. Genetic knock-out mutants of *Foxn1* and *Msx2* reduced *Notch1* halfway, which sequentially impaired differentiation of the medulla and the inner root sheath (IRS) of the hair follicle. Further experiments indicated that FOXN1 was directly targeting the *Notch1* promoter. Additionally, *Foxn1* and *Msx2* seem to be acting synergistically but in parallel downstream of BMP signalling, since recombinant human BMP4 induced *Foxn1* expression in *Msx2*$^{-/-}$ hair follicles. Despite their synergistic effect, *Foxn1* seems to exist downstream of *Msx2*, as demonstrated by reduced *Foxn1* expression in *Msx2*$^{-/-}$ mutants. Moreover, except for regulating NOTCH signalling, these two factors can be independently involved in cortical and cuticle keratins expression. Collectively, *Foxn1* and *Msx2* exist downstream of BMP signalling and upstream of *Notch1*, and together they play a pivotal role in the differentiation of the IRS, cortex and medulla within the hair matrix. In the second study (Hu et al., 2010), inhibition of NOTCH signalling in the surrounding mesenchyme was shown to influence *Foxn1* expression in the keratinocyte epithelial through *Wnt5a* positioning NOTCH signalling upstream of *Foxn1*.

The equilibrium of the stem cell activity within the bulge stem cells can be affected by the BMP-WNT interplay. Inhibition of the BMP signalling and increase of WNT signalling pushes the balance towards the activation of quiescent stem cells (Kandyba et al., 2013). Therefore, loss or down-regulation of *Foxn1* could potentially contribute towards stem cell activation and a more regenerative environment. Data availability for BMP, WNT and NOTCH signalling pathways in the epidermis is even more limited. BMP seems to regulate the cell fate of amplifying cells by promoting their differentiation (overexpression of terminally differentiated markers; K10 and involucrin; IVL) (Gosselet et al., 2007). In contrast, WNT members are secreted and targeting the basal proliferative layer of the epidermis (mainly consisting of committed progenitors), while the suprabasal and remaining layers have WNT inhibitors (Kaur, 2006). Taking into account that lack of FOXN1 up-regulates WNT signalling and down-regulates WNT inhibitors, it is possible that blockage of FOXN1 favours a proregenerative environment (Kur-Piotrowska et al., 2017).

1.3 CLASSICAL MODELS FOR STEM CELL DIFFERENTIATION AND MAINTENANCE

Early studies on hematopoietic stem cell (HSC) self-renewal and differentiation have created a central paradigm on the mechanisms that lead maintenance and repair in mammalian tissues (reviewed in Morrison et al., 1995; Orkin and Zon, 2008). In this well-established example, the HSCs constitute a rare population of cells that sits at the top of the blood lineage hierarchy and divides infrequently in an asymmetric fashion (Suda et al., 1984). Upon each division, the HSC creates two cells of differing fates. One of these cells is an exact copy of the pre-divided HSC, while the other daughter cell adopts a subsequently lineage-restricted fate. This unique defining capacity of stem cells to self-renew and maintain their long term lineage is called multi-potency. In this paradigm, cells flow in one direction from the HSC towards progenitor lineages to terminally differentiated cells which will undergo apoptosis eventually.

Nevertheless, this central example cannot always explain the underlying mechanisms in other well-studied stem cell systems that (unlike blood which is a fluid organ) contain a more rigid epithelial tissue (Watt, 1999). A main feature of the epithelial cells in these organs is that they attach to the basement membrane of the tissues, with the types of epithelial structures also variable between organs. In the examples of lung and intestine, the epithelia comprise a monolayer of epithelial cells in direct contact with the basal layer. In the mammary gland epithelia, two distinct cell types exist: a luminal population which surrounds a central lumen and myoepithelial cells which are located to the basal area (Visvader and Stingl, 2014). While all myoepithelial cells adhere to the basement membrane, only 20% of the luminal cells are attached to it. In the skin epithelium, multiple layers of epithelial cells (multilayer) are apparent, where distance of the epithelia from the membrane determines the level differentiation per cell. Additionally, it is not unusual for epithelia to contain structures with specialised functions (like hair follicles or glands).

In terms of their stem cell state, solid tissues display two types of behaviour: low proliferative activity when in a non-excited state (like lung, pancreas and liver) with exhibiting proliferation bursts upon damage, and constant self-renewing activity (like the intestine and testis) with daily, weekly and monthly rates of tissue replacement. The latter case resembles the bone marrow self-renewal process (reviewed in Clevers and Watt, 2018).

The low proliferative nature of cells has been widely used over the years as a main attribute to define stem cells. Excellent examples of adult tissue stem cells exist (dormant satellite cells in striated muscle; reviewed in Brack and Rando, 2012, bulge stem cells of the hair follicle; Cotsarelis et al., 1990), nevertheless, quiescence cannot be used as a discriminatory feature for novel stem cell identification, since the vast majority of mammalian cells are non-dividing.

Studies on the proliferative state of the intestinal crypt cells have provided an example where stem cells do not flow in an unidirectional way (reviewed in Clevers, 2013), violating the “one-way” flow rule as it has been established by the HSC main paradigm. In this case, lineage-restricted daughters constitute a persistent repertoire of cells which is able to resettle at the bottom of the intestinal crypt cell hierarchy in which a short-lived multipotent population exists (Buczacki et al., 2013). This process is activated upon loss of the intestinal crypt stem cell marker.

Under a different point of view, daughter cells may exhibit plasticity: the potential to return to a less differentiated (stem) cell state, even though they have progressed down the lineage hierarchy. This progression can range from a progenitor cell stage (see intestinal crypt cells) to the more extreme cases of terminally differentiated cells that can still revert to a stem cell state (see lung epithelium, reviewed in Tata and Rajagopal, 2017).

Unlike HSCs, which are characteristically rare, quiescent, long-living and divide asymmetrically, stem cells in skin (Clayton et al., 2007), intestinal crypt (Lopez-Garcia et al., 2010; Snippert et al., 2010), oesophagus (Doupe et al., 2012), stomach glands (Leushacke et al., 2013) and testis (Klein et al., 2010) are abundant, cycling and divide symmetrically while their life span is stochastically determined. For all the latter tissues, stem cell self-renewal and longevity can only be defined at the population level, but not at the single cell level (see Clevers and Watt, 2018).

Notably, tissues may also involve more than one stem cell type acting independently to maintain the tissue distinct lineages. In skin, for instance, stem cell niches are liable to sustain the differentiation compartments that are allocated only to their local area (Watt and Jensen, 2009). Even though that is the case under steady state conditions, during tissue injury all stem cells can give rise to all the epidermal lineages (Donati and Watt, 2015).

In strong contradiction to the HSC example, no stem cell population has yet been identified in the liver tissue over decades of studies (reviewed in Stanger, 2015), suggesting that the proliferative capacity of the liver lineages are sufficient to maintain and repair the liver tissue.

In sum, besides the well-established paradigm of HSC-like, rare and quiescent professional stem cells, deeper principles may underlie maintenance and regeneration mechanisms of individual tissues, involving cell plasticity, population self-renewal, multiple tissue stem cells and regeneration led by normal tissue cells (reviewed in Clevers and Watt, 2018).

1.4 MODELS BY WHICH NOTCH OPERATES IN OTHER CELL TYPES

NOTCH signalling is a highly conserved pathway that allows cell communication between neighbouring cells. NOTCH ligands interact with their receptors which become activated through sequential proteolytic cleavages that lead to the release of the NOTCH intracellular domain. In turn, this domain is transported inside the cell nucleus where it acts as a coactivator and regulates transcription of target genes in association with members of the CSL DNA binding transcription factor family (reviews by Artavanis-Tsakonas, 1999; Bray, 2006). A wide range of cellular processes can be regulated by NOTCH, including specification of the cell fate, formation of boundaries, stem cell maintenance, cell proliferation, migration and apoptosis. Despite the simple linearity regarding the mechanism of the NOTCH pathway itself, NOTCH signalling specifies highly complex biological outcomes (Artavanis-Tsakonas, 1999; Fortini, 2009; Lai, 2004) through its different modes of action: namely lateral inhibition, fate decisions and inductive signalling.

Lateral inhibition characterises an event where the fate of a cell is specified while this cell lies among a group of equivalent cells. One of the best characterised examples of lateral inhibition occurs in the embryonic neuroepithelium in *Drosophila*, where NOTCH determines if equivalent ectodermal cells will differentiate into epithelial cells or neuroblasts. Initially, all cells in the neuroepithelium exhibit low levels of NOTCH ligand (DELTA) and its receptor (NOTCH). However, possibly due to stochastic variation, some cells increase their levels of DELTA and this inceptive difference is magnified through a transcriptional positive feedback loop. In contrast, these high-DELTA expressing cells cannot express NOTCH (because of *cis*-inhibitory interactions; Heitzler and Simpson, 1993). Eventually, the system resolves in a salt and pepper pattern where the DELTA-expressing signal-sending cells differentiate into neuronal cells, while the NOTCH-expressing signal-receiver cells become epithelial cells. The mechanism of lateral inhibition constitutes a common mechanism through which tissues containing identical cells are patterned during development (Bate and Rushton, 1993; Greenwald, 1998; Greenwald and Rubin, 1992; Roignant and Treisman, 2009; Rushton et al., 1995).

Lineage (or fate) decisions and inductive signalling within non-equivalent cells are also operated by NOTCH signalling. In particular, cells are different from one another either because of asymmetric expression of NOTCH signalling regulators or because of ligand-receptor differential distribution in adjacent cells (review by Bray, 2006). For example, in the progeny of sensory organ progenitors, asymmetric segregation of the endocytic adaptor protein NUMB down-regulates NOTCH signalling and converts the NUMB-positive cells into NOTCH signal-senders. On the contrary, expression of DELTA in vein cells during wing-vein patterning in *Drosophila* induces NOTCH signalling in the adjacent (or intervein) cells which blocks vein fate (Huppert et al., 1997).

Overall, even though NOTCH signalling modality acts as a simple on/off switch, multiple levels of control can ensure precise signalling in the involved cells. First, post-translational modifications and trafficking of the deployed ligands and receptors can regulate their activity. Secondly, when the pathway is active for a specific period of time, oscillatory initiation/termination mechanisms are used. This is possible because the NICD cofactor is short-lived but also because NOTCH target gene messenger RNAs and proteins are very unstable and can also inhibit their own transcription (review by Fior and Henrique, 2009).

1.5 MECHANISMS OF GENE REGULATION

Gene expression in eukaryotes is controlled through a complex mechanism that involves a) presence of multiple RNA polymerases regulating distinct classes of genes, b) binding of proteins to specific regulatory sequences in order to interact and modulate the activity of these RNA polymerases and c) unfolding of the compact chromatin structure to provide a template that allows transcription to happen. The high complexity of eukaryotic transcription orchestrates the unique expression patterns of genes that are necessary to drive development, differentiation and function of different cells in multicellular organisms (reviewed in Cooper and Hausman, 2009).

Three different types of RNA polymerases exist in eukaryotic cells that modulate distinct classes of genes: RNA polymerase II (Pol II) for protein coding genes and microRNAs (Lee et al., 2004), RNA Polymerase III (Pol III) for tRNAs (Willis, 1993), RNA Polymerase I (Pol I) and Pol III for rRNAs (Engel et al., 2013; Willis, 1993) and Pol II and Pol III for snRNAs and scRNAs (Willis, 1993).

Part of the general mechanism through which these polymerases regulate gene transcription is the assembly of a pre-initiation complex formed by various proteins that is necessary to enable transcription initiation (Matsui et al., 1980; Orphanides et al., 1996). Genes transcribed by Pol II have a characteristic TATA box and initiator (Inr) element in their core promoter, among other upstream and downstream binding sites that are specifically targeted by a group of proteins called general transcription factors (review by Weis and Reinberg, 1992). Binding of these general transcription factors in the core promoter of genes enables Pol II recruitment and eventually formation of the pre-initiation complex. A minimum of five general transcription factors are required to initiate transcription in *in vitro* systems, however, within a cell additional factors are necessary to stimulate transcription. One of these factors, called the Mediator, is a large protein complex which can interact with both Pol II and the general transcription factors (Malik and Roeder, 2005). In that way, a complex transcriptional machinery is formed which controls transcriptional regulation in eukaryotes *in vivo*.

The core promoter of genes also includes *cis*-regulatory regions targeted by gene-specific transcription factors (reviewed in Maston et al., 2006). These factors regulate individual gene expression. Furthermore, many genes in mammalian cells can be regulated by far-acting sequences located in a significant distance from the transcription start site of the gene. These particular regions, called enhancers, can function from a long distance and in either orientation (Dynan and Tjian, 1983). They apply their regulation similarly to promoters, by regulating the RNA polymerase (reviewed in Maston et al., 2006). The distant enhancers are able to interact with the Pol II preinitiation complex because of DNA looping (Matharu and Ahituv, 2015).

1.5.1 Mechanisms of transcriptional activation and repression

Promoter and enhancer regions contain specific binding sites through which transcription is regulated. Binding sites are made of short DNA sequences that usually range from 6-10 base pairs. These sequences are degenerate, denoting that regulatory proteins (such as TFs) will also bind to sequences deviating from the binding motif consensus by one or more positions. Transcription regulators that bind to these motifs are divided into activators and repressors based on their function (reviewed in Kadonaga, 2004). Transcription activators stimulate transcription through their binding to the regulatory DNA sequence (Ptashne, 1988). These proteins are characterised by two domains: a DNA-binding domain that specifically binds DNA and an activation domain that interacts with the Mediator or other components of the pre-initiation complex to promote specific gene transcription. The latter domain stimulates transcription by two different mechanisms. First, by interacting with components of the pre-initiation complex to recruit Pol II and form the transcriptional complex at the promoter region. Second, by interacting with a variety of other factors (namely coactivators) that modify chromatin structure to stimulate transcription. Notably, transcription can also be regulated at the level of transcript elongation and RNA processing in which transactivators also play a role (reviewed in (Ma, 2011).

Repressor proteins are also responsible for gene regulation in eukaryotic cells. Similarly to the activator proteins, eukaryotic repressors bind to specific binding sites to apply their effect; but in this case they cause transcription inhibition. In some cases, this inhibition is mediated by interfering with the binding of the TFs to the regulatory region. For instance, proximal binding of a repressor to the TSS of a gene can deter/prevent the interaction of general TFs or the Pol II itself with the promoter. In other cases, receptors antagonize activators for binding in the same regulatory site. Some repressors may exhibit the same DNA-binding unit with an activator but lack a stimulation domain, while others may contain a functional inhibitory domain that represses gene transcription by protein-protein interactions (active repressors). Lastly, repressors can apply their inhibitory effect by interacting with corepressors, whose role is to modify the chromatin structure (reviewed in Gaston and Jayaraman, 2003).

The action of both transcriptional activators and eukaryotic repressors controls the expression or silencing of tissue specific genes, cell proliferation and differentiation and responses to external stimuli (e.g. hormones and growth factors).

1.5.2 Chromatin structure controlled by activators and repressors

Both activators and repressors may control transcription not only by interacting with various subunits of the transcriptional machinery but also by inducing structural changes to the chromatin of the eukaryotic cell. Instead of being present in a "naked" state, DNA in eukaryotes is wrapped around histones in the core particle of the nucleosome. Packaging of the chromatin in this tight structure is directly linked to its availability as a substrate for permissive transcription. Thus, gene expression is majorly dependent on the chromatin structure, with actively transcribed genes to be found in less condensed chromatin regions. Nevertheless, these genes are still bound to histones and wrapped around nucleosomes, affecting the ability of both transcription factors and RNA polymerase to bind and transcribe DNA respectively.

Several mechanisms are responsible for chromatin remodelling, including histone modifications, high mobility group (HMG) proteins-chromatin interactions and nucleosome rearrangement (Saha et al., 2006). Two domains are apparent in the histones of the nucleosome core particle (H2A, H2B, H3 and H4): a histone-folding domain, which takes part in histone-histone interactions and DNA wrapping around the core particle of the nucleosome, and an amino-terminal tail domain which stretches outside the nucleosome (Luger et al., 1997). The extended domain contains multiple lysines on its tail which can be acetylated in specific positions. Histone acetylation has been directly linked to transcriptional regulation through studies which showed associations of activators and repressors with histone acetyltransferases and deacetylases respectively (Brownell et al., 1996; Taunton et al., 1996). Several other modifications can influence the status of histones, including methylation of arginine and lysine residues and phosphorylation of serine residues. These modifications take place in the amino-terminal tails similar to histone acetylation and have been associated with alterations in transcriptional activity (Bannister and Kouzarides, 2011). It seems that histone modifications regulate gene expression by both remodelling the chromatin state and by creating accessible sites for other factors to bind and activate or repress transcription. For instance, acetylation of lysines neutralizes their positive charge which relaxes protein structure and enables transcriptional activation. Conversely, methylation of lysine residues in H3 create binding positions for factors that lead to a more condensed chromatin state, directly associating this modification with formation of heterochromatin and transcriptional repression.

Apart from histone modifications, interaction of the HMG superfamily of proteins is able to influence the chromatin architecture by bending DNA and facilitating binding of regulatory factors or alternatively by inducing chromatin unfolding and maintaining DNA in a decondensed state (Hock et al., 2007). Additionally, chromatin remodelling factors can also alter the DNA-histone contacts by a) catalysing the sliding of the nucleosomes along the DNA molecule to alter the accessibility of DNA sequences to TFs (Workman, 2006), b) changing the conformation of the nucleosomes to affect the ability of TF binding to DNA or c) ejecting histones from DNA to create a nucleosome-free region (Henikoff, 2008).

1.5.3 Chromatin modifications as indicators of promoter/enhancer activity

Histone modification marks are often used to distinguish promoter and enhancer regions in genome studies (reviewed in Andersson, 2015). This methodology has been adopted based on distinguishable properties that can separate one from the other. In particular, histone H3 tri-methylation at lysine residue 4 mark (H3K4me3) has been denoted as a specific promoter mark (as shown for instance in Barski et al., 2007), however, few studies have also associated this mark with regulatory active enhancers (as shown for instance in Koch et al., 2011 and Pekowska et al., 2011). Furthermore, even though promoter and enhancer regions are usually flanked by nucleosomes bearing both histone H3 monomethylation at lysine residue 4 (H3K4me1) and H3K4me3, the H3K4me1:H3K4me3 ratio was found to be high in enhancers but low in promoters (Heintzman et al., 2007). The patterns of histone post translational modifications (PTMs) have further been refined by subsequent studies to predict specific states of enhancer activity. In short, H3K4me1 marks many enhancers, but on its own it cannot distinguish between active and primed (to become active) enhancers (Bonn et al., 2012); histone H3 acetylation at lysine residue 27 (H3K27ac) generally marks active enhancers (Creyghton et al., 2010) and histone H3 trimethylation at lysine residue 27 (H3K27me3) and/or histone H3 trimethylation at lysine residue 9 (H3K9me3) exist at developmentally poised (silenced) enhancers, while they are substituted by H3K27ac when activated during early differentiation (Rada-Iglesias et al., 2011; Zentner et al., 2011). Finally, other histone acetylation marks have been presented to have predictive power in discerning gene promoters to enhancers (Rajagopal et al., 2014).

1.5.4 The role of pioneer factors

Large eukaryotic genomes exist in a dual compacted chromatin state: at the level of the DNA wrapped around the nucleosomes (Luger et al., 1997) and at the level of inter-nucleosome interactions (Schalch et al., 2005). This compacted DNA state into chromatin restricts the amount of freely available DNA to which regulatory proteins may have access to bind. Therefore, there is an intrinsic tendency of the chromatin to exist in a repressive state, which reinforces a stable gene expression state and limits undirected cell fate commitment. In order for the regulatory factors to gain access to DNA binding sites that reside within compacted chromatin (for instance silent genes) and control cell fate, the locally compacted chromatin must disentangle.

Chromatin decompaction can be enabled by a subset of factors called pioneer factors. The basic features that characterise a pioneer factor are the capacity to a) bind its cognate DNA site within "closed" chromatin (DNA is not accessible), b) instruct chromatin to remodel making specific DNA sequences accessible and subsequently c) enable binding of other TFs and lastly d) impose stable reformations in chromatin structure linked to epigenetic stability (reviewed by Mayran and Drouin, 2018; Zaret and Mango, 2016). However, binding of a pioneer factor can also lead to repressed chromatin, if the binding is alongside repressors or corepressors that cause local domain blanketing and impair activators access (Sekiya and Zaret, 2007; Watts et al., 2011). Altogether, these features suggest that when a pioneer factor "acts", it may establish competence for consequent induction, establish a repressed region, or direct activation of a gene promoter, while its effect on chromatin is maintained.

Nevertheless, binding of pioneer factors to chromatin is not fully unrestricted. Certain chromatin domains can actively exclude pioneer factor binding through repressive features: in particular, the second level of compaction imposed by heterochromatin (regions enriched for H3K9me3 mark; Lachner et al., 2001), as first shown in a OSKM factor study where fibroblasts were converted into induced pluripotent stem (iPS) cells (Soufi et al., 2012). Additionally, it is likely that pioneer factors will remain more stably bound in locations where they interact and experience cooperative binding with other regulatory proteins, a property which is common to both pioneer and other transcription factors but is not a defining feature of pioneers (reviewed by Zaret and Mango, 2016).

Forkhead Box A (FOXA) constitutes the paradigm of a pioneer factor. This was demonstrated by a study showing occupancy of liver-specific FoxA sites in endoderm prior to liver specification (Gualdi et al., 1996). FOXA has also been shown to bind nucleosomal DNA (Cirillo, 1998) and unfold condensed chromatin (Cirillo et al., 2002). In addition, FOXA remodelling activity and coupled nucleosome depletion have been shown by genome-wide studies (Iwafuchi-Doi et al., 2016; Li et al., 2011b; Serandour et al., 2011). Currently, the pioneer factor label can be assigned to a regulatory factor by assessing the chromatin status prior to and after the factor's action. This needs to be tested in an experimental system where the cells have never been exposed to the pioneer factor. Without this test, pioneer activity can only be inferred but not proved (reviewed by Mayran and Drouin, 2018).

1.6 NEXT GENERATION SEQUENCING TECHNOLOGIES AND DATA ANALYSIS (PRÉCIS II)

Leaps in scientific insight are closely associated with development of technology. The longstanding "Holy Grail" for regulatory biology is to understand how genomes encode and regulate the diverse gene expression patterns that designate the type and state of each cell. The fields of genomics, transcriptomics and epigenomics constitute the current leading elements in this type of research. The ability to measure gene expression and protein-DNA interactions in a cell in a genome-wise manner plays a vital role in elucidation of the defining features of a cell, the instructional changes driving cell differentiation and the genetic regulatory networks that supervise gene function overall.

1.6.1 A short history of DNA sequencing technologies

> "... [*A*] *knowledge of sequences could contribute much to our understanding of living matter.*"
>
> ***Frederick Sanger***

Since the fundamental discovery of the DNA structure (Watson and Crick, 1953), scientists envisioned to discover its sequence. A major breakthrough in the field of DNA sequencing occurred with the development of the Sanger (Sanger et al., 1977) and Maxam Gilbert (Maxam and Gilbert, 1977) sequencing methods, which defined the first-generation of DNA sequencing technologies. The Sanger method overruled the sequencing world at the time, due to its radiation-free methodology and, following developments in automation and parallelisation, finally led to the landmark completion of the first human genome sequence (Lander et al., 2001; Venter et al., 2001). Soon after, genomes of the mouse and several other key model organisms were also sequenced (Mouse Genome Sequencing Consortium et al., 2002).

Developed in the mid-to-late 1990s, microarray technology was the first to provide a highly parallel (or high-throughput) assay to measure DNA and RNA. In principle, a reference set of probes (with a reference to known transcripts) is fixed in a chip and the sample of interest is added (Lockhart et al., 1996; Schena et al., 1996; Southern et al., 1992). The probes are then hybridised with fluorescently labelled cDNA and the fluorescent signal on the chip is used to infer the relative quantity of the probed sequences. Microarrays became widely commercial and this significantly contributed to their broad application in many fields (Hughes et al., 2014; Lipshutz et al., 1999; Morozova et al., 2009; Stoughton, 2005). However, microarrays face a number of limitations: reliance upon existing knowledge for probes does not allow *de novo* transcript identification, and the limited dynamic range of detection (because of background and saturation noise) cannot adequately represent the dynamic range of gene expression (Mortazavi et al., 2008; Wang et al., 2009).

From the Human Genome Project onwards, the need to increase the throughput while reducing the cost of sequencing technologies became clear. This was achieved with the development of next generation sequencing (NGS) methods. This term is used to describe the technologies performing parallel sequencing of multiple DNA fragments to define a sequence (Rizzo and Buck, 2012). In comparison with Sanger sequencing, NGS technologies produce an exponentially greater amount of DNA sequence data at a much higher speed, while the cost is significantly reduced (Voelkerding et al., 2009). Second generation sequencing methods became available from commercial technologies, namely Roche/454, Applied Biosystems SOLiD and Illumina Genome Analyser IIX (reviewed in Heather and Chain, 2016). The NGS methodology of these technologies is based in common concepts. The fragmented sample is initially amplified through the polymerase chain reaction (PCR), which reduces cost, time and species-specific bias compared to cloning. Secondly, massive parallelisation is achieved via “sequencing-by-synthesis”. Even though different technologies come with different detection chemistries, as a ground rule nucleotides are labelled (with chemiluminescence or fluorescence) and they release a chemical signal when they are correctly incorporated to the synthesised sequence (Margulies et al., 2005; Shendure et al., 2005).

Third generation sequencing technologies are highly similar to the NGS technologies, however, instead of using amplified DNA for sequencing as a template, they use DNA molecules, thereby eliminating errors occurring in the lab at the DNA amplification stage (Munroe and Harris, 2010; Schadt et al., 2010). The newest available commercial technologies include Illumina HiSeq, Pacific Biosciences RS platform (English et al., 2012) and Oxford Nanopore (Mikheyev and Tin, 2014; Stoddart et al., 2009). In more detail, the PacBio platform uses zero-mode waveguides for single molecule real time sequencing (English et al., 2012), while Nanopore measures the signal change (i.e. current) when a molecule translocates through a protein nanopore (Stoddart et al., 2009).

These technologies have developed side-by-side with the physical development of sequencing machines, the data analysis methods and the storage capabilities of the time. One remaining limitation is the incapacity of these sequencing technologies to sequence long reads with high accuracy. When a reference genome is available, computational advances can possibly overcome this issue. However, sequencing novel genomes or novel transcripts remains challenging. The PacBio platform and more prominently nanopore-based technologies are currently making efforts in that direction (English et al., 2012; Mikheyev and Tin, 2014; Stoddart et al., 2009).

Currently, large scale sequencing can be performed for any DNA sample, whether from reverse-transcribed RNA, whole genome or amplified DNA by other methods. Integration of the generated big data enables investigation of cell regulation from many different aspects, such as transcription dynamics (RNA sequencing; RNA-seq), transcription factor binding and epigenetic modifications (Chromatin immunoprecipitation coupled with sequencing; ChIP-seq, DNA adenine methyltransferase identification assay coupled with sequencing; DamID-seq, Assay for Transposase-Accessible Chromatin using sequencing; ATAC-seq) and genome variations in a population (whole-genome sequencing and assembly). All the above applications have mostly been facilitated by improvements focusing on the library preparation and sensitivity, while newer applications are now focusing on the single cell level. Single cell RNA-seq (scRNA-seq) is perhaps the most prominent technique nowadays for the investigation of population heterogeneity that cannot be denoted from the average population profile of the RNA-seq (Grün and van Oudenaarden, 2015; Kolodziejczyk et al., 2015; Liu and Trapnell, 2016; Saliba et al., 2014). For the purposes of this thesis we are going to explore further the RNA-seq and ChIP-seq analyses.

1.6.2 RNA-seq data processing for novel biological insight

RNA-seq typically produces millions of sequenced reads that represent the average transcriptome profile of a cell population. These data must be analysed in order for useful biological insights to be extracted. A major challenge with analysis of the RNA-seq data comes with assembling the millions of short reads into a map of non-contiguous sequences on a genome-wide scale. Two approaches can be recruited to reconstruct the transcriptome from RNA-seq data. Transcribed sequenced reads can be mapped to a reference genome and overlapping alignments can be used to merge these sequences; or reads can be assembled as continuous transcripts *de novo* without the use of a reference genome.

Read mapping using a reference genome has traditionally been the most common method to reconstruct the transcriptome. This method increases sensitivity of the assembly, since lowly expressed genes may have very low coverage, making them insufficient for *de novo* assembly. One of the hardest computational tasks in RNA-seq read alignment is sequence non-contiguity due to introns (the non-coding sections of the RNA transcript) intersections. A number of sequence aligning tools, namely Tophat (Trapnell et al., 2009), STAR (Dobin et al., 2013) and HISAT (Kim et al., 2015) have introduced memory efficient and considerably fast ways to incorporate splice junctions allowing reliable merging of non-continuous reads. As a next step, overlapping aligned reads are merged into putative transcripts and quantified. For this quantification, regions that constitute a gene need to be inferred and ambiguously aligned reads or different isoforms have to be dealt with. Various quantification methods have been proposed, ranging from simple counting of reads that overlap using prior information of annotated gene regions to more complex modelling of transcript isoforms abundances. A few tools have been developed to perform “simple” counting, such as featureCounts (Liao et al., 2014) and HTSeq-count (Anders et al., 2015). More recent methods have aimed at quantifying individual transcripts abundances, namely, Cufflinks (Trapnell et al., 2012) , RSEM (Li and Dewey, 2011), BitSeq (Glaus et al., 2012), kallisto (Bray et al., 2016) and Salmon (Patro et al., 2016). These approaches offer higher resolution compared to simple counting, while by skipping the computationally costly read mapping step, some (notably, kallisto and Salmon) achieve very high execution speed. However, calculating abundance of gene isoforms can be a lot more challenging and complex due to the high extent of overlap among transcripts. Therefore, so far there is no agreement in the optimal resolution (or approach) for quantification of gene levels and downstream analysis.

Mapping reads against a reference genome may result in potential information loss due to genetic diversity that may be apparent in a species and which cannot be represented in a linear reference genome. Thus, *de novo* assembly constitutes an important alternative method to use, importantly when significant differences are apparent between the reference and the individual transcriptome. Such variations may be splice isoforms, point mutations, indels (insertions or deletions) or significant alterations of the genome/transcriptome in e.g. cancer studies. However, the main limitation of the *de novo* assembly method is that poor quality reads can result in accuracy loss of the transcriptomic output. The computational hurdle of transcriptome alignment has been overcome by breaking down the long sequenced reads into substrings of length *k* (*k*-mers) and then constructing graphs (called *de Bruijn* graphs), in which *k*-mers are represented by nodes and transcripts can be defined by following any route defined by the edges which constitute immediately overlapping *k*-mers (*k*-1 bases) (Compeau et al., 2011; Grabherr et al., 2011). Reads are then mapped to compatible transcripts creating a pseudo-alignment (in which exact coordinates of the match are not kept), and their quantification happens through expectation maximisation procedures (see kallisto; Bray et al., 2016).

The next step after transcript reconstitution and quantification is to gain useful insight from the experiment, by comparing gene expression profiles and selecting the genes that present significant changes in association with a specific condition for further analysis. To discern significant from non-significant changes (often caused by technical variation in the measurements), a statistical model needs to be incorporated to best fit the nature of the RNA-seq data. This model should account for the discrete nature of the RNA-seq data, differences in variance across the expression mean and expression intensity levels that are contingent on expression levels of other genes. The Poisson model can appropriately model RNA-seq data, by describing read sampling from a fixed pool of genes. A drawback of this model is that it considers an equal mean and variance for the data, which is not representative of RNA-seq data. The variation instead tends to be larger than the mean and also dependent on the mean (Marioni et al., 2008; Nagalakshmi et al., 2008; Rapaport et al., 2013). Therefore, tools such as DESeq and EdgeR use a negative binomial distribution that incorporates a dispersion factor and estimate dispersion based on mean expression intensity from other genes in the samples because of the generally low number of sample replicates (Anders and Huber, 2010; Robinson et al., 2010). Unlike the aforementioned tool, the *voom* function from limma represents expression intensity as a continuous variable and estimates the variance of expression intensity by applying non-parametric regression and incorporates it into a linear model (Law et al., 2014).

At present, none of the above tools (or models) have demonstrated a major advantage in terms of accuracy or sensitivity of the differential expression analysis over the others, and all of them are commonly used for RNA-seq analysis (Rapaport et al., 2013; Seyednasrollah et al., 2015).

1.6.3 ChIP-seq data processing for novel biological insight

Chromatin immunoprecipitation followed by sequencing (ChIP-seq) produces millions of sequenced reads that represent the average binding profile of a protein (transcription factor; TF, histone modification mark; HMM, DNA-binding enzyme, chaperone or nucleosome) in a population of cells (Barski et al., 2007; Johnson et al., 2007; Mikkelsen et al., 2007; Robertson et al., 2007). Main advantages of the NGS technologies are relatively high mapping resolution, low noise and adequate genomic coverage, in comparison to ChIP coupled with microarray hybridization (ChIP-chip) assays. For all the above reasons, ChIP-seq is currently the most widely used method for locating protein-DNA interactions in genome-wide assays (Furey, 2012) and locating histone modification positions in epigenetics research.

A major challenge with the analysis of ChIP-seq data is that sufficient coverage needs to be provided by the sequenced reads (depth of sequencing) for the data analysis to be meaningful. The two main factors that influence sequencing depth are the number and size of the protein binding sites and the size of the genome. For transcription factors (TFs) present in mammals and chromatin modifications associated with gene promoters that mark narrow sites at specific locations, 20 million reads may be adequate (with a much lower number of reads required for worm and fly TFs) (Landt et al., 2012). However, for factors with more binding sites (e.g. RNA Pol II), including most of the histone marks, more reads will be required, up to 60 million reads for mammalian genomes (Chen et al., 2012). Importantly, to ensure sufficient coverage of the genome, control samples should be sequenced in higher depth than the factor under investigation.

Read mapping of ChIP-seq data to a reference genome is highly similar to RNA-seq, however, detection of indels is not a prerequisite for most ChIP-seq experiments. Popular ChIP-seq aligners include Bowtie2 (Urban, 2014), BWA (Li and Durbin, 2009), SOAP (Li et al., 2008b) and MAQ (Li et al., 2008a), which use heuristics to improve mapping speed and reduce the memory footprint of the alignment. For ChIP-seq analysis, it is important to examine the percentage of the uniquely mapped reads (reads that map in only one location in the genome) reported by the aligner. A low percentage of uniquely mapped reads can be indicatory of technical bias, such as excessive PCR amplification, inadequate read length, sequencing platform-specific issues. A potential cause of a high percentage of "multi-mapping" reads (reads that confidently map in more than one locations in the same genome) for some ChIPed proteins is due to biological reasons (e.g. binding in repetitive DNA). In this latter case, mapping ambiguity can be improved using paired-end sequencing. As a ground rule, most peak-calling algorithms will discard multi-mapping reads for increased sensitivity although they can often lead to the discovery of new binding sites (Wang et al., 2013), while duplicated reads (reads map in the exact same location) also need to be removed prior to peak calling to improve specificity.

A pivotal component of the ChIP-seq analysis is to identify the ChIPed regions that are significantly enriched (peaks) above background. A fine balance needs to be achieved between specificity and sensitivity when choosing a peak-calling algorithm based on the type of protein that is ChIPed, with signal smoothing and background modelling to be the main focus areas for several peak calling tools. For point-source factors (such as most TFs), which are the most abundant type of available ChIP-seq data, peak callers have been fine-tuned to fit the nature of them. Peak callers, such as SPP (Kharchenko et al., 2008) and MACS (Zhang et al., 2008), calculate the footprint of the ChIPed protein from the mapped reads in the minus and the plus strand of the bound region. By making sure that the footprint differs from the read size selection during library preparation, they can improve accuracy of the prediction. After signal smoothing, background models remove noise either directly using a control sample, or indirectly based on features of the genomic landscape, such as mappability and GC content (Cheung et al., 2011). A user-defined threshold is set to finally call peaks. Multiple statistical models have been designed to fit the nature of the ChIP-seq data, namely, negative binomial (CisGenome Ji et al., 2008), inflated negative binomial (ZINBA; Rashid et al., 2011), Poisson (CSAR; Muiño et al., 2011), local Poisson (MACS), and more sophisticated machine learning modelling techniques (Qin et al., 2010; Spyrou et al., 2009). Additionally, a few peak callers, such as PeakSplitter (Salmon-Divon et al., 2010) and GPS (Guo et al., 2010) use the peak shape as a clue to improve the spatial resolution of the binding and avoid erroneous merging of neighbouring peaks from peak callers that use window-based approaches. For broadly enriched factors (such as histone modification marks), several peak calling tools have been adjusted to identify broadly enriched regions, including SICER (Zang et al., 2009), CCAT (Xu et al., 2010), RSEG (Song and Smith, 2011) and ZINBA. Few of the narrow peak callers such as SPP, MACS version 2 (MACS2), and PeakRanger (Feng et al., 2011) can also be used to call broad peaks by adjusting their bandwidth (longer regions) and peak "cut-off" (more relaxed thresholds), while MACS2 and Scripture (Guttman et al., 2010) allow for the option of identifying an enriched narrow peak inside an already enriched broad peak identifying a more hierarchical pattern of peaks. Lastly, for factors that bind to DNA regions with bigger variation (such as RNA Pol II) some tools (see SPP, MACS2 and ZINBA) offer both narrow and broad peak calling choices based on the question that needs to be addressed. However, with careful parameter adjustments, any algorithm that identifies broad peaks could be used in this case.

User-settable parameters among the different peak callers can greatly influence the number and the quality of the enriched peaks. One of the most important parameters, the enrichment metric of the peak calling algorithm, such as the *p*-value or the false discovery rate (FDR), can be greatly influenced by the statistical model that each tool/algorithm uses. Therefore, using the same enrichment metric thresholds does not lead to comparable results (number of peaks) among different peak callers (Szalkowski and Schmid, 2011). Instead, a more consistent approach of filtering significant peaks from non-significant ones, is to threshold the irreproducible discovery rate (IDR; Li et al., 2011a). This method assesses the rank consistency of enriched peaks among replicates and peaks consistent above a user-specified IDR threshold are considered consistently enriched and more likely to constitute true ChIPed peaks. It has been reported that the numbers of peaks declared using a reproducibility threshold (e.g. IDR) metric rather than an enrichment-based one (e.g. *p*-value or FDR) are more comparable across experiments (Landt et al., 2012). Importantly, at least two biological replicates are recommended in such ChIP-seq experiments (Landt et al., 2012) to ensure reproducibility of the identified peaks.

1.6.4 NGS and data storing

A major advantage in studies involving next generation sequencing datasets has been the standardisation of procedures for storing and sharing data. Genome browsers, such as UCSC (Kent et al., 2002) and Ensembl (Yates et al., 2016), make feasible the exploration of newly generated genomic, transcriptomic, epigenomic and proteomic datasets. These datasets can be mapped onto whole-genome coordinates that exist as centralised repositories, the genome assemblies, providing a lot of useful insight to any researcher. Continuous updates of these annotation packages make sure that the information keeps up-to-date. In 2012, the Encyclopaedia of DNA Elements (ENCODE) project (Dunham et al., 2012) was released, providing a significant update in respect to the function of 80% of the genome.

Since the information generated from a high-throughput experiment reaches beyond the detailed questions dealt in publications, these large scale sequencing data are stored in standardised repositories, such as NCBI’s Gene Expression Omnibus (Barrett et al., 2013) and ArrayExpress (Kolesnikov et al., 2015) to support reproducible research. The benefits of data storage and public availability can become more obvious when it is put in the context of other data in future research.

In summary, an extraordinary leap has been produced in the last 10-15 years in the capacity to sequence, store and share insight from sequencing experiments. Although some challenges still remain to achieve optimal experiment standardisation, including downstream information processing and storage, the next generation sequencing field is an ever growing resource with an extremely powerful reserve to the scientific community.

1.6.5 Understanding gene regulatory mechanisms through RNA-seq and ChIP-seq data integration

The collection of genome-wide data from various samples, such as cells, tissues or model organisms has been made easier and cheaper with the use of high-throughput technologies. Transcriptomics, genomics, epigenomics and proteomics data each provide an insightful, yet singular point of view. Therefore, integrative analysis of these datasets can provide a more unified, global view of gene function (discussed in Angelini and Costa, 2014). This section will briefly discuss how the interplay between gene expression and transcription factor binding/epigenetic markers presence can help to elucidate gene regulatory mechanisms.

RNA-seq data provide a clear picture of the gene expressional profile in an average cell population. However, sole nucleotide sequences of expressed genes cannot explain their function, nor their regulation. Genes DNA structure and accessibility to available transcription factors and the basal transcription machinery specifies their transcription. Transcription factors, chromatin-modifying enzymes and other accessory proteins physically interact with DNA to modulate gene expression dynamics and define the cell fate (Atkinson and Halfon, 2014). Recent studies demonstrate that TF and histone modification binding can predict expression of genes *in vitro*, cells' differentiation state and other epigenetic factors (Cheng et al., 2011, 2012; Creyghton et al., 2010; Karlić et al., 2010; McLeay et al., 2012; Ouyang et al., 2009). Similarly, changes in gene expression have been correlated to TF binding and chromatin mark modifications (Althammer et al., 2012; Klein et al., 2014). In terms of the potential impact on cell state, global (and local) chromatin changes accompany cell differentiation, leading to silencing of pluripotency genes and activation of lineage-specific ones (Chen and Dent, 2014). In this respect, multi-omic integrated data can be used to explore the heterogeneity of cell populations and explain or even potentially control differentiation of cells during development (Comes et al., 2013; Macaulay and Voet, 2014).

Prediction of gene expression is possible using a limited number of samples (with replicates), however this is not the case for large gene regulatory networks. To infer these networks, several high-throughput datasets need to be integrated together (shown by Gerstein et al., 2012). Nevertheless, less complicated networks that rely only on few factors and interactions can be predicted using fewer datasets. For instance, a minimal set of components was sufficient to describe self-renewal of embryonic stem cells (Dunn et al., 2014).

Two different approaches can be used to achieve data integration: unsupervised and supervised. The unsupervised approach excels at generating hypotheses without any prior knowledge through discovery of patterns using different types of clustering. These novel patterns can be used to make and test predictions with further datasets or experiments. This is where the supervised integration can become most informative, relying on a few dimensions of the different datasets to apply testable hypotheses of the question posed. Therefore, the more predictive the hypothesis, the more insightful the biological result. Collectively, data integration itself does not constitute the final step in an analysis: it aims to generate novel hypotheses and assist in their testing.

Overall, next generation sequencing technologies have evolved over the years and they now offer quick and in depth investigation of cells transcription profile, binding occupancy and accessibility landscape at a reduced cost. Integration of the above datasets together can lead to more predictive hypotheses and more biological insights that are necessary to formulate testable experimental models for validation and to drive the field under investigation further.

1.7 PROJECT AIMS AND THESIS STRUCTURE

This thesis aims to improve the current understanding of the transcriptional, regulatory and signalling mechanisms in fetal thymic epithelial cell (TEC) progression during thymus organogenesis in the mouse embryo, by dint of bioinformatics analysis. The objectives of this thesis and the structure of the chapters are laid out here, to provide a roadmap of the main points that are addressed in **Chapters 3-7**.

One of the main challenges in the thymus field is to decipher how thymic epithelial progenitor cells (TEPCs) differentiate towards the cortical and medullary TEC lineages (cTECs and mTECs) during thymus organogenesis in mouse. The current model (see outline in Takahama et al., 2017) of TEC differentiation describes a homogeneous bipotent progenitor population at E12.5 (when *Foxn1* expression has already been established), which differentiates in a serial fashion towards the cTEC and mTEC lineages, after going through a preliminary stage, in which all cells express cTEC-like markers. In **Chapter 3**, I use transcriptional datasets, representative of fetal TEPC and newborn TEC populations, to: a) explore the impact of different *Foxn1* levels on the progression of TEPCs, b) identify which developmental stage better resembles the bipotent TEPCs found in *Foxn1*-depleted samples, c) investigate a potential repressive role for FOXN1 in early fetal differentiation and d) predict a gene signature representative of the early bipotent progenitors. These bioinformatics findings are used in combination with experimental evidence (provided by **members of the Blackburn lab**) to inform an improved version of the current TEPC progression model.

TEPC exploration is currently limited by the small amounts of *in vivo* material that can be obtained from the developing thymus of mouse embryos. To achieve expansion of the bipotent TEPC population *in vitro*, the experimental conditions in which these cells are kept (when in culture) would have to be fine-tuned and optimised, for cells to survive and expand in a more efficient way. Thus, in **Chapter 4**, I use comparative transcriptomic analysis in conjunction with pathway enrichment analysis to predict the intrinsic and extrinsic signalling cues that act on the early undifferentiated TEPC state *in vivo*. Among the predicted pathways, NOTCH signalling is identified as dynamically regulated in early fetal TEPCs. In the same chapter, further experimental

(**Dong Liu**) and bioinformatics exploration of NOTCH signalling is performed, to explore its role in early TEPC development.

FOXN1 comprises the central regulator of the thymic system, with its expression to be indispensable for the development of a functional thymus. However, until recently, direct evidence for its immediate targets was lacking. In **Chapter 5**, I am using integrative analysis of transcriptomic, genomic and epigenomic datasets to investigate the regulatory network, through which FOXN1 establishes the transcriptional programme that underpins thymic epithelial cell identity and function. In this analysis, I will a) compare the chromatin accessibility landscape of fetal TEPCs with the binding events of FOXN1 in newborn TECs to identify a list of a high confident FOXN1 candidate direct targets in fetal TEPCs, b) anticipate regulation of active signalling pathways in TEPCs via the FOXN1 predicted targets, c) explore how FOXN1 and NOTCH signalling may interact during fetal thymus development and d) discuss a potential role for FOXN1 to act as a pioneer factor and thus have the ability to impose TEC identity.

The amount of high-throughput sequencing datasets of the various stromal subpopulations, which reside in the thymus, are currently limited and thus, field research could be set back due to insufficient expressional or regulatory insight in these populations. For this cause, in **Chapter 6**, I introduce the development of ThymiBase, an interactive thymic-specific database that provides a platform for easy access, analysis and integration of curated bioinformatics datasets focussed on the thymic epithelium.

To end, **Chapter 7** presents a summarised overview of **Chapters 3-6**. Additionally, it integrates together the FOXN1-NOTCH interplay (**Chapters 4 and 5**) with the revised version of the TEPC progression model in the fetal TEPCs (**Chapter 3**), to propose potential mechanisms through which TEPC specification in the cTEC and mTEC lineages may be deciphered.

Chapter 2

Materials and methods

Chapter 2

Materials and methods

2.1 PRÉCIS

The first part of **Chapter 2** (**Data Repository**) describes the biological designs and technical specifications of the different high-throughput sequencing datasets (provided by others) that I have computationally analysed for the purposes of this thesis. The second part of **Chapter 2** (**Bioinformatics Analysis**) presents in detail the RNA-seq and ChIP-seq pipelines that I have built to analyse the above data, as well as the post-processing normalisation steps that I have followed to fully refine these data and remove any artefacts or technical bias.

2.2 DATA REPOSITORY

The first part of **Chapter 2** is a detailed record of the experimental designs for all the next generation sequencing (NGS; Behjati and Tarpey, 2013) datasets presented in this thesis. Three main types of NGS data have been available for this project: transcriptomic, genomic and epigenomic data, which have been categorised and are described extensively below.

2.2.1 Collection of transcriptomic datasets in mouse (RNA-seq)

For the purposes of this project, RNA-seq datasets from thymic epithelial progenitor cells (TEPCs) and thymic epithelial cells (TECs) from fetal, newborn and adult mice have been **generated by other members of the Blackburn lab** or have been **collected through public repositories**. A descriptive listing of all the RNA-seq samples (grouped in various series of experiments), their experimental designs and specifications is provided below.

2.2.1.1 A Developmental series of fetal TEPC data

RNA was obtained from EpCAM$^+$PLET1$^+$ TECs from fetal mice at days E10.5, E11.5 and E12.5 of embryonic development (**biological triplicates per day**) by **Harsh Vaidya**, **Blackburn lab** (unpublished) and samples were sent for sequencing. EpCAM$^+$PLET1$^+$ mark the early developmental stages of the thymic epithelium and have been linked to progenitor activity of TECs (Depreter et al., 2008; Farr and Anderson, 1985).

2.2.1.2 A *Foxn1* Allelic series of fetal TEPC data

RNA was obtained from EpCAM$^+$PLET1$^+$ TECs from fetal mice of different genotypes at day E12.5 of embryonic development (**one biological replicate per genotype**) by **Stephanie Tetelin**, **Blackburn lab** (unpublished) and samples were sent for sequencing. In short, the mice genotypes contain variants of a *Foxn1* normal allele

(*Foxn1*^{WT}) and a *Foxn1* revertible hypomorphic allele (*Foxn1*^{R}). For the generation of the *Foxn1*^{R} allele, a LoxP flanked cassette was inserted into intron 1b of the *Foxn1* locus through homologous recombination in embryonic stem (ES) cells (see right panel in **Figure 2.1**) and this ES line was used to generate the *Foxn1*^{R} mouse strain (see Nowell et al., 2011 for more details). Precisely, five allelic combinations of the *Foxn1*^{WT} and *Foxn1*^{R} alleles resulted, *Foxn1*$^{+/+}$ (WT), *Foxn1*$^{R/+}$ (R/+), *Foxn1*$^{+/-}$ (Het), *Foxn1*$^{R/R}$ (R/R), *Foxn1*$^{R/-}$ (R/-). Additionally, E12.5 TECs were also obtained from *Foxn1*$^{nu/nu}$ (Nude) mice, in which a targeted disruption of the *Foxn1* locus has been used (Nehls et al., 1996). RNA obtained from TECs isolated from Nude mice represent the transcriptional profile of these cells in the complete absence of *Foxn1*. Relative levels of expression for the WT, R/+, R/R, R/- and Nude samples are shown in **Figure 2.1** (left panel).

2.2.1.3 An RBPJ-mutant series of fetal TEPC data

Foxn1^{Cre} mice were crossed with an *Rbpj* conditional knockout (cKO) line (**Dong Liu, Blackburn lab**), creating a mouse model line (*Foxn1*^{Cre}*Rbpj*$^{Fl/Fl}$ mice), where RBPJ was absent from all TECs (*Rbpj* exon deletion is under the control of *Foxn1* promoter). RNA was obtained from EpCAM^{+}PLET1^{+} wild type (WT) and *Foxn1*^{Cre}*Rbpj*$^{Fl/Fl}$ (RBPJ cKO; loss-of-function; LOF) TEPC at E12.5 (**3 biological replicates per dataset**) and PLET1^{+} and PLET1^{-} TEPC at E14.5 (**3 biological replicates per dataset**) by **Dong Liu**; **Blackburn lab** (Liu et al., 2017, submitted) and samples were sent for sequencing. EpCAM^{+}PLET1^{+} mark the early developmental stages of the thymic epithelium and have been linked to progenitor activity of TECs. The cell populations analysed were chosen since, at E12.5, although the PLET1^{+} TEPC population is already heterogeneous, and will contain cortical TEC (cTEC)-fated cells along with common TEPCs and potentially medullary TEC (mTEC)-restricted progenitors, it cannot be split on the basis of known cell surface markers. At E14.5 prospective mTECs appear to be contained within the PLET1^{+} population, while prospective cTECs have down-regulated this marker.

2.2.1.4 TEC subpopulations from newborn and adult mice

Raw RNA-seq data for cortical (sorted for $CD45^{-}EpCAM^{+}Ly51^{+}UEA1^{-}$) and medullary (sorted for $CD45^{-}EpCAM^{+}Ly51^{-}UEA1^{+}$) TEC populations from newborn (7 days old) mice (**one biological replicate per population**) were obtained from the GEO public repository (Barrett et al., 2013), under the GEO accession number: **GSE44945** (generated by **St-Pierre et al., 2013**). To isolate these populations, thymic stromal cells were stained for antibodies (Abs) against CD45, EpCAM and Ly51 and for the plant lectin ulex europeus agglutinin-1 (UEA-1). CD45 is a hematopoietic pan-epithelial marker expressed by both hematopoietic stromal cells and thymocytes. Unlike thymocytes, TECs that are not of hematopoietic origin do not express this marker (Rodewald, 2008). Ly51 Ab recognizes a glutamyl-aminopeptidase that is solely evident in cTECs, while mTECs uniquely bind the plant lectin UEA1 (Gray et al., 2002). Additionally, raw RNA-seq data for similarly sorted cTEC and mTEC populations (**biological duplicates per population**) from newborn (7 days old) mice were also obtained from the GEO public repository, under the GEO accession number: **GSE53110** (generated by **Sansom et al., 2014**). These cells, however, were also sorted for the MHCII marker; MHCII can be used to mark specifically mature TECs (Jenkinson et al., 1981). In respect to this, Yang et al. (2006) demonstrated that the majority of 1 week old cTECs and mTECs express MHCII markers. Thus, these data were used as biological replicates of the St-Pierre generated datasets and combined together for further analysis. Under the same GEO accession number (**GSE53110**, generated by **Sansom et al., 2014**), RNA-seq data from immature ($CD45^{-}EpCAM^{+}MHCII^{lo}Ly51^{-}UEA1^{+}$), mature ($CD45^{-}EpCAM^{+}MHCII^{hi}Ly51^{-}UEA1^{+}$), mature $AIRE^{+}$, mature $AIRE^{-}$ and mature AIRE knock-out from 4 week old mice (**biological duplicates per population**) were also retrieved. AIRE has been found to lead promiscuous gene expression of tissue-related antigens (TRAs) in mTECs, which essential for T-cell maturation and it is used here as a marker to further subdivide mature mTECs (Gray et al., 2007; Tykocinski et al., 2008).

2.2.2 Collection of regulatory datasets in mouse (ChIP-seq)

For the purposes of this project, ChIP-seq datasets of TECs from newborn mice were obtained from public repositories.

2.2.2.1 A FOXN1-flagged peptide protein dataset in adult TECs

Raw ChIP-seq data for the FOXN1 protein tagged with an octapeptide (**biological duplicates**) were obtained from the GEO public repository, under the GEO accession number: **GSE75219** (generated by **Žuklys et al., 2016**). In more detail, the FOXN1-tagged protein was expressed under the control of the *Foxn1* promoter in TECs of *nude* mice ($Foxn1^{nu/nu}$) using a bacterial artificial chromosome (BAC). The genotype of homozygous mice expressing the fused FOXN1 protein is represented as $Foxn1^{wt*/wt*}$ (details of the targeting strategy are shown in **Figure 2.2**). Chromatin immunoprecipitation was performed with the M2 anti-FLAG antibody (F1804; Sigma) against the FOXN1-flagged peptide in sonicated DNA obtained from enriched TECs (using magnetic beads) from digested thymic lobes from 1 week old $Foxn1^{wt*/wt*}$ (see **Chapter 1**, **section 1.2.5.3** for mice phenotypic details), and input samples were prepared from non-immunoprecipitated chromatin in parallel .

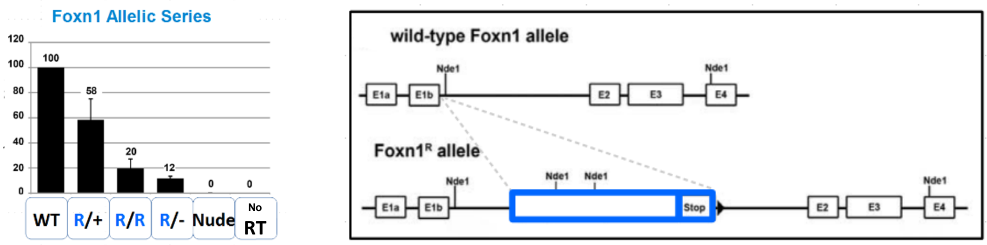


Figure 2.1: **Relative levels of expression for the *Foxn1*^{WT} and *Foxn1*^{R} variants and schematic representation of the *Foxn1*^{R} design.** (Left) QRT-PCR analysis was used to determine the relative levels of *Foxn1* expression on E12.5 with the Sybr-Green method in 5 different samples: *Foxn1*$^{+/+}$ (WT), *Foxn1*$^{R/+}$ (R/+), *Foxn1*$^{R/R}$ (R/R), *Foxn1*$^{R/-}$ (R/-) and *Foxn1*$^{nu/nu}$ (Nude) and a no RT control. The geometric mean of three housekeeping genes was used to normalise for the *Foxn1* expression levels. (Right) *Foxn1*^{R} hypomorphic allele design: Making use of ES cells and homologous recombination, a LoxP flanked cassette was inserted into intron 1b at the *Foxn1* locus. These cells were used to create a *Foxn1*^{R} mouse strain (reproduced from Nowell et al., 2011).

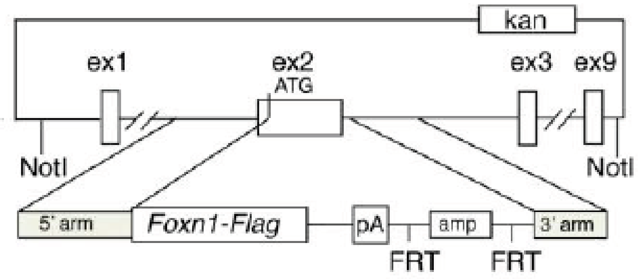


Figure 2.2: **Schematic representation of the FOXN1-tagged protein design construct.** A cDNA encoding *Foxn1* fused with three Flag sequences at its C-terminus, designated *Foxn1*-Flag was placed in exon 2 of the normal *Foxn1* locus through homologous recombination of a bacterial artificial chromosome (reproduced from Žuklys et al., 2016). FRT: Flipase recombinase target.

2.2.3 Collection of accessibility datasets in mouse (ChIP-seq)

For the purposes of this project, ChIP-seq datasets of TEPCs from fetal mice were generated in the lab.

2.2.3.1 Histone modification data from fetal TEPCs

Chromatin immunoprecipitation was performed with antibodies specific for tri-methylation of lysine 4 on histone H3 (H3K4me3) and acetylation of lysine 27 on histone H3 (H3K27ac) and core histone H3 (panH3; used as control) in sonicated DNA obtained from EpCAM$^+$PLET1$^+$ TECs from mice at day E12.5 of embryonic development (**biological duplicates per histone modification**) by **Harsh Vaidya**, **Blackburn lab** (unpublished) and samples were sent for sequencing. Histone modifications are covalent post-translational modifications that occur in histones and can influence the higher order of chromatin structure. Specific modifications lead genome compartmentalisation into distinct domains, such as transcriptionally active euchromatin or transcriptionally silent heterochromatin (summarised by Martin and Zhang, 2005). Therefore, histone modification marks can be used to assess gene expression and transcription factor occupancy. All above marks are related to open chromatin configuration and are thus characteristic of euchromatin regions. In particular, H3K4me3 marks are present in active promoters (Heintzman et al., 2007), while H3K4me1 and H3K27ac marks are often used in combination to provide a robust readout of active enhancers in the genome (Creyghton et al., 2010; Heintzman et al., 2009; Rada-Iglesias et al., 2011; Zentner et al., 2011). PanH3 marks all H3 histones in the genome, offering a uniform background to identify enrichment of histone modification marks. EpCAM$^+$PLET1$^+$ mark the early developmental stages of the thymic epithelium and have been linked to progenitor activity of TECs (Depreter et al., 2008; Farr and Anderson, 1985).

2.3 BIOINFORMATICS ANALYSIS AND WORKFLOWS

This section presents the generic bioinformatics pipelines that have been put together to pre-analyse the RNA-seq and ChIP-seq datasets provided, as well as the core processing analysis steps that have been undertaken in order to remove sources of technical bias in the aforementioned datasets. Representative graphs of the RNA-seq and ChIP-seq pre-analysis and core analysis pipelines are provided at the end of each section.

2.3.1 A common pre-analysis step in the RNA-seq and ChIP-seq analysis pipelines

2.3.1.1 Raw reads quality control

The FastQC tool (version 0.11.5, Andrews, 2010) was used to inspect the raw read files (saved in FASTQ format, Cock et al., 2010) of all the datasets described in **section 2.2** to decide if the sequenced reads require further trimming before aligning them to the genome. FastQC is a commonly used tool that assesses the sequences quality and GC content of the reads, as well as, the presence of adapters and overrepresented *k*-mers. Quality control by FastQC suggested few low-quality reads and a small amount of adapter contamination in most of the samples, so low quality reads had to be removed and the remaining reads had to be trimmed of adapter sequences with Trimmomatic using default parameters (version 0.32, Bolger et al., 2014). For single-end (SE) or paired-end (PE) reads, the SE/PE parameter was selected accordingly. Also, when adapter contamination from a specific technology was detected, the pre-set adapters were used to trim the reads; in the case that the adapter's origin was not identified, reads were trimmed for the over-represented sequence detected. Post-trimmed reads were assessed by FastQC once again to verify that adapters and low-quality reads had been removed properly and that trimming did not introduce any bias itself (for a representative diagram see upper section in **Figure 2.3** or **Figure 2.6**).

2.3.2 RNA-seq pre-analysis pipeline

2.3.2.1 Mapping reads to genome

Trimmed reads that passed QC (**section 2.3.1.1**) were aligned against the GRCm38.p5 (mm10) Mouse assembly (downloaded from GENCODE – release M12) with STAR Aligner (version 2.5.1a, Dobin et al., 2013). STAR is a splice aware read aligner that uses a RAM-intensive approach which significantly increases its mapping speed, allowing this tool to perform equally well or better than newer aligners such as HISAT2 (Kim et al., 2015), while it has also outperformed all tools of its time (reviewed in Baruzzo et al., 2016). In the same review, STAR also managed to identify the highest percentage of correctly aligned reads using its default parameters, an important feature that an aligner need to possess, since there is no clear way for parameters optimisation in real data. STAR ran with default parameters except for the *--runThreadN* parameter, which was set to 32 (the maximum number of double precision tasks that our computer facility can run in parallel; 8 processors x 4 cores each) to increase execution speed even more. The comprehensive ENCODE annotation file for the GRCm38.p5 assembly which contains gene annotation on the reference chromosomes in a general transfer format (GTF) was downloaded from GENCODE (Harrow et al., 2006, release M12) and incorporated in the run, in order to extract splice junctions of the genome and improve the accuracy of the mapping.

2.3.2.2 Quantification of gene counts

The featureCounts program (version 1.5.1, Liao et al., 2014) was used to quantify the gene expression levels from the STAR-aligned files by counting the number of mapped reads per gene based on a GTF file (the same one used in **section 2.3.2.1**) that includes the gene coordinates for the GRCm38.p5 assembly. featureCounts was executed with default parameters.

An overview of the RNA-seq pre-analysis pipeline is shown in the graph in **Figure 2.3**.

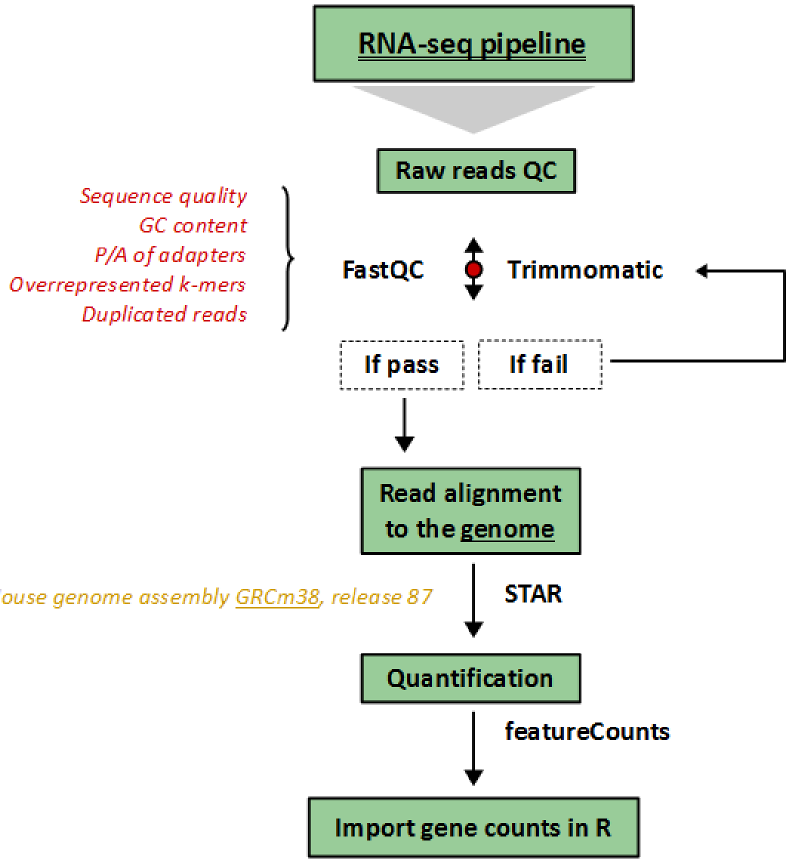

RNA-seq pipeline
Raw reads QC
Sequence quality
GC content
P/A of adapters
Overrepresented k-mers
Duplicated reads
FastQC
Trimmomatic
If pass
If fail
Read alignment to the genome
Mouse genome assembly GRCm38, release 87
STAR
Quantification
featureCounts
Import gene counts in R

Figure 2.3: Overview plot of the RNA-seq pre-analysis pipeline. In this pipeline, each step shows the type of analysis performed and the specific tool that was selected when the workflow was executed. Raw reads are assessed with FastQC (Andrews, 2010) for different types of bias (shown in red), namely, high sequence quality, organism specific GC content, absence of adapters, absence of overrepresented *k*-mers and normal duplication rate. Reads which fail to qualify for any of the quality control (QC) criteria above are removed with Trimmomatic (Bolger et al., 2014). After read trimming with Trimmomatic, reads which passed QC are reassessed with FastQC to ensure no bias has been introduced (see **section 2.3.1.1** in the main text). Reads which passed QC are then mapped to the provided genome assembly (GRCm38) with STAR aligner (Dobin et al., 2013; see **section 2.3.2.1** in the main text) and gene counts are quantified with featureCounts (Liao et al., 2014; see **section 2.3.2.2** in the main text) before raw gene counts are imported into R for further processing and analysis.

2.3.3 RNA-seq core analysis pipeline

2.3.3.1 Low gene counts

The gene count tables per data series (for details see **section 2.2.1**) calculated with featureCounts (**section 2.3.2.2**) were imported into R and were converted into DGEList objects (edgeR package, Robinson et al., 2010) for further processing. The majority of these genes were not expressed at a biologically meaningful extent across all samples (per dataset) and had to be removed prior to downstream analysis to reduce the number of genes that would be investigated, as well as the number of tests that would be performed. To define the cut-off value under which genes would be discarded, the distribution of the log2-transformed counts per million (CPM) values for each of the transcriptomic datasets was inspected with the *density* function in R. Inspection of the resulting density plots showed that a large proportion of genes were unexpressed or very lowly expressed (see example plot in the left panel of **Figure 2.4**). The CPM threshold for each dataset was chosen based on each dataset's density plot to sensibly separate non-expressed or lowly expressed genes from genes which are expressed. A threshold of 1 was selected for all data series; this is also the suggested threshold in the RNA-seq Bioconductor tutorial for limma, Glimma and edgeR (Law et al., 2016).

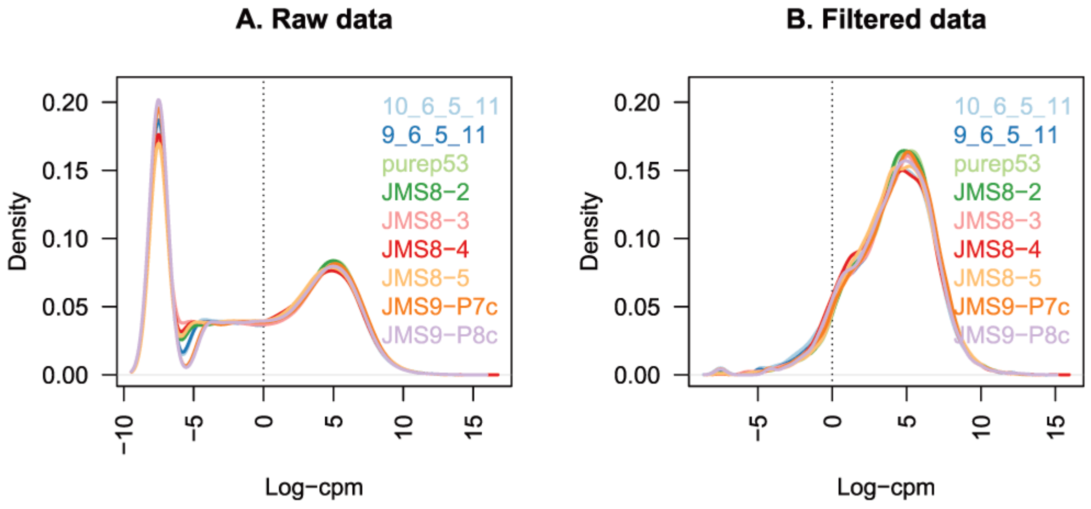


Figure 2.4: Density plots of the log2-transformed CPM values from example datasets. Density plots depict distribution of gene counts before (A) and after (B) filtering of the data based on a given expression (log2-CPM) threshold. This threshold is chosen on the grounds of removing unexpressed or very lowly expressed genes across all samples, while maintaining genes that are expressed in at least in one of the samples, or in at least one of the sample groups if replicates are available (reproduced from Law et al., 2016).

2.3.3.2 Data normalisation and differential expression analysis

Prior to gene-wise comparisons, normalisation across the samples is required to bring all distributions to the same scale in order for the future comparisons to be meaningful. The TMM normalisation (normalisation by trimmed mean of M values) was applied via the c*alcNormFactors* function (edgeR package) in R to calculate the normalisation factors which represent the scaling factors per library. These scaling factors are included in the DGEList object alongside the raw counts. The *voom* function from the limma package (Ritchie et al., 2015) converts the raw count data into log2-transformed CPM values by extracting both the library sizes and the scaling factors from the DGEList object. The *voom* function also weights the gene counts based on their mean-variance relationship and it is preferred to *limma-trend* [or simply the *cpm* function (McCarthy et al., 2012; Robinson and Smyth, 2007) from EdgeR (Robinson et al., 2010)] because its performance is more robust when library sizes vary among samples. The resulting *voom* object includes among others, a column with the normalised log-CPM counts that can be also used for visualisation purposes.

2.3.3.3 Data inspection and visualisation

Hierarchical clustering and Principal Component Analysis (PCA) were used to visualise and inspect the structure of the sample data. Both methods are unsupervised (i.e. no information of the samples relationships is provided prior to the clustering) and are therefore suitable for exploratory analysis of sample associations. Samples groupings need to be representative of the biological design of the experiment to proceed with any comparative analysis between groups, and generate results that are meaningful and reliable. Hierarchical clustering divides samples into homogeneous groups by maximising their distances based on a provided distance measure. This method can be used for inspection of datasets with a simple biological design. PCA identifies more complex sample relationships by representing multiple patterns characteristic of high levels of variance across the samples. This method is used here to inspect the relationships between more complex datasets (e.g. merged data series). Hierarchical clustering was performed using the *hclust* function (Müllner, 2013) in R using average dissimilarity of samples based on Pearson correlation (Pearson, 1895) to define groups. PCA was performed with the *prcomp* function in R while the resulting principal components were coupled with the Analysis of Variance (ANOVA) method to identify strong associations of any of these principal components with class variables (edited code for the *PrincipalComponentANOVA* function is available in the thesis electronic supplement; initial code has been provided by Jonathan Manning; inspired from http://rnbeads.mpi-inf.mpg.de/index.php). Class variables may be representative of the experiment's biological design as well as of potential extraneous or technical effects; association of principal components with the class variables can inform of potential batch effects in the inspected dataset.

2.3.3.4 Batch effect correction

Batch effect correction across samples was performed with the *Combat* function from the sva package (Leek et al., 2016) in R to remove technical sources of bias. Batch effect correction was only possible when technical effects present among the samples of a specific dataset were not confounded by the biological design of the experiment. After batch effect correction, datasets were replotted (see visualisation methods in **section 2.3.3.3**) to confirm that the extraneous effects had been removed while no new type of bias had been introduced to the data.

2.3.3.5 Differential expression analysis

Comparative analysis between groups of samples was performed using functions from the limma package (Ritchie et al., 2015). For the differential expression analysis, the TMM-normalised CPM values (see **section 2.3.3.1**) were provided to the *lmFit* function to calculate a linear model which has the optimal fit to the provided gene expression matrix. When batch effects were present across the samples, they were included in the design of the linear model to be considered for the model fitting. The *lmFit*-generated model was then provided to the *eBayes* function from the same package to perform statistical analysis between the groups of interest and calculate gene fold changes between groups (in a log2-transformed scale) and a measure of assigning confidence to each gene comparison. Genes were considered as differentially expressed when their adjusted p-value (calculated by the BH method, as known as false discovery method; FDR method, Benjamini and Hochberg, 1995) was equal to or less than 0.05 and their $|FC| \geq 1.5$ (or 0.585 in log2 format). By selecting this threshold, 5% of the differentially expressed genes could constitute false positives. The generated gene lists with the FC values per gene were also used for pathway enrichment analysis when appropriate.

2.3.3.6 Pathway enrichment analysis

The Gene Set Enrichment Analysis (GSEA, Subramanian et al., 2005) software was used to predict signalling pathways which were enriched between groups of interest. The GSEA's '*Run GSEAPreranked'* module was used to search for enriched pathways provided a differential expression gene list (generated using the limma package in R) as the ranked list of features and an "edited" version of the ConsensusPathDB database (Kamburov et al., 2011) as the pathways database to search against, while the remaining parameters were set as default (number of permutations: 1000, collapse datasets to gene symbols: false). The '*Run GSEAPreranked'* function allows numeric measurements to be incorporated into the statistical analysis, as a way to quantify the contribution of particular genes to pathway enrichment scores, in this case, by considering the fold change per gene. The ConsensusPathDB was selected on the basis that it offers an up-to-date repository of biological pathways that have been collected and integrated from multiple known databases and it included in total 2,140 mouse-specific pathways (2,140 was the number of pathways at the time of this analysis; the actual number has increased since, reaching 2,173 pathways that are mouse specific). Because FOXN1 consists one of the central regulators of the thymic system, an extra pathway was added to the ConsensusPathDB ("edited") to incorporate the 450 FOXN1 high confidence target genes in cortical thymic epithelial cells that have been published recently (Žuklys et al., 2016) to assist towards this direction. The resulted pathways were filtered for an FDR value ≤ 0.25 (default parameters of GSEA), which suggests that there is 25% probability the gene set (or pathway) that is found enriched to be a false positive result.

An overview of the RNA-seq core analysis steps pipeline is shown in the graph in **Figure 2.5**.

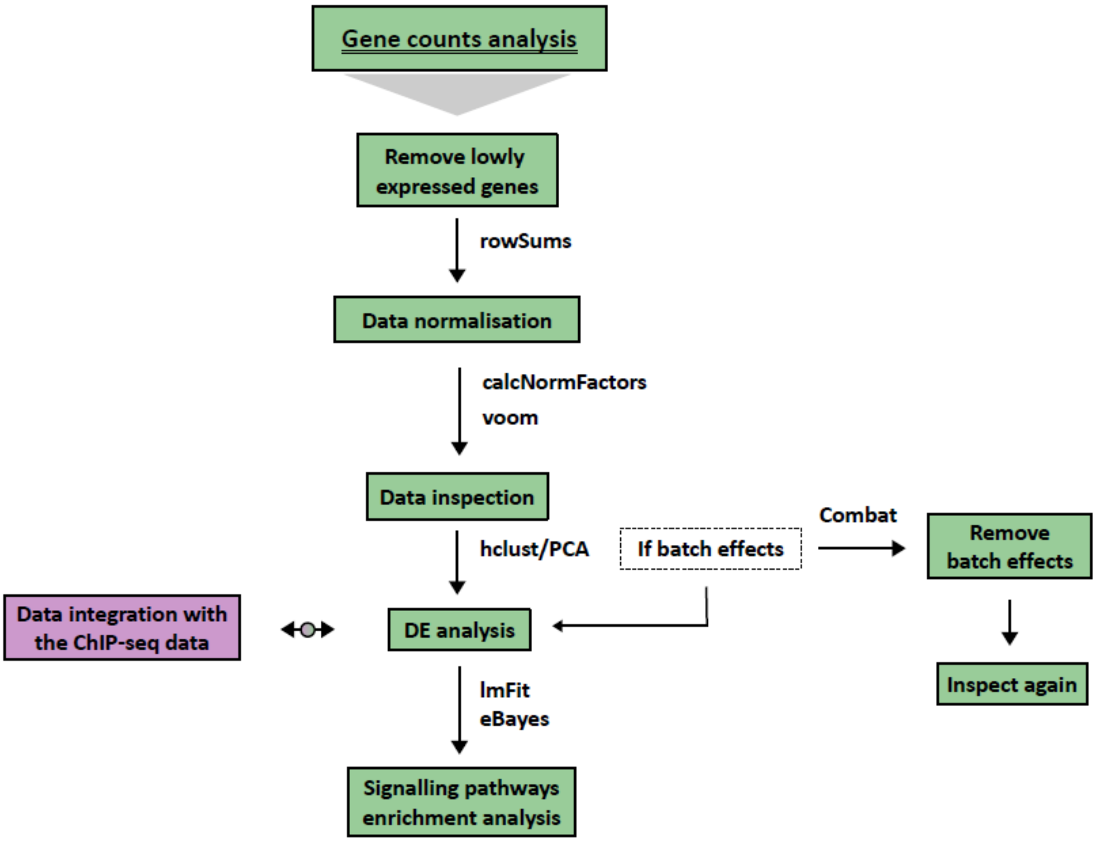

Gene counts analysis
Remove lowly expressed genes
rowSums
Data normalisation
calcNormFactors
voom
Data inspection
hclust/PCA
If batch effects
Combat
Remove batch effects
Inspect again
Data integration with the ChIP-seq data
DE analysis
lmFit
eBayes
Signalling pathways enrichment analysis

Figure 2.5: Overview plot of the RNA-seq core analysis pipeline. Each step shows the type of analysis performed and the specific function/tool that was selected in each case. After the gene counts per data series have been imported in R (see **Figure 2.3**), lowly or unexpressed genes are removed using the *rowSums* function from the edgeR package (Robinson et al., 2010; see **section 2.3.3.1** in the main text). Gene counts across samples are then normalised for their library size with the *calcNormFactors* and *voom* functions from the limma package (Ritchie et al., 2015) to make samples equal in terms of their statistical properties and allow meaningful comparisons between samples (see **section 2.3.3.2** in the main text). Hierarchical clustering and PCA plots were used to visually inspect the data (see **section 2.3.3.3** in the main text). If batch effects were apparent in the samples, the *Combat* function was used to remove the detected batch effect and the data were replotted (as before) to examine whether the artificial effect has been removed. Differential expression analysis was then performed using the *lmFit* and *eBayes* functions from the limma package (see **section 2.3.3.5** in the main text). The differentially expressed gene tables were used for integration with the ChIP-seq data (see **section 2.3.5.3** and pipeline in **Figure 2.7**). The same differentially expressed gene lists with the log2 fold change values were also used for signalling pathway enrichment analysis (see **section 2.3.3.6** in the main text).

2.3.4 ChIP-seq pre-analysis pipeline

2.3.4.1 Mapping reads to genome

Trimmed reads which passed QC were aligned against the GRCm38.p5 (mm10) Mouse assembly (downloaded from Gencode – release M12) with Bowtie2 (version 2.3.0, Langmead and Salzberg, 2012). Bowtie (Langmead et al., 2009) uses full-text minute indexing to produce ultra-fast and memory-efficient alignment of reads against any genome assembly. Bowtie2 extends Bowtie's flexible indexing with the capabilities of dynamic programming algorithms to allow alignment of longer reads against relatively long genome assemblies (e.g. mammalian) with high accuracy and sensitivity without impacting on the alignment's speed. These attributes of Bowtie2 fit the nature of the provided ChIP-seq data (see **section 2.2.2** and **2.2.3**) that are characterised by long sequences (50-70 base pairs) which will then be aligned against the mouse genome. Parameters for Bowtie2 were kept as default with both reads to be provided (*-1 read1 -2 read2*) when the datasets were of paired-end design, while the *-U* parameter was selected when reads were of single-end design. To increase sensitivity of the analysis, multi-mapping reads (reads that confidently map in more than one location in the genome) were excluded based on Bowtie's mapping quality (MAPQ) score; only uniquely aligned reads were considered for downstream analysis. A MAPQ score of ≥ 2 would exclude all true multi-mapping reads and also uniquely aligned reads with ≥ 4 mismatches (see proof from Urban, 2014). After multi-mapping reads were removed, to avoid PCR amplification bias in our results the *MarkDuplicates* function from Picard (version 1.141-1, Picard tools, 2016) was used to further remove duplicated reads (reads that map to the exact same location more than once).

2.3.4.2 Peak calling

MACS2 (version 2.1.0.20150731, Zhang et al., 2008) was used for peak calling provided the non-redundant uniquely aligned reads as the treatment file (*-t*) and their respective background as the control one (*-c*). MACS2 can be used to identify two types of enrichment: a) sharp (or narrow) domains, these are domains marked by transcription factors or histone modifications which mark only short genomic regions (e.g. gene promoters) and b) broad domains, these are domains marked by histone modifications that expand in longer stretches of the genome (e.g. full gene bodies). For sharp peaks (e.g. the FOXN1-tagged and the H3K4me3 ChIP-seq data) the *p*-value cut-off to call significant peaks was set to a more lenient threshold than the default (-*p* 1e-3) and the Irreproducible Discovery Rate (IDR) pipeline (available at https://github.com/nboley/idr) was then used to identify peaks consistent among replicates (for details over the specific thresholds and the IDR pipeline see **section 2.3.5.1**). For broad peaks (e.g. the H3K27ac ChIP-seq data), MACS2 was used with the --*broad* parameter on (IDR pipeline not applicable yet for broad peaks), a --*broad-cutoff* of 0.1 for the broader enriched regions and a -*q-value* cut-off of 0.05 for any significant narrow peaks included in that broader enriched region. For the histone modification datasets that have low sequencing depth, the MACS2 --*nomodel* parameter was selected for MACS2 not to build a shifting model when it calculates the fragment length. Instead, the fragment length was provided in the --*extsize* parameter equal to 147 nucleotides long which is representative of the DNA length that spreads across a nucleosome since the histone modification data have an underlying characteristic 147bp resolution.

An overview of the ChIP-seq pre-analysis steps pipeline is shown in the graph in **Figure 2.6**.

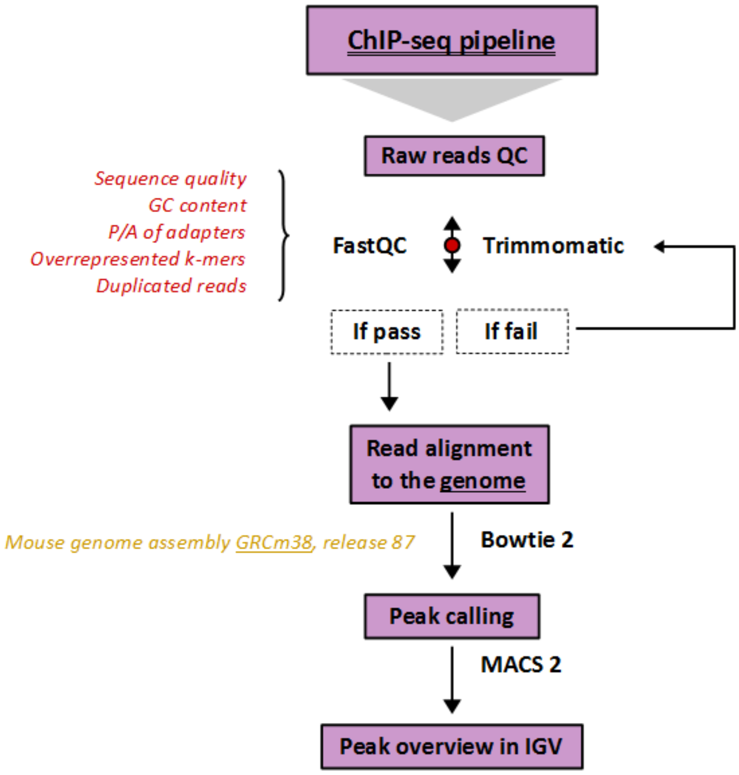

ChIP-seq pipeline
Raw reads QC
Sequence quality
GC content
P/A of adapters
Overrepresented k-mers
Duplicated reads
FastQC
Trimmomatic
If pass
If fail
Read alignment to the genome
Mouse genome assembly GRCm38, release 87
Bowtie 2
Peak calling
MACS 2
Peak overview in IGV

Figure 2.6: Overview plot of the ChIP-seq pre-analysis pipeline. In this pipeline, each step shows the type of analysis performed and the specific tool that was selected when the workflow was executed. Raw reads are inspected with FastQC (Andrews, 2010) for different types of bias (shown in red), namely, high sequence quality, organism specific GC content, absence of adapters, absence of overrepresented *k*-mers and normal duplication rate. Reads which fail to qualify for any of the quality control (QC) criteria above are removed with Trimmomatic (Bolger et al., 2014). After read trimming with Trimmomatic, reads that passed QC are reinspected with FastQC to make sure no bias has been introduced (see **section 2.3.1.1** in the main text). Reads which passed QC are then mapped to the provided genome assembly (GRCm38) with Bowtie2 (Langmead and Salzberg, 2012; see **section 2.3.4.1** in the main text) and peaks are called from the generated aligned files with MACS2 *(Zhang et al., 2008)*; see **section 2.3.4.2** in the main text). Mapped read files can be further inspected in the genome browser with IGV (Thorvaldsdottir et al., 2013; see **section 2.3.5.2** in the main text).

2.3.5 ChIP-seq core analysis pipeline

2.3.5.1 Significant peaks

When calling narrow peaks from samples with replicates, the IDR pipeline (https://sites.google.com/site/anshulkundaje/projects/idr) was used to draw a line between consistent and non-consistent peaks, based on peaks reproducibility among replicates. In order to assess peaks consistency based on the IDR value, non-consistent findings (peaks) need to be included in the called peaks list and this is why the *p*-value cut-off for MACS2 was set to 1e-3; a more relaxed threshold for peak calling. After running the *batch-consistency-analysis.r* script, included in the IDR pipeline, only consistent peaks with an IDR value of ≤ 0.05 were considered significant as suggested from the pipeline. In the case of broad peak datasets, the broad peaks *q*-value (*--broad-cutoff*) was used to assess if peaks were significantly enriched over background and it was set to 0.1 (default by MACS2 for broad peaks).

2.3.5.2 Data inspection

Mapped read files were loaded to IGV (version 2.3.72, Robinson et al., 2011; Thorvaldsdottir et al., 2013) for visual inspection of the aligned reads enrichment over the control samples (or input DNA).

2.3.5.3 Peaks to genes

The ChIPseeker package (Yu et al., 2015) in R was used to match peaks to proximal genes based on genomic distance. Genes were assigned to peaks if the peak was 5,000 (-5kb) base pairs (bp) upstream or 3,000 (+3kb) bp downstream (arbitrary threshold) genes' transcription start site (TSS) in the mouse genome.

2.3.5.4 Motif analysis

De novo motif discovery analysis was performed with the web interfaces of MEME-ChIP (Machanick and Bailey, 2011) and RSAT peak-motifs (Thomas-Chollier et al., 2011) tools with default parameters to identify enriched binding patterns under a provided set of identified binding peaks. Motif discovery tools rely on different underlying algorithms and set of parameters to identify enriched motifs in a provided set of sequences, therefore results can often vary a lot among them and a combination of tools should be used to allow better coverage of existing motifs, as was suggested by Tompa et al. (2005). Additionally, a survey of web tools for motif discovery suggested that enriched motifs that were consistent among different tools are more reliable (Tran and Huang, 2014). Commonly identified motifs by MEME-ChIP and RSAT were considered as more significant compared to motifs only identified by one of the two tools.

An overview of the ChIP-seq core analysis steps pipeline is shown in the graph in **Figure 2.7**.

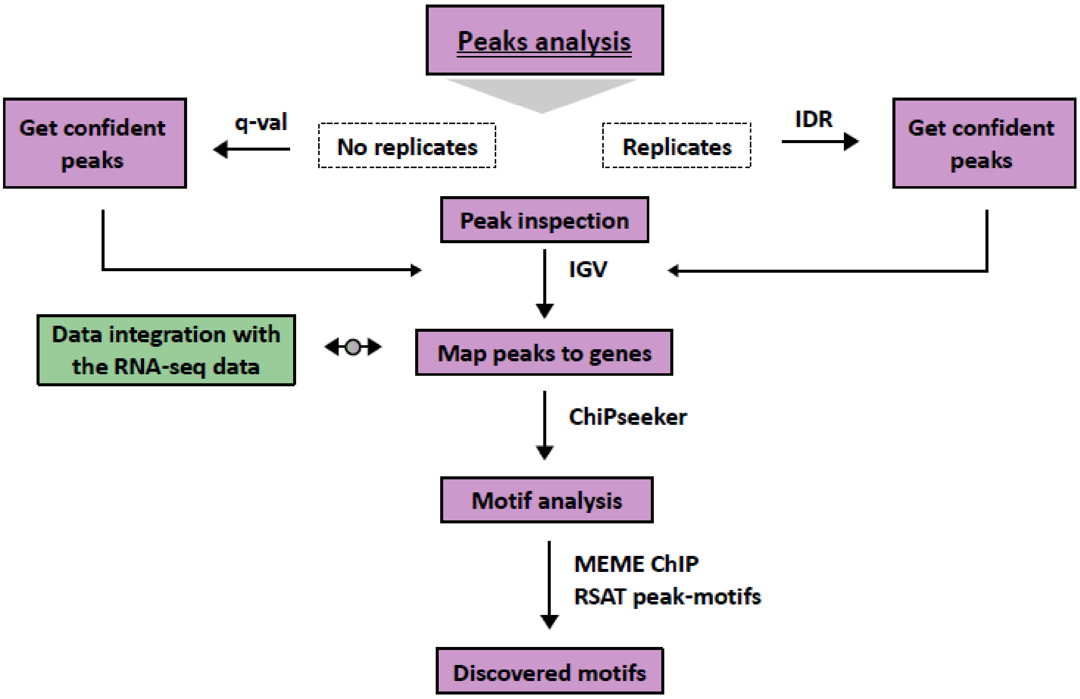

Peaks analysis
Get confident peaks
q-val
No replicates
Replicates
IDR
Get confident peaks
Peak inspection
IGV
Data integration with the RNA-seq data
Map peaks to genes
ChiPseeker
Motif analysis
MEME ChIP
RSAT peak-motifs
Discovered motifs

Figure 2.7: **Overview plot of the ChIP-seq core analysis pipeline.** Each step shows the type of analysis performed and the specific function/tool that was selected in each case. To call confident peaks, the *q*-value provided by MACS2 was used as a threshold when replicates were not available, while the IDR pipeline and an IDR threshold was used for peaks when replicates were available. Enriched peaks were further inspected in the genome browser with IGV (Thorvaldsdottir et al., 2013; see **section 2.3.5.2** in the main text). For integration of the ChIP-seq data with the RNA-seq data (described in **section 2.2.1** and analysed based on pipelines in **Figure 2.3** and **Figure 2.5**), peaks were mapped to genes based on peaks proximity to the transcriptional start site (TSS) of selected genes (-5 kb, + 3 kb from genes TSS). Peaks were also converted into DNA sequences (area under the peak) and motif discovery analysis was performed with MEME-ChIP and RSAT peak-motifs for a given a set of sequences to identify enriched motifs under the peaks of interest.

Chapter 3

Modelling the developmental progression of early TEPCs during thymus development

Chapter 3

Modelling the developmental progression of early TEPCs during thymus development

3.1 PRÉCIS

Thymic epithelial cells (TECs) constitute essential elements of the thymic microenvironment that orchestrate T lymphocyte (T-cell) differentiation and repertoire selection (Takahama, 2006) and therefore are requisite for development of a functional immune system. TECs derive from a thymic epithelial progenitor cell (TEPC) that has the capacity -bipotent- to differentiate into the functionally distinct cortical (c) and medullary (m) TEC sublineages under the control of FOXN1 (Bennett et al., 2002). Increasing levels of *Foxn1* determine phenotypically the status progression of TECs' differentiation (low *Foxn1* levels allow exit from the TEPC state and initiation of TEC differentiation, intermediate/high *Foxn1* levels allow more specialised TEC functions), while TECs lineage specification and the regulation of cell fate choice between cTEC/mTEC divergence take place even in the absence of *Foxn1* (Nowell et al., 2011). Despite the vital role of TECs in the establishment of a functionally adequate thymic system, the precise mechanisms through which TEC differentiation, and divergence of the cTEC and mTEC sublineages occur are not yet fully decrypted.

Therefore, with the work in **Chapter 3** I aim to answer a series of questions in relevance to the early TEPC status and progression upon *Foxn1* expression in early mouse development.

Question 1: What is the impact of differing *Foxn1* expression levels on the early developmental progression of TEPCs at the molecular level?

Question 2: The transcription profile of which normal TEPC developmental stage better resembles the experimentally competent bipotent TEPCs that exist in the thymic rudiments of mice lacking *Foxn1* expression?

Question 3: Is there evidence of *Foxn1* acting as a repressor in the thymic epithelium?

Question 4: Can we predict a gene expression signature capable of identifying bipotent TEPCs that can be used to better isolate putative TEPC populations in the early, newborn and adult thymus?

Question 5: Do the above findings revisit the current model of TEPC differentiation summarised in Takahama et al. (2017)?

3.2 RESULTS

3.2.1 Pseudo-timing of a *Foxn1* Allelic series dataset represents normal developmental progression of TEPCs

Foxn1 expression levels comprise a definitive factor in determining thymic epithelial cell expansion and differentiation during the early, T-cell independent, fetal developmental stages and this effect has been demonstrated by phenotypic analysis of mouse thymi obtained from a *Foxn1* Allelic series (Nowell et al., 2011), in which *Foxn1* expression levels are progressively reduced due to the presence (in some variants) of the $Foxn1^{R}$ allele, which expresses around 20% of wild type (WT) *Foxn1* mRNA levels. In the same paper, phenotypic analysis demonstrated a developmental block in TEPC differentiation that was more profound in the null or low-*Foxn1* variants compared to their allelic variants that expressed intermediate or high levels of *Foxn1* (**Figure 3.1**; image edited from (Nowell et al., 2011)).

To explore whether the phenotypes caused by impaired *Foxn1* levels during early thymus development mimic normal developmental progression of TECs, I will compare the transcriptional profiles of the E12.5 $PLET1^{+}$ TEPCs isolated from the *Foxn1* Allelic series (RNA-seq data from **singular biological samples** provided by **Stephanie Tetelin**, Blackburn lab; see also **Chapter 2**, **section 2.2.1**) with the transcriptional profile of E10.5, E11.5 and E12.5 $PLET1^{+}$ TEPCs from a TEPC Developmental series (RNA-seq data from **biological triplicates per stage** provided by **Harsh Vaidya**, Blackburn lab; see also **Chapter 2**, **section 2.2.1**).

In short, there are six allelic combinations in the *Foxn1* variants [$Foxn1^{+/+}$ (WT), $Foxn1^{R/+}$ (R/+), $Foxn1^{+/-}$ (Het), $Foxn1^{R/R}$ (R/R), $Foxn1^{R/-}$ (R/-), $Foxn1^{nu/nu}$ (Nude)], with the WT and *Nude* samples defining the highest (normal) expression and complete absence of *Foxn1* respectively (see **Figure 3.1** for R allele design and relative protein levels of *Foxn1*; image edited from (Nowell et al., 2011)). The transcriptome information contained in the TEPC Developmental series represents gene expression changes that characterise overall TEPC differentiation during this time frame. The TEPC Developmental series may include FOXN1-dependent and FOXN1-

independent elements, while the E12.5 *Foxn1* Allelic series dataset can reveal the unique impact of *Foxn1* expression levels on TEPCs (i.e. FOXN1-dependent elements) independently of their progression state since all samples were obtained at E12.5. A graphical representation of how these RNA-seq series relate to each other is illustrated in **Figure 3.2**.

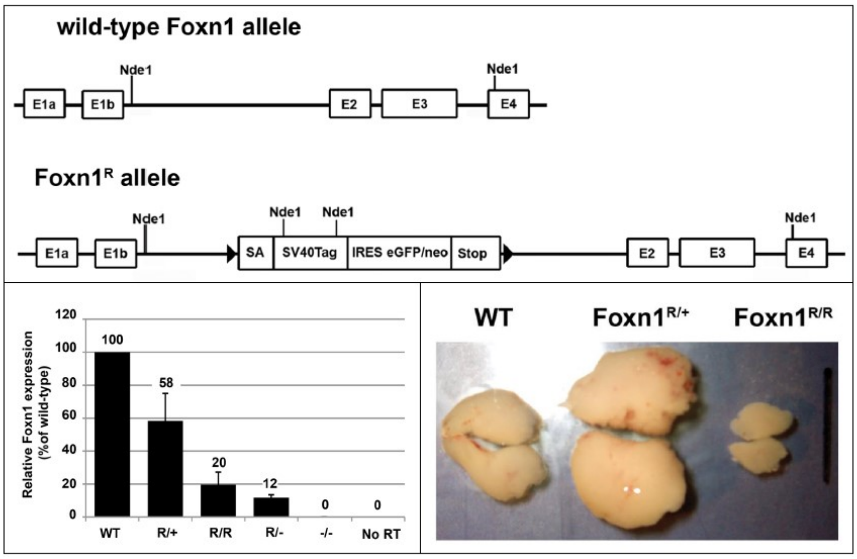

wild-type Foxn1 allele
Nde1
Nde1
E1a
E1b
E2
E3
E4
Foxn1R allele
Nde1
Nde1
Nde1
Nde1
E1a
E1b
SA
SV40Tag
IRES eGFP/neo
Stop
E2
E3
E4
Relative Foxn1 expression (% of wild-type)
120
100
80
60
40
20
0
100
58
20
12
0
0
WT
R/+
R/R
R/-
-/-
No RT
WT
Foxn1R/+
Foxn1R/R

Figure 3.1: The *Foxn1*^{R} hypomorphic allele. Making use of ES cells and homologous recombination, a LoxP flanked cassette was inserted into intron 1b of the *Foxn1* locus. These cells were used to create a *Foxn1*^{R} mouse strain (top). QRT-PCR analysis was used to determine the relative levels of *Foxn1* expression on E12.5 with the Sybr-Green method in 5 different samples. The geometric mean of three housekeeping genes was used to normalise for the *Foxn1* expression levels (bottom left). Sizes of generated thymi from the WT, R/+ and R/R samples (bottom right). Image edited from (Nowell et al., 2011).

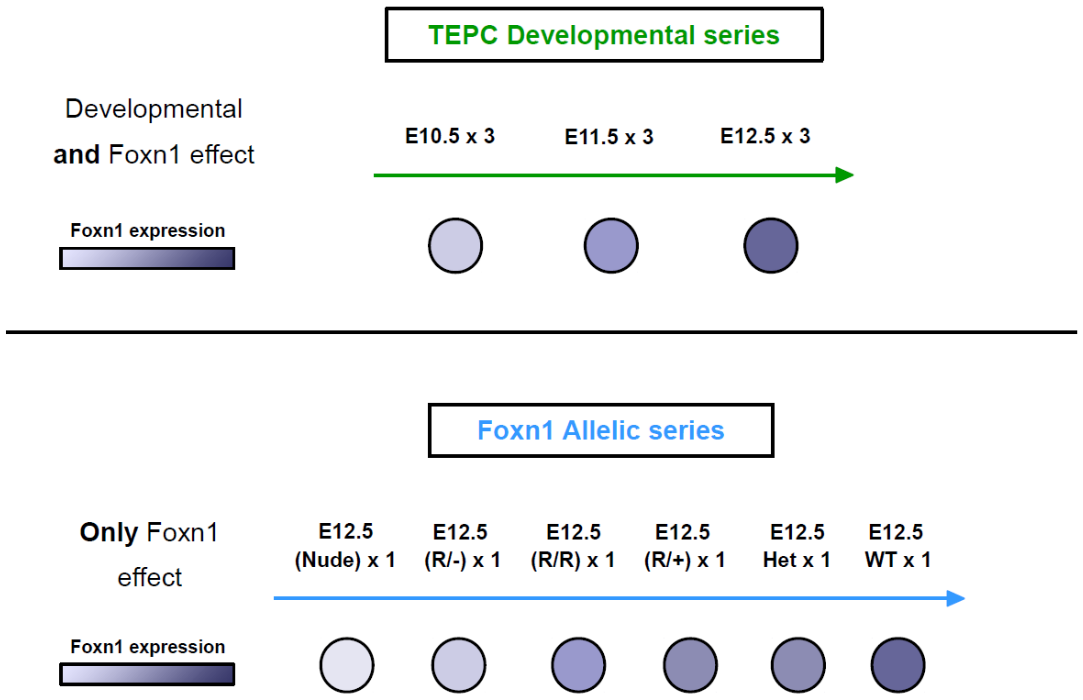
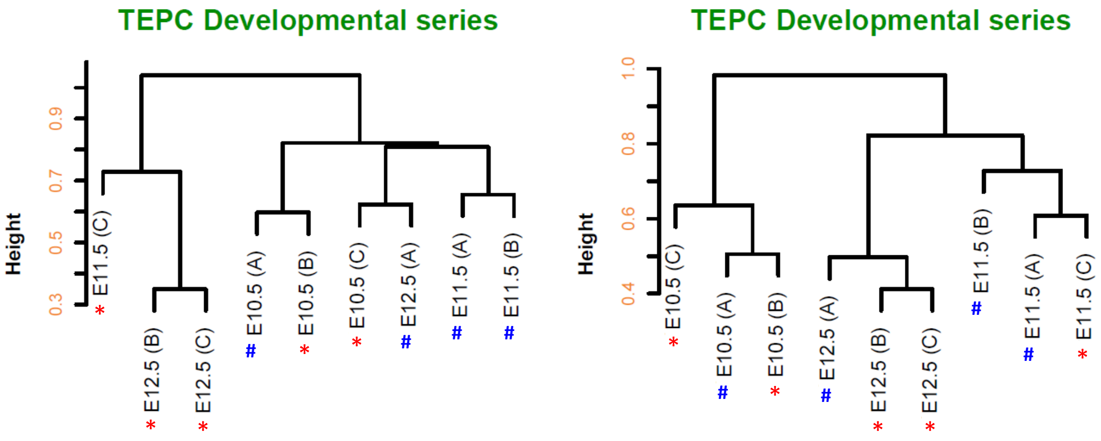

TEPC Developmental series
Developmental
and Foxn1 effect
E10.5 x 3
E11.5 x 3
E12.5 x 3
Foxn1 expression
Foxn1 Allelic series
Only Foxn1
effect
E12.5
(Nude) x 1
E12.5
(R/-) x 1
E12.5
(R/R) x 1
E12.5
(R/+) x 1
E12.5
Het x 1
E12.5
WT x 1
Foxn1 expression

Figure 3.2: Graphical representation of the TEPC Developmental and the *Foxn1* Allelic series. The image depicts the number of samples and biological replicates per data series, as well as the developmental stage at which the samples were obtained. The *Foxn1* Allelic series is linked to the TEPC Developmental series through a common developmental point that they both share. This point is day E12.5; it represents the oldest developmental point in the Developmental series, and is the timepoint at which all samples in the *Foxn1* Allelic series were collected.

For the analysis of the RNA-seq datasets in this thesis, I have put together a generic RNA-seq pre-analysis and core-analysis pipeline (described in **Chapter 2**, **section 2.3.1** and **2.3.2**) that I am using to retrieve the expression profiles (normalised gene counts) of the average cell populations in the different datasets, including the TEPC Developmental series and the *Foxn1* Allelic series datasets. Hierarchical clustering was applied to the top 1,000 most variable genes (dominant gene expression profile) per series, to examine if samples followed the biological design of each experiment. The resulting dendrogram for the TEPC Developmental series (**Figure 3.3**, left panel), shows that triplicates of each developmental stage did not cluster in their own subgroups; sample E10.5 (C) clustered with E12.5 (A) while sample E11.5 (C) clustered with the rest of the E12.5 replicates. This indicated a potential batch effect caused by samples sequenced in different lanes (different lanes marked with "#" and "*" symbols, **Figure 3.3**). Therefore, the *Combat* function from the sva package in Bioconductor (Leek et al., 2016) was used to correct for this technical effect (**Chapter 2**, **section 2.3.3.4**) and the balanced lane–subgroup design of the TEPC Developmental series allowed removal of lane effects without introducing other bias to the expression values. Batch (or lane) effect corrected data (**Figure 3.3**, right panel) clustered as expected, based on biological age.

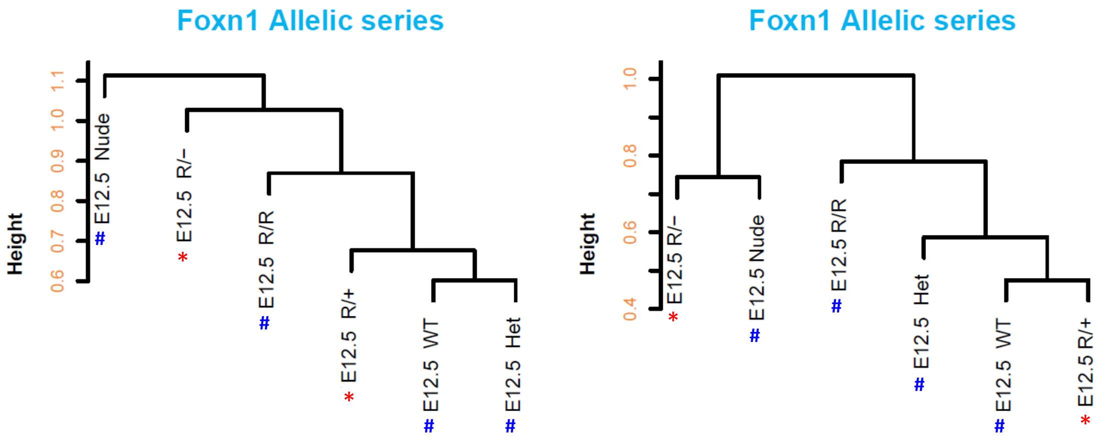
The resulting dendrogram for the *Foxn1* Allelic series dataset (top 1,000 most variable genes) revealed a sequential clustering of the samples that in general terms followed the biological design of graded drops in *Foxn1* mRNA expression levels. The lack of replicates in the *Foxn1* Allelic series made it difficult to assess extraneous effects on the data. Based on the phenotype of the generated mice, E12.5 R/- and E12.5 Nude samples would be expected to cluster separately to the rest of the samples, since only these *Foxn1* variants exhibit a complete/near complete block in TEPC differentiation and hence functional athymia. The R/R mutant generates instead a hypoplastic thymus. Furthermore, the E12.5 WT TEPC samples were expected to cluster more closely with the E12.5 R/+ rather than the E12.5 Het TEPC sample, as it expresses slightly higher levels of *Foxn1*, although all three samples generate a fully functional thymus. Despite the lack of replicates, similar phenotype samples were sequenced in different lanes creating a balanced experimental design to remove technical effects. The *Combat* function (sva package) was used again to remove the lane effect among samples (see **Figure 3.4**, right hand panel). The *Combat*-corrected data did not vary greatly from the original clustering, however, they were more consistent with known phenotypes of each allelic variant and the *Foxn1* relative expression levels (**Figure 3.1**; image edited from Nowell et al., 2011).



(before lane effect correction)

(after lane effect correction)

Figure 3.3: Cluster dendrogram of the E10.5, E11.5 and E12.5 sample triplicates from the TEPC Developmental series dataset. Triplicates did not cluster into subgroups based on the developmental stage they have been obtained (left panel), suggesting a potential lane effect among samples (lanes represented with "#" and "*" symbols). The post batch effect correction dendrogram is shown in the panel on the right, where samples have now clustered according to their biological design.



(before lane effect correction)

(after lane effect correction)

Figure 3.4: Cluster dendrogram of the *Foxn1* variant samples from the *Foxn1* Allelic series dataset. Dendrogram before batch effect correction (left panel) broadly reflects decreasing levels of *Foxn1* expression. Clustering of the samples after lane effect correction (right panel) better represents the biological design of the *Foxn1* Allelic series and the relative *Foxn1* expression levels as quantified by QRT-PCR (**Figure 3.1**, Nowell et al., 2011). Lanes represented with "#" and "*" symbols.

To compare the expression profiles of samples from the TEPC Developmental series with those from the *Foxn1* Allelic series, I further merged the two lane-effect-corrected series. Sample similarity among two experiments can be masked by technical factors that are specific to each experiment (e.g. vastly different library sizes), therefore, association of class variables (such as experiment, developmental stage and *Foxn1* levels) for the top 5,000 most variable genes from the merged dataset was plotted against the first ten principal components (code provided by Jonathan Manning, inspired from http://rnbeads.mpi-inf.mpg.de/index.php; edited code available in thesis electronic supplement), to assess if technical differences among the two different series were masking biological differences. The above ANOVA analysis suggested that PC1, accounting for 20.9% of the variance (**Figure 3.5**), is highly associated with the data series variable and therefore may be masking biologically relevant differences due to, for instance, *Foxn1* expression levels. Since neither of the two biological variables (Developmental stage and *Foxn1* levels) were confounded by the data series class and since E12.5 WT, Het and R/+ samples from the *Foxn1* Allelic series were almost identical to the E12.5 triplicates from the TEPC Developmental series, I concluded that batch effect correction would remove the technical effect caused by different data series being integrated together, without introducing any bias.

ANOVA analysis on the batch effect corrected merged dataset indicated that the data series effect had been removed. PC1 and PC3 (accounting for 17.8% and 10.4% of variation respectively) were now highly associated with the developmental stage and *Foxn1* level variables respectively (depicted in the covariate matrix, **Figure 3.6**) representing differences among samples due to biological factors and not extraneous effects.

Gene distances in PC1 were plotted against gene distances in PC3, to observe sample clustering in the integrated dataset (**Figure 3.7**). Since samples from the *Foxn1* Allelic series were all collected at E12.5, they might be expected to group together with the E12.5 triplicates from the Developmental series. However, if cell identity is significantly influenced by *Foxn1* mRNA expression levels, the samples should cluster based on the extent to which their average global expression profile has been impacted by the differing *Foxn1* levels in the *Foxn1* variants.

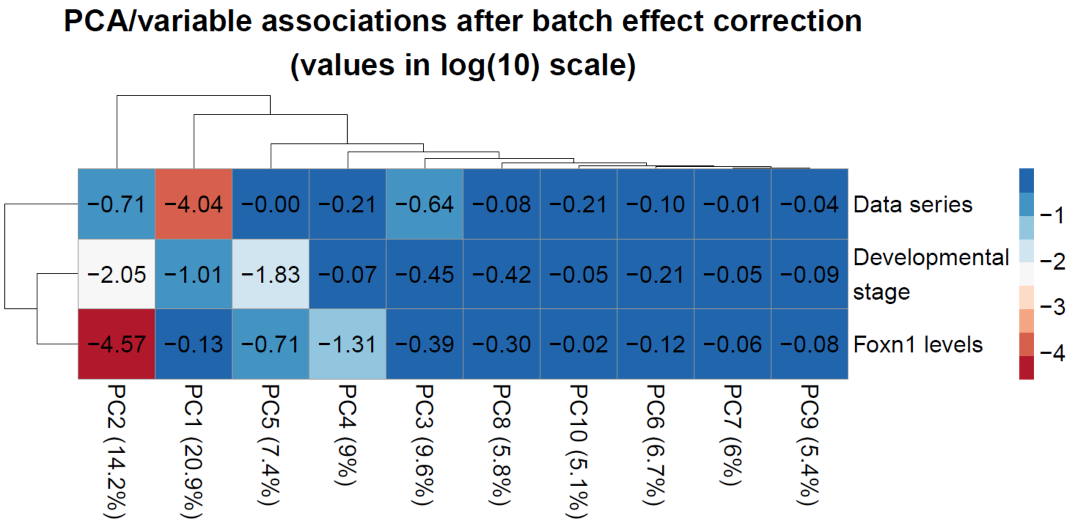


Figure 3.5: PCA/variable associations among the TEPC Developmental series and *Foxn1* Allelic series integrated dataset before batch effect correction. Heatmap of the association levels (*p*-value calculated by ANOVA function in R, original code provided by Jonathan Manning and edited version available in thesis electronic supplement) of class variables (data series, developmental stage and *Foxn1* levels) among the first ten principal components for the top 5,000 most variable genes indicates an obvious effect due to technical differences between expreriments.

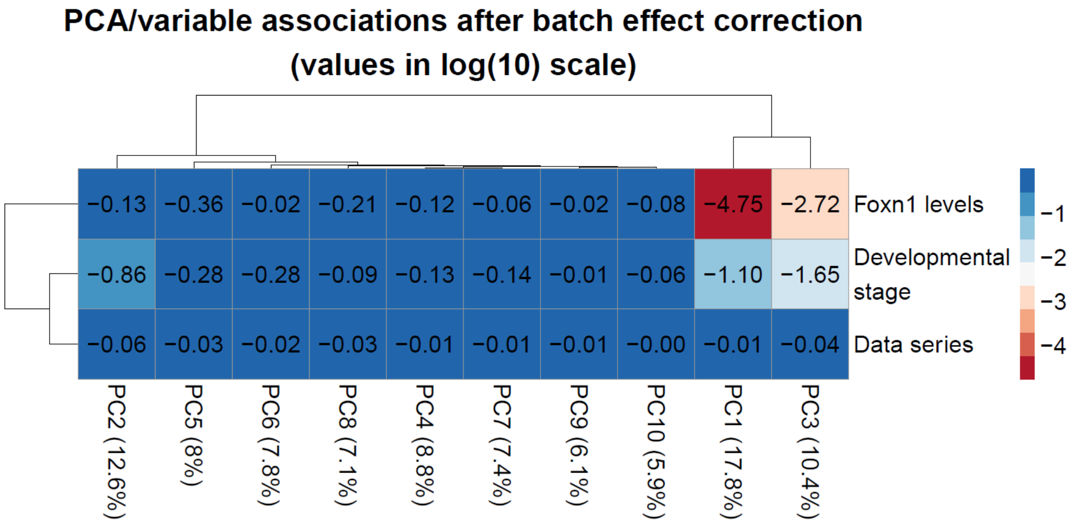


Figure 3.6: PCA/variable associations among the TEPC Developmental series and *Foxn1* Allelic series integrated dataset after batch effect correction. Heatmap of the association levels (*p*-value calculated by ANOVA function in R, original code provided by Jonathan Manning and edited version available in thesis electronic supplement) of class variables (data series, developmental stage and *Foxn1* levels) among the first ten principal components for the top 5,000 most variable genes, indicating that variation within the data reflects the biological design of the data.

In **Figure 3.7**, the *Foxn1* null/low expression thymic primordium (samples E12.5 Nude and E12.5 R/-), which comprises TEPCs in a developmentally arrested state that can function as stem cells (TESCs, see Jin et al., 2014), clusters together with the undifferentiated wild type TEPCs at E10.5 (grey circle). Sample E12.5 R/-, that expresses very low levels of *Foxn1* mRNA, groups together with samples E10.5 (C) and E11.5 (A) (**Figure 3.7**, pink circle). This is consistent with the fact that in sample E10.5 (C), *Foxn1* expression is just above the detection level, while sample E11.5 (A) contains the lowest detected levels of *Foxn1* among the E11.5 group. Sample E12.5 R/R, which expresses intermediate levels of *Foxn1*, clusters together with the two other E11.5 replicates (**Figure 3.7**, blue circle) whose profiles reflect further developmental progression than samples from the E10.5 developmental stage but less than those from E12.5. Samples E12.5 WT – Het – R/+ group clusters with the rest of the E12.5 triplicates (**Figure 3.7**, yellow circle) which agrees with the normal thymus phenotype generated by mice carrying the heterozygous and wild type genotypes.

The above groupings thus demonstrate that the E12.5 Nude and E12.5 R/- samples represent on average a very immature TEPC population, despite being isolated on the later E12.5 developmental stage. This also indicates that the average E10.5 TEPC population would be able to generate a full functional thymus if isolated and exposed to the right levels of *Foxn1*, consistent with previous findings (Gordon et al., 2004). The E12.5 developmental stage represents the latest TEPC age analysed. At this stage, TEPC are actively differentiating. Overall, **Figure 3.7** illustrates that data clustering faithfully follows the *Foxn1* expression levels design, showing a clear effect of *Foxn1* dosage on cell identity.

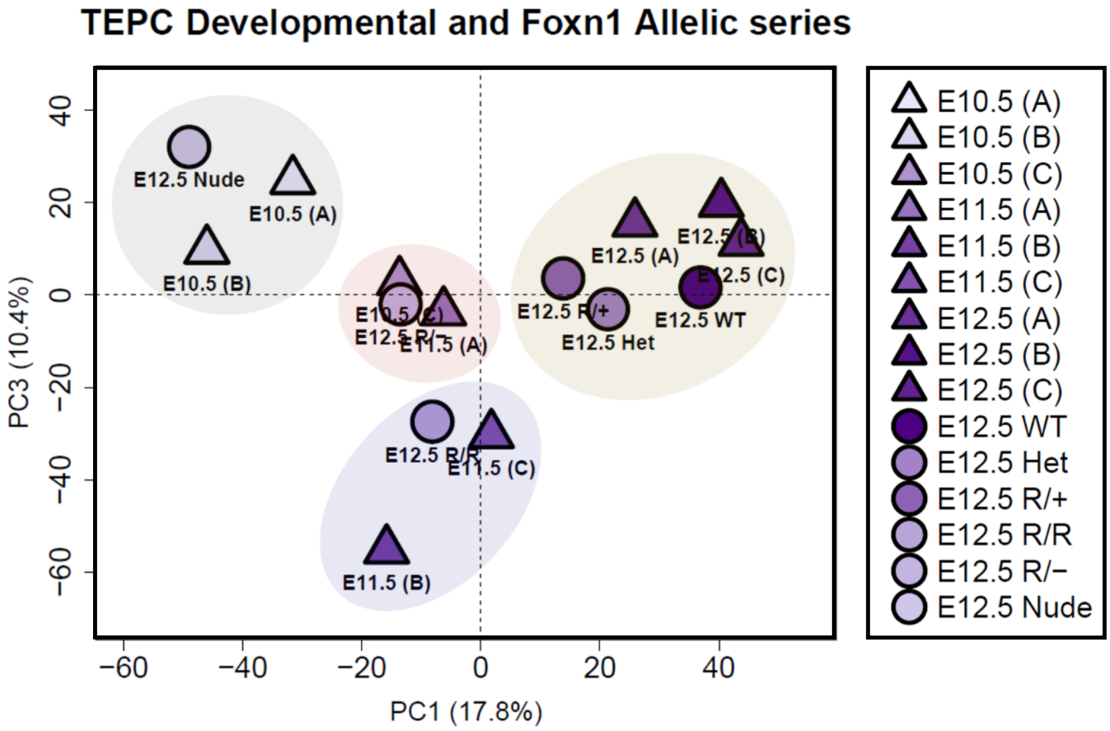


Figure 3.7: **Principal component analysis of the TEPC Developmental series and *Foxn1* Allelic series integrated dataset.** The PCA plot shows samples clustering from two different data series that have been integrated together. Samples from the TEPC Developmental series are represented with triangles, while samples from the *Foxn1* Allelic series are represented with circles. Colour gradient inside both triangles and circles illustrates *Foxn1* expression levels in each sample. PC1 (*x*-axis) accounts for 17.8% of samples variance, while PC3 (*y*-axis) describes 10.4% of samples variance. Both axes represent samples differences that, in the Allelic series dataset, is due primarily to the *Foxn1* level effect, while for the TEPC Developmental series is due to both the *Foxn1* levels and the developmental stage which is encompassed by that (shown also in **Figure 3.2**).

To further establish that batch effect correction of the integrated datasets did not bias our data and to also examine if groups' clustering (in **Figure 3.7**) is representative of genes expression profiles across samples (expressing similar *Foxn1* levels) and did not result because of samples' average clustering, gene expression patterns for TEPC and TEC lineage-specific markers (reviewed in Takahama et al., 2017) and for *Foxn1* target genes (Nowell et al., 2011; Žuklys et al., 2016) were observed across samples in each series and samples per series were clustered (see heatmaps in **Figure 3.8**) based on these markers. Even though these markers consist a very small percentage of the total number of expressed genes in TEPCs, their level of expression is able to define major subpopulations of the thymic epithelium and therefore they can be used when inspecting samples identity and expressional profiles.

Gene patterns across samples of varying *Foxn1* levels were majorly similar for the selected markers. Furthermore, the TEPC Developmental series dataset demonstrated a clear separation of samples E10.5 (A) and E10.5 (B) (similar to the PCA plot; **Figure 3.7**) from rest of the samples, with samples E10.5 (C) and E11.5 (A) grouped and slightly separated from the E11.5 (B) and E11.5 (C) ones, while the E12.5 triplicates formed their own group with most genes showing a clear expression pattern among them. The *Foxn1* Allelic series also indicated a clear separation of the Nude and R/- from rest of the samples, which show a linear trend of separation based on increasing *Foxn1* levels. Overall, heatmap clustering of the individual series supports the observed clustering in PCA (**Figure 3.7**) and suggests that gene patterns across samples are highly similar.

In summary, integration of the TEPC Developmental series with the *Foxn1* Allelic series demonstrated that variation in *Foxn1* expression level within an allelic series results in changes in gene expression that closely mimic those observed in the normal temporal development of TEPCs during fetal thymus development. This analysis further highlighted the importance of a single transcription factor, FOXN1, in defining cell identity in TEPCs. These datasets will consist the basic platform used to track differences between undifferentiated and differentiated TEPC genes, and FOXN1-dependent genes, in this thesis, and will also be used in combination with more datasets to answer further questions in respect to the TEPC state.

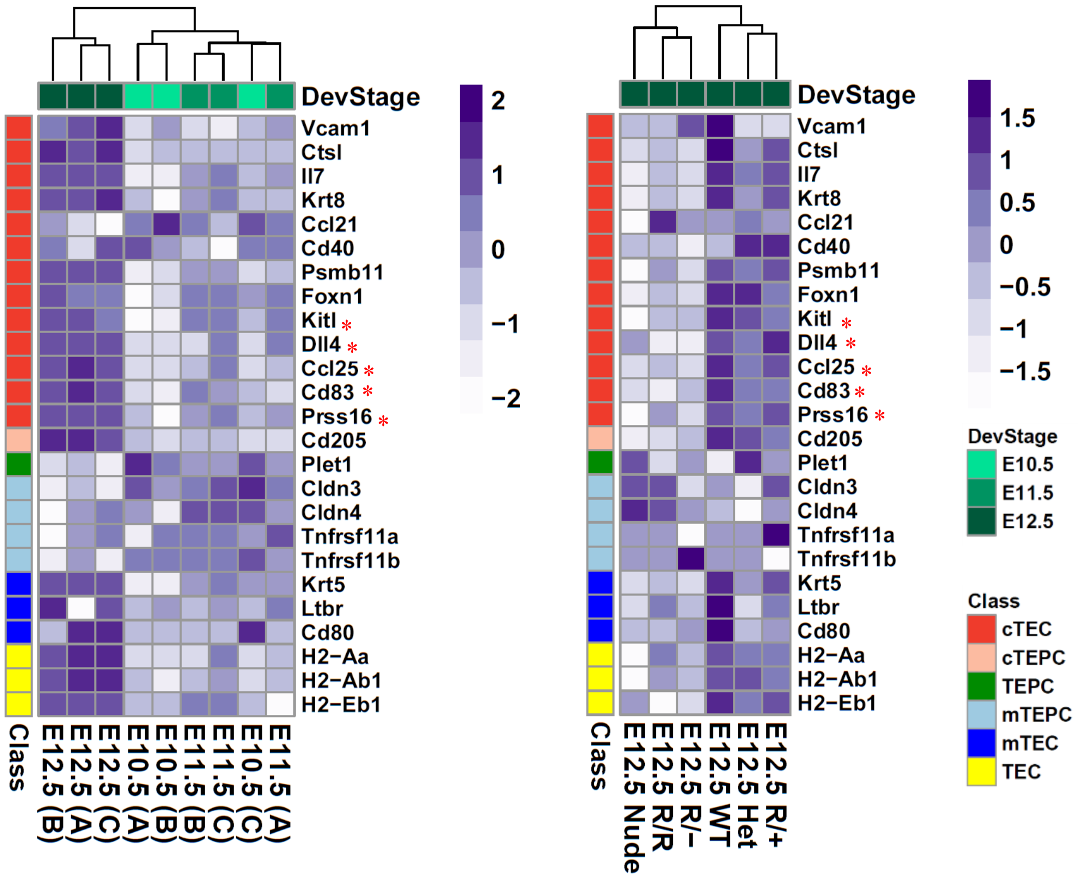


Figure 3.8: Lineage-specific gene heatmaps for the TEPC Developmental series and the *Foxn1* Allelic series datasets before data integration. Both heatmaps depict samples clustering per series, based on a selected list of markers (Takahama et al., 2017) which are representative of different TEPC and TEC sublineages present in the embryonic and postnatal thymus. Column annotations (DevStage) on top side of the heatmaps represent the developmental day on which samples have been collected (E10.5-E12.5), while row annotations (Class) describes the TEPC/TEC subpopulation that is defined according to high expression levels of the specific gene. FOXN1 target genes, including *Kitl*, (under the cTEC class) have been additionally annotated with an "*" symbol.

3.2.2 Projection of differentiating TEPC genes onto representative populations of the TEC sublineages indicates TEPC heterogeneity among the E12.5 TEPCs

To interrogate how the cTEC and mTEC sublineages are generated during TEC differentiation, genes that are differentially expressed between the E10.5 (undifferentiated progenitors) and E12.5 (differentiating TEPCs) timepoints from the TEPC Developmental series were projected on representative populations of the cTEC and mTEC sublineages (see next paragraphs for details). The data that I have used for this projection were embryonic E14.5 PLET1$^-$ and E14.5 PLET1$^+$ TECs (RNA-seq data from **biological triplicates per population** provided by **Dong Liu**, Blackburn lab; the majority of these cells will become cTECs and mTECs respectively) and 1 week old cTECs and mTECs [merged RNA-seq data from **similarly sorted individual and biological duplicates** obtained and reanalysed from GEO public repository: GEO accession codes **GSE44945 (St-Pierre et al., 2013)** and **GSE53110 (Sansom et al., 2014)**; for details see **Chapter 2**, **section 2.2.1.4**].

To identify groups of genes that change consistently between the E10.5 and E12.5 triplicates from the TEPC Developmental series, differential expression analysis was performed using the limma *voom* function (see **Chapter 2**, **section 2.3.3.2**) and I assigned statistical significance to the gene changes using the false discovery rate (FDR). The *Foxn1* expression levels across all samples are shown in **Figure 3.9**, suggesting higher variance among the E10.5 samples than the ones at E11.5 or E12.5 timepoints. An FDR threshold of ≤ 0.2 and an absolute fold change (|FC|) value ≥ 1.5 was chosen to define statistically significant genes. A total of 1,650 genes passed the FDR threshold and were considered for further analysis.

To explore how changes in the expression profile of early TEPC development is represented in later developmental points (such as E13.5 and older samples), when the cTEC and mTEC sublineages have emerged (or are starting to emerge), I further divided the 1,650 differentially expressed genes (E12.5 versus E10.5) in up and down-regulated gene groups (980 versus 670 genes respectively) and then plotted them according to the representative expression values they have acquired in the E14.5 and 1 week TEC subpopulations (**Figure 3.10**).

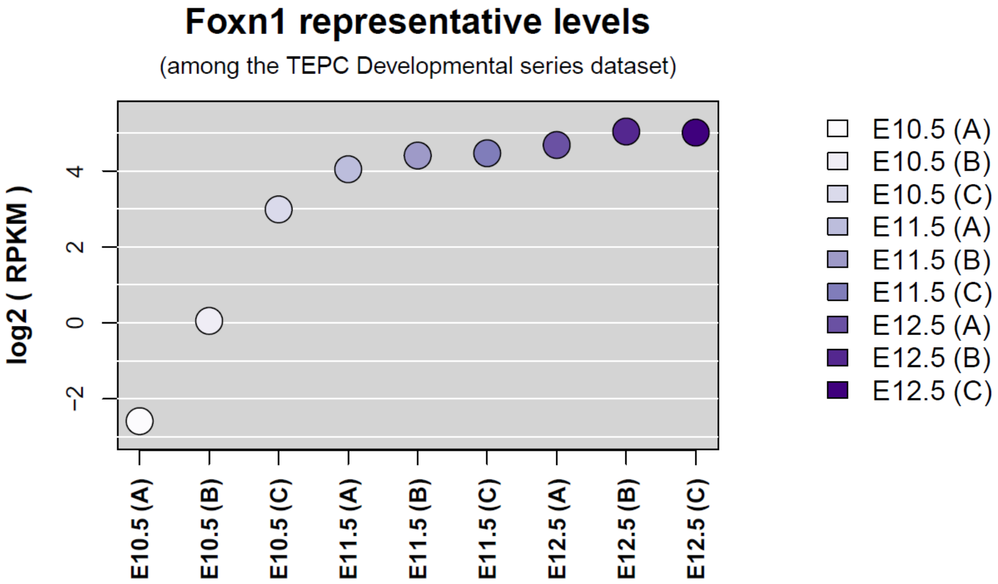


Figure 3.9: Representative levels of *Foxn1* in the TEPC Developmental series dataset. A scatterplot of *Foxn1* levels among the early TEPC developmental points (replicates per point distinguished by A-B-C lettering) demonstrates higher variability among the E10.5 triplicates, compared to the rest of the samples. The high variance may be explained by high biological variance at this very immature stage. E10.5 may comprise a high interchangeable TEPC state, where *Foxn1* levels may increase very rapidly after detection of its expression at ~E10.25 timepoint. Also, only small numbers of cells were used to generate the RNA-seq samples (data from Harsh Vaidya; Blackburn lab) which may contribute to this higher variability.

The scatterplots shown in **Figure 3.10** show a bias for up-regulated genes at E12.5 vs E10.5 to be more highly expressed in E14.5 PLET1$^-$ TECs than in E14.5 PLET1$^+$ TEC, a pattern that was even stronger when looking at the same genes in 1 week old cTECs versus 1 week old mTECs. The opposite pattern was apparent for the genes down-regulated in the E12.5 vs E10.5 comparison, which were more highly expressed in the E14.5 PLET1$^+$ and 1 week old mTECs than in the E14.5 PLET1$^-$ and 1 week old cTECs. Of note is that most of the up-regulated genes from the E10.5 to E12.5 developmental stage were also expressed in E14.5 PLET1$^-$/1 week old mTEC populations, albeit at lower levels than those in the E14.5 PLET1$^+$/1 week old cTEC populations. In contrast, a high portion of the down-regulated genes were fully absent from E14.5 PLET1$^-$/1 week old cTEC (**Figure 3.10**, bottom right panel – genes below 0 in *y*-axis).

To investigate if these differing patterns of expression were driven by FOXN1, samples from the *Foxn1* Allelic series were used to filter the 1,650 differentially expressed genes (E12.5 vs E10.5; TEPC Developmental series) for genes dependent on FOXN1. Lack of sample replicates in the *Foxn1* Allelic series dataset made it difficult to confidently compare samples to one another. However, based on the phenotypic and functional characteristics that the *Foxn1* Allelic series samples share, they could be categorised in three groups. Samples E12.5 WT, E12.5 Het and E12.5 R/+ (or Group 1) represent TEPC populations that express high levels of *Foxn1* and can give rise to a normal (or hyperplastic) fully functional thymus, sample E12.5 R/R (or Group 2) expresses intermediate levels of *Foxn1* and generates a suboptimally functional, hypoplastic thymus, while samples E12.5 Nude and E12.5 R/- (or Group 3) are representative of a TEPC arrested population that leads to athymia.

Pairwise comparison analysis was performed between the two phenotypically most distinct groups (Group 1: WT – Het – R/+ versus Group 3: Nude – R/-) to identify genes that are highly dependent on FOXN1. In this analysis, up-regulated genes from E10.5 to E12.5 would be considered FOXN1-dependent if two out of three samples in Group 1 expressed these genes more highly than both samples in Group 3, and vice versa in the case of down-regulated genes. The pairwise comparison identified a total of 850 FOXN1-dependent genes, which were used to assess the influence of FOXN1 on the expression patterns demonstrated in **Figure 3.10**.

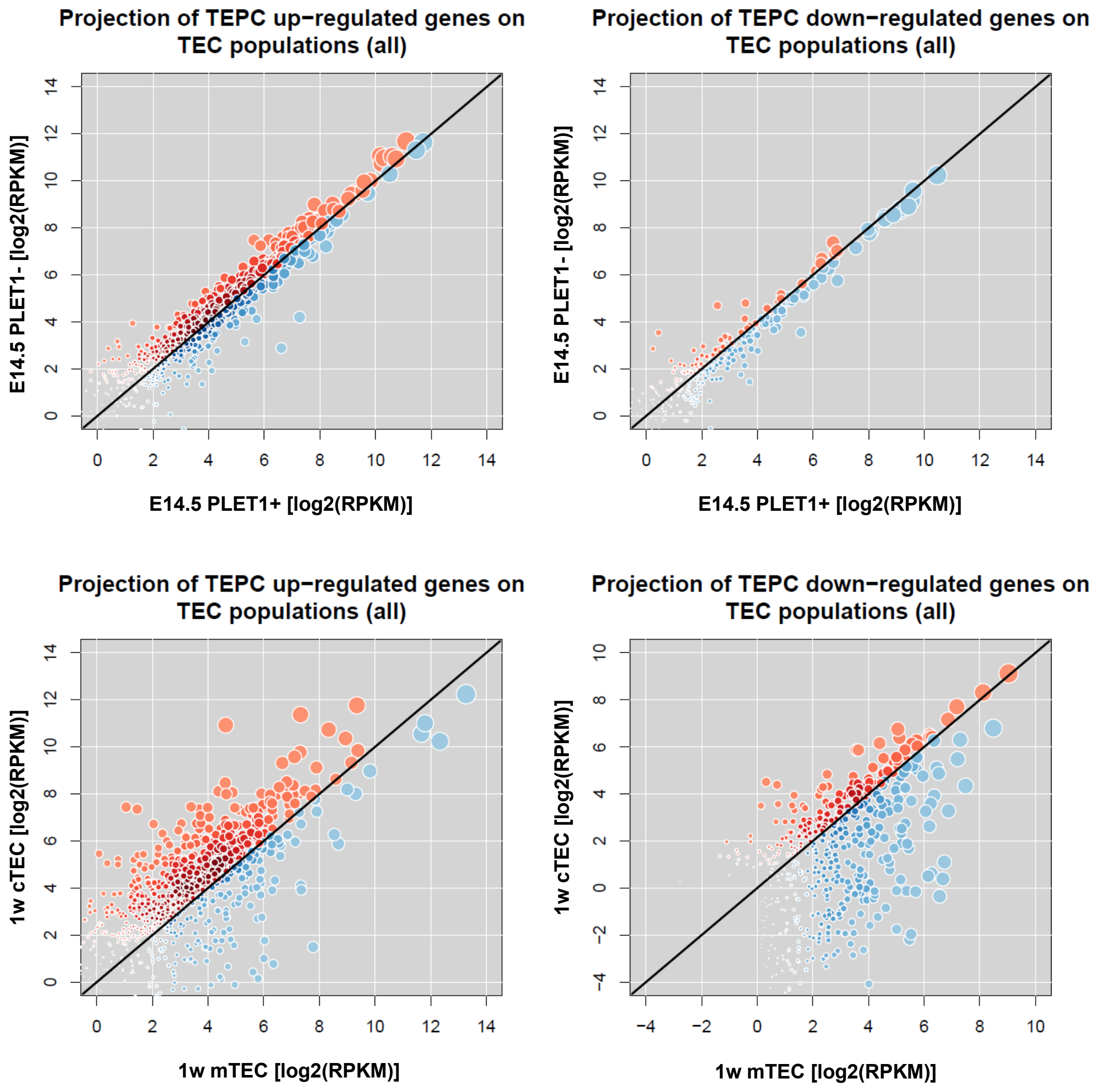


Figure 3.10: Projection of TEPC differentially expressed genes on E14.5 and 1w TEC subpopulations. The plot depicts a clear trend of the E12.5 vs E10.5 up-regulated genes to be more highly expressed in the representative cTEC compartment (E14.5 PLET1- and 1w cTECs; see circles coloured in red shades), while the down-regulated genes from the same comparison are more highly expressed in the representative mTEC compartment (E14.5 PLET1+ and 1w mTECs; see circles coloured in blue shades) with some of the genes completely missing from 1w cTECs (bottom right plot, blue circles below -1 in *y*-axis). Values in both axes are shown in log2-transformed RPKM format, size of the spot is scaled so that more highly expressed genes (either in *x* or *y*-axis) are represented with bigger circles, while colour scale (red or blue) represent spots density (darkers shades of red or blue – more overlapping spots) in graph.

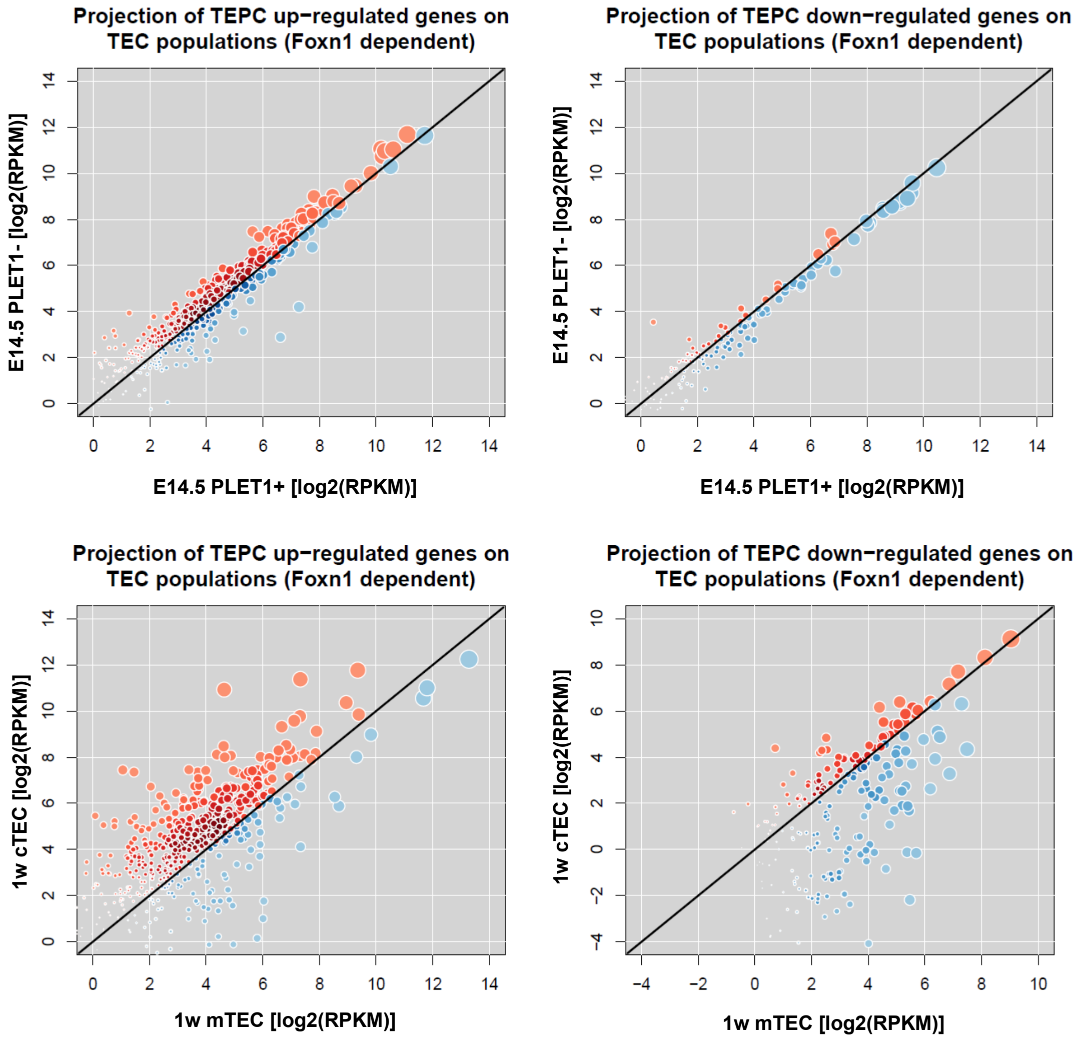


Figure 3.11: Projection of FOXN1-dependent, differentially expressed TEPC genes on E14.5 and 1w TEC subpopulations. The plot depicts a clear trend of the FOXN1-dependent up-regulated genes to be found higher expressed in the representative cTEC compartmnet (E14.5 PLET1- and 1w cTECs; see circles coloured in red shades), while the FOXN1-dependent down-regulated genes to be found higher expressed in the representative mTEC compartment (E14.5 PLET1+ and 1w mTECs; see circles coloured in blue shades) with some of the genes completely missing from 1w cTECs (bottom right plot, blue circles below -1 in *y*-axis). Values in both axes are shown in log2-transformed RPKM format, size of the spot is scaled so that more highly expressed genes (either in *x* or *y*-axis) are represented with bigger circles, while colour scale (red or blue) represent spots density (darkers shades of red or blue – more overlapping spots) in graph.

Figure 3.11 illustrates the expression profiles of the 850 FOXN1-dependent genes [586 up-regulated (60% of DE up-regulated genes) and 264 down-regulated (40% of DE down-regulated genes)] that are mapped to the later developmental E14.5 PLET1$^-$/1 week old cTEC and E14.5 PLET1$^+$/1 week old mTEC populations. The patterns observed in **Figure 3.10** for the up- and down-regulated genes are preserved in the FOXN1-dependent genes in **Figure 3.11** suggesting that this pattern reflects FOXN1 up-regulation. FOXN1-dependent genes whose expression increases with FOXN1 up-regulation (from E10.5 to E12.5) show higher levels of expression in representative populations of the cTEC than the mTEC lineage. This difference is in agreement with the higher level of *Foxn1* expression in cTECs versus mTECs (Rode et al., 2015; O'Neill et al., 2016) and it suggest the role of FOXN1 as a potential activator (also denoted by Žuklys et al., 2016). Most of the FOXN1-dependent up-regulated genes remain expressed still in the mTEC compartment, although not as highly as in cTECs.

The opposite trend is observed for genes that down-regulate from E10.5 towards E12.5. These genes down-regulate when cells progress from an early developmental point to a latter one when *Foxn1* expression increases and they are also found to be more highly expressed in the mTEC compartment (which has lower *Foxn1* levels compared to the cTEC compartment), with some of the genes to be completely absent from the cTEC compartment (genes below -1 in *y*-axis; **Figure 3.11**). All the above suggest a repressive role for FOXN1 in regulation of these genes. This repression could happen either by direct binding of FOXN1 on gene promoters, or through indirect control of other regulatory factors that can then bind and affect downstream genes. In both cases, these down-regulated genes may be expressed in a heterogeneous population of differentiating TECs (at E12.5), in which cells average profile shows an expressional decrease of these genes, however, this average expression profile (which at E12.5 better resembles a cTEC early phenotype) could be masking a few cells that would still be expressing these genes higher (less differentiated/expressing lower levels of *Foxn1*). These genes may never reach as high levels of *Foxn1* and could form the mTEC compartment by maintaining expression of these genes. To examine the above scenario and to better envision how the differentiation of TEPCs occurs, gene expression values of the 264 down-regulated genes at E10.5 and E12.5 were compared to the expression values of these genes in 1week old cTEC and mTEC (**Figure 3.12**).

Gene expression values at E12.5 for particular genes in the down-regulated gene set were almost identical with those in 1 week old cTECs (**Figure 3.12**, top right panel, genes between 150 and 264 on *x*-axis), while the same genes were expressed much more highly in 1 week old mTECs (**Figure 3.12**, bottom right panel, genes between 150 and 264 on *x*-axis). On the other hand, the same group of genes at E10.5 TEPCs demonstrate similar expression levels to the same group of genes in 1w mTECs (**Figure 3.12**, bottom left panel, genes between 150 and 264 on *x*-axis). The high similarity of the E10.5 TEPC gene expression profiles with the profiles in 1w mTECs could suggest a relation between these two populations. However, if only a few cells maintain an E10.5 TEPC expressional profile (potentially less differentiated) when reaching the E12.5 developmental stage, gene expression values of these cells would be masked by the average E12.5 population gene profile which would represent on average a more cTEC-like phenotype.

From this, I conclude that FOXN1 rapidly represses, either directly or indirectly, a number of genes on its induction in TEPCs. These genes remain repressed in cTECs, and the repression is complete by E12.5. However, since I observed a trend for these down-regulated genes to be expressed in mTECs, the possibility that a rare subpopulation of TEPCs at E12.5 continues to express these genes, and that these cells then generate the mTEC lineage, cannot be excluded. This could be tested by determining heterogeneity of FOXN1 expression at E12.5, or by analysis of single cell sequencing data.

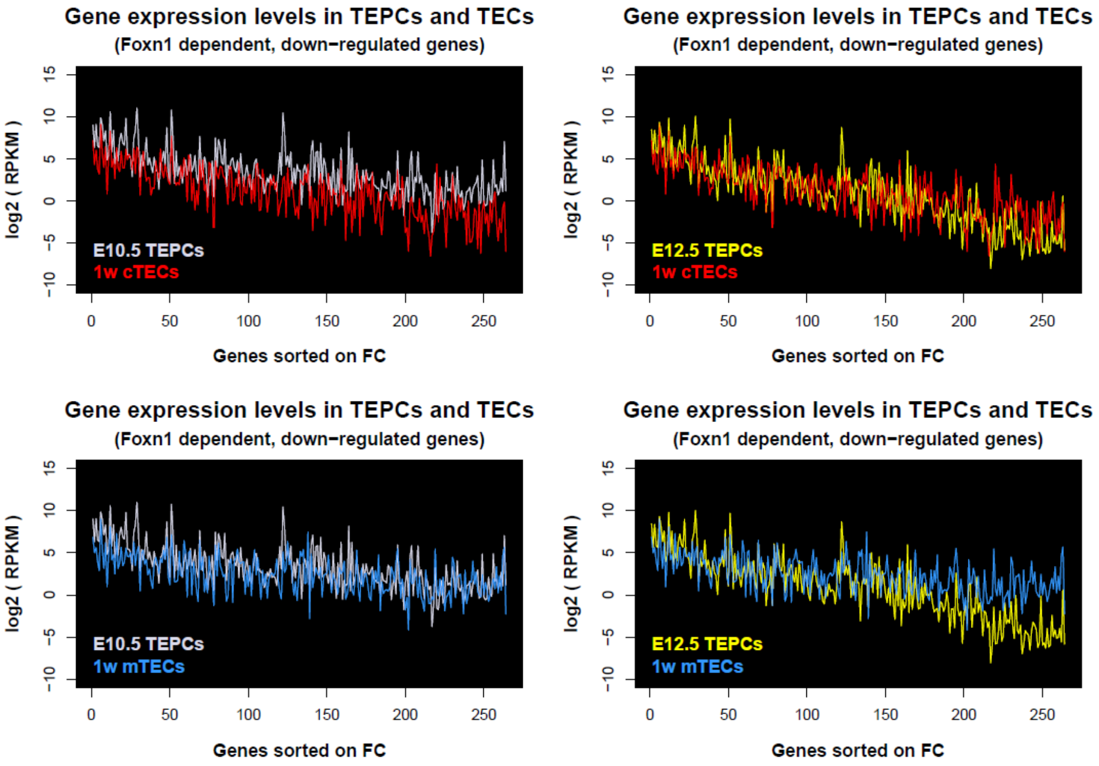


Figure 3.12: Expression levels of the FOXN1-dependent, down-regulated genes (E12.5 versus E10.5) in E10.5 TEPCs, E12.5 TEPCs, 1w old cTECs and 1w old mTECs. Expression levels of the FOXN1-dependent E12.5 versus E10.5 down-regulated genes are shown in pairwise comparisons of E10.5 TEPCs-1w cTECs (top left panel), E12.5 TEPCs-1w cTECs (top right panel), E10.5 TEPCs-1w mTECs (bottom left panel) and E12.5 TEPCs-1w mTECs (bottom right panel) to compare the expressional profiles of these genes between these populations. Genes (represented in the *x*-axis) are sorted by increasing fold change (in absolute value), with stronger changes to be shown towards the end of the *x*-axis (>150th gene), while *y*-axis shows expression values (RPKM) of genes in a log2-transformed scale.

3.2.3 *Foxn1* heterogeneity among E12.5 bipotent TEPCs

To address the issue of *Foxn1* heterogeneity among the E12.5 TEPCs (see **section 3.2.2**), *Foxn1* levels should be determined on a per cell resolution. *Foxn1* expression has already been established in TECs during prenatal development (Rode et al., 2015; O'Neill et al., 2016). In particular, analysis of thymi from a *Foxn1*-eGFP reporter mouse line has demonstrated a graded expression of *Foxn1* between cTECs (marked by the CD205 cTEC-specific marker) and mTECs (marked by the UEA1 mTEC-specific marker) at E17.5 in mouse development (when mTEC numbers become appreciable) with the mTEC population to be expressing lower levels of *Foxn1* (O'Neill et al., 2016). In earlier timepoints, there is no distinct separation among the cTEC and mTEC lineages but only a separation between the MHCII^{hi} and MHCII^{lo} cell populations (E13.5 TECs, see O'Neill et al., 2016), with the MHCII^{lo} TECs to be expressing lower levels of *Foxn1*. The above published data come in support of a possible *Foxn1* heterogeneity in E12.5 TEPCs, while they more specifically suggest the existence of a two-level graded *Foxn1* expression pattern which would agree with the presence of a rare TEC progenitor among rest of the differentiating TECs at E12.5.

If the E12.5 TEPCs are dynamically regulated based on *Foxn1* levels, it would also be interesting to examine if cells of the same stage with higher *Foxn1* level would demonstrate lower levels of the PLET1 marker whose expression has been linked to progenitor capacity of fetal TEPCs (Depreter et al., 2008; Ulyanchenko et al., 2016). Approximately 99% of E12.5 TEPCs are PLET1 positive (example FACS plot from Liu et al., 2017, submitted). PLET1 was significantly down-regulated between the E10.5 and E12.5 timepoints, with its down-regulation being FOXN1-dependent (see **section 3.2.2**). To determine how strongly PLET1 down-regulation depended on *Foxn1* expression level, I observed the expression profile of *Plet1* in relation to *Foxn1* levels in 31 thymic epithelia samples, comprising the TEPC samples from the TEPC Developmental series and *Foxn1* Allelic series integrated dataset, the 1 week old cTEC and mTEC samples and the 4 week old mTEC subpopulations (St-Pierre et al., 2013; Sansom et al., 2014) consisting of immature, Aire knock-out, Aire negative, Aire positive and mature mTECs (see **Chapter 2**, **section 2.2.1.4** for sample details).

To estimate the level of correlation between *Plet1* and *Foxn1* expression, I calculated gene-to-gene correlation across the aforementioned 31 datasets (replicates treated separately) using the Spearman method (non-parametric data) in combination with a statistics test to evaluate the confidence of the correlation. *Foxn1* expression was significantly anti-correlated to *Plet1* (rho $\simeq$ -0.85, *p*-value $\simeq$ 5 x 10^{-7}), as shown in **Figure 3.13**, with strong dependence of *Plet1* down-regulation on *Foxn1*. Moreover, association of the *Plet1*/*Foxn1* ratio with the differentiation status of the sample suggested a more immature/anti-differentiated phenotype in samples expressing more *Plet1* and less *Foxn1* (TEPCs and immature mTECs, **Figure 3.13**), and a more differentiated phenotype in the opposite case (cTECs/mature mTECs).

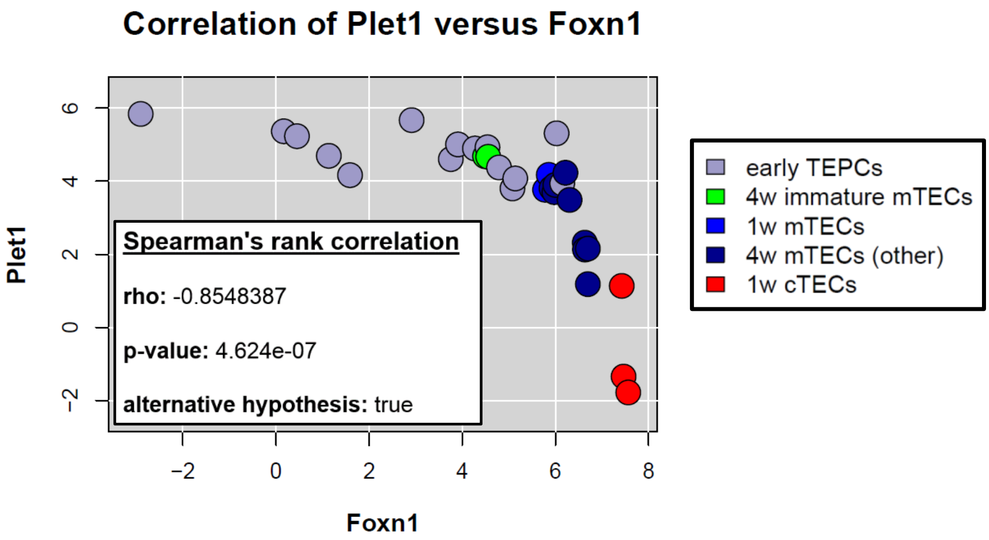


Figure 3.13: Scatterplot of *Plet1* versus *Foxn1* among various TEC samples. Correlation analysis of *Plet1* and *Foxn1* among the TEPC Developmental series and *Foxn1* Allelic series integrated dataset, E14.5 PLET1$^-$ and PLET1$^+$ TEC populations (Dong Liu, submitted), 1 week old cTECs and mTECs and 4 week old immature, *Aire* knock-out, AIRE negative, AIRE positive and mature mTECs. Level and statistical significance of the correlation between the two factors are shown under Spearman's rank correlation, while sample labels are provided in the box on the right-hand side of the plot. The null hypothesis (H_0) of the test would consider that samples are not correlated. H_0 was rejected (*p*-value ≤ 0.05) and the alternative hypothesis (*Foxn1* and *Plet1* are highly anti-correlated) was accepted.

3.2.4 A well-defined TEP/SC gene signature from scRNA-seq data

To investigate in depth the heterogeneity of *Foxn1* expression in E12.5 TEPCs, and to determine the existence of a rare subpopulation of cells that expresses lower levels of *Foxn1* and possibly mTEC-associated genes among the E12.5 TEPCs, single cell (sc) RNA-seq data from the E12.5 TEPC population needs to be generated. Identification of a rare *Foxn1*^{low} subpopulation of cells from the E12.5 TEPC scRNA-seq data would likely allow better characterisation of this population by a unique gene signature which could then be used to potentially identify the same population or its derivatives in other stages during thymus development. Characterising the expressional profile of this population would also clarify if these cells are bipotent TEPCs or they better resemble mTEPCs not previously identified before. Howbeit, if this population of (m)TEPCs is very rare, a larger number of E12.5 progenitors would be required to be sequenced to capture it.

3.2.5 Alternative markers to use in combination with PLET1 and EpCAM to identify a homogeneous population of bipotent TEP/SCs

PLET1 is able to identify and purify all TEPCs at E12.5 in prenatal mouse development (Bennett et al., 2002; Gill et al., 2002). PLET1 has also been shown to mark medulla epithelia stem/progenitor cells in early thymus ontogeny that further generate medullary islets (Rodewald et al., 2001) and a rare cTEC specific population in the adult thymus (Ulyanchenko et al., 2016) that is able to generate both thymic epithelial lineages; it has, therefore, been associated with progenitor potential. However, after E12.5, appearance of PLET1$^-$ TECs demonstrated an equal capacity to generate a fully functional thymus. This potency of both PLET1$^+$ and PLET1$^-$ populations continues until E18.5 when the PLET1$^+$ subpopulation becomes a minority (Swann and Boehm, 2007; Rossi et al., 2007b). Thus, PLET1 alone becomes an insufficient marker to isolate TEPCs that are bipotent.

In the absence of scRNA-seq data, it would be preferable to use an earlier timepoint to identify markers to use in combination with (or instead of) PLET1 for isolating bipotent TEPCs or TESCs. As discussed before, from the *Foxn1* pseudo-timing analysis (**section 3.2.1**), and the *Foxn1-Plet1* correlation analysis (**section 3.2.3**), E10.5 samples were shown to resemble on average a more undifferentiated TEP/SC population. Therefore, representative markers of that population would be more likely to identify a premature state of the TEP/SCs.

To generate a list of alternative markers to be used in combination with PLET1 for bipotent TEP/SC isolation, the integrated dataset of the TEPC Developmental series and *Foxn1* Allelic series was scanned and genes were considered as candidate markers only if they met certain criteria. The expression level of candidate TEP/SC markers should diminish from E10.5 to E12.5; thus, a more stringent FDR threshold (FDR <= 0.05) was used to identify high confidence genes in the candidate down-regulated gene list. To bias the identification of candidate TEP/SC markers that could be used for cell sorting, genes that were expressed at higher levels than *Plet1* at the early E10.5 undifferentiated progenitor stage (based on their average expression values across the E10.5 replicates from the TEPC Developmental series) and that could also bind to cell's surface (transmembrane proteins) were selected. The resulting list of the candidate markers to be further characterised is provided in **Figure 3.14**.

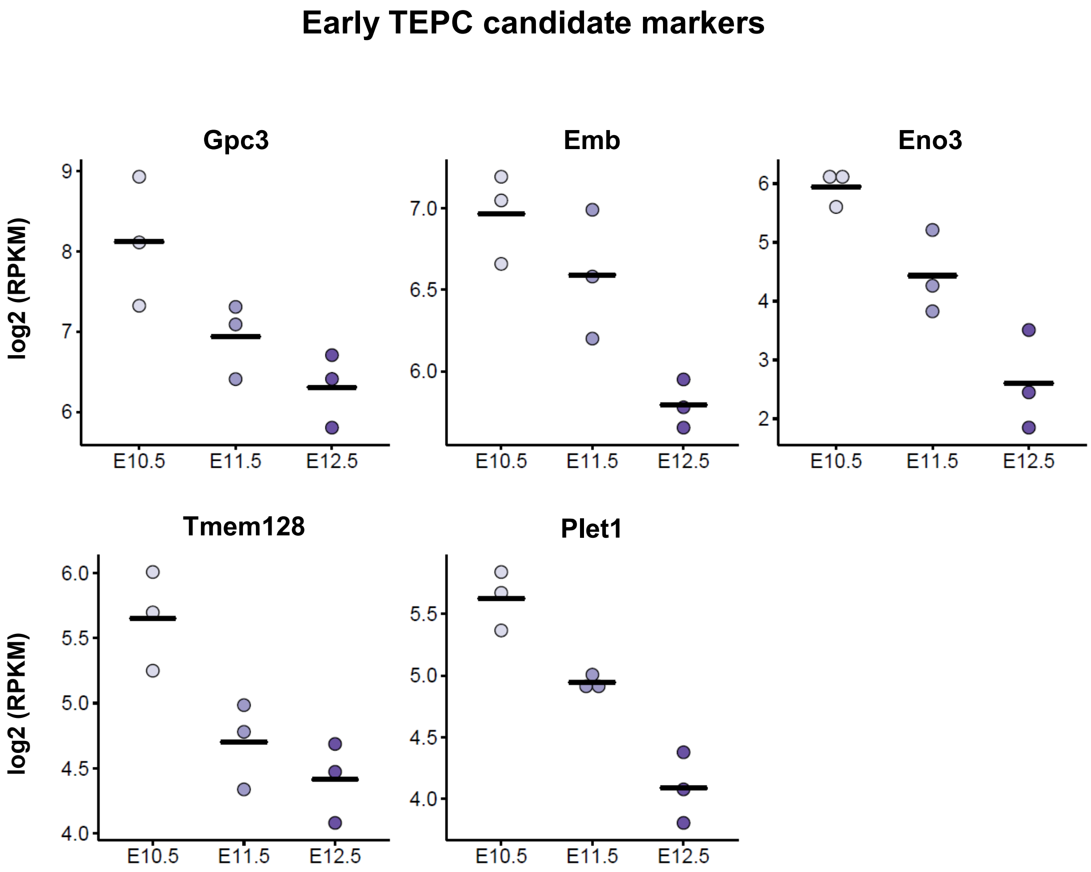


Figure 3.14: Candidate marker list to be used in combination with PLET1 for TEP/SC isolation. The screening process for the candidate markers selected genes that were more highly expressed at E10.5 compared to *Plet1*, their expression decreased with *Foxn1* increase and they could bind to the cell surface. A list of these TEP/SC markers with their mean expression levels noted by a horizontal line (across replicates; log2 (RPKM) values) is shown in the plots depicted here.

3.3 DISCUSSION

Multiple lines of evidence have demonstrated the importance of *Foxn1* in respect to the thymic system, with *Foxn1* to be indispensable for TEPC differentiation in each derived sublineage (Chen et al., 2009; Cheng et al., 2010; Corbeaux et al., 2010; Nowell et al., 2011; Su et al., 2003). In one of these studies (Nowell et al., 2011), phenotypic analysis of thymi from an allelic series of mice expressing different levels of *Foxn1* suggested that available levels of *Foxn1* can determine both the size and the functionality of the thymi produced (either directly or indirectly through lack of regulatory signals between differentiating TECs and maturing T-cells). In this chapter, I have made a list of observations based on bioinformatics analysis of RNA-seq series from early developmental stages of fetal TEPCs and newborn TECs to predict a potential TEPC differentiation model during thymus ontogeny.

Observation 1: The TEPC samples from the *Foxn1* Allelic series, sorted for *Foxn1* increasing levels, correspond to wild type E10.5, E11.5 and E12.5 TEPCs from the TEPC Developmental series (see **Figure 3.7**), and thus the developmental steps that can be achieved in each of the allelic variants resemble a natural progression step of the fetal thymic epithelium. Therefore, imposition of different maximum *Foxn1* levels effectively corresponds to 'pseudo-timing' of early TEPC differentiation, with progression at each stage representing a natural progression step of TEPCs. {this paragraph addresses **Question 1** in **section 3.1**}

Observation 2: The average transcriptional profiles of the E10.5 TEPC samples (see groupings in **Figure 3.7**) and gene expression patterns of important TEPC and TEC lineage-specific markers (depicted in **Figure 3.8**) have indicated that the E10.5 TEPCs better resemble the E12.5 R/- and E12.5 Nude samples from the TEPC Developmental series, in which *Foxn1* expression is majorly blocked. The latter ones have been shown to be able to survive *in vivo* indefinitely and produce all thymic lineages upon *Foxn1* reactivation (Jin et al., 2014). Additionally, these early E10.5 TEPCs are characterised by the highest levels of *Plet1* (early thymic progenitor marker) compared to the rest of the TEPC developmental points, as shown by the differential expression analysis between the E10.5 and E12.5 sample triplicates, in which *Plet1* was found to be down-regulated with statistical significance in the E12.5 TEPCs (see **section 3.2.1**). *Plet1* expression was also found strongly anti-correlated (rho = -0.85, *p*-value ≤ 0.05) to that of *Foxn1* (thymic pro-differentiation marker) in a correlation analysis performed among 31 TEC datasets (see **section 3.2.3**). Together, these findings suggest that E10.5 TEPCs are more likely to consist a homogeneous TESC population (or bipotent TEPC population) that can give rise to all cTEC and mTEC sublineages and under the right conditions they could be maintained and expanded *in vitro*. {this paragraph addresses **Question 2** in **section 3.1**}

Observation 3: FOXN1-dependent genes that were differentially expressed between the E12.5 and E10.5 timepoints showed divergent patterns of expression among the cTEC and mTEC populations with up-regulated genes at E12.5 (vs E10.5) to exhibit higher expression in cTECs (vs mTECs) and down-regulated genes at E12.5 (vs E10.5) to exhibit lower expression in cTECs (vs mTECs) (see **section 3.2.2, Figure 3.11** and **Figure 3.12**). These patterns mainly represent the dynamic role of FOXN1 which appears to be acting both as an activator and a repressor for different groups of genes, resulting in divergent expression patterns maintained in the postnatal cTECs and mTECs (TEC populations that express *Foxn1* in different levels). Therefore, cTEC and mTEC fated cells could already be present among the E12.5 TEPCs, when *Foxn1* expression has been well established. {this and the next paragraph address **Question 3** in **section 3.1**}

A repressive role for FOXN1 has not been demonstrated before, thus here I discuss potential mechanisms by which FOXN1 could achieve gene repression in TEPCs. In order to repress transcription of a particular gene, FOXN1 would have to bind to specific DNA sites and cause transcriptional inhibition. Repression could be imposed by FOXN1 blocking or antagonising (lack of a stimulation domain or active inhibition through protein-protein interactions) for the binding site of a regulatory factor (transcription factor or member from the transcription initiation complex), or it could repress a certain group of genes by interacting with corepressors whose role is to modify the chromatin structure (reviewed in Gaston and Jayaraman, 2003). Nonetheless, FOXN1 ChIP-seq data will be vital to confidently demonstrate whether or not gene regulation is a direct effect of this transcription factor (see analysis in **Chapter 5**).

Observation 4: Even though PLET1 has been linked to progenitor activity of TECs before (Bennett et al., 2002; Gill et al., 2002; Rodewald et al., 2001; Ulyanchenko et al., 2016), using solely PLET1 as a TEC early marker does not allow pure isolation of bipotent progenitors from the thymus in different embryonic stages of mouse development. Therefore, in **Chapter 3**, I have used the E10.5 TEPCs (which better represent bipotent progenitors [see **Observation 2**]), in a comparative analysis versus the more progressed E12.5 TEPCs to identify a list of gene candidates which we could use in combination with PLET1 for better isolation of the early bipotent TEC population. Subsequent filtering criteria used for the markers selection include equal or higher expression of these marker candidates at E10.5 compared to PLET1 expression at the same stage and ability of these markers to bind the cell surface for more efficient cell isolation. {this paragraph addresses **Question 4**, **section 3.1**}

3.3.1 A potential TEP/SC differentiation model

The *in-silico* observations from **Chapter 3** inform a TEPC differentiation model {this section addresses **Question 5** in **section 3.1**} in which E10.5 TEPCs represent (on average) an undifferentiated bipotent TEPC population (or TESCs) characterised by the highest *Plet1* levels (in agreement with the progenitor capacity of these cells) [**Observation 2**], while from the onset of *Foxn1* expression, E11.5 and E12.5 TEPCs comprise increasing proportions of differentiating TECs (lineage fated) whose progression is dependent on the *Foxn1* level [**Observation 1**], with the possibility of distinctive expression profiles of cells to already exist among the E12.5 TEPCs that are then maintained in the differentiated cTEC and mTEC lineages [**Observation 3**].

Previous analysis of the *Foxn1* levels on a per cell resolution in prenatal mouse development (O'Neill et al., 2016) has already proposed the existence of two populations with differing *Foxn1* levels as early as E13.5 in mouse development that supports the prediction of a heterogeneous (or bimodal) expression pattern for *Foxn1* among the E12.5 TEPCs. The subpopulation of cells that expresses *Foxn1* in a lesser degree could comprise: a) less differentiated progenitor cells (more similar to the ones at the E10.5 or the E11.5 developmental stage) that will eventually increase their *Foxn1* level and choose between a cTEC or an mTEC fate or b) mTEC fated progenitors (mTEPCs) that will maintain moderate levels of *Foxn1* and progressively become mTECs.

E12.5 TEPC heterogeneity comes in opposition to the clonal analysis of TEPCs by Rossi et al. (2006) that demonstrated a capacity for the E12 TEPCs to differentiate into both cortical and medullary TECs. However, unpublished data from **Alison M. Farley** (Blackburn lab, 2006) that repeated the clonal analysis for the E11.5 and E12.5 TEPC populations failed to reproduce the above results (injected cells would acquire only a cTEC phenotype). As noted from correspondence with the authors, the only difference between the two protocols was due to cells laying in an overnight culture before injected in the aged matched developing thymus. Apparently, keeping TEPCs in culture, reduces average levels of *Foxn1* in the population. It is then possible that the majority of the TEPCs in culture return to a more immature state and become again TEP/SCs. If that is the case, Rossi et al. has really demonstrated that TEPCs at E12.5 still exist in a transient state and are still capable of returning to a more immature state through adjustment of the *Foxn1* levels, while Farley demonstrated that the majority of E12.5 TEPCs lack bipotency and are most probably cTEC fated progenitors.

Collectively, the bioinformatics observations in conjunction with experimental evidence (denoted when necessary in the paragraphs above) can inform of a TEPC progression model, in which the E10.5 TEPCs comprise undifferentiated bipotent progenitors, while the TEPCs in later developmental stages (E11.5 and E12.5) are characterised by cell heterogeneity, which takes place after increase in *Foxn1* expression. Three different versions of heterogeneity can be assumed based on the available data: **Version 1** describes the E11.5 and E12.5 TEPCs to consist of (more immature) bipotent TEPCs and cTEC fated progenitors (cTEPCs), **version 2** envisions that the E11.5 and E12.5 TEPCs constitute mTEC fated progenitors (mTEPCs) and cTEPCs, while **version 3** proposes the possibility of all three populations (TEPCs, mTEPCs and cTEPCs) to be apparent within the E11.5 and E12.5 TEPCs in different proportions. The predicted TEPC progression model with the alternative heterogeneity versions of the E11.5 and E12.5 TEPC populations are illustrated unitedly in **Figure 3.15**. This model improves our current knowledge of fetal thymus development.

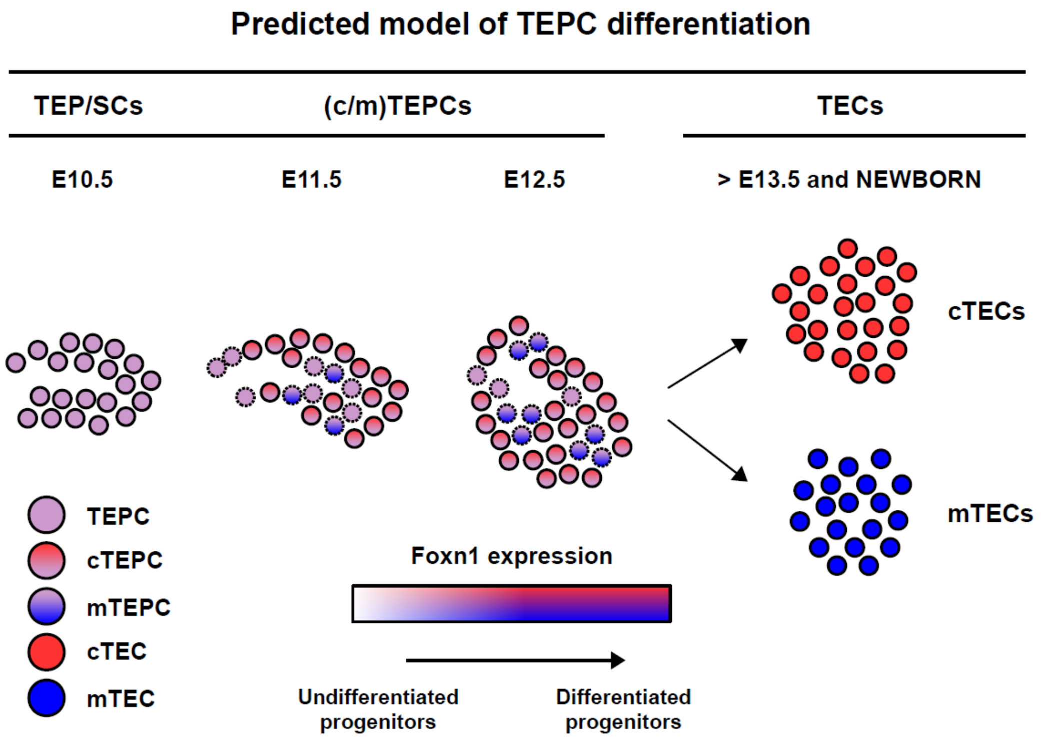

Predicted model of TEPC differentiation
TEP/SCs
(c/m)TEPCs
TECs
E10.5
E11.5
E12.5
> E13.5 and NEWBORN
cTECs
mTECs
TEPC
cTEPC
mTEPC
cTEC
mTEC
Foxn1 expression
Undifferentiated progenitors
Differentiated progenitors

Figure 3.15: Schematic representation of a potential TEPC differentiation model towards the TEC sublineages. In this model, E10.5 TEPCs (light magenta colour) represent (on average) an undifferentiated TEPC population (or a TESC population, see **section 3.2.1**), while from the onset of *Foxn1* expression, cells progressively comprise increasing populations of TEPC-fated cells (red/magenta and magenta/blue shades, see discussion in **section 3.3.1**); these TEPC-fated cells shall give rise to the cTEC (red colour) and mTEC (blue colour) main lineages. The predicted TEPC heterogeneity, which is depicted in the E11.5 and E12.5 developmental stages of this model, could be resolved by alternative versions of existing populations: In **version 1**, only bipotent TEPCs and cTEC-fated progenitors (cTEPCs) are present in the E11.5 and E12.5 TEPCs. In **version 2**, only mTEC-fated progenitors (mTEPCs) and cTEPCs exist within the above populations, and in **version 3**, TEPCs, cTEPCs and mTEPCs altogether form the above populations. The TEPC and the mTEPC populations are illustrated with a dashed line at the developmental timepoints E11.5 and E12.5, to represent the alternative versions of heterogeneity within the proposed model, because in version 1 and in version 2, mTEPCs and TEPCs respectively are absent from the E11.5 and E12.5 timepoints. Notably, the E11.5 and E12.5 cTEPCs or mTEPCs are able to return to a less immature or bipotent state upon *Foxn1* reduction (Farley 2006; Rossi et al. 2006, discussed in the main text of this section), as shown by representative arrows.

The model in **Figure 3.15** agrees with the serial progression model of TEC differentiation reviewed in Takahama et al. (2017) that describes together the current understanding around "TEPCs towards TECs" differentiation based on existing literature. Briefly, starting from a bipotent progenitor stage, TEPCs have to go through a cTEC-like TEPC stage [expressing *Ly75* (*Cd205*), *β5t* (*Psmb11*) and *Il7*] from which a Claudin^{hi} SSEA1^{+} mTEC stem cell population derives through an unknown mechanism and further differentiates into the mTEC lineage, while cells not expressing these latter markers will further differentiate into cTECs (see **Chapter 1**, **Figure 1.2**). Our model, in comparison to the one reviewed by Takahama et al., additionally suggests the existence of a (m)TEPC population among the cTEC-like TEPCs (E11.5/E12.5 TEPCs, see **Figure 3.15**) that would eventually consist some mTEC population, or act like a bipotent progenitor. If existent, this population could consist part of the unknown mechanism that gives rise to the Claudin^{hi} SSEA1^{+} mTEC stem cells, since E10.5-E12.5 TEPCs do not express the SSEA1 marker just yet. Lastly, since all mTECs seem to have experienced expression of *Ly75* (CD205), *β5t* (PSMB11) and *Il7* in early ontogeny (Ohigashi et al., 2013), I would expect this population to also express (momentarily or continuously in a low level) these genes among other genes that are found highly expressed in E10.5 TEPCs and newborn mTECs.

To examine the prospect of a heterogeneous expression of *Foxn1* in E12.5 TEPCs and the existence of a rare TEPC and/or mTEPC subpopulation, scRNA-seq data for the E12.5 TEPCs will be required. scRNA-seq data will verify current predictions/ findings presented here. If enough cells express *Foxn1* at a low level, the scRNA-seq data will also be able to generate a gene signature to allow isolation of this particular population of cells (even from later developmental stages) for expansion and functional characterisation. Because these data are not yet available, uniquely expressed genes in E10.5 TEPCs that are expressed higher than *Plet1* and are also able to bind the cell surface could be used for the isolation of a TEP/SC population [see **Observation 4**].

3.4 SUMMARY

RNA-seq data integration in conjunction with comparative bioinformatics analysis was used in **Chapter 3** to investigate the early state of TEPCs in conjunction to *Foxn1* expression and to also develop a potential model of TEPCs differentiation towards the cTEC and mTEC sublineages. In **Chapter 3**, I have shown that precise expression levels of *Foxn1* play a vital role in regulating TEPCs identity, with different maximum *Foxn1* levels to effectively correspond to 'pseudo-timing' of TEPCs natural progression in the fetal thymic epithelium. Furthermore, it is the first time that a repressive role for FOXN1 has been indicated, with additional analysis to be required to confirm whether the repressive effect is due to direct binding of FOXN1 on DNA regulatory regions or due to indirect regulation of other transcription factors. In **Chapter 3**, I also propose a new *in silico* TEPC differentiation model which encompasses the existence of a common bipotent progenitor (at E10.5) and appearance of lineage-specific progenitors (from E11.5 onwards). Therefore, based on these findings and a clonal analysis of the E12.5 TEPCs by Farley (Blackburn lab), the current bipotent E12.5 TEPCs suggested by Rossi et al. (2005) is revisited. Our data, in combination with the Rossi et al. paper, suggest that E12.5 cTEPCs and mTEPCs could exist in a transient state where lineage restriction can be amended through *Foxn1* expression reduction. Markers to use in combination with (or without) PLET1 for the isolation of TEP/SCs have also been provided in this chapter.

Chapter 4

An *in silico* model of the signalling requirements for the survival and expansion of bipotent fetal TEPCs

Chapter 4

An *in silico* model of the signalling requirements for the survival and expansion of bipotent fetal TEPCs

4.1 PRÉCIS

The thymic primordium is believed to contain bipotent thymic epithelial progenitor cells (TEPCs) from the time of its formation at least until the onset of expression of FOXN1, the transcription factor (TF) whose continuous expression is necessary for development and maintenance of a functional thymus. These undifferentiated TEPCs present in the early fetal thymus can reconstitute all known thymic epithelial cell lineages and direct the formation of a complete, fully functional thymus (Gordon et al., 2004). Differentiation arrest of TEPCs occurs in *nude* mice, which lack functional *Foxn1* due to a null mutation (Nehls et al., 1996). Recently, it has also been shown that when FOXN1 expression is blocked, at least some TEPCs are retained in a bipotent state, and these remain present *in vivo* apparently indefinitely (Jin et al., 2014). This indicates that the cell state upstream of FOXN1 initiation is inherently stable, and suggests that these cells effectively function as thymic epithelial stem cells (TESCs). Expanding this population of cells *in vitro* could assist significantly in minimising the cost and time in thymus research.

During fetal thymus development, early patterning of the 3rd pharyngeal pouch into the thymus and the parathyroid specific structures occurs in a neighbouring cell – independent way (Griffith et al., 2009), while domain spreading of both organs and separation from the pharynx are dependent on the existence of a supporting mesenchymal population around both primordia (Franz, 1989). Interactions between the thymic epithelium that resides in the expanding primordium and the surrounding mesenchyme enables proliferation and differentiation of the undifferentiated TEPCs towards cortical and medullary lineages independently of further lympho-epithelia crosstalk (Klug et al., 2002). Previous research has indicated the WNT (Balciunaite et al., 2002), FGF (Frank et al., 2002), SHH (Moore-Scott and Manley, 2005) and BMP (Bleul and Boehm, 2005; Patel et al., 2006; Tsai, 2003) signalling pathways as regulators of the early TEPC state through endocrine signalling or provision of soluble growth factors, since diminished or enhanced activity of these pathways can lead to athymia or thymic hyperplasia. Additionally, it has been proposed that FOXN1 expression may be regulated by WNT and BMP signalling pathways (Balciunaite et al., 2002; Soza-Ried et al., 2008). Despite this progress, a comprehensive model that describes dynamic regulation of signalling activity during the transition from undifferentiated TEPCs into differentiating TEPCs has not yet been described.

To achieve the expansion of TEPCs *in vitro*, understanding of the intrinsic and extrinsic regulation of this upstream cell state and of FOXN1 initiation is needed, as expansion of TEPCs *in vitro* will require imposition of a reversible early block in differentiation and also fine-tuned experimental conditions for these cells to survive and grow. Thus, in **Chapter 4**, I aimed to generate an unbiased *in silico* model of signalling pathways activity during fetal thymus development that could be used to assist with TEP/SC maintenance and proliferation *in vitro*. Additionally, to support the integrity of this *in silico* model, throughout **section 4.2.3** of **Chapter 4** I have included experimental evidence that validate one of the predicted outcomes in the *in silico* model.

4.2 RESULTS

4.2.1 Pathways reshaped at the onset of *Foxn1* expression in TEP/SCs: pathways switching off

To predict signalling pathways that reshape on the onset of FOXN1 expression, when early undifferentiated TEPCs start to differentiate, I extended the differential expression analysis (described in **Chapter 3**, **section 3.2.1**) that was performed between the E10.5 and E12.5 samples from the TEPC Developmental series (RNA-seq data from **biological triplicates per stage** provided by **Harsh Vaidya**, Blackburn lab; see also **Chapter 2**, **section 2.2.1**) using pathway enrichment analysis.

For the pathway enrichment analysis, the fold change values of the total 15,321 differentially expressed genes from the comparative analysis of RNA-seq data from PLET1^{+} TEPCs isolated at E10.5 and E12.5 were provided as a 'Pre-Ranked Gene List' in GSEA (Subramanian et al., 2005) and analysed against the edited ConsensusPathDB database (Kamburov et al., 2011, details in **Chapter 2**, **section 2.3.3.6**). This approach should predict signalling pathways that alter between these two developmental stages.

The GSEA output indicated two groups. The first is pathways demonstrating increased intracellular signalling activity while TEPCs proceed from E10.5 to E12.5; these are referred to as up-regulated pathways. The second group is pathways that demonstrate decreased or complete loss of intracellular signalling activity while TEPCs proceed from E10.5 to E12.5; these are referred to as down-regulated pathways. Since the purpose of this analysis is to identify pathways implicated in regulation of the proliferation and survival of early undifferentiated TEPCs, I focused my analysis on pathways that exhibit more activity at the early E10.5 developmental stage and whose activity is found to decrease over time (down-regulated list of pathways). From these down-regulated pathways, only those that had an FDR ≤ 0.25 (enriched with statistical significance) and were related to signalling were considered for further investigation. This analysis identified 10 pathways which are presented in **Table 4.1**.

ID	GS follow link to MSigDB	SIZE	ES	NES	NOM p-val	FDR q-val
2	SIDS SUSCEPTIBILITY PATHWAYS	35	-0.65	-1.88	0	0.054
4	SHC-MEDIATED CASCADE	18	-0.71	-1.85	0.004	0.043
14	FRS2-MEDIATED CASCADE	27	-0.62	-1.73	0.002	0.073
20	NEGATIVE REGULATION OF FGFR SIGNALING	31	-0.6	-1.71	0.004	0.068
37	NEURAL CREST DIFFERENTIATION	69	-0.47	-1.56	0.007	0.161
38	HEDGEHOG SIGNALING PATHWAY - MUS MUSCULUS (MOUSE)	36	-0.54	-1.56	0.022	0.166
44	SRP-DEPENDENT COTRANSLATIONAL PROTEIN TARGETING TO MEMBRANE	106	-0.42	-1.51	0.005	0.221
45	IRS-RELATED EVENTS TRIGGERED BY IGF1R	69	-0.45	-1.5	0.007	0.223
47	IGF1R SIGNALING CASCADE	72	-0.44	-1.49	0.012	0.235
48	SIGNALING BY TYPE 1 INSULIN-LIKE GROWTH FACTOR 1 RECEPTOR (IGF1R)	72	-0.44	-1.48	0.007	0.243

Table 4.1: List of enriched signalling pathways with decreased activity in TEPCs between E10.5 and E12.5 timepoints. In sequence, the columns show the name of each enriched pathway (GS follow link to MSigDB), the number of Pre-Ranked genes identified per pathway (size), the enrichment score of each pathway (ES), the enrichment score of each pathway normalised for the size of the pathway (NES), the nominal *p*-value (NOM *p*-val) that is not adjusted for the pathways size or multiple hypothesis testing and the FDR *q*-val (see **Chapter 2**, **section 2.3.3.6**).

Pathway enrichment analysis can often be limited by the complexity of the database used to predict enrichment for a particular gene set. By complexity we refer to the size (number of pathways) and coverage (number of organisms and cell types represented) of the database in relation to the system under investigation. Database curation, based on the current literature, defines pathways and assigns genes to them accordingly. Thus, *de novo* pathway analysis can only identify known pathways. However, current annotation is far from perfect, with signalling pathways to also often have different function in different cell types, thus relying on pathways names to identify the pathways involved in regulation of a new cell type can be misleading. Additionally, when a pathway is represented by a large number of genes, random accumulation of some genes may indicate that the pathway is enriched (or active) in that cell type, even though the main signalling molecules of the pathway (such as ligands, intermediate signalling molecules and receptors) may be missing.

To further analyse the resulting pathways in **Table 4.1**, pathways that shared a large number of genes contributing to their enrichment were grouped together (see **Figure 4.1**). Additionally, individual genes that contribute to each pathway's core enrichment were manually inspected and when appropriate the related pathway was renamed to better fit the contained genes (for detailed listing of genes see sections below). For complicity, non-differentially expressed genes were also considered (if expressed) per pathway if they consisted essential molecules for the pathway's function. In the case of NEURAL CREST DIFFERENTIATION and HEDGEHOG SIGNALING PATHWAY pathways (#37 and #38 in **Table 4.1**, genes contributing to their core enrichment could be assigned to more than one major signalling pathways, thus all related pathways were included in the renaming of the #37 and #38 pathways (**Table 4.1**) respectively (see **Figure 4.1**). From these further analyses, six main pathways have been identified from the data presented in **Table 4.1**: the Fibroblast Growth Factor (FGF) signalling pathway, the NOTCH signalling pathway, the canonical and non-canonical Wingless Integration-1 (WNT) signalling pathways, the Sonic Hedgehog (SHH) signalling pathway and the Insulin-like Growth Factor (IGF) signalling pathway (**Figure 4.1**).

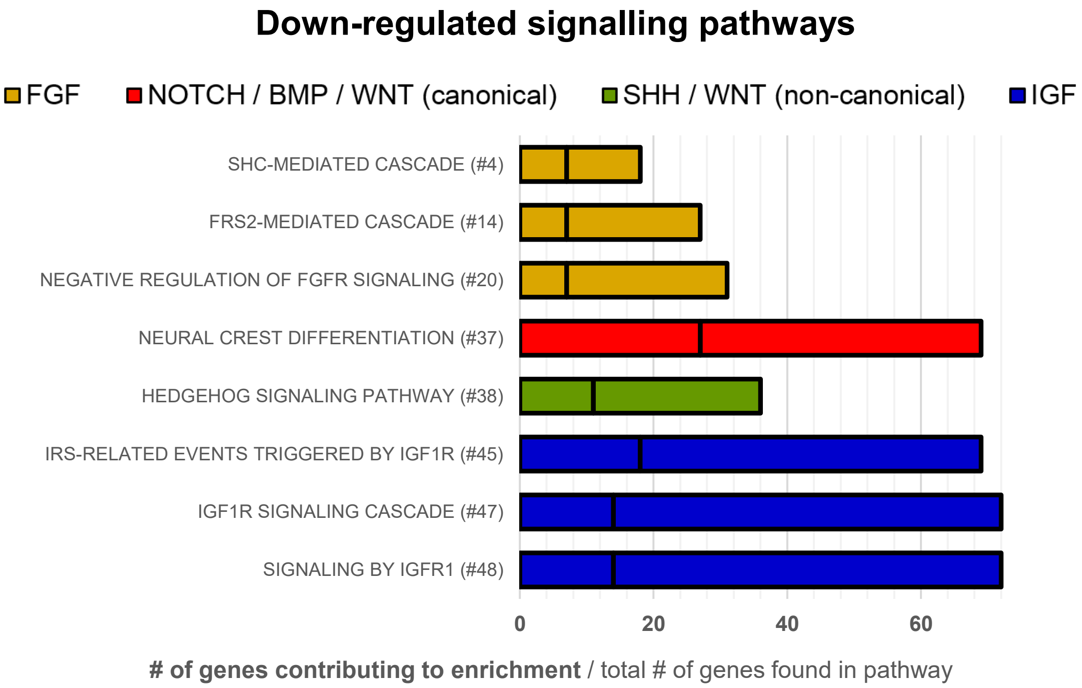


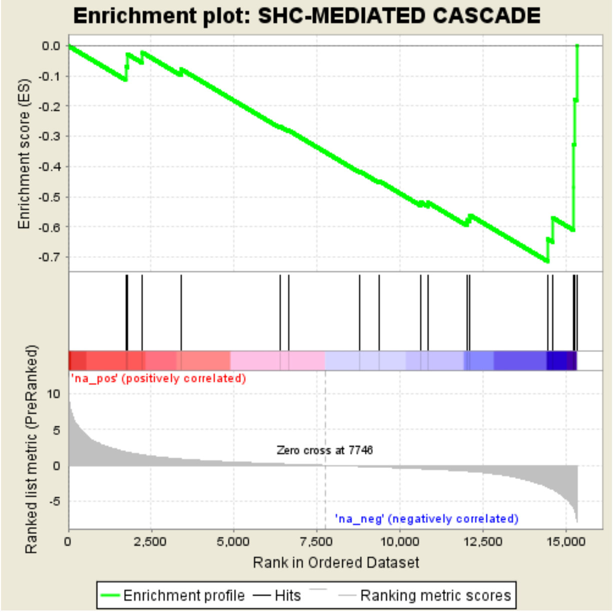
Figure 4.1: Grouped overview of the down-regulated signalling pathways based on core enrichment genes overlap. High overlap of core enrichment genes among pathways allowed grouping of the statistically significant signalling pathways (listed in **Table 4.1**) in broader categories with names that better represent their underlying signalling mechanism/process (colour coded pathways per category). In addition, the core enrichment genes of the NEURAL CREST DIFFERENTIATION (#37) and the HEDGEHOG SIGNALING PATHWAY (#38) pathways could be assigned to more than one major pathways and that is shown in the top label with major pathway names separated by slashes ("/").

The above analysis is informative about the intracellular signalling mechanisms that are active in early undifferentiated TEPCs, but it can also identify signalling pathways activated in response to ligands provided by neighbouring cells (e.g. mesenchymal cells or migrating haematopoietic progenitors) by assessing the expression of effector genes known to be activated in response to specific signalling.

In **sections 4.2.1.1** to **4.2.1.4**, I consider in detail the curated categories outlined in **Figure 4.1**, discussing the details of each category, including the expression levels of the relevant ligands, intracellular molecules and receptors, and core enrichment genes per pathway, and outlining which of these genes are differentially expressed with statistical significance (FDR ≤ 0.2, see **Chapter 3**, **section 3.2.2** for threshold selection) between the E12.5 and E10.5 from the TEPC Developmental series to represent more confident changes between the two timepoints. Representative graphs combining all resulting pathways in groups of two follow at the end of this section (**Figure 4.6**).

4.2.1.1 FGF signalling pathway

Three curated pathways, namely, the SHC-MEDIATED CASCADE (#4), FRS2-MEDIATED CASCADE (#14) and NEGATIVE REGULATION OF FGFR SIGNALING (#20) (**Table 4.1**), were found to share the same core genes contributing to their enrichment score (see **Figure 4.1** for overlap and **Figure 4.2** for the list of core enrichment genes). These were therefore grouped together under the category FGF signalling, since the enriched genes constitute the main ligands and receptors of the FGF signalling pathway (for detailed references to the FGF signalling pathway and its role in thymus development see **Chapter 1**, **section 1.2.4.3**). The top ranked core enrichment genes from the FGF signalling category (**Figure 4.2**) are considered below.



PROBE	RANK METRIC SCORE	CORE ENRICHMENT
Fgf8*	-7.535	Yes
Fgf10	-6.333	Yes
Fgf15	-5.93	Yes
Fgf3	-5.927	Yes
Fgf2	-3.359	Yes
Fgfr1*	-3.065	Yes

Figure 4.2: GSEA Enrichment plot for the SHC-MEDIATED CASCADE pathway and gene list of core enrichment. Enrichment score on the GSEA plot (top *y*-axis) indicates the degree of over-representation (in absolute value) for the provided pre-ranked gene set (differentially expressed genes – E12.5 versus E10.5) and the direction of their regulation (negative sign/down-regulated genes). The Pre-Ranked list metric (GSEA plot, bottom *y*-axis) provides information over the numeric ranking used to weight genes (in this case log fold change), while the barcode bars (GSEA plot, middle part) represent all gene hits in the particular pathway with genes contributing to core enrichment of the pathway listed in the table next to the Enrichment plot. Differentially expressed genes for an FDR ≤ 0.2 are colour-coded in blue (* FDR ≤ 0.05, ** FDR ≤ 0.01).

Based on the expression data from the TEPC Developmental serie, *Fgf8* was strongly expressed in the undifferentiated TEPCs at E10.5, but was not expressed in E11.5 or E12.5 TEPCs. *Fgf10* was moderately expressed in E10.5 TEPCs and was also absent at E11.5 and E12.5. *Fgf3*, despite its apparent down-regulation in terms of fold-change, exhibited very low level expression at all three TEPC developmental stages. *Fgfr1* was expressed in E10.5 TEPCs, albeit at lower levels than *Fgf8*, with its expression progressively to decrease at E11.5 while it is fully absent in the E12.5 TEPCs. Notably, *Fgf10*, *Fgf7* and *Fgf3* comprise the main ligands of the *Fgfr2-IIIb* receptor, and therefore expression levels of *Fgfr2-IIIb* are also noted in the collective pathways representation (**Figure 4.6**).

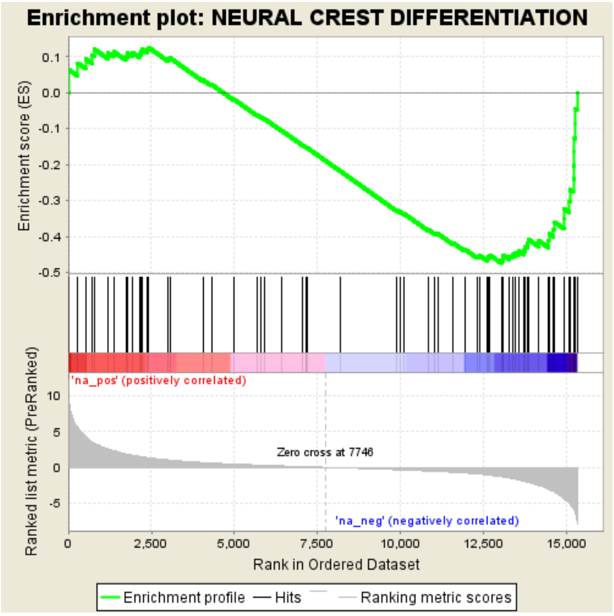
4.2.1.2 NOTCH, BMP and canonical WNT signalling pathways

Inspection of the NEURAL CREST DIFFERENTIATION (#37) pathway indicated that its core enrichment genes (see top right panel in **Figure 4.3**) mainly comprised regulators of the NOTCH signalling pathway.

Of NOTCH pathway genes, the receptor *Notch3*, ligands *Dll1*, and *Dll3,* effector *Rbpj* and target *Msx2* contributed significantly to the enrichment of the NEURAL CREST DIFFERENTIATION pathway. These genes demonstrate a continuous down-regulation from E10.5 to E12.5 stage in TEPCs, indicating a gradual decrease in pathway activity. However, NOTCH signalling seems to exhibit more complex regulation in the early development of the thymic epithelium since one of its main ligands, *Dll4*, is targeted by FOXN1 and reaches high levels of expression at E12.5, in contrast to the above trend. In agreement with the *Dll4* up-regulation, two of NOTCH's main target genes [*Hes1* (FDR ≤ 0.05) and *Heyl* (FDR ≤ 0.2)] also up-regulate when cells proceed towards the E12.5 stage. Finally, one of NOTCH's main receptors (*Notch1*) seems to be constitutively expressed in all three timepoints, suggesting that the pathway could be active in all of them but its activity may be dynamically regulated based on provision of ligands by different groups of cells at distinct timepoints.

Other genes contributing to the NEURAL CREST DIFFERENTIATION pathway enrichment, and also strongly down-regulated with statistical significance between

E10.5 and E12.5 (FDR ≤ 0.2), include *Fgf8* and *Fgfr1* (see **section 4.2.1.1** above), Homeobox B1 (*Hoxb1*), Peripheral myelin protein 22 (*Pmp22*), Transcription factor AP-2 alpha (*Tfap2a*), T-Cell leukemia homeobox 2 (*Tlx2*), Paired like homeobox 2b (*Phox2b*), Collagen type II alpha 1 chain (*Col2a1*) and Protogenin (*Prtg*). From the above, only *Col2a1* and *Prtg* were expressed at E12.5, while only *Col2a1*, *Prtg* and *Phox2b* were expressed at E11.5. The remaining genes were only present in E10.5 TEPCs.



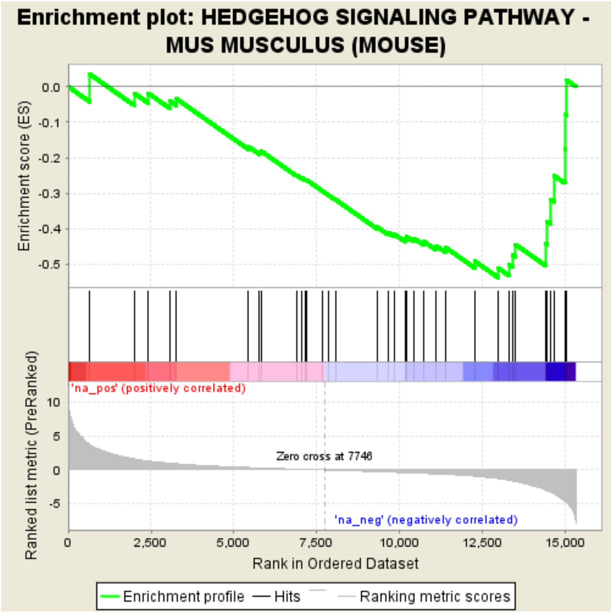
PROBE	RANK METRIC SCORE	CORE ENRICHMENT
Fgf8*	-7.535	Yes
Hoxb1*	-6.542	Yes
Pmp22*	-6.514	Yes
Msx2	-6.035	Yes
Tfap2a	-6.008	Yes
Fgf15	-5.93	Yes
Tlx2	-5.777	Yes
Phox2b	-5.083	Yes
Dll1	-4.99	Yes
Neurog1	-4.372	Yes
Snai1	-4.368	Yes
Cdh6	-3.426	Yes
Fgf2	-3.359	Yes
Tfap2b	-3.123	Yes
Fgfr1*	-3.065	Yes
Dlx5	-2.54	Yes
Dll3	-2.098	Yes
Col2a1*	-2.074	Yes
Rbpj	-2.003	Yes
Hoxa1	-1.941	Yes
Hdac5	-1.756	Yes
Mpz	-1.659	Yes
Cdh2	-1.562	Yes
Mbp	-1.488	Yes
Prtg	-1.355	Yes
Myb	-1.338	Yes
Notch3	-1.312	Yes

Figure 4.3: GSEA Enrichment plot for the NEURAL CREST DIFFERENTIATION pathway and gene list of core enrichment. Enrichment score on the GSEA plot (top *y*-axis) indicates the degree of over-representation (in absolute value) for the provided pre-ranked gene set (differentially expressed genes – E12.5 versus E10.5) and the direction of their regulation (negative sign/down-regulated genes). The Pre-Ranked list metric (GSEA plot, bottom *y*-axis) provides information over the numeric ranking used to weight genes (in this case log fold change), while the barcode bars (GSEA plot, middle part) represent all gene hits in the particular pathway with genes contributing to core enrichment of the pathway listed in the table next to the Enrichment plot. Differentially expressed genes for an FDR ≤ 0.2 are colour-coded in blue (* FDR ≤ 0.05, ** FDR ≤ 0.01).

4.2.1.3 SHH and WNT (non-canonical) signalling pathways

The HEDGEHOG SIGNALING PATHWAY (#38, **Table 4.1**) enrichment plot and list of genes contributing to the pathway's core enrichment are shown in **Figure 4.4**. Patched 1 (*Ptch1)*, Patched 2 (*Ptch2)*, Serine/Threonine Kinase 36 (*Stk36)* and GLI Family Zinc Finger 1 (*Gli1)* are all major effectors of the SHH pathway. Of these genes, *Ptch1* (the main receptor of *Shh*) was down-regulated between E10.5 and E12.5 with statistical significance (FDR ≤ 0.2), while closer inspection of the mRNA levels for all the above genes revealed very low level of expression at E10.5, which progressively became incompetent or fully absent, suggesting that this pathway is potentially switched off before TEPCs progress to E12.5. Inactivation of the SHH signalling in TEPCs is consistent with former literature that shows SHH signalling to drive cells differentiation towards the parathyroid fate by restricting the expression of the *Gcm2* main regulator of the parathyroid (Grevellec et al., 2011).

Further overview of the core list of genes for the HEDGEHOG SIGNALING PATHWAY suggested that some of the genes allocated to this pathway could also activate the WNT non-canonical pathway. WNT Family Member 5A (*Wnt5a)* and WNT Family Member 5B (*Wnt5b)* are components of the non-canonical WNT pathway and were both strongly down-regulated with significance from E10.5 to E12.5.



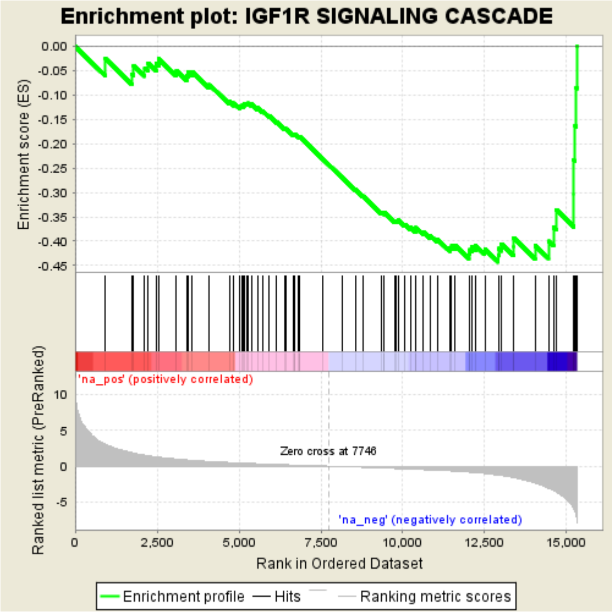
PROBE	RANK METRIC SCORE	CORE ENRICHMENT
Wnt9b	-4.758	Yes
Wnt5b	-4.699	Yes
Wnt3	-4.616	Yes
Stk36	-3.538	Yes
Ptch2	-3.287	Yes
Ptch1*	-3.112	Yes
Wnt5a*	-3.002	Yes
Wnt6	-1.7	Yes
Lrp2	-1.625	Yes
Hhip	-1.51	Yes
Gli1	-1.273	Yes

Figure 4.4: GSEA Enrichment plot for the HEDGEHOG SIGNALING PATHWAY and gene list of core enrichment. Enrichment score on the GSEA plot (top *y*-axis) indicates the degree of over-representation (in absolute value) for the provided pre-ranked gene set (differentially expressed genes – E12.5 versus E10.5) and the direction of their regulation (negative sign/down-regulated genes). The Pre-Ranked list metric (GSEA plot, bottom *y*-axis) provides information over the numeric ranking used to weight genes (in this case log fold change), while the barcode bars (GSEA plot, middle part) represent all gene hits in the particular pathway with genes contributing to core enrichment of the pathway listed in the table next to the Enrichment plot. Differentially expressed genes for an FDR ≤ 0.2 are colour-coded in blue (* FDR ≤ 0.05, ** FDR ≤ 0.01).

4.2.1.4 IGF signalling pathway

The SIGNALING BY IGFR1 (#48), IGF1R SIGNALING CASCADE (#47) and IRS-RELATED EVENTS TRIGGERED BY IGF1R (#45) pathways were also found to share the majority of core genes contributing to their enrichment (see **Figure 4.5**) and were grouped together under the IGF signalling category (**Figure 4.1**), since Insulin Like Growth Factor 1 (*Igf1)* and Insulin Like Growth Factor 2 (*Igf2)* are the main ligands of the IGF signalling pathway, while SHC Adaptor Protein 2 (Shc2), Insulin Receptor Substrate 1 (*Irs1*) and Insulin Receptor Substrate 4 (*Irs4*) consist part of the IGF signalling cascade. The genes enriched in these categories also included all those that contribute to the FGF pathway core enrichment (see **section 4.2.1.1**).

Both IGF ligands showed decreased expression at E12.5 compared to E10.5, with *Igf1* reaching very low levels of expression by E12.5. To form a more complete picture of the IGF intracellular activity, levels of the IGF receptors (IGF1R and IGF2R), as well as, IGF binding proteins (IGFBPs) were also analysed in addition to the IGF ligands.



PROBE	RANK METRIC SCORE	CORE ENRICHMENT
Fgf8*	-7.535	Yes
Igf1	-6.981	Yes
Fgf10	-6.333	Yes
Fgf15	-5.93	Yes
Fgf3	-5.927	Yes
Shc2	-3.588	Yes
Fgf2	-3.359	Yes
Fgfr1*	-3.065	Yes
Igf2**	-2.365	Yes
Fgf18	-1.592	Yes
Pde3b	-1.588	Yes
Them4	-1.291	Yes
Irs1	-1.254	Yes
Irs4	-1.24	Yes

Figure 4.5: GSEA Enrichment plot for the IGF1R SIGNALING CASCADE and gene list of core enrichment. Enrichment score on the GSEA plot (top *y*-axis) indicates the degree of over-representation (in absolute value) for the provided pre-ranked gene set (differentially expressed genes – E12.5 versus E10.5) and the direction of their regulation (negative sign/down-regulated genes). The Pre-Ranked list metric (GSEA plot, bottom *y*-axis) provides information over the numeric ranking used to weight genes (in this case log fold change), while the barcode bars (GSEA plot, middle part) represent all gene hits in the particular pathway with genes contributing to core enrichment of the pathway listed in the table next to the Enrichment plot. Differentially expressed genes for an FDR ≤ 0.2 are colour-coded in blue (* FDR ≤ 0.05, ** FDR ≤ 0.01).

4.2.2 A testable model of the signalling pathways required to maintain and expand TEP/SCs *in vivo*

From the outcomes of the pathway enrichment analysis presented in **section 4.2.1** and its subsections, I set out to create an *in silico* model that collectively summarises the pathways that down-regulate or switch off when cells proceed towards differentiation. The main regulators of the signalling pathways predicted in the enrichment analysis above (**section 4.2.1**) including the ligands, transient signalling molecules and targets are illustrated in **Figure 4.6**. The core enrichment genes exhibiting consistent changes (**Figure 4.2**-**Figure 4.5**, coloured in blue) between the two developmental timepoints were then used to guide assembly of a signalling pathways model over all pathways. To improve the consistency of the model, when main pathway regulators were not part of the core enrichment gene list, they were added to the graph, with representative expression values from the TEPC Developmental series dataset.

The SHH, WNT (canonical and non-canonical), NOTCH, BMP, FGF and IGF signalling pathways are thus depicted in **Figure 4.6**, with SHH effectively absent from all TEPC stages, the IGF, FGF and non-canonical WNT pathways to show a substantial decrease in ligand-expression (note: extrinsic provision of these ligands is possible), while NOTCH activity exhibits more complex regulation. Previous literature has already demonstrated roles for the SHH, FGF and IGF pathways in early fetal thymus development, partially validating the outcomes of this model.

The nature of *Wnt5a,* as a secreted protein, suggests that it could affect distinct regulatory pathways downstream in TEPCs but also in surrounding cells. Previous research has identified multiple alternative receptors (He et al., 2008; Keeble, 2006; Martinez et al., 2015) through which WNT5A may establish its action, as well as partner proteins (Matsuyama et al., 2009) that can regulate cells by interacting with WNT5A. In the model illustrated in **Figure 4.6**, *Wnt5a* is expressed very highly in E10.5 TEPCs and it subsequently down-regulates (FDR ≤ 0.03) in the later developmental stages. A list of potentially involved receptors and interactors based on current literature is shown in **Figure 4.7**, where the log2 (RPKM) normalised expression values from the TEPC Developmental series dataset were used to depict the expression levels of these genes during early TEPC development. Even though both *Ryk* and *Sfrp1* genes are expressed in a high enough level to allow signal transduction, the only gene that mimics *Wnt5a* pattern of expression and down-regulates with statistical significance (FDR ≤ 0.05) towards the E12.5 timepoint is WNT5A potential interactor, *Sfrp2*. Based on this, we can speculate that WNT5A-SFRP2 (and/or SFRP1) protein interaction, as well as binding of WNT5A to the RYK receptor could control signalling processes in early undifferentiated TEPCs.

It is apparent in **Figure 4.6** that NOTCH signalling activity (as represented from the early TEPC samples) cannot describe a definitive trend of action. The NOTCH ligand genes *Dll1* and *Dll3*, receptor gene *Notch3* and intermediate molecule *Rbpj* are clearly "switching off" by E12.5, when the NOTCH ligand *Dll4*, and target genes *Hes1* and *Heyl* exhibit their highest expression level, with the NOTCH receptor *Notch2* to be constitutively highly expressed in all three developmental stages. It is possible that different levels of activity and involved molecules take part in distinct developmental stages. Throughout these stages, the undifferentiated TEPCs have evolved towards more fated progenitors, now possibly constituting a more progressed cell type in the thymic lineage which may require NOTCH signalling for different purposes. It is also likely that NOTCH may affect cells in a non-equal manner, since a number of cells that exhibit NOTCH signalling need to be the signal senders, while their neighbouring cells should be receiving the signal through cell-to-cell interactions. It is evident that the role of NOTCH in early TEPCs requires further exploration to generate a more definitive picture of its action.

Notably, *Msx2* has been linked to the NOTCH, BMP and WNT pathways during differentiation of other cell types other than the thymic epithelium. In the most relevant example, *Msx2* together with *Foxn*1 have been shown to act upstream of the NOTCH pathway and promote hair shaft differentiation (Cai et al., 2009). Furthermore, in multipotent mesenchymal progenitors, BMP signalling has been shown to up-regulate *Msx2* which in turn promotes differentiation of the osteogenic lineage that contributes into calcification of the vasculature (Cheng et al., 2003). Additionally, BMP-*Msx2* signalling has been demonstrated to up-regulate the canonical WNT signalling through up-regulation of the WNT ligands in artery calcification (Shao et al., 2007). Lastly, during osteogenic differentiation of vascular smooth muscle cells, *Msx2* is regulated through NOTCH signalling in an independent manner of that of the BMP signalling (Shimizu et al., 2009). In short, *Msx2* consists a common target or regulator of BMP, NOTCH and WNT signalling pathways and its expression is highly related to the cells differentiation state.

In summary, in **Figure 4.6**, I provide a collective *in silico* model of signalling activity for the early undifferentiated TEPCs that can assist in better understanding TEPC differentiation *in vivo*. Detailed interpretation of this model in combination with the links to known literature are provided in **section 4.3** of this chapter.

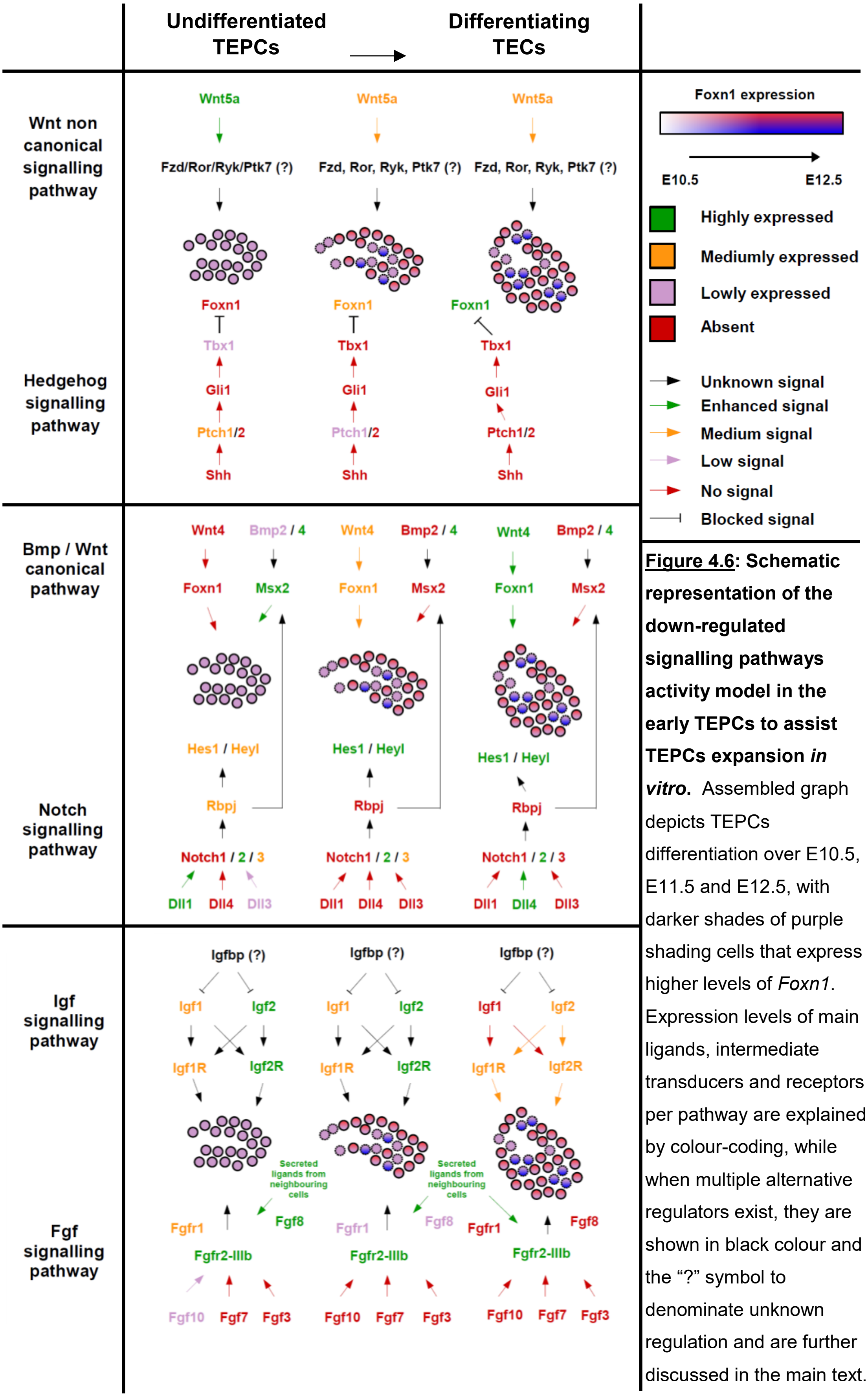

Undifferentiated TEPCs
Differentiating TECs
Wnt non canonical signalling pathway
Wnt5a
Wnt5a
Wnt5a
Fzd/Ror/Ryk/Ptk7 (?)
Fzd, Ror, Ryk, Ptk7 (?)
Fzd, Ror, Ryk, Ptk7 (?)
Foxn1
Foxn1
Foxn1
Tbx1
Tbx1
Tbx1
Hedgehog signalling pathway
Gli1
Gli1
Gli1
Ptch1/2
Ptch1/2
Ptch1/2
Shh
Shh
Shh
Bmp / Wnt canonical pathway
Wnt4
Bmp2 / 4
Wnt4
Bmp2 / 4
Wnt4
Bmp2 / 4
Foxn1
Msx2
Foxn1
Msx2
Foxn1
Msx2
Hes1 / Heyl
Hes1 / Heyl
Hes1 / Heyl
Notch signalling pathway
Rbpj
Rbpj
Rbpj
Notch1 / 2 / 3
Notch1 / 2 / 3
Notch1 / 2 / 3
Dll1
Dll4
Dll3
Dll1
Dll4
Dll3
Dll1
Dll4
Dll3
Igf signalling pathway
Igfbp (?)
Igfbp (?)
Igfbp (?)
Igf1
Igf2
Igf1
Igf2
Igf1
Igf2
Igf1R
Igf2R
Igf1R
Igf2R
Igf1R
Igf2R
Secreted ligands from neighbouring cells
Secreted ligands from neighbouring cells
Fgf signalling pathway
Fgfr1
Fgf8
Fgfr1
Fgf8
Fgfr1
Fgf8
Fgfr2-IIIb
Fgfr2-IIIb
Fgfr2-IIIb
Fgf10
Fgf7
Fgf3
Fgf10
Fgf7
Fgf3
Fgf10
Fgf7
Fgf3
Foxn1 expression
E10.5
E12.5
Highly expressed
Mediumly expressed
Lowly expressed
Absent
Unknown signal
Enhanced signal
Medium signal
Low signal
No signal
Blocked signal

Figure 4.6: **Schematic representation of the down-regulated signalling pathways activity model in the early TEPCs to assist TEPCs expansion *in vitro*.** Assembled graph depicts TEPCs differentiation over E10.5, E11.5 and E12.5, with darker shades of purple shading cells that express higher levels of *Foxn1*. Expression levels of main ligands, intermediate transducers and receptors per pathway are explained by colour-coding, while when multiple alternative regulators exist, they are shown in black colour and the “?” symbol to denominate unknown regulation and are further discussed in the main text.

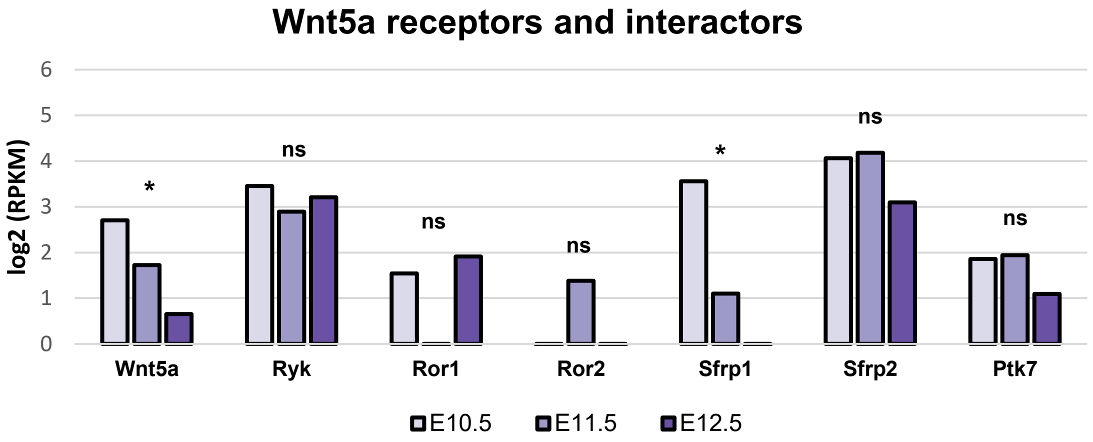


Figure 4.7: Expression profiles of *Wnt5a* receptors and interactors in TEPCs. Average log2 RPKM expression values of the WNT5A receptors and interactor genes (including *Wnt5a* itself) over the E10.5, E11.5 and E12.5 timepoints (sample triplicates) from the TEPC Developmental series. Statistical significance calculated from the differential expression analysis between the E10.5 and E12.5 timepoints using the limma *voom* function is denoted as follows: ns (not significant), * ($P \leq 0.05$),** ($P \leq 0.01$), *** ($P \leq 0.001$).

4.2.3 Exploring NOTCH signalling in the thymic epithelium: regulation of the medullary thymic epithelial cell lineage via specification of medullary progenitor cells.

NOTCH signalling has a crucial role in thymocyte development (Shah and Zuniga-Pflucker, 2014) and its down-regulation has recently been shown to be required to promote the later stages of mTEC development (Goldfarb et al., 2016; Jiang et al., 1998; Masuda et al., 2009). However, the potential for NOTCH to regulate early TEPC development has not yet been studied in detail. The pathway enrichment analysis in **section 4.2.1** has indicated NOTCH signalling to exhibit complex expression in the PLET1$^+$ TEPCs at E10.5, E11.5 and E12.5 developmental stages, with *Rbpj* (the transcriptional effector required to transduce NOTCH signals) showing decreased expression when proceeding towards the E12.5 stage (see **4.2.1.2**), complementary to the decreased activity of the other NOTCH pathway genes *Notch3*, *Dll1*, *Dll3*, and *Msx2* (**Figure 4.3**), while the NOTCH ligand *Dll4* and the NOTCH target genes *Hes1* and *Heyl* demonstrate an opposite pattern of behaviour with significant up-regulation of their expression towards the E12.5 TEPCs. This implicates NOTCH signalling as a potential regulator of the early fetal TEPC state, and therefore I selected this pathway for further investigation.

4.2.3.1 Genetic ablation of NOTCH signalling in early TEPCs under the control of *Foxn1* demonstrates an effect only in the mTEC compartment

To validate and better understand the importance of NOTCH activity in early fetal TEPCs and differentiating TECs, NOTCH signalling had to be blocked in all TECs. RBPJ is the main signal transducer of NOTCH signalling and deletion of it should block NOTCH activity in cells that express NOTCH. Therefore, *Foxn1*^{Cre} mice were crossed with an *Rbpj* conditional knockout (cKO) line (**Dong Liu**, Blackburn lab), creating a mouse model line (*Foxn1*^{Cre}*Rbpj*$^{Fl/Fl}$ mice), where RBPJ was absent from all TECs (*Rbpj* exon deletion is under the control of *Foxn1* promoter). Analysis of 2 week old mice (**Dong Liu**, Blackburn lab) demonstrated a reduced proportion of mTECs among total TECs in both male and female mice, while numbers of cTECs remained unaffected. T-cell differentiation proceeded normally in the *Rbpj* cKO mice, with T regulatory cells showing no apparent differences when compared to the control mice suggesting that the model was specific for the TEC compartment of the developing thymus (Liu et al., 2017, submitted). These data suggested a unique impact of the early loss of NOTCH signalling activity (loss-of-function assay) in the differentiating mTEC population.

Therefore, to probe at the transcriptional level the effect of loss of RBPJ on TEPC, RNA-seq datasets from wild type (WT) and *Foxn1*^{Cre}*Rbpj*$^{Fl/Fl}$ (RBPJ cKO; loss-of-function; LOF) PLET1^{+} TEPC at E12.5 (**biological triplicates per population**; see also **Chapter 2**, **section 2.2.1.3**), and PLET1^{+} and PLET1^{-} TEPC at E14.5 (**biological triplicates per population**; see also **Chapter 2**, **section 2.2.1.3**) were used; these datasets were generated by **Dong Liu** (Blackburn lab, University of Edinburgh). The cell populations analysed were chosen since, at E12.5, although the PLET1^{+} TEPC population is already heterogeneous, and will contain cTEC-fated cells along with common TEPCs and potentially mTEC-restricted progenitors, it cannot be split on the basis of known cell surface markers. At E14.5, prospective mTECs appear to be contained within the PLET1^{+} population, while prospective cTECs have down-regulated this marker. NOTCH resides in both TEC compartments.

To retrieve the normalised gene counts of the WT and RBPJ cKO (LOF) datasets, I ran the RNA-seq pre-analysis and core-analysis pipelines (described in **Chapter 2**, **sections 2.3.2** and **2.3.3)** that I have put together for the analysis of RNA-seq datasets in this thesis. I next used the generated log2 RPKM normalised values of the above datasets to observe samples similarity based on their expressional profiles and I also further applied differential expression analysis between selected samples (see next paragraphs) to identify genes that change with statistical significance.

No obvious differences were observed from the differential expression analysis performed between the PLET1^{+} WT and PLET1^{+} LOF samples at E12.5. As expected considering the timing of deletion of *Rbpj* using *Foxn1*^{Cre} (approximately at E12, Liu et al., 2017, submitted), the time period between samples collection and *Rbpj*'s exon deletion is too little to allow any differences to become apparent. In accordance to this, Principal Component Analysis (PCA) of the top 2,000 most variable genes from the E14.5 RBPJ PLET1$^+$ and PLET1$^-$ WT and LOF samples (samples collected two days later by **Dong Liu**, Blackburn lab) revealed a slight separation of the E14.5 PLET1$^+$ WT from the E14.5 PLET1$^+$ LOF samples on PC2, while no differences were found among the PLET1$^-$ samples (**Figure 4.8**). This analysis therefore verified that loss of RBPJ, and consequently NOTCH signalling activity, affected only the TEC compartment containing prospective mTECs and undifferentiated TEPCs. Additionally, clustering of the E14.5 PLET1$^+$ LOF samples closer to the PLET1$^-$ samples (PC2), suggested that TEPC samples with an ablated NOTCH signal during early TEPC development (after E12) acquires a more cTEC-like phenotype similarly to the PLET1$^-$ samples. In the same PCA plot, biological variability among samples is represented in PC1. This effect is apparent when observing PLET1$^+$ and PLET1$^-$ groupings per replicate across the *x*-axis (note that the PLET1$^+$ and PLET1$^-$ populations were sorted from the same cell preparations). Overall, clustering analysis of the E14.5 PLET1$^+$ and PLET1$^-$ samples shows that ablation of the NOTCH signalling pathway in early TEPCs under the *Foxn1* promoter uniquely impacts on the mTEC lineage with also PLET1$^+$ LOF samples to better resemble on average PLET1$^-$ samples of the same day. The specific impact of the mTEC lineage was also supported by flow cytometric analysis (**Dong Liu**, Blackburn lab), demonstrating a proportional and numerical decrease in mTEC numbers in E14.5 PLET1$^+$ LOF samples (Liu et al., 2017, submitted).

To identify consistent expressional changes of genes between the E14.5 PLET1$^+$ RBPJ cKO samples and their controls, I performed differential expression analysis between the E14.5 PLET1$^+$ RBPJ cKO and the E14.5 PLET1$^+$ WT groups using the limma *voom* function (**Chapter 2**, **section 2.3.3.2**). This revealed very few consistent differences with gene names to be outlined in the text here [Desmoglein 3 (*Dsg3*) and Regulator of G protein signalling 4 (*Rgs4*) (FDR < 0.1), Myosin light chain 7 (*Myl7*) (FDR < 0.2)] and a clear trend of NOTCH targets to be down-regulated in the E14.5 PLET1$^+$ RBPJ cKO samples, though their FDR values were higher than the pre-set threshold of 0.2 (**Table 4.2**). Down-regulation of these targets was verified by RT-qPCR analysis (Liu et al., 2017, submitted). From the RT-qPCR analysis it became apparent that *Notch1*, *Notch2*, *Notch3* and *Jag1* in both E14.5 PLET1$^+$ and PLET1$^-$ TECs were significantly more lowly expressed in cKO thymi compared to controls (data from **Dong Liu**). This pattern suggested that NOTCH activation may be causing up-regulation of its own receptors and JAG1 ligand, in a positive feedback loop. One could contemplate that this positive feedback loop may lead to a sharper responsiveness to further boost NOTCH signalling. Of note is that average *Foxn1* and *Plet1* expression levels were unaffected by loss of RBPJ (**Table 4.2**).

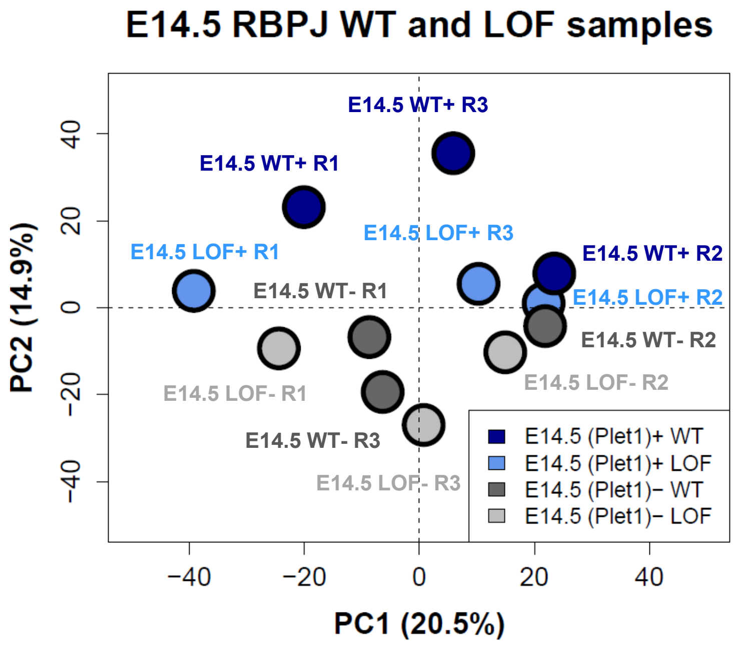


Figure 4.8: PCA plot of the E14.5 RBPJ WT and LOF (RBPJ cKO) samples. Samples on PC2 separate at a first level into PLET1$^+$ (dark blue/light blue) and PLET1$^-$ (dark grey, light grey), while on a second level the E14.5 PLET1$^+$ WT samples (dark blue) are distinct to the E14.5 PLET1$^-$ WT samples demonstrating the effects of RBPJ cKO specifically in the PLET1$^+$ compartment. PC1 is representative of the high variability among the biological replicates, since samples are clustered based on their biological replicate.

Developmental stage		E14.5											
PLET1 status		**PLET1+**						**PLET1-**					
Genotype & sample name		**WT**			**LOF**			**WT**			**LOF**		
Gene symbol	**Ensembl ID**	**Rep1**	**Rep2**	**Rep3**	**Rep1**	**Rep2**	**Rep3**	**Rep1**	**Rep2**	**Rep3**	**Rep1**	**Rep2**	**Rep3**
Notch1*	ENSMUSG00000026923	0.19	0.06	0.58	-8.69	-1.03	-1.74	-2.39	0.01	-2.30	-1.56	-1.25	-2.05
Notch2*	ENSMUSG00000027878	4.91	3.40	4.02	6.43	3.29	3.44	4.78	3.64	3.46	5.20	3.27	3.68
Notch3*	ENSMUSG00000038146	-0.16	-0.14	0.48	-8.20	-3.79	-5.55	-1.78	-0.45	-1.72	-4.82	-1.50	-2.28
Jag1*	ENSMUSG00000027276	5.51	4.49	5.83	6.34	4.18	3.73	4.52	3.40	3.85	4.86	3.32	3.49
Hes1*	ENSMUSG00000022528	4.32	4.10	4.43	2.97	4.10	4.49	4.50	4.63	4.82	3.88	4.06	3.41
Heyl*	ENSMUSG00000032744	1.17	2.36	2.18	-5.52	0.18	-2.96	1.77	2.76	2.72	-2.39	2.03	1.25
Ascl1	ENSMUSG00000020052	-5.37	-0.65	0.54	0.37	-0.97	-3.54	-0.32	-0.61	-6.48	-5.83	-1.83	-6.48
Foxn1	ENSMUSG00000002057	4.89	5.95	4.94	2.64	5.51	4.58	5.06	6.30	6.58	4.42	5.82	5.52
Plet1	ENSMUSG00000032068	3.01	3.36	3.54	2.36	3.57	4.00	-5.85	1.77	-0.54	-5.85	0.67	-2.49

Table 4.2: Expression values of NOTCH targets for the E14.5 RBPK cKO (LOF) and WT samples. Log2 RPKM values are provided for the NOTCH gene targets in the E14.5 PLET1^{+} and PLET1^{-} LOF and WT samples. "*" symbol marks NOTCH target genes whose down-regulation is supported by RT-qPCR analysis (Liu et al., 2017, submitted).

4.2.3.2 Time-dependence of mTECs ablation during early thymus development

The functional analyses conducted by **Dong Liu** have demonstrated that loss of NOTCH signalling in E12.5 TEPCs causes a numerical decrease in mTECs, and that NOTCH signalling is required for mTEC specification in a limited time-window prior to E16.5 (Liu et al., 2017, submitted). While some mTECs are generated in the *Foxn1*^{Cre}; *Rbpj*$^{Fl/Fl}$ model, this likely reflects that by the time of onset of deletion (around E12), some TEPCs may have already acquired an mTEC fate.

My RNA-seq analyses of the TEPC Developmental series and the *Foxn1* Allelic series (described in **Chapter 3, section 3.2.1**) shows that *Rbpj* expression levels are lower in the E11.5 and E12.5 TEPCs, with *Plet1* having statistically significant (FDR ≤ 0.05) down-regulation between the E10.5 and E12.5 timepoints. This suggests that the NOTCH signalling operation through RBPJ would normally be reduced by E11.5 and can partially explain the lack of more consistent differences among the E14.5 PLET1$^+$ RBPJ cKO and E14.5 PLET1$^+$ WT samples. The lower levels of *Rbpj* observed at the later timepoints (E11.5 and E12.5) may represent a general decrease of NOTCH signalling in all TEPCs that express NOTCH. However, if a few TEPCs retain high levels of *Rbpj,* this would be masked by the population analysis (results in support of TEPCs heterogeneity after *Foxn1* expression are described in **Chapter 3**). In either case, impacting *Rbpj* levels at a stage where its expression has already been limited should not lead to very distinguishable differences, as observed in **Figure 4.8**. The findings presented here provide an explanation for the time-dependency of NOTCH function observed in experiments using pharmacological inhibition of NOTCH activity (Liu et al, submitted).

Previous literature has suggested that divergence of mTEC progenitors can occur independently of *Foxn1* expression and earlier than E12.5 (Hamazaki et al., 2007; Nowell et al., 2011). To study the impact of loss of NOTCH signalling on mTECs at a FOXN1-independent stage when *Rbpj* is more highly expressed, **Dong Liu** cultured E10.5 pharyngeal pouches (PP) in control and DAPT conditions and demonstrated an almost complete inhibition of the mTEC lineage in the presence of DAPT (Liu et al., 2017, submitted). In other words, blockage of NOTCH signalling before *Foxn1* expression fully ablated the mTEC fate and maintained the cTEC fate of TEPCs. As will be demonstrated in **Chapter 5**, **section 5.2.3**, *Rbpj* is predicted to be a direct target of *Foxn1* in TEPCs, a result that comes in agreement with the more severe phenotype caused by ablation of NOTCH signalling prior to *Foxn1* expression; which would itself cause down-regulation of *Rbpj*. Therefore, these data suggested that high NOTCH activity in early undifferentiated TEPCs is necessary for mTECs specification, a process possibly controlled by *Foxn1* levels per cell. This prediction could be further reinforced by the single-cell (sc) RNA-seq data of the E12.5 TEPCs.

Collectively, the bioinformatics analyses and clustering of the PLET1$^{+/-}$ RBPJ cKO and WT samples, that I have performed in **sections 4.2.3.1** and **4.2.3.2** have demonstrated a consistent (though not statistically significant) decrease in the expression of multiple genes involved in NOTCH signalling and have further implicated a unique role of NOTCH signalling in the mTEC compartment (see separation of PLET1^{+} LOF and WT samples on PC2, **Figure 4.8**). Additionally, observation of *Rbpj* expression levels across the TEPC Development series and the *Foxn1* Allelic series integrated dataset in conjunction with ChIP-seq analysis of a FOXN1-tagged protein dataset (Žuklys et al., 2016) presented and reanalysed in **Chapter 5** have suggested a potential time-dependence of NOTCH's signalling impact in the mTEC lineage, with a more severe phenotype (mTEC ablation) to be expected in earlier deactivation (prior to *Foxn1* expression) of the NOTCH signalling. Data generated by **Dong Liu** (Blackburn lab, University of Edinburgh) have fully validated and further expanded the findings from the bioinformatics analyses here, verifying the expressional differences in the NOTCH target genes and the unique impact on mTECs, as well as the suggested time dependence model, implicating NOTCH as an important regulator of the mTEC specification in early TEC development in a short-defined time window.

4.2.3.3 Genetic reinforcement of NOTCH signalling in early TEPCs perturbs the exit from the early TEPC state

Since loss of NOTCH signalling leads to loss of the mTEC lineage, it was interesting to determine whether enforced NOTCH signalling activity might convert all TEPCs into mTECs. This was addressed using a TEC specific gain-of-function (GOF) model, in which *Foxn1*^{Cre} mice were crossed with *R26-LoxP-stop-LoxP-NICD-IRES-eGFP* (NICD) mice and generated *Foxn1*^{Cre}*;*NICD mice that would express constitutively active NOTCH signalling and high (but normal) levels of NICD (**Dong Liu**, Blackburn lab). These NOTCH gain-of-function experiments led to NICD thymi with higher levels of PLET1 and lower levels of MHCII in comparison to the E14.5 controls, indicating a delay in TEC differentiation (Liu et al., 2017, submitted).

To probe the effect of sustained NOTCH signalling activity on early TEPCs at the transcriptional level, I analysed the RNA-seq datasets from five E14.5 PLET1^{+} NICD TEPC samples (data provided by **Dong Liu**, Blackburn lab; cell numbers for E14.5 PLET1^{-} NICD TEPC samples were extremely low) using the RNA-seq pre-analysis and core-analysis pipelines that I have put together for the RNA-seq data processing in this thesis (described in **Chapter 2**, **section 2.2.1.3**) in order to retrieve the normalised gene counts for all expressed genes of these datasets. PCA of the E12.5 and E14.5 WT and RBPJ cKO (LOF) samples (**biological triplicates per population**) with the E14.5 NICD (GOF) samples revealed three groups (**Figure 4.9**); Group 1: E14.5 PLET1^{+} NICD; Group 2: E14.5 PLET1^{+} and PLET1^{-} WT and LOF samples; and Group 3: E12.5 PLET1^{+} WT and LOF samples. Comparison of these groups showed that Group 2 and 3 shared some similarity in PC2, while Group 1 was found to cluster in-between Groups 2 and Group 3 in PC1 (but closer to Group 2). This suggested that the E14.5 NICD samples had undergone a block in differentiation, since they all remain PLET1^{+} (Liu et al., 2017, submitted), or that over activation of NOTCH signalling had inducted a rare or aberrant TEC subtype.

Since within batch effect correction was not possible between the three groups due to confounded batch effect with the biological differences across groups, the profiles of housekeeping genes were plotted among all groups to track preferences towards any specific group (**Figure 4.9**). No significant differences were observed among groups for the housekeeping genes, allowing for across group comparisons. I additionally performed differential expression analysis with limma (for details see **Chapter 2**, **section 2.3.3.5**) between the E14.5 PLET1$^+$ GOF and E14.5 PLET1$^+$ WT samples which clearly predicted several NOTCH targets as up-regulated in the GOF samples with statistical significance (FDR ≤ 0.2), while again none of the housekeeping genes were found differentially expressed in the same comparison.

Finally, I performed an independent signalling pathway enrichment analysis using the total number of differentially expressed genes between the E14.5 PLET1$^+$ GOF versus the E14.5 PLET1$^+$ WT samples and their fold changes (calculated by limma *voom*) as a 'Pre-Ranked Gene List' in GSEA (Subramanian et al., 2005) against the edited ConsensusPathDB database (Kamburov et al., 2011, details in **Chapter 2**, **section 2.3.3.6**), which also predicted NOTCH signalling pathway as enriched between the GOF and the WT groups among two other signalling pathways that were identified (shown in **Table 4.3**). The GSEA Enrichment plot and the list of genes contributing to its core enrichment are shown in **Figure 4.10**.

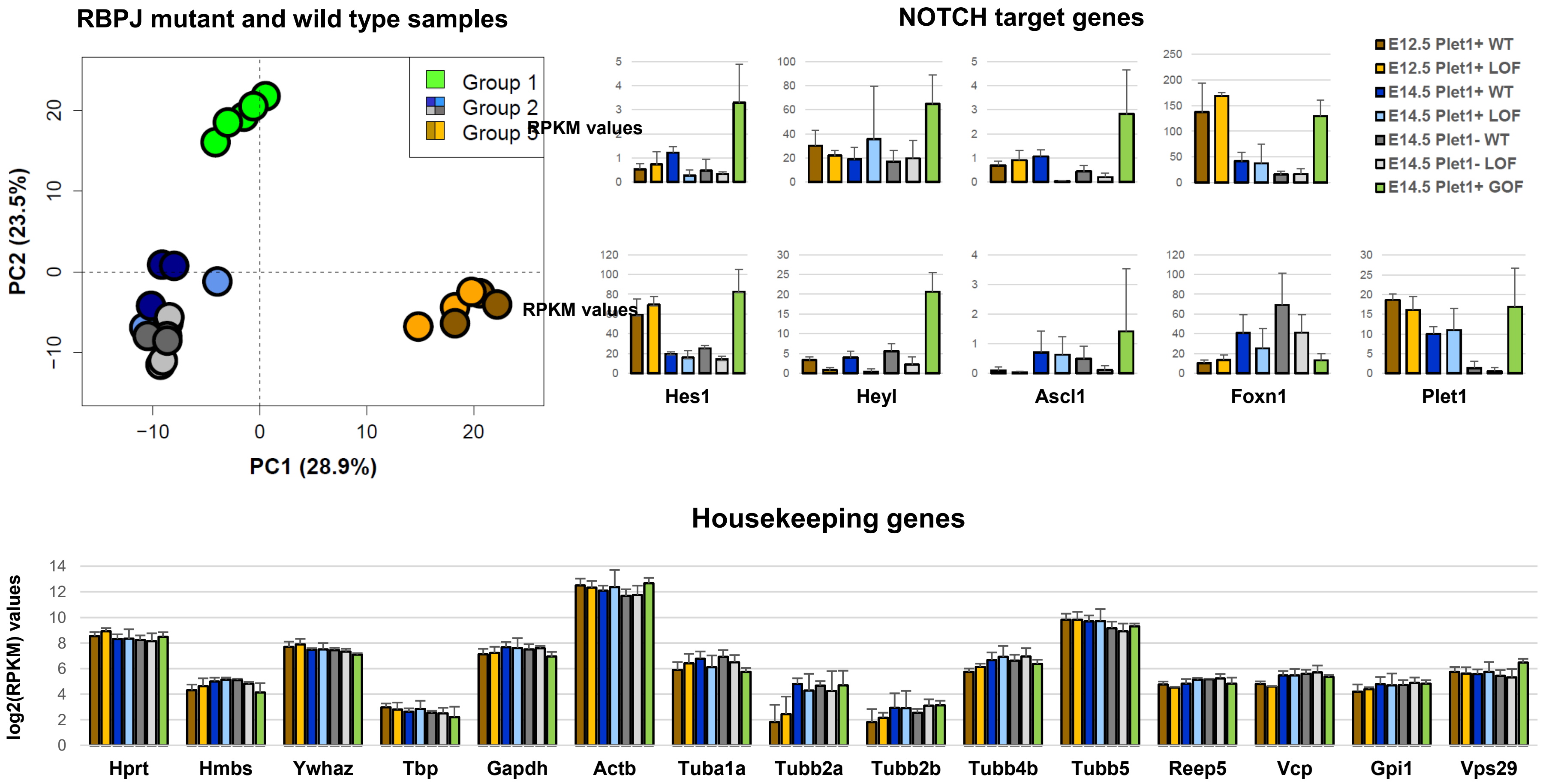
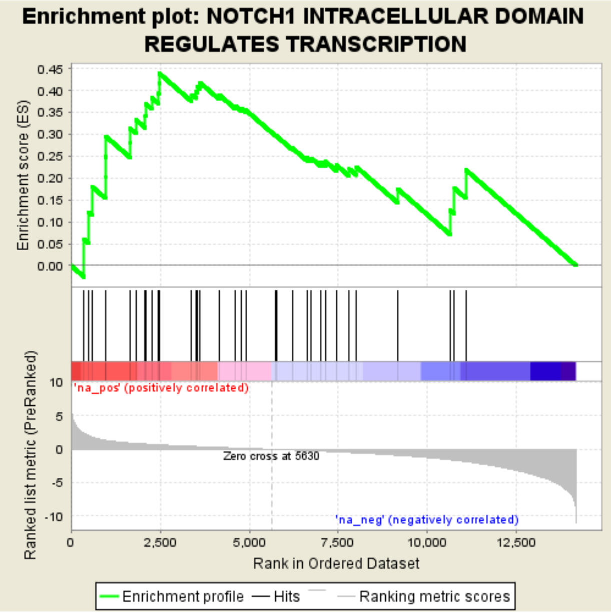


Figure 4.9: Differences among RBPJ cKO, NICD and WT groups. PCA plot (top right panel) depicts a clear separation among the E12.5 WT/RBPJ cKO (orange/brown), the E14.5 WT/RBPJ cKO (blues/greys) and the NICD group (green). The top right panel shows how NOTCH genes are impacted in the aforementioned samples with *Notch3*, *Hes1* and *Heyl* found to be statistically significant between the GOF versus the E14.5 PLET1$^+$ WT group. The bottom barplot shows that no strong patterns were observed for the housekeeping genes among these datasets.

ID	GS follow link to MSigDB	SIZE	ES	NES	NOM p-val	FDR q-val
2	INFLAMMATORY RESPONSE PATHWAY	16	0.54	1.71	0.017	0.053
4	NOTCH1 INTRACELLULAR DOMAIN REGULATES TRANSCRIPTION	37	0.44	1.66	0	0.067
14	SIGNALING BY CONSTITUTIVELY ACTIVE EGFR	18	0.43	1.39	0.054	0.232

Table 4.3: Up-regulated signalling pathways in the E14.5 PLET1$^+$ GOF versus E14.5 PLET1$^+$ WT comparison. In sequence, the table provides name of each pathway (GS follow link to MSigDB), number of Pre-Ranked genes identified per pathway (size), enrichment score of each pathway (ES), enrichment score of each pathway normalised for the size of the pathway (NES), NOM *p*-val and FDR *q*-val (see **Chapter 2**, **section 2.3.3.6** for more details).



PROBE	RANK METRIC SCORE	CORE ENRICHMENT
Heyl	2.43	Yes
Hes1**	2.039	Yes
Hey2	1.832	Yes
Maml3*	1.349	Yes
Gm10093	1.347	Yes
Notch1	1.338	Yes
Ubb	0.943	Yes
Tle1	0.94	Yes
Hdac10	0.887	Yes
Ncor1	0.791	Yes
Maml2	0.788	Yes
Hey1	0.728	Yes
Gm8430	0.673	Yes
Rbx1	0.668	Yes
Cdk8	0.667	Yes

Figure 4.10: GSEA Enrichment plot for NOTCH1 INTRACELLULAR DOMAIN REGULATES TRANSCRIPTION and gene list of core enrichment. Enrichment score on the GSEA plot (top *y*-axis) indicates the degree of over-representation (in absolute value) for the provided pre-ranked gene set (differentially expressed genes – E14.5 PLET1$^+$ GOF versus E14.5 PLET1$^+$ WT) and the direction of their regulation (positive sign/up-regulated genes). The Pre-Ranked list metric (GSEA plot, bottom *y*-axis) provides information over the numeric ranking used to weight genes (in this case log fold change), while the barcode bars (GSEA plot, middle part) represent all gene hits in the particular pathway with genes contributing to core enrichment of the pathway listed in the table next to the Enrichment plot. Differentially expressed genes for an FDR ≤ 0.2 are colour-coded in red (* FDR ≤ 0.05, ** FDR ≤ 0.01).

To better characterise the phenotype of the E14.5 GOF samples, I observed the expression levels of markers associated with the early TEPC and mature TEC state, as well as, the cTEC and mTEC derived sublineages across all E12.5 WT and LOF, E14.5 WT and LOF and E14.5 GOF samples (shown in **Figure 4.11**). This revealed that cTEC-associated genes in Group B (**Figure 4.11**), including *Foxn1*, were up-regulated from E12.5 to E14.5 in the WT samples, but that the E14.5 GOF samples retained similar levels of these markers to those observed in the E12.5 progenitors. This suggested that they retained some similarity to the early TEPCs/early cTECs, and that expression of these differentiation markers was blocked as a result of enforced NOTCH signalling. From this, I hypothesised that the lack of *Foxn1* up-regulation might explain the lack of up-regulation of the other cTEC markers. Relevant to this is that *Foxn1* overexpression has previously been shown to result in down-regulation of NOTCH target genes in both mTEC and cTEC populations isolated from E17.5 *Foxn1*^{Cre}*;iFoxn1* mice (**O'Neill and Blackburn**, unpublished). These mice provide a model where *Foxn1* overexpression can be induced under the control of *Cre* (Bredenkamp et al., 2014b) (see also **Chapter 5**, **section 5.2.3** for more details). Supporting this idea, the ChIP-seq analysis of a FOXN1-tagged protein dataset (Žuklys et al., 2016) that I will present in **Chapter 5** of this thesis shows that FOXN1 directly negatively regulates *Rbpj*, as well as a number of known NOTCH target genes (*Fbxw7*, *Hey1*, *Hes6*). TEC differentiation may therefore be driven by a reciprocal inhibition between *Foxn1* and *Rbpj,* in which increasing levels of FOXN1 drive cTEC differentiation by down-regulating the response to NOTCH signalling, while high levels of NOTCH signalling are required in some TEPCs to protect the mTEC phenotype and possibly down-regulate *Foxn1*. The same pattern is also retained in one week old cTEC and mTEC populations (merged RNA-seq data from **similarly sorted individual and biological duplicates** obtained from **public repositories**; see **Chapter 2**, **section 2.2.1.4** for details).

Expression of genes related to both cTEC and mTEC differentiation in Group C (**Figure 4.11**) would remain high in all E14.5 samples with little variation among them, while they were much lower expressed in the E12.5 samples, suggesting a partial progression of the E14.5 PLET1^{+} GOF samples despite their high expression of *Foxn1* and *Plet1*. Additionally, early ontogeny lineage-restricted mTEC markers (such as

Krt5, Cld3 and Cl4; Group A) were substantially up-regulated in the GOF samples. Lastly, genes in Group D (such as *Fgfr2*, *Kitl* and *Pax9*) showed lower expression in the GOF samples than controls, despite being expressed highly in E12.5 TEPCs.

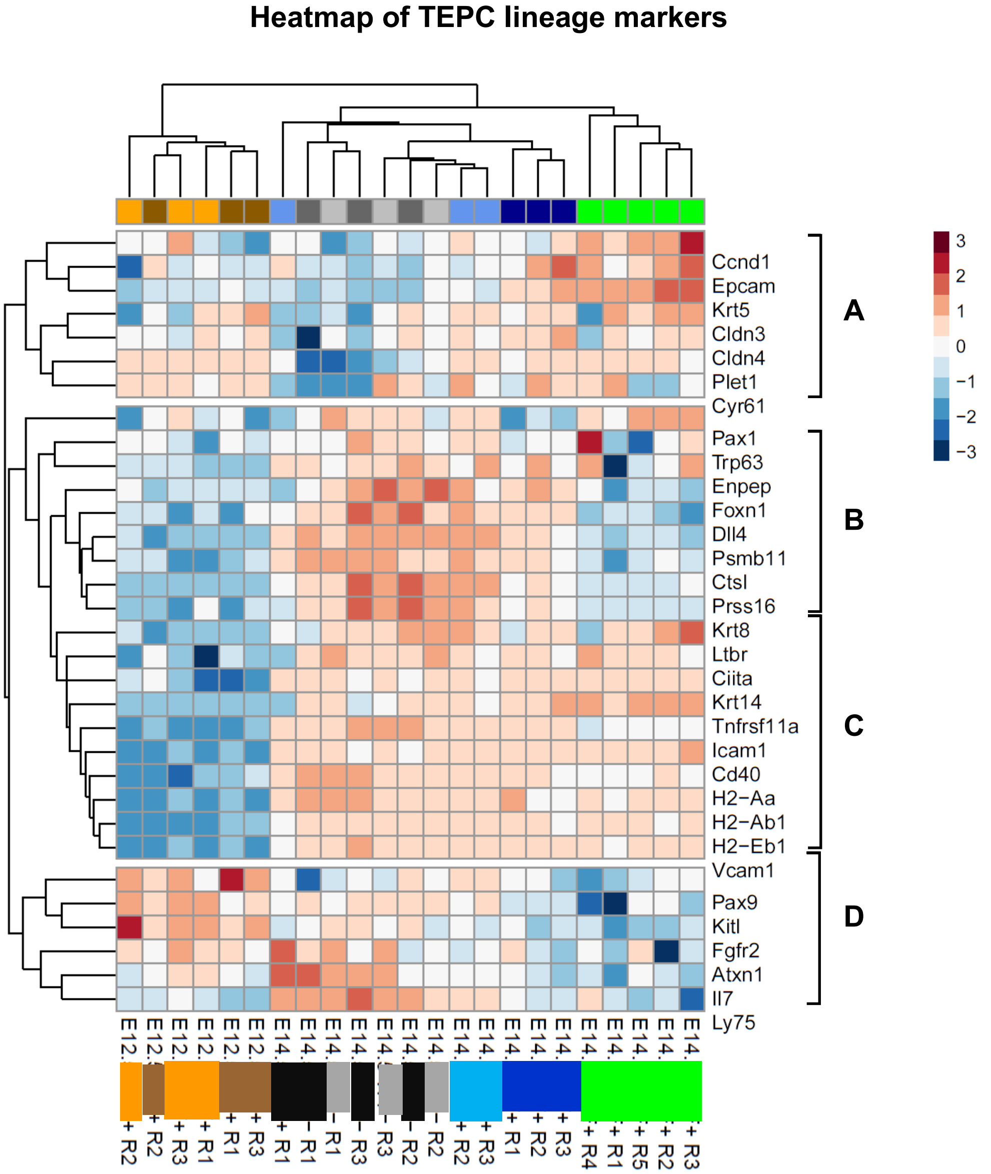


Figure 4.11: Heatmap of TEPC and TEC-specific lineage markers for NOTCH mutant and control samples. Row scaled values for a panel of TEPC and TEC markers among the E12.5 WT/LOF, E14.5 PLET1$^+$ WT/LOF, E14.5 PLET1$^-$ WT/LOF and E14.5 PLET1$^+$ GOF are shown to identify clusters [A-D] of genes that change similarly per group. Group **A**: early ontogeny lineage-restricted mTEC markers. Group **B**: FOXN1 targets/cTEC-like genes. Group **C**: cTEC and mTEC lineage differentiation genes. Group **D**: early ontogeny TEPC genes.

From these data, I have concluded that overexpression of NOTCH in TEPCs causes at least a partial block in cTEC lineage progression (ablation of PLET1$^-$ TECs) and assists but does not impose mTEC development, suggesting an additional role of NOTCH in TEPC maintenance.

Overall, experimental validation of NOTCH signalling activity in fetal TEPCs has verified that E10.5 TEPCs experience high levels of NOTCH activity, which, at a population level, progressively decreases concomitant with the onset of differentiation such that by E12.5 NOTCH signalling activity is lower in the TEPC population. These data have informed detailed functional testing of the role of NOTCH signalling in early TEC development by **Dong Liu** and **Dr Kathy O'Neill** in the Blackburn lab, and form part of a resulting manuscript (Liu et al., 2017 submitted). From these studies, we conclude that NOTCH signalling is necessary for specification of the mTEC compartment in a defined time window, prior to E16.5. Withal, these functional data are partially consistent with the model of extracellular signalling pathways regulation of early TEPCs that I established through bioinformatics analyses (**section 4.2.2**, **Figure 4.6**).

4.3 DISCUSSION

Nude mice, in which TEC differentiation is blocked prior to the onset of TEC differentiation, retain at least some bipotent TEPCs into adulthood. These cells remain in a stable state *in vivo* apparently indefinitely. Remarkably, release of the block in *Foxn1* expression in a single cell, is sufficient to direct formation of a fully functional thymus (Bleul et al., 2006; Jin et al., 2014). To improve the culture conditions and acquire the ability to maintain TEPCs *in vitro*, I investigated the signalling mechanisms active in the early pre-*Foxn1* state. Pathway enrichment analysis was performed between the undifferentiated and more differentiated TEPCs during the E10.5-E12.5 time-period, to identify signalling pathways active in the thymic epithelium prior to *Foxn1* expression. From this analysis, I generated an *in silico* model of signalling pathways activity in TEPCs just prior to the onset of differentiation that provides increased understanding of the signalling mechanisms supporting cells before their transition. Insights gained from this predicted signalling pathway analysis are discussed below.

4.3.1 A clear separation of TEC and parathyroid fate

The Sonic Hedgehog (SHH) signalling pathway is known to act as a control gate of the thymic-parathyroid fate, with its ablated or enforced activity to disrupt the patterning of the thymus-parathyroid primordium. Explicitly, absence of SHH (the main ligand of the pathway) in the endoderm of the 3rd pharyngeal pouch during organogenesis (thymus-parathyroid domain spreading), allows extension of the thymus into the parathyroid compartment (as demonstrated by Moore-Scott and Manley, 2005) suggesting that SHH signalling must be restricted in a particular region at a particular time point for thymus formation to take place. Alternatively, ectopic expression of SHH in early development induced TBX1 which blocked the expression of *Foxn1* and subsequently imposing of the thymic fate without though the cells to acquire a parathyroid fate (Bain et al., 2016; Reeh et al., 2014).

The clear separation from the parathyroid fate is evident in the RNA-seq dataset of the E10.5 and E12.5 TEPCs from the TEPC Developmental series, with the main receptors of the pathway (*Ptch1* and *Ptch2*) expressed only at a very low level in the PLET1$^+$ population analysed at E10.5, and being undetected at subsequent timepoints. This is consistent with silencing of the SHH intracellular or extracellular signalling in the early TEPCs to allow normal *Foxn1* expression and progression of the thymic fate (**Figure 4.6**). Since PLET1 is expressed throughout the parathyroid at E10.5 and E11.5, there is a minor possibility that few parathyroid cells are present in the RNA-seq samples. This seems not to be the case here though, at least by E11.5 onwards since none of the SHH involved genes are expressed at E11.5 or E12.5 TEPCs (*Ptch1/2*, *Gli1*, *Tbx1*). Additionally, *Shh* itself is fully absent from all datasets, consistent with the above observation.

4.3.2 Changes in TEPC identity could be reflecting changes in cell morphology

The mechanisms which orchestrate the diverse morphogenetic events in the thymus during early mouse ontogeny are not yet fully understood. Originating from the endoderm of the 3rd pharyngeal pouch at E10.5, the thymic anlage initially forms as essentially a polarised monolayer of columnar epithelial cells. This monolayer transits through a multilayered pseudostratified epithelial structure around E11.5 (which maintains an apico-basal polarity around a central lumen) before it forms the highly orchestrated three dimensional structure of the thymus where cells are no longer polarised (Itoi et al., 2001). During this transition, the thymus and the parathyroid primordia are clearly patterned by E11.5, while they detach sequentially from the pharynx by apoptotic cell death around E11.5-E11.75 (Gordon et al., 2004). In the end, both organs start migrating towards their final anatomical positions at E12.5 (summarised in Gordon and Manley, 2011).

Cell polarity is the asymmetric division of various components of a cell (cytoskeleton, plasma membrane or organelles) and it is used to maintain a barrier between cells or help cells to find their proper position for their normal function. The polarity of specialised structures can be established by morphogens. These include members of the BMP, WNT and SHH signalling pathways which can work cooperatively or in opposition to each other (Lee et al., 2001; Marcelle et al., 1997; Zhang et al., 2002) to sculpt tissues. In particular, *Bmp4*, *Fgf8* and *Shh* have been implicated with the patterning of the pharyngeal region, since their loss or impaired regulation results in hypoplastic or absent pharyngeal arches (Abu-Issa et al., 2002; Ahlgren and Bronner-Fraser, 1999; Bachiller, 2003; Frank et al., 2002; Ohnemus et al., 2002; Revest et al., 2001; Stottmann et al., 2001) which subsequently lead to malfunctions in the thymus formation due to disturbed pouch patterning or thymus-pharynx-parathyroid separation.

WNT5A is a known morphogen ascribed to the non-canonical WNT signalling pathway. Non-canonical WNT signalling branches include: a) the Planar Cell Polarity (PCP)/Convergent Extension (CE) pathway, b) the WNT/Ca2+ pathway and c) the WNT/FZD2-ROR2 pathway. Primarily, *Wnt5a* is an important regulator of embryonic development since its disruption leads to prenatal death in mice (Yamaguchi et al., 1999). The same study demonstrated that proximal distal outgrowth of several structures in vertebrate early development requires *Wnt5a* expression (cell cycling), with the outgrowing tissues of homozygous mice (at a late embryonic stage) to reveal gross morphological defects. Examples of *Wnt5a* regulation of cell polarity and directional cell migration exist in various cell types. During the mammalian palate development, this is orchestrated by the graded expression of *Wnt5a* and its receptor *Ror2* along the anteroposterior (AP) axis (He et al., 2008), and a potential mechanism of regulation could be by the asymmetric accumulation of a WNT-mediated receptor-actin-myosin polarity structure (W-RAMP) as demonstrated in a melanoma cell line study (Witze et al., 2008). In a more recent study, WNT5A expression was shown to contribute to AP polarity of epithelial cells through its asymmetric secretion (basolaterally) in polarised kidney epithelial cells in mice (Yamamoto et al., 2015). Additionally, interaction of SFRP proteins with WNT5A can regulate the apicobasal polarity and oriented cell division in the gut epithelium during embryonic mouse development (Matsuyama et al., 2009). Roles of *Wnt5a* in promotion of self-renewal in tissue stem cells have also been demonstrated in spermatogonial stem cells in mouse (Yeh et al., 2011). Taken together, WNT5A demonstrates a versatile role in the regulation of various processes such as self-renewal, differentiation, polarity and migration of the cells, while the outcome of its regulation depends on the provision of several receptors in the regulated cells.

In my analysis, *Wnt5a* demonstrates high expression levels at E10.5, while its expression gradually decreases over the E11.5 and E12.5 subsequent timepoints. As discussed before (**Chapter 3**), E10.5 TEPCs exist in an undifferentiated TEPC state, in which they *can* survive *in vitro* long-term upon *Foxn1* blockage. The high levels of *Wnt5a* expression in E10.5 TEPCs may indicate a potential role in the control of cells self-renewal. In that direction, *Wnt5a* was shown to maintain HSCs in a quiescent G0 state which increased their repopulating ability (Nemeth et al., 2007).

Moreover, E10.5 TEPCs and E11.5 TEPCs exhibit an apicobasal polarity while very few cells may be still polarised at E12.5 (Itoi et al., 2001). The timing of *Wnt5a* down-regulation after *Foxn1* initiation is consistent with their differentiation and their morphological switch from a mono layer of polarised cells that surrounds the thymic lumen to an apolar, complex and well-defined structure that characterises the thymus. Towards this direction, the expression pattern of the *Sfrp* genes among the stages of the TEPC Developmental series (shown in **Figure 4.7**) resembles the expression pattern of *Wnt5a*, with *Sfrp2* to down-regulate with statistical significance (FDR $\leq$ 0.05) when cells proceed towards the E12.5 timepoint. Thus, similar to the example in the gut epithelium (Matsuyama et al., 2009), *Wnt5a* may be interacting with these proteins to orchestrate cell polarity. In that scenario, epithelial cells would have their basal layer at the lumen surface and their apical layer would be extended away from the lumen during E11.5. Further down-regulation of *Wnt5a* at E12.5 could represent the loss of the apicobasal polarity for the majority of TEPCs.

Because *Wnt5a* comprises a secreted protein and a morphogen, existence of multiple available receptors (based on the bioinformatics data presented in this thesis) make it difficult to distinguish between the specific mechanisms that WNT5A may be controlling in early undifferentiated TEPCs. Nevertheless, to test if WNT5A may support self-renewal of the early undifferentiated TEPCs, provision of WNT5A in cell culture should be able to promote TEPCs survival even in the absence of feeder cells (mesenchyme) and/or it could aid in the formation of a pseudostratified structure since TEPCs survive better in a 3D structure. Additionally, conditional deletion of *Wnt5a* in TEPCs in conjunction with immunofluorescent analysis for tight junctions and adherent proteins would enable testing of the above hypothesis and establish a role for *Wnt5a* in the maintenance/control of TEPCs polarity and differentiation. If *Wnt5a* is related to the delayed separation of the thymus/parathyroid from each other and/or both from the pharynx, a migration assay could be performed to identify cell polarisation and movement through imagining *in vitro*. Finally, it is always possible that the surrounding mesenchyme or the incoming hematopoietic progenitors could also provide the WNT5A ligand and maintain the ongoing signalling required for different functions of TEPCs, as discussed in the next section.

4.3.3 Switching to a neighbouring cell-dependent fate

The role of the FGF signalling pathway (FGF) has been widely investigated in the early undifferentiated TEPC stages and proceeding thymus development (summarised in Gardiner et al., 2012). Previous research has highlighted FGF8 as necessary for normal thymus development, with high expression of *Fgf8* apparent in the distal posterior presumptive thymus region before the onset of *Foxn1*, and null or hypoplastic *Fgf8* mutations leading to thymic hypoplasia and athymia (Abu-Issa et al., 2002; Frank et al., 2002). FGFR2-IIIb comprises another essential factor for thymus development since its expression is required for TEC proliferation after E12.5 (Revest et al., 2001). FGFR2-IIIb expression escalates after *Foxn1* initiation (tested in thymic epithelium and skin samples) suggesting that *Fgfr2IIIb* is a direct target of FOXN1. *Fgfr2IIIb null* mutation results in reduced thymus size, but normal anatomical position, T-cell development and *Foxn1* expression (Revest et al., 2001). FGFR2-IIIb earliest point of detection is E13.5, approximately the time that TEC proliferation will reach to a block in *Fgfr2IIIb*$^{-/-lacZ}$. Unlike its ligands (FGF3, FGF7 and FGF10) that are expressed in the mesenchymal tissue surrounding the developing thymus, FGFR2-IIIb expression is limited in the thymic epithelium. Of these ligands, FGF10 is thought to be the main ligand required for normal thymus development, since the absence of FGF10 leads to athymia (Ohuchi et al., 2000).

In agreement with the literature, my analyses revealed high expression levels of *Fgf8* and *Fgfr1* in E10.5 TEPCs, with no further expression of these factors at E11.5 or E12.5. Of the ligands, *Fgf7* and *Fgf3* were essentially absent from TEPC at all stages analysed. Previous *in situ* hybridisation (ISH) analysis had indicated that *Fgf10* was not expressed in TEPCs, but in the neural crest cell (NCC)-derived thymic mesenchyme. However, through the more sensitive technique of RNA-seq, the analyses presented herein revealed moderate expression of *Fgf10* in TEPCs at E10.5 possibly explaining the greater effect of loss of FGF10 compared to FGF7 in thymus development. Potential contamination of the cell populations analysed by RNA-seq with surrounding neural crest cells is unlikely, since the cells were sorted using TEPC markers and since *bona fide* neural crest genes *FoxD3* and *Sox10* (Simoes-Costa and Bronner, 2015) were also not detected. Furthermore, our RNA-seq data have demonstrated expression of the *Fgfr2-IIIb* as early as E10.5, reestablishing the earliest detection point for *Fgfr2-IIIb* to exist prior to *Foxn1* initiation. In our data, *Fgfr2-IIIb*'s expression escalates in E11.5 and E12.5 agreeing with the suggested role of FOXN1 as a regulator of FGFR2-IIIb. Further evidence in support of this suggestion comes from the ChIP-seq reanalysis of a FOXN1-tagged protein dataset (Žuklys et al., 2016) presented in **Chapter 5** that identifies *Fgfr2-IIIb* as a highly confident direct target of FOXN1. Lastly, *Spry1* and *Spry2*, FGF feedback antagonists, whose expression is required for FGF inhibition maintenance at a specific location of the pouch and subsequent patterning formation (Gardiner et al., 2012), were found expressed in similar levels across the developmental points.

Up-regulation of FGFR2-IIIb after *Foxn1* initiation potentially allows for stronger activation of the FGF signalling pathway in TEPCs, signifying the moment when TEPCs become more reactive to FGFR2-IIIb ligands. The absence of most FGF ligands from TEPC at all stages analysed suggests that TEPCs depend on an appropriate supporting surrounding environment to provide the appropriate proliferation and survival signals. Down-regulation of *Fgf10*, the only FGFR2-IIIb ligand expressed in E10.5 TEPCs, potentially marks a complete transition of developing TEC to dependence on neighbouring cells for FGF signalling inputs.

IGF factors are also known to play an important role in thymus development (reviewed in Lee et al., 2010; Smith, 2010), with IGF1 increasing survival and differentiation of TECs and thymocytes by activating IGF-1R, an effect that is delivered through TECs (Chu et al., 2008). My analysis has indicated that both *Igf1* and *Igf2* mRNA levels decrease significantly from E10.5 to E12.5, with *Igf1* being expressed at extremely low levels at E12.5, suggesting decreased intracellular signalling activity. Similar to FGF signalling, IGF ligands are expressed at low levels at E12.5, but the receptor is still expressed at a high level, allowing activation of the pathway through ligands provided by neighbouring cells. Contrary to the decrease of the FGF ligands, that were absent from both cTEC and mTEC populations from one week old mice, both IGF factors were found to be almost uniquely expressed in the one week old mTEC population and was almost absent in cTECs, suggesting that a loss of the IGF ligands could potentially impact more the mTEC compartment. Lastly, IGF availability can be controlled by the presence of IGFBPs that can bind to the IGF ligands and prevent them from activating their receptor. Multiple *Igfbp* genes seem to be expressed in the different timepoints of the TEPC Developmental series at variable mRNA levels, with no obvious trend of up-regulation or down-regulation across these developmental stages. Therefore, it is difficult to evaluate their contribution in restricting pathway's activation.

In summary, investigation of the core enrichment genes involved in the FGF and IGF signalling pathways has suggested a switch of TEPCs from a partially cell autonomous system to a system depending on extracellular provision of ligands from the NCC-derived mesenchyme and potentially haematopoietic progenitor cells. The data suggest that this switch is driven by FOXN1 and permitted by the presence of FGFR2-IIIb and IGF1R. However, the impact of decreased IGF signalling activity in TEPCs may not equally influence both TEC lineages.

4.3.4 Establishing the mTEC lineage from a common TEPC progenitor

NOTCH regulators hold a dominant role in different stages of TEC and thymocyte development (Shah and Zuniga-Pflucker, 2014), however, contribution of NOTCH signalling during early ontogeny of the thymic epithelium has not yet been studied in detail. In brief, pathway enrichment analysis in **section 4.2.1** has identified NOTCH pathway (among other signalling pathways) as highly active in the undifferentiated E10.5 TEPCs but consistently down-regulated as TEPCs progress towards the E12.5 timepoint, suggesting a potential regulatory role for NOTCH signalling during fetal TEPC differentiation. Nevertheless, a closer look at NOTCH ligand *Dll4* and target genes *Hes1* and *Heyl* (whose expression increases when cells proceed towards the

E12.5 stage) made obvious that NOTCH pathway may be dynamically regulated in different developmental stages and possibly by different groups of cells. Thus, NOTCH was put under the microscope for further exploration.

In agreement with the decreased expression of *Rbpj* in E12.5 TEPCs, comparative bioinformatics analysis of NOTCH loss-of-function (LOF, *Rbpj* exon deletion under the control of the *Foxn1* promoter around E12) samples versus stage-matched controls demonstrated only a few expressional differences between these groups (**section 4.2.3.1**). Though few, differences between the E14.5 PLET1$^+$ LOF TEPCs and their controls were still apparent, while comparison of the E14.5 PLET1$^-$ LOF TEPCs to their controls did not reveal any differences (see **Figure 4.8**), demonstrating an effect (even subtle) only in the PLET1$^+$ TEPC population. The PLET1$^+$ compartment of E14.5 TEPCs constitutes undifferentiated TEC progenitors and potentially mTEC specific progenitors (mTEPCs), while the PLET1$^-$ compartment comprises of mostly cTEC-fated progenitors (cTEPCs), suggesting that it is the TEPCs or the mTEPCs rather than the cTEPCs that are impacted by NOTCH ablation.

Additionally, comparative analysis of the transcriptional profiles of the E12.5 vs E10.5 samples from the TEPC Developmental series (RNA-seq data from **biological triplicates per stage** provided by **Harsh Vaidya**, Blackburn lab; see also **Chapter 2**, **section 2.2.1**) in conjunction with comparative analysis of high-*Foxn1* and low-*Foxn1* groups from the *Foxn1* Allelic series (RNA-seq data from **singular biological samples** provided by **Stephanie Tetelin**, Blackburn lab; see also **Chapter 2**, **section 2.2.1**) presented in **Chapter 3** (**section 3.2.2**) has identified *Rbpj* among other genes as FOXN1-dependent that down-regulates with *Foxn1* increase. This finding suggested that *Rbpj* down-regulation may be regulated (directly or indirectly) by FOXN1, thus ablation of *Rbpj* expression after *Foxn1* initiation would be expected to have minimum impact.

Experimental evidence from **Dong Liu** (Blackburn lab, University of Edinburgh) demonstrated that inhibition of NOTCH signalling at E10.5 in TEPCs led to the complete loss of the mTEC lineage, in total agreement with the above predictions, and also verified NOTCH signalling as a potent regulator of mTECs specification. Conversely, overexpression of NOTCH intracellular domain in all TEPCs blocked

TEC lineage differentiation, evidenced by retention of PLET1 staining (**Dong Liu**) and lack of differentiation markers (**Figure 4.11**), but did not force the differentiation of mTECs (**Figure 4.11**), suggesting roles for NOTCH in both mTEC specification and TEPC maintenance.

A potential regulatory mechanism between FOXN1 and NOTCH signalling is extensively described in **Chapter 5** (**section 5.3.3**) while how this interplay fits with the revised TEPC differentiation model (presented in **Chapter 3**, **section 3.3.1**) is devised in **Chapter 7** (**section 7.1.4**).

In summary, the bioinformatics analysis presented herein (**section 4.2.3**), in conjunction with extensive experimental investigation and validation of the NOTCH signalling pathway in the Blackburn lab by **Dong Liu**, has concluded that “a NOTCH gate controls differentiation of thymic epithelial progenitors and medullary thymic epithelial lineage specification” (Liu et al., 2017, submitted).

4.4 SUMMARY

In this chapter, I have presented a comparative bioinformatics analysis in conjunction with pathway enrichment analysis of RNA-seq data obtained from TEPCs at timepoints just prior to, during and immediately following the initiation of the TEC differentiation programme. This integrative analysis was used to determine which major signalling pathways are highly active in fetal TEPCs, focusing on those pathways whose activity is decreasing while they are transiting towards differentiation.

The identified pathways indicate a) the separation of the thymus from the parathyroid fate (**section 4.3.1**), b) potential roles for WNT5A in cell renewal and cell polarity (maintenance of TEPCs and/or guidance of TEPCs into the 3D complex architecture of the functional thymus, **section 4.3.2**) and c) a shift from a TEPC autocrine niche to a TEPC niche dependent on neighbouring cells (**section 4.3.3**). Moreover, computational and experimental validation of the importance of NOTCH activity in early TEPCs (ablation of NOTCH signalling), highlighted NOTCH as a potent regulator of mTEC specification and maintenance of the TEPC state in a short but well-defined time window. Additionally, overexpression of NOTCH signalling suggested that E12.5 TEPCs may still exist in a fairly interchangeable state which could be fine-tuned or potentially pushed back to a less differentiated state (a concept also discussed in **Chapter 3**, **section 3.3.1**) by an altered signalling reciprocity involving (in this case) FOXN1 and NOTCH.

Within the *in silico* model predicted in **Chapter 4**, the master thymus regulator, FOXN1, seems to control the signalling pathways that dictate the TEPC proliferation and differentiation state. This is evidenced by FOXN1 directly targeting FGFR2-IIIb, that controls TEPC expansion, alongside with *Rbpj* and other regulators involved in the NOTCH pathway, that are involved in the specification of the mTEC lineage (the role of FOXN1 and its regulatory landscape are discussed extensively in **Chapter 5**). Overall, this analysis has improved the current picture of the signalling mechanisms that govern maintenance and specification of TEPCs prior to their sublineage progression orchestrated by FOXN1.

Chapter 5

FOXN1: A master regulator of TEC differentiation

Chapter 5

FOXN1: A master regulator of TEC differentiation

5.1 PRÉCIS

FOXN1 comprises the master transcriptional regulator of TEC differentiation, with continuous expression of *Foxn1* also being required for maintenance of the thymic epithelium (Chen et al., 2009; Cheng et al., 2010; Nowell et al., 2011). Evidence for its significance is that absence of functional FOXN1 protein arrests differentiation of thymic epithelial cells (TECs), blocks initial formation of the organ, and prevents colonisation of thymic primordium by hematopoietic progenitors (T-cell precursors). This early block can be reversed by reestablishing *Foxn1* expression in TECs, and this alone then leads to normal thymus development (Bleul et al., 2006). In a more extreme example, "forced" expression of FOXN1 in a thymus-unrelated cell population, mouse embryonic fibroblasts (MEFs), is capable of reprogramming these cells into functional TECs ("induced" TECs or iTECs), further demonstrating the extraordinary potency of this factor (Bredenkamp et al., 2014a).

Despite the core role that FOXN1 plays in the thymus and 20 years of research since FOXN1 (*nude* gene product) was first identified (Nehls et al., 1994, 1996), the full picture of the functional network around this regulator is incomplete. In **Chapter 5,** I attempt to better understand how FOXN1 imposes its dominant role in the thymic system, by identifying direct FOXN1 targets in early TEPCs and investigating the pathways that FOXN1 controls through them, via a comparative study between fetal and newborn thymic epithelial populations.

5.2 RESULTS

5.2.1 Prediction of candidate direct FOXN1 targets in mouse fetal TEPCs

In this chapter, I will describe the integrative analysis undertaken to identify candidate direct FOXN1 targets in mouse fetal TEC progenitors (TEPCs). A brief summary of the above analysis is provided here. Starting from the already identified list of FOXN1-dependent genes from the fetal TEPC RNA-seq series (described extensively in **Chapter 3**, **section 3.2.1**), I will first subselect the genes which show binding evidence for FOXN1 in their distal promoter sites (peaks located within -5kb and +3kb of the gene's transcription start site) using ChIP-seq data available from 1 week old (1w) TECs. Next, I will retain only the genes that also show evidence for an active promoter or enhancer mark in TEPCs using histone modification ChIP-seq data from E12.5 TEPCs that overlap with the 1w TEC FOXN1 enriched peaks. This final list of genes will define the (first-described) high-confidence candidate direct FOXN1 targets in fetal TEPCs.

5.2.1.1 FOXN1-dependent genes in fetal TEPCs

To identify a list of FOXN1-dependent genes, the differentially expressed genes between E12.5 and E10.5 timepoints ($|FC| \geq 1.5$ and $FDR \leq 0.2$) from the TEPC Developmental series (RNA-seq data from **biological triplicates per stage** provided by **Harsh Vaidya**, Blackburn lab; see also **Chapter 2**, **section 2.2.1**) were integrated with the differentially expressed genes (pairwise comparisons between *Foxn1*^{high} vs *Foxn1*^{low} phenotypic sample groups) from the *Foxn1* Allelic series (RNA-seq data from **singular biological samples** provided by **Stephanie Tetelin**, Blackburn lab; see also **Chapter 2**, **section 2.2.1**) and the overlap of both differentially expressed gene lists was taken (details on data integration are given in **Chapter 3**, **section 3.2.1**). As described previously, the TEPC Developmental series dataset depicts gene expression profiles of E10.5, E11.5 and E12.5 TEPCs during mouse development. These timepoints relate to the developmental stages at which: *Foxn1* expression is majorly absent, *Foxn1* expression has been initiated and *Foxn1* expression has been established. The gene expression changes that occur among these stages could be assigned to either FOXN1 expression itself or FOXN1-independent developmental progression. The *Foxn1* Allelic series dataset shows, specifically at the latest timepoint of the TEPC Developmental series (E12.5), how variation of *Foxn1* levels alone impacts on the expressional profile of TEPCs. **Chapter 3 section 3.2.1** has demonstrated the dominant effect of *Foxn1* dosage on cell identity, with *Foxn1* expression level influencing TEC developmental progression, suggesting that many of the changes occurring in the early TEPC development might be FOXN1-dependent. In order to focus on direct FOXN1 targets, I therefore focused the analysis on those genes that are dependent on FOXN1 (independently of the developmental stage) to limit the span of the gene inventory search.

In particular, a total of 15,321 genes were expressed in at least in one of the E10.5 and E12.5 TEPC sample triplicates from the TEPC Developmental series based on a log2-transformed counts per million (CPM) cut-off value of zero, which was selected based on samples' expression density plot, as described in **Chapter 2**, **section 2.3.3.1**. As described before (**Chapter 3**, **section 3.2.1**), the limma package from Bioconductor was used to calculate differential gene expression between E10.5 and E12.5 timepoints and to assign statistical significance to the gene changes. From the 15,321 genes expressed in E10.5 or E12.5 TEPC samples, only 1,650 (980 up-regulated and 670 down-regulated) were differentially expressed between the E10.5 and E12.5 triplicates with $|FC| \geq 1.5$ and $FDR \leq 0.2$. These 1,650 genes were selected for further analysis (**Figure 5.1**, step 1). Pairwise comparison analysis was performed between a *Foxn1*^{high} and a *Foxn1*^{low} phenotypic group, consisting of the *Foxn1*$^{+/+}$ (wild type; WT), *Foxn1*$^{+/-}$ (Het) and *Foxn1*$^{R/+}$ (R/+) samples, and the *Foxn1*$^{R/-}$ (R/-) and *Foxn1*$^{nu/nu}$ (Nude) samples from the *Foxn1* Allelic series dataset respectively. This comparison identified genes that are highly dependent on FOXN1 (for details see **Chapter 3, section 3.2.1**). The latter filtering for FOXN1-dependent genes decreased the number of differentially expressed genes from 1,650 to 850 genes (586 up-regulated and 264 down-regulated) that were fully dependent on FOXN1 (**Figure 5.1**, step 2). These 850 genes change consistently between the pre- (E10.5) and post-FOXN1 (E12.5) TEPC developmental stages as defined by a $|FC| \geq 1.5$ and a $FDR \leq 0.2$ criteria, however, their change can be most likely explained by direct dependence on FOXN1, since they also change consistently between the *Foxn1*^{high} and a *Foxn1*^{low} phenotypic groups, independently of the developmental stage of TEPCs.

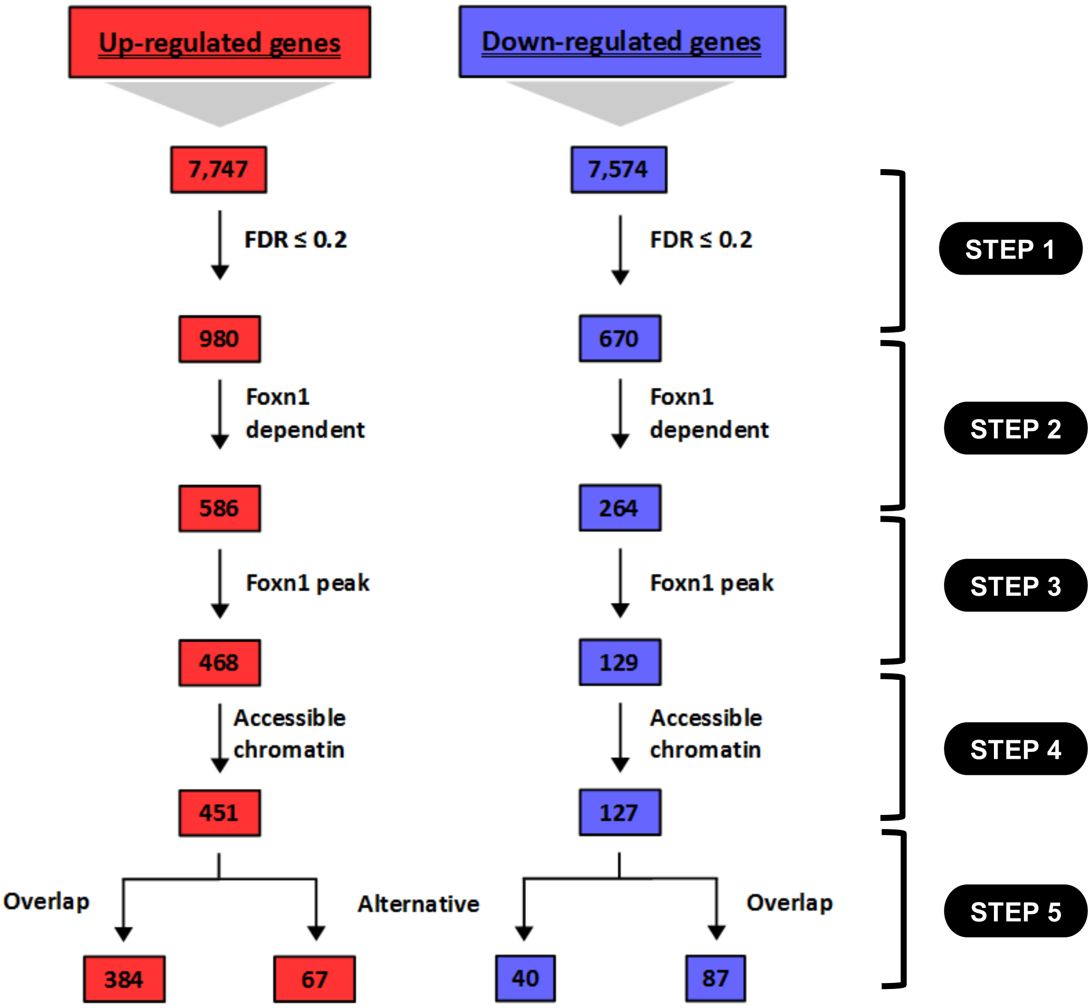

Up-regulated genes
Down-regulated genes
7,747
7,574
FDR ≤ 0.2
FDR ≤ 0.2
STEP 1
980
670
Foxn1 dependent
Foxn1 dependent
STEP 2
586
264
Foxn1 peak
Foxn1 peak
STEP 3
468
129
Accessible chromatin
Accessible chromatin
STEP 4
451
127
Overlap
Alternative
Overlap
STEP 5
384
67
40
87

Figure 5.1: Overview of the integrative analysis for identification of FOXN1 target candidates in TEPCs. The total number of up-regulated (in red) and down-regulated (in blue) genes between the E10.5 and E12.5 timepoints from the TEPC Developmental series dataset is shown in the top of this schema, with step 1 retaining only differentially expressed genes with FDR ≤ 0.2. In step 2, genes that are not dependent on FOXN1 are excluded from further analysis, while in step 3 genes are filtered on the basis of having a FOXN1 binding peak in their distal promoter area (-5kb to +3kb from the TSS of the gene). Step 4 shows how many of these genes have an accessible promoter or enhancer region with either an H3K4me3 or an H3K27ac mark, while step 5 shows how many of these two accessibility marks co-reside with the FOXN1 peaks (or simply mark alternative sites in the distal, TSS of the gene).

5.2.1.2 Direct binding events of FOXN1 in newborn TECs

Until recently, lack of suitable antibodies meant that data on direct FOXN1 binding to the promoter/enhancer regions of expressed genes were not available. Recently, a FOXN1-tag ChIP-seq dataset became available (Žuklys et al., 2016), which identified direct FOXN1 targets in newborn mouse cTECs.

In their study, Žuklys et al. express a FOXN1 protein, tagged with an octapeptide (for details of the design see cited paper or **Chapter 2**, **section 2.2.2.1**), under the *Foxn1* promoter of *nude* mice (*Foxn1*$^{nu/nu}$) using a bacterial artificial chromosome (BAC). Mice homozygous for the BAC (*Foxn1*$^{wt*/wt*}$) demonstrated a normal coat and thymus architecture, as well as normal *Foxn1* expression levels, but a mildly reduced cellularity which was still capable of supporting thymocyte development. Chromatin immunoprecipitation (ChIP) was performed in isolated thymic lobes from homozygous mice and DNA samples were pooled to create two replicates that were sent for sequencing. After sequencing, contaminating adapters were removed using Trimmomatic (Bolger et al., 2014), alignment was performed with BWA (pre-alignment; Li and Durbin, 2009) and Stampy (Lunter and Goodson, 2011) against the mouse genome assembly (mm10) and MACS2 (Zhang et al., 2008) in combination with the IDR pipeline (https://github.com/nboley/idr) was used to call confident FOXN1 peaks. 9,012 peaks passed the IDR threshold, of which a third marked the distal upstream TSS region of genes and were enriched for H3K4me3 marks. Furthermore, *de novo* motif discovery using the aforementioned IDR-selected peaks predicted the canonical (5'-GACGC-3') FOXN1 binding motif and an alternative motif (5'-GAAGC-3') which were both over-represented in the DNA sequences under the peaks.

5.2.1.3 Inferring candidate direct FOXN1 targets in fetal TEPCs from FOXN1 binding events in newborn TECs

Since no other ChIP-seq datasets were available for FOXN1 in fetal TEPCs, I used the FOXN1-tagged ChIP-seq dataset in newborn TECs (Žuklys et al., 2016) to obtain a list of FOXN1 binding sites, which I then further integrated with the list of FOXN1-

dependent genes in TEPCs (from the differential expression analyses described in **section 5.2.1.1)**, to infer potential FOXN1 gene regulatory events in fetal TEPCs. In more detail. I obtained the raw FASTQ files from the FOXN1-tagged ChIP-seq data from newborn TECs through the GEO repository (Barrett et al., 2013) and reanalysed them (as described in **Chapter 2**, **section 2.3.1 and section 2.3.4**) to provide a list of FOXN1 binding sites. In general, the ChIP-seq reanalysis steps were similar to the Žuklys et al. ChIP-seq pipeline, including adapter removal with Trimmomatic, peak calling with MACS2 and identification of confident peaks with the IDR pipeline. However, I used Bowtie2 (Langmead and Salzberg, 2012) for read mapping in the ChIP-seq reanalysis instead of the BWA-Stampy combination that was used by Žuklys et al., so that the same alignment tool was used for the analyses of the ChIP-seq datasets described in my thesis. A total of 4,858 IDR confident peaks (IDR ≤ 0.05) resulted from the FOXN1-tagged protein ChIP-seq reanalysis. As also shown by Žuklys et al., FOXN1 confident peaks (IDR ≤ 0.05) were found to mostly bind at the promoter region of genes and also at some distal intergenic regions (to a lesser extent); these intergenic regions could represent enhancer regions across the genome (**Figure 5.2**).

To infer candidate direct target genes for FOXN1 in the fetal TEPCs, I integrated the 4,858 FOXN1 confident peaks (IDR ≤ 0.05) from the FOXN1 ChIP-seq reanalysis in newborn TECs based on their genomic location (in reference to mouse genome assembly; version GRCm38.p5 from GENCODE – release M12) with the FOXN1-dependent genes in TEPCs from the TEPC Developmental series and the *Foxn1* Allelic series RNA-seq dataset analyses (identification of these genes is described extensively in **section 5.2.1.1**, see also **Figure 5.1**, step 1-2). For this integration, only TEPC FOXN1-dependent genes with a FOXN1 confident peak around their distal TSS site [5 kilobases (kb) upstream and 3kb downstream of genes TSSs] were chosen as potential FOXN1 direct targets in TEPCs. A subset of 597 FOXN1-dependent genes demonstrated FOXN1 binding evidence at their distal promoter region (-5kb, +3kb), with 468 of them (~78.4%) to be up-regulated in association with increase in *Foxn1* expression (**Figure 5.1**, step 3), suggesting a predominant role for FOXN1 as a transcription activator in TEPCs. This conclusion was consistent with findings showing a prevailing role of FOXN1 in genes up-regulation in newborn cTECs (Žuklys et al., 2016).

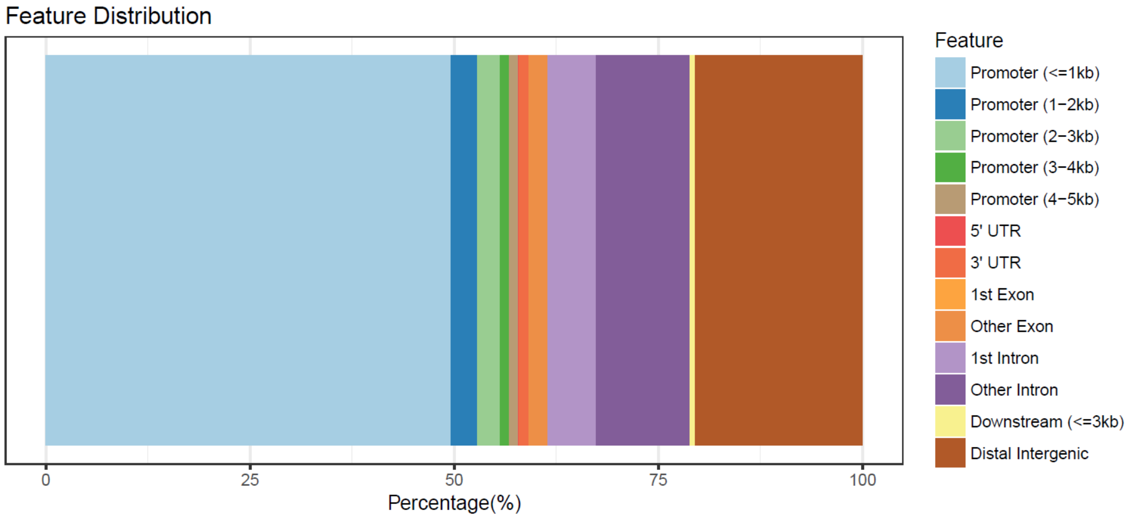


Figure 5.2: FOXN1 ChIP-seq peaks distribution in the mouse reference genome assembly gene regions. The 4,858 FOXN1 confident peaks (IDR ≤ 0.05) from the FOXN1 ChIP-seq re-analysis in newborn TECs were annotated in terms of genomic features with the *peakAnno* function from ChIPseeker package in R and plotted in a barplot format with *plotAnnoBar* function from the same package. Approximately 50% of the peaks reside in the proximal promoter region (light blue section) of the reference genome, with the second biggest percentage of peaks to fall into distal intergenic regions (brown section).

5.2.1.4 Insights in the TEPC chromatin accessibility landscape based on promoter and enhancer histone modification marks in E12.5 TEPCs

The FOXN1 binding landscape marked by the FOXN1-tagged ChIP-seq dataset is representative of a mixed TEC population in newborn mice. When a transcription factor binds at a promoter or enhancer region of a gene and regulates its expression in a specific cell type, a fair assumption is that it will probably regulate the same gene in another cell type if the gene is expressed in the new cell type and the promoter or enhancer region of this gene is also accessible and active.

To assess the chromatin regulatory landscape of TEPCs, histone modification data were generated from E12.5 fetal TEPCs [Vaidya and Blackburn (unpublished)]. These datasets comprise ChIP-seq for the H3K4me3 and H3K27ac chromatin modifications (ChIP-seq data from **biological duplicates per mark** provided by **Harsh Vaidya**, Blackburn lab; see also **Chapter 2**, **section 2.2.3.1**): the H3K4me3 modification mainly marks active promoters (Heintzman et al., 2007), while the H3K27ac modification marks mostly accessible and active enhancers (Creyghton et al., 2010; Heintzman et al., 2009; Rada-Iglesias et al., 2011; Zentner et al., 2011). Combined, these histone modification marks provide a robust readout of accessible and active promoter and enhancer regions of TEPCs on day E12.5 in embryonic development. By making use of the above histone modification datasets, TEPC genes with accessible and active promoters or enhancers in their distal promoter regions in the E12.5 TEPC average population landscape can be identified.

The average binding profile of the H3K4me3 and H3K27ac marks in TEPCs was thus plotted for all regions marked by the 4,858 FOXN1 confident peaks (IDR ≤ 0.05) from the FOXN1 ChIP-seq reanalysis in newborn TECs, to broadly assess accessibility of the regions located under the FOXN1 IDR peaks in TEPCs, by looking at the modification profiles of those regions in TEPCs (**Figure 5.3**). In more detail, **Figure 5.3** depicts genomic region profiles (top section) and respective heatmaps (bottom section) of the FOXN1 ChIP-seq peaks in newborn TECs, and the H3K4me3 and H3K27ac histone modification marks in fetal TEPCs for all the regions defined by the FOXN1 IDR peaks in newborn TECs, centred in the middle of the FOXN1 IDR peak and extended by ± 2 kb from the centre of the peak. The genomic profiles represent cumulatively the peaks enrichment scores over a set of genomic regions (the FOXN1 IDR peaks), while heatmaps show for each FOXN1 IDR peak-defined region the enrichment score of selected datasets (FOXN1 peaks from newborn TECs, and H3K4me3 and H3K27ac from fetal TEPCs) per row (Ramírez et al., 2016). As expected, distributions of the FOXN1 peaks are centred in the middle of the IDR FOXN1 peaks, while the histone modification marks show a binomial distribution over the IDR FOXN1 defined regions with a depletion in the centre of the IDR FOXN1 peaks. Transcription factor binding sites (TFBS) reside in the open chromatin region between flanking nucleosomes that carry these histone modification marks (Chai et al., 2013), thus this depletion signature is expected in IDR FOXN1 binding site. Overall, samples H3K4me3 (A) and H3K27ac (B) demonstrated the strongest signals (fold-enrichment versus control) around the FOXN1 peak centres, marking approximately two thirds of the FOXN1 IDR peaks. This analysis therefore suggested that a common landscape might exist for FOXN1 targets in fetal TEPC and newborn TEC samples. Because replicate samples for H3K4me3 and H3K27ac revealed lower signal intensity for the FOXN1 peaks marked by their own replicates (H3K4me3 (B) and H3K27ac (A); **Figure 5.3**) IDR analysis was not performed for these samples. Instead significant peaks were determined for each replicate for a 0.05 (H3K4me3 mark; narrow peaks) and 0.1 (H3K27ac mark; broad peak) q-value threshold (MACS2) and these significant peaks were mapped to genes from the GRCm38.p5 (mm10) Mouse assembly.

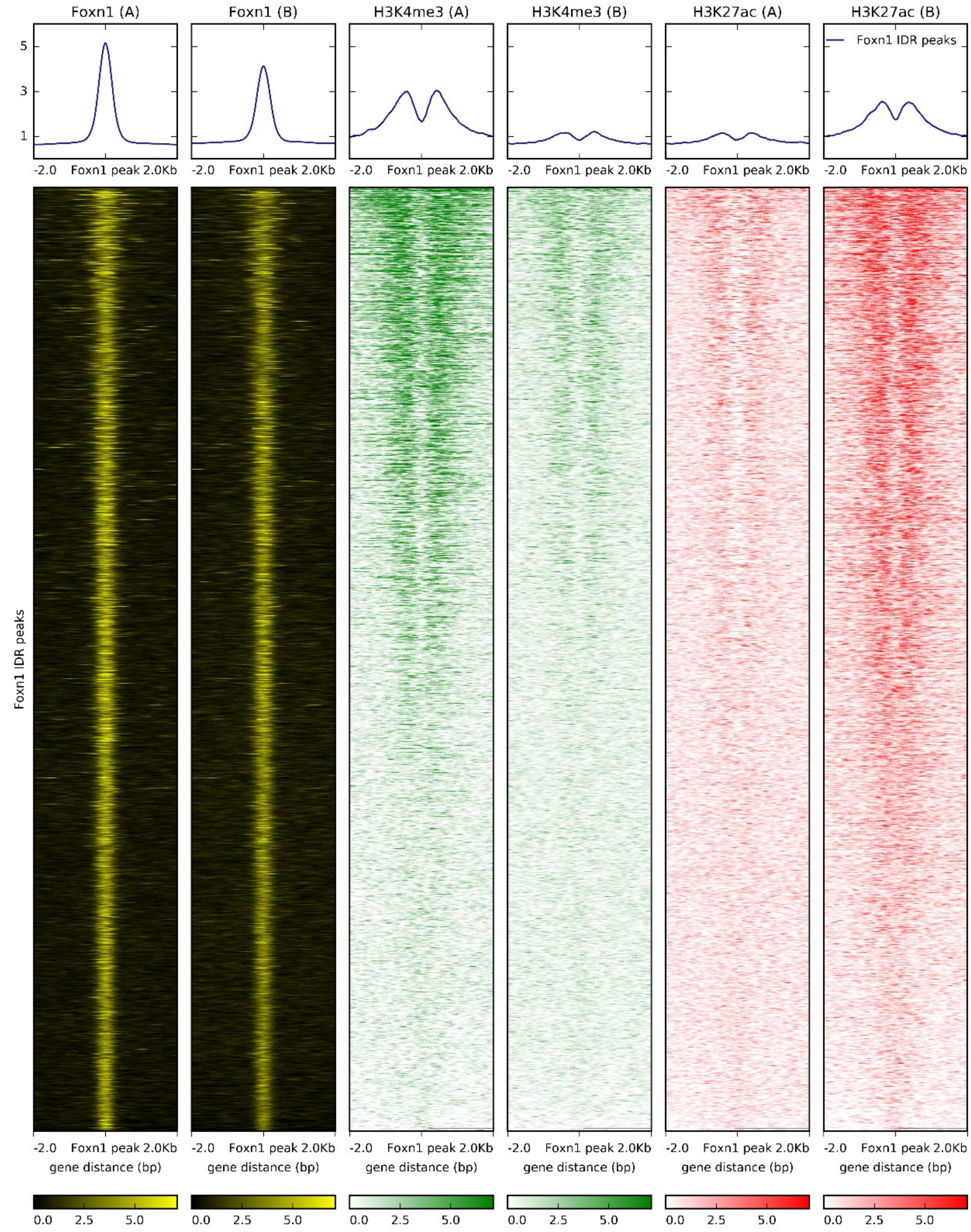


Figure 5.3: Genomic region profile plots and heatmaps over the FOXN1 IDR peak regions in newborn TECs for the FOXN1 and histone modification ChIP-seq datasets. Total signal profiles (top section) over the FOXN1 IDR peak-defined regions [± 2kb from centre of the FOXN1 peak] show cumulatively enrichment scores for FOXN1 peaks from 1 week old TECs (FOXN1 biological replicates A and B: FOXN1 (A) and FOXN1 (B)), and for H3K4me3 and H3K27ac peaks for E12.5 TEPCs (H3K4me3 biological replicates A and B; H3K4me3 (A) and H3K4me3 (B), H3K27ac biological replicate A and B; H3K27ac (A) and H3K27ac (B)). Heatmap plots (bottom section) show for each FOXN1 IDR peak-

defined region from 1 week old TECs the enrichment score for both FOXN1 replicates from 1 week old TECs (shown in gold-black colour scale) and for both replicates of each histone modification mark from E12.5 TEPCs: H3K4me3, and H3K27ac (shown in green-white, blue-white and red-white colour scales respectively). Heatmaps have been sorted in descending order based on the mean value per region.

5.2.1.5 Prediction of high confidence candidate direct FOXN1 targets in fetal TEPCs from overlapping FOXN1 binding sites in newborn TECs with accessibility histone marks in fetal TEPCs

As shown in **section 5.2.1.3**, candidate direct FOXN1 targets were inferred based on the proximity (-5kb, +3kb) of the promoters of FOXN1-dependent genes in TEPCs to the FOXN1 confident peaks (IDR ≤ 0.05) from the FOXN1-tagged ChIP-seq dataset in newborn TECs (see **Figure 5.1**, steps 1-2 and 3). This integrative analysis identified 597 genes as candidate direct FOXN1 targets. When the significant peaks of the histone modification marks detected in E12.5 TEPCs (q-value for H3K4me3 ≤ 0.05 and for H3K27ac ≤ 0.1; see **section 5.2.1.4**) were mapped to genes, the vast majority of the 578 candidate direct FOXN1 targets in TEPCs (451 up-regulated and 127 down-regulated, as shown **Figure 5.1**, step 4) possessed an H3K4me3 and/or an H3K27ac mark in their distal promoter area (-5kb, +3kb) as expected, since the list comprised TEPC expressed genes.

Although the 578 candidate direct FOXN1 targets in TEPCs exhibited a FOXN1 binding peak from newborn TECs as well as an H3K4me3 peak and/or an H3K27ac peak from fetal TEPCs in the region (-5kb, +3kb) around their TSS, these peaks did not necessarily coincide. When the H3K4me3 and H3K27ac peaks were overlaid with the distal FOXN1 binding peaks, 107 (67 up-regulated and 40 down-regulated) genes exhibited marks at a different location from the FOXN1 binding site (**Figure 5.1**, step 5). This suggested that the non-overlapping FOXN1 peaks from the newborn TEC population were not located in an open chromatin area in TEPCs, and that they could not be regulating the particular genes via direct binding. The remaining genes in which the FOXN1 and H3K4me3 and/or FOXN1 and H3K27ac peaks overlapped included: 269 (231 up-regulated, 38 down-regulated) with FOXN1 overlapping with both histone modification marks, 146 (107 up-regulated, 39 down-regulated) with FOXN1 overlapping only with the H3K4me3 mark and 56 (46 up-regulated, 10 down-regulated) with FOXN1 overlapping only with the H3K27ac mark (**Figure 5.4**). All 471 genes with one or more FOXN1 peaks from newborn TECs overlapping with one or both histone modification marks from fetal TEPCs are present in my most confident candidate list for FOXN1 direct targets in fetal TEPCs. Among these candidates were the widely known FOXN1 targets, *Ccl25* and *Dll4*, which were shown to be up-regulated after transient transfection of mouse embryonic stem (ES) cells with *Foxn1* by QRT-PCR analysis (Nowell et al., 2011), as well as the recently identified FOXN1 direct targets in 1 week old cTECs, *Psmb11* and *Cd83*, whose expression was lost or reduced in HEK293 cells expressing FOXN1 if their FOXN1 binding sites located in genes near promoter were mutated (Žuklys et al., 2016) (**Figure 5.4**). Further evaluation of these targets, including comparative analysis versus FOXN1 targets specifically predicted in newborn cTECs, and pathway enrichment analysis between the E10.5 and E12.5 TEPCs is described in **sections 5.2.2** and **5.2.3** below.

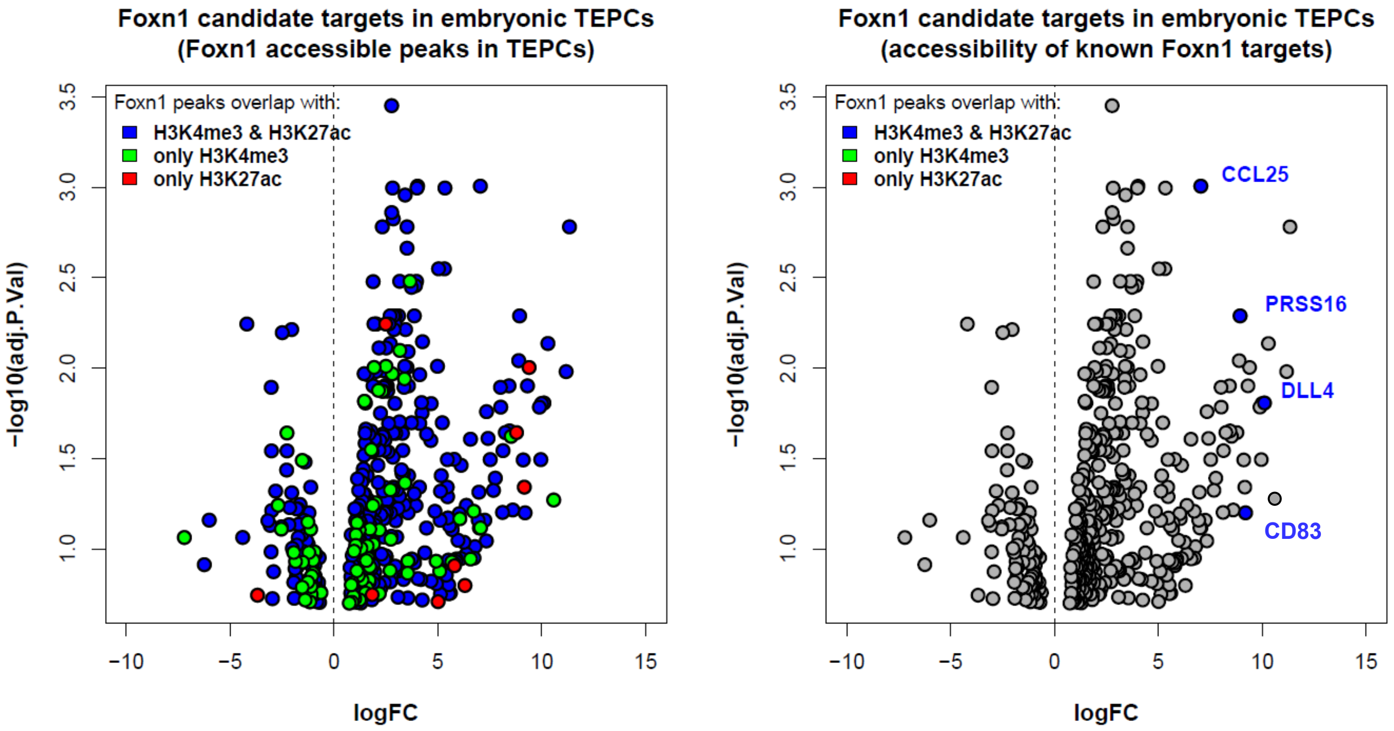


Figure 5.4: Volcano plots of the FOXN1 candidate targets in fetal TEPCs. Each point represents one of the 471 FOXN1 candidate targets in fetal TEPCs that have an overlapping FOXN1 binding peak from newborn TECs with a histone modification mark from fetal TEPCs around gene's distal promoter region (-5kb to +3kb from gene's TSS). In the left graph, genes with FOXN1 peaks overlapping with **A**) both H3K4me3 and H3K27ac marks are shown in blue, **B**) only an H3K4me3 mark are shown in green and **C**) only an H3K27ac mark are shown in red. Location of FOXN1 known targets is shown in the right graph and gene names have been coloured based on the FOXN1 peak overlap. The *x*-axis shows the log2-transformed fold changes as calculated from the differential expression analysis between E10.5 and E12.5 timepoints from the TEPC Developmental series (**Chapter 3**, **section 3.2.2**) with limma package in R. The *y*-axis depicts the adjusted *p*-value (here the false discovery rate), with a threshold of 0.7 in the –log10 scale (or a threshold of 0.2 in the log2 scale) to define confident gene changes (**Chapter 3**, **section 3.2.2**).

5.2.2 Comparative analysis and insights on the highly confident direct candidate FOXN1 targets in fetal TEPCs

In their study, Žuklys et al. (2016) integrated the IDR FOXN1 peaks from 1 week old (newborn) TECs with transcriptomic data depicting gene changes in 1 week old cTECs and 1 week old mTECs between *Foxn1*$^{wt*/wt*}$ and *Foxn1*$^{wt*/-}$ mice. As previously described, the *Foxn1*^{wt*} allele expresses the FOXN1 protein tagged with a flagged peptide, and mice homozygous for this allele (*Foxn1*$^{wt*/wt*}$) demonstrated an overall normal phenotype and *Foxn1* expression levels, with a mildly reduced cellularity which was still capable of supporting thymocyte development (Žuklys et al., 2016). In comparison, the heterozygous mice for the *Foxn1*^{wt*} allele (*Foxn1*$^{wt*/-}$) demonstrated a thymus of significantly reduced size, decreased *Foxn1* expression, disorganised corticomedullary segregation and multiple large cysts. A large number of genes were found to be differentially expressed (DE) from this analysis in newborn cTECs (8,378 DE genes) and mTECs (11,690 DE genes). Furthermore, to get a more confident set of genes that are regulated by FOXN1 in newborn cTECs, Žuklys et al. designed triple mutant mice (*Psmb11-rtTA::tetO-Cre::Foxn1*$^{7,8loxP/loxP}$; designated *iFoxn1*$^{\Delta 7,8}$), in which induction with doxycycline (Dox) drives a cTEC-specific deletion of exon 7 and 8 in the *Foxn1* locus. Prior to Dox exposure, mice demonstrated an overall normal thymus phenotype, TEC architecture and intrathymic T-cell development. However, the thymi of Dox-treated mice demonstrated significantly reduced *Foxn1* levels, decreased cellularity and specific reduction of thymocytes subpopulations (Lin$^-$CD44$^+$CD25$^+$CD4$^-$CD8$^-$ and Lin$^-$CD44$^-$CD25$^+$CD4$^-$CD8$^-$) by day 3 after treatment. Differential expression analysis of the cTEC transcriptomic datasets from 1 week old *iFoxn1*$^{\Delta 7,8}$ mice prior to, and 3 days after, Dox treatment identified 2,506 DE genes regulated by FOXN1. Finally, the gene overlap of the differentially expressed genes between cTECs in *Foxn1*$^{wt*/wt*}$ and *Foxn1*$^{wt*/-}$ mice, and the differentially expressed genes in cTECs from *iFoxn1*$^{\Delta 7,8}$ mice prior to and 3 days after Dox treatment was integrated with the 9,012 IDR FOXN1 peaks (IDR) from the analysis of the FOXN1-tagged ChIP-seq dataset by Žuklys et al. based on distance of genes TSS to the FOXN1 peak (5kb upstream or 100 bases downstream) and identified 450 highly confident candidate direct FOXN1 targets in cTECs.

Comparison of the 450 highly confident direct FOXN1 targets in newborn cTECs (Žuklys et al., 2016) with my highly confident 471 candidate direct FOXN1 targets in fetal TEPCs showed that 102 of these targets were commonly shared between the two datasets, with these genes having a FOXN1 peak overlapping with an H3K4me3 and/or an H3K27ac mark in fetal TEPCs. Apart from these 102 common targets between TEPCs and cTECs, 369 candidate direct FOXN1 targets in fetal TEPCs (identified from this analysis) did not constitute candidate direct FOXN1 targets in the 1 week cTEC population. Possible reasons for the absence of these 369 candidate targets from the newborn cTEC population include differential binding or regulation of these genes between cTECs and mTECs since the FOXN1-tagged ChIP dataset comes from a mixed cTEC-mTEC population and it would be possible that some of the FOXN1 binding sites result from unique FOXN1 binding only in one of the two populations. In this respect, 65 of the 369 genes were more strongly or equally expressed in newborn (1 week) mTECs compared to cTECs of the same age, suggesting potential regulation of these genes only in mTECs. Additionally, technical reasons could explain why genes with FOXN1 binding sites proximal to their promoter may not result as differentially expressed between cTEC samples that express normal and defective levels of FOXN1. For instance, even though *Foxn1* mRNA levels were significantly reduced by day 3 after Dox treatment, FOXN1 protein levels may still reside in the cells allowing part of genes to continue to be regulated from *iFoxn1*$^{\Delta 7,8}$ mice prior to and 3 days after Dox treatment. In support of this, on day 3 after Dox exposure, *iFoxn1*$^{\Delta 7,8}$ mice maintained a fairly normal intrathymic T-cell development. Alternatively, genes may be not strongly impacted by the absence of FOXN1 in the given period of time. Finally, it is possible that our reanalysis of the FOXN1-tagged ChIP-seq dataset by Žuklys et al. assigned more distal FOXN1 peaks to genes, since I have defined a more extended window (-5kb, +3kb) to match genes TSSs to FOXN1 regulatory peaks compared to the Žuklys study (-5kb, +100b).

To better understand the behaviour of the predicted FOXN1 targets in fetal TEPCs, I plotted the 471 highly confident candidate direct FOXN1 targets across the TEPC Developmental series dataset (E10.5, E11.5 and E12.5 timepoints) from fetal TEPCs (**Figure 5.5**) using the *pheatmap* function in R and observed them in accordance to a) the expression of the same genes in 1 week old cTEC/mTEC populations, b) the provided highly confident cTEC targets (Žuklys et al., 2016) and c) the FOXN1, H3K4me3 and H3K27ac binding overlap landscape. Gene targets with overlapping or alternative binding sites for FOXN1 and H3K4me3/H3K27ac were equally distributed across the genes (shown in **Figure 5.5**) showing no obvious patterns of preference for promoter or enhancer marks towards up-regulated or down-regulated groups of genes. Approximately 80% of the direct FOXN1 target genes that were differentially expressed between E10.5 and E12.5 timepoints from the TEPC Developmental series dataset were up-regulated with *Foxn1* increase supporting the predominantly activating role of FOXN1 in fetal TEPCs, as similarly suggested in the newborn cTEC population (Žuklys et al., 2016). The vast majority of the direct FOXN1 target genes in fetal TEPCs, which are up-regulated with increase in *Foxn1* expression from day E10.5 to E12.5 (TEPC Developmental series), were found more highly expressed in cTECs, and the opposite trend was observed in mTECs as expected, since FOXN1 is more highly expressed in the cTEC population. Lastly, overlapping FOXN1 targets with the 450 highly confident direct FOXN1 targets in newborn cTECs are shown to mark only the up-regulated proportion of the gene list, overlapping mostly with genes found higher expressed in the cTEC population (see label "whichTEC" in **Figure 5.5**).

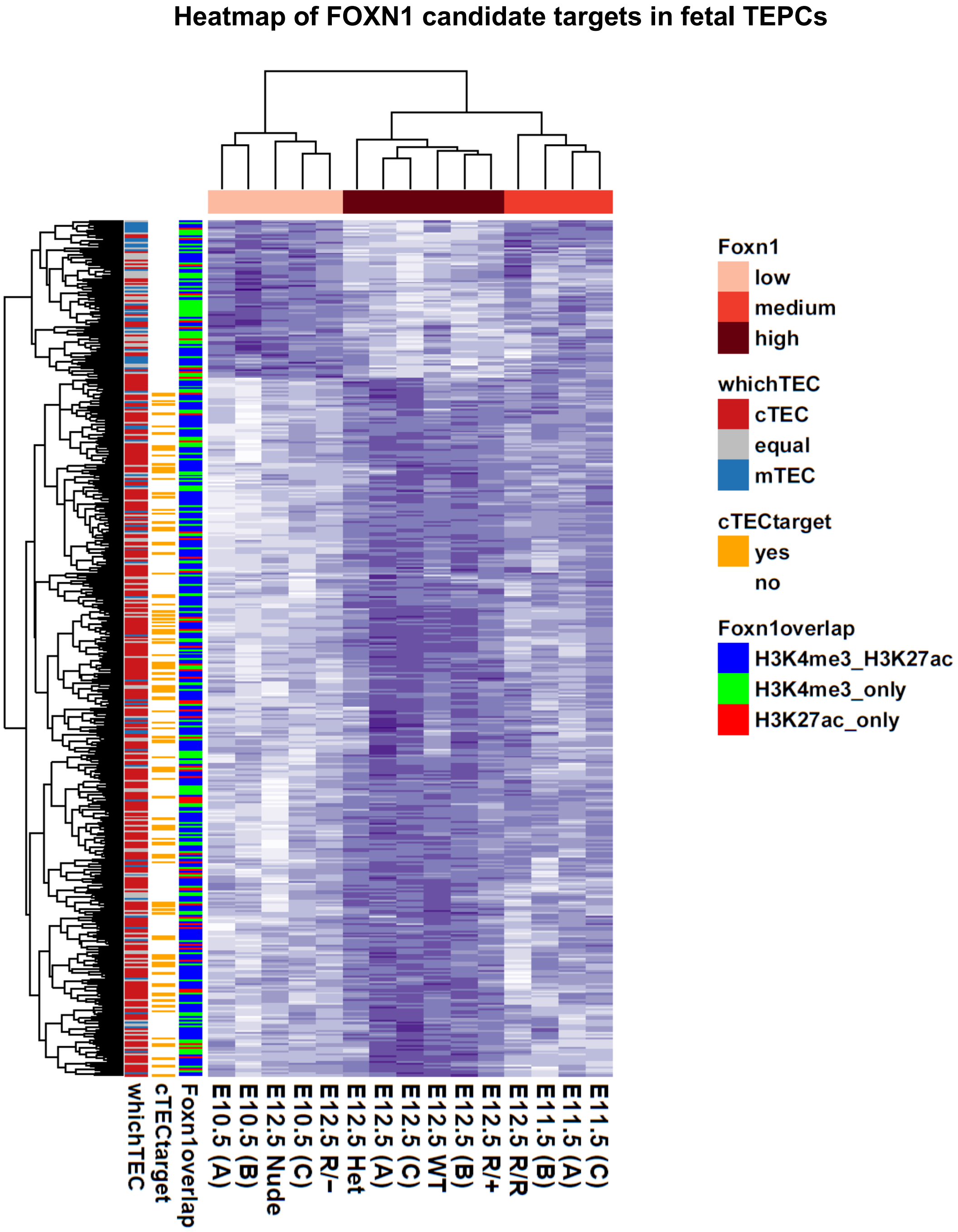


Figure 5.5: Heatmap of the high confidence FOXN1 target candidates in fetal TEPCs. Expression levels of the FOXN1 target candidates in TEPCs are shown across the TEPC Developmental series and *Foxn1* Allelic series datasets. Samples cluster according to their *Foxn1* expression level (red bar on top of the heatmap; also in agreement with **Chapter 3**, **section 3.2.1**). Rows (representative of genes) have been colour-annotated based on differential expression in TECs for an FDR ≤ 0.05 and |FC| ≥ 1.5 (red: higher expression in cTECs, blue: higher expression in mTECs, grey: not differentially expressed or not statistically significant), overlap with the high confidence FOXN1 targets in cTECs (yellow bars; Žuklys et al., 2016) and presence of FOXN1-H3K4me3/H3K27ac overlapping/alternative peaks around their TSS (blue/green/red; see also **Figure 5.4**).

5.2.3 Comparative analysis and insight of the highly confident direct candidate FOXN1 targets in newborn cTECs

To complement the analysis in **section 5.2.2** and better understand why the high confidence candidate direct FOXN1 targets in newborn cTECs (Žuklys et al., 2016) were not identified from the integrative analysis of candidate FOXN1 targets in early TEPCs, the unique FOXN1 targets in cTECs were analysed across the embryonic data series (**Chapter 3, section 3.2.1**). When the 348 FOXN1 cTEC-specific targets (450 excluding the 102 common targets) were mapped to the TEPC Developmental series and the *Foxn1* Allelic series datasets, 12% of these genes (42/348 genes; see **Table 5.1**) were not detected at all in the TEPC datasets. Of those genes that were expressed, 270 (see **Table 5.1**) did not pass the significance threshold in the differential expression analysis between the E10.5 and E12.5 timepoints and therefore were excluded from been considered as FOXN1 targets in our analysis. These genes may have a different dependence on FOXN1 in the fetal and newborn cTEC populations or they may have been too variable among replicates in the fetal TEPC Developmental series and did not pass the FDR and FC defined thresholds. An additional 22 genes (see **Table 5.1**) were not FOXN1-dependent, based on the pairwise comparison between $Foxn1^{high}$ and $Foxn1^{low}$ phenotypic samples from the fetal *Foxn1* Allelic series independently of the developmental stage (for details see **Chapter 3**, **section 3.2.2**). Again, differential regulation or variability may explain the lack of these genes from our list. A minority of 10 genes (both dependent and with a FOXN1 peak) had an IDR FOXN1 peak ($IDR \leq 0.05$) in their distal promoter (-5kb, +3kb), however, this peak did not overlap with any of the histone modification marks (H3K4me3, H3K27ac) in fetal TEPCs; these genes are potentially regulated by a different factor in early development that binds to a different accessible region close to genes promoter, while FOXN1 could become involved in their regulation later on in development (newborn stage). Lastly, 4 genes (see **Table 5.1**) did not have an IDR FOXN1 peak, based on our reanalysis of the published FOXN1-tagged ChIP-seq data. The reanalysis did not identify as many peaks as the ones identified by Žuklys et al., so it is likely that the peaks close to these genes were not identified as enriched in our reanalysis.

Unique* FOXN1 *targets in cTECs (versus TEPCs)

Not detected	42 / 348
Not differentially expressed	270 / 348
Not solely FOXN1*-dependent*	22 / 348
Non-overlapping FOXN1 *peak*	10 / 348
No FOXN1 *peak*	4 / 348

Table 5.1: Unique FOXN1 targets in newborn cTECs compared to targets in fetal TEPCs. List of potential causes of the absence of uniquely identified FOXN1 targets in newborn cTECs from the fetal TEPC population. Genes per list are mutually exclusive.

Taken together, the comparative analyses presented in **section 5.2.2** and **5.2.3** have described two major classes of genes that are directly regulated by FOXN1: a) genes regulated in both fetal TEPCs and newborn cTECs and b) genes regulated only in newborn cTECs. Genes in category (a) could be subdivided into i) genes at which the FOXN1 ChIP-seq peaks from newborn cTECs overlapped with both H3K4me3 and H3K27ac marks from fetal TEPCs, ii) genes at which the FOXN1 ChIP-seq peaks from newborn cTECs overlapped only with the H3K4me3 mark from fetal TEPCs, iii) genes at which the FOXN1 ChIP-seq peak from newborn TECs overlapped only with the H3K27ac mark from fetal TEPCs and iv) genes at which the FOXN1 ChIP-seq peak from newborn cTECs did not overlap with either H3K4me3 or H3K27ac from fetal TEPCs (see **Figure 5.4**).

5.2.4 Differential FOXN1 binding profiles in fetal TEPCs and newborn TECs

To better understand the differences between the genes regulated directly by FOXN1 in fetal TEPCs and newborn cTECs, and genes regulated directly by FOXN1 only in newborn cTECs, the overall signals from the ChIP-seq FOXN1 (newborn mixed TEC population), H3K4me3, and H3K27ac (fetal TEPCs) datasets were plotted for 4 different clusters of regions: the FOXN1 IDR peaks that overlap with a) H3K4me3, b) H3K27ac, c) both H3K4me and H3K27ac marks (clusters a, b and c define common candidate direct FOXN1 targets in fetal TEPCs and newborn TECs) or d) none of the histone modification marks (cluster d defines unique candidate direct FOXN1 targets only in TECs, and suggest differential regulation of these genes in fetal TEPCs by other factors) and peak intensity for FOXN1 was compared between these groups (see **Figure 5.6**). FOXN1 IDR peaks from newborn TECs that have no overlap with either of the histone modification marks ("only FOXN1"; **Figure 5.6**) from fetal TEPCs demonstrated the strongest signal intensity for FOXN1 compared to all other groups, suggesting stronger binding for FOXN1 in regions that are accessible in newborn TECs and not in fetal TEPCs. No major differences were observed among the other groups for the FOXN1 dataset, suggesting similar binding intensity of FOXN1-H3K4me3, FOXN1-H3K27ac and FOXN1-H3K4me3/H3K27ac sites, allowing merging these datasets together for further analysis.

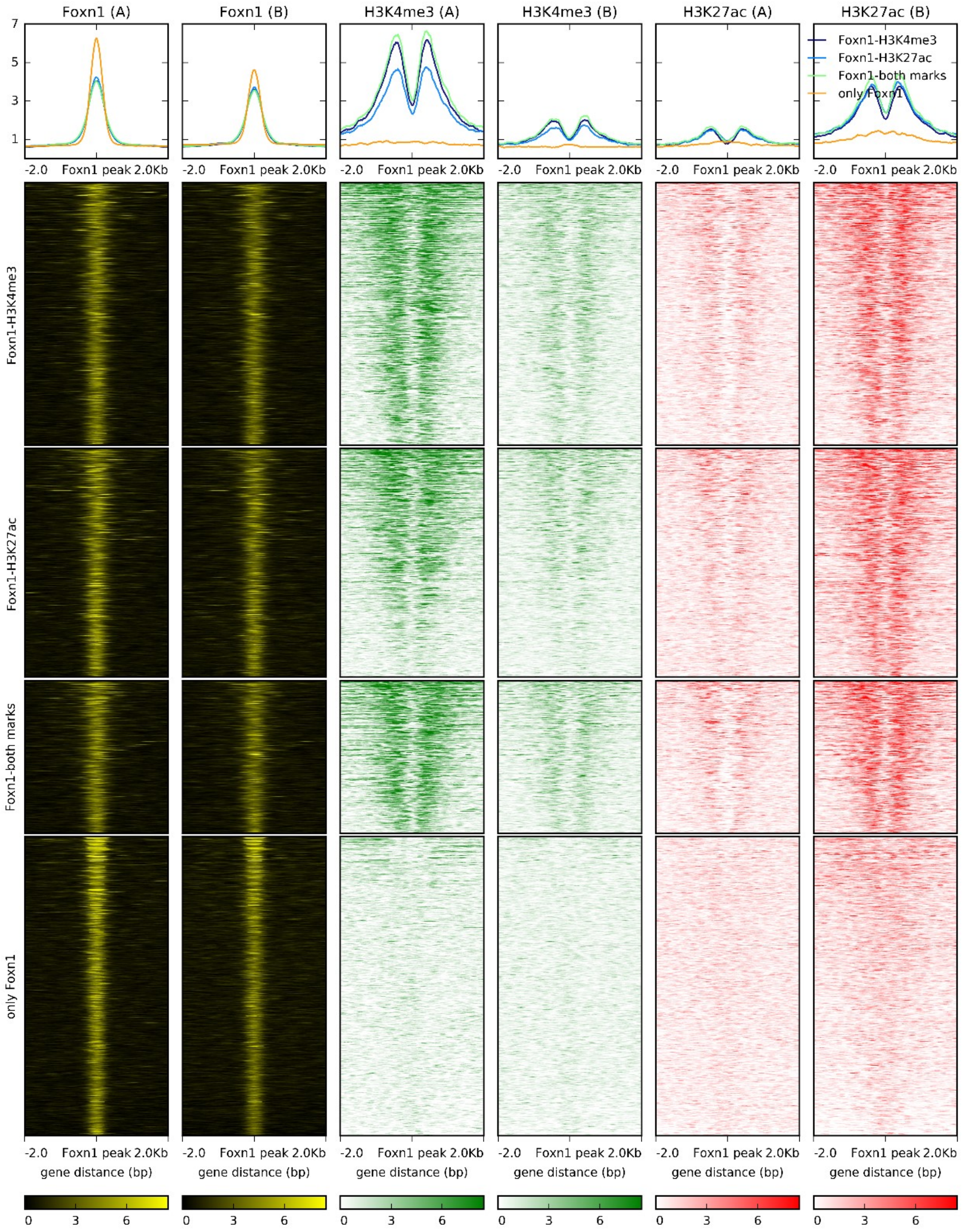


Figure 5.6: Genomic region profile plots and heatmap clusters over the FOXN1-H3K4me3, FOXN1-H3K27ac, FOXN1-H3K4me3-H3K27ac and FOXN1 IDR overlapping peaks for the FOXN1 and histone modification ChIP-seq datasets. Total signal profiles (top section) show cumulatively enrichment scores for FOXN1 peaks from 1 week old TECs (FOXN1 biological replicates A and B: FOXN1 (A) and FOXN1 (B)), and for H3K4me3 and H3K27ac peaks from E12.5 TEPCs (H3K4me3 biological replicates A and B; H3K4me3 (A) and H3K4me3 (B), H3K27ac biological replicate A and B; H3K27ac (A) and H3K27ac (B)) over regions defined by computed *k*-means clusters, Cluster 1: FOXN1-

H3K4me3 (dark blue) , Cluster 2: FOXN1-H3K27ac (light blue), Cluster 3: FOXN1-H3K4me3/H3K27ac (green) and Cluster 4: only FOXN1 (orange) ± 2kb from the centre of the FOXN1 peak. Heatmap plots (bottom section) show for each cluster the enrichment score for both FOXN1 replicates from 1 week old TECs (gold-black colour scale) and for both replicates of each histone modification mark from E12.5 TEPCs: H3K4me3 and H3K27ac (green-white, blue-white and red-white colour scales respectively). Heatmaps have been sorted in descending order based on the mean value per cluster.

5.2.5 Motif discovery in regions defined by FOXN1 peaks in newborn TECs and histone modification marks in fetal TEPCs

Following the differential FOXN1 binding analysis presented in **section 5.2.4**, I undertook a *de novo* motif discovery analysis to test if the differential peak intensity of FOXN1 in regions uniquely bound by FOXN1 in newborn TECs and regions bound by FOXN1 in newborn TECs that overlap with accessible histone modification marks in fetal TEPCs resulted from a different FOXN1 binding motif under the aforementioned regions. FOXN1 IDR peaks in newborn TECs that did not overlap with the histone modification marks in fetal TEPCs were used to predict enriched binding motifs uniquely bound by FOXN1 in TECs, while overlapping FOXN1 IDR peaks in newborn TECs with either an H3K4me and/or an H3K27ac mark in fetal TEPCs were used to predict enriched binding motifs predicted to be bound by FOXN1 in both fetal TEPCs and newborn TECs. To examine the integrity of our analysis, the H3K4me3/H3K27ac peaks in fetal TEPCs that were not overlapping with the FOXN1 IDR peaks in newborn TECs were used. Since FOXN1 did not ChIP on these regions (even though FOXN1 is expressed already at E12.5 TEPCs and the regions are accessible based on the ChIP-seq analysis of the histone modification marks in the same TEPC population), these regions should be controlled by other direct regulators, therefore no FOXN1 binding motifs should be identified in the area under the histone modification mark. Additionally, motif discovery analysis of the non-FOXN1-overlapping histone modification regions could potentially identify alternative binding sites of FOXN1 (differential regulation in fetal mouse) or binding motifs of other regulators.

Motif discovery analysis was performed using two different motif discovery tools (MEME-ChIP; Machanick and Bailey, 2011 and RSAT; Thomas-Chollier et al., 2011). MEME-ChIP uses two complementary discovery algorithms to perform *ab initio* motif discovery analysis. These algorithms are MEME (Bailey et al., 2006) and DREME (Bailey, 2011). MEME uses expectation maximisation to discover probabilistic models of single or combined transcription factor binding motifs (TFBM), while DREME uses a less complex, non-probabilistic algorithm to predict single TF binding events. The RSAT *peak-motifs* tool (Thomas-Chollier et al., 2011) implements complimentary word based approaches to perform *ab initio* motif discovery analysis. *Oligo-analysis* (van Helden et al., 1998) and *dyad-analysis* (van Helden et al., 2000a) algorithms discover over-represented oligonucleotides or spaced pairs respectively, while *position-analysis* (van Helden et al., 2000b) and *local-word-analysis* algorithms identify positionally biased oligonucleotides. Previous literature has discussed extensively how results from different motif discovery algorithms can vary greatly even if they are executed with same parameters because the underlying method is different and has suggested using more than one algorithm to improve motif discovery (Tompa et al., 2005). A recent survey focusing only on online accessible motif discovery tools showed that results consistent between tools are more reliable (Tran and Huang, 2014). In the same study, results from MEME-ChIP and RSAT in a benchmarking study were found to highly overlap. Thus, I have selected these two tools to use for this analysis.

The motif discovery analysis revealed the previously identified canonical FOXN1 binding site 5'-GACGC-3' (Nakagawa et al., 2013; Schlake et al., 1997; Žuklys et al., 2016) and a newly identified slightly extended FOXN1 motif with the same penta-peptide core sequence (GACGC) under the IDR FOXN1 peaks (IDR $\leq$ 0.05) overlapping with either or both histone modification marks (H3K4me3, H3K27ac) in fetal TEPCs, as well as, the ones that did not overlap and uniquely defined FOXN1 binding sites in newborn TECs. *De novo* motif discovery analysis for the H3K4me3 and H3K27ac marks in fetal TEPCs that do not overlap with the FOXN1 IDR peaks in newborn TECs did not identify the canonical FOXN1 motif. In the latter case, accessible regions in fetal TEPCs without FOXN1 binding evidence are potentially controlled by other transcription factors.

From this, I conclude that FOXN1 bound peaks in newborn TECs that overlap with histone modification marks in fetal TEPCs identify the very same FOXN1 motif with the FOXN1 bound peaks in newborn TECs that are uniquely accessible in TECs (no overlap with histone modification marks from fetal TEPCs). Importantly though, a slight variation of the RSAT identified motif, with the adenosine in the 6th position of the *y*-axis in **Table 5.2** (green "A") to be not as conserved, could suggest that other regulators may bind co-operatively in some of the uniquely accessible FOXN1 regions in TECs and this motif could be the outcome of the averaged profiles of two different existing motifs. Therefore, differences in binding intensity between the overlapping and non-overlapping IDR FOXN1 peaks could be partially explained by differential binding of the overlapping and the non-overlapping regions. Additionally, the lack of an enriched FOXN1 motif in the accessible regions of the fetal TEPCs marked by the histone modification marks suggests that the FOXN1 binding motif is not generally enriched in the accessible regions of the genome, but it is specifically located under the FOXN1 ChIPed peaks.

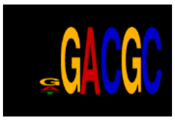
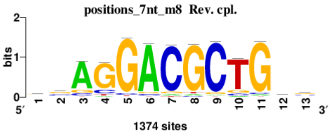
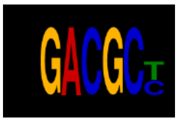
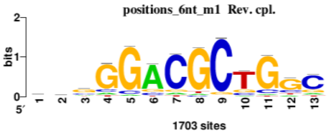
	MEME		**RSAT**	
Peaks	**FOXN1 motif**	**E-value**	**FOXN1 motif**	**E-value**
Common peaks		6.5e-025		3.6e-08
Unique FOXN1		3.5e-026		8.7e-31
Unique H3K4me3 - H3K27ac	NONE	-	NONE	-

Table 5.2: ***De novo*** **identification of canonical and extended FOXN1 binding motifs in the regions under the IDR FOXN1 peaks in newborn TECs that overlap (or not) with histone modification marks in fetal TEPCs.** Table depicts FOXN1-identified motifs in the regions under the FOXN1 bound peaks in newborn TECs that overlap with histone modification marks (H3K4me3 and H3K27ac) in fetal TEPCs (common peaks) and under the FOXN1 bound peaks in newborn TECs that are uniquely accessible in TECs (no overlap with histone modification marks in fetal TEPCs; unique FOXN1). The dash ("-") symbol is used when no similar motifs to the previously identified canonical and alternative FOXN1 motif are identified (regions under the histone modification marks that do not overlap with IDR FOXN1 peaks from newborn TECs; Unique H3K4me3-H3K27ac). The statistical significance of the enrichment (E-value) assigned by the *de novo* motif discovery tools, MEME-ChIP and RSAT is also provided in the table.

5.2.6 Nodal points that FOXN1 regulates in fetal TEPCs during thymus organogenesis

In **sections 5.2.1 to 5.2.5,** I have described an integrated analysis of transcriptional, regulatory and accessibility datasets, which identified confident candidate FOXN1 target genes in fetal TEPCs, while FOXN1 peak and binding motif differences among fetal TEPCs and newborn TECs were also discussed thoroughly. I next wished to investigate the molecular mechanisms that FOXN1 may regulate via its targets and how these may control TEPC progression.

As described in **Chapter 4** (**section 4.2.1**), to predict the signalling pathways that are enriched between the E10.5 and E12.5 TEPC developmental stages, all differentially expressed genes (in number 15,321) from the comparison above (analysis with limma) were identified and used as a 'Pre-Ranked Gene List' in GSEA (Subramanian et al., 2005) against the edited ConsensusPathDB database (Kamburov et al., 2011) to predict biological pathways (or processes) that alter between these two developmental stages. Only pathways with an FDR ≤ 0.25 were considered as enriched and these were selected for further analysis (see **Table 5.3** and **Table 4.1** for pathways and **Chapter 2**, **section 2.3.3.6** for interpretation of the GSEA results).

FOXN1, similar to other transcription factors, can interact with specific pathways by targeting their intrinsic regulators. Since this analysis aims to identify nodal points in early TEPC progression directly regulated by FOXN1, I compared the genes contributing to the core enrichment of each pathway (**Table 5.3** and **Chapter 4**, **Table 4.1**) to the 471 high confidence candidate direct FOXN1 targets in fetal TEPCs identified in **section 5.2.1** (see **Figure 5.7**) and I used the overlap of this comparison to select the enriched pathways that are directly regulated by FOXN1 during the undifferentiated (E10.5) – differentiating TEPCs (E12.5) transition. The generated list of 39 candidate direct FOXN1 targets was used to group the selected pathways into broader biological categories (**Figure 5.7**, A-F gene clusters) based on the presence of ligands, receptors, signalling molecules or machinery subunits.

Among the pathways in **Table 5.3**, the TCR SIGNALING (#10), DOWNSTREAM TCR SIGNALING (#13), COSTIMULATION BY THE CD28 FAMILY (#27) and T-

CELL RECEPTOR SIGNALING PATHWAY (#51) candidates suggest increased activity of T-cell specific pathways in the E12.5 TEPCs, an enrichment that could indicate T-cell contamination. However, a potential contamination of the RNA-seq samples would be difficult based on the fact that hematopoietic progenitors migrate to the mesenchymal capsule of the thymus around E11.5, but they are not observed among epithelial cells until around E12 (Itoi et al., 2001). Moreover, even if T-cell precursors were present among the obtained cell population of thymic epithelial cells, sorting these cells for EPCAM and PLET1 should remove any non-epithelial cells. Additionally, a closer look to the core enrichment genes of these pathways can resolve any dispute, since the involved genes are not T-cell specific genes. On the contrary, they seem to be mostly ligands, receptors, transcription factors and intermediate molecules (*H2-Aa*, *Prkcq*, *Pag1*, *H2-Eb1*, *Nfkb1*, *Nfkbia*, *Zap70*, *Ube2v1* in #10; *H2-Aa, Prkcq*, *H2-Eb1*, *Nfkb1*, *Nfkbia*, *Ube2v1* in #13; *H2-Aa*, *Cd274*, *Pdcd1lg2*, *H2-Eb1*, *Map3k8*, *Lyn*, *Ptpn6*, *Cd80* in #27 and *Prkcq*, *Nfatc2*, *Pak6*, *Nfatc1*, *Map3k8*, *Tec*, *Fos*, *Map3k14*, *Nfkbie*, *Nfkbia*, *Vav3*, *Zap70*, *Jun*, *Ptpn6*, *Cblc*, *Il4* in #51) that could be involved in alternative signalling pathways, however, in this case these genes (together) have qualified to surpass the enrichment analysis threshold and result in enrichment of the aforementioned pathways. This observation highlights again the inefficacy of the manually curated pathways and/or the pathway enrichment analysis tools to accurately predict the biological reality in the different cell types (previously discussed in **Chapter 4**, **section 4.2.1**), since existence of T-cell specific genes should be required to allocate enrichment to T-cell specific pathways.

Based on the enriched pathways between the E10.5 and the E12.5 developmental stages during TEPC differentiation (shown in **Table 5.3**) and their core enrichment genes that were also identified to be high-confidence FOXN1 direct targets (shown in **Figure 5.7**, I conclude that FOXN1 could directly regulate the downstream cytokine signalling, the antigen processing and presentation program, NF-κB and NOTCH signalling pathways and protein degradation (gene groups depicted in **Figure 5.7**). These pathways are discussed further in **sections 5.3.1** and **5.3.2**.

ID	GS follow link to MSigDB	SIZE	ES	NES	NOM p-val	FDR q-val
1	FOXN1 HIGH CONFIDENCE TARGETS IN CTECS (NEWBORN MOUSE)	406	0.62	2.43	0	0
3	NF-KAPPA B SIGNALING PATHWAY - MUS MUSCULUS (MOUSE)	67	0.62	1.95	0	0.008
6	TYPE II INTERFERON SIGNALING (IFNG)	21	0.72	1.83	0.002	0.039
8	INTERLEUKIN-1 SIGNALING	33	0.65	1.82	0	0.036
10	TCR SIGNALING	33	0.64	1.79	0	0.042
12	SIGNALING BY INTERLEUKINS	79	0.54	1.76	0	0.051
13	DOWNSTREAM TCR SIGNALING	22	0.69	1.75	0.004	0.06
14	INTESTINAL IMMUNE NETWORK FOR IGA PRODUCTION - MUS MUSCULUS (MOUSE)	20	0.71	1.75	0.004	0.056
16	CYTOKINE SIGNALING IN IMMUNE SYSTEM	121	0.51	1.73	0.002	0.057
19	ANTIGEN PROCESSING AND PRESENTATION - MUS MUSCULUS (MOUSE)	43	0.59	1.67	0.005	0.103
22	RIG-I(OR)MDA5 MEDIATED INDUCTION OF IFN-ALPHA(OR)BETA PATHWAYS	41	0.57	1.66	0.004	0.113
25	CYTOKINE-CYTOKINE RECEPTOR INTERACTION - MUS MUSCULUS (MOUSE)	94	0.49	1.64	0.003	0.112
27	COSTIMULATION BY THE CD28 FAMILY	36	0.58	1.61	0.011	0.151
29	IL-1 SIGNALING PATHWAY	31	0.58	1.6	0.005	0.154
30	TRAF6 MEDIATED NF-KB ACTIVATION	20	0.63	1.59	0.028	0.175
34	TNF SIGNALING PATHWAY - MUS MUSCULUS (MOUSE)	83	0.48	1.58	0.005	0.163
35	PROTEASOME - MUS MUSCULUS (MOUSE)	40	0.54	1.58	0.02	0.159
37	JAK-STAT SIGNALING PATHWAY - MUS MUSCULUS (MOUSE)	81	0.49	1.57	0.003	0.167
40	APOPTOSIS - MUS MUSCULUS (MOUSE)	65	0.5	1.55	0.009	0.186
43	TRAF6 MEDIATED IRF7 ACTIVATION	16	0.65	1.54	0.039	0.194
44	TARGETED PROTEIN DEGRADATION	55	0.5	1.53	0.014	0.199
46	INTERFERON SIGNALING	40	0.53	1.51	0.025	0.222
48	IL-2 SIGNALING PATHWAY	66	0.48	1.5	0.021	0.235
51	T CELL RECEPTOR SIGNALING PATHWAY - MUS MUSCULUS (MOUSE)	74	0.47	1.5	0.02	0.227

Table 5.3: List of signalling pathways whose activity increases during the developmental progression from E10.5 undifferentiated TEPCs to E12.5 TEPCs. In sequence, the table provides name of each pathway (GS follow link to MSigDB), number of Pre-Ranked genes identified per pathway (size), enrichment score of each pathway (ES), enrichment score of each pathway normalised for the size of the pathway (NES), NOM *p*-val and FDR *q*-val (see Material and Methods for more details).

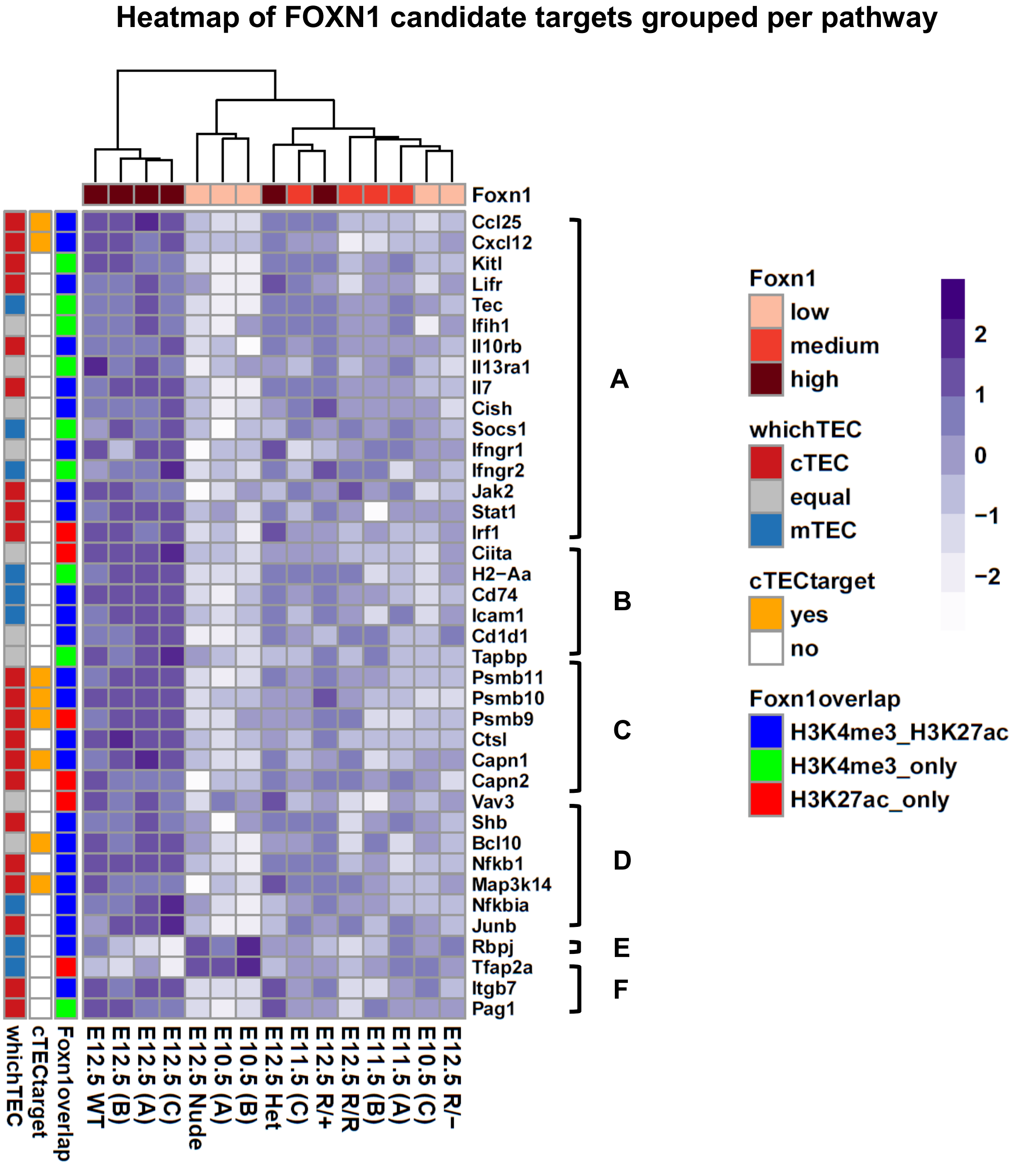


Figure 5.7: Heatmap of the high confidence FOXN1 target candidates that regulate enriched signalling pathways in fetal TEPCs. Expression levels of the FOXN1 target candidates involved in signalling pathways in TEPCs are shown across the TEPC Developmental series and *Foxn1* Allelic series datasets. Genes (rows) have been colour-annotated based on presence of an FOXN1-H3K4me3 overlapping/alternative peak around their TSS (green/blue; see also **Figure 5.4**), overlap with the high confidence FOXN1 targets in cTECs (yellow bars; Žuklys et al., 2016) and differential expression in TECs with statistical significance for an FDR ≤ 0.05 and |logFC| ≥ 0.585 (red: higher expression in cTECs, blue: higher expression in mTECs, grey: not differentially expressed or not statistically significant). Genes were ordered and categorised in 6 groups [**A-F**] based on their biological role or signalling activity. **A:** Cytokines with group subclustering into FOXN1 targets (*Ccl25*, *Kitl*), IL-7 signalling (*Il7*, *Cish*) and IFN-γ signalling (*Ifngr1*, *Ifngr2*, *Jak2*, *Irf1*, *Stat1*), **B:** Antigen processing and presentation programme, **C:** Degradation with group subclustering into proteasome (*Psmb11*, *Psmb10*, *Psmb9*) and proteases (*Ctsl*, *Capn1*, *Capn2*), **D:** T-cell regulation with group subclustering into NF-κB signalling (*Bcl10*, *Nfkb1*, *Map3k14*, *Nfkbia*, *Junb*) and VAV-SHB signalling, **E:** NOTCH signalling and **F:** Other.

5.3 DISCUSSION

5.3.1 FOXN1: a potential pioneer factor

The analysis presented in this chapter has provided further insight into the FOXN1 binding landscape and the FOXN1 gene regulatory network in fetal TEPCs. It has also highlighted the central role of FOXN1 in the establishment of the thymic system through comparative analysis of the FOXN1 binding landscape in newborn TECs with the chromatin accessibility landscape in fetal TEPCs. The latter analysis has demonstrated that the binding landscapes of FOXN1 in these two developmental stages are highly similar, suggesting regulation of the same programmes in both early (fetal TEPCs) and late (newborn TECs) TEC differentiation stages. The highly similar predicted landscape for FOXN1 binding in combination with known literature where reversion of a *null* allele of FOXN1 can initiate the thymic programme in *nude* TECs (Nowell et al., 2011) and where forced FOXN1 expression in an un-related population of cells (MEFs) converted these cells into TECs (Bredenkamp et al., 2014b) suggests that FOXN1 may be able to open chromatin in the regions that it binds, in order to apply the same regulatory programme in the above cell types. This would be more obvious in the case of MEFs, since they constitute a TEC-unrelated population and FOXN1 would have to bind and initiate the expression of genes that are not normally expressed in MEFs to establish the thymic program.

In other words, FOXN1 may be acting as a 'pioneer' factor: pioneer transcription factors can bind to specific DNA motifs, open chromatin and formulate the transcriptional landscape of the cells accordingly (Zaret and Carroll, 2011). In more detail, FOXN1 could bind FOXN1-specific DNA sequences within "closed" chromatin and instruct the chromatin to remodel. Chromatin remodelling would then make "hidden" DNA sequences accessible to regulatory factors (general transcription factors, activators, RNA polymerase II) which would bind and interact with each other to enable initiation of gene transcription (see also **Chapter 1**, **section 1.5.1**). Even though FOXN1 has been shown to act as an activator, analysis in **section 5.2.1.3** identified 129 genes as candidate direct FOXN1 targets whose expression drops with increase in *Foxn1* expression, demonstrating a potential repressive role for FOXN1 that has not been observed before (this was suggested in **Chapter 3**, **section 3.3**). Taking into consideration its potency as a pioneer factor, FOXN1 could cause gene silencing by binding alongside repressors or corepressors. These factors could in turn cause local domain blanketing that would impair activator recruitment and would repress transcription. Nevertheless, to verify or reject the hypothesis of FOXN1 exhibiting pioneer factor activity, chromatin accessibility datasets in fetal TEPCs pre and post FOXN1 initiation would be required to identify FOXN1 regulated regions that are in closed formation in the absence of FOXN1 and become accessible just after FOXN1 is activated through FOXN1 direct binding to chromatin.

5.3.2 A first-described high-confidence list of FOXN1 direct targets in fetal TEPCs and dynamic regulation of pathways

The analysis described in this chapter has identified 471 candidate direct FOXN1 target genes that are expressed in fetal TEPCs, and a "most" confident list of 102 FOXN1 targets that have been predicted to be directly regulated by FOXN1 in both fetal TEPCs and newborn cTECs. Several of the FOXN1 targets in TEPCs were also shown, via complimentary pathway enrichment analysis, to directly regulate a list of important biological pathways for thymic function. TEPC pathways directly regulated by FOXN1 include downstream cytokine signalling (activation of cytokine receptors in TECs), the antigen processing and presentation program (AP&P programme), protein degradation, NF-κB signalling and NOTCH signalling. Based on the current literature, most of these pathways can be interconnected through loops of positive/negative regulation (see below). Additionally, the role of FOXN1 in the maintenance of the antigen processing and presentation programme and protein degradation pathways has already been demonstrated in newborn cTECs (Žuklys et al., 2016).

In the next paragraphs, the FOXN1-regulated pathways are described in more detail in an effort to define a testable *in silico* FOXN1 regulatory model that acts during early TEPC progression.

5.3.3 FOXN1 drives and maintains the antigen processing and presentation programme via multiple mechanisms

One of the main regulators of the antigen processing and presentation (AP&P) program in multiple cell types is the interferon gamma IFN-γ signalling pathway (Dong et al., 1999; Muhlethaler-Mottet et al., 1998; Piskurich et al., 1998, 1998; Rohn et al., 1999). Below, I discuss the potential role of IFN-γ in FOXN1-mediated regulation of the AP&P pathway in TECs.

5.3.3.1 General mechanism of interferon gamma signalling

The interferon gamma signalling pathway (**Figure 5.8**; Reith et al., 2005) involves signalling via the JAK-STAT pathway (reviewed in Schroder, 2003). IFN-γ binds to the IFN-γ receptor (IFN-γR; consisting of the IFNGR1 and IFNGR2 subunits). This changes the conformation of the receptor, leading to auto-phosphorylation (and activation) of JAK2, which in turn transphosphorylates (and activates) JAK1 (Briscoe et al., 1996). These events lead to the release of the STAT1 homodimer from the receptor (Greenlund et al., 1995). Free STAT1 reaches the nucleus where it binds to gamma activated site (GAS) elements at the promoter of genes to initiate or suppress IFN-γ regulated gene transcription (Decker et al., 1997; Meraz et al., 1996). Many of the STAT1 regulated genes are transcription factors (such as IRF1; Pine et al., 1994) which drive the next wave of transcription by binding to IFN-stimulated response elements (ISRE) at the promoter regions of target genes (Harada et al., 1989; Nelson et al., 1993), with IRF1 to be able to promote transcription of STAT1. A number of genes involved in the IFN-γ signalling pathway, *Ifngr1*, *Ifngr2*, *Jak2*, *Irf1*, *Stat1*, have been all predicted to be candidate direct FOXN1 targets in fetal TEPCs based on the analysis in this chapter (**section 5.2.1**), indicating a very consistent regulation of the IFN-γ signalling by FOXN1 at multiple levels.

With respect to the antigen processing and presentation program, IFN-γ regulates both constitutive and induced expression of MHC class II (MHCII) (Boehm et al., 1997), as described below.

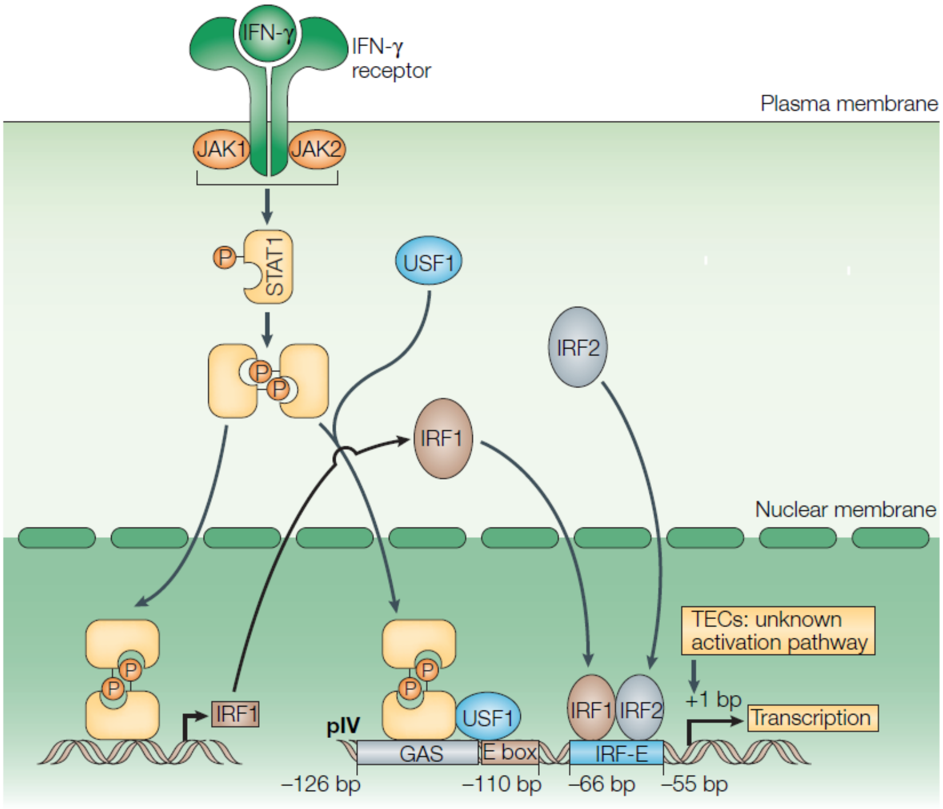

IFN-γ
IFN-γ receptor
Plasma membrane
JAK1
JAK2
P
STAT1
P
P
USF1
IRF2
IRF1
Nuclear membrane
P
P
IRF1
P
P
pIV
USF1
IRF1
IRF2
TECs: unknown activation pathway
+1 bp
Transcription
GAS
E box
IRF-E
−126 bp
−110 bp
−66 bp
−55 bp

Figure 5.8: Regulation the MHCII programme through direct binding of the interferon gamma activated STAT1 and IRF1 on the pIV promoter of *Ciita*. The general mechanism of IFN-γ is depicted in upper section: IFN-γ binds to the IFN-γ receptor (IFN-γR; consisting of the IFNGR1 and IFNGR2 subunits) and changes IFNGR1 receptor conformation (through activation of JAK1 and sequentially JAK2) which enables the recruitment of a STAT1 pair. STAT1-homodimer phosphorylation releases the dimer from the receptor and free STAT1 reaches the nucleus where it binds to GAS elements at the promoter of genes and can activate or suppress IFN-γ regulated gene transcription. IRF1 is one of the main targets of STAT1 that will then drive the next wave of transcription by binding to IFN-stimulated response elements (ISRE) at the promoter regions of target genes. Direct binding on the pIV CIITA promoter is shown in lower section: IFN-γ activated STAT1 and IRF1 bind on the TEC specific promoter (pIV) that includes a GAS and an IRF-specific binding motif within the proximal promoter of CIITA and induce its transcription (reviewed in (Schroder, 2003). Image reproduced from Reith et al., 2005.

5.3.3.2 MHC class II antigen presentation pathway and IFN-γ

IFN-γ is a key regulator of the MHCII antigen processing and presentation programme (Schroder, 2003). Professional antigen presenting cells (B-cells, macrophages and dendritic cells) and TECs are the only cell types that constitutively express MHCII (LeibundGut-Landmann et al., 2004), and treatment with IFN-γ can lead to further up-regulation of the MHCII molecules in the majority of these cells. Additionally, in cells that do not constitutively express MHCII, IFN-γ comprises the most potent inducer of MHCII expression (LeibundGut-Landmann et al., 2004).

5.3.3.3 The MHC class II transactivator (CIITA)

Typically, IFN-γ regulates expression of MHCII genes by targeting the pIV promoter of the class II transactivator (CIITA, see **Figure 5.8**). This transactivator has been characterised as the "master control factor" of the MHCII antigen presentation programme because it can activate on its own both genes encoding classical MHCII molecules and those encoding accessory proteins needed for MHCII antigen presentation (LeibundGut-Landmann et al., 2004). The pIV promoter contains a GAS element and an E-box that are bound cooperatively by STAT1 and USF-1, as well as an IRF1-consensus motif that is occupied by IRF1. These elements exist within a 300bp region and are necessary for CIITA induction in most cells (LeibundGut-Landmann et al., 2004).

5.3.3.4 CIITA and TECs

The pIV promoter regulates expression of the MHCII molecules in TECs (Muhlethaler-Mottet, 1997) and it is essential for maintenance of MHCII gene expression in both cTECs and mTECs (Waldburger, 2003; Waldburger et al., 2001). pIV knockout mice (MHCII$^-$ TECs) demonstrate strong reduction of CD4$^+$ T-cells due to ablation of positive selection since this process requires MHCII$^+$ cTECs (Waldburger, 2003). The same mice also showed a reduction in genes that encode peripheral tissue antigens in mTECs. This process normally promotes tolerance to autoantigens and absence of these peripheral antigens causes autoimmunity (Waldburger, 2003).

5.3.3.5 FOXN1 consists part of the unknown mechanism that maintains CIITA's expression and the MHC class II programme in the absence of IFN-γ signalling

The mechanism by which IFN-γ induces CIITA and MHCII expression in TECs (constitutive expression of MHCII molecules) appears to differ from that operating in other cells that constitutively express MHCII (APCs). Although treatment with IFN-γ further up-regulates CIITA in TECs, loss of essential components of the IFN-γ signalling – including IFN-γ, IFN-γR, IRF1 and STAT1; did not alter positive selection of $CD4^+$ cells, suggesting that an alternative mechanism must mediate pIV activation in TECs (Reith et al., 2005). The unknown mechanism of pIV activation is believed to depend on signals provided by the thymic microenvironment because TECs lose expression of MHCII if cultured in a monolayer *ex vivo*, while they maintain MHCII expression in reaggregate thymic organ cultures (Anderson et al., 1993; Reith et al., 2005) or *in vivo*. A potential model for FOXN1-mediated regulation of MHCII is discussed below.

Thymic epithelial cells seem to possess a unique mechanism for maintaining/ recovering the MHCII antigen presentation programme if IFN-γ signalling is disrupted. FOXN1 is the master regulator of TEC differentiation, and is expressed only in thymic epithelial cells and keratinocytes (which do not constitutively express MHCII). FOXN1 expression is also lost when TECs are cultured in a monolayer and this observation has suggested that its maintenance may require signals from a surrounding mesenchymal population, cell-cell contact and proper vasculature. This suggests that FOXN1 could potentially regulate the IFN-γ independent mechanism that maintains MHCII antigen presentation. Furthermore, the analysis presented above shows that FOXN1 can directly bind *Ciita; Ciita* is not expressed in E10.5 TEPCs, but is expressed once *Foxn1* has been initiated. Thus, FOXN1 might be responsible not only for maintenance of the MHCII antigen presentation program in TECs, but also for its initiation.

Finally, in agreement with the role of FOXN1 in regulating the MHCII AP&P, *Ctsl*, which facilitates MHCII maturation through processing of the MCHII molecules invariant chain (Honey et al., 2002; Hsieh et al., 2002), was predicted to belong in the high confidence candidate direct FOXN1 targets list

Our data suggest that FOXN1 regulates the MHCII AP&P program at two levels: by targeting CIITA directly and via the IFN-γ signalling pathway; since STAT1 and IRF1 (which bind to the CIITA promoter region) are both candidate direct targets of FOXN1 in TEPCs. Although loss of each of these factors individually has been reported not to affect CD4$^+$ thymocyte selection and therefore by extension MHCII antigen processing and presentation, the effect of loss of both (*Irf1*$^{-/-}$ and *Stat1*$^{-/-}$ mice) on activation of CIITA in TEC has not been tested.

In conclusion, this analysis strongly suggests that FOXN1 directly activates and maintains the MHCII program through FOXN1-targeted IFN-γ signalling or through direct binding to the *Ciita* pIV promoter under both normal and disrupted IFN-γ signalling conditions.

5.3.4 The FOXN1-NOTCH interplay in determining TEC specification

The importance of the NOTCH signalling pathway in early TEPC differentiation has been extensively discussed in **Chapter 4**, **section 4.3.4** with the NOTCH canonical pathway shown to play a critical role in mTEC specification. In brief, ablation of NOTCH signalling (NOTCH loss-of-function; LOF) in undifferentiated E10.5 TEPCs (when FOXN1 is still inactive) led to a complete loss of the mTEC compartment, while blocking NOTCH signalling at later stages (RBPJ deletion under the control of the *Foxn1* promoter in E12.5 TEPCs) caused a lesser reduction in the mTEC number (Liu et al., 2017, submitted). On the other hand, reinforcement of NOTCH signalling by overexpression of NOTCH intracellular domain (NOTCH gain-of-function; NOTCH GOF) restricted TEC lineage differentiation (retention of PLET1 staining, **Dong Liu**), and even though cells also lacked differentiation markers (**Figure 4.11**), it did not lead to forced mTEC differentiation (**Figure 4.11**).

NOTCH signalling was identified as one of the signalling pathways down-regulated with significance in TEPCs between E10.5 and E12.5 (**section 4.2.1.2**; **Figure 4.1**,

Figure 4.6) with *Notch3*, *Dll1, Dll3*, *Rbpj* and *Msx2* genes (**Figure 4.3**) contributing to the core enrichment of the pathway. Unlike the rest of the genes contributing to the pathway's core enrichment, *Rbpj* was also identified as a high confidence candidate direct FOXN1 target in fetal TEPCs (**Figure 5.7**; Group E), and one of the very few genes to be seemingly repressed by FOXN1 up-regulation. This finding is consistent with the less severe effect on mTECs of *Rbpj* deletion via *Cre* under the control of *Foxn1* promoter (see **Chapter 4**, **section 4.3.4**), since if FOXN1 is a natural repressor of RBPJ, the *Rbpj* knockout would simply slightly amplify that effect.

Additionally, clustering of the TEPC and TEC-specific lineage markers for NOTCH loss-of-function (LOF; RBPJ deletion under the control of the *Foxn1* promoter in E12.5 TEPCs), NOTCH gain-of-function (GOF: NICD overexpression) and control samples presented in **Chapter 4**, **section 4.2.3.3** (**Figure 4.11**) identified multiple cTEC markers (*Ctsl, Krt8, Ly75 and Psmb11*) to show a reduced expression pattern in the E14.5 GOF samples compared to the E14.5 controls. Both *Psmb11* and *Ly75* have been predicted to be high confidence candidate direct FOXN1 targets in newborn cTECs, with *Psmb11* to also be experimentally verified (via mutation of the FOXN1 binding site analysis) as a direct FOXN1 target in cTECs (Žuklys et al., 2016). Additionally, *Ctsl* and *Krt8* have been predicted as high confidence candidate direct FOXN1 targets in fetal TEPCs. Taken together, these results suggest that blockage of the cTEC progression by NOTCH may be mediated by repression of the expression and/or activity of FOXN1. Partial FOXN1 repression by NOTCH expression may explain how NOTCH protects the mTEC lineage from higher levels of FOXN1 induction.

The relationship between FOXN1 and NOTCH was further explored by **Dr Kathy O'Neil** who has generated *Rosa26*$^{CAG\text{-}FL\text{-}STOP\text{-}FL\text{-}Foxn1\text{-}IRES\text{-}GFP}$ transgenic mice (designated *iFoxn1*), which provide a model where *Foxn1* overexpression can be induced under the control of *Cre* (Bredenkamp et al., 2014b). In particular, male *Foxn1*^{Cre} mice were mated with female *iFoxn1* mice to produce double transgenic litters which constitutively overexpress *Foxn1* in TECs (designated *TgFoxn1*). Analysed E17.5 *TgFoxn1* thymi revealed the presence of very few mTECs, and a highly similar phenotype with the one observed by the NOTCH LOF model (Liu et al., 2017, submitted). Additionally, RT-qPCR analysis demonstrated a three-fold increase in *Foxn1* expression, while a two-fold reduction of *Notch1* and *Notch3* in mTECs. Therefore, molecular interactions seem to support the phenotypic similarities between FOXN1 gain-of-function and NOTCH loss-of-function. Importantly, analysis in this chapter has identified among the high confident candidate direct FOXN1 targets in fetal TEPCs, *Foxn1* to also constitute a FOXN1 target itself. Thus, after *Foxn1* initiation, FOXN1 may be able to further boost its own expression through auto-regulation.

In summary, investigation of NOTCH signalling pathway and NOTCH gain-of-function experiments reveal cross repressive modulations between FOXN1 and NOTCH, suggesting a regulatory model in early TEPC differentiation, in which NOTCH signalling protects the emergence of the mTEC compartment during an early defined time-window, FOXN1 expression represses NOTCH and a FOXN1-NOTCH interplay determines if TEPCs will acquire the cTEC or the mTEC fate.

5.3.5 A potential role for *Tfap2a* in the early TEPCs

Analysis in this chapter has revealed a similar expression pattern of *Rbpj* with the transcription factor *Tfap2a* (**Figure 5.7**), with high *Tfap2a* expression levels to be associated with immature cells and *Tfap2a* expression drop to relate to cells developmental progression. The role of TFAP2A (also commonly known as AP2) has not yet been investigated in the thymic epithelium. Nevertheless, *Tfap2a* has been shown to play an important role in both the epidermis and the neural crest (NC) cell gene regulation. In the epidermis, AP2 is necessary (but not sufficient) for epidermal gene expression (Leask et al., 1991). Early studies have shown evidence for AP2 protein binding directly to the proximal promoter of Keratin 14 (*K14*) and other epidermal-specific genes, however, additional binding of other transcription factors in a distal *K14* element contributes to tissue-specific expression of *K14* (Leask et al., 1991). Since keratins K14 and K5 have been associated with mitotically active keratinocytes (Nelson and Sun, 1983), and their down-regulation occurs upon further keratinocyte differentiation (different pairs of keratins are switched on then) (Fuchs and Green, 1980; Moll et al., 1982; Sun et al., 1984), it could be inferred that AP2 expression is linked to a more immature cell phenotype. Furthermore, Wang et al.

(2011) reported that *Tfap2a* acts within the same genetic pathway with Forkhead Box D3 (*Foxd3*) during early neural crest cells development, with combined loss of these factors (*Tfap2a*^{mob} and *Foxd3*^{mos}) to cause nearly complete absence of NC-derived tissues. Their findings indicated that both factors are necessary for NC specification, but their role is still important during the earliest steps of the NC precursor cell development. Taken together, the above studies suggest that *Tfap2a* plays a significant role in early progenitor cell progression, with tissue-specificity possibly determined by the action of other regulatory factors. Similarly, it is possible that in the thymic epithelium, expression of *Tfap2a* is required in the early specification and development of TEPCs, with FOXN1 up-regulation (which belongs in the same protein family with FOXD3) to lead to *Tfap2a* down-regulation by direct binding to the *Tfap2a* enhancer region upon further developmental progression.

5.4 SUMMARY

In conclusion, **Chapter 5** has investigated the mechanisms via which FOXN1 regulates early fetal thymic epithelial progenitor cell differentiation, in order to better understand its dynamic role in the establishment of a fully functional thymus. Comparative analysis of FOXN1 ChIP-seq datasets in newborn TECs with histone modification ChIP-seq datasets in fetal TEPCs demonstrated high similarity in the FOXN1 binding landscape in newborn TECs with the chromatin accessible landscape in fetal TEPCs. This finding, in combination with predicted biological pathways in fetal TEPCs and newborn cTECs that are commonly regulated by FOXN1, suggests that FOXN1 may create a similar transcriptional landscape in these two different developmental stages in the mouse thymus. This finding could enhance the already suggested (Vaidya et al., 2016) role of FOXN1 as a pioneer factor which can establish and maintain a FOXN1-specific thymic landscape given a chromatin substrate; however, further experiments would be necessary to prove the pioneer role of FOXN1. This model of pioneer activity could also explain how FOXN1 is able to lead the conversion of an unrelated cell type into TECs which are also able to support the development of a fully functional system.

Chapter 6

ThymiBase

Chapter 6

ThymiBase

6.1 PRÉCIS

The large amount of bioinformatics datasets generated and analysed during this study has rendered the need to store and share these data in an easy and manageable way, but to also allow downstream analysis and visualisation of the presented datasets that may assist in other ongoing research projects. Currently, there are only two online databases providing expressional profiling of immune cell types: the Gene expression commons (Seita et al., 2012) and the Immunological Genome Project (Immgen; Heng et al., 2008) databases. The majority of the expressional information available in both platforms has been derived mostly from microarray experiments, with only Immgen providing RNA-seq profiling for 34 immune cell types, from which only one dataset relates to the thymic epithelium (mTEC population isolated from 6 week old C57BL/6J mice, sorted on $CD45^{-}Ly51^{lo}MHCII^{hi}EpCAM^{hi}$). Gene expression commons uses a very large number (>10,000) of various microarray datasets as a common reference to overcome the variable sensitivities of each probeset and in that way the intrinsic bias in the expressional range of genes per probeset is normalised, providing reliable gene expression levels among the different microarray experiments (Seita et al., 2012). However, downstream analysis and visualisation capabilities for these microarray profiles is limited. On the other hand, Immgen offers more flexibility in terms of downstream investigation methods (differential expression and gene correlation analysis) of the existent datasets, however, this functionality has been built to fit mostly the nature of the microarrays technology (Heng et al., 2008). Therefore, there is an obvious gap in the availability of whole transcriptome profiling datasets for immune cell populations, in particular representative populations of the thymic stroma, and a further need for appropriate subsequent analysis and normalisation of these RNA-seq datasets to create new hypotheses and answer different biological questions.

In **Chapter 6**, I present an online database for easy access, analysis and visualisation of curated bioinformatics datasets of the thymus (ThymiBase). ThymiBase currently consists of RNA-seq datasets which represent the global transcriptomic profiles of several thymic epithelia cell subsets.

6.2 METHODS

ThymiBase is an interactive web application that I have built under the Shiny package (Chang et al., 2016) in R. The Shiny package provides an online application framework that enables easy building of interactive web application with R. The Shiny package also offers a "reactive" environment that automatically binds inputs and outputs together and in combination with the available prebuild widgets that are available, users can build applications that are responsive, powerful and with a delicate design with minimal effort. Details on ThymiBase's specific inputs – outputs and methods are described in the commented code that launches the database (available in the thesis electronic supplement), while the functionality of ThymiBase's main components is also discussed in the **Results / Discussion** section below.

6.3 RESULTS / DISCUSSION

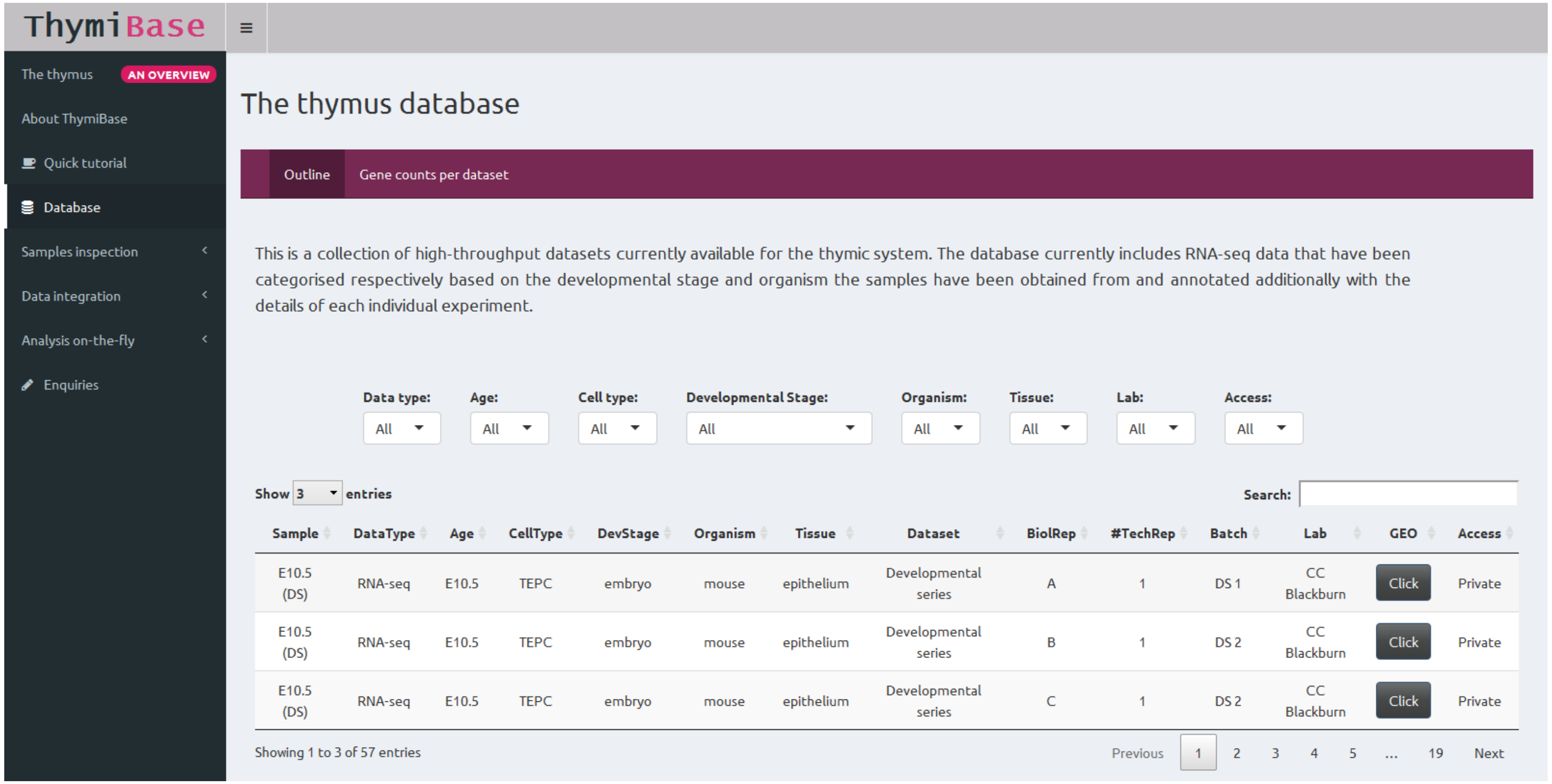
6.3.1 A curated collection of thymic datasets

In order to meet the needs for a data storage and analysis repository, I have generated a bespoke database, named ThymiBase, which is a curated, thymus-specific collection of next generation sequencing datasets (outline shown in **Figure 6.1**). ThymiBase was created to accommodate the bioinformatics data generated during this PhD project and to provide a platform for easy access of post-analysed bioinformatics datasets to other researchers. The graphical environment of this platform allows scientists with a basic bioinformatics knowledge to analyse and integrate already normalised RNA-seq data. Because my PhD studies were part of a bigger European consortium, designated ThymiSTEM (www.thymistem.org), this database could also assist substantially in terms of data sharing across collaborative labs in the future.

At the moment, ThymiBase consists of 7 broad RNA-seq datasets of multiple samples, namely, a Developmental series of fetal TEPC samples [RNA-seq data from **biological triplicates per stage** provided by **Harsh Vaidya**, Blackburn lab (unpublished); see also **Chapter 2**, **section 2.2.1**], a *Foxn1* Allelic series of fetal TEPC data [RNA-seq data from **singular biological samples** provided by **Stephanie Tetelin**, Blackburn lab (unpublished); see also **Chapter 2**, **section 2.2.1**], an E12.5 RBPJ LOF series of fetal TEPC data, an E14.5 RBPJ LOF series of fetal TEPC data, an E14.5 RBPJ GOF series of fetal TEPC data [all RBPJ series were provided by **Dong Liu**; Blackburn lab (Liu et al., 2017, submitted); see also **Chapter 2**, **section 2.2.1.3**], cTEC and mTEC subpopulations from 1 week old (newborn) mice and mTEC subpopulations from 4 week old (adult) mice [merged RNA-seq data from **similarly sorted individual and biological duplicates** obtained and reanalysed from GEO public repository: GEO accession codes **GSE44945 (St-Pierre et al., 2013)** and **GSE53110 (Sansom et al., 2014)**; for details see **Chapter 2**, **section 2.2.1.4**]. Taken together the above datasets are representative of several distinct thymic epithelial subpopulations in mouse and of different developmental stages of the thymic epithelium during mouse development (fetal, newborn, adult).

The raw sequenced files of the aforementioned datasets have been processed following the generic pre-analysis (described in **Chapter 2**, **sections 2.3.1-2.3.2**) and core-analysis (described in **Chapter 2**, **sections 2.3.3.1** and **2.3.3.2**) pipelines to obtain data tables of normalised gene counts per dataset. These normalised data tables can be further inspected using several visualisation methods or they can also be used for downstream types of analysis as described in **section 6.3.2** below.

.


ThymiBase
The thymus
AN OVERVIEW
About ThymiBase
Quick tutorial
Database
Samples inspection
Data integration
Analysis on-the-fly
Enquiries
Outline
Gene counts per dataset
Previous
1
2
3
4
5
...
19
Next

The thymus database

This is a collection of high-throughput datasets currently available for the thymic system. The database currently includes RNA-seq data that have been categorised respectively based on the developmental stage and organism the samples have been obtained from and annotated additionally with the details of each individual experiment.

Data type: All
Age: All
Cell type: All
Developmental Stage: All
Organism: All
Tissue: All
Lab: All
Access: All

Show 3 entries
Search:

Sample	DataType	Age	CellType	DevStage	Organism	Tissue	Dataset	BiolRep	#TechRep	Batch	Lab	GEO	Access
E10.5 (DS)	RNA-seq	E10.5	TEPC	embryo	mouse	epithelium	Developmental series	A	1	DS 1	CC Blackburn	Click	Private
E10.5 (DS)	RNA-seq	E10.5	TEPC	embryo	mouse	epithelium	Developmental series	B	1	DS 2	CC Blackburn	Click	Private
E10.5 (DS)	RNA-seq	E10.5	TEPC	embryo	mouse	epithelium	Developmental series	C	1	DS 2	CC Blackburn	Click	Private

Showing 1 to 3 of 57 entries

Figure 6.1: ThymiBase database panel. ThymiBase's generic preface consists of a sidebar menu (shown on the left in black background) and different main panels (shown always on the right in light blue background) which can be selected from the former sidebar menu. The "Database" main panel here includes a lay summary of the available RNA-seq samples available in ThymiBase (listed in **section 6.3.1**). Samples have been manually annotated for multiple categories (shown here in table header). The experiment to which each sample belongs can be seen from the Dataset column in the table. The normalised gene counts (data normalisation explained in **Chapter 2**, **sections 2.3.3.1** and **2.3.3.2**) for each of these datasets are provided in the "Gene counts per dataset" tab (shown in the purple bar).

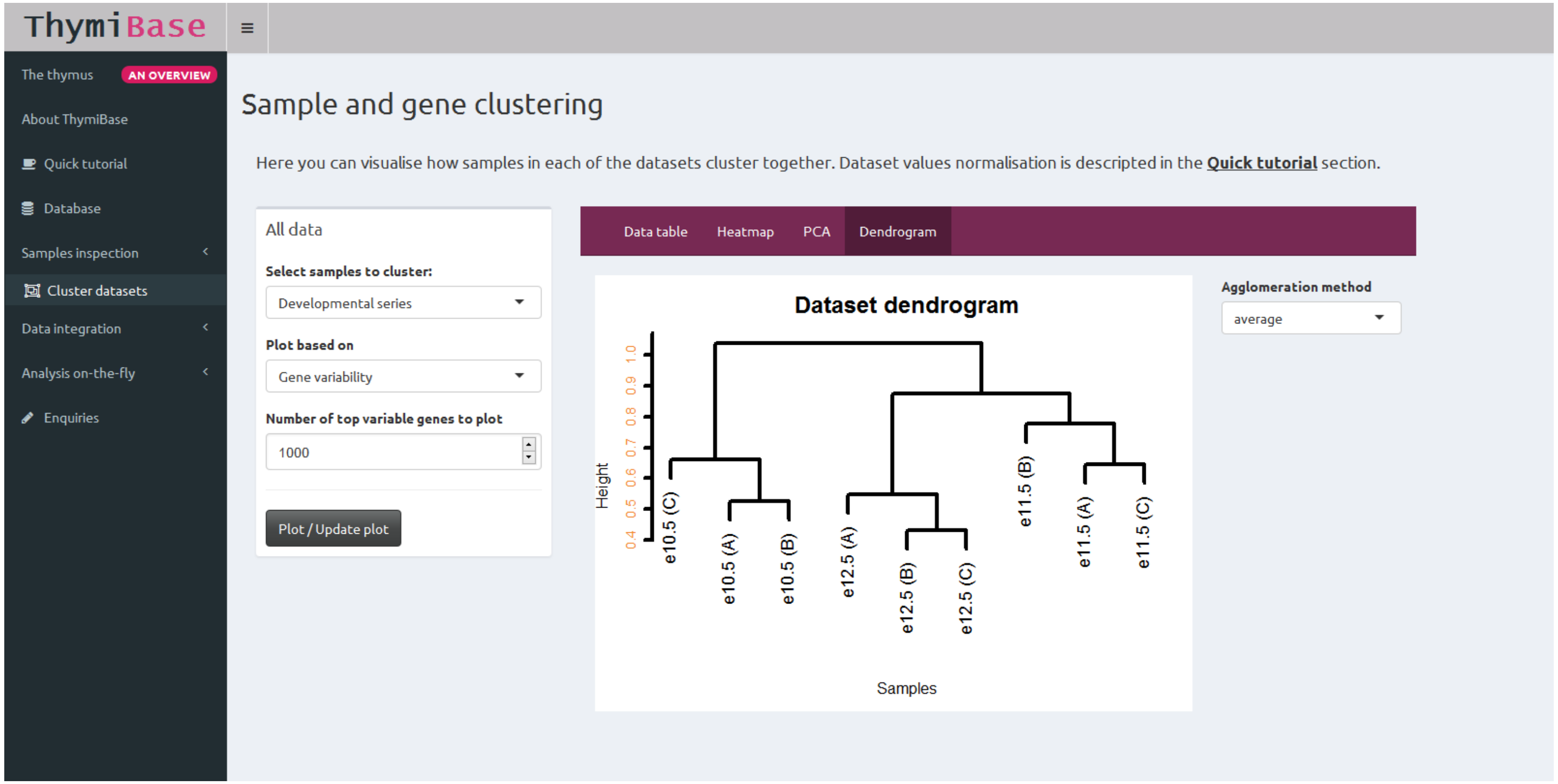
6.3.2 An interface to support easy reanalysis of data

ThymiBase currently provides different types of visualisation and data analysis for advanced investigation of the available (or user-provided) gene expression count tables from RNA-seq datasets. The current preface of ThymiBase is divided into two sections: a side-bar menu that allows navigation through the different panels and a main panel section, where the sidebar-selected main panel is displayed (**Figure 6.1**). In the sidebar menu, the user can select which main panel to be displayed from eight options: The thymus (AN OVERVIEW), About ThymiBase, Quick Tutorial, Database, Samples Inspection, Data Integration, Analysis on-the-fly, Enquiries (see sidebar menu in **Figure 6.1**). To simplify the navigation process of the ThymiBase components (panels) in the sidebar menu, the Samples Inspection, Data Integration and Analysis on-the-fly main panels include a selection of subpanels in relevance to the main panel.

Under Samples Inspection in the side-bar menu exists the "Cluster datasets" option (see snapshot in **Figure 6.2**) which enables users to visualise selected datasets as hierarchical clustering dendrograms, heatmaps, or Principal Component Analysis (PCA) plots. The R respective functions, *hclust*, *pheatmap* and *prcomp* (described in **Chapter 2**, **section 2.3.3.3**) are executed in the background to generate the plot, when the user clicks the Plot/Update plot button. The users get to choose between the 7 broad datasets (listed in **section 6.3.1**) or to upload their own gene counts tables.

The "Analysis on-the-fly" panel in the sidebar menu includes a "Differential Expression" and a "Pathway Enrichment" analysis subpanels that allow users to run a) differential expression analysis (see snapshot in **Figure 6.3**) to investigate significant differences in the expression of genes between sample groups and b) gene set enrichment analysis (see snapshot in **Figure 6.4**) to predict enriched pathways from comparisons between selected sample groups. For the differential expression analysis, the user can choose which samples to compare from one of the provided broad datasets (see **section 6.3.1**), and the analysis executes on-the-fly with the *voom* function from the limma package (Law et al., 2014) in R which runs in the background (for more details on the differential expression analysis with limma *voom*, see **Chapter 2**, **section 2.3.3.5**). For the pathway enrichment analysis, users need to upload a ranked gene list file (similar to GSEA's ranked file, Subramanian et al., 2005) and a pathways database file to search against and identify enriched pathways based on the provided ranked list. The "edited" ConsensusPathDB file contains 2,140 biological pathways in mouse (Kamburov et al., 2011; see also **Chapter 2**, **section 2.3.3.6**) and is provided from a scroll-down menu in the main panel as a pre-set database option (**Figure 6.4**). This analysis also executes on-the-fly, with the *fgsea* function (Sergushichev, 2016) in R to run in the background when the Run GSEA button is pressed. The *fgsea* function produces similar results to the GSEA's '*Run GSEAPreranked'* module analysis (described in **Chapter 2**, **section 2.3.3.6**).

Importantly, although this is a curated database focused on the cell types found into the thymus, most (if not all) of its components, including data inspection and analysis, could be used independently to house and analyse for any externally provided datasets.



Sample and gene clustering

Here you can visualise how samples in each of the datasets cluster together. Dataset values normalisation is descripted in the **Quick tutorial** section.

All data

Select samples to cluster: Developmental series

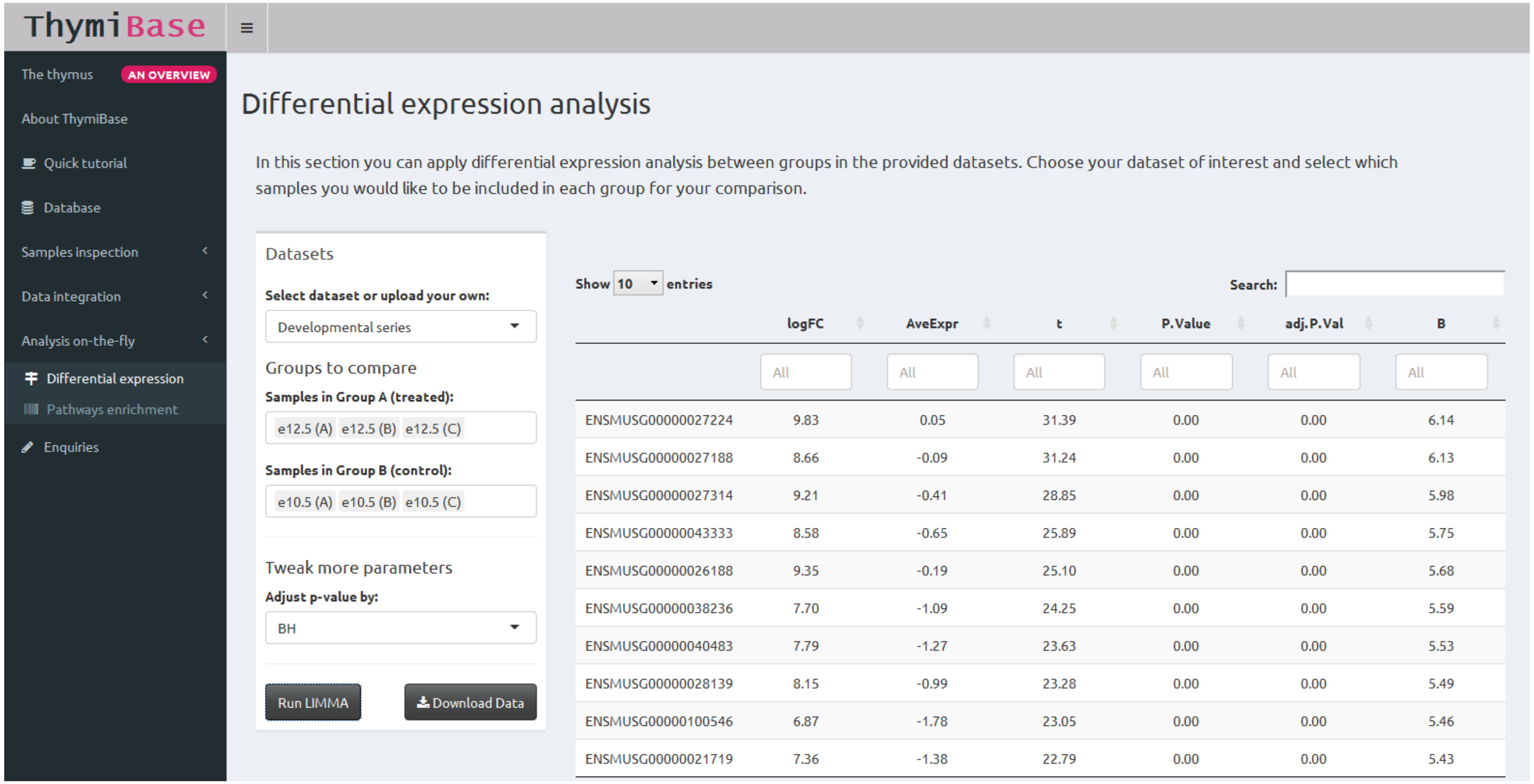
Plot based on Gene variability

Number of top variable genes to plot 1000

Agglomeration method average

ThymiBase
≡
The thymus
AN OVERVIEW
About ThymiBase
Quick tutorial
Database
Samples inspection
Cluster datasets
Data integration
Analysis on-the-fly
Enquiries
Plot / Update plot
Data table
Heatmap
PCA
Dendrogram
Dataset dendrogram
Height
0.4
0.5
0.6
0.7
0.8
0.9
1.0
e10.5 (C)
e10.5 (A)
e10.5 (B)
e12.5 (A)
e12.5 (B)
e12.5 (C)
e11.5 (B)
e11.5 (A)
e11.5 (C)
Samples

Figure 6.2: ThymiBase visualisation panel. In the "Cluster datasets" main panel here (see legend from **Figure 6.1** for ThymiBase's generic preface), users can inspect the available datasets (listed in **section 6.3.1**) from a scroll down menu (includes user-upload option) inside the "All data" box under the "Select samples to cluster" subheader. In the same box under the "Plot based on" subheader, the "Number of top variable genes to plot" can be selected to cluster the datasets based on "Gene variability" with the *rowVars* function from the genefilter package (Gentleman et al., 2017) in R. A number of graphical representations, namely heatmaps, PCA plots or dendrograms (shown in the purple bar) are finally used to cluster the datasets on-the-fly using the *pheatmap*, *hclust*,or *prcomp* functions in R (see **Chapter 2**, **section 2.3.3.3**).


ThymiBase
≡
The thymus
AN OVERVIEW
About ThymiBase
Quick tutorial
Database
Samples inspection
Data integration
Analysis on-the-fly
Differential expression
Pathways enrichment
Enquiries
Run LIMMA
Download Data

Differential expression analysis

In this section you can apply differential expression analysis between groups in the provided datasets. Choose your dataset of interest and select which samples you would like to be included in each group for your comparison.

Datasets

Select dataset or upload your own: Developmental series

Groups to compare

Samples in Group A (treated): e12.5 (A) e12.5 (B) e12.5 (C)

Samples in Group B (control): e10.5 (A) e10.5 (B) e10.5 (C)

Tweak more parameters

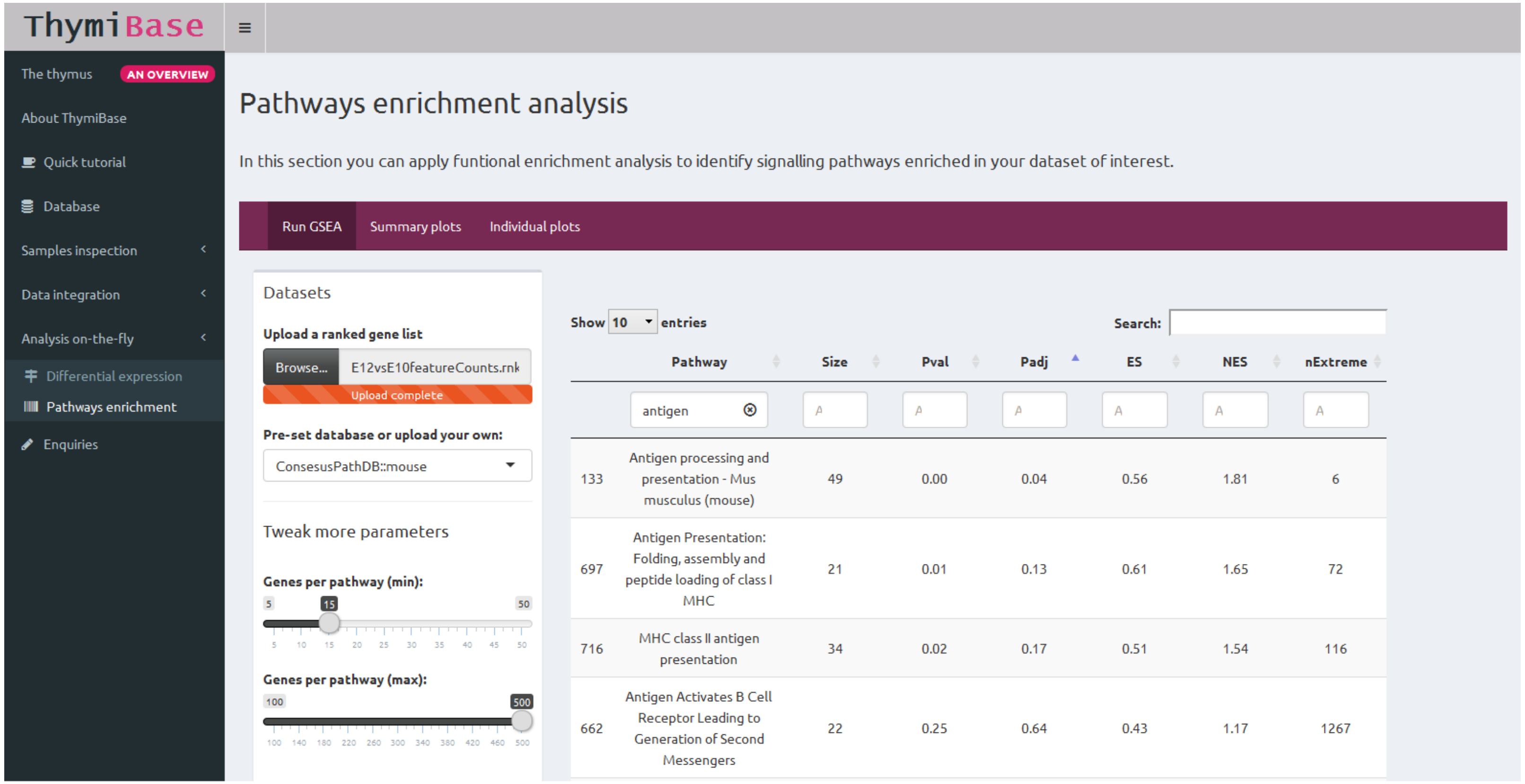
Adjust p-value by: BH

Show 10 entries

Search:

	logFC	AveExpr	t	P.Value	adj.P.Val	B
	All	All	All	All	All	All
ENSMUSG00000027224	9.83	0.05	31.39	0.00	0.00	6.14
ENSMUSG00000027188	8.66	-0.09	31.24	0.00	0.00	6.13
ENSMUSG00000027314	9.21	-0.41	28.85	0.00	0.00	5.98
ENSMUSG00000043333	8.58	-0.65	25.89	0.00	0.00	5.75
ENSMUSG00000026188	9.35	-0.19	25.10	0.00	0.00	5.68
ENSMUSG00000038236	7.70	-1.09	24.25	0.00	0.00	5.59
ENSMUSG00000040483	7.79	-1.27	23.63	0.00	0.00	5.53
ENSMUSG00000028139	8.15	-0.99	23.28	0.00	0.00	5.49
ENSMUSG00000100546	6.87	-1.78	23.05	0.00	0.00	5.46
ENSMUSG00000021719	7.36	-1.38	22.79	0.00	0.00	5.43

Figure 6.3: ThymiBase panel for differential expression analysis. In the "Differential expression" main panel here (see legend from **Figure 6.1** for ThymiBase's generic preface), users can run differential expression analysis between selected groups of samples from the available (listed in **section 6.3.1**) or user-uploaded datasets using the limma *voom* function (Law et al., 2014) in R which executes on-the-fly in the background (for details see **Chapter 2**, **section 2.3.3.5**). The total table of the differentially expressed genes with genes log2 fold change (logFC), average log2 xpression for gene over all genes (AveExp), moderated t-statistic (t), raw *p*-value (P.value), adjusted *p*-value (adj.P.value; calculated by the FDR method; Benjamini and Hochberg, 1995) is displayed in the right of the main panel (for details see limma *topTable* function).


ThymiBase
≡
The thymus
AN OVERVIEW
About ThymiBase
Quick tutorial
Database
Samples inspection
Data integration
Analysis on-the-fly
Differential expression
Pathways enrichment
Enquiries
Run GSEA
Summary plots
Individual plots
Browse...
Upload complete
5
50
5 10 15 20 25 30 35 40 45 50
100
100 140 180 220 260 300 340 380 420 460 500

Pathways enrichment analysis

In this section you can apply funtional enrichment analysis to identify signalling pathways enriched in your dataset of interest.

Datasets

Upload a ranked gene list: E12vsE10featureCounts.rnk

Pre-set database or upload your own: ConsesusPathDB::mouse

Tweak more parameters

Genes per pathway (min): 15

Genes per pathway (max): 500

Show 10 entries

Search:

	Pathway	Size	Pval	Padj	ES	NES	nExtreme
	antigen						
133	Antigen processing and presentation - Mus musculus (mouse)	49	0.00	0.04	0.56	1.81	6
697	Antigen Presentation: Folding, assembly and peptide loading of class I MHC	21	0.01	0.13	0.61	1.65	72
716	MHC class II antigen presentation	34	0.02	0.17	0.51	1.54	116
662	Antigen Activates B Cell Receptor Leading to Generation of Second Messengers	22	0.25	0.64	0.43	1.17	1267

Figure 6.4: ThymiBase panel for pathway enrichment analysis. In the “Pathways enrichment” main panel here (see legend from **Figure 6.1** for ThymiBase’s generic preface), users can run pathway enrichment analysis (see **Chapter 2**, **sections 2.3.3.6**) by uploading a ranked gene list file and selecting the ConsesusPathDB (Kamburov et al., 2011) or uploading their own database in the “Datasets” box. The *fgsea* function (Sergushichev, 2016) in R will run in the background on-the-fly to generate the list of enriched pathways based on the provided inputs. The total pathways table with the number of genes per pathway (size), enrichment *p*-value (Pval), BH-adjusted *p*-value (Padj; Benjamini and Hochberg, 1995), enrichment score (ES; same as in GSEA, Subramanian et al., 2005), normalised enrichment score for the mean value of random samples of the same size (NES) and number of times a random gene set had a more extreme ES (nExtreme) is displayed in the right part of the main panel.

6.4 SUMMARY

In a nutshell, **Chapter 6** demonstrates the development of a curated database of whole genome transcriptome (RNA-seq) datasets related to the thymus that allows easy access, sharing and downstream analysis of these stored datasets. Currently, available expressional profiling databases of immune cells (Gene expression commons; Seita et al., 2012) and Immgen; Heng et al., 2008) are limited to microarray datasets (very few RNA-seq datasets included in Immgen, only one related to the thymic epithelium) while they offer none or mostly microarray-specific downstream analysis options for these datasets. Another advantage of ThymiBase is that it offers multiple data visualisation methods to allow more precise data inspection prior to subsequent data analyses. These methods are provided through a user-friendly graphical interface instead of the R environment to make easier the use of these analysis methods to non-bioinformaticians. Additionally, all incorporated analyses in the ThymiBase platform for all included or user-added datasets are executed "on-the-fly" and can run almost instantly in a laptop. Finally, ThymiBase is a generic tool for inspection and analysis of RNA-seq datasets that is thymus independent (user-added datasets) and can also be of great use to non-thymus specialists. The intention is to convert this platform into a publicly available repository to assist other scientists in gaining further biological insight in the thymus field.

Chapter 7

Concluding remarks

Chapter 7

Concluding remarks

7.1 CONCLUSIONS

This thesis sets out to further our understanding of the transcriptional, regulatory and signalling networks that regulate fetal thymic epithelial cell (TEC) development, via *in silico* analyses. In this chapter, I provide a summary of the main bioinformatics findings that have been presented in **Chapters 3-6**, accompanied by experimental data (provided by others as stated), when necessary, to validate or complement the computational predictions. I also include a section of future work which may be required to solidify or validate some of these outcomes.

7.1.1 A revised serial progression model of TEPCs into cortical and medullary TECs

The main focus of **Chapter 3** was to investigate the serial progression model (reviewed in **Chapter 1**, **section 1.2.3.3**) of the thymic epithelial progenitor cells (TEPCs) towards their functionally distinct differentiated lineages, cortical (c) and medullary (m) TECs, during early mouse development. In **Chapter 3**, I showed that different maximum levels of *Foxn1* in E12.5 TEPCs (from a *Foxn1* Allelic series; raw sequence data provided by **Stephanie Tetelin**, Blackburn lab), effectively capture 'pseudo-timed' snapshots of TEPC natural progression (TEPC Developmental series; raw sequence data provided by **Harsh Vaidya**, Blackburn lab), indicating that the precise expression levels of *Foxn1* play a vital role in regulating TEPC identity. Part of this analysis highlighted the high resemblance of the E10.5 TEPCs with the E12.5 TEPCs, in which *Foxn1* expression has been majorly blocked (E12.5 R/- and E12.5 *Nude* samples). These *Foxn1*-deprived samples are capable of surviving *in vivo* indefinitely and giving rise to all TEC lineages upon *Foxn1* reactivation (Jin et al., 2014). The highly similar transcriptional profiles of the E10.5 TEPCs with the *Foxn1*-deficient samples, in combination with the high levels of PLET1 that E10.5 TEPCs exhibit (compared to the E12.5 TEPCs), strongly suggest that E10.5 TEPCs are more likely to consist a homogeneous TESC/TEPC (TEP/SC) population.

Further comparative analysis of the E10.5 and E12.5 samples from the TEPC Developmental series revealed distinct patterns of expression for the differentially expressed genes when these genes were projected into newborn cTEC and mTEC representative populations (with cTECs expressing higher levels of *Foxn1* compared to mTECs; raw sequence data obtained from GEO: **GSE44945**, St-Pierre et al., 2013 and **GSE53110**, Sansom et al., 2014). The above patterns were still apparent when only FOXN1-dependent genes were selected. These results propose a both activatory and repressive role for FOXN1 in fetal TEPCs (with a repressive role for FOXN1 not to have been acknowledged before) and also suggest that variable levels of *Foxn1* expression per cell can lead to transcriptional heterogeneity. Experimental data by **O'Neill et al.** (2016), showing a graded expression pattern for *Foxn1* across the E13.5 TECs, suggest that *Foxn1* exhibits heterogeneous expression among the E13.5 TECs and hence imply that this heterogeneity may be already apparent among the E12.5 TEPCs. The above observation and a clonal analysis of the E12.5 TEPCs by **Alison M. Farley** (Blackburn lab), in which single E12.5 TEPCs gave rise mostly to cTECs, strongly suggest that E12.5 TEPCs exhibit heterogeneity, with the majority of them to comprise cTEC-fated progenitors. The above evidence comes in contradiction with the E12 TEPC bipotency model proposed by Rossi et al. (2006), which demonstrated E12 TEPC bipotency at the single cell level. However, this inconsistency can be explained by the drop in *Foxn1* levels when TEPCs are kept in culture (Rossi et al., 2006), from which we can argue that E12.5 TEPCs exist in a fated but transient state, which can be pushed back upon a reduction in *Foxn1* levels.

As indicated in the previous paragraph, the majority of the E12.5 TEPCs constitute cTEC-fated progenitor cells (cTEPCs). These cTEPCs would most likely exhibit higher levels of *Foxn1* expression, compared to other existing bipotent or mTEC-fated cells; undifferentiated TEPCs are characterised by low *Foxn1* levels, while the cortical component exhibits higher *Foxn1* expression compared to the medullary component.

The remaining cells among the E12.5 TEPCs are thus characterised by lower levels of *Foxn1* and may constitute undifferentiated TEPCs and/or mTEC-fated progenitor cells (mTEPCs). It is also possible that mTEPCs among the E12.5 TEPCs exhibit a transcriptional profile more similar to the profile of the E10.5 undifferentiated TEPCs (the bipotent progenitors). Notably, the mTEPC population, if existent, must appear prior to the Claudin^{hi}SSEA1^{+} mTEC stem cell population included in the TEC serial progression model summarised by Takahama et al. (2017), since E10.5-E12.5 TEPCs do not yet express the SSEA1 marker.

In summary, in **Chapter 3**, I have used *in silico* analysis, in conjunction with experimental evidence (provided as stated in the text by **others**, **Blackburn lab**), to demonstrate an early TEPC progression model (see **Figure 3.15**) which encompasses the existence of a common bipotent TEPC at day E10.5 of mouse development (when cells still exist at an undifferentiated state and express only low levels of *Foxn1*) and a more immature TEPC type and/or an mTEPC one among cTEPCs at day E12.5 of mouse development. Our data in combination with the Rossi et al. paper (2006) suggest that the E12.5 cTEPCs or (m)TEPCs exist in a transient state, where lineage restriction can be amended by reduction of *Foxn1* expression.

7.1.2 An *in silico* model of the intrinsic and extrinsic cues acting in the early undifferentiated TEPC state: NOTCH signalling in mTEC specification

In **Chapter 4**, I have examined the intrinsic and extrinsic signalling cues that act on the early undifferentiated TEPC state, through pathway enrichment analysis of the expressional differences between TEPCs at timepoints just prior to and after the establishment of the TEC differentiation programme. In this chapter, I generated an unbiased *in silico* model that describes signalling pathway activity during fetal thymus development *in vivo* (**Chapter 4**, **Figure 4.6**), and identified candidate genes which could improve TEP/SC maintenance and proliferation *in vitro*. The predicted signalling pathways from this analysis denote the complete deactivation of the SHH pathway, which normally protects the formation of the parathyroid, and the early specification of the progenitor cells towards the thymic fate. This analysis has also identified the WNT non-canonical ligand, *Wnt5a*, as being strongly down-regulated between E10.5 and E12.5 TEPCs (when cells become more differentiated), the timing of which is consistent with the morphological switch of TEPCs from a monolayer of polarised cells (E10.5) to a more complex and well-defined structure (E12.5) which characterises the thymic system. The high levels of *Wnt5a* expression in E10.5 TEPCs may also reflect a potential role for WNT5A in the regulation of cells self-renewal. Therefore, ablation or reinforcement of WNT5A expression in the early thymic epithelium could reveal interesting insights regarding its biological function *in vivo*, while provision of WNT5A in TEPC culture could aid TEPC survival and expansion *in vitro*. The signalling pathway analysis in **Chapter 4** has predicted down-regulation of ligands that normally activate the FGF and IGF signalling pathways. Since these pathways are active during early thymus development (see **Chapter 1**, **section 1.2.4**), absence of their ligands may represent a TEPC switch from a partially cell autonomous system to a system depending on extracellular provision of ligands from neighbouring cells (for instance from the NCC-derived mesenchyme and potentially incoming haematopoietic progenitor cells).

Among the predicted pathways, this analysis has identified the NOTCH signalling pathway to be highly active in the early undifferentiated TEPCs. Further bioinformatics analysis, accompanied by experimental validation (**Dong Liu**) of NOTCH loss-of-function data (raw sequence data provided by **Dong Liu**, Blackburn lab), has highlighted NOTCH as a potent regulator of mTEC specification in a short but well-defined early TEPC time window (Liu et al., 2017, submitted). Additionally, the RNA-seq analysis of NOTCH gain-of-function data (raw sequence data provided by **Dong Liu**, Blackburn lab), indicated a developmental block in cTECs and a high expression of the PLET1 marker (related to progenitor activity of cells), suggesting that the transient state of TEPCs could be fine-tuned or potentially pushed back to a less differentiated state through altered signalling of NOTCH (**Chapter 4**, **section 4.2.3.3**). Collectively, the bioinformatics analyses that I present in **Chapter 4**, have: a) improved the current picture of the signalling mechanisms that govern maintenance and specification of TEPCs before FOXN1 orchestrates differentiation towards both TEC lineages, b) identified WNT5A as a potential regulator of cell polarity and proliferation in early TEPCs and c) highlighted the importance of NOTCH signalling in early mTEC specification and TEPC maintenance.

7.1.3 FOXN1: A master regulator of the thymic system and a potential pioneer factor

In **Chapter 5**, I explored the role of the central thymus regulator, FOXN1, in TEC differentiation. For this purpose, I applied an integrative analysis of FOXN1-dependent genes in fetal TEPCs, FOXN1 binding sites in newborn TECs (raw sequence data obtained from GEO: **GSE75219**, **Žuklys et al., 2016**) and histone modification marks in fetal TEPCs (raw sequence data provided by **Harsh Vaidya**, Blackburn lab). With this analysis, I aimed to predict high confidence candidate direct FOXN1 targets in fetal TEPCs. In summary, 471 genes were predicted as candidate direct FOXN1 targets in fetal TEPCs. Also, a "most" confident list of 102 genes from the above targets has been shown to be directly regulated by FOXN1 in both fetal TEPCs and newborn cTECs. Genes involved in Cytokine signalling, the Antigen processing and presentation (AP&P) programme, protein degradation, NF-κB signalling, VAV-SHB signalling and NOTCH signalling exist among the 471 candidates. From these, the AP&P programme and NOTCH signalling are further discussed in this thesis. With evidence of FOXN1 binding in the proximal promoters of *Irf1*, *Stat1* and *Ciita* (genes involved in the IRF signalling), it is likely that FOXN1 consists part of the unknown mechanism which regulates expression of the MHC class II AP&P programme in TECs. FOXN1 has also been found to target genes involved in the NOTCH signalling pathway, which has been shown to control mTEC specification and TEPC maintenance in **Chapter 4** (**section 4.3.4**); the FOXN1-NOTCH interplay is overviewed in **section 7.1.4**.

High similarity of the chromatin accessibility landscape in fetal TEPCs with the FOXN1 binding sites in newborn TECs, in combination with known literature (Bleul et al., 2006; Bredenkamp et al., 2014b), have indicated a potential pioneer role for FOXN1 in the thymus. Even though FOXN1 has been shown to act as an activator, 129 of the 471 high confidence candidate direct FOXN1 targets in fetal TEPCs were down-regulated with an increase in *Foxn1* expression, showing that FOXN1 could also directly repress these genes. Thus, FOXN1, as a pioneer factor, could cause gene activation/silencing by binding nearby (co)activators/repressors, and induce or block the expression of proximal genes through them. Further bioinformatics analyses and/ or experimental validation would be required to prove the pioneer potency of FOXN1.

7.1.4 Role of the FOXN1-NOTCH interplay in fetal TEPCs and potential mechanisms of how it deciphers fate specification during TEPC progression

The analysis presented in **Chapter 4** demonstrated an important role for NOTCH signalling in mTEC specification and TEPC maintenance, while the analysis in **Chapter 5** highlighted FOXN1 as a potent regulator of NOTCH in fetal TEPCs. In particular, a NOTCH positive feedback loop has been shown to safeguard the mTEC population (**Chapter 4**, **section 4.2.3.1)**, with blockage of its expression causing a partial or total developmental block in the emergence of mTEC expression (**Chapter 4**, **section 4.2.3**). Additionally, a partial block in cTEC differentiation and reduction of *Foxn1* levels occurred (without an imposition of the mTEC fate) when NOTCH was overexpressed in TEPCs (**Chapter 4**, **section 4.2.3.3)**, suggesting a potential role for NOTCH in TEPC maintenance. FOXN1 has been predicted to down-regulate RBPJ, the main transcriptional mediator of NOTCH, through direct binding on the distal *Rbpj* promoter (**Chapter 5**, **section 5.3.2**), succeeding down-regulation of NOTCH signalling upon TEPC differentiation. The analysis of FOXN1-tagged ChIP-seq data in newborn TECs (**Chapter 5**, **section 5.3.2**) also provided direct binding evidence of FOXN1 at the *Foxn1* locus, indicating that FOXN1 could regulate its own expression. Notably, enforced FOXN1 expression in TEPCs (**Dr Kathy O'Neil**) led to a phenotype similar to the one observed by NOTCH ablation, demonstrating a FOXN1 repressive role over NOTCH. Taken together, the above findings propose an *in silico* regulatory model, in which FOXN1-NOTCH exhibit mutual repressive roles in configuring the balance in early fetal TEPC differentiation in mouse, with NOTCH permitting the emergence of mTECs, and FOXN1 promoting cTEC and mTEC differentiation. The collective model of FOXN1-NOTCH interplay in early TEPC differentiation is summarised in **Figure 7.1**.

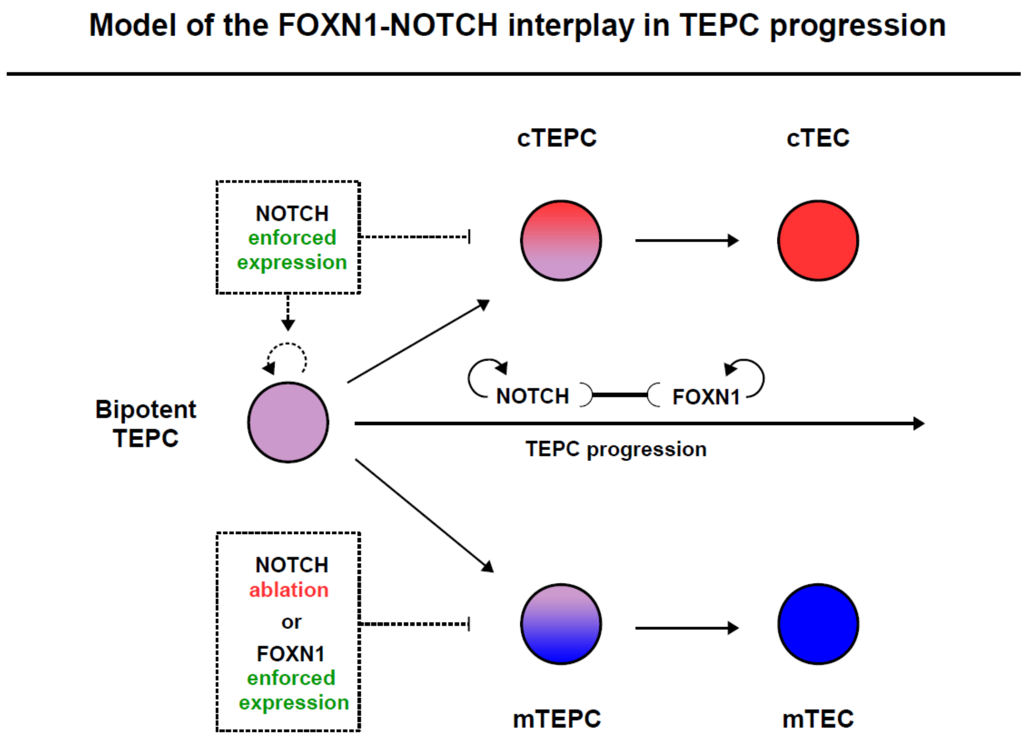

Model of the FOXN1-NOTCH interplay in TEPC progression

Figure 7.1: **Model of the FOXN1-NOTCH interplay in TEPC progression.** In this model, the mutual repressive roles FOXN1-NOTCH, in combination with their auto-regulation (FOXN1 through direct binding of its own promoter, see **Chapter 5**, **section 5.3.2**; and NOTCH via a positive feedback loop through its targets, see **Chapter 4**, **section 4.2.3.1**), are shown to drive normal TEPC progression towards both TEC lineages. NOTCH signalling has been shown to safeguard the mTEC population, with ablation of its expression during early thymus development causing a partial or total developmental block in the mTEC emergence (**Chapter 4**, **section 4.2.3**). In an opposite fashion, enforced expression of FOXN1 has been experimentally demonstrated to also block mTEC progression (**Dr Kathy O'Neil**). NOTCH down-regulation with an increase in FOXN1 expression could be achieved by direct binding of FOXN1 on RBPJ's distal promoter (as shown in **Chapter 5**, **section 5.3.2**). Additionally, NOTCH enforced expression did not enforce mTEC fate; however, it partially blocked cTEC emergence, proposing a role for NOTCH in the maintenance of TEPCs. Together, the above findings contributed to a FOXN1-NOTCH mutual repressive model that configures the balance in early fetal TEC regulation, with NOTCH permitting mTECs emergence and maintaining the TEPC state, and FOXN1 promoting cTEC and mTEC differentiation.

Despite the invaluable role that the FOXN1-NOTCH interplay exhibits in the early progression of TEPCs, the mechanisms by which fate specification is deciphered among the E10.5 TEPCs when ***Foxn1*** expression is initiated are still unknown. This section proposes two alternative versions of how TEPC specification may occur, based on current knowledge and the analyses presented in this thesis. NOTCH signalling is well-known to be involved in the specification of highly complex biological outcomes (Artavanis-Tsakonas, 1999; Fortini, 2009; Lai, 2004), with NOTCH achieving its regulation through different modes of action, namely: lateral inhibition, fate decisions and inductive signalling. Lateral inhibition is used to describe lineage specification of a cell among a group of equivalent cells, while the other two modes take place within non-equivalent cells. Since TEPCs are likely to comprise a very homogeneous population prior to FOXN1 expression, I am going to explore how lateral inhibition could be establishing lineage-restricted fates among these bipotent progenitors. Similar to the well-characterised example in the embryonic neuroepithelium in *Drosophila* (see **Chapter 1**, **section 1.4**), stochastic variation (possibly driven by stochastic up-regulation of *Foxn1*) may lead to the increase of NOTCH ligands (DLL4 is a known FOXN1 direct target), with a transcriptional positive feedback loop magnifying this difference (**Figure 7.2**, **Mechanism A**). Because of their high level of DELTA expression, these TEPCs would not be able to express NOTCH (*cis*-inhibitory interactions; Heitzler and Simpson, 1993), resulting in a salt and pepper pattern where the DELTA-high, signal-sender cells will differentiate into cTECs, while the NOTCH-high, signal-receiver cells will become mTECs. Alternatively, despite E10.5 TEPC homogeneity, because NOTCH signalling receptors are already lowly expressed as early as E10.5 (**Chapter 3**), these cells could exhibit asymmetric expression of these receptor proteins (potentially activated by neighbouring cells), forcing the adjacent TEPCs to become the signal receivers (**Figure 7.2**, **Mechanism B**), and enable higher expression of FOXN1 in those cells, leading again to a salt and pepper pattern as mentioned above.

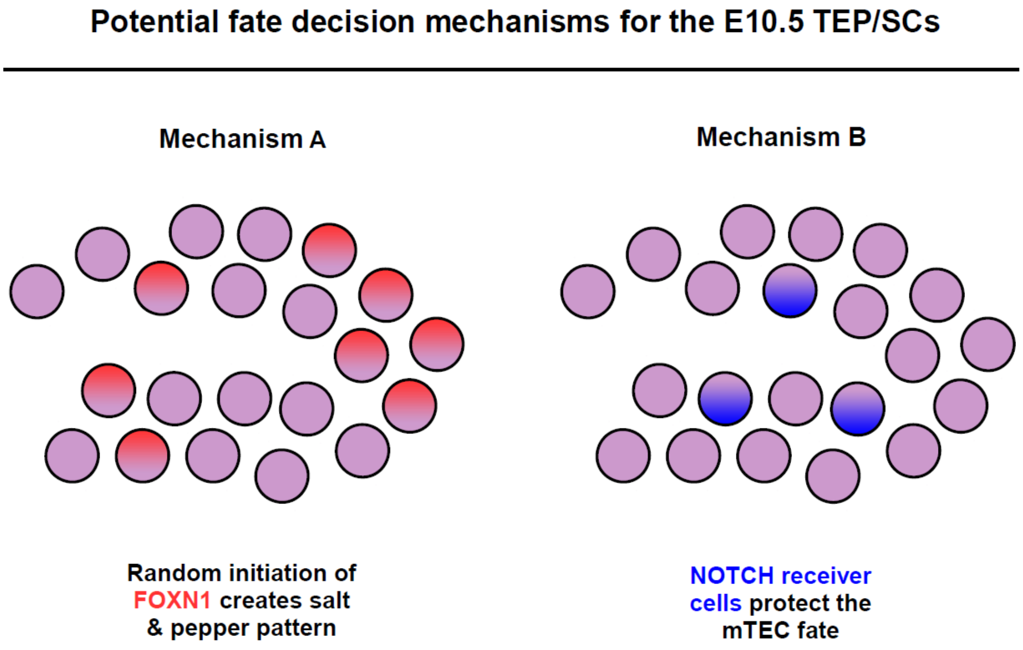

Potential fate decision mechanisms for the E10.5 TEP/SCs
Mechanism A
Mechanism B
Random initiation of FOXN1 creates salt & pepper pattern
NOTCH receiver cells protect the mTEC fate

Figure 7.2: **Potential fate decision mechanisms for the E10.5 TEP/SCs.** The figure presents two alternative mechanisms by which TEPC specification may be resolved within the E10.5 undifferentiated TEPCs. **Mechanism A** illustrates the scenario in which stochastic expression of *Foxn1* may increase the expression levels of the NOTCH ligand (DELTA) in the same cells, forcing their adjacent cells to become the NOTCH signal receivers and creating a salt and pepper pattern with an increase in *Foxn1* expression. Alternatively, **mechanism B** proposes a model in which few of the E10.5 TEPCs already express NOTCH receptors (or other intermediate molecules), which have been activated by neighbouring cells (NOTCH ligands are not expressed yet within the thymic epithelium). NOTCH-receiver cells would possibly safeguard the mTEC fate of themselves, allowing the remaining cells to reach high *Foxn1* levels which will then up-regulate the NOTCH ligands and convert these cells into the NOTCH-sender cells (salt and pepper pattern). In both scenarios of this model, the DELTA-high, signal-sender cells differentiate into cTECs, while the NOTCH-high, signal-receiver cells become mTECs.

7.1.5 ThymiBase

In **Chapter 6**, I introduce the development of ThymiBase, an interactive database with a user-friendly graphical interface (see **Figure 6.1**), to provide a platform for easy access, analysis and integration of curated bioinformatics datasets. This platform can yield further insight for other collaborative or public research groups which may put this information in the context of other data. This database demonstrates an advantage against other currently available immune cell expressional profiling databases, because (unlike the other databases) it grants access to whole transcriptome expression datasets focused on the thymus system and also offers different analysis components specialised in further examining this type of data. ThymiBase's encapsulated analysis components can run on-the-fly and almost instantly on a laptop. While this platform may currently focus on thymus-related data, it is not limited to this data.

7.2 FUTURE WORK

7.2.1 Single cell RNA-seq analysis of E12.5 TEPCs

To validate the revised serial progression model of TEPCs into cortical and medullary TECs proposed in **Chapter 3**, in which *Foxn1* exhibits (a bimodal; O'Neill et al., 2016) heterogeneous expression among the E12.5 TEPCs with a predicted rare mTEPC population to exist among E12.5 cTEC-fated cells, single cell (sc) RNA-seq data from the E12.5 TEPC population need to be generated. These data will provide a per cell quantification of the *Foxn1* levels in E12.5 fetal TEPCs, to confirm or contradict the presence of FOXN1^{low} and FOXN1$^{medium/high}$ TEPCs. Identification of a FOXN1^{low} population among the E12.5 TEPCs will allow better characterisation of this population, to determine if indeed these cells comprise an mTEPC population expressing high levels of Claudins and mTEC markers or a bipotent TEPC population. Prediction of a unique gene signature for the potential (c/m)TEPC populations among the E12.5 fetal TEPCs could then be used to potentially identify the same populations or their derivatives in other stages during thymus development.

7.2.2 Validation of FOXN1's pioneer activity in TEPCs

The analysis in **Chapter 5** in combination with known literature (Bleul et al., 2006; Bredenkamp et al., 2014b) have suggested FOXN1 to demonstrate pioneer activity in TECs. In order to prove FOXN1's pioneer role in the thymus, evidence of chromatin "opening" events due to FOXN1 expression is necessary. To collect this evidence, chromatin accessibility datasets (ChIP-seq, ATAC-seq) need to be generated from a TEC population prior to FOXN1 initiation (at E10.5 developmental stage; undifferentiated TEPCs) and after establishment of its expression (at E12.5 developmental stage; differentiating TEPCs). Since ChIP-seq data from histone modification marks obtained from E12.5 fetal TEPCs and a reference FOXN1 binding site dataset from newborn TECs have already been analysed in **Chapter 5**, generation of the same histone modification marks in E10.5 TEPCs would give a direct measure of determining whether FOXN1 bound sites in fetal TEPCs are in close formation prior to FOXN1 initiation and become accessible with FOXN1 expression. This would be sufficient evidence to propose that FOXN1 is acting as a pioneer factor in TECs.

7.2.3 ThymiBase: additional components and ChIP-seq data

Future plans for ThymiBase include expanding the current downstream analysis components of the platform to allow integration of RNA-seq datasets with a) other available RNA-seq datasets (through correlation to lineage-specific markers) or b) newly added ChIP-seq datasets (through assignment of peaks to the nearest gene) for identification of potential gene targets. Integration of the available (or user-provided) transcriptomics datasets with other expressional datasets can be used to infer sample similarities or similar gene regulation across different data based on co-expression analysis of multiple datasets for specific markers. Selection of lineage-specific markers based on prior knowledge will allow tracing the developmental and differentiation stages of samples, for instance, if two different datasets highly correlate with the same marker (e.g. an early differentiation marker). Relying on markers correlation for dataset integration may also be an intelligent option when batch effect correction is not possible due to confounded factors of the different experiments and cannot be applied. Furthermore, integration of the available (or user-provided) transcriptomics datasets with ChIP-seq added datasets can be used to infer regulation of genes based on transcription factor binding and/or chromatin accessibility.

References

Abu-Issa, R., Smyth, G., Smoak, I., Yamamura, K., and Meyers, E.N. (2002). Fgf8 is required for pharyngeal arch and cardiovascular development in the mouse. Dev. Camb. Engl. *129*, 4613–4625.

Ahlgren, S.C., and Bronner-Fraser, M. (1999). Inhibition of Sonic hedgehog signaling in vivo results in craniofacial neural crest cell death. Curr. Biol. *9*, 1304–1314.

Akiyama, T., Shimo, Y., Yanai, H., Qin, J., Ohshima, D., Maruyama, Y., Asaumi, Y., Kitazawa, J., Takayanagi, H., Penninger, J.M., et al. (2008). The tumor necrosis factor family receptors RANK and CD40 cooperatively establish the thymic medullary microenvironment and self-tolerance. Immunity *29*, 423–437.

Alpdogan, O. (2006). Keratinocyte growth factor (KGF) is required for postnatal thymic regeneration. Blood *107*, 2453–2460.

Althammer, S., Pagès, A., and Eyras, E. (2012). Predictive models of gene regulation from high-throughput epigenomics data. Comp. Funct. Genomics *2012*, 1–13.

Anders, S., and Huber, W. (2010). Differential expression analysis for sequence count data. Genome Biol. *11*, R106.

Anders, S., Pyl, P.T., and Huber, W. (2015). HTSeq--a Python framework to work with high-throughput sequencing data. Bioinforma. Oxf. Engl. *31*, 166–169.

Anderson, G., Jenkinson, E.J., Moore, N.C., and Owen, J.J.T. (1993). MHC class II-positive epithelium and mesenchyme cells are both required for T-cell development in the thymus. Nature *362*, 70–73.

Andersson, R. (2015). Promoter or enhancer, what's the difference? Deconstruction of established distinctions and presentation of a unifying model: Prospects & Overviews. BioEssays *37*, 314–323.

Andrews, S. (2010). FastQC: a quality control tool for high throughput sequence data.

Angelini, C., and Costa, V. (2014). Understanding gene regulatory mechanisms by integrating ChIP-seq and RNA-seq data: statistical solutions to biological problems. Front. Cell Dev. Biol. *2*.

Arnold, J.S., Werling, U., Braunstein, E.M., Liao, J., Nowotschin, S., Edelmann, W., Hebert, J.M., and Morrow, B.E. (2006). Inactivation of Tbx1 in the pharyngeal endoderm results in 22q11DS malformations. Dev. Camb. Engl. *133*, 977–987.

Artavanis-Tsakonas, S. (1999). Notch signaling: Cell fate control and signal integration in development. Science *284*, 770–776.

Atkinson, T.J., and Halfon, M.S. (2014). Regulation of gene expression in the genomic context. Comput. Struct. Biotechnol. J. *9*, e201401001.

Auerbach, R. (1960). Morphogenetic interactions in the development of the mouse thymus gland. Dev. Biol. *2*, 271–284.

Bachiller, D. (2003). The role of chordin/Bmp signals in mammalian pharyngeal development and DiGeorge syndrome. Development *130*, 3567–3578.

Baik, S., Jenkinson, E.J., Lane, P.J.L., Anderson, G., and Jenkinson, W.E. (2013). Generation of both cortical and Aire(+) medullary thymic epithelial compartments from CD205(+) progenitors. Eur. J. Immunol. *43*, 589–594.

Baik, S., Sekai, M., Hamazaki, Y., Jenkinson, W.E., and Anderson, G. (2016). Relb acts downstream of medullary thymic epithelial stem cells and is essential for the emergence of RANK $^{+}$ medullary epithelial progenitors. Eur. J. Immunol. *46*, 857–862.

Bailey, T.L. (2011). DREME: motif discovery in transcription factor ChIP-seq data. Bioinformatics *27*, 1653–1659.

Bailey, T.L., Williams, N., Misleh, C., and Li, W.W. (2006). MEME: discovering and analyzing DNA and protein sequence motifs. Nucleic Acids Res. *34*, W369–W373.

Bain, V.E., Gordon, J., O'Neil, J.D., Ramos, I., Richie, E.R., and Manley, N.R. (2016). Tissue-specific roles for sonic hedgehog signaling in establishing thymus and parathyroid organ fate. Development *143*, 4027–4037.

Balciunaite, G., Keller, M.P., Balciunaite, E., Piali, L., Zuklys, S., Mathieu, Y.D., Gill, J., Boyd, R., Sussman, D.J., and Holländer, G.A. (2002). Wnt glycoproteins regulate the expression of FoxN1, the gene defective in nude mice. Nat. Immunol. *3*, 1102–1108.

Bannister, A.J., and Kouzarides, T. (2011). Regulation of chromatin by histone modifications. Cell Res. *21*, 381–395.

Barbarulo, A., Lau, C.-I., Mengrelis, K., Ross, S., Solanki, A., Saldaña, J., and Crompton, T. (2016). Hedgehog signalling in the embryonic mouse thymus. J. Dev. Biol. *4*, 22.

Barrett, T., Wilhite, S.E., Ledoux, P., Evangelista, C., Kim, I.F., Tomashevsky, M., Marshall, K.A., Phillippy, K.H., Sherman, P.M., Holko, M., et al. (2013). NCBI GEO: archive for functional genomics data sets--update. Nucleic Acids Res. *41*, D991-995.

Barski, A., Cuddapah, S., Cui, K., Roh, T.-Y., Schones, D.E., Wang, Z., Wei, G., Chepelev, I., and Zhao, K. (2007). High-resolution profiling of histone methylations in the human genome. Cell *129*, 823–837.

Baruzzo, G., Hayer, K.E., Kim, E.J., Di Camillo, B., FitzGerald, G.A., and Grant, G.R. (2016). Simulation-based comprehensive benchmarking of RNA-seq aligners. Nat. Methods *14*, 135–139.

Bate, M., and Rushton, E. (1993). Myogenesis and muscle patterning in Drosophila. C. R. Acad. Sci. III *316*, 1047–1061.

Baxter, R.M., and Brissette, J.L. (2002). Role of the nude gene in epithelial terminal differentiation. J. Invest. Dermatol. *118*, 303–309.

Benjamini, Y., and Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *57*, 289–300.

Bennett, A.R., Farley, A., Blair, N.F., Gordon, J., Sharp, L., and Blackburn, C.C. (2002). Identification and characterization of thymic epithelial progenitor cells. Immunity *16*, 803–814.

Berger, M.F., and Bulyk, M.L. (2009). Universal protein-binding microarrays for the comprehensive characterization of the DNA-binding specificities of transcription factors. Nat. Protoc. *4*, 393–411.

Berger, M.F., Philippakis, A.A., Qureshi, A.M., He, F.S., Estep, P.W., and Bulyk, M.L. (2006). Compact, universal DNA microarrays to comprehensively determine transcription-factor binding site specificities. Nat. Biotechnol. *24*, 1429–1435.

Blackburn, C.C., and Manley, N.R. (2004). Developing a new paradigm for thymus organogenesis. Nat. Rev. Immunol. *4*, 278–289.

Blackburn, C.C., Augustine, C.L., Li, R., Harvey, R.P., Malin, M.A., Boyd, R.L., Miller, J.F., and Morahan, G. (1996). The nu gene acts cell-autonomously and is required for differentiation of thymic epithelial progenitors. Proc. Natl. Acad. Sci. U. S. A. *93*, 5742–5746.

Blake, J.A., and Ziman, M.R. (2014). Pax genes: regulators of lineage specification and progenitor cell maintenance. Development *141*, 737–751.

Bleul, C.C., and Boehm, T. (2005). BMP signaling is required for normal thymus development. J. Immunol. Baltim. Md 1950 *175*, 5213–5221.

Bleul, C.C., Corbeaux, T., Reuter, A., Fisch, P., Mönting, J.S., and Boehm, T. (2006). Formation of a functional thymus initiated by a postnatal epithelial progenitor cell. Nature *441*, 992–996.

Boehm, T., and Swann, J.B. (2013). Thymus involution and regeneration: two sides of the same coin? Nat. Rev. Immunol. *13*, 831–838.

Boehm, U., Klamp, T., Groot, M., and Howard, J.C. (1997). CELLULAR RESPONSES TO INTERFERON-γ. Annu. Rev. Immunol. *15*, 749–795.

Bolger, A.M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics *30*, 2114–2120.

Bonn, S., Zinzen, R.P., Girardot, C., Gustafson, E.H., Perez-Gonzalez, A., Delhomme, N., Ghavi-Helm, Y., Wilczyński, B., Riddell, A., and Furlong, E.E.M. (2012). Tissue-specific analysis of chromatin state identifies temporal signatures of enhancer activity during embryonic development. Nat. Genet. *44*, 148–156.

Boyd, R.L., Tucek, C.L., Godfrey, D.I., Izon, D.J., Wilson, T.J., Davidson, N.J., Bean, A.G.D., Ladyman, H.M., Ritter, M.A., and Hugo, P. (1993). The thymic microenvironment. Immunol. Today *14*, 445–459.

Brack, A.S., and Rando, T.A. (2012). Tissue-specific stem cells: Lessons from the skeletal muscle satellite cell. Cell Stem Cell *10*, 504–514.

Bray, S.J. (2006). Notch signalling: a simple pathway becomes complex. Nat. Rev. Mol. Cell Biol. *7*, 678–689.

Bray, N.L., Pimentel, H., Melsted, P., and Pachter, L. (2016). Near-optimal probabilistic RNA-seq quantification. Nat. Biotechnol. *34*, 525–527.

Bredenkamp, N., Nowell, C.S., and Blackburn, C.C. (2014a). Regeneration of the aged thymus by a single transcription factor. Development *141*, 1627–1637.

Bredenkamp, N., Ulyanchenko, S., O'Neill, K.E., Manley, N.R., Vaidya, H.J., and Blackburn, C.C. (2014b). An organized and functional thymus generated from FOXN1-reprogrammed fibroblasts. Nat. Cell Biol. *16*, 902–908.

Bredenkamp, N., Jin, X., Liu, D., O'Neill, K.E., Manley, N.R., and Blackburn, C.C. (2015). Construction of a functional thymic microenvironment from pluripotent stem cells for the induction of central tolerance. Regen. Med. *10*, 317–329.

Briscoe, J., Rogers, N.C., Witthuhn, B.A., Watling, D., Harpur, A.G., Wilks, A.F., Stark, G.R., Ihle, J.N., and Kerr, I.M. (1996). Kinase-negative mutants of JAK1 can sustain interferon-gamma-inducible gene expression but not an antiviral state. EMBO J. *15*, 799–809.

Brownell, J.E., Zhou, J., Ranalli, T., Kobayashi, R., Edmondson, D.G., Roth, S.Y., and Allis, C.D. (1996). Tetrahymena histone acetyltransferase A: a homolog to yeast Gcn5p linking histone acetylation to gene activation. Cell *84*, 843–851.

Brunk, F., Augustin, I., Meister, M., Boutros, M., and Kyewski, B. (2015). Thymic epithelial cells are a nonredundant source of Wnt ligands for thymus development. J. Immunol. *195*, 5261–5271.

Buczacki, S.J.A., Zecchini, H.I., Nicholson, A.M., Russell, R., Vermeulen, L., Kemp, R., and Winton, D.J. (2013). Intestinal label-retaining cells are secretory precursors expressing Lgr5. Nature *495*, 65–69.

Buono, M., Facchini, R., Matsuoka, S., Thongjuea, S., Waithe, D., Luis, T.C., Giustacchini, A., Besmer, P., Mead, A.J., Jacobsen, S.E.W., et al. (2016). A dynamic niche provides Kit ligand in a stage-specific manner to the earliest thymocyte progenitors. Nat. Cell Biol. *18*, 157–167.

Cai, J., Lee, J., Kopan, R., and Ma, L. (2009). Genetic interplays between Msx2 and Foxn1 are required for Notch1 expression and hair shaft differentiation. Dev. Biol. *326*, 420–430.

Calderón, L., and Boehm, T. (2012). Synergistic, context-dependent, and hierarchical functions of epithelial components in thymic microenvironments. Cell *149*, 159–172.

Cardarelli, P.M., and Pierschbacher, M.D. (1986). T-lymphocyte differentiation and the extracellular matrix: identification of a thymocyte subset that attaches specifically to fibronectin. Proc. Natl. Acad. Sci. U. S. A. *83*, 2647–2651.

Cardarelli, P.M., Crispe, I.N., and Pierschbacher, M.D. (1988). Preferential expression of fibronectin receptors on immature thymocytes. J. Cell Biol. *106*, 2183–2190.

Chai, X., Nagarajan, S., Kim, K., Lee, K., and Choi, J.K. (2013). Regulation of the boundaries of accessible chromatin. PLoS Genet. *9*, e1003778.

Chang, W., Cheng, J., Allaire, J., Xie, Y., and McPherson, J. (2016). shiny: Web application framework for R. R package.

Chen, T., and Dent, S.Y.R. (2014). Chromatin modifiers and remodellers: regulators of cellular differentiation. Nat. Rev. Genet. *15*, 93–106.

Chen, L., Xiao, S., and Manley, N.R. (2009). Foxn1 is required to maintain the postnatal thymic microenvironment in a dosage-sensitive manner. Blood *113*, 567–574.

Chen, Y., Negre, N., Li, Q., Mieczkowska, J.O., Slattery, M., Liu, T., Zhang, Y., Kim, T.-K., He, H.H., Zieba, J., et al. (2012). Systematic evaluation of factors influencing ChIP-seq fidelity. Nat. Methods *9*, 609–614.

Cheng, C., Yan, K.-K., Yip, K.Y., Rozowsky, J., Alexander, R., Shou, C., and Gerstein, M. (2011). A statistical framework for modeling gene expression using chromatin features and application to modENCODE datasets. Genome Biol. *12*, R15.

Cheng, C., Alexander, R., Min, R., Leng, J., Yip, K.Y., Rozowsky, J., Yan, K.-K., Dong, X., Djebali, S., Ruan, Y., et al. (2012). Understanding transcriptional regulation by integrative analysis of transcription factor binding data. Genome Res. *22*, 1658–1667.

Cheng, L., Guo, J., Sun, L., Fu, J., Barnes, P.F., Metzger, D., Chambon, P., Oshima, R.G., Amagai, T., and Su, D.-M. (2010). Postnatal tissue-specific disruption of transcription factor Foxn1 triggers acute thymic atrophy. J. Biol. Chem. *285*, 5836–5847.

Cheng, S.-L., Shao, J.-S., Charlton-Kachigian, N., Loewy, A.P., and Towler, D.A. (2003). Msx2 promotes osteogenesis and suppresses adipogenic differentiation of multipotent mesenchymal progenitors. J. Biol. Chem. *278*, 45969–45977.

Cheung, M.-S., Down, T.A., Latorre, I., and Ahringer, J. (2011). Systematic bias in high-throughput sequencing data and its correction by BEADS. Nucleic Acids Res. *39*, e103.

Chidgey, A.P., Layton, D., Trounson, A., and Boyd, R.L. (2008). Tolerance strategies for stem-cell-based therapies. Nature *453*, 330–337.

Chojnowski, J.L., Masuda, K., Trau, H.A., Thomas, K., Capecchi, M., and Manley, N.R. (2014). Multiple roles for HOXA3 in regulating thymus and parathyroid differentiation and morphogenesis in mouse. Development *141*, 3697–3708.

Chu, Y.-W., Schmitz, S., Choudhury, B., Telford, W., Kapoor, V., Garfield, S., Howe, D., and Gress, R.E. (2008). Exogenous insulin-like growth factor 1 enhances thymopoiesis predominantly through thymic epithelial cell expansion. Blood *112*, 2836–2846.

Cirillo, L.A. (1998). Binding of the winged-helix transcription factor HNF3 to a linker histone site on the nucleosome. EMBO J. *17*, 244–254.

Cirillo, L.A., Lin, F.R., Cuesta, I., Friedman, D., Jarnik, M., and Zaret, K.S. (2002). Opening of compacted chromatin by early developmental transcription factors HNF3 (FoxA) and GATA-4. Mol. Cell *9*, 279–289.

Clayton, E., Doupé, D.P., Klein, A.M., Winton, D.J., Simons, B.D., and Jones, P.H. (2007). A single type of progenitor cell maintains normal epidermis. Nature *446*, 185–189.

Clevers, H. (2013). The intestinal crypt, a prototype stem cell compartment. Cell *154*, 274–284.

Clevers, H., and Watt, F.M. (2018). Defining adult stem cells by function, not by phenotype. Annu. Rev. Biochem. *87*.

Cock, P.J.A., Fields, C.J., Goto, N., Heuer, M.L., and Rice, P.M. (2010). The Sanger FASTQ file format for sequences with quality scores, and the Solexa/Illumina FASTQ variants. Nucleic Acids Res. *38*, 1767–1771.

Coleman, W. (1969). Galen on the Usefulness of the Parts of the Body. (Peri Chreias Morion.)(De Usu Partium.) Translated from the Greek with an introduction and commentary by Margaret Tallmadge May. Cornell University Press, Ithaca, N. Y., 1968. 2 vols., boxed, xiv + 804 pp. $25. Cornell Publications in the History of Science. Science *163*, 1439–1440.

Comes, S., Gagliardi, M., Laprano, N., Fico, A., Cimmino, A., Palamidessi, A., De Cesare, D., De Falco, S., Angelini, C., Scita, G., et al. (2013). L-Proline induces a mesenchymal-like invasive program in embryonic stem cells by remodeling H3K9 and H3K36 methylation. Stem Cell Rep. *1*, 307–321.

Compeau, P.E.C., Pevzner, P.A., and Tesler, G. (2011). How to apply de Bruijn graphs to genome assembly. Nat. Biotechnol. *29*, 987–991.

Conway, S.J., Henderson, D.J., and Copp, A.J. (1997). Pax3 is required for cardiac neural crest migration in the mouse: evidence from the splotch (Sp2H) mutant. Dev. Camb. Engl. *124*, 505–514.

Cook, A.M. (2010). Proliferation and lineage potential in fetal thymic epithelial progenitor cells.

Cooper, G.M., and Hausman, R.E. (2009). Regulation of transcription in eukaryotes. In the cell : A molecular approach (5th Edition), p.

Corbeaux, T., Hess, I., Swann, J.B., Kanzler, B., Haas-Assenbaum, A., and Boehm, T. (2010). Thymopoiesis in mice depends on a Foxn1-positive thymic epithelial cell lineage. Proc. Natl. Acad. Sci. *107*, 16613–16618.

Cordier, A.C., and Haumont, S.M. (1980). Development of thymus, parathyroids, and ultimo-branchial bodies in NMRI and nude mice. Am. J. Anat. *157*, 227–263.

Cordier, A.C., and Heremans, J.F. (1975). Nude mouse embryo: ectodermal nature of the primordial thymic defect. Scand. J. Immunol. *4*, 193–196.

Cotsarelis, G., Sun, T.T., and Lavker, R.M. (1990). Label-retaining cells reside in the bulge area of pilosebaceous unit: implications for follicular stem cells, hair cycle, and skin carcinogenesis. Cell *61*, 1329–1337.

Creyghton, M.P., Cheng, A.W., Welstead, G.G., Kooistra, T., Carey, B.W., Steine, E.J., Hanna, J., Lodato, M.A., Frampton, G.M., Sharp, P.A., et al. (2010). Histone H3K27ac separates active from poised enhancers and predicts developmental state. Proc. Natl. Acad. Sci. *107*, 21931–21936.

Decker, T., Kovarik, P., and Meinke, A. (1997). GAS Elements: A few nucleotides with a major impact on cytokine-induced gene expression. J. Interferon Cytokine Res. *17*, 121–134.

Depreter, M.G.L., Blair, N.F., Gaskell, T.L., Nowell, C.S., Davern, K., Pagliocca, A., Stenhouse, F.H., Farley, A.M., Fraser, A., Vrana, J., et al. (2008). Identification of Plet-1 as a specific marker of early thymic epithelial progenitor cells. Proc. Natl. Acad. Sci. *105*, 961–966.

Dietrich, S., and Gruss, P. (1995). Undulated phenotypes suggest a role of Pax-1 for the development of vertebral and extravertebral structures. Dev. Biol. *167*, 529–548.

Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., and Gingeras, T.R. (2013). STAR: ultrafast universal RNA-seq aligner. Bioinforma. Oxf. Engl. *29*, 15–21.

Donati, G., and Watt, F.M. (2015). Stem cell heterogeneity and plasticity in epithelia. Cell Stem Cell *16*, 465–476.

Dong, Y., Rohn, W.M., and Benveniste, E.N. (1999). IFN-gamma regulation of the type IV class II transactivator promoter in astrocytes. J. Immunol. Baltim. Md 1950 *162*, 4731–4739.

Doupe, D.P., Alcolea, M.P., Roshan, A., Zhang, G., Klein, A.M., Simons, B.D., and Jones, P.H. (2012). A single progenitor population switches behavior to maintain and repair esophageal epithelium. Science *337*, 1091–1093.

Dudakov, J.A., Hanash, A.M., Jenq, R.R., Young, L.F., Ghosh, A., Singer, N.V., West, M.L., Smith, O.M., Holland, A.M., Tsai, J.J., et al. (2012). Interleukin-22 drives endogenous thymic regeneration in mice. Science *336*, 91–95.

Duijvestijn, A.M., and Hoefsmit, E.C. (1981). Ultrastructure of the rat thymus: the micro-environment of T-lymphocyte maturation. Cell Tissue Res. *218*, 279–292.

Dunham, I., Kundaje, A., Aldred, S.F., Collins, P.J., Davis, C.A., Doyle, F., Epstein, C.B., Frietze, S., Harrow, J., Kaul, R., et al. (2012). An integrated encyclopedia of DNA elements in the human genome. Nature *489*, 57–74.

Dunn, S.-J., Martello, G., Yordanov, B., Emmott, S., and Smith, A.G. (2014). Defining an essential transcription factor program for naïve pluripotency. Science *344*, 1156–1160.

Dynan, W.S., and Tjian, R. (1983). The promoter-specific transcription factor Sp1 binds to upstream sequences in the SV40 early promoter. Cell *35*, 79–87.

Engel, C., Sainsbury, S., Cheung, A.C., Kostrewa, D., and Cramer, P. (2013). RNA polymerase I structure and transcription regulation. Nature *502*, 650–655.

English, A.C., Richards, S., Han, Y., Wang, M., Vee, V., Qu, J., Qin, X., Muzny, D.M., Reid, J.G., Worley, K.C., et al. (2012). Mind the gap: Upgrading genomes with Pacific Biosciences RS long-read sequencing technology. PLoS ONE *7*, e47768.

Epstein, J.A., Li, J., Lang, D., Chen, F., Brown, C.B., Jin, F., Lu, M.M., Thomas, M., Liu, E., Wessels, A., et al. (2000). Migration of cardiac neural crest cells in Splotch embryos. Dev. Camb. Engl. *127*, 1869–1878.

van Ewijk, W., Shores, E.W., and Singer, A. (1994). Crosstalk in the mouse thymus. Immunol. Today *15*, 214–217.

van Ewijk, W., Wang, B., Hollander, G., Kawamoto, H., Spanopoulou, E., Itoi, M., Amagai, T., Jiang, Y.-F., Germeraad, W.T.V., Chen, W.-F., et al. (1999). Thymic microenvironments, 3-D versus 2-D? Semin. Immunol. *11*, 57–64.

van Ewijk, W., Hollander, G., Terhorst, C., and Wang, B. (2000). Stepwise development of thymic microenvironments in vivo is regulated by thymocyte subsets. Development *127*, 1583.

Fanu, W.R.L. (1963). The origin of medical terms, by Henry Alan Skinner, second edition, Baltimore, The Williams and Wilkins Company, 1961, pp. x, 437, illus., $12.50. Med. Hist. *7*, 94–95.

Farr, A.G., and Anderson, S.K. (1985). Epithelial heterogeneity in the murine thymus: fucose-specific lectins bind medullary epithelial cells. J. Immunol. Baltim. Md 1950 *134*, 2971–2977.

Feng, X., Grossman, R., and Stein, L. (2011). PeakRanger: a cloud-enabled peak caller for ChIP-seq data. BMC Bioinformatics *12*, 139.

Fior, R., and Henrique, D. (2009). “Notch-Off”: a perspective on the termination of Notch signalling. Int. J. Dev. Biol. *53*, 1379–1384.

Fischbein, A. (1984). Medical meanings: A glossary of word origins. William S. Haubrich, MD. Harvest/Harcourt Brace Jovanovich, New York, 1984, 304 pages, $9.95. Am. J. Ind. Med. *6*, 399–399.

Flanagan, S.P. (1966). “Nude”, a new hairless gene with pleiotropic effects in the mouse. Genet. Res. *8*, 295–309.

Fortini, M.E. (2009). Notch signaling: The core pathway and its posttranslational regulation. Dev. Cell *16*, 633–647.

Frank, D.U., Fotheringham, L.K., Brewer, J.A., Muglia, L.J., Tristani-Firouzi, M., Capecchi, M.R., and Moon, A.M. (2002). An Fgf8 mouse mutant phenocopies human 22q11 deletion syndrome. Dev. Camb. Engl. *129*, 4591–4603.

Franz, T. (1989). Persistent truncus arteriosus in the Splotch mutant mouse. Anat. Embryol. (Berl.) *180*, 457–464.

Fuchs, E., and Green, H. (1980). Changes in keratin gene expression during terminal differentiation of the keratinocyte. Cell *19*, 1033–1042.

Furey, T.S. (2012). ChIP–seq and beyond: new and improved methodologies to detect and characterize protein–DNA interactions. Nat. Rev. Genet. *13*, 840–852.

Gardiner, J.R., Jackson, A.L., Gordon, J., Lickert, H., Manley, N.R., and Basson, M.A. (2012). Localised inhibition of FGF signalling in the third pharyngeal pouch is required for normal thymus and parathyroid organogenesis. Development *139*, 3456–3466.

Gaston, K., and Jayaraman, P.-S. (2003). Transcriptional repression in eukaryotes: repressors and repression mechanisms. Cell. Mol. Life Sci. CMLS *60*, 721–741.

Gentleman, R., Carey, V., Huber, W., and Hahne, F. (2017). Genefilter: Methods for filtering genes from high-throughput experiments.

Gerstein, M.B., Kundaje, A., Hariharan, M., Landt, S.G., Yan, K.-K., Cheng, C., Mu, X.J., Khurana, E., Rozowsky, J., Alexander, R., et al. (2012). Architecture of the human regulatory network derived from ENCODE data. Nature *489*, 91–100.

Gill, J., Malin, M., Holländer, G.A., and Boyd, R. (2002). Generation of a complete thymic microenvironment by MTS24+ thymic epithelial cells. Nat. Immunol. *3*, 635–642.

Glaus, P., Honkela, A., and Rattray, M. (2012). Identifying differentially expressed transcripts from RNA-seq data with biological variation. Bioinformatics *28*, 1721–1728.

Goldfarb, Y., Kadouri, N., Levi, B., Sela, A., Herzig, Y., Cohen, R.N., Hollenberg, A.N., and Abramson, J. (2016). HDAC3 is a master regulator of mTEC development. Cell Rep. *15*, 651–665.

Gordon, J., and Manley, N.R. (2011). Mechanisms of thymus organogenesis and morphogenesis. Dev. Camb. Engl. *138*, 3865–3878.

Gordon, J., Bennett, A.R., Blackburn, C.C., and Manley, N.R. (2001). Gcm2 and Foxn1 mark early parathyroid- and thymus-specific domains in the developing third pharyngeal pouch. Mech. Dev. *103*, 141–143.

Gordon, J., Wilson, V.A., Blair, N.F., Sheridan, J., Farley, A., Wilson, L., Manley, N.R., and Blackburn, C.C. (2004). Functional evidence for a single endodermal origin for the thymic epithelium. Nat. Immunol. *5*, 546–553.

Gordon, J., Patel, S.R., Mishina, Y., and Manley, N.R. (2010). Evidence for an early role for BMP4 signaling in thymus and parathyroid morphogenesis. Dev. Biol. *339*, 141–154.

Gosselet, F.P., Magnaldo, T., Culerrier, R.M., Sarasin, A., and Ehrhart, J.-C. (2007). BMP2 and BMP6 control p57Kip2 expression and cell growth arrest/terminal differentiation in normal primary human epidermal keratinocytes. Cell. Signal. *19*, 731–739.

Grabherr, M.G., Haas, B.J., Yassour, M., Levin, J.Z., Thompson, D.A., Amit, I., Adiconis, X., Fan, L., Raychowdhury, R., Zeng, Q., et al. (2011). Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nat. Biotechnol. *29*, 644–652.

Grabowska, A.I., and Wilanowski, T. (2017). FOXN1 transcription factor in epithelial growth and wound healing. Mol. Cell. Biol. *37*, e00110-17.

Gray, D., Abramson, J., Benoist, C., and Mathis, D. (2007). Proliferative arrest and rapid turnover of thymic epithelial cells expressing Aire. J. Exp. Med. *204*, 2521–2528.

Gray, D.H.D., Chidgey, A.P., and Boyd, R.L. (2002). Analysis of thymic stromal cell populations using flow cytometry. J. Immunol. Methods *260*, 15–28.

Greenlund, A.C., Morales, M.O., Viviano, B.L., Yan, H., Krolewski, J., and Schreiber, R.D. (1995). Stat recruitment by tyrosine-phosphorylated cytokine receptors: An ordered reversible affinity-driven process. Immunity *2*, 677–687.

Greenwald, I. (1998). LIN-12/Notch signaling: lessons from worms and flies. Genes Dev. *12*, 1751–1762.

Greenwald, I., and Rubin, G.M. (1992). Making a difference: The role of cell-cell interactions in establishing separate identities for equivalent cells. Cell *68*, 271–281.

Grevellec, A., Graham, A., and Tucker, A.S. (2011). Shh signalling restricts the expression of Gcm2 and controls the position of the developing parathyroids. Dev. Biol. *353*, 194–205.

Griffith, A.V., Cardenas, K., Carter, C., Gordon, J., Iberg, A., Engleka, K., Epstein, J.A., Manley, N.R., and Richie, E.R. (2009). Increased thymus- and decreased parathyroid-fated organ domains in Splotch mutant embryos. Dev. Biol. *327*, 216–227.

Grün, D., and van Oudenaarden, A. (2015). Design and analysis of single-cell sequencing experiments. Cell *163*, 799–810.

Gruver, A., Hudson, L., and Sempowski, G. (2007). Immunosenescence of ageing. J. Pathol. *211*, 144–156.

Gualdi, R., Bossard, P., Zheng, M., Hamada, Y., Coleman, J.R., and Zaret, K.S. (1996). Hepatic specification of the gut endoderm in vitro: cell signaling and transcriptional control. Genes Dev. *10*, 1670–1682.

Guo, Y., Papachristoudis, G., Altshuler, R.C., Gerber, G.K., Jaakkola, T.S., Gifford, D.K., and Mahony, S. (2010). Discovering homotypic binding events at high spatial resolution. Bioinforma. Oxf. Engl. *26*, 3028–3034.

Guttman, M., Garber, M., Levin, J.Z., Donaghey, J., Robinson, J., Adiconis, X., Fan, L., Koziol, M.J., Gnirke, A., Nusbaum, C., et al. (2010). Ab initio reconstruction of cell type-specific transcriptomes in mouse reveals the conserved multi-exonic structure of lincRNAs. Nat. Biotechnol. *28*, 503–510.

Hamazaki, Y., Fujita, H., Kobayashi, T., Choi, Y., Scott, H.S., Matsumoto, M., and Minato, N. (2007). Medullary thymic epithelial cells expressing Aire represent a unique lineage derived from cells expressing claudin. Nat. Immunol. *8*, 304–311.

Harada, H., Fujita, T., Miyamoto, M., Kimura, Y., Maruyama, M., Furia, A., Miyata, T., and Taniguchi, T. (1989). Structurally similar but functionally distinct factors, IRF-1 and IRF-2, bind to the same regulatory elements of IFN and IFN-inducible genes. Cell *58*, 729–739.

Harrow, J., Denoeud, F., Frankish, A., Reymond, A., Chen, C.-K., Chrast, J., Lagarde, J., Gilbert, J.G.R., Storey, R., Swarbreck, D., et al. (2006). GENCODE: Producing a reference annotation for ENCODE. Genome Biol. *7 Suppl 1*, S4.1-9.

He, F., Xiong, W., Yu, X., Espinoza-Lewis, R., Liu, C., Gu, S., Nishita, M., Suzuki, K., Yamada, G., Minami, Y., et al. (2008). Wnt5a regulates directional cell migration and cell proliferation via

Ror2-mediated noncanonical pathway in mammalian palate development. Development *135*, 3871–3879.

Heather, J.M., and Chain, B. (2016). The sequence of sequencers: The history of sequencing DNA. Genomics *107*, 1–8.

Heinonen, K.M., Ruiz Vanegas, J., Brochu, S., Shan, J., Vainio, S.J., and Perreault, C. (2011). Wnt4 regulates thymic cellularity through the expansion of thymic epithelial cells and early thymic progenitors. Blood *118*, 5163–5173.

Heintzman, N.D., Stuart, R.K., Hon, G., Fu, Y., Ching, C.W., Hawkins, R.D., Barrera, L.O., Van Calcar, S., Qu, C., Ching, K.A., et al. (2007). Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. Nat. Genet. *39*, 311–318.

Heintzman, N.D., Hon, G.C., Hawkins, R.D., Kheradpour, P., Stark, A., Harp, L.F., Ye, Z., Lee, L.K., Stuart, R.K., Ching, C.W., et al. (2009). Histone modifications at human enhancers reflect global cell-type-specific gene expression. Nature *459*, 108–112.

Heitzler, P., and Simpson, P. (1993). Altered epidermal growth factor-like sequences provide evidence for a role of Notch as a receptor in cell fate decisions. Dev. Camb. Engl. *117*, 1113–1123.

van Helden, J., André, B., and Collado-Vides, J. (1998). Extracting regulatory sites from the upstream region of yeast genes by computational analysis of oligonucleotide frequencies. J. Mol. Biol. *281*, 827–842.

van Helden, J., Rios, A.F., and Collado-Vides, J. (2000a). Discovering regulatory elements in non-coding sequences by analysis of spaced dyads. Nucleic Acids Res. *28*, 1808–1818.

van Helden, J., del Olmo, M., and Pérez-Ortín, J.E. (2000b). Statistical analysis of yeast genomic downstream sequences reveals putative polyadenylation signals. Nucleic Acids Res. *28*, 1000–1010.

Heng, T.S.P., Painter, M.W., Elpek, K., Lukacs-Kornek, V., Mauermann, N., Turley, S.J., Koller, D., Kim, F.S., Wagers, A.J., Asinovski, N., et al. (2008). The Immunological Genome Project: Networks of gene expression in immune cells. Nat. Immunol. *9*, 1091–1094.

Henikoff, S. (2008). Nucleosome destabilization in the epigenetic regulation of gene expression. Nat. Rev. Genet. *9*, 15–26.

Hetzer-Egger, C., Schorpp, M., Haas-Assenbaum, A., Balling, R., Peters, H., and Boehm, T. (2002). Thymopoiesis requires Pax9 function in thymic epithelial cells. Eur. J. Immunol. *32*, 1175–1181.

Hikosaka, Y., Nitta, T., Ohigashi, I., Yano, K., Ishimaru, N., Hayashi, Y., Matsumoto, M., Matsuo, K., Penninger, J.M., Takayanagi, H., et al. (2008). The cytokine RANKL produced by positively selected thymocytes fosters medullary thymic epithelial cells that express autoimmune regulator. Immunity *29*, 438–450.

Hock, R., Furusawa, T., Ueda, T., and Bustin, M. (2007). HMG chromosomal proteins in development and disease. Trends Cell Biol. *17*, 72–79.

Holländer, G., Gill, J., Zuklys, S., Iwanami, N., Liu, C., and Takahama, Y. (2006). Cellular and molecular events during early thymus development. Immunol. Rev. *209*, 28–46.

Honey, K., Nakagawa, T., Peters, C., and Rudensky, A. (2002). Cathepsin L regulates CD4+ T cell selection independently of its effect on invariant chain: a role in the generation of positively selecting peptide ligands. J. Exp. Med. *195*, 1349–1358.

Hozumi, K., Mailhos, C., Negishi, N., Hirano, K., Yahata, T., Ando, K., Zuklys, S., Holländer, G.A., Shima, D.T., and Habu, S. (2008). Delta-like 4 is indispensable in thymic environment specific for T cell development. J. Exp. Med. *205*, 2507–2513.

Hsieh, C.-S., deRoos, P., Honey, K., Beers, C., and Rudensky, A.Y. (2002). A role for cathepsin L and cathepsin S in peptide generation for MHC class II presentation. J. Immunol. Baltim. Md 1950 *168*, 2618–2625.

Hu, B., Lefort, K., Qiu, W., Nguyen, B.-C., Rajaram, R.D., Castillo, E., He, F., Chen, Y., Angel, P., Brisken, C., et al. (2010). Control of hair follicle cell fate by underlying mesenchyme through a CSL-Wnt5a-FoxN1 regulatory axis. Genes Dev. *24*, 1519–1532.

Hughes, J.R., Roberts, N., McGowan, S., Hay, D., Giannoulatou, E., Lynch, M., De Gobbi, M., Taylor, S., Gibbons, R., and Higgs, D.R. (2014). Analysis of hundreds of cis-regulatory landscapes at high resolution in a single, high-throughput experiment. Nat. Genet. *46*, 205–212.

Huppert, S.S., Jacobsen, T.L., and Muskavitch, M.A. (1997). Feedback regulation is central to Delta-Notch signalling required for Drosophila wing vein morphogenesis. Dev. Camb. Engl. *124*, 3283–3291.

Itoi, M., Kawamoto, H., Katsura, Y., and Amagai, T. (2001). Two distinct steps of immigration of hematopoietic progenitors into the early thymus anlage. Int. Immunol. *13*, 1203–1211.

Iwafuchi-Doi, M., Donahue, G., Kakumanu, A., Watts, J.A., Mahony, S., Pugh, B.F., Lee, D., Kaestner, K.H., and Zaret, K.S. (2016). The pioneer transcription factor FoxA maintains an accessible nucleosome configuration at enhancers for tissue-specific gene activation. Mol. Cell *62*, 79–91.

Jacobs, M.T., Frush, D.P., and Donnelly, L.F. (1999). The right place at the wrong time: Historical perspective of the relation of the thymus gland and pediatric radiology. Radiology *210*, 11–16.

Jameson, S.C., Hogquist, K.A., and Bevan, M.J. (1995). Positive selection of thymocytes. Annu. Rev. Immunol. *13*, 93–126.

Janes, S.M. (2004). Transient activation of FOXN1 in keratinocytes induces a transcriptional programme that promotes terminal differentiation: contrasting roles of FOXN1 and Akt. J. Cell Sci. *117*, 4157–4168.

Jenkinson, E.J. (1992). Studies on T cell maturation on defined thymic stromal cell populations in vitro. J. Exp. Med. *176*, 845–853.

Jenkinson, E.J., Van Ewijk, W., and Owen, J.J. (1981). Major histocompatibility complex antigen expression on the epithelium of the developing thymus in normal and nude mice. J. Exp. Med. *153*, 280–292.

Jenkinson, W.E., Jenkinson, E.J., and Anderson, G. (2003). Differential requirement for mesenchyme in the proliferation and maturation of thymic epithelial progenitors. J. Exp. Med. *198*, 325–332.

Jerome, L.A., and Papaioannou, V.E. (2001). DiGeorge syndrome phenotype in mice mutant for the T-box gene, Tbx1. Nat. Genet. *27*, 286–291.

Ji, H., Jiang, H., Ma, W., Johnson, D.S., Myers, R.M., and Wong, W.H. (2008). An integrated software system for analyzing ChIP-chip and ChIP-seq data. Nat. Biotechnol. *26*, 1293–1300.

Jiang, R., Lan, Y., Chapman, H.D., Shawber, C., Norton, C.R., Serreze, D.V., Weinmaster, G., and Gridley, T. (1998). Defects in limb, craniofacial, and thymic development in Jagged2 mutant mice. Genes Dev. *12*, 1046–1057.

Jin, X., Nowell, C.S., Ulyanchenko, S., Stenhouse, F.H., and Blackburn, C.C. (2014). Long-term persistence of functional thymic epithelial progenitor cells in vivo under conditions of low FOXN1 expression. PLoS ONE *9*, e114842.

Johnson, D.S., Mortazavi, A., Myers, R.M., and Wold, B. (2007). Genome-wide mapping of in vivo protein-DNA interactions. Science *316*, 1497–1502.

Kadonaga, J.T. (2004). Regulation of RNA polymerase II transcription by sequence-specific DNA binding factors. Cell *116*, 247–257.

Kamburov, A., Pentchev, K., Galicka, H., Wierling, C., Lehrach, H., and Herwig, R. (2011). ConsensusPathDB: Toward a more complete picture of cell biology. Nucleic Acids Res. *39*, D712-717.

Kandyba, E., Leung, Y., Chen, Y.-B., Widelitz, R., Chuong, C.-M., and Kobielak, K. (2013). Competitive balance of intrabulge BMP/Wnt signaling reveals a robust gene network ruling stem cell homeostasis and cyclic activation. Proc. Natl. Acad. Sci. *110*, 1351–1356.

Karlić, R., Chung, H.-R., Lasserre, J., Vlahovicek, K., and Vingron, M. (2010). Histone modification levels are predictive for gene expression. Proc. Natl. Acad. Sci. U. S. A. *107*, 2926–2931.

Kaur, P. (2006). Interfollicular epidermal stem cells: Identification, challenges, potential. J. Invest. Dermatol. *126*, 1450–1458.

Keeble, T.R. (2006). The Wnt receptor Ryk is required for Wnt5a-mediated axon guidance on the contralateral side of the corpus callosum. J. Neurosci. *26*, 5840–5848.

Kent, W.J., Sugnet, C.W., Furey, T.S., Roskin, K.M., Pringle, T.H., Zahler, A.M., and Haussler, D. (2002). The human genome browser at UCSC. Genome Res. *12*, 996–1006.

Kharchenko, P.V., Tolstorukov, M.Y., and Park, P.J. (2008). Design and analysis of ChIP-seq experiments for DNA-binding proteins. Nat. Biotechnol. *26*, 1351–1359.

Kim, D., Langmead, B., and Salzberg, S.L. (2015). HISAT: a fast spliced aligner with low memory requirements. Nat. Methods *12*, 357–360.

Klein, A.M., Nakagawa, T., Ichikawa, R., Yoshida, S., and Simons, B.D. (2010). Mouse germ line stem cells undergo rapid and stochastic turnover. Cell Stem Cell *7*, 214–224.

Klein, H.-U., Schäfer, M., Porse, B.T., Hasemann, M.S., Ickstadt, K., and Dugas, M. (2014). Integrative analysis of histone ChIP-seq and transcription data using Bayesian mixture models. Bioinformatics *30*, 1154–1162.

Klug, D.B., Carter, C., Crouch, E., Roop, D., Conti, C.J., and Richie, E.R. (1998). Interdependence of cortical thymic epithelial cell differentiation and T-lineage commitment. Proc. Natl. Acad. Sci. U. S. A. *95*, 11822–11827.

Klug, D.B., Crouch, E., Carter, C., Coghlan, L., Conti, C.J., and Richie, E.R. (2000). Transgenic expression of Cyclin D1 in thymic epithelial precursors promotes epithelial and T cell development. J. Immunol. *164*, 1881–1888.

Klug, D.B., Carter, C., Gimenez-Conti, I.B., and Richie, E.R. (2002). Cutting edge: thymocyte-independent and thymocyte-dependent phases of epithelial patterning in the fetal thymus. J. Immunol. Baltim. Md 1950 *169*, 2842–2845.

Koch, F., Fenouil, R., Gut, M., Cauchy, P., Albert, T.K., Zacarias-Cabeza, J., Spicuglia, S., de la Chapelle, A.L., Heidemann, M., Hintermair, C., et al. (2011). Transcription initiation platforms and GTF recruitment at tissue-specific enhancers and promoters. Nat. Struct. Mol. Biol. *18*, 956–963.

Koch, U., Fiorini, E., Benedito, R., Besseyrias, V., Schuster-Gossler, K., Pierres, M., Manley, N.R., Duarte, A., MacDonald, H.R., and Radtke, F. (2008). Delta-like 4 is the essential, nonredundant ligand for Notch1 during thymic T cell lineage commitment. J. Exp. Med. *205*, 2515–2523.

Kolesnikov, N., Hastings, E., Keays, M., Melnichuk, O., Tang, Y.A., Williams, E., Dylag, M., Kurbatova, N., Brandizi, M., Burdett, T., et al. (2015). ArrayExpress update--simplifying data submissions. Nucleic Acids Res. *43*, D1113-1116.

Kolodziejczyk, A.A., Kim, J.K., Svensson, V., Marioni, J.C., and Teichmann, S.A. (2015). The technology and biology of single-cell RNA sequencing. Mol. Cell *58*, 610–620.

Kur-Piotrowska, A., Kopcewicz, M., Kozak, L.P., Sachadyn, P., Grabowska, A., and Gawronska-Kozak, B. (2017). Neotenic phenomenon in gene expression in the skin of Foxn1- deficient (nude) mice - a projection for regenerative skin wound healing. BMC Genomics *18*, 56.

Kvell, K., Fejes, A.V., Parnell, S.M., and Pongracz, J.E. (2014). Active Wnt/beta-catenin signaling is required for embryonic thymic epithelial development and functionality ex vivo. Immunobiology *219*, 644–652.

Lachner, M., O'Carroll, D., Rea, S., Mechtler, K., and Jenuwein, T. (2001). Methylation of histone H3 lysine 9 creates a binding site for HP1 proteins. Nature *410*, 116–120.

Laclef, C., Souil, E., Demignon, J., and Maire, P. (2003). Thymus, kidney and craniofacial abnormalities in Six 1 deficient mice. Mech. Dev. *120*, 669–679.

Lai, E.C. (2004). Notch signaling: control of cell communication and cell fate. Development *131*, 965–973.

Lander, E.S., Linton, L.M., Birren, B., Nusbaum, C., Zody, M.C., Baldwin, J., Devon, K., Dewar, K., Doyle, M., FitzHugh, W., et al. (2001). Initial sequencing and analysis of the human genome. Nature *409*, 860–921.

Landt, S.G., Marinov, G.K., Kundaje, A., Kheradpour, P., Pauli, F., Batzoglou, S., Bernstein, B.E., Bickel, P., Brown, J.B., Cayting, P., et al. (2012). ChIP-seq guidelines and practices of the ENCODE and modENCODE consortia. Genome Res. *22*, 1813–1831.

Langmead, B., and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. Nat. Methods *9*, 357–359.

Law, C.W., Chen, Y., Shi, W., and Smyth, G.K. (2014). voom: Precision weights unlock linear model analysis tools for RNA-seq read counts. Genome Biol. *15*, R29.

Law, C.W., Alhamdoosh, M., Su, S., Smyth, G.K., and Ritchie, M.E. (2016). RNA-seq analysis is easy as 1-2-3 with limma, Glimma and edgeR. F1000Research *5*, 1408.

Le Douarin, N.M., and Jotereau, F.V. (1975). Tracing of cells of the avian thymus through embryonic life in interspecific chimeras. J. Exp. Med. *142*, 17–40.

Leask, A., Byrne, C., and Fuchs, E. (1991). Transcription factor AP2 and its role in epidermal-specific gene expression. Proc. Natl. Acad. Sci. U. S. A. *88*, 7948–7952.

Lee, C.S., Buttitta, L., and Fan, C.-M. (2001). Evidence that the WNT-inducible growth arrest-specific gene 1 encodes an antagonist of sonic hedgehog signaling in the somite. Proc. Natl. Acad. Sci. *98*, 11347–11352.

Lee, D., Prowse, D.M., and Brissette, J.L. (1999). Association between mouse nude gene expression and the initiation of epithelial terminal differentiation. Dev. Biol. *208*, 362–374.

Lee, D.K., Hakim, F.T., and Gress, R.E. (2010). The thymus and the immune system: Layered levels of control. J. Thorac. Oncol. *5*, S273–S276.

Lee, Y., Kim, M., Han, J., Yeom, K.-H., Lee, S., Baek, S.H., and Kim, V.N. (2004). MicroRNA genes are transcribed by RNA polymerase II. EMBO J. *23*, 4051–4060.

Leek, J., Johnson, W., Parker, H., Fertig, E., Jaffe, A., and Storey, J. (2016). sva: Surrogate Variable Analysis.

LeibundGut-Landmann, S., Waldburger, J.-M., Krawczyk, M., Otten, L.A., Suter, T., Fontana, A., Acha-Orbea, H., and Reith, W. (2004). Mini-review: Specificity and expression of CIITA, the master regulator of MHC class II genes. Eur. J. Immunol. *34*, 1513–1525.

Leushacke, M., Ng, A., Galle, J., Loeffler, M., and Barker, N. (2013). Lgr5+ gastric stem cells divide symmetrically to effect epithelial homeostasis in the pylorus. Cell Rep. *5*, 349–356.

Li, B., and Dewey, C.N. (2011). RSEM: Accurate transcript quantification from RNA-Seq data with or without a reference genome. BMC Bioinformatics *12*, 323.

Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinforma. Oxf. Engl. *25*, 1754–1760.

Li, H., Ruan, J., and Durbin, R. (2008a). Mapping short DNA sequencing reads and calling variants using mapping quality scores. Genome Res. *18*, 1851–1858.

Li, Q., Brown, J., Huang, H., and Bickel, P. (2011a). Measuring reproducibility of high-throughput experiments. Ann Appl Stat *5*, 1752–1779.

Li, R., Li, Y., Kristiansen, K., and Wang, J. (2008b). SOAP: short oligonucleotide alignment program. Bioinforma. Oxf. Engl. *24*, 713–714.

Li, Z., Schug, J., Tuteja, G., White, P., and Kaestner, K.H. (2011b). The nucleosome map of the mammalian liver. Nat. Struct. Mol. Biol. *18*, 742–746.

Liao, Y., Smyth, G.K., and Shi, W. (2014). featureCounts: An efficient general purpose program for assigning sequence reads to genomic features. Bioinformatics *30*, 923–930.

Lindsay, E.A., Vitelli, F., Su, H., Morishima, M., Huynh, T., Pramparo, T., Jurecic, V., Ogunrinu, G., Sutherland, H.F., Scambler, P.J., et al. (2001). Tbx1 haploinsufficiency in the DiGeorge syndrome region causes aortic arch defects in mice. Nature *410*, 97–101.

Linton, P.J., and Dorshkind, K. (2004). Age-related changes in lymphocyte development and function. Nat. Immunol. *5*, 133–139.

Lipshutz, R.J., Fodor, S.P., Gingeras, T.R., and Lockhart, D.J. (1999). High density synthetic oligonucleotide arrays. Nat. Genet. *21*, 20–24.

Liu, C. (2005). The role of CCL21 in recruitment of T-precursor cells to fetal thymi. Blood *105*, 31–39.

Liu, S., and Trapnell, C. (2016). Single-cell transcriptome sequencing: recent advances and remaining challenges. F1000Research.

Liu, C., Saito, F., Liu, Z., Lei, Y., Uehara, S., Love, P., Lipp, M., Kondo, S., Manley, N., and Takahama, Y. (2006). Coordination between CCR7- and CCR9-mediated chemokine signals in prevascular fetal thymus colonization. Blood *108*, 2531–2539.

Liu, D., Kousa, A.I., O'Neill, K.E., Guillemot, F., Popis, M., Farley, A.M., Tomlinson, S.R., Koch, U., Radtke, F., and Blackburn, C.C. (2017). A Notch gate controls differentiation of thymic epithelial progenitors and medullary thymic epithelial lineage specification. (submitted)

Lockhart, D.J., Dong, H., Byrne, M.C., Follettie, M.T., Gallo, M.V., Chee, M.S., Mittmann, M., Wang, C., Kobayashi, M., Norton, H., et al. (1996). Expression monitoring by hybridization to high-density oligonucleotide arrays. Nat. Biotechnol. *14*, 1675–1680.

Lopez-Garcia, C., Klein, A.M., Simons, B.D., and Winton, D.J. (2010). Intestinal stem cell replacement follows a pattern of neutral drift. Science *330*, 822–825.

Lucas, B., McCarthy, N.I., Baik, S., Cosway, E., James, K.D., Parnell, S.M., White, A.J., Jenkinson, W.E., and Anderson, G. (2016). Control of the thymic medulla and its influence on αβT-cell development. Immunol. Rev. *271*, 23–37.

Luger, K., Mäder, A.W., Richmond, R.K., Sargent, D.F., and Richmond, T.J. (1997). Crystal structure of the nucleosome core particle at 2.8 Å resolution. Nature *389*, 251–260.

Luis, T.C., Luc, S., Mizukami, T., Boukarabila, H., Thongjuea, S., Woll, P.S., Azzoni, E., Giustacchini, A., Lutteropp, M., Bouriez-Jones, T., et al. (2016). Initial seeding of the embryonic thymus by immune-restricted lympho-myeloid progenitors. Nat. Immunol. *17*, 1424–1435.

Lunter, G., and Goodson, M. (2011). Stampy: A statistical algorithm for sensitive and fast mapping of Illumina sequence reads. Genome Res. *21*, 936–939.

Ma, J. (2011). Transcriptional activators and activation mechanisms. Protein Cell *2*, 879–888.

Macaulay, I.C., and Voet, T. (2014). Single cell genomics: advances and future perspectives. PLoS Genet. *10*, e1004126.

Machanick, P., and Bailey, T.L. (2011). MEME-ChIP: motif analysis of large DNA datasets. Bioinformatics *27*, 1696–1697.

Mackall, C.L., Fry, T.J., Bare, C., Morgan, P., Galbraith, A., and Gress, R.E. (2001). IL-7 increases both thymic-dependent and thymic-independent T-cell regeneration after bone marrow transplantation. Blood *97*, 1491–1497.

Malik, S., and Roeder, R.G. (2005). Dynamic regulation of pol II transcription by the mammalian Mediator complex. Trends Biochem. Sci. *30*, 256–263.

Manley, N.R., and Capecchi, M.R. (1995). The role of Hoxa-3 in mouse thymus and thyroid development. Dev. Camb. Engl. *121*, 1989–2003.

Marcelle, C., Stark, M.R., and Bronner-Fraser, M. (1997). Coordinate actions of BMPs, Wnts, Shh and noggin mediate patterning of the dorsal somite. Dev. Camb. Engl. *124*, 3955–3963.

Margulies, M., Egholm, M., Altman, W.E., Attiya, S., Bader, J.S., Bemben, L.A., Berka, J., Braverman, M.S., Chen, Y.-J., Chen, Z., et al. (2005). Genome sequencing in microfabricated high-density picolitre reactors. Nature *437*, 376–380.

Marioni, J.C., Mason, C.E., Mane, S.M., Stephens, M., and Gilad, Y. (2008). RNA-seq: an assessment of technical reproducibility and comparison with gene expression arrays. Genome Res. *18*, 1509–1517.

Martin, C., and Zhang, Y. (2005). The diverse functions of histone lysine methylation. Nat. Rev. Mol. Cell Biol. *6*, 838–849.

Martinez, S., Scerbo, P., Giordano, M., Daulat, A.M., Lhoumeau, A.-C., Thomé, V., Kodjabachian, L., and Borg, J.-P. (2015). The PTK7 and ROR2 protein receptors interact in the vertebrate WNT/Planar cell polarity (PCP) pathway. J. Biol. Chem. *290*, 30562–30572.

Maston, G.A., Evans, S.K., and Green, M.R. (2006). Transcriptional regulatory elements in the human genome. Annu. Rev. Genomics Hum. Genet. *7*, 29–59.

Masuda, K., Germeraad, W.T.V., Satoh, R., Itoi, M., Ikawa, T., Minato, N., Katsura, Y., van Ewijk, W., and Kawamoto, H. (2009). Notch activation in thymic epithelial cells induces development of thymic microenvironments. Mol. Immunol. *46*, 1756–1767.

Matharu, N., and Ahituv, N. (2015). Minor loops in major folds: Enhancer–promoter looping, chromatin restructuring, and their association with transcriptional regulation and disease. PLOS Genet. *11*, e1005640.

Matsui, T., Segall, J., Weil, P.A., and Roeder, R.G. (1980). Multiple factors required for accurate initiation of transcription by purified RNA polymerase II. J. Biol. Chem. *255*, 11992–11996.

Matsuyama, M., Aizawa, S., and Shimono, A. (2009). Sfrp controls apicobasal polarity and oriented cell division in developing gut epithelium. PLoS Genet. *5*, e1000427.

Maxam, A.M., and Gilbert, W. (1977). A new method for sequencing DNA. Proc. Natl. Acad. Sci. U. S. A. *74*, 560–564.

Mayran, A., and Drouin, J. (2018). Pioneer transcription factors shape the epigenetic landscape. J. Biol. Chem. jbc.R117.001232.

McCarthy, D.J., Chen, Y., and Smyth, G.K. (2012). Differential expression analysis of multifactor RNA-Seq experiments with respect to biological variation. Nucleic Acids Res. *40*, 4288–4297.

McLeay, R.C., Lesluyes, T., Cuellar Partida, G., and Bailey, T.L. (2012). Genome-wide in silico prediction of gene expression. Bioinforma. Oxf. Engl. *28*, 2789–2796.

Meier, N., Dear, T.N., and Boehm, T. (1999). Whn and mHa3 are components of the genetic hierarchy controlling hair follicle differentiation. Mech. Dev. *89*, 215–221.

Meraz, M.A., White, J.M., Sheehan, K.C., Bach, E.A., Rodig, S.J., Dighe, A.S., Kaplan, D.H., Riley, J.K., Greenlund, A.C., Campbell, D., et al. (1996). Targeted disruption of the Stat1 gene in mice reveals unexpected physiologic specificity in the JAK-STAT signaling pathway. Cell *84*, 431–442.

Mikheyev, A.S., and Tin, M.M.Y. (2014). A first look at the Oxford Nanopore MinION sequencer. Mol. Ecol. Resour. *14*, 1097–1102.

Mikkelsen, T.S., Ku, M., Jaffe, D.B., Issac, B., Lieberman, E., Giannoukos, G., Alvarez, P., Brockman, W., Kim, T.-K., Koche, R.P., et al. (2007). Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. Nature *448*, 553–560.

Milićević, N.M., Milićević, Z., Colic, M., and Mujović, S. (1987). Ultrastructural study of macrophages in the rat thymus, with special reference to the cortico-medullary zone. J. Anat. *150*, 89–98.

Miller, J.F. a. P. (2002). The discovery of thymus function and of thymus-derived lymphocytes. Immunol. Rev. *185*, 7–14.

Moll, R., Franke, W.W., Schiller, D.L., Geiger, B., and Krepler, R. (1982). The catalog of human cytokeratins: patterns of expression in normal epithelia, tumors and cultured cells. Cell *31*, 11–24.

Moore-Scott, B.A., and Manley, N.R. (2005). Differential expression of Sonic hedgehog along the anterior-posterior axis regulates patterning of pharyngeal pouch endoderm and pharyngeal endoderm-derived organs. Dev. Biol. *278*, 323–335.

Morasso, M.I., and Tomic-Canic, M. (2005). Epidermal stem cells: The cradle of epidermal determination, differentiation and wound healing. Biol. Cell *97*, 173–183.

Morozova, O., Hirst, M., and Marra, M.A. (2009). Applications of new sequencing technologies for transcriptome analysis. Annu. Rev. Genomics Hum. Genet. *10*, 135–151.

Morrison, S.J., Uchida, N., and Weissman, I.L. (1995). The biology of hematopoietic stem cells. Annu. Rev. Cell Dev. Biol. *11*, 35–71.

Mortazavi, A., Williams, B.A., McCue, K., Schaeffer, L., and Wold, B. (2008). Mapping and quantifying mammalian transcriptomes by RNA-Seq. Nat. Methods *5*, 621–628.

Mouse Genome Sequencing Consortium, Waterston, R.H., Lindblad-Toh, K., Birney, E., Rogers, J., Abril, J.F., Agarwal, P., Agarwala, R., Ainscough, R., Alexandersson, M., et al. (2002). Initial sequencing and comparative analysis of the mouse genome. Nature *420*, 520–562.

Muhlethaler-Mottet, A. (1997). Expression of MHC class II molecules in different cellular and functional compartments is controlled by differential usage of multiple promoters of the transactivator CIITA. EMBO J. *16*, 2851–2860.

Muhlethaler-Mottet, A., Di Berardino, W., Otten, L.A., and Mach, B. (1998). Activation of the MHC class II transactivator CIITA by interferon-gamma requires cooperative interaction between Stat1 and USF-1. Immunity *8*, 157–166.

Muiño, J.M., Kaufmann, K., van Ham, R.C., Angenent, G.C., and Krajewski, P. (2011). ChIP-seq Analysis in R (CSAR): An R package for the statistical detection of protein-bound genomic regions. Plant Methods *7*, 11.

Müllner, D. (2013). fastcluster: Fast hierarchical, agglomerative clustering routines for R and Python. J. Stat. Softw. *53*.

Munroe, D.J., and Harris, T.J.R. (2010). Third-generation sequencing fireworks at Marco Island. Nat. Biotechnol. *28*, 426–428.

Murata, S., Sasaki, K., Kishimoto, T., Niwa, S. -i., Hayashi, H., Takahama, Y., and Tanaka, K. (2007). Regulation of CD8+ T cell development by thymus-specific proteasomes. Science *316*, 1349–1353.

Nagalakshmi, U., Wang, Z., Waern, K., Shou, C., Raha, D., Gerstein, M., and Snyder, M. (2008). The transcriptional landscape of the yeast genome defined by RNA sequencing. Science *320*, 1344–1349.

Nakagawa, S., Gisselbrecht, S.S., Rogers, J.M., Hartl, D.L., and Bulyk, M.L. (2013). DNA-binding specificity changes in the evolution of forkhead transcription factors. Proc. Natl. Acad. Sci. U. S. A. *110*, 12349–12354.

Nakagawa, T., Roth, W., Wong, P., Nelson, A., Farr, A., Deussing, J., Villadangos, J.A., Ploegh, H., Peters, C., and Rudensky, A.Y. (1998). Cathepsin L: critical role in Ii degradation and CD4 T cell selection in the thymus. Science *280*, 450–453.

Nehls, M., Pfeifer, D., Schorpp, M., Hedrich, H., and Boehm, T. (1994). New member of the winged-helix protein family disrupted in mouse and rat nude mutations. Nature *372*, 103–107.

Nehls, M., Kyewski, B., Messerle, M., Waldschütz, R., Schüddekopf, K., Smith, A.J., and Boehm, T. (1996). Two genetically separable steps in the differentiation of thymic epithelium. Science *272*, 886–889.

Nelson, W.G., and Sun, T.T. (1983). The 50- and 58-kdalton keratin classes as molecular markers for stratified squamous epithelia: cell culture studies. J. Cell Biol. *97*, 244–251.

Nelson, N., Marks, M.S., Driggers, P.H., and Ozato, K. (1993). Interferon consensus sequence-binding protein, a member of the interferon regulatory factor family, suppresses interferon-induced gene transcription. Mol. Cell. Biol. *13*, 588–599.

Nemeth, M.J., Topol, L., Anderson, S.M., Yang, Y., and Bodine, D.M. (2007). Wnt5a inhibits canonical Wnt signaling in hematopoietic stem cells and enhances repopulation. Proc. Natl. Acad. Sci. *104*, 15436–15441.

Nowell, C.S., Bredenkamp, N., Tetélin, S., Jin, X., Tischner, C., Vaidya, H., Sheridan, J.M., Stenhouse, F.H., Heussen, R., Smith, A.J.H., et al. (2011). Foxn1 regulates lineage progression in cortical and medullary thymic epithelial cells but is dispensable for medullary sublineage divergence. PLoS Genet. *7*, e1002348.

Ohigashi, I., Zuklys, S., Sakata, M., Mayer, C.E., Zhanybekova, S., Murata, S., Tanaka, K., Holländer, G.A., and Takahama, Y. (2013). Aire-expressing thymic medullary epithelial cells originate from β5t-expressing progenitor cells. Proc. Natl. Acad. Sci. U. S. A. *110*, 9885–9890.

Ohnemus, S., Kanzler, B., Jerome-Majewska, L.A., Papaioannou, V.E., Boehm, T., and Mallo, M. (2002). Aortic arch and pharyngeal phenotype in the absence of BMP-dependent neural crest in the mouse. Mech. Dev. *119*, 127–135.

Ohuchi, H., Hori, Y., Yamasaki, M., Harada, H., Sekine, K., Kato, S., and Itoh, N. (2000). FGF10 acts as a major ligand for FGF receptor 2 IIIb in mouse multi-organ development. Biochem. Biophys. Res. Commun. *277*, 643–649.

Okubo, T., Kawamura, A., Takahashi, J., Yagi, H., Morishima, M., Matsuoka, R., and Takada, S. (2011). Ripply3, a Tbx1 repressor, is required for development of the pharyngeal apparatus and its derivatives in mice. Development *138*, 339–348.

O'Neill, K.E., Bredenkamp, N., Tischner, C., Vaidya, H.J., Stenhouse, F.H., Peddie, C.D., Nowell, C.S., Gaskell, T., and Blackburn, C.C. (2016). Foxn1 is dynamically regulated in thymic epithelial cells during embryogenesis and at the onset of thymic involution. PLOS ONE *11*, e0151666.

Orkin, S.H., and Zon, L.I. (2008). Hematopoiesis: An evolving paradigm for stem cell biology. Cell *132*, 631–644.

Orphanides, G., Lagrange, T., and Reinberg, D. (1996). The general transcription factors of RNA polymerase II. Genes Dev. *10*, 2657–2683.

Osada, M., Ito, E., Fermin, H.A., Vazquez-Cintron, E., Venkatesh, T., Friedel, R.H., and Pezzano, M. (2006). The Wnt signaling antagonist Kremen1 is required for development of thymic architecture. Clin. Dev. Immunol. *13*, 299–319.

Ouyang, Z., Zhou, Q., and Wong, W.H. (2009). ChIP-Seq of transcription factors predicts absolute and differential gene expression in embryonic stem cells. Proc. Natl. Acad. Sci. U. S. A. *106*, 21521–21526.

Palmer, E. (2003). Cell death and immunity: Negative selection — clearing out the bad apples from the T-cell repertoire. Nat. Rev. Immunol. *3*, 383–391.

Pantelouris, E.M. (1968). Absence of thymus in a mouse mutant. Nature *217*, 370–371.

Pantelouris, E.M., and Hair, J. (1970). Thymus dysgenesis in nude (nu nu) mice. J. Embryol. Exp. Morphol. *24*, 615–623.

Parent, A.V., Russ, H.A., Khan, I.S., LaFlam, T.N., Metzger, T.C., Anderson, M.S., and Hebrok, M. (2013). Generation of functional thymic epithelium from human embryonic stem cells that supports host T cell development. Cell Stem Cell *13*, 219–229.

Patel, S.R., Gordon, J., Mahbub, F., Blackburn, C.C., and Manley, N.R. (2006). Bmp4 and Noggin expression during early thymus and parathyroid organogenesis. Gene Expr. Patterns *6*, 794–799.

Patro, R., Duggal, G., Love, M.I., Irizarry, R.A., and Kingsford, C. (2016). Salmon provides accurate, fast, and bias-aware transcript expression estimates using dual-phase inference.

Pearson, K. (1895). Notes on regression and inheritance in the case of two parents. Proc R Soc Lond *58*, 240–242.

Pekowska, A., Benoukraf, T., Zacarias-Cabeza, J., Belhocine, M., Koch, F., Holota, H., Imbert, J., Andrau, J.-C., Ferrier, P., and Spicuglia, S. (2011). H3K4 tri-methylation provides an epigenetic signature of active enhancers: Epigenetic signature of active enhancers. EMBO J. *30*, 4198–4210.

Peters, H., Neubüser, A., Kratochwil, K., and Balling, R. (1998). Pax9-deficient mice lack pharyngeal pouch derivatives and teeth and exhibit craniofacial and limb abnormalities. Genes Dev. *12*, 2735–2747.

Picard tools (2016). Picard tools.

Pine, R., Canova, A., and Schindler, C. (1994). Tyrosine phosphorylated p91 binds to a single element in the ISGF2/IRF-1 promoter to mediate induction by IFN alpha and IFN gamma, and is likely to autoregulate the p91 gene. EMBO J. *13*, 158–167.

Piskurich, J.F., Wang, Y., Linhoff, M.W., White, L.C., and Ting, J.P. (1998). Identification of distinct regions of 5’ flanking DNA that mediate constitutive, IFN-gamma, STAT1, and TGFbeta-regulated expression of the class II transactivator gene. J. Immunol. Baltim. Md 1950 *160*, 233–240.

Plotkin, J., Prockop, S.E., Lepique, A., and Petrie, H.T. (2003). Critical role for CXCR4 signaling in progenitor localization and T cell differentiation in the postnatal thymus. J. Immunol. Baltim. Md 1950 *171*, 4521–4527.

Prowse, D.M., Lee, D., Weiner, L., Jiang, N., Magro, C.M., Baden, H.P., and Brissette, J.L. (1999). Ectopic expression of the nude gene induces hyperproliferation and defects in differentiation: Implications for the self-renewal of cutaneous epithelia. Dev. Biol. *212*, 54–67.

Ptashne, M. (1988). How eukaryotic transcriptional activators work. Nature *335*, 683–689.

Qin, Z.S., Yu, J., Shen, J., Maher, C.A., Hu, M., Kalyana-Sundaram, S., Yu, J., and Chinnaiyan, A.M. (2010). HPeak: an HMM-based algorithm for defining read-enriched regions in ChIP-Seq data. BMC Bioinformatics *11*, 369.

Rada-Iglesias, A., Bajpai, R., Swigut, T., Brugmann, S.A., Flynn, R.A., and Wysocka, J. (2011). A unique chromatin signature uncovers early developmental enhancers in humans. Nature *470*, 279–283.

Rajagopal, N., Ernst, J., Ray, P., Wu, J., Zhang, M., Kellis, M., and Ren, B. (2014). Distinct and predictive histone lysine acetylation patterns at promoters, enhancers, and gene bodies. G3amp58 GenesGenomesGenetics *4*, 2051–2063.

Ramírez, F., Ryan, D.P., Grüning, B., Bhardwaj, V., Kilpert, F., Richter, A.S., Heyne, S., Dündar, F., and Manke, T. (2016). deepTools2: a next generation web server for deep-sequencing data analysis. Nucleic Acids Res. *44*, W160–W165.

Rapaport, F., Khanin, R., Liang, Y., Pirun, M., Krek, A., Zumbo, P., Mason, C.E., Socci, N.D., and Betel, D. (2013). Comprehensive evaluation of differential gene expression analysis methods for RNA-seq data. Genome Biol. *14*, R95.

Rashid, N.U., Giresi, P.G., Ibrahim, J.G., Sun, W., and Lieb, J.D. (2011). ZINBA integrates local covariates with DNA-seq data to identify broad and narrow regions of enrichment, even within amplified genomic regions. Genome Biol. *12*, R67.

Rebay, I., Silver, S., and Tootle, T. (2005). New vision from Eyes absent: Transcription factors as enzymes. Trends Genet. *21*, 163–171.

Reeh, K.A.G., Cardenas, K.T., Bain, V.E., Liu, Z., Laurent, M., Manley, N.R., and Richie, E.R. (2014). Ectopic TBX1 suppresses thymic epithelial cell differentiation and proliferation during thymus organogenesis. Development *141*, 2950–2958.

Reith, W., LeibundGut-Landmann, S., and Waldburger, J.-M. (2005). Regulation of MHC class II gene expression by the class II transactivator. Nat. Rev. Immunol. *5*, 793–806.

Revest, J.M., Suniara, R.K., Kerr, K., Owen, J.J., and Dickson, C. (2001b). Development of the thymus requires signaling through the fibroblast growth factor receptor R2-IIIb. J. Immunol. Baltim. Md 1950 *167*, 1954–1961.

Ritchie, M.E., Phipson, B., Wu, D., Hu, Y., Law, C.W., Shi, W., and Smyth, G.K. (2015). limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res. *43*, e47.

Rizzo, J.M., and Buck, M.J. (2012). Key principles and clinical applications of "next-generation" DNA sequencing. Cancer Prev. Res. (Phila. Pa.) *5*, 887–900.

Robertson, G., Hirst, M., Bainbridge, M., Bilenky, M., Zhao, Y., Zeng, T., Euskirchen, G., Bernier, B., Varhol, R., Delaney, A., et al. (2007). Genome-wide profiles of STAT1 DNA association using chromatin immunoprecipitation and massively parallel sequencing. Nat. Methods *4*, 651–657.

Robinson, M.D., and Smyth, G.K. (2007). Moderated statistical tests for assessing differences in tag abundance. Bioinforma. Oxf. Engl. *23*, 2881–2887.

Robinson, J.T., Thorvaldsdóttir, H., Winckler, W., Guttman, M., Lander, E.S., Getz, G., and Mesirov, J.P. (2011). Integrative genomics viewer. Nat. Biotechnol. *29*, 24–26.

Robinson, M.D., McCarthy, D.J., and Smyth, G.K. (2010). edgeR: A Bioconductor package for differential expression analysis of digital gene expression data. Bioinforma. Oxf. Engl. *26*, 139–140.

Rode, I., Martins, V.C., Küblbeck, G., Maltry, N., Tessmer, C., and Rodewald, H.-R. (2015). Foxn1 protein expression in the developing, aging, and regenerating thymus. J. Immunol. *195*, 5678–5687.

Rodewald, H.-R. (2008). Thymus organogenesis. Annu. Rev. Immunol. *26*, 355–388.

Rodewald, H.R., Paul, S., Haller, C., Bluethmann, H., and Blum, C. (2001). Thymus medulla consisting of epithelial islets each derived from a single progenitor. Nature *414*, 763–768.

Rohn, W., Tang, L.P., Dong, Y., and Benveniste, E.N. (1999). IL-1 beta inhibits IFN-gamma-induced class II MHC expression by suppressing transcription of the class II transactivator gene. J. Immunol. Baltim. Md 1950 *162*, 886–896.

Roignant, J.-Y., and Treisman, J.E. (2009). Pattern formation in the Drosophila eye disc. Int. J. Dev. Biol. *53*, 795–804.

Rossi, S.W., Jenkinson, W.E., Anderson, G., and Jenkinson, E.J. (2006). Clonal analysis reveals a common progenitor for thymic cortical and medullary epithelium. Nature *441*, 988–991.

Rossi, S.W., Kim, M.-Y., Leibbrandt, A., Parnell, S.M., Jenkinson, W.E., Glanville, S.H., McConnell, F.M., Scott, H.S., Penninger, J.M., Jenkinson, E.J., et al. (2007a). RANK signals from CD4(+)3(-) inducer cells regulate development of Aire-expressing epithelial cells in the thymic medulla. J. Exp. Med. *204*, 1267–1272.

Rossi, S.W., Chidgey, A.P., Parnell, S.M., Jenkinson, W.E., Scott, H.S., Boyd, R.L., Jenkinson, E.J., and Anderson, G. (2007b). Redefining epithelial progenitor potential in the developing thymus. Eur. J. Immunol. *37*, 2411–2418.

Rushton, E., Drysdale, R., Abmayr, S.M., Michelson, A.M., and Bate, M. (1995). Mutations in a novel gene, myoblast city, provide evidence in support of the founder cell hypothesis for Drosophila muscle development. Dev. Camb. Engl. *121*, 1979–1988.

Saha, A., Wittmeyer, J., and Cairns, B.R. (2006). Chromatin remodelling: The industrial revolution of DNA around histones. Nat. Rev. Mol. Cell Biol. *7*, 437–447.

Saldaña, J.I., Solanki, A., Lau, C.-I., Sahni, H., Ross, S., Furmanski, A.L., Ono, M., Holländer, G., and Crompton, T. (2016). Sonic Hedgehog regulates thymic epithelial cell differentiation. J. Autoimmun. *68*, 86–97.

Saliba, A.-E., Westermann, A.J., Gorski, S.A., and Vogel, J. (2014). Single-cell RNA-seq: advances and future challenges. Nucleic Acids Res. *42*, 8845–8860.

Salmon-Divon, M., Dvinge, H., Tammoja, K., and Bertone, P. (2010). PeakAnalyzer: Genome-wide annotation of chromatin binding and modification loci. BMC Bioinformatics *11*, 415.

Sanger, F., Nicklen, S., and Coulson, A.R. (1977). DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. U. S. A. *74*, 5463–5467.

Sansom, S.N., Shikama-Dorn, N., Zhanybekova, S., Nusspaumer, G., Macaulay, I.C., Deadman, M.E., Heger, A., Ponting, C.P., and Holländer, G.A. (2014). Population and single-cell genomics reveal the Aire dependency, relief from Polycomb silencing, and distribution of self-antigen expression in thymic epithelia. Genome Res. *24*, 1918–1931.

Sasaki, K., Takada, K., Ohte, Y., Kondo, H., Sorimachi, H., Tanaka, K., Takahama, Y., and Murata, S. (2015). Thymoproteasomes produce unique peptide motifs for positive selection of CD8+ T cells. Nat. Commun. *6*, 7484.

Schadt, E.E., Turner, S., and Kasarskis, A. (2010). A window into third-generation sequencing. Hum. Mol. Genet. *19*, R227-240.

Schalch, T., Duda, S., Sargent, D.F., and Richmond, T.J. (2005). X-ray structure of a tetranucleosome and its implications for the chromatin fibre. Nature *436*, 138–141.

Schena, M., Shalon, D., Heller, R., Chai, A., Brown, P.O., and Davis, R.W. (1996). Parallel human genome analysis: microarray-based expression monitoring of 1000 genes. Proc. Natl. Acad. Sci. U. S. A. *93*, 10614–10619.

Schlake, T., Schorpp, M., Nehls, M., and Boehm, T. (1997). The nude gene encodes a sequence-specific DNA binding protein with homologs in organisms that lack an anticipatory immune system. Proc. Natl. Acad. Sci. U. S. A. *94*, 3842–3847.

Schluep, M., Willcox, N., Ritter, M.A., Newsom-Davis, J., Larché, M., and Brown, A.N. (1988). Myasthenia gravis thymus: clinical, histological and culture correlations. J. Autoimmun. *1*, 445–467.

Schreiber, L., Eshel, I., Meilin, A., Sharabi, Y., and Shoham, J. (1991). Analysis of thymic stromal cell subpopulations grown in vitro on extracellular matrix in defined medium. III. Growth conditions of human thymic epithelial cells and immunomodulatory activities in their culture supernatant. Immunology *74(4)*, 621–629.

Schroder, K. (2003). Interferon-g : An overview of signals, mechanisms and functions. J. Leukoc. Biol. *75*, 163–189.

Seita, J., Sahoo, D., Rossi, D.J., Bhattacharya, D., Serwold, T., Inlay, M.A., Ehrlich, L.I.R., Fathman, J.W., Dill, D.L., and Weissman, I.L. (2012). Gene expression commons: An open platform for absolute gene expression profiling. PLoS ONE *7*, e40321.

Sekai, M., Hamazaki, Y., and Minato, N. (2014). Medullary thymic epithelial stem cells maintain a functional thymus to ensure lifelong central T cell tolerance. Immunity *41*, 753–761.

Sekiya, T., and Zaret, K.S. (2007). Repression by Groucho/TLE/Grg proteins: Genomic site recruitment generates compacted chromatin in vitro and impairs activator binding in vivo. Mol. Cell *28*, 291–303.

Serandour, A.A., Avner, S., Percevault, F., Demay, F., Bizot, M., Lucchetti-Miganeh, C., Barloy-Hubler, F., Brown, M., Lupien, M., Metivier, R., et al. (2011). Epigenetic switch involved in activation of pioneer factor FOXA1-dependent enhancers. Genome Res. *21*, 555–565.

Sergushichev, A. (2016). An algorithm for fast preranked gene set enrichment analysis using cumulative statistic calculation.

Seyednasrollah, F., Laiho, A., and Elo, L.L. (2015). Comparison of software packages for detecting differential expression in RNA-seq studies. Brief. Bioinform. *16*, 59–70.

Shah, D.K., and Zuniga-Pflucker, J.C. (2014). An overview of the intrathymic intricacies of T cell development. J. Immunol. *192*, 4017–4023.

Shakib, S., Desanti, G.E., Jenkinson, W.E., Parnell, S.M., Jenkinson, E.J., and Anderson, G. (2009). Checkpoints in the development of thymic cortical epithelial cells. J. Immunol. Baltim. Md 1950 *182*, 130–137.

Shao, J.-S., Aly, Z.A., Lai, C.-F., Cheng, S.-L., Cai, J., Huang, E., Behrmann, A., and Towler, D.A. (2007). Vascular Bmp Msx2 Wnt signaling and oxidative stress in arterial calcification. Ann. N. Y. Acad. Sci. *1117*, 40–50.

Shendure, J., Porreca, G.J., Reppas, N.B., Lin, X., McCutcheon, J.P., Rosenbaum, A.M., Wang, M.D., Zhang, K., Mitra, R.D., and Church, G.M. (2005). Accurate multiplex polony sequencing of an evolved bacterial genome. Science *309*, 1728–1732.

Sheridan, J.M., Taoudi, S., Medvinsky, A., and Blackburn, C.C. (2009). A novel method for the generation of reaggregated organotypic cultures that permits juxtaposition of defined cell populations. Genesis *47*, 346–351.

Shimizu, T., Tanaka, T., Iso, T., Doi, H., Sato, H., Kawai-Kowase, K., Arai, M., and Kurabayashi, M. (2009). Notch signaling induces osteogenic differentiation and mineralization of vascular smooth muscle cells: Role of Msx2 gene induction via Notch-RBP-Jk signaling. Arterioscler. Thromb. Vasc. Biol. *29*, 1104–1111.

Simoes-Costa, M., and Bronner, M.E. (2015). Establishing neural crest identity: a gene regulatory recipe. Development *142*, 242–257.

Smith, T.J. (2010). Insulin-like growth factor-I regulation of immune function: A potential therapeutic target in autoimmune diseases? Pharmacol. Rev. *62*, 199–236.

Snippert, H.J., van der Flier, L.G., Sato, T., van Es, J.H., van den Born, M., Kroon-Veenboer, C., Barker, N., Klein, A.M., van Rheenen, J., Simons, B.D., et al. (2010). Intestinal crypt homeostasis results from neutral competition between symmetrically dividing Lgr5 stem cells. Cell *143*, 134–144.

Song, Q., and Smith, A.D. (2011). Identifying dispersed epigenomic domains from ChIP-Seq data. Bioinforma. Oxf. Engl. *27*, 870–871.

Soriano, P. (1997). The PDGF alpha receptor is required for neural crest cell development and for normal patterning of the somites. Dev. Camb. Engl. *124*, 2691–2700.

Soufi, A., Donahue, G., and Zaret, K.S. (2012). Facilitators and impediments of the pluripotency reprogramming factors' initial engagement with the genome. Cell *151*, 994–1004.

Southern, E.M., Maskos, U., and Elder, J.K. (1992). Analyzing and comparing nucleic acid sequences by hybridization to arrays of oligonucleotides: Evaluation using experimental models. Genomics *13*, 1008–1017.

Soza-Ried, C., Bleul, C.C., Schorpp, M., and Boehm, T. (2008). Maintenance of thymic epithelial phenotype requires extrinsic signals in mouse and zebrafish. J. Immunol. *181*, 5272–5277.

Spyrou, C., Stark, R., Lynch, A.G., and Tavaré, S. (2009). BayesPeak: Bayesian analysis of ChIP-seq data. BMC Bioinformatics *10*, 299.

Stanger, B.Z. (2015). Cellular homeostasis and repair in the mammalian liver. Annu. Rev. Physiol. *77*, 179–200.

Stoddart, D., Heron, A.J., Mikhailova, E., Maglia, G., and Bayley, H. (2009). Single-nucleotide discrimination in immobilized DNA oligonucleotides with a biological nanopore. Proc. Natl. Acad. Sci. U. S. A. *106*, 7702–7707.

Stottmann, R.W., Anderson, R.M., and Klingensmith, J. (2001). The BMP antagonists Chordin and Noggin have essential but redundant roles in mouse mandibular outgrowth. Dev. Biol. *240*, 457–473.

Stoughton, R.B. (2005). Applications of DNA microarrays in biology. Annu. Rev. Biochem. *74*, 53–82.

St-Pierre, C., Brochu, S., Vanegas, J.R., Dumont-Lagacé, M., Lemieux, S., and Perreault, C. (2013). Transcriptome sequencing of neonatal thymic epithelial cells. Sci. Rep. *3*, 1860.

Su, D.M., and Manley, N.R. (2000). Hoxa3 and pax1 transcription factors regulate the ability of fetal thymic epithelial cells to promote thymocyte development. J. Immunol. Baltim. Md 1950 *164*, 5753–5760.

Su, D., Ellis, S., Napier, A., Lee, K., and Manley, N.R. (2001). Hoxa3 and Pax1 regulate epithelial cell death and proliferation during thymus and parathyroid organogenesis. Dev. Biol. *236*, 316–329.

Su, D., Navarre, S., Oh, W., Condie, B.G., and Manley, N.R. (2003). A domain of Foxn1 required for crosstalk-dependent thymic epithelial cell differentiation. Nat. Immunol. *4*, 1128–1135.

Subramanian, A., Tamayo, P., Mootha, V.K., Mukherjee, S., Ebert, B.L., Gillette, M.A., Paulovich, A., Pomeroy, S.L., Golub, T.R., Lander, E.S., et al. (2005). Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc. Natl. Acad. Sci. U. S. A. *102*, 15545–15550.

Suda, T., Suda, J., and Ogawa, M. (1984). Disparate differentiation in mouse hemopoietic colonies derived from paired progenitors. Proc. Natl. Acad. Sci. U. S. A. *81*, 2520–2524.

Sun, T.-T., Eichner, R., Schermer, A., Cooper, D., Nelson, W., Weiss, R., and Sun, T. (1984). Classification, expression, and possible mechanisms of evolution of mammalian epithelial keratins: A unifying model T.-T. Sun, R Eichner, A Schermer, D. COOPER, W. NELSON, R. WEISS, TT Sun, J. A. Schermer, D Cooper, WG Nelson, RA Weiss, TT Sun.

Sun, X., Xu, J., Lu, H., Liu, W., Miao, Z., Sui, X., Liu, H., Su, L., Du, W., He, Q., et al. (2013). Directed differentiation of human embryonic stem cells into thymic epithelial progenitor-like cells reconstitutes the thymic microenvironment in vivo. Cell Stem Cell *13*, 230–236.

Swann, J.B., and Boehm, T. (2007). Back to the beginning – the quest for thymic epithelial stem cells. Eur. J. Immunol. *37*, 2364–2366.

Swann, J.B., Happe, C., and Boehm, T. (2017). Elevated levels of Wnt signaling disrupt thymus morphogenesis and function. Sci. Rep. *7*.

Szalkowski, A.M., and Schmid, C.D. (2011). Rapid innovation in ChIP-seq peak-calling algorithms is outdistancing benchmarking efforts. Brief. Bioinform. *12*, 626–633.

Takahama, Y. (2006). Journey through the thymus: stromal guides for T-cell development and selection. Nat. Rev. Immunol. *6*, 127–135.

Takahama, Y., Ohigashi, I., Baik, S., and Anderson, G. (2017). Generation of diversity in thymic epithelial cells. Nat. Rev. Immunol. *17*, 295–305.

Tata, P.R., and Rajagopal, J. (2017). Plasticity in the lung: making and breaking cell identity. Development *144*, 755–766.

Taunton, J., Hassig, C.A., and Schreiber, S.L. (1996). A mammalian histone deacetylase related to the yeast transcriptional regulator Rpd3p. Science *272*, 408–411.

Thomas-Chollier, M., Defrance, M., Medina-Rivera, A., Sand, O., Herrmann, C., Thieffry, D., and van Helden, J. (2011). RSAT 2011: regulatory sequence analysis tools. Nucleic Acids Res. *39*, W86-91.

Thorvaldsdottir, H., Robinson, J.T., and Mesirov, J.P. (2013). Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. Brief. Bioinform. *14*, 178–192.

Tompa, M., Li, N., Bailey, T.L., Church, G.M., De Moor, B., Eskin, E., Favorov, A.V., Frith, M.C., Fu, Y., Kent, W.J., et al. (2005). Assessing computational tools for the discovery of transcription factor binding sites. Nat. Biotechnol. *23*, 137–144.

Tran, N.T.L., and Huang, C.-H. (2014). A survey of motif finding Web tools for detecting binding site motifs in ChIP-Seq data. Biol. Direct *9*, 4.

Trapnell, C., Pachter, L., and Salzberg, S.L. (2009). TopHat: Discovering splice junctions with RNA-Seq. Bioinforma. Oxf. Engl. *25*, 1105–1111.

Trapnell, C., Roberts, A., Goff, L., Pertea, G., Kim, D., Kelley, D.R., Pimentel, H., Salzberg, S.L., Rinn, J.L., and Pachter, L. (2012). Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. Nat. Protoc. *7*, 562–578.

Tsai, P.T. (2003). BMP4 acts upstream of FGF in modulating thymic stroma and regulating thymopoiesis. Blood *102*, 3947–3953.

Tykocinski, L.-O., Sinemus, A., and Kyewski, B. (2008). The thymus medulla slowly yields its secrets. Ann. N. Y. Acad. Sci. *1143*, 105–122.

Ulyanchenko, S., O'Neill, K.E., Medley, T., Farley, A.M., Vaidya, H.J., Cook, A.M., Blair, N.F., and Blackburn, C.C. (2016). Identification of a bipotent epithelial progenitor population in the adult thymus. Cell Rep. *14*, 2819–2832.

Urban, J. (2014). How does bowtie2 assign MAPQ scores?

Vaidya, H.J., Briones Leon, A., and Blackburn, C.C. (2016). FOXN1 in thymus organogenesis and development. Eur. J. Immunol. *46*, 1826–1837.

Venter, J.C., Adams, M.D., Myers, E.W., Li, P.W., Mural, R.J., Sutton, G.G., Smith, H.O., Yandell, M., Evans, C.A., Holt, R.A., et al. (2001). The sequence of the human genome. Science *291*, 1304–1351.

Visvader, J.E., and Stingl, J. (2014). Mammary stem cells and the differentiation hierarchy: current status and perspectives. Genes Dev. *28*, 1143–1158.

Vitelli, F., Huynh, T., and Baldini, A. (2009). Gain of function of Tbx1 affects pharyngeal and heart development in the mouse. Genes. N. Y. N 2000 *47*, 188–195.

Voelkerding, K.V., Dames, S.A., and Durtschi, J.D. (2009). Next-generation sequencing: from basic research to diagnostics. Clin. Chem. *55*, 641–658.

Waldburger, J.-M. (2003). Promoter IV of the class II transactivator gene is essential for positive selection of CD4+ T cells. Blood *101*, 3550–3559.

Waldburger, J.-M., Suter, T., Fontana, A., Acha-Orbea, H., and Reith, W. (2001). Selective abrogation of major histocompatibility complex class II expression on extrahematopoietic cells in mice lacking promoter IV of the class II transactivator gene. J. Exp. Med. *194*, 393–406.

Wallin, J., Eibel, H., Neubüser, A., Wilting, J., Koseki, H., and Balling, R. (1996). Pax1 is expressed during development of the thymus epithelium and is required for normal T-cell maturation. Dev. Camb. Engl. *122*, 23–30.

Wang, R., Hsu, H.-K., Blattler, A., Wang, Y., Lan, X., Wang, Y., Hsu, P.-Y., Leu, Y.-W., Huang, T.H.-M., Farnham, P.J., et al. (2013). LOcating Non-Unique matched Tags (LONUT) to improve the detection of the enriched regions for ChIP-seq data. PLoS ONE *8*, e67788.

Wang, Z., Gerstein, M., and Snyder, M. (2009). RNA-Seq: A revolutionary tool for transcriptomics. Nat. Rev. Genet. *10*, 57–63.

Watson, J.D., and Crick, F.H.C. (1953). Molecular structure of nucleic acids: A structure for deoxyribose nucleic acid. Nature *171*, 737–738.

Watt, F.M. (1999). Stem cell manifesto. Cell *96*, 470–473.

Watt, F.M., and Jensen, K.B. (2009). Epidermal stem cell diversity and quiescence: Epidermal stem cells. EMBO Mol. Med. *1*, 260–267.

Watt, S.M., Thomas, J.A., Murdoch, S.J., Kearney, L., Chang, S.E., and Bartek, J. (1991). Human thymic epithelial cells are frequently transformed by retroviral vectors encoding simian virus 40. Cell. Immunol. *138*, 456–472.

Watts, J.A., Zhang, C., Klein-Szanto, A.J., Kormish, J.D., Fu, J., Zhang, M.Q., and Zaret, K.S. (2011). Study of FoxA pioneer factor at silent genes reveals Rfx-repressed enhancer at Cdx2 and a potential indicator of esophageal adenocarcinoma development. PLoS Genet. *7*, e1002277.

Wei, Q., and Condie, B.G. (2011). A focused in situ hybridization screen identifies candidate transcriptional regulators of thymic epithelial cell development and function. PLoS ONE *6*, e26795.

Weiner, L., Han, R., Scicchitano, B.M., Li, J., Hasegawa, K., Grossi, M., Lee, D., and Brissette, J.L. (2007). Dedicated epithelial recipient cells determine pigmentation patterns. Cell *130*, 932–942.

Weis, L., and Reinberg, D. (1992). Transcription by RNA polymerase II: Initiator-directed formation of transcription-competent complexes. FASEB J. *6*, 3300–3309.

Willis, I.M. (1993). RNA polymerase III. Genes, factors and transcriptional specificity. Eur. J. Biochem. *212*, 1–11.

Witze, E.S., Litman, E.S., Argast, G.M., Moon, R.T., and Ahn, N.G. (2008). Wnt5a control of cell polarity and directional movement by polarized redistribution of adhesion receptors. Science *320*, 365–369.

Workman, J.L. (2006). Nucleosome displacement in transcription. Genes Dev. *20*, 2009–2017.

Wortis, H.H., Nehlsen, S., and Owen, J.J. (1971). Abnormal development of the thymus in "nude" mice. J. Exp. Med. *134*, 681–692.

Wu, L., Antica, M., Johnson, G.R., Scollay, R., and Shortman, K. (1991). Developmental potential of the earliest precursor cells from the adult mouse thymus. J. Exp. Med. *174*, 1617–1627.

Xu, H., Handoko, L., Wei, X., Ye, C., Sheng, J., Wei, C.-L., Lin, F., and Sung, W.-K. (2010). A signal-noise model for significance analysis of ChIP-seq with negative control. Bioinforma. Oxf. Engl. *26*, 1199–1204.

Xu, P.-X., Zheng, W., Laclef, C., Maire, P., Maas, R.L., Peters, H., and Xu, X. (2002). Eya1 is required for the morphogenesis of mammalian thymus, parathyroid and thyroid. Dev. Camb. Engl. *129*, 3033–3044.

Yamaguchi, T.P., Bradley, A., McMahon, A.P., and Jones, S. (1999). A Wnt5a pathway underlies outgrowth of multiple structures in the vertebrate embryo. Dev. Camb. Engl. *126*, 1211–1223.

Yamamoto, H., Awada, C., Matsumoto, S., Kaneiwa, T., Sugimoto, T., Takao, T., and Kikuchi, A. (2015). Basolateral secretion of Wnt5a in polarized epithelial cells is required for apical lumen formation. J. Cell Sci. *128*, 1051–1063.

Yang, S.J., Ahn, S., Park, C.S., Holmes, K.L., Westrup, J., Chang, C.H., and Kim, M.G. (2006). The quantitative assessment of MHC II on thymic epithelium: implications in cortical thymocyte development. Int. Immunol. *18*, 729–739.

Yates, A., Akanni, W., Amode, M.R., Barrell, D., Billis, K., Carvalho-Silva, D., Cummins, C., Clapham, P., Fitzgerald, S., Gil, L., et al. (2016). Ensembl 2016. Nucleic Acids Res. *44*, D710–D716.

Yeh, J.R., Zhang, X., and Nagano, M.C. (2011). Wnt5a is a cell-extrinsic factor that supports self-renewal of mouse spermatogonial stem cells. J. Cell Sci. *124*, 2357–2366.

Yu, G., Wang, L.-G., and He, Q.-Y. (2015). ChIPseeker: An R/Bioconductor package for ChIP peak annotation, comparison and visualization. Bioinformatics *31*, 2382–2383.

Zamisch, M., Moore-Scott, B., Su, D., Lucas, P.J., Manley, N., and Richie, E.R. (2005). Ontogeny and regulation of IL-7-expressing thymic epithelial cells. J. Immunol. Baltim. Md 1950 *174*, 60–67.

Zang, C., Schones, D.E., Zeng, C., Cui, K., Zhao, K., and Peng, W. (2009). A clustering approach for identification of enriched domains from histone modification ChIP-Seq data. Bioinforma. Oxf. Engl. *25*, 1952–1958.

Zaret, K.S., and Carroll, J.S. (2011). Pioneer transcription factors: establishing competence for gene expression. Genes Dev. *25*, 2227–2241.

Zaret, K.S., and Mango, S.E. (2016). Pioneer transcription factors, chromatin dynamics, and cell fate control. Curr. Opin. Genet. Dev. *37*, 76–81.

Zentner, G.E., Tesar, P.J., and Scacheri, P.C. (2011). Epigenetic signatures distinguish multiple classes of enhancers with distinct cellular functions. Genome Res. *21*, 1273–1283.

Zhang, Y., Liu, T., Meyer, C.A., Eeckhoute, J., Johnson, D.S., Bernstein, B.E., Nussbaum, C., Myers, R.M., Brown, M., Li, W., et al. (2008). Model-based analysis of ChIP-seq (MACS). Genome Biol. *9*, R137.

Zhang, Z., Song, Y., Zhao, X., Zhang, X., Fermin, C., and Chen, Y. (2002). Rescue of cleft palate in Msx1-deficient mice by transgenic Bmp4 reveals a network of BMP and Shh signaling in the regulation of mammalian palatogenesis. Dev. Camb. Engl. *129*, 4135–4146.

Zou, D., Silvius, D., Davenport, J., Grifone, R., Maire, P., and Xu, P.-X. (2006). Patterning of the third pharyngeal pouch into thymus/parathyroid by Six and Eya1. Dev. Biol. *293*, 499–512.

Zuklys, S., Gill, J., Keller, M.P., Hauri-Hohl, M., Zhanybekova, S., Balciunaite, G., Na, K.-J., Jeker, L.T., Hafen, K., Tsukamoto, N., et al. (2009). Stabilized b-catenin in thymic epithelial cells blocks thymus development and function. J. Immunol. *182*, 2997–3007.

Žuklys, S., Handel, A., Zhanybekova, S., Govani, F., Keller, M., Maio, S., Mayer, C.E., Teh, H.Y., Hafen, K., Gallone, G., et al. (2016). Foxn1 regulates key target genes essential for T cell development in postnatal thymic epithelial cells. Nat. Immunol. *17*, 1206–1215.