

# **Mechanistic insights into the effect of RUNX1/ETO knockdown in t(8;21) AML**

Anna Pickin

A thesis submitted to the University of Birmingham for the degree of  
DOCTOR OF PHILOSOPHY

Institute of Biomedical Research  
Birmingham Centre for Genomic Biology  
College of Medical and Dental Sciences  
University of Birmingham

September 2016

UNIVERSITY OF  
BIRMINGHAM

**University of Birmingham Research Archive**

**e-theses repository**

This unpublished thesis/dissertation is copyright of the author and/or third parties. The intellectual property rights of the author or third parties in respect of this work are as defined by The Copyright Designs and Patents Act 1988 or as modified by any successor legislation.

Any use made of information contained in this thesis/dissertation must be in accordance with that legislation and must be properly acknowledged. Further distribution or reproduction in any format is prohibited without the permission of the copyright holder.

## ABSTRACT

The mutation of transcription factor genes is a main cause for acute myeloid leukaemia. RUNX1/ETO, the product of the t(8;21) chromosomal translocation, subverts normal blood cell development by impairing myeloid differentiation. RUNX1/ETO knockdown alleviates this block, with a global reprogramming of transcription factor binding and initiation of myeloid differentiation. Co-depletion of the myeloid transcription factor C/EBP $\alpha$  with RUNX1/ETO suppressed this differentiation response. Furthermore, C/EBP $\alpha$  overexpression largely phenocopied the effect of RUNX1/ETO knockdown. Our data show that low levels of C/EBP $\alpha$  are critical to the maintenance of t(8;21) AML and that C/EBP $\alpha$  drives the response to RUNX1/ETO depletion.

To examine how changes in transcription factor binding impact on the activity of cis-regulatory elements we mapped genome wide promoter-distal-element interactions in a t(8;21) AML cell line, via Capture HiC, and found that hundreds of interactions were altered by RUNX1/ETO knockdown. Differentially interacting elements exhibited changes in C/EBP $\alpha$  binding and were enriched for the CTCF motif. Our results demonstrate that the presence or absence of RUNX1/ETO has a profound impact on the intra-nuclear organisation of t(8;21) AML cells, and indicate which transcription factors are driving these changes. This work provides a novel mechanism for the RUNX1/ETO mediated differentiation block in t(8;21) AML.

## AKNOWLEDGEMENTS

Foremost, I express my sincerest gratitude to Professor Constanze Bonifer for her constant support, encouragement and immense knowledge. She has been an excellent supervisor and, despite an extremely busy schedule, always found time for me.

I thank Dr Cameron Osborne for his generosity, patience and stimulating scientific discussion during my stay in London to learn Capture HiC. I would also like to thank Professor Ruud Delwel and Dr Stefan Gröschel for giving me the opportunity to work in the Netherlands and for their very thorough training. These experiences have been invaluable.

A special thank you goes to Dr Anetta Ptasinska. Not only has she been a great academic supervisor, she has also given me unwavering emotional support. It has been a pleasure working with her during these four years.

Thank you also to Dr Salam Assi for all of her bioinformatic analysis. I really appreciate the time spent getting to grips with the complicated methods in this study.

I thank the other members of the Bonifer-Cockerill laboratory and last, but by no means least, I would like to thank my friends and family. Although they could not always completely understand the trials and tribulations of laboratory life, they were always there for me.

This work was funded by Cancer Research UK

## TABLE OF CONTENTS

<b>Chapter 1. INTRODUCTOIN.....</b>	<b>1</b>
1.1 Chromatin structure .....	1
1.1.1 The nucleosome .....	1
1.1.2 Histone variants.....	4
1.1.3 Chromatin remodelers .....	5
1.1.4 Chromatin modifications.....	6
1.2 Eukaryotic transcription and its regulation .....	9
1.2.1 The Eukaryotic transcriptional machinery .....	10
1.2.2 Role of transcription factors.....	13
1.2.3 Transcriptional regulatory elements.....	15
1.3 Higher-order chromatin structure .....	18
1.3.1 Topologically associating domains .....	18
1.3.2 Chromosome compartments .....	20
1.3.3 Chromosome territories.....	22
1.3.4 Transcription factories .....	22
1.3.5 The beta-globin locus and the HOX gene cluster .....	23
1.4 Interrogating nuclear organisation and chromosome conformation.....	24
1.4.1 Assays based on microscopy .....	25
1.4.2 Chromosome conformation capture.....	27

1.5 Haematopoiesis .....	31
1.5.1 Origin of haematopoietic stem cells .....	32
1.5.2 The maintenance of haematopoietic stem cells - the stem cell niche.....	33
1.5.3 Hierarchical differentiation of haematopoietic stem cells.....	33
1.5.4 Myelopoiesis is controlled by transcription factors .....	40
1.6 Acute myeloid leukaemia .....	43
1.6.1 AML can be caused by mutations in haematopoietic transcription factors .....	43
1.6.2 <i>CEBPA</i> mutations and expression levels in AML .....	46
1.6.3 Clonal evolution of leukaemia.....	49
1.6.4 AML with the t(8;21) translocation .....	50
1.6.5 Secondary mutations are required for the development of t(8;21) AML.....	51
1.7 Molecular pathogenesis of t(8;21) leukaemia .....	54
1.7.1 The function of RUNX1 transcription factor.....	54
1.7.2 RUNX1/ETO exhibits dominant inhibition of RUNX1 function.....	56
1.7.3 RUNX1/ETO has effects distinct from RUNX1 inhibition.....	57
1.7.4 RUNX1/ETO interferes with the activity of various transcription factors .....	59
1.7.5 RUNX1/ETO alters the epigenetic landscape .....	61
1.7.6 RUNX1/ETO functions in an oligomeric complex.....	63
1.7.7 The effect of RUNX1/ETO knockdown in t(8;21) AML .....	64
1.7.8 The role of C/EBP $\alpha$ in t(8;21) AML .....	69
1.8 The role of SP1 in t(8;21) AML.....	71

1.9 t(8;21) AML is dependent on wild type RUNX1 for survival .....	73
1.9.1 Targeting transcription factor interaction.....	73
1.9.2 Targeting RUNX1 with small molecule inhibitors.....	75
1.10 Aims and Objectives .....	77
<b>Chapter 2. METHODS .....</b>	<b>80</b>
2.1 Cell line culture .....	80
2.2 Small molecule inhibitor treatment.....	80
2.3 siRNA and shRNA mediated RUNX1/ETO and C/EBP $\alpha$ depletion .....	81
2.4 Western blotting.....	82
2.5 Flow cytometry .....	83
2.6 Apoptosis detection assay .....	84
2.7 Extraction of RNA, cDNA synthesis and analysis of gene expression analysis .....	84
2.8 Chromatin Immunoprecipitation (ChIP) .....	87
2.9 Retroviral production.....	89
2.9.1 Transfection of HEK293T cells for retroviral production.....	89
2.9.2 Virus concentration .....	90
2.9.3 Retroviral transduction with RetroNectin.....	90
2.10 Circularised chromosome conformation capture (4C-seq) .....	91
2.10.1 Fixation and cell lysis .....	91
2.10.2 Digestion with primary restriction enzyme.....	92
2.10.3 Ligation and DNA purification .....	92

2.10.4 Digestion with secondary restriction enzyme .....	93
2.10.5 Second ligation and DNA purification .....	93
2.10.6 View-point fragment selection and primer design .....	94
2.10.7 PCR amplification and sequencing .....	95
2.10.8 Differential analysis of 4C interactions .....	95
2.11 Capture HiC.....	99
2.11.1 Fixation and lysis of cells .....	99
2.11.2 HindIII digestion, biotinylation and ligation of digested DNA ends .....	100
2.11.3 Crosslink reversal and DNA purification.....	101
2.11.4 HiC ligation efficiency and quality controls.....	102
2.11.5 Removal of biotin from non-ligated DNA ends .....	104
2.11.6 DNA shearing and end repair .....	104
2.11.7 A-tailing and size selection.....	105
2.11.8 Biotin-streptavidin pulldown and adapter ligation .....	105
2.11.9 Test PCRs to determine conditions for Hi-C library amplification.....	106
2.11.10 Final PCR amplification of Hi-C library.....	107
2.11.11 Hybridization of Hi-C library with biotin-RNA .....	107
2.11.12 Capture HiC library amplification .....	109
2.11.13 Capture HiC data analysis.....	109



<b>Chapter 3. RESULTS.....</b>	<b>111</b>
3.1 CBF complex inhibition in t(8;21) AML cells .....	111
3.1.1 Treatment of HL60 and Kasumi-1 cells with CBF complex inhibitor induces apoptosis in a dose dependant manner .....	111
3.1.2 The CBF complex inhibitor has no significant effect on transcription factor binding..	117
3.2 The effect of RUNX1/ETO depletion on gene expression and the role of Sp1 in t(8;21) AML .	119
3.2.1 RUNX1/ETO depletion leads to the upregulation of genes involved in myelomonocytic differentiation .....	120
3.2.2 The SP1 motif is more protected from DNaseI digestion after RUNX1/ETO knockdown .....	123
3.2.3 RUNX1/ETO knockdown has no effect on the expression of Sp1 .....	124
3.2.4 Sp1 and RUNX1/ETO bind to distinct sites in the genome.....	126
3.3 The importance of C/EBP $\alpha$ expression levels in t(8;21) AML .....	129
3.3.1 C/EBP $\alpha$ is required for the full upregulation of myeloid genes after RUNX1/ETO depletion .....	130
3.3.2 Activation of a $\beta$ -Estradiol inducible form of C/EBP $\alpha$ in t(8;21) AML cells alleviates differentiation block.....	132
3.3.3 Genes that are differentially expressed after RUNX1/ETO knockdown are correlated with genes that are differentially expressed by C/EBP $\alpha$ induction.....	135
3.3.4 Almost half of the genes differentially expressed by RUNX1/ETO knockdown are also differentially expressed by C/EBP $\alpha$ induction .....	140
3.4 The effect of RUNX1/ETO knockdown on selected promoter-enhancer interactions .....	144
3.4.1 A reciprocal interaction between the <i>SP1</i> promoter and its upstream regulatory element (URE) was detected.....	147

3.4.2 A reciprocal interaction between the CD34 promoter and its URE was detected.....	149
3.4.3 4C-seq data was highly reproducible .....	151
3.4.4 Genomic proximity cannot always be used as a predictor of enhancer function.....	151
3.4.5 RUNX1/ETO depletion has no significant effect on the interactions between cis-regulatory elements at <i>SPI1</i> and <i>CD34</i> loci .....	153
3.5 The effect of RUNX1/ETO knockdown on genome wide <i>cis</i> -regulatory element interactions	156
3.5.1 The majority of DNA interactions in Kasumi-1 cells are intrachromosomal .....	157
3.5.2 Interaction frequency decreases with genomic distance .....	159
3.5.3 RUNX1/ETO knockdown led to alterations in specific promoter- <i>cis</i> element interactions .....	163
3.5.4 Capture HiC data support the 4C-seq data at the <i>CD34</i> and <i>SPI1</i> loci .....	166
3.5.5 The gene expression changes after RUNX1/ETO knockdown are associated with alterations in cis-regulatory element interactions but are not directly correlated .....	168
3.5.6 RUNX1/ETO depletion led to differential interactions at the <i>CEBPA</i> locus.....	175
3.5.7 C/EBP $\alpha$ and CTCF play a major role in mediating differential cis-regulatory element interactions before and after RUNX1/ETO knock-down .....	178
<b>Chapter 4. DISCUSSION .....</b>	<b>185</b>
4.1 CBF complex inhibition in t(8;21) AML cells .....	185
4.1.1 Treatment of HL60 and Kasumi-1 cells with a CBF complex inhibitor induces apoptosis in a dose dependant manner .....	185
4.1.2 The CBF complex inhibitor has no significant effect on transcription factor binding..	186
4.2 RUNX1/ETO depletion led to the up regulation of genes involved in myeloid differentiation	189

4.2.1 RUNX1/ETO knockdown has no significant effect on SP1 DNA binding .....	190
4.3 The importance of C/EBP $\alpha$ expression levels in t(8;21) AML .....	192
4.3.1 C/EBP $\alpha$ is required for the full upregulation of myeloid genes after RUNX1/ETO depletion .....	192
4.3.2 Activation of a $\beta$ -Estradiol inducible form of C/EBP $\alpha$ in t(8;21) AML cells alleviates the RUNX1/ETO mediated differentiation block .....	194
4.4 The effect of RUNX1/ETO knockdown on specific cis-regulatory element interactions .....	197
4.4.1 RUNX1/ETO knockdown had no effect on specific cis-regulatory element interactions at the <i>SPI1</i> and <i>CD34</i> loci.....	197
4.5 The effect of RUNX1/ETO knockdown on genome wide DNA interactions .....	200
4.5.1 The majority of DNA interactions are intrachromosomal with a trend of decreased interaction frequency with genomic distance.....	201
4.5.2 RUNX1/ETO knockdown led to alterations in specific promoter- <i>cis</i> -element interactions that did not correlate with gene expression change .....	203
4.5.3 DNA interactions at the <i>CEBPA</i> locus .....	203
4.5.4 CTCF and C/EBP $\alpha$ may play a role in reshaping the promoter- <i>cis</i> -element interaction profile following RUNX1/ETO knockdown .....	205
4.6 Future work.....	207
4.6.1 Targeting RUNX1 in t(8;21) AML .....	207
4.6.2 The role of C/EBP $\alpha$ in t(8;21) AML .....	207
4.6.3 Promoter- <i>cis</i> -element interactions in t(8;21) .....	208

## LIST OF FIGURES

Figure 1-1: Nucleosome structure .....	2
Figure 1-2: The chromatin fibre .....	3
Figure 1-3: The basal eukaryotic transcriptional machinery.....	12
Figure 1-4: Chromatin is folded at different length scales by combinations of structural proteins	19
Figure 1-5: Topologically associating domains are organised into two separate compartments. ...	21
Figure 1-6: Fluorescence in situ hybridization imaging of chromosome territories.....	26
Figure 1-7: Fluorescence in situ hybridization imaging of specific promoter-enhancer interactions. .....	26
Figure 1-8: Simplistic overview of 3C based chromosome conformation capture techniques .....	28
Figure 1-9: Classical Weissman model of haematopoietic differentiation .....	35
Figure 1-10: The updated model of haematopoietic differentiation.....	37
Figure 1-11: A new model of haematopoiesis - multipotent cells differentiate directly into unipotent cells by adulthood .....	39
Figure 1-12: C/EBP $\alpha$ dysregulation in acute myeloid leukaemia.....	49
Figure 1-13: The t(8;21) chromosomal translocation .....	51
Figure 1-14: Schematic of interactions within the RUNX1/ETO complex.....	64
Figure 1-15: RUNX1 and RUNX1/ETO compete for the same DNA binding sites .....	67
Figure 1-16: RUNX1/ETO knockdown leads to a genome wide increase in the binding of CEBP $\alpha$ and RUNX1 .....	68
Figure 1-17: Bar graph demonstrating the fold change of DNaseI peaks after RUNX1/ETO knockdown.....	70
Figure 1-18: The CEBP motif is enriched at DNaseI hypersensitive sites that are unique to RUNX1/ETO knockdown.....	70
Figure 1-19: Targeting Runt domain to CBF $\beta$ -SMMHC interaction with allosteric small molecule inhibitors .....	76
Figure 2-1: Overview of the 4C method.....	98
Figure 2-2: Schematic of Capture Hi-C experimental procedure.....	99
Figure 2-3: Capture-HiC quality controls.....	103

Figure 3-1: Treatment of HL60 and Kasumi-1 cells with CBF complex inhibitor induced apoptosis in a dose dependant manner.....	116
Figure 3-2: The CBF complex inhibitor had no significant effect on transcription factor binding.	118
Figure 3-3: RUNX1/ETO depletion led to the up regulation of genes involved in haematopoietic differentiation .....	122
Figure 3-4: RUNX1/ETO knockdown increased protection of the Sp1 motif.....	125
Figure 3-5: RUNX1/ETO knockdown had no effect on the expression of Sp1 .....	125
Figure 3-6: Sp1 and RUNX1/ETO bound to distinct sites in the genome and RUNX1/ETO knockdown had no effect on Sp1 binding.....	128
Figure 3-7: C/EBP $\alpha$ is required for the full upregulation of myeloid genes after RUNX1/ETO depletion .....	131
Figure 3-8: Activation of a $\beta$ -Estradiol inducible form of C/EBP $\alpha$ in t(8;21) AML cells alleviated differentiation block with a reduction in self renewal.....	134
Figure 3-9: Genes that are differentially expressed by RUNX1/ETO knockdown were correlated with genes that were differentially expressed by C/EBP $\alpha$ induction .....	139
Figure 3-10: Almost half of the genes differentially expressed by RUNX1/ETO knockdown were also differentially expressed by C/EBP $\alpha$ induction .....	142
Figure 3-11: RUNX1/ETO target genes that were inhibited by C/EBP $\alpha$ knockdown were upregulated by induction of C/EBP $\alpha$ .....	143
Figure 3-12: <i>SPI1</i> and <i>CD34</i> loci were selected for 4C experiments as they are bound by RUNX1/ETO, RUNX1 and C/EBP $\alpha$ and are differentially expressed after RUNX1/ETO knockdown .....	146
Figure 3-13: 4C-seq detected a reciprocal interaction between the <i>SPI1</i> promoter and URE .....	148
Figure 3-14: 4C-seq detected specific interactions at the <i>CD34</i> locus.....	150
Figure 3-15: 4C-seq data was highly reproducible.....	152
Figure 3-16: Differential analysis revealed no significant difference in interaction frequency after RUNX1/ETO depletion.....	155
Figure 3-17: 10 Mb resolution genome wide contact matrix showed the majority of DNA interactions are intrachromosomal .....	158
Figure 3-18: The 1 Mb resolution contact matrix of chromosome 8 showed a trend of decreased interaction frequency with genomic distance .....	160
Figure 3-19: Capture HiC interaction data was analysed to extract statistically significant interactions. ....	162

Figure 3-20: Capture HiC interactions were filtered using DNaseI-seq data .....	164
Figure 3-21: RUNX1/ETO knockdown led to statistically significant differences in interaction strength.....	165
Figure 3-22: Capture HiC data agreed with 4C data .....	167
Figure 3-23: Differential interactions were not directly correlated with differential gene expression. ....	169
Figure 3-24: Differentially expressed haematopoietic gene promoters, such as <i>SELPLG</i> , <i>SRGN</i> and <i>PRAM1</i> , had both increased and decreased <i>cis</i> -element interactions after RUNX1/ETO knockdown .....	174
Figure 3-25: The C/EBPA promoter engaged in a strong interaction with a putative enhancer at +29 kb following RUNX1/ETO knockdown .....	176
Figure 3-26: The <i>CEBPG-CEBPA</i> promoter-promoter interaction was detected with either <i>CEBPA</i> or <i>CEBPG</i> promoters as the 'viewpoint' .....	177
Figure 3-27: Differential interactions were bound by RUNX1/ETO, RUNX1 and C/EBP $\alpha$ , and contained the CTCF motif .....	180
Figure 3-28: Differential interactions exhibited increased C/EBP $\alpha$ and RUNX1 binding after RUNX1/ETO knockdown, particularly those that increased in interaction strength. ....	181
Figure 3-29: Differentially interacting DNaseI hypersensitive sites were enriched for CTCF motifs .....	183
Figure 3-30: The density of CTCF motifs in DNaseI footprints decreased following RUNX1/ETO knockdown .....	183
Figure 4-1: CBF complex inhibitors block <i>de novo</i> RUNX1 binding domain.....	188

## LIST OF TABLES

Table 1: Eukaryotic histone variants and their functions. (table adapted from (16)) .....	5
Table 2: Overview of histone modifications, their site and role in transcription .....	9
Table 3: The cytogenetic abnormalities and gene mutations associated with t(8;21) leukaemia ..	54
Table 4: Small molecule inhibitors of the core binding complex.....	81
Table 5: Western blot antibodies.....	82
Table 6: Flow Cytometry Antibodies .....	83
Table 7: Primers used for qPCR.....	86
Table 8: Primers used for ChIP-qPCR .....	86
Table 9: ChIP Antibodies .....	89
Table 10: Viewpoint specific 4C-seq PCR primers.....	96
Table 11: Genomic co-ordinates of selected viewpoint restriction fragments .....	97
Table 12: 3C PCR primers .....	102

## LIST OF ABBREVIATIONS

°C	Degrees Celsius
µg	Microgram
AGM	Aorta-gonad mesonephros
AML	Acute myeloid leukaemia
ATP	Adenosine triphosphate
BMCP	Basophil-mast cell progenitor
bp	Base pair
BSA	Bovine serum albumin
CBF	Core binding factor
CBP	CREB-binding protein
CD34	Cluster of differentiation 34
cDNA	Complementary deoxyribonucleic acid
CEBP	CCAAT-enhancer-binding proteins
Ch18	Chromosome 18
CLP	Common lymphoid progenitor
CMP	Common myeloid progenitor
CT	Chromosome territory
CTCF	CCCTC-binding factor
dATP	Deoxyadenosine triphosphate
DCE	Downstream core element
dCTP	Deoxycytidine triphosphate
dGTP	Deoxyguanosine triphosphate
DNA	Deoxyribonucleic acid
DNase I	Deoxyribonuclease I
dNTP	Deoxynucleotide triphosphate



DPE	Downstream promoter element
dTTP	Deoxythymidine triphosphate
EDTA	Ethylenediaminetetraacetic acie
EGTA	Ethylene glycol-bis( $\beta$ -aminoethyl ether)- N,N,N',N'-tetraacetic acid
ESC	Embryonic stem cell
FACS	Florescence activated cell sorting
FISH	Florescence in situ hybridisation
FITC	Fluorescein isothiocyanate
Fli-1	Friend leukaemia integration 1 transcription factor
FLT-3	Fms-like tyrosine kinase 3
FRET	Fluorescence resonance energy transfer
Gag	Group antigens
GAPDH	Glyceraldehyde 3-Phosphate dehydrogenase
GFI1	Growth factor independent 1
GFP	Green fluorescent protein
HAT	Histone acetyltransferase complexes
HEK	Human Embryonic Kidney
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
hrs	Hours
HSC	Hematopoietic stem cell
iPSC	Induced pluripotent stem cell
IRES	Internal ribosome entry site
IRF8	Interferon regulatory factor 8
ISWI	Imitation SWI
IVL	Ivolucrin
Kb	Kilobase
LBD1	Lim domain binding 1

LCR	Locus control region
LiCl	Lithium chloride
LMPP	Lymphoid-primed multi-potent progenitors
LMO2	LIM domain only 2
LT-HSC	Long-term hematopoietic stem cell
M	Molar
Mb	Megabase
mA	Micro amp
MAPK	Mitogen-activated protein kinase
ml	Millilitres
MLV	Murine leukaemia virus
MPP	Multipotential progenitors
ms	Milliseconds
MSCV	Murine stem cell virus
MTE	Motif ten element
NaAC	Sodium acetate
NaCl	Sodium chloride
NEB	New England Biolabs
Ng	Nanogram
nm	Nanometer
NP-40	Nonyl phenoxyethoxyethanol
oligo dT	Oligo deoxythymidine
PBS	Phosphate buffered saline
PCR	Polymerase chain reaction
PCR	Polymerase chain reaction
Pen/Strep	Penicillin-streptomycin
PIC	Pre-initiation complex

PMSF	Phenylmethane sulfonyl fluoride
Prot K	Proteinase K
qPCR	Quantitative PCR
REP	Replicate
RNA	Ribonucleic acid
RNase	Ribonuclease
RPMI	Roswell Park Memorial Institute
RUNX1	Runt-related transcription factor 1
SEM	Standard error of mean
SDS	Sodium dodecyl sulfate
SMMHC	Smooth-muscle myosin heavy chain
SOX4	Sry-box 4
SP1	Specificity protein 1
SPRI	Solid phase reversible immobilization
ST-HSC	Short-term hematopoietic stem cell
TAD	Topologically associating domain
TBST	Tris buffered saline with Tween 20
TAF	TBP-associated factor
TBP	TATA-box-binding protein
TFBS	Transcription factor binding site
TFIID	Transcription Factor II D
TFIIE	Transcription Factor II E
TFIIH	Transcription Factor II H
Tris-HCl	Tris(hydroxymethyl)aminomethane Hydrochloric acid
V	Volts
μl	Microlitres

## **Chapter 1. INTRODUCTION**

### **1.1 Chromatin structure**

Chromosomes were visualised for the first time in the 19<sup>th</sup> century, but it was not until much later it was discovered that they contain the genetic material of the cell, the genome. The genome is made up of billions of DNA bases, which carry the information to create complex organisms, made up of hundreds of different cell types. All of this information is compressed into the nucleus of a single cell. This is no easy feat; roughly two metres of linear DNA must be packaged into the nucleus, which is a mere six thousandths of a millimetre in diameter.

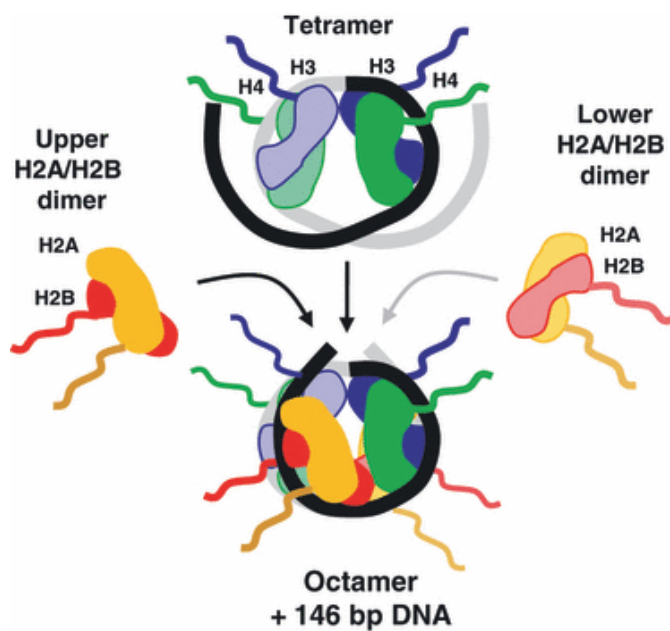
The way in which it is organised must also be highly dynamic, to facilitate the complex processes which take place during the cell cycle. For example, during mitosis, the chromosomes are very condensed to ensure transmission of genetic material to daughter cells. The requirements are more complex during interphase; the genome must be organised in such a way that gene transcription, DNA replication and DNA repair can be tightly regulated. The following section will outline how this is achieved.

#### **1.1.1 The nucleosome**

The first step of DNA packaging is nucleosome assembly. The DNA double helix is spooled around a protein complex, 147 base pairs (bp) at a time, to form the smallest component of chromatin - the nucleosome. The protein core of the nucleosome consists of 8 histone molecules; two molecules of histone H2A, H2B, H3 and H4. H2A and H2B form dimers and

H3 and H4 form dimers, which combine to form a compact complex. Each of the histones has a long, protruding amino acid 'tail' that extends clear of the central core (figure 1-1). These tails can be covalently modified, which enables regulation of various aspects of chromatin structure (1). This is discussed in section 1.1.4.

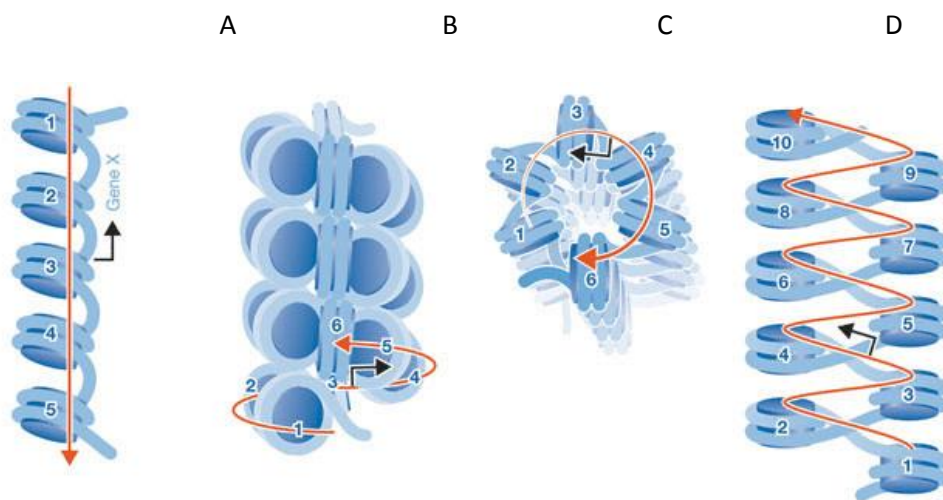
Multiple nucleosomes are linked with an approximately 20 bp to 80 bp linker region to form a chromatin thread, commonly described as having a 'beads on a string' appearance. This first level of packaging already reduces the length of the genome by approximately one third (1).



**Figure 1-1: Nucleosome structure**

This schematic illustrates the assembly of the core histone octamer. Two H3/H4 dimers are incorporated, followed by two H2A/H2B dimers (image modified from (1)).

The 'beads on a string' structure is condensed further by packing nucleosomes on top of each other, in a highly ordered fashion, to form an approximately 30 nm chromatin fibre (2, 3) (figure 1-2). The way in which nucleosomes are packed together resembles a 'zig-zag' pattern (4). An additional histone, histone H1, binds to the nucleosome, contacting both protein and DNA. It protects linker DNA from nuclease digestion and promotes the packing of nucleosomal DNA into this structure (5, 6). The way in which it assists in packaging is still unclear. One theory is that histone H1 can bend DNA, thus increasing DNA compaction rate (7).



**Figure 1-2: The chromatin fibre**

A) 10 nm fibre B) Horizontal view of a 30 nm fibre C) Vertical view of the 30 nm fibre D) Zig-zag model of the 30 nm fibre organisation (figure adapted from (8)).

### **1.1.2 Histone variants**

Each of the canonical histones found in the nucleosome has a repertoire of variants that differ in their amino acid sequence, primarily at the N-terminus. The canonical histones are mostly made during DNA replication in the S-phase of the cell cycle, so are incorporated into chromatin in a DNA replication-dependent fashion. In contrast, the histone variants can be incorporated independently to DNA replication and are expressed during all stages of the cell-cycle (9).

The histone variants are brought to the nucleosome by histone chaperones where they can then recruit chromatin modifiers (9-13). The replacement of a canonical histone with its variant introduces structural changes. This can alter the interaction strength between histone proteins within the nucleosome, which therefore affects nucleosome stability and level of chromatin compaction. For example, the deposition of histone variants H2A.Z and H3.3 are associated with open and active chromatin (14). Whereas incorporation of macroH2A leads to stabilisation of the nucleosome and is usually associated with a repressive chromatin state (15). The availability of histone variants is therefore considered to add another level of complexity to the genome and plays an important role in modulating chromatin function.

Histone variant	Function	Conserved?
CENP-A/CID/cse4	Epigenetic marker of the centromere	Yes
H3.3	Transcription	Yes
H2A.Z/H2AV	Transcription/double strand break repair	Yes
H2A.X	Double strand break repair/meiotic remodelling of sex chromosomes	Yes
macroH2A	Gene silencing/X chromosome inactivation	Yes
H2A.Bbd	Epigenetic mark of active chromatin	Yes
H3.Z	Regulation of cellular response to outside stimuli	No
H3.Y	Regulation of cellular response to outside stimuli	No

**Table 1: Eukaryotic histone variants and their functions.** (table adapted from (16))

### 1.1.3 Chromatin remodelers

The positioning of nucleosomes must be highly dynamic and there are times when the DNA needs to be temporarily less compacted. This is achieved by an active process termed 'nucleosome remodeling'. This is usually implemented to enable access of proteins to the DNA to drive important biological processes, such as gene transcription and DNA replication. The rearrangement is orchestrated by 'chromatin remodeling complexes', which use ATP to temporarily alter the nucleosome structure to reduce how tightly bound the DNA is to the histone core (17). There are several different remodeling complexes, but most are very large complexes made up of many protein subunits. The best known are the ISWI, SWI/SNF, and SWR1 complex families. The exact way in which they remodel varies depending on the type of complex. Briefly, the ISWI family is thought to 'slide' nucleosomes along the DNA, SWI/SNF proteins displace nucleosomes and the SWR1 family replace core histones with a histone variant (H2A/H2B dimers are replaced with H2AZ/H2B dimers) (18-20).



#### 1.1.4 Chromatin modifications

Core histone protein tails are subject to a multitude of covalent modifications, at several different amino acid residues (see table 1 for key examples), by specific nuclear enzymes. Thus far, 60 residues have been detected, but it is likely that there are many more. The modifications include lysine acetylation, lysine methylation and serine phosphorylation. Further complexity is added by the fact these residues come in different forms – mono, di or tri and mono or di for methylation and acetylation respectively. Their effect on the stability of the chromatin and therefore accessibility of the DNA to other proteins is tremendously influential to gene expression and therefore the regulation of cellular processes.

There is an increasing body of literature which highlights both the variety and biological specificity linked to particular patterns of modifications (21-24). This has sparked the analogy that histone modifications are a ‘language’ which is ‘read’ by other proteins to trigger downstream events. This language has been termed ‘the histone code’ (25). The different combinations of histone modifications (the histone code) allow the cell to carefully regulate contacts with the underlying DNA. The enzymes which transduce the information can be very specific to the modification, and its amino acid positions, thus adding an extra level of information to the genetic (DNA) code.

Acetylation of histone H3 and histone H4, and di/tri methylation of H3K4 are associated with active transcription, and are commonly termed ‘euchromatin modifications’. The modifications associated with transcriptional repression, ‘heterochromatin

modifications', include methylation of H3K9 and H3K27. It is not only the particular modification and the residue which is important for regulation, the localisation on the linear genomic sequence also plays a role – for example, whether a modification is at the promoter, the 3' end or the 5' end of an open reading frame will alter its effect on transcription (26, 27).

Histone modifying enzymes are responsible for carefully controlling this dynamic arrangement of modifications on chromatin. In general, acetylation is laid down by histone acetyltransferase complexes (HATs), of which there are many. Key examples of HATs include HAT1, p300 and GCN5 (28). Acetylation can then be removed by histone deacetylases (HDACs). There are approximately 18 mammalian HDACs which can be divided into two main classes. Class I are ubiquitously expressed in human cells and have a nuclear localisation. Class I includes HDACs 1,2,3 and 8. Class II can be found in the cytoplasm as well as the nucleus, suggesting that they may deacetylate non-histone proteins. Class II are expressed in a tissue specific fashion and include HDACs 4,6,7,9 and 10 (29). Similarly, methyl groups are added by histone methyl transferases, and removed by histone demethylases.

A long standing model is that histone modifications alter the net charge of nucleosomes, and thus weaken inter or intra-nucleosomal interactions. The strongest support for this comes from studies with acetylated histones. Several groups have found that acetylated histones are easier to separate from DNA both in vivo and in vitro (30, 31) (32, 33).

Histone modification can also serve as binding sites for effector proteins. The earliest example is the recognition and selective interaction of protein bromodomains with acetylated histones (34). Once bound to the chromatin via the acetylated histones, bromodomain containing proteins can then recruit other proteins to form a complex (34). Bromodomains therefore allow the regulation of protein-protein interaction via lysine acetylation. Recent research has highlighted the importance of this mechanism to the regulation of gene transcription and chromatin organisation (reviewed in (35)).

The histone code is also 'read' by components of the polycomb pathway, which silences gene expression. Polycomb proteins recognise methylated histones which then triggers downstream processes. Polycomb repressive complex 2 (PRC2) is targeted to polycomb response elements (PRE) where it methylates local histones. This leads to the accumulation of H3K27Me<sub>3</sub>, which in turn attracts polycomb repressive complex 1 (PRC1) and leads to a block of transcription (36)

Modification	Role in transcription	Modification site
Acetylation	Activation	H3 (K9,K14,K18,K56) H4 (K5,K8,K12,K16) H2B (K6,K7,K16,K17)
Methylation	Activation	H3(K4me2,K4me3, K36me3,K79me2)
Methylation	Repression	H3(K9me3,K27me3) H4 (K20me3)
Phosphorylation	Activation	H3(S10)

**Table 2: Overview of histone modifications, their site and role in transcription**  
(adapted from (37))

## 1.2 Eukaryotic transcription and its regulation

Every eukaryotic organism has a vast diversity of cell phenotypes, despite each cell carrying an identical genome. These differences are largely driven by difference in gene expression, which leads to the controlled production of a specified set of proteins in the cell, via the processes of transcription and translation. This complement of proteins defines the cell type, and further variations to the protein levels and timing of expression can alter the cells behaviour during different biological processes. The following section will outline how gene expression is regulated.

### 1.2.1 The Eukaryotic transcriptional machinery

RNA polymerase II (Pol II), along with several other proteins, is responsible for the transcription of protein coding genes in eukaryotes. The other proteins include general transcription factors, activator proteins and co-activators. General transcription factors assemble on the core promoter to form the pre-initiation complex (PIC). The assembly of the pre-initiation complex usually begins with the binding of TFIID, a complex made up of TATA-box-binding protein (TBP) and several other TBP-associated factors (TAFs) (figure 1-3). The PIC then recruits Pol II to the transcription start site; recruitment of Pol II is a rate limiting step for transcription (38). The recruitment of Pol II to the DNA is often termed 'initiation'.

Once recruited, Pol II can begin the process of elongation, which can be divided into two main parts. It begins with the incorporation of nucleotides to the 3' end of Pol II, so it can transcribe the first 20 to 50 nucleotides of RNA (39). At almost 50% of active genes, Pol II then pauses after this initial early transcription. Three models have been proposed to explain how and why this promoter proximal pausing occurs. The first model, the 'Kinetic model' states that the pause is due to the fact that elongation is energetically unfavourable, due to the recruitment of pausing factors to the DNA (40). The 'barrier model' proposes that Pol II pauses due to nucleosomes acting as a physical barrier which hinders elongation (41). The 'interaction model' involves factors which bind to DNA and Pol II at the same time, 'tethering' them and pausing elongation (42). Progression beyond

this 'promoter proximal' pausing is often tightly controlled so is therefore recognized as an important step in gene regulation (43).

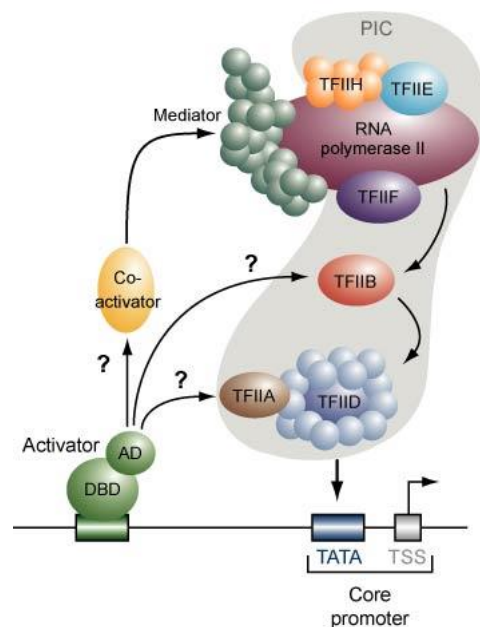
The escape of Pol II from pausing is dependent on two processes taking place during its residence at the promoter; the 5' capping of the nascent RNA (39) and the phosphorylation of the paused complex (44). Phosphorylation is mediated by P-TEFb, a heterodimer complex of Cyclin-dependent kinase 9 (Cdk9) and Cyclin T (CycT) (44). P-TEFb phosphorylates serine 2 of Pol II (45), as well as pausing factors DSIF and NELF, leading to their release (46). P-TEFb can be recruited via the coactivator Brd4 (47, 48).

Once Pol II is released from its paused state it can begin productive elongation, where it progresses along the gene body. The elongating Pol II inevitably encounters nucleosome barriers, which it can overcome with the help of elongation factors, for example FACT and Spt6 (49). FACT does not displace nucleosomes. Instead, it removes H2A-H2B dimers from the octamer, which allows Pol II to transcribe through the remaining components (50). Spt6 is thought to directly interact with histones, displacing the nucleosome in front of Pol II then reassembling it behind (51). Spt6 has also been shown to increase the elongation rate of Pol II, even when no nucleosomes are present (52, 53).

Pol II continues elongation through the gene body until it eventually reaches the end of the gene. To mark the end of the transcript, the nascent RNA is then cleaved and polyadenylated, usually approximately 8 kb downstream from the 3' end of the gene (54).

After Pol II has escaped the core promoter and began the process of elongation, the PIC does not completely disassemble. TFIID, TFIIE, TFIIH and Mediator remain as a 'scaffold' for

efficient subsequent reinitiation of transcription via recruitment of Pol II (55). Furthermore, RNA polymerase can be maintained at the same gene through several cycles of transcription, via a facilitated recycling pathway (56). There is also evidence that DNA looping could bring the same polymerase molecule from the end to the beginning of the gene, to reinitiate transcription (57).



**Figure 1-3: The basal eukaryotic transcriptional machinery**

The PIC, consisting of general transcription factors (including RNA polymerase II, TFIIA, TFIIB, TFIID, TFIIE, TFIIF and TFIIH), assembles at the core promoter in a stepwise fashion. This targets RNA polymerase II to the transcription start site (TSS). The transcriptional activity can be increased by activators (green), which bind to distal elements. Activators usually contain a distinct DNA-binding domain (DBD) and an activation domain (AD). Question marked arrows represent the fact that the direct targets of activators have not yet been defined (58).

### 1.2.2 Role of transcription factors

The rate of transcription can be dramatically increased or decreased by other transcription factors. Transcription factors are modular proteins that increase (activators) or decrease (repressors) the rate of gene transcription. Transcription factors are often members of multiprotein families, for example, CCAAT-enhancer-binding proteins and Sp family proteins. The members of the family usually have very similar DNA binding properties but varied effects on transcription. In general, they repress transcription by blocking the general transcriptional machinery, and activate by the following main mechanisms:

1. Promoting the assembly of the PIC at the core promoter (12).
2. Stimulating the activity of the general transcriptional machinery by, for example, post-translation modification, subsequent conformational change and the recruitment of co activators (59-61).
3. Interacting with chromatin remodelers to open chromatin, thus allowing access of other general transcription factors to *cis* elements (62).
4. Bringing together enhancers and promoters by physically 'looping' out intervening DNA. RUNX1 and GATA2 are examples of transcription factors that mediate DNA looping (63) (64)



Transcription factors usually consist of a DNA binding domain, transactivation domain and interaction domain. The three dimensional shape, and amino acid composition, of the DNA binding domains determines the affinity to specific bases of DNA, thereby providing sequence specific DNA binding. There are several main categories of DNA binding domain or 'motifs'. These include; helix-turn-helix, zinc finger, basic leucine zipper and the basic-loop-helix (65).

Coactivators are proteins which can modulate the activity of activator proteins. In contrast to activators, they generally do not have sequence-specific DNA binding activity, so are recruited by protein-protein interaction with DNA bound activators. They function via two main mechanisms. The first class of activators stimulate PIC assembly. The second class are chromatin remodeling or modifying enzymes. Coactivators and activators cooperate to act synergistically, presumably via post binding interactions (66).

A key example of the first class of activators is TBP associating factors (TAFs). They are proposed to act as a link between the sequence specific DNA binding activators and the basal transcriptional machinery, to stimulate activator dependant transcription. This model is based on experiments which detected a direct interaction of TAFs with various activation domains (67). Furthermore, mutation of the TAF binding sites leads to the failure to activate transcription (68). Interestingly, select TAF mutations can lead to defects in specific subsets of genes, suggesting that TAFs help to regulate genes in a gene/activator specific manner (69-71).

The second class of activators are associated with histone modification. An early experiment in yeast, in which the *Tetrahymena* histone acetyltransferase A was found to be homologous to the coactivator Gcn5, was one of the first to link coactivator function with histone acetylation (72). Several mammalian coactivators have since been identified as having HAT activity. One of the most well studied is the CREB-binding protein (CBP) and the E1A-interacting protein p300 (reviewed in (73)). The TBP associating factor TAF250 has also been shown to have intrinsic HAT activity (74) .

### **1.2.3 Transcriptional regulatory elements**

Only approximately 1.5% of the eukaryotic genome is protein-coding. However, the remaining DNA is not redundant. The typical protein-coding gene is associated with several transcriptional regulatory elements that can be located immediately before the transcription start site (promoter elements) or distributed over many kilobases (long-range elements). These regulatory elements are specific DNA sequences that are recognised by transcription factors. Many cells in an organism will express the same transcription factors. However, the unique combinations of accessible elements, and the particular combination of transcription factors, can lead to great complexity and diversity of gene expression patterns.

## **Promoters**

The core promoter element defines the transcription start site. It is the binding site for PIC and the basic transcription machinery. The core promoter can contain several distinct elements. The first to be described was the TATA box – the binding site for TBP, a subunit of TFIID. Other core promoter elements include the downstream core element (DCE), the downstream promoter element (DPE) and the motif ten element (MTE). Like the TATA box, all of these serve as docking sites for TFIID. The different combinations and organisation of these elements is diverse, which is believed to serve a regulatory role. The region immediately upstream of the core promoter is called the ‘proximal promoter’ (75). It usually contains several elements which have binding sites for activator proteins.

## **Enhancers**

Gene promoter activity can be amplified by elements called ‘enhancers’ (76). Enhancer activity is often tissue and/or developmental stage specific so they play a very important role in differentiation and many biological processes. They are typically approximately 500 bp in length and contain several clusters of transcription factor binding sites (TFBS), which work together to increase transcription. In contrast to promoter elements, they can be located upstream, downstream, and within introns, sometimes over a megabase away from the promoter (77). This long range regulation is mediated by the physical interaction between the two elements, with the intervening DNA looping out. These interactions are specific, highlighted by the fact genes are often not regulated by the nearest element (78). Multiple enhancers will often act on the same promoter to coordinate expression in different cell types (79)

## Silencers

In contrast to enhancers, silencers repress transcription of their target genes. However, like enhancer elements, they can act over long distances and often work in concert. They contain binding sites for repressive cofactors, sometimes called corepressors (80).

Key examples of transcriptional corepressors are SMRT (Silencing Mediator of Retinoic acid and Thyroid hormone receptors) and N-CoR (Nuclear hormone receptor Co-Repressors). As their names suggest, they mediate repression via nuclear hormone receptors (81, 82). However, they also repress via several different transcription factors, for example PLZF, BCL-6 and ETO (83-85).

There are numerous possible mechanisms for repressor function. One proposed model is that repressors block and/or compete with coactivator binding (86). Repressors may also promote a repressive chromatin structure via the recruitment of histone modifying enzymes, thereby blocking activator binding (87). For example, ETO recruits N-CoR and SMRT, which in turn recruit histone deacetylases (HDACs) (see section 1.1.4 for more on HDACs).

Experiments in yeast have shown that repressors can hinder PIC assembly, thus suppressing transcription (88). This mechanism has since been shown in human B cells. The transcription factor PAX5 was found to inhibit transcription of the *c-fms* gene by binding to the major transcriptional start site thus blocking the assembly of the basal transcriptional machinery (89)

## **Insulators**

Insulators are also known as ‘boundary elements’ as they function by hindering interactions between neighbouring loci by sectioning the genome into discrete domains of transcriptional regulation. They function in two main ways; by blocking promoter - enhancer physical interaction and by limiting the spread of repressive chromatin (90). They are commonly bound by CCCTC-binding factor (CTCF), which maintains DNA interaction and genome partitioning, however their precise mechanism of blockade is not yet known (91).

### **1.3 Higher-order chromatin structure**

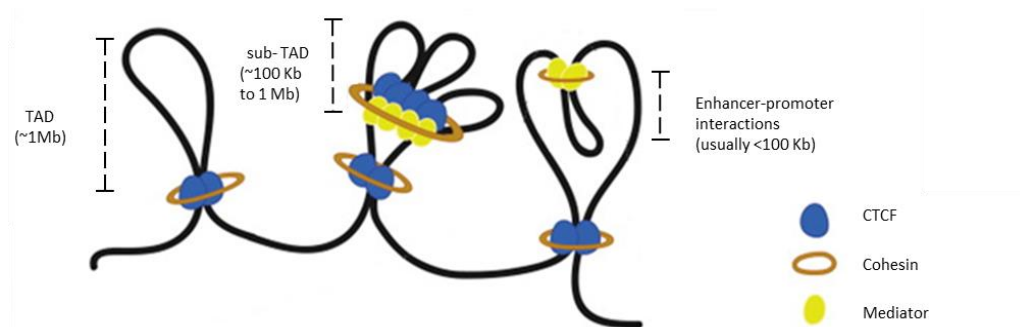
So far in this chapter, chromatin organisation has been described up to the level of the chromatin fibre. However, much further compaction is required to fit the entire genome into the nucleus. This section will outline higher order chromatin structure.

#### **1.3.1 Topologically associating domains**

The chromatin fibre is folded into a hierarchy of domains. The most robust pattern is the presence of well defined, sub-megabased sized regions termed ‘topologically associating domains’ (TADs). TADs are characterized by a high frequency of chromosomal contacts within the domain, but infrequent contacts between TADs (92). The boundaries sectioning off these regions of interaction are largely conserved between cell types of the same species and even during transcriptional activity. Genes within the same block are generally in close physical proximity, have many epigenetic features in common, and tend

to be up and down regulated in concert (92). It is due to these characteristics that the TAD is viewed as a unit of transcriptional regulation, enabling the coordination of the epigenetic status of several genes simultaneously.

TAD boundaries are defined by specific DNA interactions; either DNA to DNA loops or DNA to the nuclear laminar (93). These contacts are mediated and stabilised by several different structural proteins. For example CTCF is a major player in genome architecture by bringing together DNA and acting as an insulator between TADs (94). These DNA-CTCF interactions are commonly stabilised by cohesin (95). The transcriptional co-activator mediator can join the complex to stabilise DNA to DNA loops at active promoter regions, partitioning off 'sub-TADs' to enable cell type specific promoter-enhancer interactions. This CTCF-cohesin-mediator complex has the ability to hold together several different loops, enabling coordinated transcriptional activity (96) (figure 1-4).



**Figure 1-4: Chromatin is folded at different length scales by combinations of structural proteins** TADs are demarcated by CTCF and cohesin binding. Others factors can act in combination with these proteins to divide the Mb-sized TADs into smaller sub-TADS, facilitating shorter range DNA looping between promoters and enhancers (figure adapted from (96)).

In addition to these ubiquitous structural proteins, more cell type specific proteins have been shown to play an important role in genome architecture. For example, the Yamanaka reprogramming transcription factors mediate embryonic stem cell (ESC) specific DNA interactions, driving reprogramming into 'induced pluripotent stem cells' (iPSCs) (97). Another example comes from work on the *CD34* locus. It was shown that the transcription factor RUNX1 is required for the interaction between the *CD34* promoter and enhancer, and this interaction was needed for *CD34* gene expression (63). Furthermore, studies in *Drosophila melanogaster* have identified a role for the polycomb complex proteins in mediating DNA interactions, to bring genes into so called 'polycomb bodies' (98).

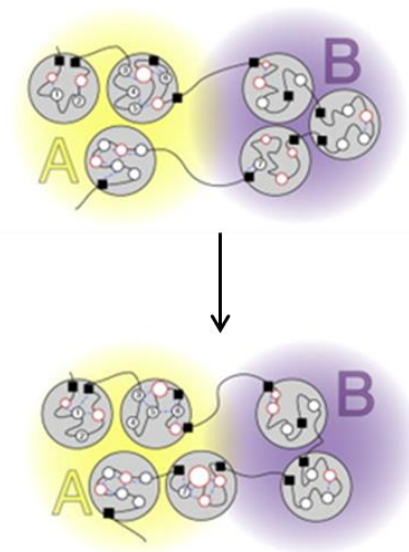
### **1.3.2 Chromosome compartments**

With the development of genome wide DNA interaction mapping (HiC), Lieberman-Aiden, E. *et al.* demonstrated that the genome is divided into two compartments –A and B. At approximately 5 Mb in size, these compartments are significantly larger than TADs. They alternate along chromosomes represent regions of active (A) and inactive (B) chromatin. A-compartments of chromosomes preferentially interact with other A compartments and vice-versa for B compartments (99) (figure 1-5).

Lieberman-Aiden, E. *et al.* demonstrated that, in contrast to TADs, these compartments are cell type specific (99). However, the genome wide differences between cell types were not comprehensively described. A more recent study from Dixon, JR *et al.* highlighted that over a third of the genome changes compartment during stem cell differentiation. These reorganisations were associated with changes in gene expression

(100). As a result of these findings they state “that the A and B compartments have a contributory but not deterministic role in determining cell-type-specific patterns of gene expression” (100).

In short, TADs are regarded as the very robust features of chromosomes, which can then be arranged into either an A or B compartments. Hence, differences in expression between cell types are more likely to be mediated via re-arrangement of TADS into alternative compartments, rather than changing TAD boundaries (see figure 1-5).



**Figure 1-5: Topologically associating domains are organised into two separate compartments.** Chromatin is packed into TADs (grey circles) which are in turn positioned into either compartment A or B, depending on their transcriptional activity. The arrangement of TADs can change in order to alter expression level of genes within a TAD. Note that TAD organisation does not change upon change of expression status (figure adapted from (101)).



### **1.3.3 Chromosome territories**

During cell division, all chromosomes are replicated and equally distributed between the two daughter cells. After division, the nuclear membrane reforms around the chromosomes and the chromosomes are distributed within sub-volumes of the nucleus – chromosome territories (CTs). The way in which the chromosomes are distributed within these CTs is not random (102-104). To a certain extent, chromosome size affects the radial position of chromosomes due to size-dependent mitotic forces acting on the chromosomes (102, 103, 105). Radial positions are also related to gene density (102). For example, gene – poor chromosome 18 is usually located closer to the outside of the nucleus than gene-dense chromosome 19 (102).

The activity states of distinct loci within a CT effects their positions in nuclear space and position in relation to each other (106-109). Genes which are active can migrate outside of their defined CT and undergo transcriptional ‘bursts’ when they join so called ‘transcription factories’ (regions with highly concentrated, active RNA polymerase II) (106, 110). Inactive genes, on the other hand, are usually located at the nuclear periphery, close to the nuclear lamina (111, 112). They can, however, migrate away from the lamina and become active, for example during differentiation (112).

### **1.3.4 Transcription factories**

The term ‘transcription factory’ was coined in 1993 (113). They are regions of the nucleus that have a micro-environment which favours transcription. In these hubs, RNA polymerase II (pol II) is active and at high levels. It is thought that pol II may be stabilised

at these discrete sites, and during transcription genes are passed through the polymerases (114, 115). Genes at a particular transcription factory are often co-regulated and are bound by the same transcription factors. Dynamic transcription regulation can be achieved by moving genes in or out of these factories (116). It is due to these properties that transcription factories are believed to improve the efficiency of transcription and its regulation.

### **1.3.5 The beta-globin locus and the HOX gene cluster**

The beta-globin locus is an excellent example of the principles described previously in this chapter. Beta and alpha globin chains combine to make haemoglobin and their regulation is very carefully controlled, as free globin chains are toxic. The beta globin locus contains a number of genes encoding beta subunits which are used at different stages of development. A far upstream locus control region (LCR) consisting of several regulatory elements, helps orchestrate the tightly controlled expression of those genes (117). When silent during the progenitor stage of erythroid development, the locus is positioned at the nuclear periphery. However, as the cells differentiate into mature blood cells, one of the beta globin genes is expressed and the locus moves away from the periphery and associates with a transcription factory, where expression levels can significantly increase. During this process, an enhancer within the LCR is brought into physical contact with the beta globin promoter via DNA looping (118-120).

The HOX gene locus is another excellent example of a link between 3D genomic arrangement and the temporal and spatial expression of genes. Work over the past few years has identified several long-range enhancers within the two gene deserts either side of the HoxD gene cluster which, during limb development, control the expression of the HoxD genes (reviewed in (121)). These flanking regulatory regions overlap with the TAD boundaries established by Dixon *et al.* (92). They are referred to as the T-DOM (telomeric TAD) and the C-DOM (centromeric TAD). Interestingly, chromosome conformation capture and gene expression analysis revealed that the T-DOM and C-DOM control the early and late phases of development respectively (122, 123). Thus, a switch between which TAD regulates HoxD expression is required for normal limb development. A very recent study from the Denis Debole group showed that HOX13 is critically involved in mediating this switch between TAD activities (124).

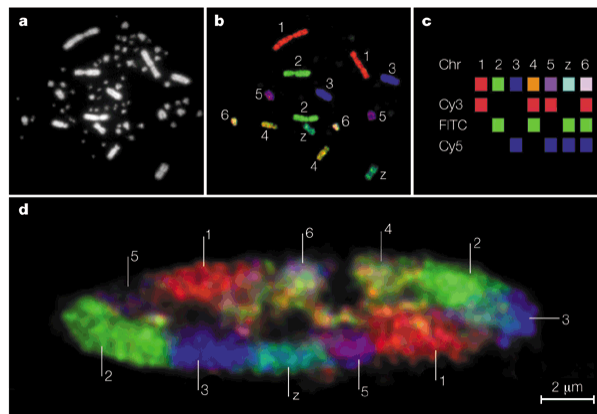
#### **1.4 Interrogating nuclear organisation and chromosome conformation**

It is now very clear that genome organisation is fundamental to many crucial biological processes, such as transcriptional regulation and DNA replication. Hence the development of techniques to interrogate chromosome conformation has been, and still is, a highly active area of research. The above findings regarding 3D genome structure were made possible with the relatively recent advent of chromosome conformation capture technologies.

#### **1.4.1 Assays based on microscopy**

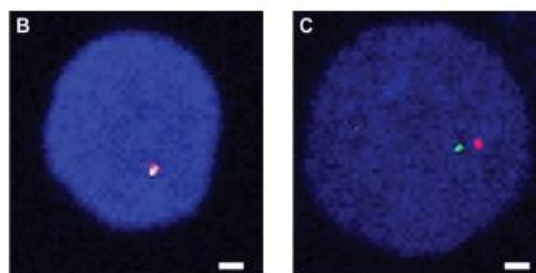
At the turn of the 20<sup>th</sup> century, it was from observations with early microscopes that Theodor Boveri introduced the term 'chromosomal territory' as a result of his work with the blastomere stages of the horse round worm (125). It wasn't until the late 1970s that more concrete evidence came from laser-UV-micro irradiation experiments (126, 127). A laser micro-beam was directed at a small part of the nucleus to induce DNA damage. The cells were allowed to enter metaphase and then, once re-condensed, the chromosomes were assessed for signs of damage. Different results were predicted to transpire, depending on the way chromosomes were arranged. If chromosomes were randomly arranged and inter-mingled together, multiple inter-chromosomal arrangements are likely to occur. In contrast, if chromosomes resided in distinct sub-compartments with minimal inter-chromosomal contact, directed damage would result in predominantly intra-chromosomal rearrangements and damage on one particular chromosome. The latter was found to be true (126, 128, 129).

With the advent of fluorescence in situ hybridisation (FISH) technology, these chromosomal territories could be directly observed, using fluorescently labelled probes complementary to entire chromosomes (130, 131) (figure 1-6). Promoter-enhancer interactions can also be directly visualised with the same technology, using probes complementary to specific loci (132, 133) (figure 1-7).



**Figure 1-6: Fluorescence in situ hybridization imaging of chromosome territories**

A) Chicken diploid metaphase spread of chromosomes. Nuclei were stained with DAPI. B) The metaphase spread was subject to multicolour fluorescence *in situ* hybridization. The probes were labelled by a combinatorial scheme with Estradiol (1, 4, 5, 6), Digoxigenin (2, 4, 6, Z) and Biotin (3, 5, 6, Z). C) These probes were detected using secondary antibodies labelled with the indicated fluorophore. D) Fluorescence microscopy was used to visualise the chromosomes in an optical section of the nucleus. You can clearly see that homologous chromosomes are positioned in separate locations – chromosome territories(131) .

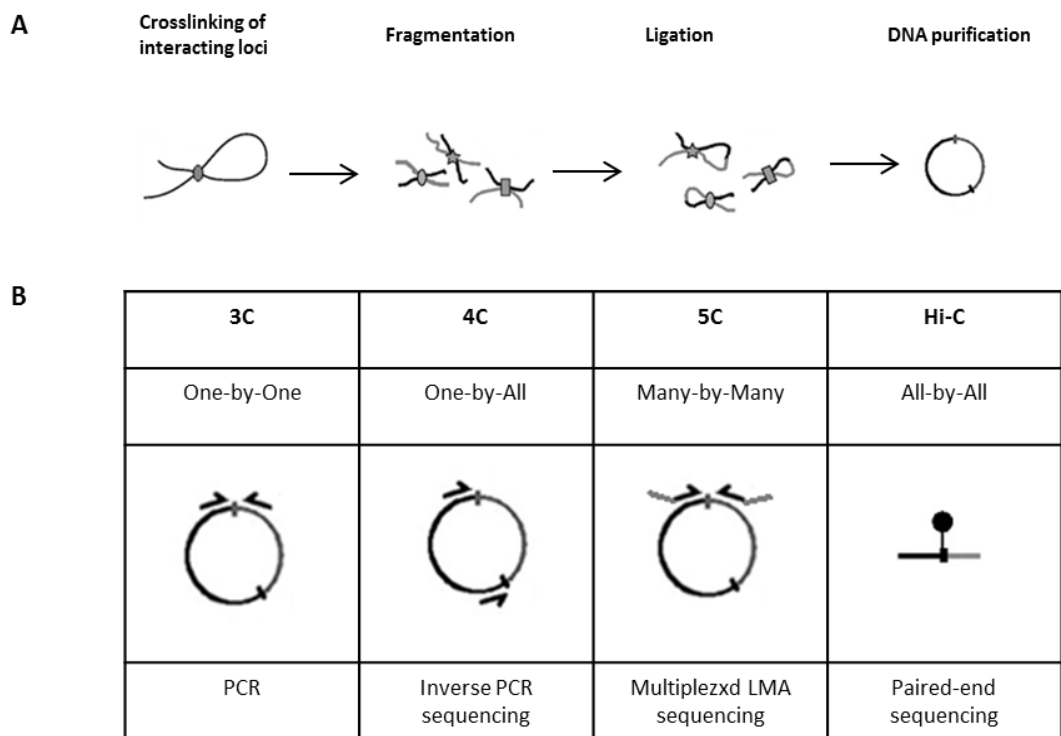


**Figure 1-7: Fluorescence in situ hybridization imaging of specific promoter-enhancer interactions.**

Sonic hedgehog (*Shh*) in mouse limb buds is regulated by a long-range enhancer 1 Mb upstream of the *Shh* promoter. B) A nucleus in which the *Shh* the promoter and enhancer are colocalised, presumably via DNA looping. C) A nucleus in which the two loci are separate. TOPRO-3 (blue) was used to stain nuclei and the white bar represent 1 µm (132).

### 1.4.2 Chromosome conformation capture

Experiments using imaging techniques have resulted in landmark discoveries. However, these techniques are limited. They have low resolution, they require cell fixation so only display 'snapshot' data and they are very low throughput. To circumvent some of these issues, molecular techniques have been developed. Rather than directly measuring interactions in a single cell via microscopy, these molecular methods deduce interactions in a population of cells, *in vivo*. The original assay is called 'Chromosome Conformation Capture' or 3C (see figure 1-8). This assay assesses the frequency that two particular loci are in close physical proximity, at a single time point, in a population of nuclei. This is achieved by formaldehyde crosslinking to fix the conformation of the genome. The DNA is then digested with a restriction enzyme and then any restriction fragment ends of the crosslinked fragments are ligated together. The result is a collection of chimeric DNA molecules consisting of two restriction fragments that were in close physical proximity. This is effectively a one dimensional demonstration of the 3D nuclear structure. In the traditional 3C experiments, to determine interaction frequency, the collection of fragments is subject to PCR amplification with primers specific to the DNA fragments of interest.



**Figure 1-8: Simplistic overview of 3C based chromosome conformation capture techniques**

A) All of the 3C based techniques follow the same fundamental steps to convert chromatin interaction into ligation products. B) The different methods vary in the way ligation products (which represent interactions) are detected and quantified (see methods section for a detailed description and depiction of the 4C and Hi-C methods)(134).

The development of this method by Dekker *et al.* in 2002 was ground breaking (107). However, this original 3C method is low throughput and requires prior prediction of which two loci are interacting; it does not allow unbiased screening. It is perhaps due to these limitations that significant efforts have been made to modify this protocol, and as a result, it is now possible to analyse several interactions in a single experiment.

One such modified protocol is called '4C', a 'one-vs-all' strategy (see figure 1-8). This refers to the fact that the technique allows the analysis of all interactions which are taking place with one selected region of the genome, the 'view-point'. This protocol

follows the same fundamental principles of the original 3C experiment; however following 3C library generation the DNA is subject to a second restriction digestion to enable the formation of small circles upon ligation. Primers specific to the view-point are then used to amplify, via inverse PCR, all fragments which are ligated to the viewpoint i.e. all fragments which were in contact with the viewpoint. This 4C library can then be analysed via either micro array or next generation sequencing (135, 136). The use of next generation sequencing allows genome wide analysis of interactions made by the loci of interest.

One limitation of 4C is that it is only possible to determine all the interactions made by one select region of the genome at a time. Therefore, with this protocol, one cannot obtain conformation information for an entire domain of a chromosome. It is for this that 5C was developed, which is referred to as a 'many vs many' approach. Although it is not a genome-wide analysis, it allows the experimenter to detect many interactions made by several restriction fragments, simultaneously. In this method, a set of oligos are designed adjacent to the restriction site of all restriction fragments in the genomic region to be interrogated. The primers are then used to amplify a 3C library. If, due to a ligation junction, two primers are next to each other, they are ligated together by *taq* ligase. This generates new DNA molecules; a combination of the two primers. This 5C library is then amplified using primers specific to a shared sequence on all 5C oligos. The resultant library is then subject to high throughput sequencing. The end result is a quantitation of interactions made by all regions covered by the original 5C oligos. The scale of 5C data can vary, depending on the number of regions covered by 5C primers.



5C has been used to interrogate large, even chromosome sized, regions of the genome (137-139). However, it wasn't until the development of HiC that scientists could obtain a true genome wide analysis of chromosome conformation. The HiC technique involves direct quantitation of all ligation junctions. Again, the protocol follows the same initial steps – a population of nuclei are fixed and digested with a restriction enzyme. In this case, the resultant sticky overhang is filled in with a deoxynucleoside triphosphates (dNTPs) mix that includes a biotinylated nucleotide. The now blunt-ended restriction fragments, within the same DNA-protein complex, are then ligated to each other. The biotinylated nucleotide enables the subsequent enrichment of the library for ligation junctions via streptavidin pull down. Illumina's paired end next generation sequencing is then used to identify the ligation junctions (140).

Since a HiC library is extremely complex, in order to obtain an informative level of signal at any given point in the genome, a huge depth of sequencing is required. Furthermore, data from bioinformatic 'windows' of the genome must be pooled together in order to obtain sufficient signal. Therefore the assay is of relatively low resolution. It is for this reason that this technique is more suitable for mapping larger, domain structures of the genome, rather than specific promoter-enhancer interactions.

To circumvent these issues; Capture HiC was developed. In this method the complexity of the library is reduced, thus improving resolution, whilst maintaining the valuable promoter-enhancer interaction data. This is achieved by enriching the HiC library so that it only contains DNA fragments with at least one gene promoter. Biotinylated oligos complementary to gene promoters are hybridized to the library, and the complexes are

captured with streptavidin beads. The technique was initially used to capture a few hundred promoters and more recently almost all promoters in the genome (22,000 promoters) (see methods sections 2.10 and 2.11 for a detailed description and depiction of the 4C and Hi-C experimental procedures) (141) (77).

## **1.5 Haematopoiesis**

Haematopoiesis is the process during which blood cells are generated within the bone marrow. The process begins with a type of cell that has the potential to give rise to each of the mature blood cell types. These cells were first purified from mouse bone marrow in 1986 (142). Subsequent experimentation demonstrated that the cell population was unique in its ability to completely reconstitute the hematopoietic system in lethally irradiated mice (143). These pluripotent cells, now termed hematopoietic stem cells (HSCs), also have the ability to self-renew, thus ensuring that the relatively short-lived mature blood cells are replenished. During haematopoiesis, the HSCs first differentiate into more restricted progenitor cells, which gradually become progressively more restricted to a specific cell fate. This is accompanied by a loss of self-renewal capacity. The whole process requires tight regulation of gene expression by various transcription factors. If not appropriately regulated, haematological disorders, such as leukaemia, can occur (144, 145).

### **1.5.1 Origin of haematopoietic stem cells**

In vertebrates, the HSCs emerge in the embryo at various sites. The initial site is the extra-embryonic yolk sack, where 'primitive haematopoiesis' takes place. This process generates nucleated primitive erythrocytes (EryP) which ensure the embryo is sufficiently oxygenated, enabling its rapid growth (145, 146). Primitive haematopoiesis is followed by definitive haematopoiesis, which takes place in the in the aorta-gonad mesonephros region (AGM), foetal liver, thymus, spleen and finally the bone marrow. In contrast to primitive haematopoiesis, definitive haematopoiesis gives rise to pluripotent cells which have the potential to differentiate into cells of all haematopoietic lineages; haematopoietic stem cells (HSCs). These cells arise from the 'haemogenic endothelium', a precursor tissue for both endothelial cells and hematopoietic cells. In a recent study, dynamic real time imaging captured the emergence of cells from the endothelium of mouse and fish embryos (147, 148). This study, along with others, resolved debates regarding the existence of the haemogenic endothelium. Although the site of haematopoiesis changes during embryonic development, bone marrow remains the major site of haematopoiesis throughout adult life, where the HSCs are maintained in a stem cell niche (149-152).

### **1.5.2 The maintenance of haematopoietic stem cells - the stem cell niche**

The adult bone marrow niche is a highly specialised microenvironment which plays a crucial role in regulating the balance between self-renewal, differentiation and proliferation of HSCs. HSCs are a subject of intense research; however they are yet to be maintained *in vitro*. This is primarily due to their requirement for this highly specialised, dynamic microenvironment which includes various cellular components and signalling pathways. It is the lack of complete understanding of all the factors involved which may be hindering their successful culture *in vitro* (153).

The HSCs within the bone marrow can exist in either of the two identified anatomical niches; the vascular or the osteoclastic niche. These spatially distinct niches are also functionally distinct, although the understanding of their precise roles is incomplete. It has been proposed that the two niches work together, performing complementary roles, to ensure the correct regulation of the cells within. The hypoxic osteoclastic niche is believed to promote a quiescent cellular state, whereas the vascular niche is oxygen rich and promotes the proliferation and further differentiation of the HSCs (154).

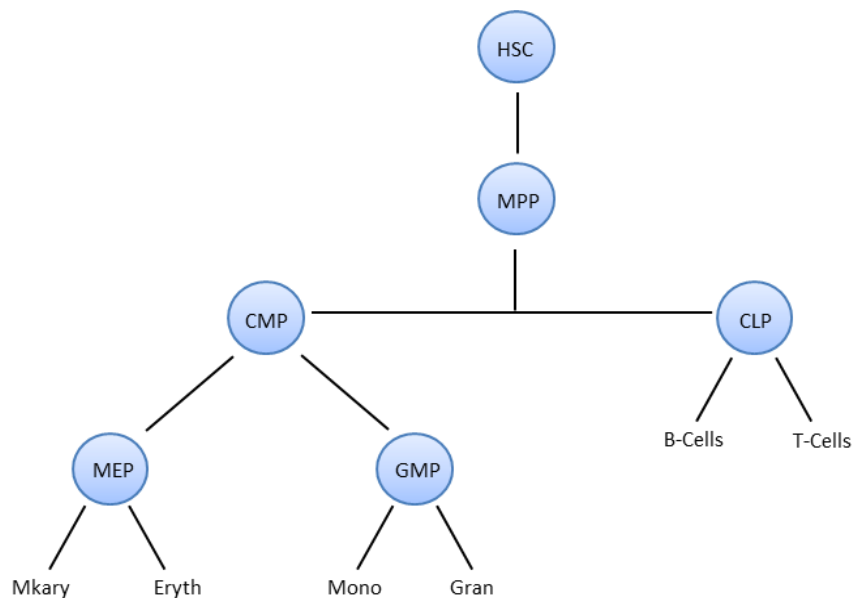
### **1.5.3 Hierarchical differentiation of haematopoietic stem cells**

The blood stem cell differentiation process is the best-defined adult stem cell system, and has long served as a model for stem cell research. In recent years considerable progress has been made in characterising the transition from a HSC into a mature blood cell. Primarily via multicolour flow cytometry and transgenic mouse studies, a hierarchical model of differentiation has been established. Much of the pioneering research was

conducted by the Weismann laboratory, who successfully identified the phenotypically distinct precursor cell populations that a HSC will progress through on its differentiation path (155, 156).

Several models of haematopoietic differentiation have since been proposed. In the original Weissman model (figure 1-9), the first group of cells, at the 'top' of the hierarchy, are the long-term HSCs (LT-HSC). These are the only population able to differentiate into any lineage and self-renew throughout adult life. The cells maintain multilineage potential, but lose some self-renewal capacity, as they transition to short-term HSCs (ST-HSC) (157). Self-renewal is then lost completely in the next cell population, the multipotential progenitors (MPPs). As the name suggests, these cells still have the potential to differentiate into any lineage. However, the developmental potential of these cells is already restricted as they cannot de-differentiate into a ST-HSC or LT-HSC (158). The next stage in this model is the transition into either a common lymphoid progenitor (CLP) or common myeloid progenitor (CMP); the earliest branch point in the differentiation process. CLPs give rise to blood cells of the lymphoid lineage (B-cells, T-cells and dendritic cells) and have lost the potential to generate cells from the myeloid lineage. Conversely, GMPs can give rise to blood cells of the myeloid lineage, but cannot differentiate into lymphoid cells (155, 156).

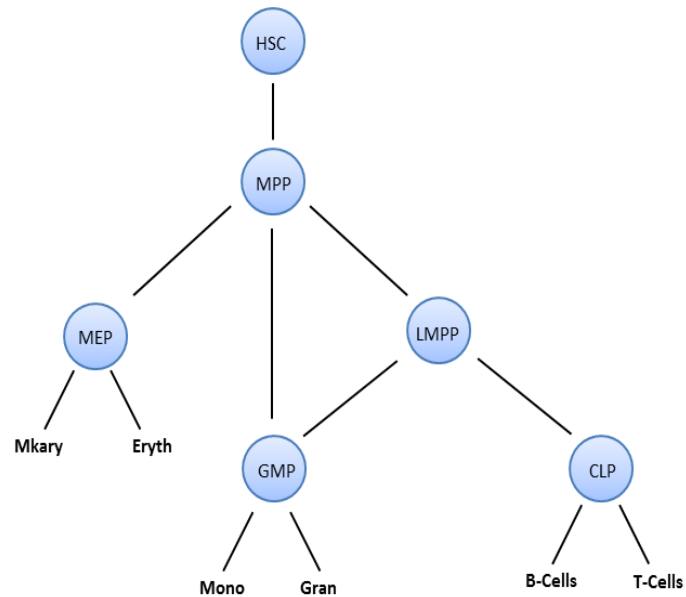
Prior to the generation of mature blood cells, CMPs progress into more specified progenitors; granulocyte/macrophage progenitors (GMPs), megakaryocyte/erythroid progenitors (MEPs) and basophil progenitors (159, 160). The origin of mast cells has been under scrutiny. There is evidence that mast cells can arise directly from MPPs, however an additional progenitor population has recently been identified in the spleens of mice; the basophil-mast cell progenitor (BMCP)(161).



**Figure 1-9: Classical Weissman model of haematopoietic differentiation**

The HSC undergoes a step wise transition into a mature blood cell. The first step of lineage commitment is the divergence into either a CMP or CLP. HSC, hematopoietic stem cell; CLP, common lymphoid progenitors; CMP, common myeloid progenitor; MEP, megakaryocyte/erythrocyte progenitors; GMP, granulocyte/monocyte progenitors; MKary, megakaryocytes; Eryth, erythrocytes; Mono, monocytes; Gran, granulocytes (162).

The initial model outlined above suggests that the first step of lineage commitment is the binary decision between myeloid or lymphoid potential. Since this model was proposed, a subset of progenitor cells, marked by the expression of lymphoid genes and Fms-like tyrosine kinase 3 (FLT-3), has been identified. It was observed that these cells had lost erythroid and megakaryocytic potential but had maintained the ability to generate lymphocytes, monocytes and granulocytes. This subset of progenitor cells was termed lymphoid-primed multi-potent progenitors (LMPPs). The new model therefore suggests that the first branch point is the decision of a MPP to differentiate into a CMP, MEP or LMPP (figure 1-10). The megakaryocytic and erythroid cells are generated from the MEPs, whilst the LMPPs have the potential to generate all other hematopoietic cells (163).



**Figure 1-10: The updated model of haematopoietic differentiation**

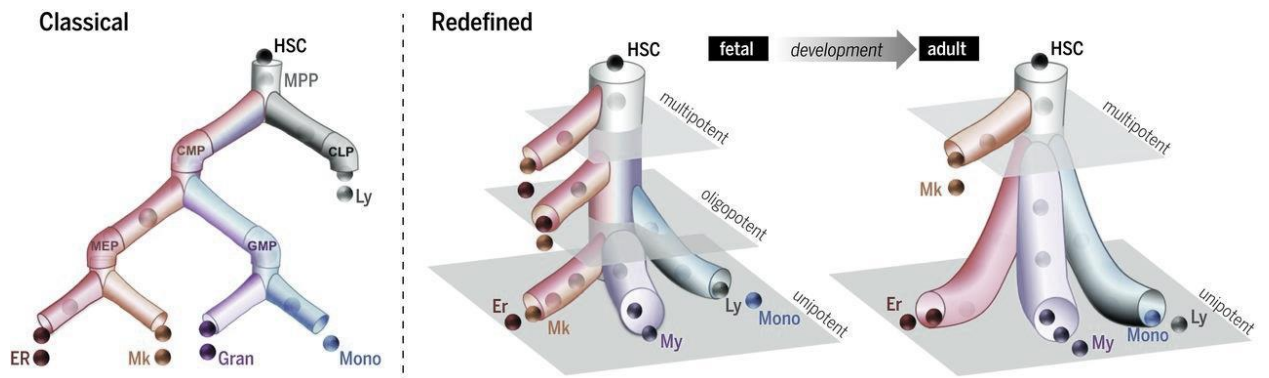
Here the first step of lineage commitment is the formation of GMPs, LMPPs or MEPs. LMPPs can give rise to any blood cell, excluding megakaryocytic or erythroid cells. HSC, hematopoietic stem cell; LMPP, lymphoid primed progenitors; CMP, common myeloid progenitor; MEP, megakaryocyte/erythrocyte progenitors; GMP, granulocyte/monocyte progenitors; MKary, megakaryocytes; Eryth, erythrocytes; Mono, monocytes; Gran, granulocyte (163).

However, there is still controversy regarding the correct scheme of differentiation. A subsequent study showed that a significant proportion of the LMPP cell population can differentiate into megakaryocytic and erythroid cells, which is contrary to the model outlined above (164). Furthermore, Arinobu Y, *et al.* propose that LMPPs are too heterogeneous to be deemed a separate stage in the differentiation tree and, as the classical model suggest, the CMP and CLP remain the first 'branching point' in hematopoiesis (165).



In the above studies haematopoietic progenitor populations were primarily defined using surface marker expression, detected via flow cytometry. Single cell RNA-seq analysis and single-cell fate determination methods have since been developed. With these exciting new technologies researchers have highlighted the heterogeneity that resides in each progenitor compartment. Therefore, new models have been proposed in which there are no progenitor cells with mixed potential, but instead multiple sub-groups transcriptionally primed towards alternative differentiation fates (166) (167, 168).

Notta *et al.*, for example, dispute the dogma that haematopoiesis always begins with multipotent stem cells and progresses through a series of progenitor 'tiers' of increasingly restricted lineage potential. They propose a new 'two tier' model (figure 1-11). Using single cell assays they mapped the origins of myeloid, erythroid and megakaryocytic cells in the foetus and in adults. They found that the previously defined MPP, CMP and MEP populations were all heterogeneous. Furthermore, they identified a change between foetal and adult haematopoiesis. In the adult, multipotent cells were only found in the stem cell compartment and oligopotent progenitors were almost non-existent. In contrast, in the foetus the ratio of multilineage to unilineage cells was constant between the stem and progenitor cell compartments (167).



**Figure 1-11: A new model of haematopoiesis - multipotent cells differentiate directly into unipotent cells by adulthood**

Classical models of haematopoiesis include oligopotent progenitors (left). The model has been redefined (right). The new model proposes that there is a shift between foetal and adult development, which changes the progenitor cell architecture. In the foetus there are multipotent progenitors whereas in the adult only the stem cell compartment is multipotent. Progenitors are unipotent (figure taken from (167)).

Drissen R, *et al.* focused on redefining the model of myelopoiesis. Using single cell analysis, they found that the previously defined GMP progenitor population actually contains two main myeloid progenitor populations; one gives rise to eosinophils and mast cells, and the other gives rise to neutrophils and monocytes-macrophages. The separation of these two populations is driven by expression levels of GATA-1 and FLT3 in multipotent progenitors (168).

#### **1.5.4 Myelopoiesis is controlled by transcription factors**

The process of stem cell specification is characterised by a loss of self-renewal and gradual acquisition of lineage specificity. The process is brought about by alterations in gene expression; differentiation is mediated by the down-regulation of self-renewal genes and up-regulation of lineage specific genes. These gene sets both counteract and co-operate with each other in a finely controlled network. This strict control is orchestrated by the carefully timed expression of specific transcription factors. This concept has been most clearly demonstrated in the processes of myelopoiesis; the regulated production of myeloid blood cells.

The first step in myeloid differentiation is the decision made by the MPP to differentiate into either an LMPP (has both myeloid and lymphoid potential) or GMPs (has only myeloid potential). A long standing theory was that the decision is primarily determined by the relative expression levels of two important transcription factors; PU.1 and GATA1. Early transgenic mouse studies revealed that PU.1 is needed for the production of LMPPs, whereas GATA1 is crucial for generation of GMPs (169, 170).

This theory was based on *In vitro* experimentation in which the forced expression of PU.1 led to a block in the erythroid pathway, and likewise the forced expression of GATA1 blocks myelopoiesis (171-173). Co-immunoprecipitation and Chromatin Immunoprecipitation (ChIP) experiments suggested that the observed antagonistic relation was due to a direct protein-protein physical interaction (173).

The paradigm that PU.1 and GATA-1 levels determine cell fate has been disputed. A very recent publication using advance single cell analysis, states that the early myeloid lineage fate decisions is not triggered by random relative levels of GATA1 and PU.1 proteins (174). Instead of driving the lineage decision, the role of PU.1 and GATA.1 expression is to reinforce the choice once it has been made.

More complicated regulatory networks are required for the further specification of MEPs and LMPPs. The MEPs can either differentiate into a megakaryocytic or an erythrocytic precursor cell. The expression levels of the transcription factors EKLF and Fli-1 are the key determinants of this decision. EKLF is implicated in the inhibition of the transcription factor activity of Fli-1, thus inhibiting the production of megakaryocytic genes. Similarly, Fli-1 can inhibit the EKLF mediated transcription and therefore megakaryocytic cell fate (175).

The resolution of an LMPP into either the lymphoid (CLPs) or the myeloid lineage (GMPs) is determined by the level of PU.1 expression. Via the gradual, targeted depletion of PU.1 from hematopoietic progenitor cells it was established that low levels of PU.1 favour B cell development, whilst a higher concentration encouraged myeloid cell development and inhibits B cell development (176). It has since been discovered that the different expression levels in different cells types is achieved by the ability of PU.1 to regulate its own expression in a cell type specific manner, by associating with cell type specific transcription factors (163, 177). C/EBP $\alpha$  levels are also very important; without C/EBP $\alpha$  mice show a complete absence of GMPs and therefore lack granulocytic differentiation (178).

At the GMP stage, cells have only the potential to generate cells of the myeloid lineage. They then branch, specifying into either a neutrophil or macrophage precursor. This decision is thought to be determined by PU.1 and C/EBP $\alpha$  levels (179). At the GMP stage, PU.1 levels are at low 'sub-threshold' levels and the transcription factor regulates genes characteristic of both neutrophils and macrophages. Once PU.1 levels are elevated, the gene expression profile is tailored towards macrophage development, via a complicated network involving the transcription factors GFI1 and ERG. PU.1 acts in a 'feed-forward' loop with ERG, activating macrophage specific genes and repressing neutrophilic genes. The neutrophilic gene expression profile is activated via a similar feed-forward loop involving C/EBP $\alpha$  and GFI1 (180).

An additional transcription factor, IRF8, is also believed to be fundamental to the generation of macrophages. IRF8 null mice have fewer macrophages relative to wild type. Furthermore, the subsequent expression of IRF8 in null mice restores macrophage differentiation. IRF8 was also found to repress the expression of granulocytic specific genes and inhibit granulocytic differentiation. This suggests IRF8 has a key role in the specification of myeloid cells (181).

## **1.6 Acute myeloid leukaemia**

Acute myeloid leukaemia (AML) is a haematopoietic malignancy defined by more than 20% of the bone marrow consisting of immature precursor cells. The incidence is most common in adults, with AML accounting for 25% of all adult leukaemias. It is caused by genetic abnormalities in HSCs, which render them unable to undergo the normal differentiation process. The result is an accumulation of immature cells and an interference with the generation of normal blood cells. This aggressive disease will spread rapidly and, if left untreated, bone marrow failure and infection can cause fatality within weeks (182).

A system devised by the World Health Organization (WHO) is used for the classification of AML. It takes into account the differentiation state of the cell and more detailed classification such as cytogenetic abnormalities. In 2008 the WHO classification was revised in order to incorporate more recently characterized genetic features (183).

### **1.6.1 AML can be caused by mutations in haematopoietic transcription factors**

As outlined in section 1.4.3, the transcription factors involved in myelopoiesis display a clear stage and lineage restricted expression pattern, which highlights the importance of their careful regulation. If any of these factors are mutated or de-regulated, the result can be an arrest in differentiation. In normal conditions, progenitor cells proliferate rapidly then undergo differentiation. The failure or inhibition of differentiation prolongs this proliferative stage, thus predisposing cells to further mutation and the development of myeloid leukaemia.

The most commonly perturbed transcription factor is RUNX1. It is frequently involved in chromosomal translocation, for example RUNX1/ETO (t(8;21)), RUNX1/EVI1 (t(3;21)), RUNX1/ETV6 (t(12;21)). In fact, the chromosomal translocation of RUNX1 is the most common cytogenic abnormality in leukaemia. Furthermore, several different point mutations of RUNX1 have been found in myeloid malignancies (184). The cofactor of RUNX1, CBF $\beta$ , is also mutated in AML. For example, inv(16) leads to the fusion of CBF $\beta$  with smooth muscle myosin heavy chain (SMMHC) (185) (details regarding RUNX1 function will be given later in section 1.7.1).

Another commonly mutated transcription factor is C/EBP $\alpha$ ; C/EBP $\alpha$  mutations are found in approximately 9% of acute myeloid leukaemia (AML) patients (186). In addition to point mutation there are several ways in which C/EBP $\alpha$  function can be altered in haematological malignancies. These include aberrant gene expression and post-transcriptional or post-translational suppression (186). The role of C/EBP $\alpha$  in leukaemia development will be explained in more detail in section 1.6.2 and section 1.7.8.

GATA-1 exists in long and short isoforms and its function can be impaired via disruption to the ratio of their expression. A mutation in the N-terminus of GATA-1 results in the loss of expression of the long isoform. The short isoform is expressed and binds target genes; however it lacks the transactivation domain thus significantly hindering its ability to regulate gene expression (187). The dominant activity of the GATA-1 mutation, and its involvement in disease generation, has been demonstrated in a transgenic mouse study (188).

Thus far, no mutations in the IRF8 transcription factor gene have been reported in myeloid leukaemia patients, despite its important role in myelopoiesis (181). However, there is substantial evidence supporting its involvement in pathogenesis. For example, IRF8 is significantly down regulated in AML and chronic myeloid leukaemia (CML) patients. In addition, knock down of the gene from mice results in a CML-like phenotype (189). Furthermore, IRF8 has also been shown to encourage malignant myelopoiesis via its synergy with the RUNX1/ETO fusion protein, the result of the t(8;21) translocation (190).

In contrast to the other transcription factors involved in myelopoiesis, details of the link between PU.1 and the transformation process are less clear. An *in vivo* study involving the gradual reduction of PU.1 expression in mice demonstrated that low PU.1 levels initiate a leukemic state. Interestingly, the complete abolition and a 50% reduction of PU.1 had no effect, whereas a clear pathogenic effect was seen with an 80% reduction. This emphasizes the requirement for precise PU.1 levels for correct myelopoiesis (191). In this study, the PU.1 levels were reduced via targeted disruption of the enhancer driving PU.1 expression. A subsequent study identified a single nucleotide polymorphism (SNP) within this highly conserved region which leads to reduced PU.1 expression levels. This SNP is common in human AML (192). This suggests that aberrant PU.1 expression levels, possibly via hindered enhancer activity, could account for, or contribute to, malignant myelopoiesis.



### 1.6.2 *CEBPA* mutations and expression levels in AML

One of the most commonly mutated transcription factors is the CCAAT enhancer binding protein alpha (C/EBP $\alpha$ ). C/EBP $\alpha$  is a member of the basic region leucine zipper family of transcription factors. It is encoded by an intronless gene and consists of two transactivation domains at the N-terminal region, and a leucine zipper region at the C-terminus (193). The leucine zipper enables dimerization dependant DNA binding. C/EBP $\alpha$  can promiscuously dimerise with other C/EBP $\alpha$  family members. Dimerisation is dependent on specific orientations of the basic amino acid residues. Any mutations that effect these exact positions can hinder DNA binding. *CEBPA* mRNA can be translated at two different AUG codons, giving rise to two different isoforms; p30 and p42. The short p30 protein begins at the AUG codon further downstream, so lacks the N terminal sequences. This means the p30 isoform lacks the regions which interact with the transcriptional machinery, whereas the ability to dimerise and bind to the DNA is preserved (194). The cellular ratio of p42/p30 is regulated by extracellular signalling via the protein kinase R and target of rapamycin signalling pathways (195).

As outlined previously, the role of mutations in lineage-specific transcription factors to leukaemia development has been an intense area of research over the past decade. *CEBPA* mutation is the most frequently studied, probably, at least in part, due to the fact that *CEBPA* mutations are found in approximately 9 % of all AML cases (196). This high prevalence of *CEBPA* mutation in myeloid-lineage leukaemia is consistent with the effect of *CEBPA* knockouts in mice. C/EBP $\alpha$  deficient mice display impaired myelopoiesis, specifically a block in granulocyte maturation, and die at birth due to hepatic dysfunction

(197). A conditional *CEBPA* knockout in adult mice blocks the CMP to GMP transition and, as a consequence, myeloid blasts accumulate. In patients, null mutations are extremely rare. Instead the *CEBPA* gene can be subject to a variety of different point mutations.

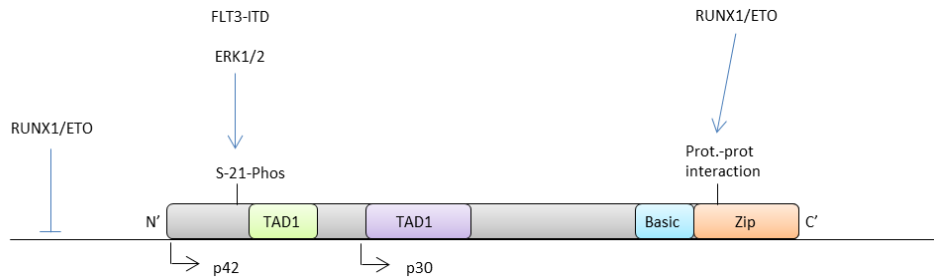
There are two main groups of point mutation. The first are N-terminal frame shift mutations which truncate the p42 isoform but leave the p30 is unaffected. The p30 isoform then inhibits the remaining wildtype p42 in a dominant negative fashion. The second group are mutations at the C-terminal which affect the structure of the basic zipper region and therefore DNA binding. Most AML patients with *CEBPA* mutation have more than one, usually a combination of both N and C terminal mutations, positioned on different *CEBPA* alleles.

As well as abnormalities effecting genomic sequence, C/EBP $\alpha$  function can be significantly affected at the expression level. For example, expression levels are extremely low in t(8;21) AML. This is due to repression of the *CEBPA* promoter by RUNX1/ETO and a physical interaction between C/EBP $\alpha$  and RUNX1/ETO proteins, which inhibits *CEBPA* autoregulation (198). Furthermore, analysis in t(8;21) cell lines and patient samples has shown binding of the repressive RUNX1/ETO at the +42 kb enhancer, an enhancer essential for *CEBPA* expression in myeloid cells specifically (199, 200).

In addition, C/EBP $\alpha$  protein is subject to posttranslational modifications which can affect its activity. Phosphorylation at specific serine residues on the transactivation domain affects its structure and means the C/EBP $\alpha$  molecules in a dimer are further apart from each other, thus activity is inhibited. The differentiation block in leukaemia with

constitutively active FLT3 may be mediated by this phosphorylation, and therefore inhibition, of C/EBP $\alpha$  by downstream ERK1/2 (201).

It is well established that *CEBPA* mutations, and aberrant expression, are linked to leukemic transformation. How these abnormalities mediate this oncogenic effect is now under investigation. One theory is that the tumour suppressive role of C/EBP $\alpha$  comes from its repression of stem cell genes, like *SOX4*. HSCs from *CEBPA* null mice have upregulated *SOX4* expression. Upon knockdown of *SOX4* in these cells, self-renewal was blocked. Consistent with this finding, *SOX4* was found to be highly expressed in patients with abnormal C/EBP $\alpha$  function (202). Therefore a dual role for C/EBP $\alpha$  has been proposed; the upregulation of the myeloid gene expression programme and the suppression of genes associated with stemness and self-renewal. In patients with abnormal C/EBP $\alpha$  activity, the inhibition of both of these processes is likely to be the driver of leukemic transformation.



**Figure 1-12: C/EBP $\alpha$  dysregulation in acute myeloid leukaemia.**

This image shows the functional domains of C/EBP $\alpha$  and some of the ways in which C/EBP $\alpha$  can be dysregulated in human AML. There are two transactivation domains (TAD1 and TAD2). The basic region (Basic) mediates DNA binding and the leucine zipper region (Zip) mediates dimerization. C/EBP $\alpha$  mRNA can be translated from either of the two ATGs (p42 or p30) to yield a 42-KDa and 30-KDa C/EBP $\alpha$  protein respectively. The p42 protein can be truncated by frame shift mutation at the N-terminus. Mutations at the C-terminus can affect DNA binding. Phosphorylation at serine 21 can lead to C/EBP $\alpha$  inactivation. This can be caused by constitutive activation of the Mitogen-activated protein kinase (MAPK) pathway by FLT3-ITD. Physical interaction with RUNX1/ETO, and RUNX1/ETO mediated repression of the C/EBP $\alpha$  promoter, inhibits C/EBP $\alpha$  autoregulation and leads to low expression levels.

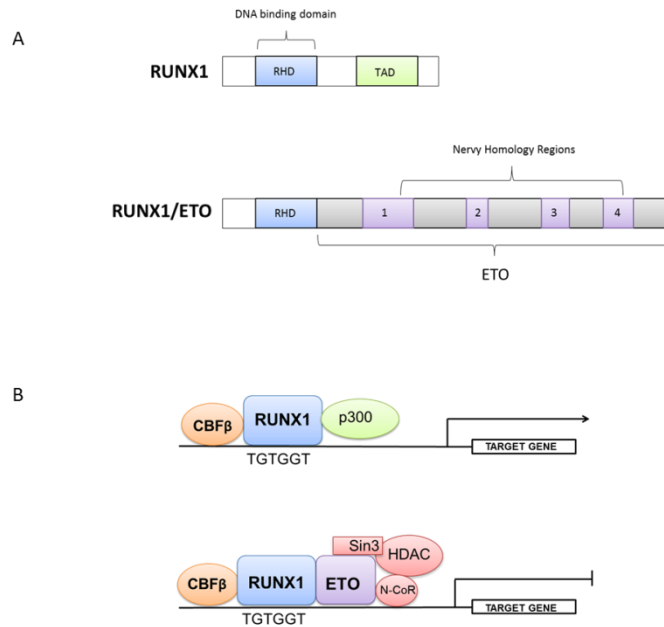
### 1.6.3 Clonal evolution of leukaemia

The clonal evolution theory of leukemia development is based on the expansion of a single cell after the initiating mutation. This initial mutation in the origin cell blocks its maturation and increases proliferation, driving clonal expansion and cancer progression. As the leukemia continues to grow, some cells will obtain additional mutations that are favorable for survival, generating sub-clonal populations. As a consequence, all cells of the leukemia will contain the initiating mutation, but the entire population is heterogeneous. Due to this, drugs targeting the products of initiation mutations, such as the t(8;21) translocation, are attractive drug targets (203).

#### **1.6.4 AML with the t(8;21) translocation**

The first of the four WHO subtypes is 'AML with recurrent genetic abnormalities'. The majority of AML cases fall into this category and are associated with non-random chromosomal translocations (204). One such translocation is the t(8;21) translocation; the most common and first chromosomal abnormality to be identified in AML (205). The abnormality is identified in approximately 10% of all AML patients; the exact figure depends on the geographical location and genetic background of the considered population (206). The figure is also dependent on the age of the patients, as the t(8;21) translocation is more common in children/younger patients (207).

The t(8;21) translocation event leads to the fusion of the RUNX1 gene on chromosome 21 with almost the entire ETO gene on chromosome 8. As a result of the translocation, RUNX1 loses its transactivation domain but keeps its runt homology domain (RHD), thus retaining the ability to bind DNA. The transactivation domain is primarily responsible for the recruitment of transcriptional activators, such as p300/CBP. Its replacement, ETO, possesses neryv homology regions which recruit transcriptional repressors such as the N-CoR/mSin3/HDAC1 complex (208). The result is the transcriptional repression of genes which would normally be activated by RUNX1 (figure 1-13). As many RUNX1 target genes are critically involved in granulocytic differentiation, it is not surprising that this abnormality results in a block in differentiation and inhibits the maturation of progenitor cells. Cell survival is also increased, thus cells are predisposed to develop leukemia (209).



### Figure 1-13: The t(8;21) chromosomal translocation

A) The t(8;21) translocation event results in the replacement of the transactivation domain (TAD) with almost the entire ETO gene. The runt homology domain (RHD) is retained. B) RUNX1 can act as a transcriptional activator, via the recruitment of the p300 and transcription factors (TF). The fusion with ETO leads to its transformation into a transcriptional repressor, via the recruitment of the N-CoR/mSin3/HDAC1 complex. Both RUNX1 and RUNX1/ETO heterodimerise with CBFβ (redrawn from (210)).

#### 1.6.5 Secondary mutations are required for the development of t(8;21) AML

The t(8;21) translocation is not sufficient to cause leukaemia; secondary mutagenic events are required for the development of overt disease. Evidence of this is that, in some cases, the abnormality is detected in utero but disease does not occur until later life (211). Furthermore, adult mice engineered to express RUNX1/ETO via an inducible system (to circumvent embryonic lethality) do not develop the disease (212). However, upon administration of mutagenic agents, mice harboring the fusion protein develop AML whereas wild type mice do not (213). This is supported by findings by Nina Cabezas Wallscheid *et al.* which demonstrated, using a mouse model, the very slow disease

progression following RUNX1/ETO expression (214). These experiments suggest that the translocation is necessary, but insufficient, for the development of AML; there is a clear requirement for one or more additional mutations.

Further evidence for the requirement of additional mutation is the disproportionate number of t(8;21) AML patients harbouring additional cytogenetic abnormalities, such as the loss of a sex chromosome (215). Table 3 presents the frequency of various cytogenetic abnormalities in t(8;21) AML, as well as the gene mutations associated with t(8;21) leukaemia (216). The secondary mutations can occur in pathways involved in proliferation, differentiation, avoidance of apoptosis, escape from anti-growth signals and enhanced self-renewal (the hall marks of cancer) (217). The most common are mutations in growth factor receptors such as stem cell growth factor receptor (c-KIT) and FMS-related tyrosine kinase 3 (FLT3). Mutations in the transcription factor gene *SPI1* and the oncogene *N-RAS* are also associated with t(8;21) AML (216).

<b>Mutations</b>	<b>Frequency, no./total (%)</b>
<b>Cytogenetic abnormalities</b>	
-X in female patients	115/331 (35)
-Y in male patients	235/419 (56)
Del (9q)	80/454 (18)
Trisomy 4	8/75 (11)
Trisomy 8	28/454 (6)
Others	77/454 (16)

<b>Molecular genetic mutations</b>	
FLT3-length mutations: FLT3/ITD	27/386 (6.9)
FLT3 D853 activating mutations	5/184 (2.7)
c-KIT (D>Y, D>V, D>H, D>I)	43/351 (12.3)
c-KIT N822K	10/54 (19)
NRAS: codons 12, 13, 61	40/469 (8.5)
PU.1	1/19 (5.3)
RUNX1	1/26 (3.8)



**Table 3: The cytogenetic abnormalities and gene mutations associated with t(8;21) leukaemia** (table modified from (189)).

### **1.7 Molecular pathogenesis of t(8;21) leukaemia**

It is well established that RUNX1/ETO expression leads to a differentiation block and primes hematopoietic cells for oncogenic transformation. The exact mechanism by which it does so is currently under investigation by various research groups. The classical model suggests RUNX1/ETO is oncogenic due to dominant inhibition of RUNX1 function, as a result of the fusion of RUNX1 with a transcriptional repressor (218). More recent research has led to the questioning of this model, with the suggestion of a more complex disturbance in gene regulation, gene expression and chromatin structure (217) (219). This will be outlined in the rest of this section.

#### **1.7.1 The function of RUNX1 transcription factor**

Runt Related Transcription Factor 1 (RUNX1), belongs to the Runt family of transcription factors. These transcription factors share a runt homology domain, a region homologous to the *Drosophila* gene *Runt*, which is essential for normal *Drosophila* embryogenesis (220). It is via this Runt homology region that RUNX1 binds DNA. The Runt domain is also required for the interaction of RUNX1 with CBF $\beta$ , forming a heterodimeric complex termed the core binding factor (CBF) complex. This interaction enhances the DNA binding ability of RUNX1 (221). The (CBF) complex is essential for definitive haematopoiesis and its deregulation can lead to haematological disorders, such as t(8;21) leukaemia (222).

RUNX1 is versatile in its interactions and can cooperate with various transcriptional co-regulators in addition to CBF $\beta$ . As a result, it has the ability to both activate and repress genes. For example, RUNX1 can interact with the histone acetyltransferase, p300, to activate genes. Conversely it can interact with co-repressors, such as the histone deacetylase mSin3A, an interaction which is regulated by RUNX1 phosphorylation (223, 224).

RUNX1 plays a critical role in the specification of the definitive HSC. The presence of RUNX1 marks the earliest HSC; it is detected in the AGM and foetal liver during embryogenesis (225). The fundamental role of RUNX1 in definitive haematopoiesis was clearly demonstrated using knock out mouse studies. Mice lacking RUNX1 had no significant problems with primitive haematopoiesis, whereas definitive haematopoiesis was completely absent and mice exhibited embryonic lethality between 12.5 and 13.5 days. Furthermore, in a chimeric mouse experiment the RUNX1 null embryonic stem cells gave no contribution to adult haematopoiesis (226). More recent studies have shown this is because RUNX1 is required for the generation of haematopoietic progenitors and HSCs from the endothelial cells of the vasculature (227).

It is well established that RUNX1 is absolutely required for the generation of HSC. However, the importance of RUNX1 during adult life is less clear. A study in which RUNX1 function was inhibited using a conditional gene-targeting strategy showed that RUNX1 is not absolutely crucial for adult haematopoiesis; mature blood cells were still present in

peripheral blood and bone marrow of mice (227, 228). However, RUNX1 must play a role as its absence in adult mice leads to several hematopoietic abnormalities. Such as the significant expansion of putative HSC and myeloid progenitor cells, demonstrative of a myeloproliferative phenotype (228). Moreover, the RUNX1 gene locus is commonly mutated in AML, which strongly suggest RUNX1 has a role in adult haematopoiesis.

The transcription factor activity of RUNX1 may help explain the phenotypic difference between wild type and RUNX1 null mice. During adult haematopoiesis RUNX1 is known to regulate the expression of genes involved in myeloid growth factor signalling, by directly binding to their promoters at the consensus DNA sequence PyGPyGGT. These include *IL-3*, *GM-CSF*, *CSF1R* and *c-MPL*. (206). Furthermore, RUNX1 regulates *PU.1*, a transcription factor gene which is critical for terminal myeloid differentiation; PU.1 null mice exhibit a complete block in myeloid differentiation (229, 230).

### **1.7.2 RUNX1/ETO exhibits dominant inhibition of RUNX1 function**

The classical model of RUNX1/ETO driven leukaemogenesis suggests that the presence of RUNX1/ETO results in malignant haematopoiesis by interfering with RUNX1 function. RUNX1/ETO is believed to exhibit dominant inhibition of RUNX1 function via the recruitment of co-repressors to RUNX1 target genes. This simple model was based on various observations. The principal observation being that the translocation event does not affect the DNA binding ability; RUNX1/ETO can still bind RUNX1 target genes. This lead to the presumption that it is through these genes that RUNX1/ETO has a leukemic effect (231).

The classical model was supported by a transgenic mouse study. A RUNX1/ETO 'knock-in' mouse demonstrated embryonic lethality and a block in definitive haematopoiesis. This is a phenotype almost identical to that of a RUNX1 knockout mouse (232, 233). The similarity of the two phenotypes was regarded as *in vivo* evidence of dominant RUNX1 inhibition.

The ability of RUNX1/ETO to repress RUNX1 target genes, via the recruitment of transcriptional repressors, has also been demonstrated experimentally. Using amino-terminal deletions in a transcriptional repression assay, a research group characterised the ETO modular structure and its interaction with the corepressors N-CoR and mSin3A (234). Further support is offered via experiments with inhibitors of HDACs; one group found that treatment of RUNX1/ETO expressing cells with trichostatin and phenylbutyrate partially relieved the ETO-mediated repression and differentiation block (235).

### **1.7.3 RUNX1/ETO has effects distinct from RUNX1 inhibition**

There is significant evidence for the classical model of RUNX1/ETO mediated leukaemogenesis. However, more recent reports suggest that this model, although logical, is an oversimplification of RUNX1/ETO action. For example, the observed phenotypic similarity between RUNX1/ETO knock-in and RUNX1 null mice does indeed suggest a mechanism of dominant RUNX1 inhibition. However, further investigation revealed the phenotypes are not identical. RUNX1/ETO knock-in mice lived one day longer than RUNX1 null mice. In addition, the foetal liver contained absolutely no

progenitor cells in the RUNX1 null, whereas a small number were detected in RUNX1/ETO knock-in mice. This suggests that the presence of RUNX1/ETO has effects distinct from the deregulation of RUNX1 target genes (233). This theory was further supported by a study which showed that RUNX1/ETO can affect the expression of genes not normally controlled by RUNX1. The study identified 24 RUNX1/ETO target genes and of these 24 genes, only 10 were also RUNX1 regulated targets (236).

The classical model suggests the leukemic effect of RUNX1/ETO is mediated via the repression of genes normally activated by RUNX1, via the recruitment of various co-repressors. However, unexpectedly, micro-array analysis revealed RUNX1/ETO can also lead to the activation of genes, and the proportion of activated and repressed genes is actually similar. Furthermore, the genes interrogated were all directly bound by RUNX1/ETO, suggesting the phenomenon seen was not exclusively a downstream effect of RUNX1 inhibition (199). This offers convincing evidence that RUNX1/ETO mediated pathogenesis is likely to be far more complex than the simple direct repression of RUNX1 targets.

Based on the classical model, repression of RUNX1 genes is pivotal to the leukaemogenicity of RUNX1/ETO. It was therefore hypothesised that if RUNX1/ETO lost its repressor function, the leukemic effect of the fusion protein would dramatically decrease. One research group investigated this question by truncating RUNX1/ETO at the C-terminus, rendering it unable to recruit co-repressors. They obtained very interesting results; the truncated RUNX1/ETO in fact promoted leukaemia development (237). This offers more evidence that RUNX1/ETO has other pathogenic activities beside the down

regulation of RUNX1 targets. However, it must be noted that subsequent experiments by Link, K *et al.* demonstrated that the levels of truncated RUNX1/ETO required to cause leukaemia are supraphysiological, in both murine and human experimental systems (238). Another research group investigated if RUNX1/ETO mediates its oncogenic effects primarily via gene repression, via the use of histone deacetylase inhibitors (HDACi) in leukemic cells. In this case, the inhibition of RUNX1/ETO repressor function partly alleviated the myeloid differentiation block (236, 239). However, a more recent *in vivo* study showed that the treatment of RUNX1/ETO expressing leukemic mice with HDAC inhibitors gave no reduction in disease progression and no survival benefit (240). This suggests that, although RUNX1/ETO does repress genes via co-repressor recruitment, it must have additional oncogenic activities.

#### **1.7.4 RUNX1/ETO interferes with the activity of various transcription factors**

There is now substantial evidence that RUNX1/ETO does indeed have additional activities beside the modulation of RUNX1 activity. Numerous studies have demonstrated that RUNX1/ETO can alter the behaviour of several other transcription factors that are implicated in haematopoiesis, such as PU.1, GATA and C/EBP $\alpha$  (8).

PU.1 is essential for normal myelopoiesis; without PU.1, mice have a severe block in differentiation and no mature myeloid cells are present. RUNX1/ETO has been shown to repress PU.1 via a direct physical interaction. This interaction leads to the dissociation of PU.1's co-activator JUN, dramatically hindering the transcription factor activity of PU.1. As

PU.1 regulates the expression of several genes involved in haematopoiesis, it is likely that this mechanism contributes to the leukaemogenicity of RUNX1/ETO (241).

Another transcription factor key to normal haematopoiesis is C/EBP $\alpha$ . It has been observed that AML patients with the t(8;21) translocation have very low C/EBP $\alpha$  expression relative to other AML subtypes. Research has provided a mechanism to explain this observation. Via conditional expression of RUNX1/ETO *in vitro*, it was established that the fusion protein suppresses C/EBP $\alpha$  mRNA levels via the inhibition of positive autoregulation at the C/EBP $\alpha$  promoter (242). The role of C/EBP $\alpha$  in AML development is discussed in more details in sections 1.6.2 and 1.7.8.

GATA-1 is a transcription factor crucial for erythroid lineage commitment. During normal erythropoiesis, GATA-1 is acetylated by p300, a mechanism which is essential for appropriate GATA-1 function. RUNX1/ETO binds to GATA-1 and blocks this acetylation step, resulting in an inhibition of normal differentiation. This RUNX1/ETO mediated interference with erythroid lineage commitment is believed to contribute to leukaemia development (243).

Another mechanism featuring p300 has been described, which involves the E protein transcription factors, E2A and HEB. E2A is essential for early B cell differentiation and is a potential tumor suppressor. HEB has been implicated in both myogenesis and hematopoiesis. Both transcription factors interact with p300 via their AD1 domain, an interaction which is required for their activity. RUNX1/ETO stably binds to the AD1 domains of E2A and HEB, thus displacing p300 and inhibiting their transcriptional activation (244).

### **1.7.5 RUNX1/ETO alters the epigenetic landscape**

This chapter has already described several examples of how RUNX1/ETO interferes with normal haematopoiesis via inhibition of key transcription factors. We know that transcription factors interact with epigenetic modifiers (see section 1.2.2). It is therefore perhaps not surprising that RUNX1/ETO presence can interfere with the epigenetic environment of the cell. This has been described by several publications over the last decade and it is believed that these epigenetic alteration are pivotal to the pathogenesis of t(8;21) leukaemia (199, 204, 206, 245-247).

The first investigation of RUNX1/ETO mediated chromatin modification was focused on the c-FMS locus. The gene is normally bound by RUNX1 at the promoter, however in t(8;21) AML the gene is also bound by RUNX1/ETO at an intronic regulatory region. RUNX1/ETO binds as part of a complex (see section 1.7.6) and is associated with the reduction of gene expression. The binding of this complex correlated with specific epigenetic modifications; the de-acetylation of histones and the tri methylation of histone H3K27. A very interesting observation was that, although RUNX1/ETO binds to the intronic regulatory region and represses gene expression, the promoter region is still accessible to, and is bound by, other transcription factors. This demonstrates that, as the RUNX1/ETO complex does not completely block binding of other transcription factors, RUNX1/ETO target genes should not be irreversibly silenced. Therefore, the continuous presence of RUNX1/ETO is perhaps required to maintain the established epigenetic



modifications (206). This prompted investigation into the effect of RUNX1/ETO knockdown in t(8;21) AML (see section 1.6.6).

With the advent of affordable next generation sequencing, genome wide analysis of the t(8;21) epigenetic landscape is now possible. One research group used ChIP-seq and unsupervised clustering analysis to measure histone modifications, DNA methylation and p300 localisation. These data were used to define 6 classes of chromatin accessible regions; each class was distinct in its functional make up. It was shown that RUNX1/ETO bound to the class characterised by p300 enrichment and lower than normal histone acetylation levels. This offers evidence that alterations to the histone modification pattern, via the abnormal recruitment of epigenetic modifiers, is a key part of RUNX1/ETO mediated pathogenesis (245).

Another genome-wide study confirmed that RUNX/ETO binds to acetylated sites involved in haematopoiesis, and provided a mechanism by which it does so. ERG and FLI1 are ETS transcription factors that play a key role in normal haematopoiesis, by the acetylation of histones and subsequent gene activation. The study proposed a model in which RUNX1/ETO is recruited to these acetylated sites, via an interaction with ERG/FLI1, and removes the acetyl groups thus repressing haematopoietic genes (204). This is another example of how RUNX1/ETO disrupts the epigenetic landscape via its aberrant recruitment of histone modifiers.

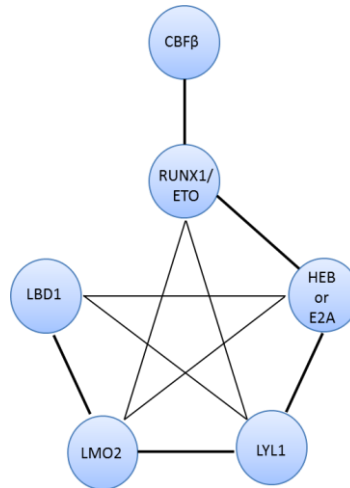
### 1.7.6 RUNX1/ETO functions in an oligomeric complex

Figure 1-13A demonstrates that RUNX1/ETO is made up of different domains. It has a RUNT domain for DNA binding and four neryv homology regions. The neryv homology region 2 (NHR2) is responsible for the oligomerisation of RUNX1/ETO, which is essential for RUNX1/ETO mediated leukaemogenesis (248-251). As individual molecules of RUNX1/ETO complex with each other, RUNX1/ETO has a preference for double RUNX1 binding motifs (252).

Over several years separate studies showed that RUNX1/ETO has many interaction partners, such as the corepressor N-CoR/mSin3/HDAC1 complex (208), C/EBP $\alpha$  (198) and GATA1 transcription factors (243, 253) and E-proteins (204, 254, 255) . Following these findings, Sun, XJ *et al.* published detailed analysis of RUNX1/ETOs interaction partners. They found that RUNX1/ETO forms a high molecular weight, oligomeric complex which is stable *in vivo*. To determine the components of the RUNX1/ETO complex, the authors isolated the complex and performed SDS-polyacrylamide gel electrophoresis and mass spectrometry analysis (256).

The main components were found to be the E proteins HEB and LYL1, the RUNX1 binding partner CBF $\beta$ , LIM-domain binding protein LMO2 and its interacting partner LBD1. They characterised the pairwise interactions within the complex and discovered a network of strong interactions that join all these constituents, one by one. It is likely that this interaction network plays a role in the assembly of the complex (figure 1-14). Weaker

interactions (indicated by thinner lines in the figure) are likely to add stability to the complex.



**Figure 1-14: Schematic of interactions within the RUNX1/ETO complex**

Thick lines represent strong interactions and thin lines represent weak interactions. All components of the complex may homodimerise (figure adapted from (256)).

### 1.7.7 The effect of RUNX1/ETO knockdown in t(8;21) AML

The above findings suggest that the RUNX1/ETO mediated epigenetic reprogramming may not be permanent. To investigate this, Dunne, J *et al.* designed a small interfering RNA (siRNA) specific to RUNX1/ETO for depletion of RUNX1/ETO expression. They used this knockdown system in both t(8;21) AML cell lines and patient blasts, followed by gene expression profiling (257). They found that RUNX1/ETO depletion led to alterations in the expression of 76 genes. The gene expression changes were indicative of initiation of myeloid differentiation and inhibited leukemic cell proliferation (258). Furthermore, cell staining showed an increase in myeloid maturation following RUNX1/ETO depletion.

These results indicated that the maintenance of t(8;21) AML is dependent on the continued expression of RUNX1/ETO, and highlights RUNX1/ETO as a potential therapeutic target.

It was now clear that the presence of RUNX1/ETO disrupts the gene expression profile of t(8;21) AML cells (258). Ptasinska, A *et al.* investigated how RUNX1/ETO does so, in a genome wide fashion. The binding pattern of RUNX1/ETO and RUNX1 was assessed via ChIP-seq analysis, before and after RUNX1/ETO knockdown. As their shared DNA binding domain would suggest, the analysis showed binding of both transcription factors to a subset of RUNX1 targets. Post RUNX1/ETO depletion there was an increase of RUNX1 binding. Interestingly, this increase in RUNX1 enrichment was not only at sites previously blocked by RUNX1/ETO. RUNX1 was found to bind at thousands of new, 'de novo' sites that were not previously bound by RUNX1/ETO (see figure 1-16 B) (199). These results gave more insight into how RUNX1/ETO interferes with wild type RUNX1 to reprogram AML cells.

The RUNX1/ETO ChIP-seq data was used to identify RUNX1/ETO target genes. The effect of RUNX1/ETO knockdown on the expression of these target genes was then investigated. The depletion of RUNX1/ETO resulted in both the upregulation and downregulation of target gene expression. In agreement with Dunne, J *et al.*, functional analyses revealed that a disproportionate number of upregulated genes are involved in differentiation, whereas down regulated genes are involved in cell cycle progression.

To gain further insight into the mechanism of these complex gene expression changes, the researchers assessed the effect of RUNX1/ETO knockdown on the epigenetic profile.

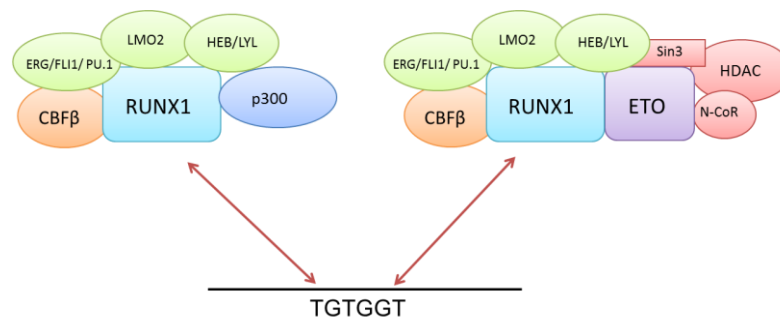
To this end, they conducted a genome wide analysis of the enrichment of RNA polymerase II and Histone H3K9 with CHIP-seq. The results showed that a significant proportion of genes upregulated by RUNX1/ETO knockdown had increased RNase Polymerase II enrichment. There was also a genome wide increase in Histone H3K9 acetylation, which is in accordance with the recruitment of HDACs by RUNX1/ETO (199).

The above findings suggested that the RUNX1/ETO mediated epigenetic reprogramming and the leukemic phenotype is dependent on the persistent expression of RUNX1/ETO. The complexity of the response to RUNX1/ETO knockdown prompted further investigation into the mechanisms involved. It was clear that the myeloid differentiation response is more complex than the alleviation of repression, as wild type RUNX1 is recruited to new sites distinct from those previously bound by RUNX1/ETO. Furthermore, the gene expression alterations were strikingly complex. The researchers wish to find out how RUNX1 was redistributed, and how such extensive changes in the epigenetic profile were mediated. It was hypothesised that other transcription factors and chromatin modifying regulators were likely to be involved.

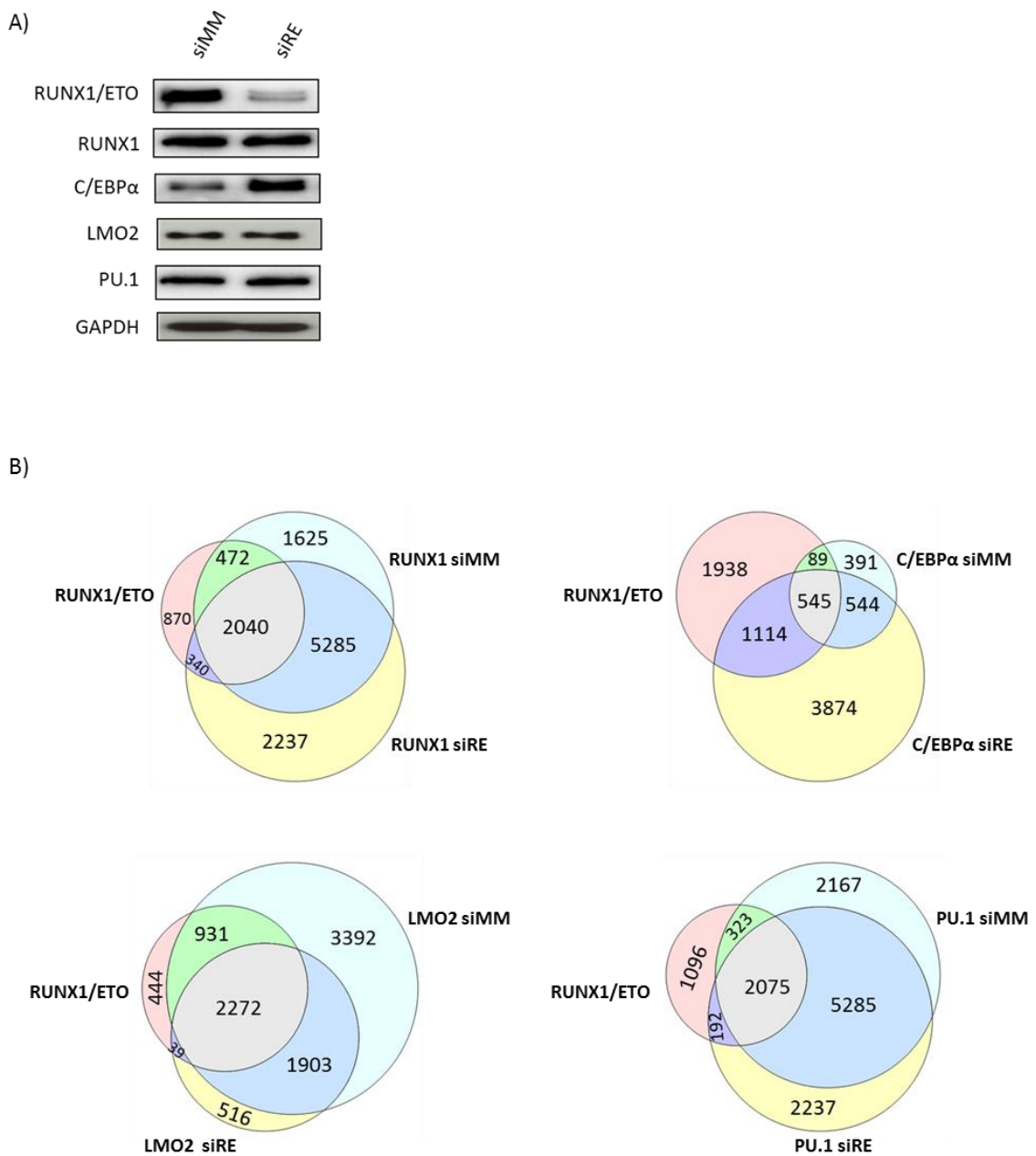
In the next study the relationship between RUNX1 and RUNX1/ETO binding was investigated in more detail. Previous CHIP-seq analysis detected both RUNX1 and RUNX1/ETO binding at the same sites in the genome (199). However, as these results came from analyses on a population of cells, it was not possible to determine whether they bind together at the DNA as a complex. To circumvent this issue, re-CHIP experiments were conducted. The results of these experiments showed that the binding of RUNX1 and RUNX1/ETO is mutually exclusive (158). These findings suggest that RUNX1

and RUNX1/ETO may bind the DNA in a dynamic equilibrium, taking it in turns to bind genomic sites (figure 1-15).

ChIP-seq then used to determine which other transcription factors and regulatory proteins RUNX1 and RUNX1/ETO associate with. It was shown, for the first time *in vivo*, that both proteins differ in their preference for transcriptional repressors and activators. However, they form complexes with similar accessory transcription factors. The factors included LMO2, ERG, FLI1, PU.1 and HEB/LYL. These ChIP sequencing analyses supported proteomic experiments which show co-occupancy of RUNX1 and RUNX1/ETO with the transcription factors HEB and LYL1 and the bridging factors LMO2 and LDB1 at their target sites (see section 1.7.6 for more details on the RUNX1/ETO complex) (256, 259). Below is a diagram showing competing RUNX1 and RUNX1/ETO complexes.



**Figure 1-15: RUNX1 and RUNX1/ETO compete for the same DNA binding sites**



**Figure 1-16: RUNX1/ETO knockdown leads to a genome wide increase in the binding of CEBPα and RUNX1**

A) Western blots detecting RUNX1/ETO, CEBPα, LMO2 and PU.1 protein expression in Kasumi-1 cells treated with control (siMM) and with RUNX1/ETO siRNA (siRE) for 48 hours. B) Venn diagrams showing the overlap between RUNX1/ETO, RUNX1, CEBPα, LMO2 and PU.1 ChIP-seq peaks in Kasumi-1 cells treated for 48 hrs with siMM and with siRE. Experiments were conducted by Dr Anetta Ptasinska (199, 219)

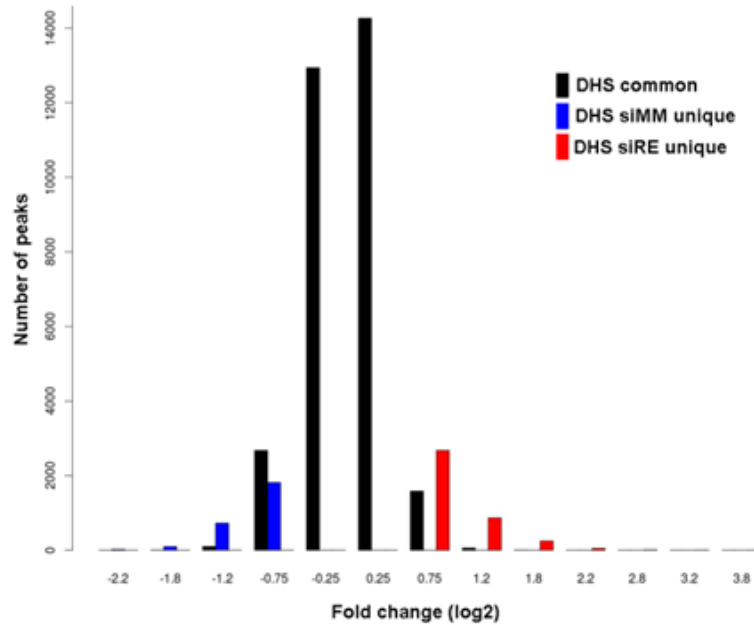
### **1.7.8 The role of C/EBP $\alpha$ in t(8;21) AML**

The role of other transcription factors in reprogramming the transcriptional network after RUNX1/ETO knockdown was investigated. C/EBP $\alpha$  was an obvious candidate for investigation as, out of the several transcription factors analysed, it was the only transcription factor which showed dramatic and rapid changes in protein expression following RUNX1/ETO knockdown (figure 1-16 A). Furthermore, as outlined in earlier sections, it is a key driver of myeloid differentiation during normal haematopoiesis and RUNX1/ETO is known to interfere with its activity (see section 1.7.4).

ChIP sequencing experiments revealed that C/EBP $\alpha$  binding increased considerably after RUNX1/ETO knockdown; C/EBP $\alpha$  bound to almost four times more sites in the genome. This was not a phenomenon universal to all transcription factors. PU.1 and LMO2 binding was also analysed and they exhibited no change and a loss of binding respectively (figure 1-16 B)

DNaseI-seq was then conducted and the results show that the majority of the C/EBP $\alpha$  and RUNX1 binding alterations occurred within pre-existing open chromatin sites. However, there was a subset of DNaseI sites that showed increased hypersensitivity after RUNX1/ETO depletion (figure 1-17). Intriguingly, these sites had an over representation of the C/EBP binding motif (see figure 1-18).





**Figure 1-17: Bar graph demonstrating the fold change of DNaseI peaks after RUNX1/ETO knockdown**

The majority of DNaseI peaks exhibit a low fold change after RUNX1/ETO knockdown. Subsets of DNaseI peaks are reduced (blue bars) and increased (red bars) after knock down. DNaseI was conducted by Dr Anetta Ptasinska (219).

Promoter footprints				Distal footprints			
Motif	best match	score	% target	Motif	best match	score	% target
	SP1	-197.5	8.7		ETS	-1016.8	10.3
	ETS	-118.1	2.5		ETS	-862.1	17.0
	NFY	-89.6	1.5		C/EBP	-708.7	8.9
	KLF4	-48.9	1.1		CTCF	-604.4	2.6
					RUNX	-452.8	5.2
					AP1	-276.2	1.2
					TAL1_TCF <sub>3</sub>	-102.3	12.9
					ETS:Ebox	-94.9	0.32

**Figure 1-18: The CEBP motif is enriched at DNaseI hypersensitive sites that are unique to RUNX1/ETO knockdown**

Genome wide DNaseI site mapping was conducted on cells transfected with siMM and cells transfected with siRE for 10 days. Motif enrichment was conducted, using HOMER software, on the DNaseI footprints that show an increase in hypersensitivity after RUNX1/ETO knockdown. The table indicates the enrichment score of various transcription factor motifs. DNaseI was conducted by Dr Anetta Ptasinska (219). The motif analysis is not published.

## **1.8 The role of SP1 in t(8;21) AML**

Sp1 is a transcription factor involved in haematopoietic differentiation (178, 260, 261). The first study into its role in t(8;21) AML found that Sp1 physically interacts with RUNX1/ETO via the Runt domain. This interaction inhibits the transcriptional activity of Sp1. It was therefore proposed that Sp1 deregulation could be involved in the RUNX1/ETO mediated oncogenic differentiation block (262).

Another group investigated the relationship between Sp1 and RUNX1/ETO and, using ChIP-on-ChIP, mapped the gene promoters which are subject to RUNX1/ETO driven epigenetic repression. They found that many of these genes were involved in haematopoiesis. Furthermore, a significant proportion of these genes were co-occupied with RUNX1/ETO and Sp1. Using an Sp1 inhibitor, they demonstrated that co-occupancy of RUNX1/ETO with Sp1 may promote the repression of these genes, thus suggesting an important role for Sp1 in RUNX1/ETO mediated pathogenesis and identifying SP1 as a new potential therapeutic target (263).

Work from the same research group further investigated the role of Sp1 in t(8;21) AML (264). Their results suggest that Sp1 is important, not only for differentiation block, but also for leukemic cell maintenance. The authors found that the elevated MAPK signalling in t(8;21) AML stabilises Sp1 and thus increases its intracellular levels. They suggest that this increase in Sp1 protein levels enhances leukemic cells growth, as knockdown of Sp1 in these cells led to reduced cell growth and increased apoptosis. These findings suggest an oncogenic role for SP1 in t(8;21) and provide more evidence that it would be a good target for treatment of t(8;21) AML.

Recent data from our lab appears to offer support for a role of Sp1 in the differentiation block which we see after RUNX1/ETO knockdown. DNaseI foot printing analysis suggests that there were alterations in Sp1 binding after RUNX1/ETO knockdown. When the DNaseI cutting frequency was aligned around the Sp1 motif across all binding sites, there appears to be an increase in protection of the Sp1 motif after RUNX1/ETO knockdown, presumably by Sp1. This suggests that RUNX1/ETO presence effects Sp1 binding, which is in accordance with the previously published work (figure 3-4).

### **1.9 t(8;21) AML is dependent on wild type RUNX1 for survival**

RUNX1 is commonly mutated in myeloid malignancies, so is generally considered to be a tumour suppressor. However, several recent studies have shown that the presence of wild type RUNX1 promotes t(8;21) AML cell survival (265, 266). This result is supported by the fact RUNX1 mutation is almost always heterogeneous in AML with CBF abnormalities, whereas RUNX1 can be fully inactivated in other types of leukemia (267). Ben Ami *et al.* showed that knockdown of RUNX1 from a t(8;21) AML cells line leads to apoptosis, which could be rescued with knockdown of RUNX1/ETO (265). Goyama *et al.* propose that a certain level of wild type RUNX1 is crucial for leukaemogenesis. RUNX1 was knocked down from an engineered t(8;21) AML cells line (cord blood cells expressing RUNX1/ETO). Cell growth was inhibited and the same was true when mutant RUNX1 was expressed (266). This is suggestive of an oncogenic role for RUNX1. However, interestingly, over expression of wild type RUNX1 also inhibited AML cell growth by driving myeloid differentiation. They therefore propose a 'dosage-dependent' function of RUNX1 in t(8;21) AML. These unexpected findings highlighted the potential for using wild type RUNX1 as a therapeutic target in t(8;21) AML.

#### **1.9.1 Targeting transcription factor interaction**

RUNX1 heterodimerises with CBF $\beta$ , via its Runt domain - an interaction which stabilizes DNA binding and protects RUNX1 from proteasomal degradation (268, 269). The dimerisation is essential for adequate RUNX1 function. This is clearly demonstrated by the shared phenotype between CBF $\beta$  mutant mice and RUNX1 null mice (270). The Runt domain is retained after t(8;21) translocation and RUNX1/ETO also interacts with CBF $\beta$ . In

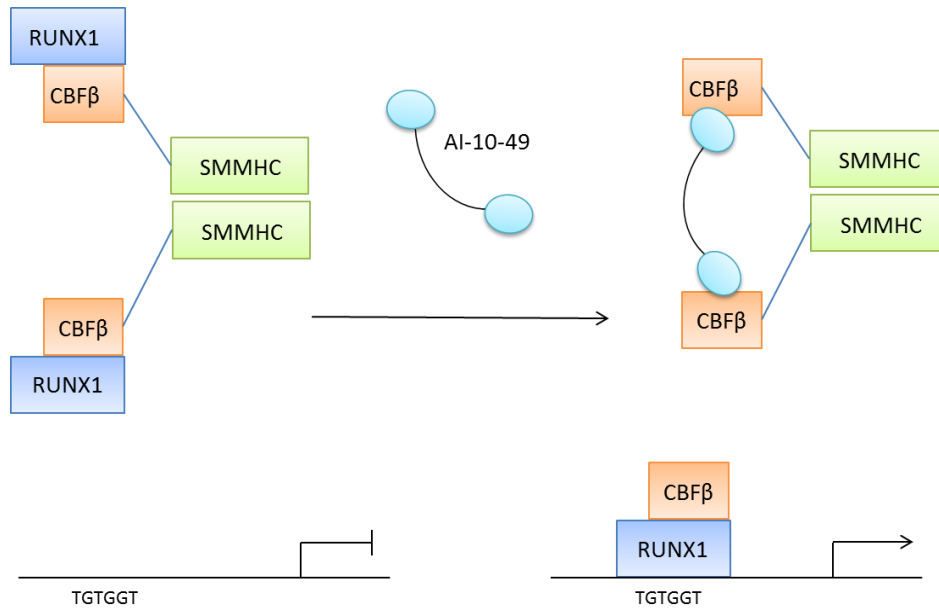
contrast to the wild type RUNX1, the importance of CBF $\beta$  to RUNX1/ETO function is less certain. Roudaia *et al.* state that CBF $\beta$  is critical for RUNX1/ETO activity (271), whereas Kwok *et al.* found that the interaction of RUNX1/ETO to CBF $\beta$  is not needed for RUNX1/ETO mediated transformation of primary cells (250). Furthermore, although a recently published genome-wide study showed that CBF $\beta$  enrichment was detected at thousands of RUNX1/ETO binding sites, the majority RUNX1/ETO sites (59%) were not co-occupied by CBF $\beta$  (204). It is therefore reasonable to predict that blocking the interaction between the Runt Domain and CBF $\beta$  could abolish RUNX1 activity, and may hinder RUNX1/ETO activity. Considering the literature outlined above, inhibiting the CBF $\beta$ -Runt interaction in t(8;21) AML cells could trigger apoptosis and/or differentiation, thus making this an attractive therapeutic approach.

The targeting of protein-protein interaction is a relatively new field, and transcription factors have long been considered 'undruggable' (272, 273). However, attempts have been made to disrupt the Runt domain- CBF $\beta$  interaction therapeutically. RNA aptamers were designed against recombinant Runt- $\beta$  complex and successfully inhibited RUNX1 DNA binding *in vitro* (269). However, unfortunately no subsequent *in vivo* work has been published. More recent, pioneering work from the Bushweller lab has had more promising results.

### 1.9.2 Targeting RUNX1 with small molecule inhibitors

John Bushweller's group designed small molecules that inhibit the Runt- CBF $\beta$  interaction by an allosteric mechanism. The compounds bind to CBF $\beta$  at a region distinct from the Runt domain interaction interface, and induce a conformational change that is transmitted through the protein structure. This alters the dynamics of residues within the interface, dramatically decreasing interaction strength. These residues were previously shown to be vital for interaction via mutagenesis studies (274). *In vitro* experiments proved an inhibition of RUNX1 DNA binding upon treatment with the compound (275). Since the initial *in vitro* studies, the chemical structure of the compound has been subtly modified several times to increase its efficacy and decrease toxicity.

Impressive results were seen with another allosteric inhibitor compound designed to block the interaction between the Runt domain and the CBF $\beta$  fusion protein found in *inv(16)* AML cells. This inversion leads to the fusion of the smooth-muscle myosin heavy chain (SMMHC) to CBF $\beta$ . This CBF $\beta$ -SMMHC is oligomeric and out competes wild type CBF $\beta$  for RUNX1 binding. When initially tested, the compound had only moderate potency and specificity for blocking CBF $\beta$ -SMMHC interaction. They therefore generated a bivalent derivative with the aim to target the oligomeric nature of the fusion complex (figure 1-19). Treating *inv(16)* cells with the bivalent molecule (AI-10-49) lead to an inhibition of the CBF $\beta$ -SMMHC interaction with restoration of wild type CBF $\beta$ -RUNX1 heterodimers. This was accompanied by cell death of AML patient samples and delayed leukaemia progression in mice (276). This offers proof of principle that targeting interactions between RUNX1 and CBF $\beta$  is a feasible therapeutic option.



**Figure 1-19: Targeting Runt domain to CBFβ-SMMHC interaction with allosteric small molecule inhibitors**

The CBFβ – SMMHC fusion protein binds to the Runt domain of RUNX1, blocking wild type CBFβ-Runt domain interaction. The polyvalent inhibitor AI-10-49 interacts with the CBFβ – SMMHC fusion protein, altering its conformation allosterically, rendering it unable to bind RUNX1. RUNX1 can then associate with wild type CBFβ and can drive the expression of its target genes (figure adapted from (277)).

Since the above study, the focus has been targeting the Runt domain-CBFβ interaction in other AMLs. The researchers found that the monomeric version of AI-10-49 had a modest inhibitory effect on the interaction between the RUNT domain and wild type CBFβ; the interaction between RUNT domain and wildtype CBFβ is dimeric, so a monovalent compound is more appropriate. From this compound a series of analogues were prepared. They then established which features of the compound are necessary for inhibition of the Runt domain-CBFβ interaction. They did this via sequential substitution of the functional groups, followed by assessment of inhibition via FRET. The next challenge was metabolic stability *in vivo*. This was achieved via the introduction of a

trifluoromethoxy group, to yield the compound named AI-10-104 (278), which will be tested in this study. Administration of AI-10-104 to mice unfortunately had significant sedative effects. Therefore attempts were made to decrease the toxicity of AI-10-104.

Analogues of AI-10-104 were engineered by appending a morpholine ring to the pyridine ring. *In vitro* experiments to validate inhibition of Runt domain-CBF $\beta$  interaction showed that the analogue 'AI-14-91' was as effective as the original (AI-10-104) but was well tolerated by mice. AI-14-91 and AI-10-104 were found to inhibit the growth of both leukemic and breast cancer cell lines. In embryonic stem cells, AI-14-91 was shown to reduce the binding of RUNX1 to RUNX1 target genes and decrease their expression (see section 4.1.2).

### **1.10 Aims and Objectives**

The Bonifer group have lead the way in characterising the effect of RUNX1/ETO knockdown from t(8;21) AML. We have established that the persistent expression of RUNX1/ETO is required to maintain the leukemic phenotype; RUNX1/ETO knockdown triggers myeloid differentiation and suppressed self-renewal. We have also determined, in a genome wide fashion, that there are epigenetic and gene expression changes which are accompanied by the extensive reorganisation of transcription factor binding, primarily wild type RUNX1 and C/EBP $\alpha$ . We also see the appearance of thousands of new DNaseI hypersensitive sites, which are enriched for the C/EBP motif. However, several questions still need to be answered so we can gain more insight into the mechanisms RUNX1/ETO uses to reprogram haematopoietic cells.

This thesis will systematically present the findings to the questions below:



**1) Can we target wild type RUNX1 and/or RUNX1/ETO with small molecules inhibitors of the Runt domain-CBF $\beta$  interaction?**

Other research groups have recently found that t(8;21) AML depends on wild type RUNX1 for survival (265, 266). We wish to determine whether the inhibition of wild type RUNX1 and RUNX1/ETO with small molecule inhibitors of the Runt domain-CBF $\beta$  interaction is possible in t(8;21) AML cell (278). We will test the ability of the AI-10-104 compound to inhibit survival of t(8;21) AML cells and the effect of the compound on RUNX1 and RUNX1/ETO DNA binding. The experiments have the potential induce cell death via inhibition of RUNX1 and/or trigger differentiation via inhibition of RUNX1/ETO, which has exciting therapeutic implications.

**2) Which transcription factors drive the differentiation response after RUNX1/ETO knockdown?**

Our existing genome wide data points towards an important role for the myeloid transcription factor C/EBP $\alpha$  in the reprogramming of cells after RUNX1/ETO knockdown. We wished to determine whether this is the case by performing RUNX1/ETO knockdown in the absence of C/EBP $\alpha$  and determining how the response to knockdown is affected. We will also investigated if C/EBP $\alpha$  upregulation is the key event in reprogramming t(8;21) AML cells, by over expressing C/EBP $\alpha$  and analysing the effect on the transcriptional programme.

Data primarily from the Alvarez lab suggest that the association between Sp1 and RUNX1/ETO may contribute to RUNX1/ETO mediated leukemic transformation. Furthermore, we have found that RUNX1/ETO knockdown may have an effect on the binding of proteins to the Sp1 motif. To further investigate the role of Sp1 in t(8;21) AML we will assess the effect of RUNX1/ETO knockdown on genome wide Sp1 binding.

**3) Are the epigenetic changes triggered by RUNX1/ETO knockdown reflected by alterations to specific promoter-*cis* regulatory element interactions?**

It is well established that alterations in gene expression can be reflected by changes in DNA looping interactions (63, 118). There is an increasing body of evidence that transcription factors initiate and maintain these DNA loops (63, 279-281). We therefore hypothesise that the changes to gene expression and transcription factor binding after RUNX1/ETO knockdown may be accompanied by alterations in specific promoter-enhancer interactions. We will test this by performing chromatin conformation capture experiments on t(8;21) AML cells with and without RUNX1/ETO expression.

## Chapter 2. METHODS

### 2.1 Cell line culture

Kasumi-1 cells were maintained at  $5 \times 10^5$  cells/ml in Roswell Park Memorial Institute (RPMI) 1640 (Sigma) supplemented with 15% heat-inactivated foetal calf serum, 1% Pen/Strep and 1% glutamine (GIBCO). HL60 cells were maintained at  $5 \times 10^5$  cells/ml in Iscove's modified Dulbecco's medium (Sigma) with 10% heat-inactivated foetal calf serum, 1% pen/strep and 1% glutamine (GIBCO). Human Embryonic Kidney (HEK) 293T cells were maintained in Dulbecco's Modified Eagle Medium, supplemented with 1% pen/strep and 1% glutamine (GIBCO). All cells lines were incubated at 37° C, with humidified air and 5% CO<sub>2</sub> (Carbon Dioxide) supplementation.

### 2.2 Small molecule inhibitor treatment

The small molecule inhibitors and control compounds (see table 4 for details) were dissolved in DMSO and added to cell media at the concentration stated in the results (gift from Bushweller JH, University of Virginia). All 'DMSO control' cells were treated with 1% DMSO.

Compound	Target
AI-10-104	CBF $\beta$
AI-14-91	CBF $\beta$
AI-4-88	Control

Table 4: Small molecule inhibitors of the core binding complex

### 2.3 siRNA and shRNA mediated RUNX1/ETO and C/EBP $\alpha$ depletion

Cells were electroporated with 200nm siRNA, in 700  $\mu$ l of culture medium at a density of  $1 \times 10^7$ /ml. The Fischer EPI 3500 electroporator (Fischer, Heidelberg, Germany) was used for all transfections with the parameters set as 350V for 10ms. Post transfection, cells were maintained as described in section 2.1. RUNX1/ETO knockdown was conducted with the following siRNAs: RUNX1/ETO siRNA (sense, 5'-CCUCGAAAUCGUACUGAGAAG-3' antisense, 5'-UCUCAGUACGAUUUCGAGGUU-3'), mismatch control siRNA (sense, 5'-CCUCGAAUUCGUUCUGAGAAG-3' antisense, 5'-UCUCAGAACGAAUUCGAGGUU-3'). C/EBP $\alpha$  knockdown was conducted with the following siRNAs: GAAGUCGGUGGACAAGAAG and mismatch control UAGGAGCUGGUGAACAAGAC

RUNX1/ETO was depleted from SKNO-1 R/E cells via an doxycycline inducible shRNA expression system. The SKNO-1 R/E cells (courtesy of Olaf Heidenreich) are SKNO-1 cells stably transduced with a pTRIPZ-derived vector (Open Biosystems) encoding shRNA for RUNX1/ETO. Cells were incubated with 1 $\mu$ g/ml doxycycline to induce expression of the shRNA, via a Tet-On<sup>®</sup> system.

## 2.4 Western blotting

Cells were lysed in RIPA buffer (Cell Signalling). Western Blotting was conducted following a standard protocol (13). In short, proteins were separated using gel electrophoresis with a 10% SDS polyacryl amide gel at 40 mA for ~1.5 hrs. The proteins were transferred on to a nitrocellulose membrane (Thermo Scientific Pierce) at 70 V for 1 hour 15 minutes at 4 °C using transfer buffer (192 mM glycine, 25 mM Tris). The membrane was blocked with 5 % non-fat dry milk (Marvel) in TBST (Tris-buffered saline, 0.1% Tween 20) solution for 1 hour and then incubated with a primary antibody overnight at 4°C. A horse radish peroxidase-conjugated antibody (Abcam) was added and detected via enhanced chemiluminescence after incubation with Super-Signal® west Pico Chemiluminescent substrate (Thermo Scientific) and visualised by exposure to an X-ray film. Films were developed with a Compact X4 imaging system (Xograph Healthcare Limited). The primary antibodies used were as follows:

**Table 5: Western blot antibodies**

Antibody target	Manufacturer	Serial number
ETO	Santa Cruz	sc-9737
RUNX1	Abcam	23980
C/EBP $\alpha$	Abcam	40761
SP1	Santa Cruz	Sc-17824 X
GAPDH	Abcam	8245
H3	Abcam	1791

## 2.5 Flow cytometry

For each analysis,  $2 \times 10^5$  cells were counted and centrifuged for 5 minutes, at 300 x g at room temperature. The pellet was washed twice with MACS buffer (PBS, 0.5% BSA and 2 mM EDTA). After the second wash, the pellet was resuspended in the remainder of the supernatant (approximately 200  $\mu$ l). 2  $\mu$ l of antibody was added and the cell suspension was incubated at 4 °C, in the dark, for 30 minutes. The cell suspension was then washed with MACs buffer and analysed with a CyAn™ ADP flow cytometer (Beckman Coulter).

**Table 6: Flow Cytometry Antibodies**

Antibody target	Manufacturer	Serial number
Annexin-FITC/PI kit	BD Pharmingen	556547
CD11b – APC	Miltenyi-biotec	130091241
CD117-APC	Miltenyi-biotec	130091733
IgG FITC	Miltenyi-biotec	130093192
IgG APC	Miltenyi-biotec	130093194

## **2.6 Apoptosis detection assay**

Apoptosis was assessed using an Annexin V-FITC Apoptosis Detection Kit II (BD Pharmingen). At the time points stated in the results,  $10^5$  cells were collected, washed with phosphate buffered saline, and then stained with Annexin V-FITC and propidium iodide (PI) according to the manufacturer's protocol. Analysis of Annexin V-FITC binding was determined by flow cytometry with a CyAn™ ADP Analyser (Beckman Coulter).

## **2.7 Extraction of RNA, cDNA synthesis and analysis of gene expression analysis**

Approximately  $5 \times 10^6$  cells were centrifuged (300xg, 5 minutes, room temperature) and lysed with 350µl of RLT buffer with β- mercaptoethanol (Qiagen). Total RNA was extracted using RNeasy columns, according to the manufacturers protocol (Qiagen). An on-column DNase I digestion was conducted using an RNase-Free DNase kit, according to the manufacturers protocol (Qiagen). The RNA was quantified using a NanoDrop™ 2000 (Thermo Scientific) and cDNA was synthesized as follows: 2 µg of total RNA extracted was annealed to 1 µl of oligo dT (deoxythymidine) primers (80µM) in a total volume of 15µl and incubated at 70°C for 5 minutes. Samples were then placed on ice for 5 minutes. Reverse transcription was conducted via the addition of 5 µl of 5 x Reaction Buffer (Invitrogen), 5 µl of 10mM dNTPs, 1 µl of 200 U/µl MLV- Reverse Transcriptase (Invitrogen), 0.625µl of 400 U/µl of RNase inhibitor (Invitrogen) and the reaction was incubated at 42 °C for 60 minutes.

cDNA was quantified by Real-Time qPCR, with the primers listed below and SYBR® Green PCR master mix (Applied Bio systems). Analysis was conducted using an ABI 7500 Real-Time PCR system (Applied Biosystems), with the step one software. The expression of individual genes was calculated relative to GAPDH expression. A cDNA standard curve was made using cDNA synthesized from mRNA extracted from the cell line undergoing analysis. The standard cDNA was diluted to 1:10, 1:50, 1:250, 1:1250 and 1:6250 with water. cDNA to be tested was diluted 1:50 times. Genomic DNA derived from chromatin immunoprecipitation was analysed by qPCR as stated above, but standards were made using genomic DNA.



**Table 7: Primers used for qPCR**

Gene target	Forward Primer Sequence	Reverse Primer Sequence
<i>CEBPA</i>	GAGGGACCGGAGTTATGACA	AGACGCGCACATTCACATT
<i>CEBPE</i>	ATGTCCCACGGGACCTACTACGA	ACAGTGTGCCACTTGGTACTGCAG
<i>CSF1R</i>	AGCACGAGAACATCGTCAACC	TTCGAGAAAGTTGAGCAGGT
<i>CST7</i>	CCAACCACACCTTGAAGCAGA	GGGTCAGTGACAACGGAGAAC
<i>CTSG</i>	TCAGTTGCTGCTGTGCTTC	TTCTCAATCCCCTGTCCCCAC
<i>ERG</i>	ATGGAGGAGAAGCACATGCC	ATAGCGTAGGATCTGCTGGC
<i>GAPDH</i>	CCTGGCCAAGGTCATCCAT	GGGGCCATCCACAGTCTT
<i>IGFBP7</i>	GAAGTAACTGGCTGGGTGCTG	GCTGATGCTGAAGCCTGTCC
<i>MPL</i>	TCAGCAGCCAAGATGTCTCC	TGCCTCTTCTCATCCCAGA
<i>MS4A3</i>	CCAAGCCATAAACAACCCCA	TTCTGGTCCCGTCTCACTGC
<i>NFE2</i>	CCAAGGTGTGTTCAAAGAGGC	GGAGCCGAGTCAGGGAAGAC
<i>NKG7</i>	CTGATTGCTTTGAGCACCGA	CCTGATATGATGTCCCCATGC
<i>PU.1</i>	TCTTGCCACCAGGTCTCCTA	CGCCCTCCTCCTCATCTGA
<i>RNASE2</i>	CCCCTGAACCCCAAGAA	ACCATGTTTCCCAGTCTCCG
<i>RUNX1/ETO</i>	TCAAAATCACAGTGGATGGGC	CAGCCTAGATTGCGTCTTCACA
<i>SOX4</i>	AAGATCATGGAGCAGTCGCC	CGCCTCTCGAATGAAAGGGA
<i>SP1</i>	GCACCTGCCCTACTGTAAA	TGGATGTGGCAAATATGCTGT
<i>UBASH3B</i>	ACCATCAAGCATGGATCGGC	GGTCACCGACATGGGAGAAT

**Table 8: Primers used for ChIP-qPCR**

Gene target	Forward Primer Sequence	Reverse Primer Sequence
IVL promoter	GCCGTGCTTTGGAGTTCTTA	CCTCTGCTGCTGCCACTT
PU.1 14Kb enhancer	AACAGGAAGCGCCCAGTCA	TGTGCGGTGCCTGTGGTAAT
IGFBP7 promoter	GTCAAGCACTAAAAGGACAAACCG	TGAATGCCACTGGGAGACAAAG

## 2.8 Chromatin Immunoprecipitation (ChIP)

$2 \times 10^7$  cells were counted, centrifuged (300 xg, 5 minutes, 4 °C) and resuspended in media. Whole cells were treated with 1% formaldehyde (Sigma) at room temperature for 10 minutes to crosslink protein to DNA. 0.5M glycine was added and the reaction was quenched for 5 minutes at room temperature. Cells were pelleted via centrifugation (300 xg, 5 minutes, 4 °C) and washed twice with ice cold PBS and resuspended in buffer A (10 mM HEPES pH 8.0, 10 mM EDTA, 0.5 mM EGTA, 0.25% Triton X-100, protease inhibitor cocktail (Roche UK) and 0.1 mM PMSF)) and rotated at 4 °C for 10 minutes. The solution was centrifuged (500 xg, 5 minutes, 4 °C) and pelleted nuclei were resuspended in buffer B (10 mM HEPES pH 8.0, 200 mM NaCl, 1 mM EDTA, 0.5 mM EGTA, 0.01% Triton X-100, protease inhibitor cocktail and 0.1 mM PMSF ) by rotation for 10 minutes at 4 °C. Chromatin was pelleted (500g, 5minutes, 4 °C), snap frozen with liquid nitrogen and stored at -80 °C. For immunoprecipitation, the chromatin pellet was resuspended in 600 µl ChIP buffer 1 (25 mM Tris-HCl pH 8.0, 150 mM NaCl, 2 mM EDTA, 1% Triton X-100, 0.25% SDS, protease inhibitor cocktail and 0.1 mM PMSF) and sonicated twice, for 10 minutes of 30s ON 30s OFF each time with the Bioruptor™ sonicator (Diagenode). Chromatin was pelleted (16,000xg, 5 minutes, 4 °C) and the supernatant was diluted 1:3 with ChIP buffer 3 (25 mM Tris-HCl pH 8.0, 150 mM NaCl, 2 mM EDTA, 1% Triton X-100, 7.5% glycerol, protease inhibitor cocktail and 0.1 mM PMSF).

Dynabeads® Protein G beads (Invitrogen) were used for immunoprecipitation. 15µl beads were added to 0.5% BSA, phosphate buffer and 2 µg of antibody; RUNX1 (Abcam), ETO (Santa Cruz), CBFβ (Abcam) and PU.1 (Santa Cruz). The immunoprecipitation mixture

was incubated at 4 °C for 1 hour, with rotation and then added to chromatin and incubated for 4 hours at 4 °C with rotation (5% of the volume of chromatin added was taken as an input control). The following wash steps were performed (a magnet was used to collect the beads); once with wash buffer 1 (150 mM NaCl, 1% Triton X-100, 20 mM Tris-HCl, 0.1% SDS and 2mM EDTA), twice with wash buffer 2 (500 mM NaCl,1% Triton X-100, 20 mM Tris-HCl, 0.1% SDS and 2mM EDTA), once with LiCl buffer (250 mM LiCl , 0.5% NP-40, 10 mM Tris-HCl pH 8.0, 0.5% Na-deoxycholate and 1 mM EDTA) and finally twice with 1 ml TE/NaCl buffer (50 mM NaCl,10 mM Tris-HCl pH 8.0 and 1 mM EDTA). Using a shaker for 15 minutes at room temperature, DNA was eluted twice with 50 µl of elution buffer (1% SDS, 100mM NaHCO<sub>3</sub>). The input control was made up to 100 µl with the same elution buffer. The crosslink was reversed overnight at 65 °C with 50 µg proteinase K. DNA purification was conducted using Agencourt AMPure beads (Beckman Coulter) following the manufacturers standard protocol. DNA was eluted with 50 µl of water (6).

Genomic DNA was assessed via qPCR (see section 2.7). Relative enrichment was calculated by normalizing data to the input chromatin. Data is also presented with normalisation against the input chromatin and the enrichment observed at negative control regions (Ch18 and Involucrin (IVL) promoter).

**Table 9: CHIP Antibodies**

<b>Antibody target</b>	<b>Manufacturer</b>	<b>Serial number</b>
RUNX1	Abcam	23980
ETO	Santa Cruz	sc-9737
PU.1	Santa Cruz	sc-352 X
SP1	Santa Cruz	sc-17824 X

## **2.9 Retroviral production**

A retrovirus was used to stably transduce Kasumi-1 cells with the C/EBPA-ER vector (see figure 3-8 A), a generous gift from Dr Chris Van Oevelen (University Pompeu Fabre, Barcelona). This vector was constructed using the MSCV-IRES-GFP retroviral vector backbone. Viral particles were made in HEK293T cells. The virus was collected and concentrated and then used to transduce Kasumi-1 cells. The genes encoding the packing and envelope proteins are on a separate plasmid. Therefore, virus cannot be produced in the Kasumi-1 cells following transduction.

### **2.9.1 Transfection of HEK293T cells for retroviral production**

For virus production HEK293T were cultured to a 80-90% confluency. Following the manufacturers protocol, Trans-IT (Mirus, USA) was used to transfect three 10 cm plates of HEK293T cells with three construct: 30µg C/EBPA-ER vector, 30µg gag/pol, 9 µg Envelope (gift from Dr James Mulloy, Cincinnati, USA). Viral supernatant was harvested at 12 hours, 36 hours and 48 hours after transfection.

### **2.9.2 Virus concentration**

To pellet cell debris, the viral supernatant was centrifuged for 15 minutes at 4 °C at 3000 rpm. The supernatant was collected and filtered (0.45 µM filter disc). Following the manufacturer's instructions, a Centricon Plus 70 100kDa filter column (Millipore, USA) was used for concentration. Centrifugation was conducted at 1000 x g for 25 minutes at 4 °C.

### **2.9.3 Retroviral transduction with RetroNectin**

RetroNectin (Takara, Japan) assists viral transduction efficiency by bringing together viral particles and the target cell. A 6 well, non-tissue culture plate was coated with RetroNectin. To do this a 24 µg/ml solution of RetroNectin in PBS was added to the plate for 2 hours at room temperature. The solution was subsequently removed and the plate was blocked with BSA (2% BSA in PBS) at room temperature for 20 minutes. The plate was then washed with 2.5% HEPES in HBSS (Hank's Balanced Salt Solution, SIGMA).

The concentrated virus was then bound to the RetroNectin coated plates, by two 45 minutes centrifugations at 2000 xg. Finally,  $1 \times 10^6$  Kasumi-1 cells were added to each well of the plate, in a 1:1 ratio with the remaining concentrated virus. Polybrene was then added at a final concentration of 8 µg/ml.

The plate was incubated overnight at 37 °C with 5% CO<sub>2</sub> (standard tissue culture conditions). The following day, cells were pelleted at 300 xg for 5 minutes at room temperature. The cell pellet was resuspended in standard tissue culture medium (see 1.1). The efficiency of transduction is indicated by the proportion of GFP positive cells. GFP expressing cells were isolated by Florescence Activated Cell Sorting (FACS) with the MoFlo Astriso EQ (Beckman Coulter Life Sciences).

## **2.10 Circularised chromosome conformation capture (4C-seq)**

### **2.10.1 Fixation and cell lysis**

1x10<sup>7</sup> Kasumi-1 cells, transfected with mismatch siRNA (siMM) or siRNA specific to siRUNX1/ETO (siRNA), were fixed with 2% formaldehyde and incubated for 10 minutes at room temperature. 1.425 ml of 1M glycine was added to quench the cross-linking reaction. Fixed cells were immediately centrifuged for 8 minutes at 4°C, 500 xg. Supernatant was removed and the pellet resuspended in 1ml lysis buffer (500µl 1M TRIS pH 7.5, 300µl 5M NaCl, 100µl 0.5M EDTA, 250µl 20% NP-40 and 100µl Triton X-100 made up to 10ml with H<sub>2</sub>O) and incubated at room temperature for 5 minutes, followed by 5 minutes at 65°C. Cells were then kept on ice whilst complete cell lysis was determined via Trypan blue (Gibco) staining. Cells were centrifuged at 800 xg for 5 minutes and the pellet was taken up in 440 µl H<sub>2</sub>O and 60 µl 10X RE buffer 2 (NEB).

### **2.10.2 Digestion with primary restriction enzyme**

15  $\mu$ l of SDS was added and the tube placed at 37°C for 1 hour. 75  $\mu$ l of 20% Triton X-100 was added and the tube incubated at 37°C for 1 hour. A 5  $\mu$ l aliquot was removed as an 'undigested control' sample before 200 units of the restriction enzyme DpnII was added. The tube was incubated for 4 hours at 37°C, and then another 200 units of DpnII was added, followed by overnight 37°C incubation. The following day 200 units of DpnII was added for 4 hrs at 37°C. A 5  $\mu$ l aliquot was removed as a 'digested control' sample. To this, along with the 'undigested' sample, 90  $\mu$ l of 10mM Tris pH 7.5 and 5 $\mu$ l Proteinase K (10 mg/ml) was added to reverse the cross links. These control samples were run on a 0.6% agarose gel to assess the digestion efficiency. All 37°C incubations were conducted in a heated block, shaking at 900 RPM. DpnII was selected as the restriction enzyme as it functions in SDS, and combined with the second restriction enzyme (Csp6I) it generates restriction fragments near the target loci, with a suitable size for efficient ligation and PCR amplification. Both of these enzymes are 4bp cutters, so will cut the genome into 256 bp fragments, on average. This allows for a high resolution assay.

### **2.10.3 Ligation and DNA purification**

The DpnII was inactivated by incubation at 65°C for 20 minutes. On ice, 700  $\mu$ l of 10X ligation buffer, 7 ml of milli-Q H<sub>2</sub>O and 10  $\mu$ l T4 Ligase (Roche 5U/ $\mu$ l) were added then samples were incubated overnight at 16 °C. The following day, to assess ligation efficiency, a 100  $\mu$ l aliquot of the sample was taken as the 'ligated control'. The crosslinks were reversed as above and the sample run on a 0.6% agarose gel (as previously described).

To reverse the crosslinks, 30 µl Proteinase K (10mg/ml) was added and samples were left overnight at 65°C. The next day, 30 µl RNase A (10mg/ml) was added and samples were incubated for 45 minutes at 37°C. DNA was extracted by adding 7 ml phenol-chloroform. Samples were mixed thoroughly then centrifuged at 3000 xg at room temperature. The water phase was transferred to a new 50 ml tube to which 7 ml of milli-Q H<sub>2</sub>O, 7 µl of glycogen, 1.5 ml 2M NaAC pH 5.7 and 35 ml ethanol was added. Samples were placed at –80°C overnight. The next day samples were centrifuged at 4°C for 30 min, 3000 xg. The supernatant was removed and 10 ml of cold 70% ethanol was added. Samples were centrifuged again for 15 min, 3000 xg at 4°C. The supernatant was removed and the pellet left to dry at room temperature. The pellet was dissolved in 150µl 10mM Tris pH 7.5.

#### **2.10.4 Digestion with secondary restriction enzyme**

Each sample was transferred to a 1.7 ml tube, 50 µl 10X restriction buffer and 50 units of the restriction enzyme Csp6I (Fermentas # ER0211) was added and the volume made up to 500 µl with milli-Q H<sub>2</sub>O. After an overnight incubation, 500 RPM shaking, at 37°C, a 5 µl aliquot of the sample was taken. This ‘digestion control’ was run on a 0.6% agarose gel.

#### **2.10.5 Second ligation and DNA purification**

The enzyme was inactivated as previously describe and the samples transferred to a 50 ml tube. 1.4 ml of 10X ligation buffer and 20 µl of ligase (100 U) (Roche Catalogue # 10799009001) was added, then the reaction made up to 14ml with milli-Q H<sub>2</sub>O. After an overnight ligation at 16 °C, 1.4ml 2M NaAC pH 5.6, 14µl glycogen and 35ml of 100% ethanol were added. Samples were stored at –80°C overnight. The next day samples



were centrifuged at 4 °C for 45 minutes, at 3750 RPM. The supernatant was removed and 15 ml of cold 70 % ethanol was added. The samples were then centrifuged again for 15 minutes, at 20°C and 3750 RPM. Again, the supernatant was removed and the pellet then left to dry at room temperature. Once dry the pellet was dissolved in 150 µl 10mM Tris pH 7.5 at 37°C. Samples were then purified using a QIAquick PCR purification kit, according to the manufacturer's protocol. Samples were eluted in 50 µl 10mM Tris pH 7.5 and pool samples. DNA concentration of each 4C template was determined via analysis with a NanoDrop™ 2000 (Thermo Scientific).

#### **2.10.6 View-point fragment selection and primer design**

Restriction fragments greater than 350 bp and within 2kb of the target genomic region were selected as viewpoint fragments, dependent on the ability to design specific primers. A 5' Illumina adapter sequence was added so the inverse-PCR products did not need further processing prior to sequencing. Reading primers were designed as close to the primary restriction site as possible, to reduce reads from the known viewpoint sequence. Non-reading primers were designed to regions less than 120kb from the secondary restriction site (table 10).

### **2.10.7 PCR amplification and sequencing**

200 ng of 4C template was used per PCR reaction. For each viewpoint and template, 16 PCR reactions were conducted using an Expand Long Template system (ROCHE # 11681834001) (see table 10 for primer sequences). The pooled PCR products (total volume 800  $\mu$ l) were then purified using the High Pure PCR Product Purification Kit (Roche # 11732676001), to remove any adaptor containing primers (<120 bp). Samples were centrifuged to pellet any beads that escaped the column. The supernatant was taken, then the concentration and purity of this 4C template was assessed by a NanoDrop™ 2000 (Thermo Scientific) (260/280 ration > 2 and 260/230 ratio > 1.8 was required). The libraries were then visualized on a 1.5% agarose gel. All 8 of the 4C libraries were pooled, and then multiplexed sequencing was performed on the HiSeq 2500 platform. The libraries were de-multiplexed, mapped and normalized using a published bioinformatic pipeline (282).

### **2.10.8 Differential analysis of 4C interactions**

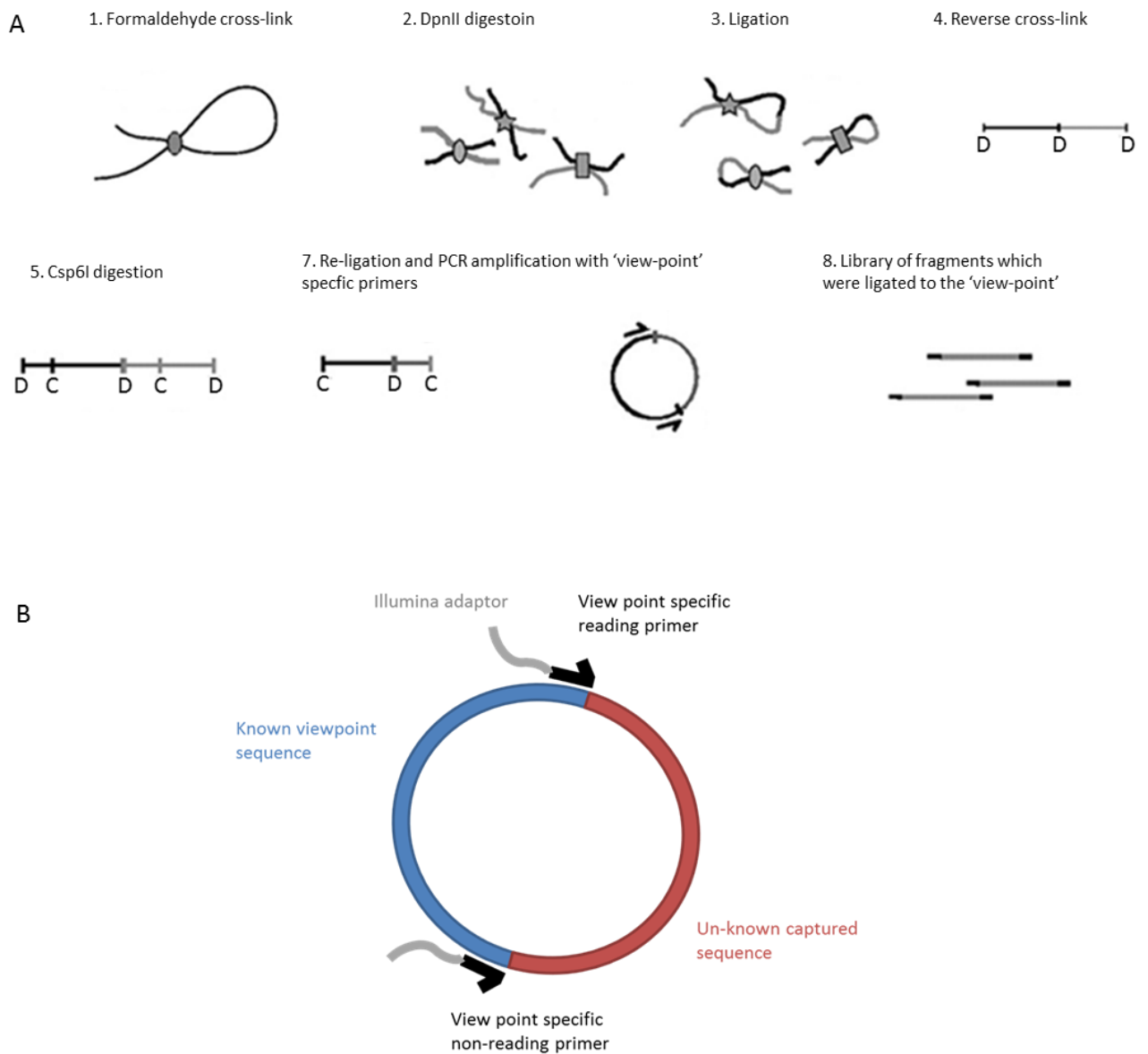
Individual fragment counts were calculated for every 1kb bin. A median was calculated, with a 3kb sliding window, and data from both biological replicates was merged. The R package DESeq2 was used to calculate the log<sub>2</sub> fold change (RUNX1/ETO knockdown vs control) at the local genomic coordinates

**Table 10: Viewpoint specific 4C-seq PCR primers**

Sample	View point	Primer sequence
siMM REP 1	SPI1 PROMOTER FORWARD	AATGATACGGCGACCACCGAACACTCTTCCCTACACGACGCTCTCCGATCTAGTGAGGAGCAGTGGCGATC
siRUNX1/ETO REP 1	SPI1 PROMOTER FORWARD	AATGATACGGCGACCACCGAACACTCTTCCCTACACGACGCTCTCCGATCTAGTGAGGAGCAGTGGCGATC
siMM REP 2	SPI1 PROMOTER FORWARD	AATGATACGGCGACCACCGAACACTCTTCCCTACACGACGCTCTCCGATCTAAGTGAGGAGCAGTGGCGATC
siRUNX1/ETO REP 2	SPI1 PROMOTER FORWARD	AATGATACGGCGACCACCGAACACTCTTCCCTACACGACGCTCTCCGATCTTAGTGAGGAGCAGTGGCGATC
	SPI1 PROMOTER REVERSE	CAAGCAGAAGACGGCATAACGAACTATGCCCTGGCTCAGA
siMM REP 1	SPI1 URE FORWARD	AATGATACGGCGACCACCGAACACTCTTCCCTACACGACGCTCTCCGATCTGTCCAGGGAAGCCAGATC
siRUNX1/ETO REP 1	SPI1 URE FORWARD	AATGATACGGCGACCACCGAACACTCTTCCCTACACGACGCTCTCCGATCTGTGTCCAGGGAAGCCAGATC
siMM REP 2	SPI1 URE FORWARD	AATGATACGGCGACCACCGAACACTCTTCCCTACACGACGCTCTCCGATCTATGTCCAGGGAAGCCAGATC
siRUNX1/ETO REP 2	SPI1 URE FORWARD	AATGATACGGCGACCACCGAACACTCTTCCCTACACGACGCTCTCCGATCTGTGTCCAGGGAAGCCAGATC
	SPI1 ENHANCER REVERSE	CAAGCAGAAGACGGCATAACGATTCAACAGGCACAGACTT
siMM REP 1	CD34 PROMOTER FORWARD	AATGATACGGCGACCACCGAACACTCTTCCCTACACGACGCTCTCCGATCTAGTGATAGCCTCACCAGATC
siRUNX1/ETO REP 1	CD34 PROMOTER FORWARD	AATGATACGGCGACCACCGAACACTCTTCCCTACACGACGCTCTCCGATCTCAGTGATAGCCTCACCAGATC
siMM REP 2	CD34 PROMOTER FORWARD	AATGATACGGCGACCACCGAACACTCTTCCCTACACGACGCTCTCCGATCTGAGTGATAGCCTCACCAGATC
siRUNX1/ETO REP 2	CD34 PROMOTER FORWARD	AATGATACGGCGACCACCGAACACTCTTCCCTACACGACGCTCTCCGATCTTAGTGATAGCCTCACCAGATC
	CD34 PROMOTER REVERSE	CAAGCAGAAGACGGCATAACGAGTGCCTCTCTAGGAGC
siMM REP 1	CD34 URE FORWARD	AATGATACGGCGACCACCGAACACTCTTCCCTACACGACGCTCTCCGATCTCGTGTGGCCCCATACAGATC
siRUNX1/ETO REP 1	CD34 URE FORWARD	AATGATACGGCGACCACCGAACACTCTTCCCTACACGACGCTCTCCGATCTACGTGTGGCCCCATACAGATC
siMM REP 2	CD34 URE FORWARD	AATGATACGGCGACCACCGAACACTCTTCCCTACACGACGCTCTCCGATCTGCGTGTGGCCCCATACAGATC
siRUNX1/ETO REP 2	CD34 URE FORWARD	AATGATACGGCGACCACCGAACACTCTTCCCTACACGACGCTCTCCGATCTCGTGTGGCCCCATACAGATC
	CD34 URE REVERSE	CAAGCAGAAGACGGCATAACGAAGGAATAGAGTGGAGG
siMM REP 1	CD34 DRE FORWARD	AATGATACGGCGACCACCGAACACTCTTCCCTACACGACGCTCTCCGATCTCTTTTCCCCTGCTCGATC
siRUNX1/ETO REP 1	CD34 DRE FORWARD	AATGATACGGCGACCACCGAACACTCTTCCCTACACGACGCTCTCCGATCTACTTTTCCCCTGCTCGATC
siMM REP 2	CD34 DRE FORWARD	AATGATACGGCGACCACCGAACACTCTTCCCTACACGACGCTCTCCGATCTGCTTTTCCCCTGCTCGATC
siRUNX1/ETO REP 2	CD34 DRE FORWARD	AATGATACGGCGACCACCGAACACTCTTCCCTACACGACGCTCTCCGATCTCTTTTCCCCTGCTCGATC
	CD34 DRE REVERSE	CAAGCAGAAGACGGCATAACGAAAACTGCCAAGCAGCAG

**Table 11: Genomic co-ordinates of selected viewpoint restriction fragments**

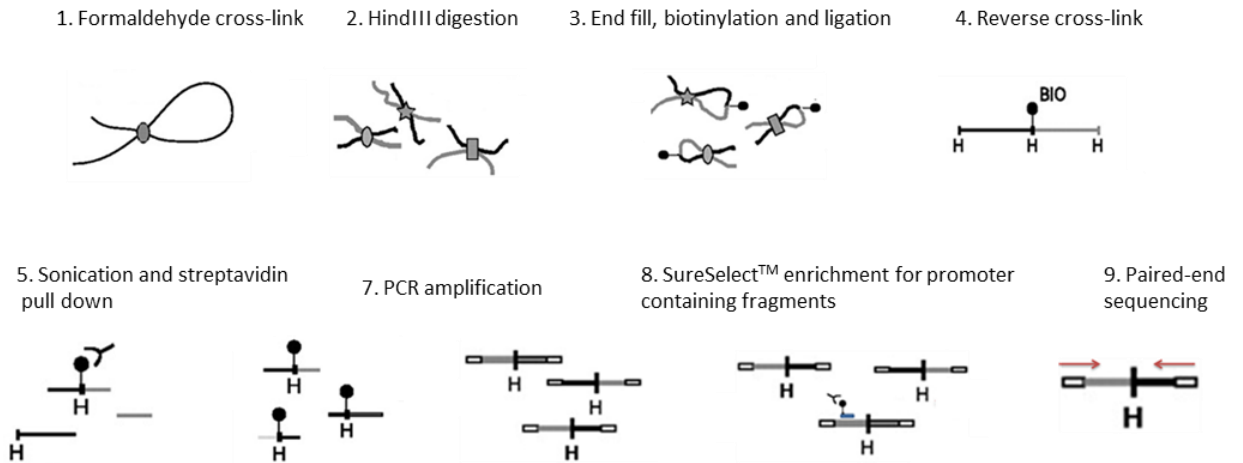
<b>View point</b>	<b>Primer sequence</b>
SPI1 Promoter	chr11:47399040-47398740
SPI1 Upstream Regulatory Element (URE)	chr11: 47414944-47414201
CD34 Promoter	chr1:208,081,333-208082113
CD34 Upstream Regulatory Element (URE)	chr1:208,095,318-208,095,781
CD34 Downstream Regulatory Element (DRE)	chr1:208,034,222-208,035,512



**Figure 2-1: Overview of the 4C method**

A) Schematic of 4C-seq experimental procedure. After formaldehyde cross-linking and digestion with the first restriction enzyme (DpnII) the chromatin is religated under dilute conditions to fuse interacting fragments. The DNA is digested with a second restriction enzyme (Csp6I) then ligated to form small DNA circles, which should now contain only one primary ligation junction. Inverse PCR primers designed to the view-point (restriction fragment of interest) allow specific amplification of fragments 'captured' by the viewpoint. The resultant library of fragments is subject to high throughput sequencing. B) Outline of 4C-seq primer and inverse PCR design. Inverse PCR primers specific to the viewpoint carry an illumina sequencing adaptor which means the resultant library does not require further processing. Sequencing reads consequently begin with the reading primer binding site. Therefore reading primers are designed as close to the primary restriction site as possible, to maximise coverage of the un-known sequence. Non-reading primers are designed to a sequence less than 120kb from the secondary restriction site.

## 2.11 Capture HiC



### Figure 2-2: Schematic of Capture Hi-C experimental procedure

Chromatin is fixed and digested as in 4C. In this case, biotin is incorporated into the ligation junction, allowing enrichment for ligation junctions via streptavidin pull down. Sequencing adapters are added during PCR amplification. The library is then enriched for promoter containing ligated fragments with biotinylated baits, designed to hybridize to 22,000 promoters. Streptavidin pull down is used to select promoter containing fragments. The resultant library is subject to paired-end sequencing.

### 2.11.1 Fixation and lysis of cells

$5 \times 10^7$  Kasumi-1 cells were fixed via suspension in 37 ml of RPMI-1640 supplemented with 15% FBS and 2% formaldehyde. The cells were fixed for 10 minutes at room temperature while mixing. 6 ml of 1M glycine (0.125 M final concentration) was added to quench the reaction. Cells were incubated at room temperature for 5 min, followed by 15 minutes on ice. Cells were then pelleted by centrifugation at  $400 \times g$  for 10 minutes at  $4^\circ\text{C}$ . The

supernatant was discarded and cells were washed in ice cold PBS. The cells were then pelleted and after removing the supernatant, each sample was flash frozen in liquid nitrogen, and stored at  $-80^{\circ}\text{C}$ .

Cells were lysed in a tight dounce homogeniser (ten cycles) with 3ml of cold lysis buffer (10 mM Tris-HCl pH 8, 10 mM NaCl, 0.2% Igepal CA-630, one tablet protease inhibitor cocktail (Roche complete, EDTA-free, 11873580001)). Cells were left on ice for five minutes then homogenised another ten times. The lysed cells, in 3 ml lysis buffer, were added to 47ml of lysis buffer and incubated on ice for 30 minutes with occasional mixing.

The chromatin was pelleted at 1800 rpm for 5 minutes at  $4^{\circ}\text{C}$  and the supernatant discarded. The pellet was resuspended in 1ml of 1.25x NEBuffer 2 and split into four. Each sample was then pelleted at 1000 rpm and resuspended in 358  $\mu\text{l}$  of 1.25x NEBuffer 2. 11  $\mu\text{l}$  10% SDS was added and each tube was incubated at  $37^{\circ}\text{C}$  for 60 minutes, rotating at 950 rpm. Samples were mixed by pipetting up and down every 15 minutes. The SDS was quenched with 75 $\mu\text{l}$  10% Triton X-100 and incubated at  $37^{\circ}\text{C}$  for 60 minutes, rotating at 950 rpm. During the incubation, samples were mixed by pipetting up and down every 15 minutes.

### **2.11.2 HindIII digestion, biotinylation and ligation of digested DNA ends**

The chromatin was digested overnight with 1500 units of HindIII (NEB R0104T), rotating (950 rpm) at  $37^{\circ}\text{C}$ . The HindIII cleavage of its restriction site (5' AAGCTT 3') generates a 5' overhang (5'-AGCT-3'), which was repaired to include a biotinylated nucleotide, allowing the enrichment of Hi-C ligation products. To do this, 6 $\mu\text{l}$  10x NEB2, 2 $\mu\text{l}$   $\text{H}_2\text{O}$ , 1.5 $\mu\text{l}$  10mM

dCTP, 1.5 $\mu$ l 10mM dGTP, 1.5 $\mu$ l 10mM dTTP, 37.5 $\mu$ l 0.4mM biotin-14-dATP (Life Technologies 19524-016), and 10 $\mu$ l 5U/ $\mu$ l Klenow (DNA polymerase I large fragment, NEB M0210L) were added to each tube. Samples were mixed carefully and incubated for 60 minutes at 37°C, with mixing by pipetting every ten minutes.

The digested chromatin mixture was added to a falcon tube with 8 ml of ligation buffer (750 $\mu$ l 10x ligation buffer (NEB B0202S), 75  $\mu$ l 10mg/ml BSA (NEB B9001S) and 50  $\mu$ l 1U/ $\mu$ l T4 DNA ligase (Invitrogen 15224-025). These dilute conditions favour ligation of fragments crosslinked within the same chromatin complex (intra-molecular ligation). The tubes were mixed by inversion and incubated overnight at 16 °C. The following day, samples were incubated for a further 30 minutes at room temperature.

### **2.11.3 Crosslink reversal and DNA purification**

Crosslinks were reversed and proteins degraded by incubating the tubes over night at 65 °C, with 60 $\mu$ l 10mg/ml proteinase K (Roche 03115879001). After the overnight incubation, another 60  $\mu$ l of 10 mg/ml proteinase K per tube was added, followed by a 2 hour incubation at 65°C.

The reaction mixtures were cooled to room temperature and 12.5  $\mu$ l of 10 mg/ml RNase A (Roche 10109142001) was added. Samples were incubated at 37°C for 60 minutes. The DNA in these tubes was purified with two phenol extractions. After ethanol precipitation, the DNA pellets were resuspended in 25 $\mu$ l 1x TE and the four samples were pooled. The Quant-iT assay was used to determine the DNA concentration of the resultant HiC library using a Qubit® 3.0 Fluorometer (Thermo Scientific).

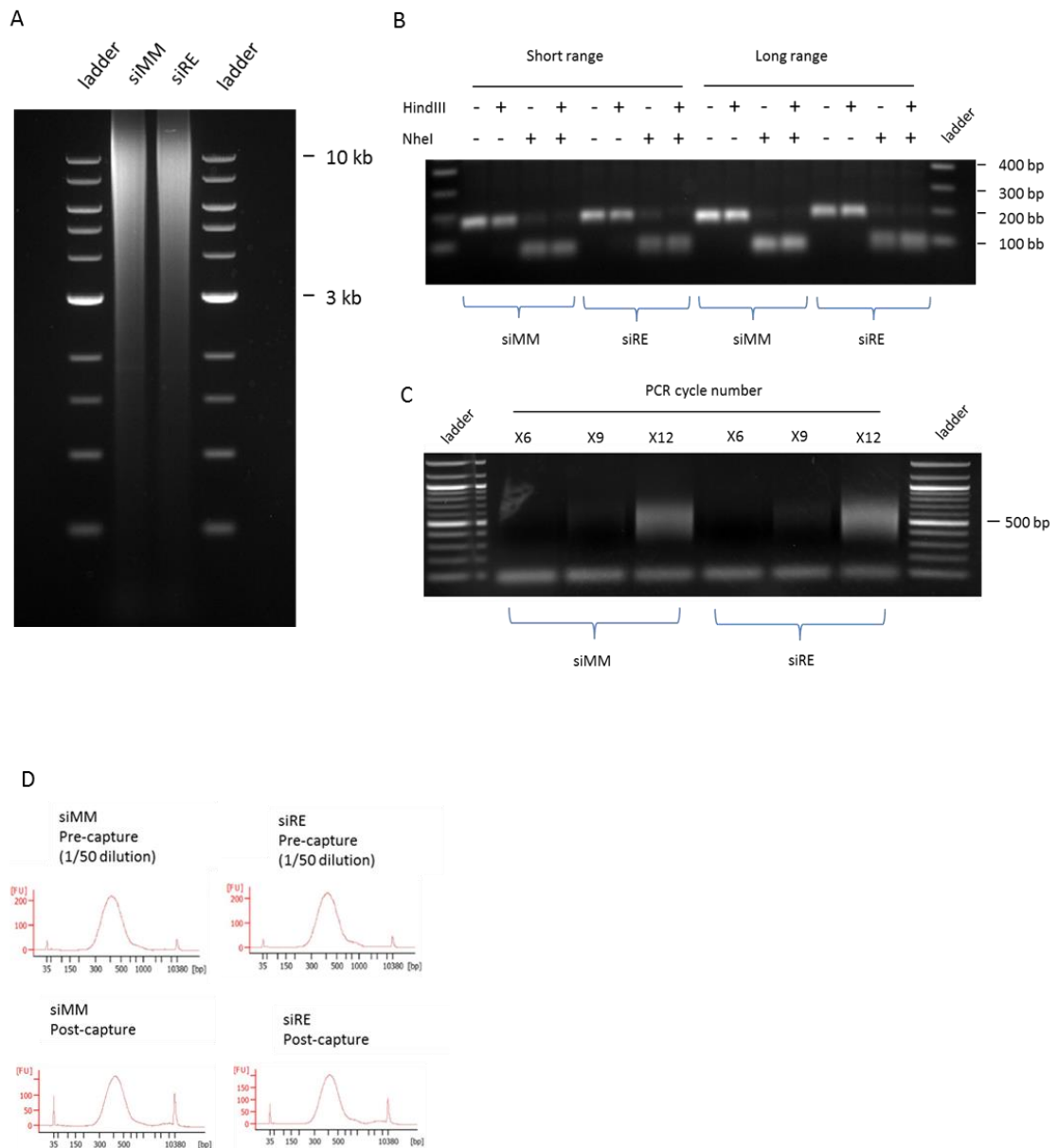


#### 2.11.4 HiC ligation efficiency and quality controls

To check the library was sufficiently ligated, a 1/10 dilution was resolved on a 0.8% agarose gel. The majority of the library should run at around 10 kb. To confirm that blunt-end ligation was successful, the fact that fill-in and ligation of a HindIII site (AAGCTT) creates a NheI restriction site (GCTAGC) was exploited. A particular ligation product formed by the MYC promoter and a known 1.8 MB long range interaction was amplified with 3C PCR (see table 12 for primer sequences). The same was done for a ligation product formed by the MYC promoter and an adjacent restriction fragment. 200ng of the purified PCR products were split into four and digested with HindIII, NheI or both enzymes, with undigested DNA is used as a control. The resultant DNA samples were run on a 1.5% gel (see figure 2-3)

**Table 12: 3C PCR primers**

Target restriction fragment	Primer Sequence
MYC promoter	GGAGAACCGGTAATGGCAAA
Restriction fragment < 1 kb from MYC promoter	TGAGGTCCCAGGCATTCTTT
Restriction fragment 1.8 MB from MYC promoter	AATAACAAGGCCCCCAATTCT



### Figure 2-3: Capture-HiC quality controls

A) Hi-C libraries were resolved on a 0.8% agarose gel. The majority of DNA runs at approximately 10 kb, which is indicative of sufficient ligation. The ligation efficiency is not 100%, hence the smear below. B) The purified HiC DNA is used as a template for PCR amplification. The DNA fragments generated from ligation of the MYC promoter with a neighbouring restriction fragment, and the MYC promoter with a 1.8MB distal fragment, is amplified using standard 3C PCR conditions. The amplicon is then digested with HindIII and/or NheI. Ligation products which have been efficiently biotinylated can be cut by NheI but not HindIII. C) After biotin pull down, one tenth of the immobilised Hi-C library was amplified from the streptavidin beads with 6, 9 or 12 PCR cycles. The resultant DNA was resolved on a 1.5% agarose gel. A smear is just about visible at 6 cycles. Therefore 7 cycles was selected for bulk amplification. D) Bioanalyser traces of HiC libraries pre and post sure select promoter enrichment.

### **2.11.5 Removal of biotin from non-ligated DNA ends**

To avoid pulling down any non-ligated fragments, biotin was removed from free ends using the exonuclease activity of T4 DNA polymerase. 8 reactions of 5µg of Hi-C library, 0.5µl 10 mg/ml BSA, 5µl 10x NEBuffer 2, 2µl 2.5mM dATP, and 5µl T4 DNA polymerase (NEB M0203L) in a total volume of 50µl were made and incubated at 20°C for 4 hours. The reaction was stopped by adding 2ml of 0.5M EDTA pH 8.0. Two reactions were pooled to generate four tubes of approximately 10ug DNA. The DNA was purified via phenol:chloroform extraction and ethanol precipitation. Each sample was reconstituted in 130 µl H<sub>2</sub>O

### **2.11.6 DNA shearing and end repair**

An E220 focused-ultrasonicator (Covaris) was used to fragment the DNA with the intensity set to 4, the duty cycle set to 10% and with 200 cycles per burst for 55 seconds. After sonication, the entire volume of each sample was transferred into a fresh Eppendorf tube and 18µl 10x ligation buffer (NEB B0202S), 18µl 2.5mM dNTP mix, 6.5µl T4 DNA polymerase (NEB M0203L), 6.5µl T4 DNA Polynucleotide kinase (NEB M0201L) and 1.3µl Klenow (NEB M0210L) were added. The reaction was left for 30 minutes at room temperature. Each sample was then split into two and purified with a MinElute column (Qiagen 28004), according to the manufacturer's instructions. Each column was eluted twice with 15 µl TLE (10mM Tris pH8.0, 0.1mM EDTA).

### **2.11.7 A-tailing and size selection**

To the sheared and end repaired DNA , 5 µl of NEBuffer 2 10x, 11.5 µl dATP 1mM and 3.5 µl Klenow exo- (NEB M0212L) were added. The reaction was incubated for 30 minutes at 37 °C. To inactivate the enzyme, each reaction was incubated at 65°C for 20 minutes, and put on ice immediately afterwards. Fragments between 200 and 650 base pairs were size selected by double-sided SPRI bead (ampure) size selection (0.6x followed by 0.9x), following the manufacturers protocol. All Hi-C library samples were then pooled and quantified with the Quant-iT™ High-Sensitivity DNA Assay Kit (Invitrogen) and assayed on the Qubit™ Fluorometer (Invitrogen).

### **2.11.8 Biotin-streptavidin pulldown and adapter ligation**

The biotin marked fragments were immobilized via MyOne Streptavidin C1 DynaBeads (Invitrogen) following manufacturers protocol. The fragments were ligated to illumina paired-end adaptors. The immobilized Hi-C fragments were amplified via PCR, using PE PCR primers 1 and 2 with 7 cycles.

PE PCR 1 - AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCT

PE PCR 2- CAAGCAGAAGACGGCATACGAGATCGGTCTCGGCATTCCTGCTGAACCGCTCTTCCGA

The Hi-C library was then purified twice with solid phase reversible immobilization (SPRI) beads (Beckman Coulter Ampure XP beads A63881) with a 1:1.8 DNA to SPRI bead ratio.

### 2.11.9 Test PCRs to determine conditions for Hi-C library amplification

To determine the optimal number of PCR cycles for Hi-C library, test PCRs were set up with 6, 9 and 12 amplification cycles. Each reaction contained 2.5 µl of Hi-C library DNA on beads, 5 µl of phusion buffer 5x (NEB F531), 0.7 µl dNTP mix, 0.75 ul PE PCR primer 1, 0.75 µl PE PCR primer 2, 0.3 µl Phusion polymerase (NEB F531) and 16.35 µl of H<sub>2</sub>O. The PCR programme was set as follows:

1 cycle	98°C	30 seconds
	65°C	30 seconds
	72°C	30 seconds
n - 2 cycles	98°C	10 seconds
	65°C	30 seconds
	72°C	30 seconds
1 cycle:	98°C	10 seconds
	65°C	30 seconds
	72°C	7 minutes

The amount of amplified DNA was assessed by running the entire reaction (25 µl) on a 1.5 % agarose gel. A smear in the range of 300 bp to 600 bp was be just about visible at 6 cycles of amplification and increased in intensity with increasing number of PCR cycles. 7 cycles were selected for bulk PCR amplification.

#### **2.11.10 Final PCR amplification of Hi-C library**

The remaining volume of HiC library DNA on beads was amplified following the above PCR reaction set up (i.e. 2.5 µl of beads per reaction). The beads were separated from the reaction mixture with a magnet and the supernatant was purified twice with solid phase reversible immobilization (SPRI) beads (Beckman Coulter Ampure XP beads A63881), following the manufacturer's instructions. The library was quantitated with the Quant-iT™ High-Sensitivity DNA Assay Kit (Invitrogen) and assayed on the Qubit™ Fluorometer (Invitrogen) and run on the bioanalyzer (Agilent).

#### **2.11.11 Hybridization of Hi-C library with biotin-RNA**

750 ng of each library was desiccated with a SpeedVac. After evaporation of all liquid, the HiC DNA pellet was suspended in 3.4 µl of H<sub>2</sub>O. 2.5 µl of custom block 1, 2.5 µl of custom block 2 µl and 0.6 µl of custom oligo block (Agilent Technologies) were added. After thorough mixing the solution was transferred into a PCR strip and kept on ice. 49 µl of hybridisation buffer was made per sample. Hybridisation buffer consists of 25 µl SureSelect Hybridization solution 1 (Agilent Technologies), 1 µl SureSelect Hybridization solution 2 (Agilent Technologies), 10 µl SureSelect Hybridization solution 3 (Agilent Technologies) and 13 µl SureSelect Hybridization solution 4 (Agilent Technologies). The buffer was mixed thoroughly, heated to 65°C for 5 minutes, then transferred into a PCR tube strip and kept at room temperature.

The SureSelect RNase Block (Agilent Technologies) was diluted 1:4 with nuclease free water. 2  $\mu$ l of the dilution was added to 5  $\mu$ l biotinylated RNA baits (custom made, Agilent Technologies (77)). After thorough mixing, the whole 7  $\mu$ l was transferred to a PCR tube strip and kept on ice.

A PCR machine (Thermocycler, Biometra) was set to 95 °C for 5 minutes at 65 °C forever, with the lid heated to 98 °C. The PCR strip containing the HiC library and custom blockers was added to the pre-heated PCR machine and the protocol was started. After just over 5 minutes (once the temperature had reached 65°C) the PCR strip with the hybridization buffer was added to the PCR machine. After 5 minutes (10 minutes since the start of the PCR program), the PCR strip with the biotinylated RNA bait was transferred to the PCR machine. After 2 minutes, 13  $\mu$ l of hybridization buffer from the PCR strip was added to the 7  $\mu$ l of RNA bait (grey into cross-hatched). The PCR strip containing the hybridization buffer was discarded. Immediately after, all of the HiC library and custom blockers (9  $\mu$ l) was added into the 20  $\mu$ l of RNA bait with hybridization buffer.

After 24 hours at 65°C, a biotin pulldown and washes were performed as follows. Per capture HiC sample, 60  $\mu$ l of Dynabeads MyOne Streptavidin T1 (Life Technologies 65601) were washed three times with binding buffer (Agilent). With the streptavidin beads in 200  $\mu$ l of binding buffer, the entire hybridization reaction was taken from the PCR machine and transferred into the tube containing the streptavidin beads. This was incubated on a rotating wheel for 30 minutes at room temperature.

After 30 minutes, the beads were washed with 500  $\mu$ l wash buffer I (agilent). The beads were incubated in wash buffer I at room temperature for 15 minutes and vortexed every 2 to 3 minutes for 5 seconds. The beads were then washed with wash buffer II (pre-warmed to 65°C). The beads with wash buffer II were incubated at 65°C for 10 minutes and vortexed for 5 seconds every 2 to 3 minutes. This was repeated for a total of 3 washes. After removing the supernatant, the beads were washed in 200  $\mu$ l 1xNEB2, then resuspend in 30  $\mu$ l of 1xNEB2 and transferred to a fresh tube.

#### **2.11.12 Capture HiC library amplification**

A post-capture PCR amplification step was carried out, using PE PCR 1.0 and PE PCR 2.0 primers, following the previously describe protocol with only 4 PCR cycles. DNA was purified twice with 1.8x volume of SPRI beads (Beckman Coulter Ampure XP beads A63881), following the manufacturer's instructions and resuspend in nuclease free H<sub>2</sub>O. The library was quantitated with the Quant-iT™ High-Sensitivity DNA Assay Kit (Invitrogen) and assayed on the Qubit™ Fluorometer (Invitrogen) and run on the bioanalyzer (Agilent).

#### **2.11.13 Capture HiC data analysis**

The capture Hi-C paired-end sequencing reads were put through a publically available pipeline called HiCUP (283). The raw sequencing reads are separated and mapped against the positions of the human genome (hg19). The reads were then filtered for experimental artefacts and duplicate reads, and then re-paired. By aligning the reads a read count per



restriction fragment was obtained. Statistically significant interactions were called with another bioinformatic package called GOTHIC (284). This uses a cumulative binomial test to detect interactions between distal genomic loci that have significantly more reads than expected by chance, by using a background model of random interactions. This analysis assigns each interaction with a p-value, which represents its significance. Differential interactions (control vs RUNX1/ETO depletion) were determined by comparing the p-values with HOMER bioinformatic software (285). A difference with a p-value of less than 0.1 was deemed to be significant.

## Chapter 3. RESULTS

### 3.1 CBF complex inhibition in t(8;21) AML cells

#### 3.1.1 Treatment of HL60 and Kasumi-1 cells with CBF complex inhibitor induces apoptosis in a dose dependant manner

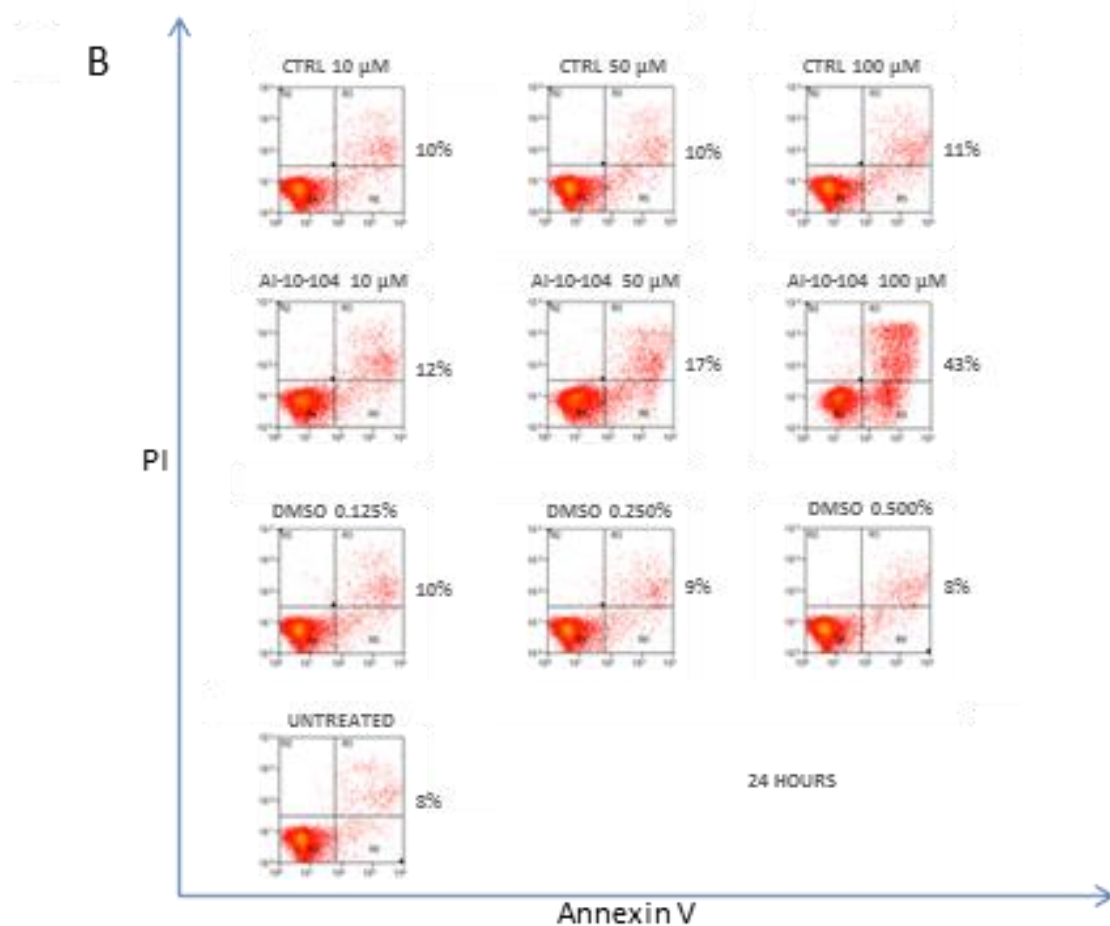
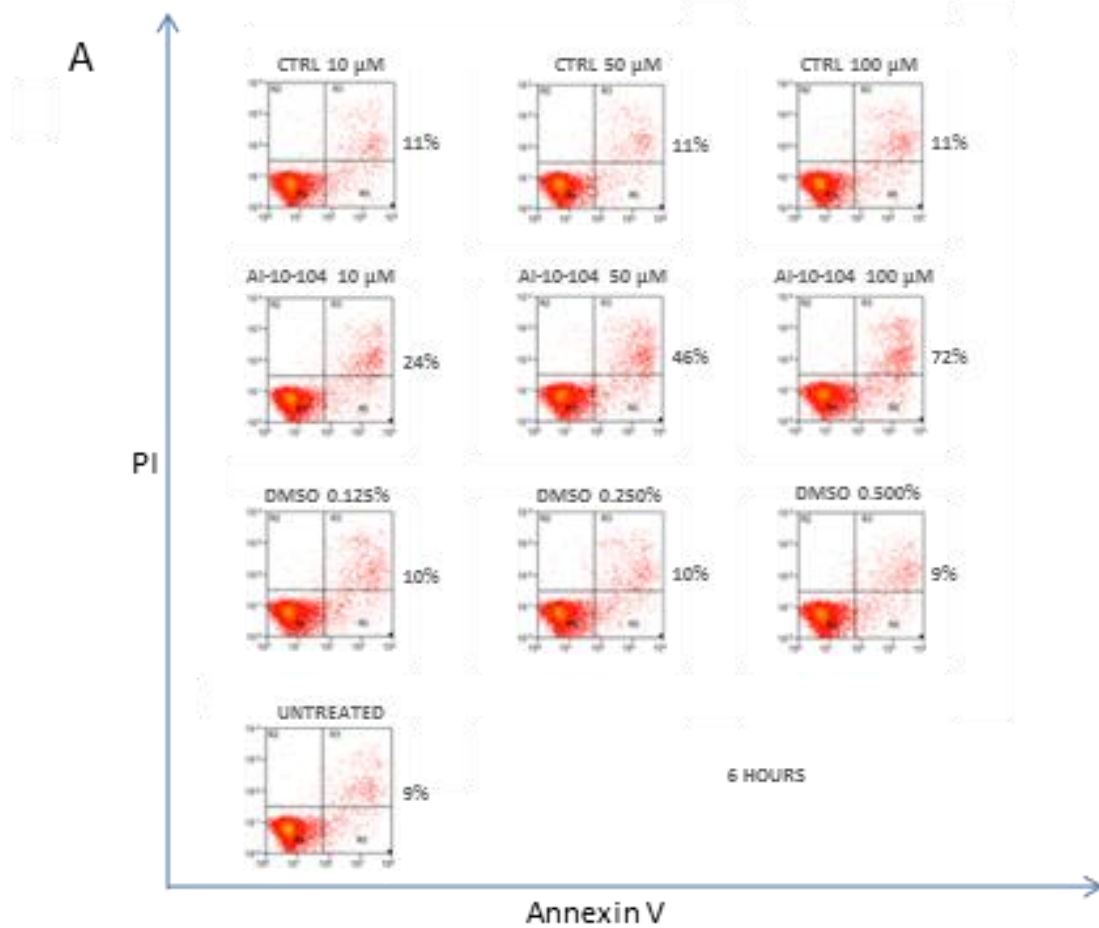
Experiments using RUNX1 specific siRNA have shown that RUNX1 is essential for the viability of t(8;21) AML cells (265, 266). This identified wild type RUNX1 as a novel therapeutic target in t(8;21) AML. However siRNA is not an easily feasible therapeutic option. For this reason, as well as several others, efforts have been made to generate small molecule inhibitors of RUNX1. One such compound is an allosteric inhibitor designed to block the Runt domain-CBF $\beta$  interaction, and thus inhibit RUNX1 DNA binding and transcriptional activity.

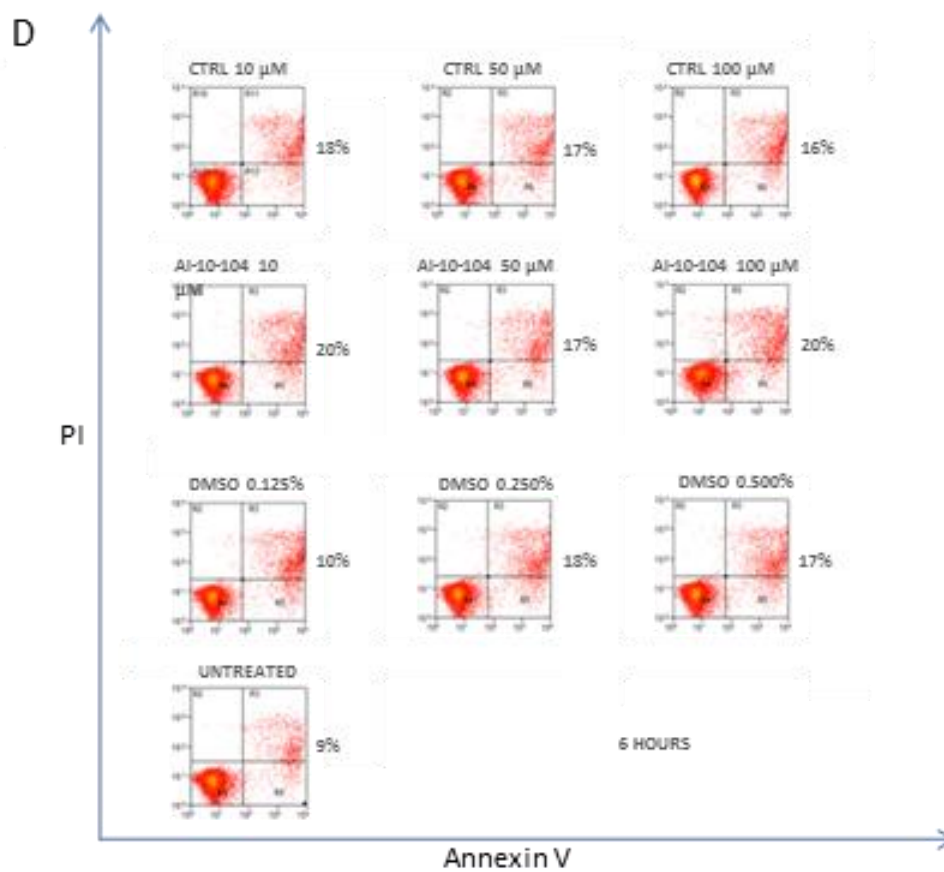
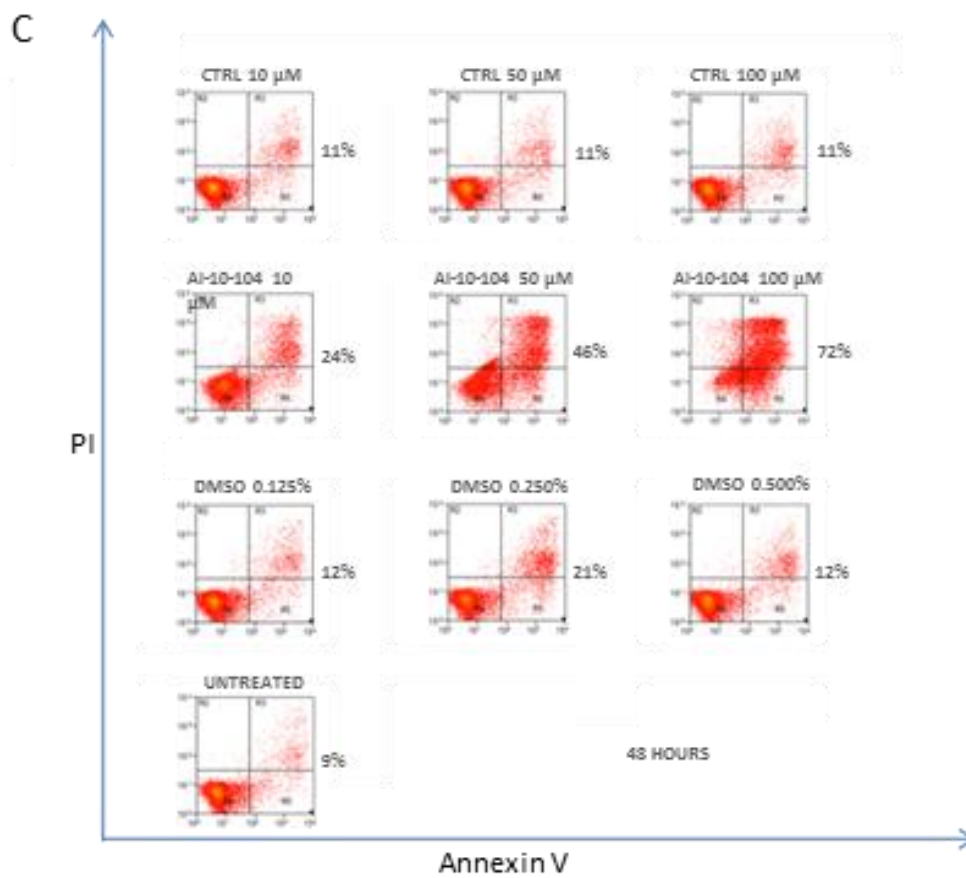
As RUNX1/ETO also interacts with CBF $\beta$  via its runt domain, the compound may also inhibit RUNX1/ETO function. The effect of RUNX1 knockdown on cell viability is not seen with a combined knockdown of RUNX1/ETO and RUNX1 (266). This suggests that the extent to which the compound affects either protein will determine the outcome on cell viability. We therefore set out to test whether these compounds were efficient at inhibiting DNA binding of RUNX1 and RUNX1/ETO.

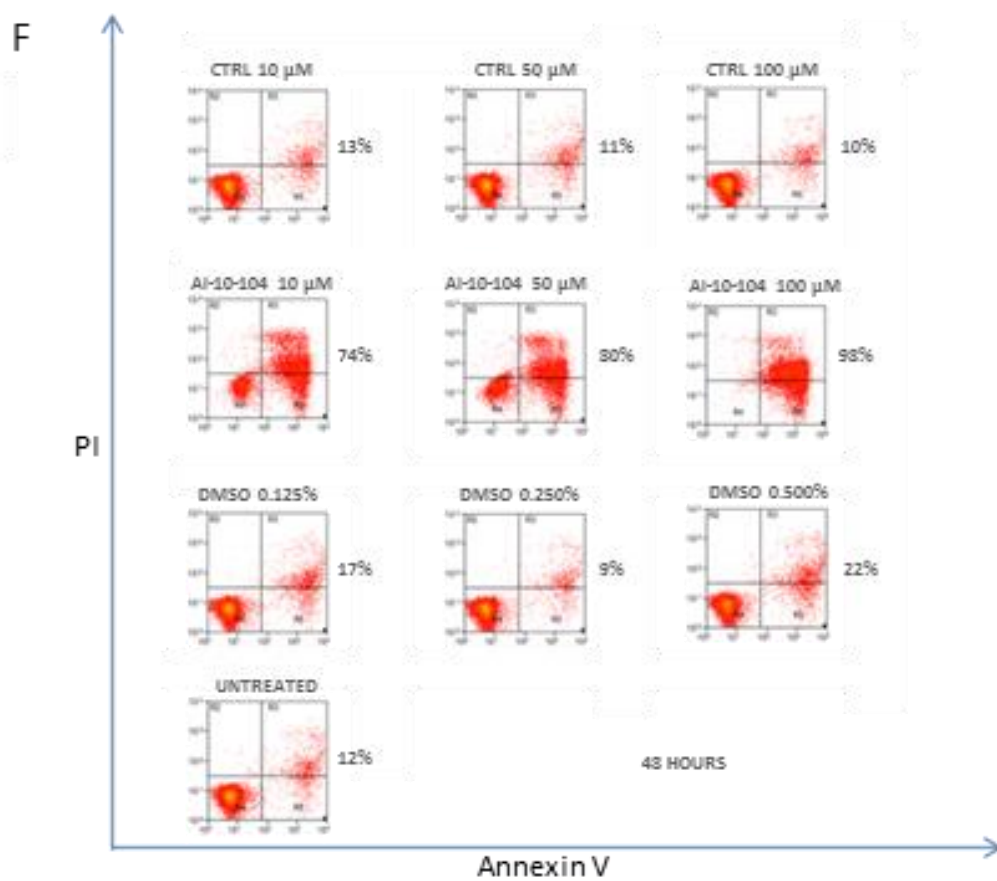
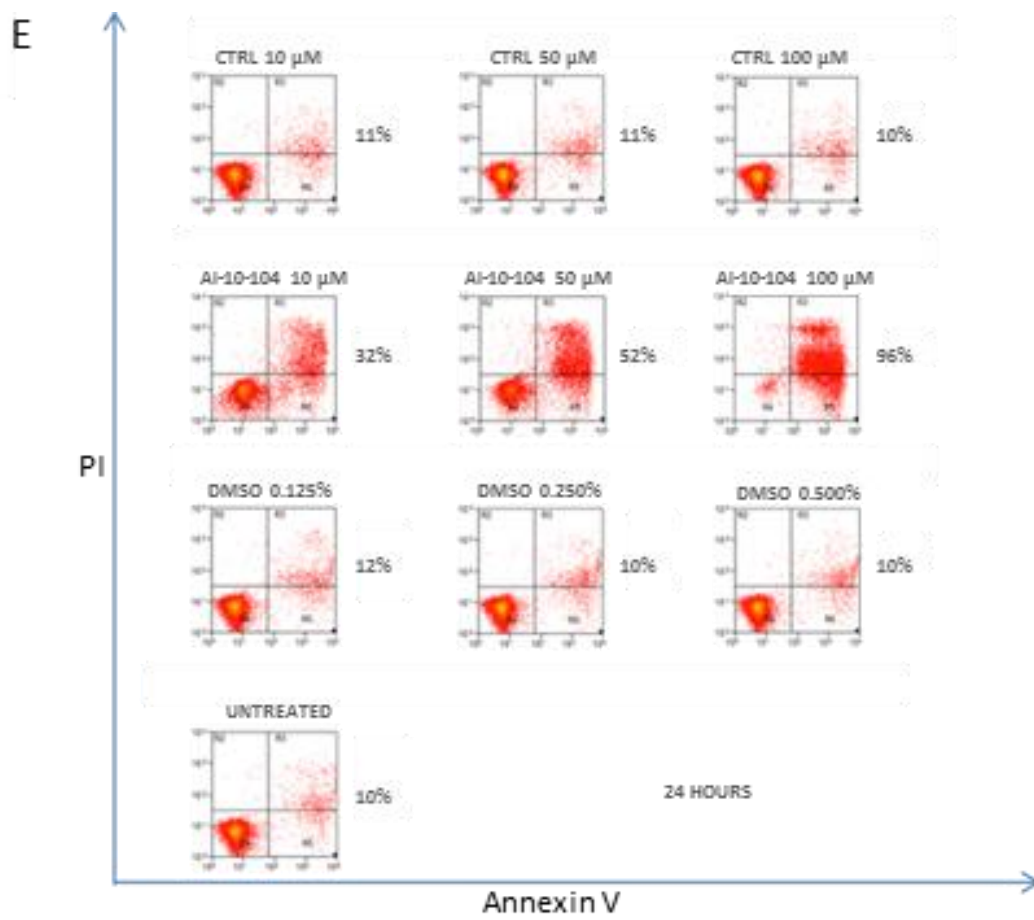
Firstly, it was necessary to determine the cytotoxicity of the compounds, in order to accurately assess the effect of the inhibitor in subsequent gene expression and Chromatin immunoprecipitation (ChIP) experiments. To this end, Kasumi-1 and HL60 cells were

treated with increasing concentrations of the inhibitor compound, in increments of 10  $\mu\text{M}$ . Kasumi-1 cells are a widely used and well characterised model of t(8;21) AML (199, 204, 219, 256, 265). HL60 cells, a promyelocytic leukaemia cell line with only wild type RUNX1, was included in the analysis to examine whether RUNX1 dependence was specific to AML cells with CBF abnormalities. The same titration was conducted with the inactive control compound. An additional DMSO vehicle control was included. In this condition, the DMSO concentration applied was equal to the amount of DMSO administered with the inhibitor and control compounds.

The proportion of apoptotic cells was assessed by staining cells with PI and Annexin V followed by flow cytometry analysis at 6 hrs, 24 hrs and 48 hrs after inhibitor application. The results from treatment with 10  $\mu\text{M}$ , 50  $\mu\text{M}$  and 100  $\mu\text{M}$  concentrations are presented in figure 3-1. An increase in inhibitor concentration was associated with an elevated percentage of apoptotic cells at 24 hrs and 48 hrs. In contrast, at 6 hrs, cell viability was unaffected. HL60 cells were more sensitive to the inhibitor than Kasumi-1 cells (figure 3-1).







**Figure 3-1: Treatment of HL60 and Kasumi-1 cells with CBF complex inhibitor induced apoptosis in a dose dependant manner**

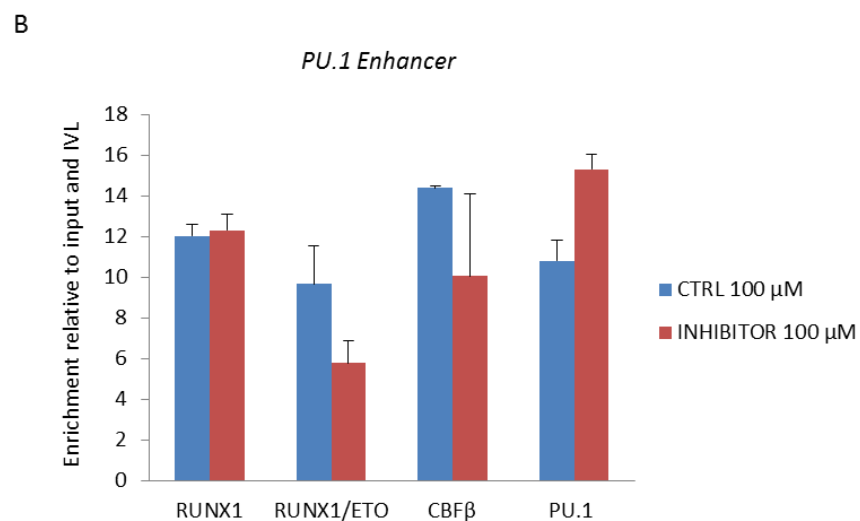
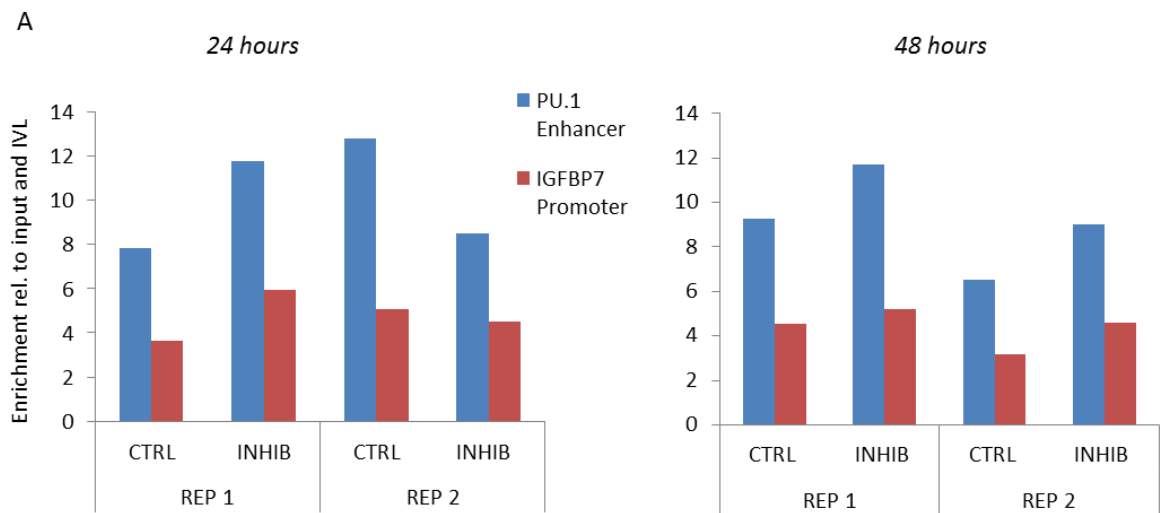
Titration of 10  $\mu$ M -100  $\mu$ M CBF complex inhibitor, the control compound, and a vehicle control (the concentration of DMSO used in each condition) were applied to Kasumi-1 and HL60 cells. Annexin V and propidium iodide (PI) staining followed by flow cytometry analysis was used to assess cell viability at 6 hours, 24 hours and 48 hours. Annexin V-FITC (*x-axis; log scale*) and PI (*y-axis; log scale*) plots are shown from 10  $\mu$ M, 30  $\mu$ M and 100  $\mu$ M of each compound, and from the DMSO control for each concentration. An untreated control was also included. The percentage stated to the right of each plot represents the sum of the proportion of cells in quadrants R3 (cells positive to for Annexin V) and R5 (cells positive for Annexin V and PI). A) Kasumi-1 cells treated for 6 hours. B) Kasumi-1 cells treated for 24 hours. C) Kasumi-1 cells treated for 48 hours. D) HL60 cells treated for 6 hours. E) HL60 cells treated for 24 hours. F) HL60 cells treated for 48 hours. The

### 3.1.2 The CBF complex inhibitor has no significant effect on transcription factor binding

Application of the CBF complex inhibitor led to cell death in both Kasumi-1 and HL60 cells. To determine whether this was due to the inhibition of Runt domain to DNA binding, Chromatin immunoprecipitation (ChIP) analysis was conducted and RUNX1 binding was measured at the *Pu.1* enhancer where this factor, as well as RUNX1/ETO, is known to bind (199). ChIP analysis cannot be conducted on apoptotic cells since cell death will affect transcription factor binding, and mask the specific effects of the compound. To circumvent this issue a low, non-toxic concentration was applied for 24 hours and 48 hours, and binding of RUNX1 to the target loci was assessed at these time points. Inhibitor treatment had no effect on RUNX1 enrichment (figure 3-2 A). We hypothesised that an effect was not seen due to an insufficient concentration of inhibitor. 100  $\mu$ M of inhibitor killed both cell lines at 24 hours, thus suggesting the inhibitor is potent at this concentration. Therefore, ChIP analysis was conducted after 6 hr incubation with 100  $\mu$ M compound; under these conditions inhibitor treated Kasumi-1 and HL60 cells had not yet gone into apoptosis. We also measured PU.1 binding as a control, as it is a protein that should not be affected by the compound. The ChIP results, in both cell lines, suggested the inhibitor still had no effect on the binding of RUNX1, CBF $\beta$ , PU.1 or RUNX1/ETO to the *Pu.1* enhancer, despite the increase to the concentration of inhibitor applied (figure 3-2 B).

In conclusion, the application of the CBF complex inhibitor led to the death of Kasumi-1 and HL60 cells. However, the inhibitor treatment resulted in no detectable effect on the DNA binding of RUNX1 or RUNX1/ETO.





**Figure 3-2: The CBF complex inhibitor had no significant effect on transcription factor binding**  
 A) ChIP-qPCR showing RUNX1 enrichment at the PU.1 enhancer and IGFBP7 promoter in Kasumi-1 cells treated with 10 μM control compound and 10 μM CBF complex inhibitor for 24 hours (left) and 48 hours (right). Bar graph presents two biological replicates (REP 1 and REP 2). B) To circumvent issues with cell death; a high concentration (100 μM) was applied for a short incubation time (6 hours). Cell viability was not effected (data not shown). Bar graphs present ChIP-qPCR data showing RUNX1, RUNX1/ETO, CBFβ and PU.1 enrichment at a shared target, the - 14 kb PU.1 enhancer.

### **3.2 The effect of RUNX1/ETO depletion on gene expression and the role of Sp1 in t(8;21) AML**

Work published by the Alvarez lab suggests that an association between the transcription factor Sp1 and RUNX1/ETO may contribute to RUNX1/ETO mediated leukemic transformation. Furthermore, our DNaseI data revealed that RUNX1/ETO knockdown may have an influence on the binding of Sp-factor family binding proteins to the SP1 motif. Therefore, to investigate the role of Sp1 in t(8;21) AML we assessed the effect of RUNX1/ETO knockdown on genome wide Sp1 binding via CHIP-seq.

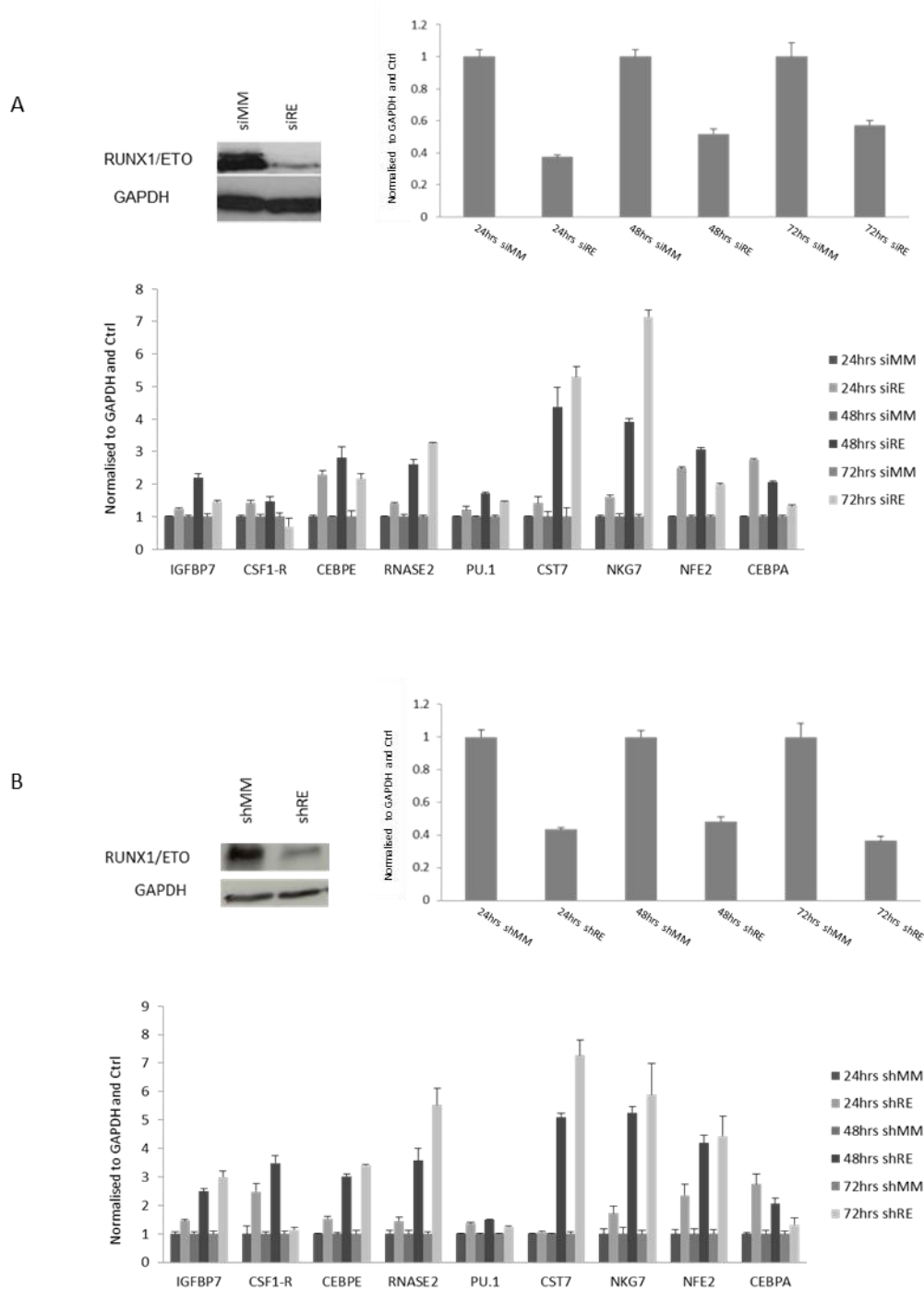
These experiments involved depletion of RUNX1/ETO from t(8;21) AML cells. An efficient siRNA mediated knockdown system has already been established (199, 219, 257). We recently obtained an inducible shRNA system for the depletion of RUNX1/ETO from SKNO-1 R/E cells. These are SKNO-1 cells (a t(8;21) AML cell line) transduced with a doxycycline inducible shRNA specific to RUNX1/ETO (courtesy of Olaf Heidenreich). This is a favourable system as it allows prolonged, stable RUNX1/ETO depletion and circumvents the need of repeat transfections, which will allow more accurate time course analysis of gene expression. Furthermore, it will facilitate the manipulation of t(8;21) AML cells in the presence and absence of RUNX1/ETO.

### **3.2.1 RUNX1/ETO depletion leads to the upregulation of genes involved in myelomonocytic differentiation**

The reversible nature of the RUNX1/ETO mediated block in differentiation has already been demonstrated via targeted depletion of RUNX1/ETO with siRNA (199). Here we wish to test a doxycycline inducible shRNA specific to RUNX1/ETO and see if comparable results are obtained.

We performed RUNX1/ETO knockdown with both systems (siRNA and shRNA), followed by quantitative PCR. These experiments were conducted to confirm efficient RUNX1/ETO knockdown for use in future experiments and will also manually validate our existing microarray data (obtained with the siRNA system) (199). Quantitative PCR analysis was conducted to quantify RUNX1/ETO mRNA levels at 24 hrs, 48 hrs and 72 hrs after transfection or doxycycline induction. In order to avoid depletion of wild type RUNX1, the siRNA and shRNA sequences were designed to target the RUNX1-ETO junction within the transcript of RUNX1/ETO (199). The specificity of the siRNA and shRNA to RUNX1/ETO was assessed by RUNX1 mRNA quantification. As shown in Figure 3-3, effective knockdown of RUNX1/ETO mRNA was achieved with both systems, on the protein and mRNA level; a greater than 50% mRNA knockdown was seen at each time point. Knockdown efficiency was similar with both systems. With the exception of SKNO-1 R/E cells at 72 hours, RUNX1/ETO siRNA and shRNA appeared to have no effect on RUNX1 expression.

Next we investigated the effect of RUNX1/ETO depletion on the expression of known RUNX1/ETO target genes and control genes. The level of mRNA was measured at 24 hrs, 48 hrs and 72 hrs after siRNA transfection/shRNA induction in Kasumi-1 or SKNO-1 R/E cells respectively. We observed a significant up regulation of RUNX1/ETO target genes in both (t8;21) cell lines. With the exception of *CEBPA*, all genes showed the greatest up regulation 48 hrs and 72 hrs post siRNA/shRNA induction. For example *NFE2*, a transcription factor gene involved in haematopoietic differentiation, displayed a 3-4 fold increase in both cell lines (286). There was also upregulation of *CSF1R* and *IGFBP7*. *CSF1R* is a gene encoding the receptor for colony-stimulating factor-1, a growth factor involved in the control of macrophage differentiation (287). *IGFBP7* is gene that has recently been associated with normal hematopoiesis and acute leukemia (288) (figure 3-3).



**Figure 3-3: RUNX1/ETO depletion led to the up regulation of genes involved in haematopoietic differentiation**

Upper panels shows, via western blot and qPCR, the expression level of RUNX1/ETO following three days of RUNX1/ETO knockdown via siRNA transfection (A) or induction of RUNX1/ETO specific shRNA with doxycycline (B). Lower graphs show myeloid gene expression in Kasumi-1 and SKNO-1-R/E cells, 24 hours, 48 hours and 72 hours after transfection with RUNX1/ETO specific siRNA for Kasumi-1 cells (A) or induction of RUNX1/ETO specific shRNA with doxycycline for SKNO-1-R/E cells (B). All data is normalised to GAPDH and relative to the control; the control for Kasumi-1 is cells treated with the control siRNA, the control for SKNO-1-R/E is non induced cells (- doxycycline). Results shown are the average of three biological replicates. Error bars represent the standard deviation between the three replicates.

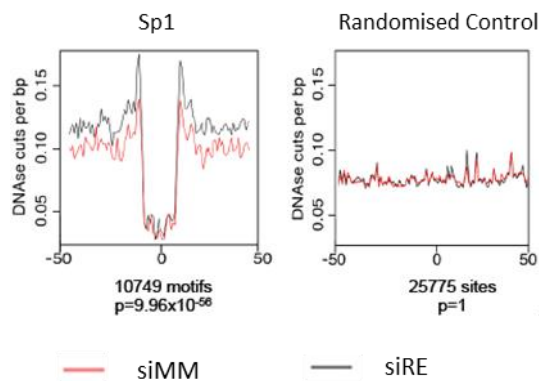
### **3.2.2 The SP1 motif is more protected from DNaseI digestion after RUNX1/ETO knockdown**

Given the extensive gene expression changes and transcription factor binding alterations, we hypothesised that RUNX1/ETO knockdown may have an effect on chromatin accessibility. DNaseI digestion was therefore conducted on cells before and after RUNX1/ETO knockdown in order to determine the effect of RUNX1/ETO knockdown on chromatin accessibility (see section 1.8 and figure 1-13). This DNaseI experiment was conducted by Dr Anetta Ptasinska (219).

Work primarily from Alvarez et al. has suggested that Sp1 may play an important role in t(8;21) AML leukaemogenesis, via an interplay with RUNX1/ETO (247, 262). We therefore used the above DNaseI-seq data to investigate whether Sp1 has a role in the response to RUNX1/ETO knockdown. Using footprinting analysis, we determined whether the Sp1 motif (which is also bound by other members of the Sp-family) was differentially protected from digestion after RUNX1/ETO knockdown, which would indicate that Sp-factor family binding was affected by RUNX1/ETO. Interestingly, when all DNaseI cuts are aligned around the Sp1 motif, we see an increase in protection of the motif after RUNX1/ETO knockdown. This suggests that following RUNX1/ETO knockdown the genome wide binding of Sp proteins might increase (figure 3-4).

### **3.2.3 RUNX1/ETO knockdown has no effect on the expression of Sp1**

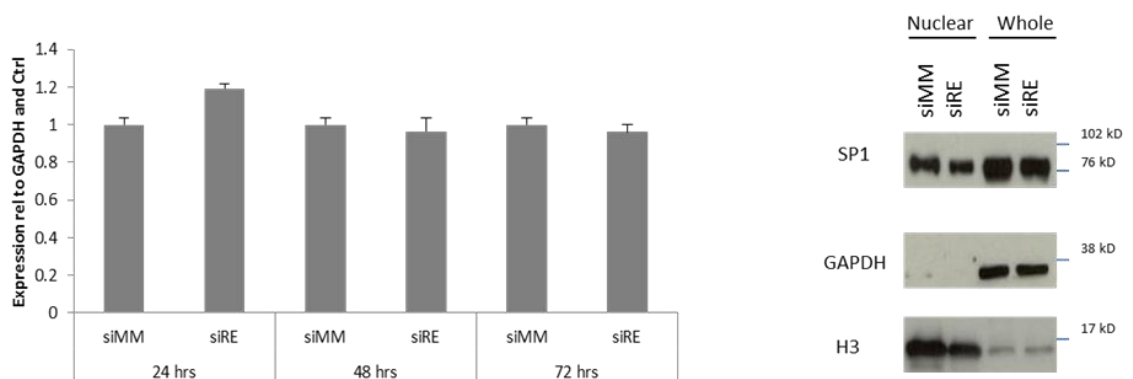
The apparent increase in protection of the Sp1 motif after RUNX1/ETO knockdown could be due to an increase in Sp1 expression. To determine the effect of RUNX1/ETO knockdown on Sp1 expression, Kasumi-1 cells were transfected with either control or RUNX1/ETO specific siRNA and after two days protein and mRNA levels were assessed. RUNX1/ETO knockdown appeared to have no effect on *SP1* mRNA or protein expression (figure 3-5).



**Figure 3-4: RUNX1/ETO knockdown increased protection of the Sp1 motif.**

DNaseI cutting frequency data was aligned around the SP1 motif, across all binding sites. Data from siMM (red) and from siRE (grey) cells were compared. The reduction in cuts at the centre represents the protection of the motif from cleavage by DNaseI, presumably by transcription factor binding.

DNaseI experiment conducted by Dr Anetta Ptasinska.



**Figure 3-5: RUNX1/ETO knockdown had no effect on the expression of Sp1**

A) qRT-PCR showing expression of SP1 over a 3 day time course of siRNA mediated RUNX1/ETO knockdown. Error bars represent the standard deviation between three biological replicates.  
 B) Western blot detecting Sp1 protein in Kasumi-1 cells treated for 48 hrs with mismatch control siRNA (siMM) and with RUNX1/ETO siRNA (siRE). Both nuclear protein and protein from whole cell lysate are shown. GAPDH and H3 expression were used as loading controls for whole and nuclear protein respectively.

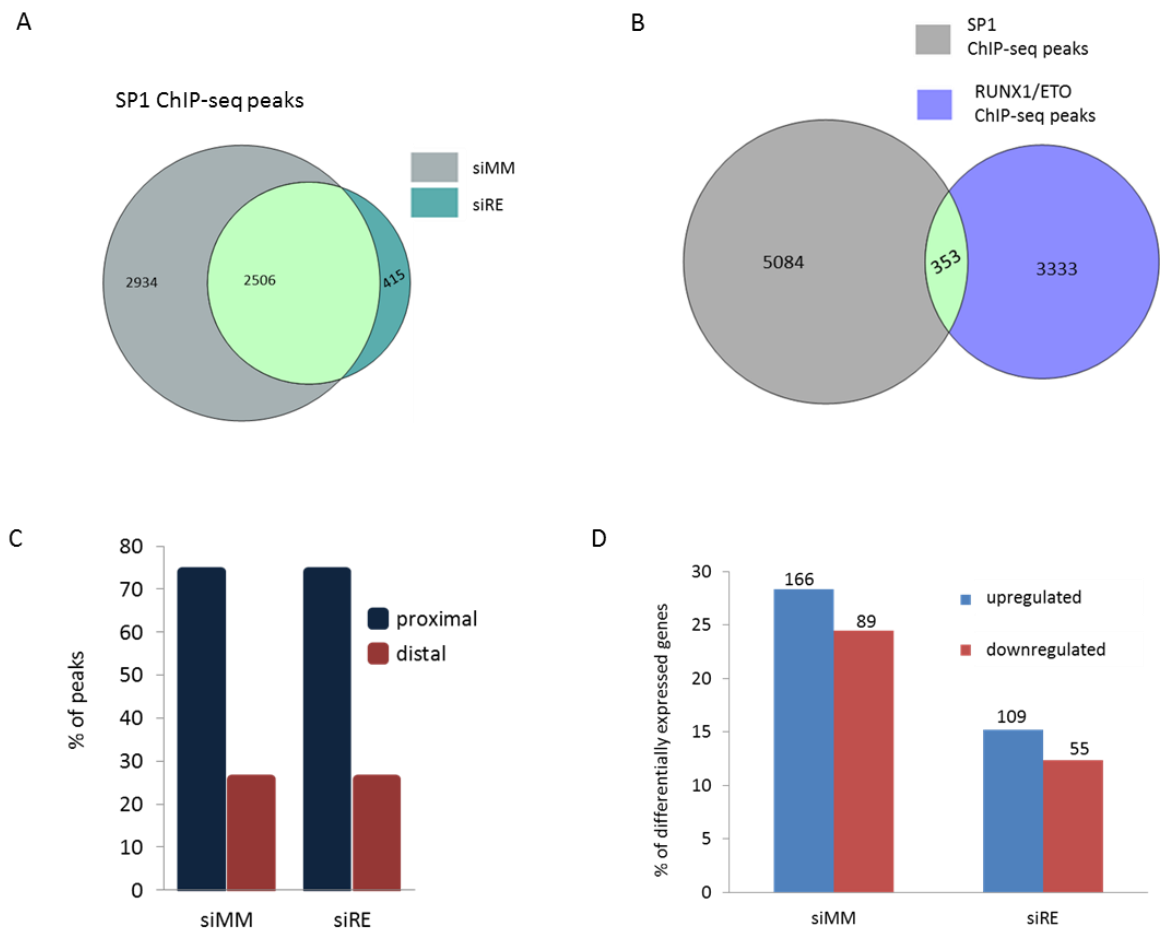


### 3.2.4 Sp1 and RUNX1/ETO bind to distinct sites in the genome

The DNaseI footprinting analysis suggested that RUNX1/ETO knockdown has an effect on Sp1 binding (figure 3-4). To assess whether this was the case, we performed Sp1 ChIP-seq on Kasumi-1 cells transfected with either mismatch or RUNX1/ETO specific siRNA. RUNX1/ETO knockdown led to the loss of 2934 Sp1 binding sites, with the acquisition of only 415 new sites (3-6 A). Maiques-Diaz *et al.* found that there is an enrichment of Sp1 binding at RUNX1/ETO target genes and propose an important role for Sp1 in the DNA binding pattern of RUNX1/ETO (264). In order to determine whether, in our system, RUNX1/ETO and Sp1 are associated with each other at the DNA, we overlapped Sp1 ChIP-seq peaks with RUNX1/ETO ChIP-seq peaks from control Kasumi-1 cells. We found that RUNX1/ETO and Sp1 do not bind the same genomic sites; less than 7% of Sp1 binding sites were co-occupied by RUNX1/ETO (3-6 B). This suggests that Sp1 and RUNX1/ETO do not associate with each other at the DNA.

We then wished to see whether differential Sp1 binding was associated with the gene expression changes which occur after RUNX1/ETO knockdown. This would help us to determine if Sp1 plays a role in RUNX1/ETO mediated transformation and transcriptional reprogramming. To do this, we determined the number of genes differentially expressed by RUNX1/ETO knockdown that were within 1.5 Kb of a Sp1 peak. In control and RUNX1/ETO knockdown cells, Sp1 binds to less than 30% of genes which respond to RUNX1/ETO knockdown. This is the case for both up and down regulated genes (figure 3.2.4 D). Sp1 is predominantly a promoter binding transcription factor, this was confirmed by our analysis (figure 3-6 C) (289).

In summary, we found no association between RUNX/ETO and Sp1 binding. RUNX1/ETO knockdown did not have a considerable effect on Sp1 binding and the two proteins do not associate with each other at the DNA. Sp1 is therefore unlikely to contribute to RUNX1/ETO mediated leukaemogenesis.



**Figure 3-6: Sp1 and RUNX1/ETO bind to distinct sites in the genome and RUNX1/ETO knockdown had no effect on Sp1 binding**

A) RUNX1/ETO knockdown has little effect on genome wide binding of Sp1. Kasumi-1 cells were electroporated with either RUNX1/ETO specific siRNA (siRE) or control siRNA (siMM) . Two days after siRNA electroporation, SP1 binding was measured by ChIP sequencing in both populations. The Venn diagram demonstrates the overlap between siMM SP1 ChIP-seq peaks and siRE SP1 ChIP-seq peaks. B) Sp1 and RUNX1/ETO bind to primarily distinct sites in the genome. The Venn diagram shows the overlap between Sp1 ChIP-seq (siMM data set) and RUNX1/ETO ChIP-seq peaks. RUNX1/ETO ChIP-seq data from Ptasinska *et al.* was used in the analysis (198). C) RUNX1/ETO knockdown has no effect on the distribution of Sp1 peaks relative to transcription start site (TSS) of the nearest gene. The plot demonstrates the proportion of Sp1 ChIP-seq peaks that are proximal or distal to the TSS in both siMM and siRE data sets. D) There is no correlation between Sp1 binding and gene expression. The plot demonstrates the proportion of Sp1 ChIP-seq peaks in Kasumi-1 cells electroporated with either RUNX1/ETO specific siRNA (siRE) or control siRNA (siMM) which are within 1.5 kb of a gene differentially expressed by RUNX1/ETO knockdown. Blue bars are genes upregulated after RUNX1/ETO knockdown. Red bars are genes down regulated after RUNX1/ETO knockdown.

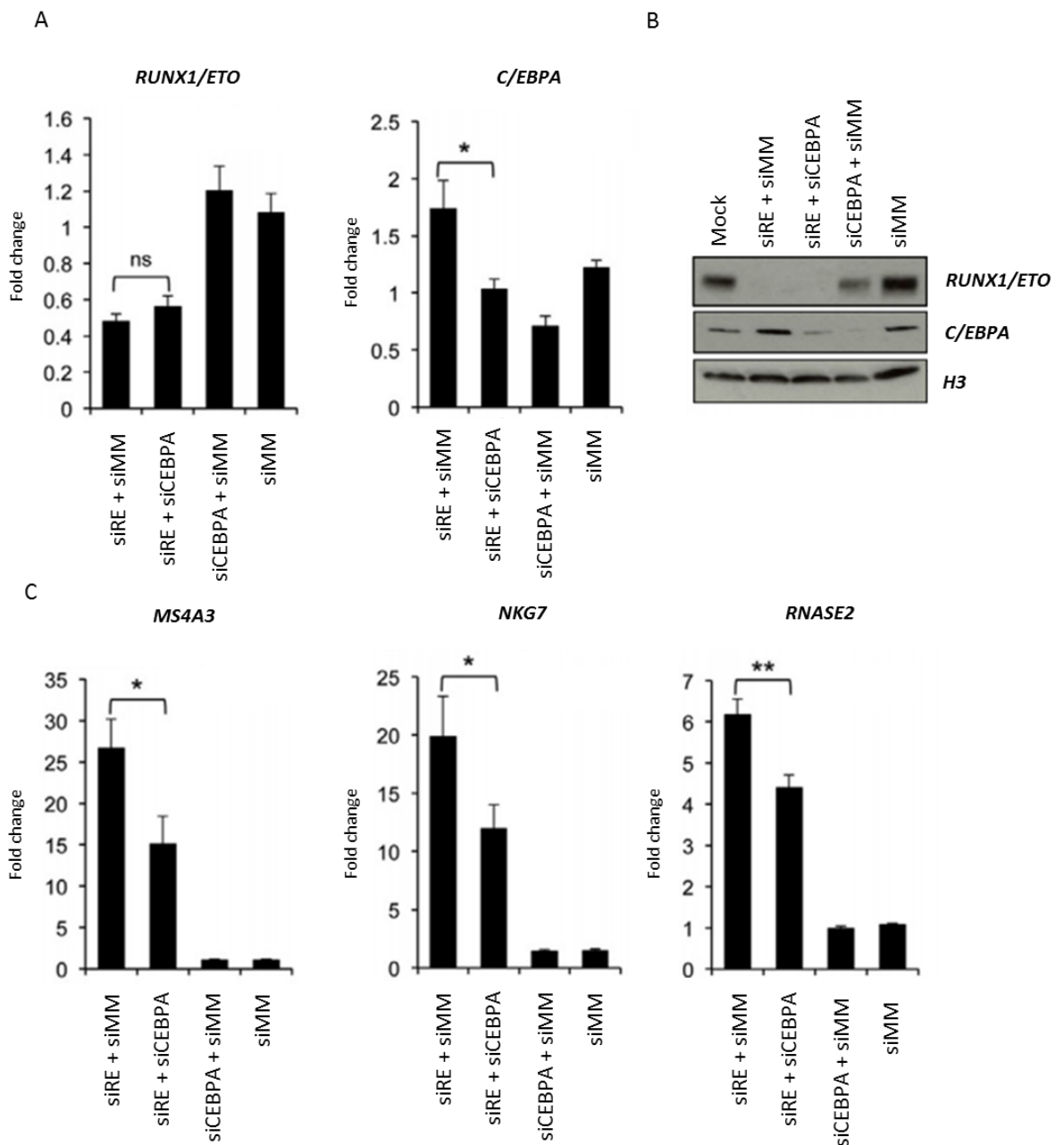
### 3.3 The importance of C/EBP $\alpha$ expression levels in t(8;21) AML

Previous experiments have shown that *CEBPA* is strongly down-regulated in t(8;21) cells. The repression comes from inhibition of *CEBPA* auto-regulation via a direct interaction between the RUNX1/ETO and C/EBP $\alpha$ , as well as the direct repression of the *CEBPA* promoter and enhancer by RUNX1/ETO binding (199) (198). Of the several transcription factors analysed, C/EBP $\alpha$  was the only protein significantly upregulated after RUNX1/ETO knockdown. This was accompanied by an increase in genome wide C/EBP $\alpha$  binding by more than fourfold (figure 1-12), indicating that this factor plays an important role in re-establishing a myelomonocytic gene expression program after knock-down. Combined with the already established crucial role of C/EBP $\alpha$  in myeloid differentiation, these data led us to question whether C/EBP $\alpha$  is the key driver of the differentiation response seen after RUNX1/ETO knockdown.

### **3.3.1 C/EBP $\alpha$ is required for the full upregulation of myeloid genes after RUNX1/ETO depletion**

To test the above hypothesis, we knocked down RUNX1/ETO with and without concomitant knockdown of C/EBP $\alpha$ . As seen in previous experiments (3-3), there was an approximately two fold increase in *CEBPA* expression following RUNX1/ETO knockdown (figure 3-7 A and B) and an increase in the expression of other RUNX1/ETO target genes (*MS4A3*, *NKG7*, *RNASE2*) (figure 3-7 C). The addition of C/EBP $\alpha$  knockdown to RUNX1/ETO knockdown significantly inhibited the upregulation of these genes. These data show that the alleviation of *CEBPA* repression by RUNX1/ETO knockdown is required for the full upregulation of at least a subset of RUNX1/ETO target genes.

*MS4A3*, *NKG7*, *RNASE2* genes had inhibited upregulation in response to the addition of C/EBP $\alpha$  knockdown to RUNX1/ETO knockdown. We found that these genes were upregulated in response to C/EBP $\alpha$  overexpression (see section 1.3.3 and figure 3-11). This validates the role of C/EBP $\alpha$  in the upregulation of myeloid genes after RUNX1/ETO knockdown.



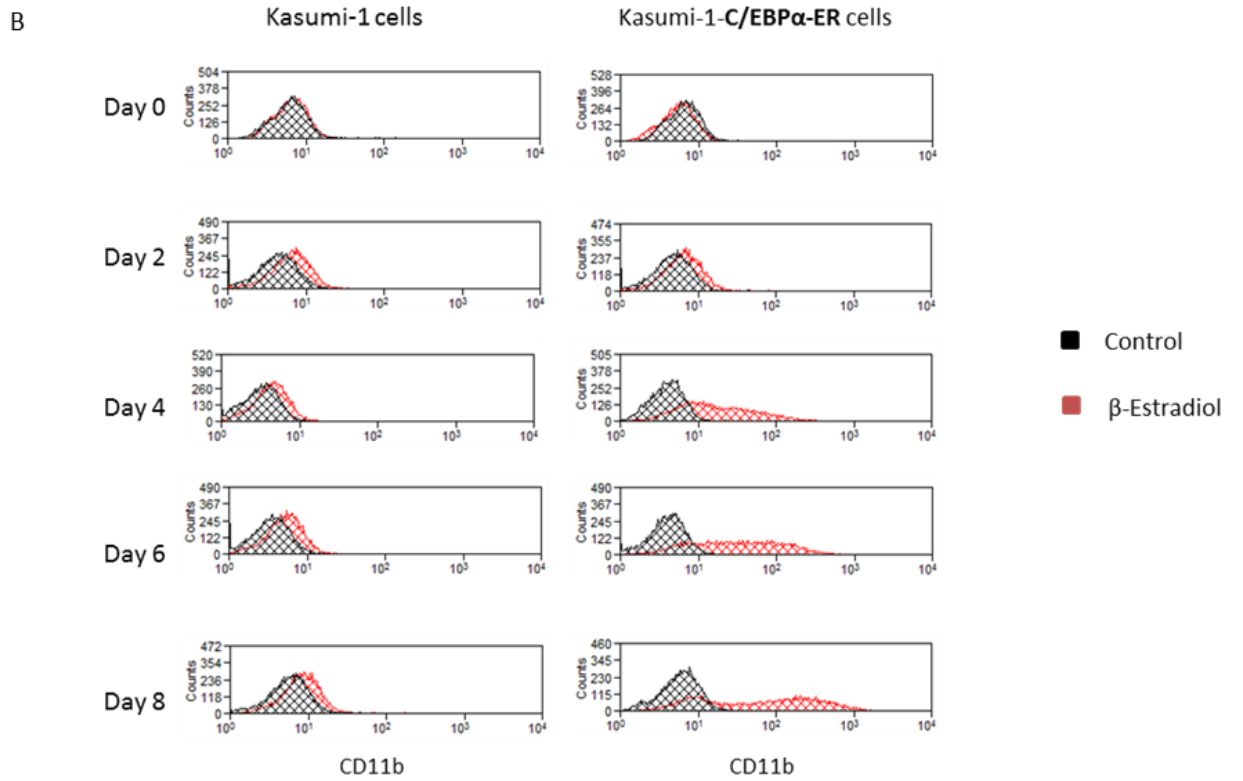
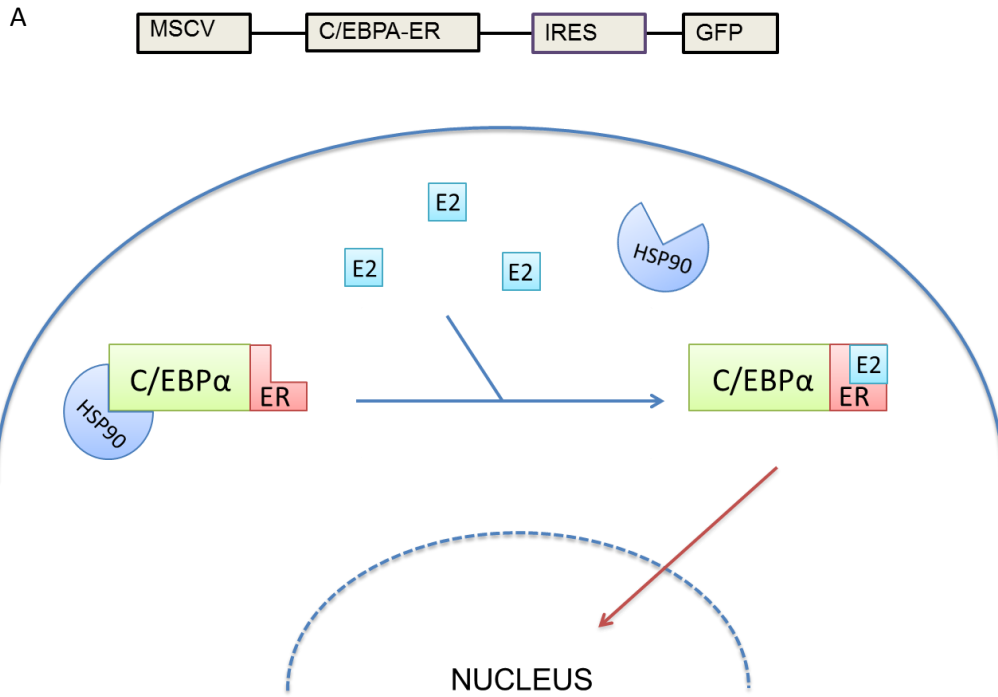
**Figure 3-7: C/EBP $\alpha$  is required for the full upregulation of myeloid genes after RUNX1/ETO depletion**

A) qRT-PCR showing RUNX1/ETO and CEBPA mRNA expression levels in Kasumi-1 cells 72 hrs after electroporation with the indicated siRNAs. Results represent the mean  $\pm$  SEM of five independent experiments. \* $p < 0.05$  by paired student's t test. B) Western blot indicating RUNX1/ETO and C/EBP $\alpha$  protein expression levels in mock transfected (no siRNA), RUNX1/ETO knockdown, C/EBP $\alpha$  knockdown and double knockdown Kasumi-1 cells. Cells were transfected with the indicated siRNAs. An antibody against H3 was used as a loading control. C) mRNA levels of MS4A3, NKG7, and RNASE2 72 hrs after electroporation with the indicated siRNAs. The bars represent the mean  $\pm$  SEM of five independent experiments. \* $p < 0.05$ , \*\* $p < 0.01$  by paired Student's t test. Data published in Ptasińska *et al* 2014.

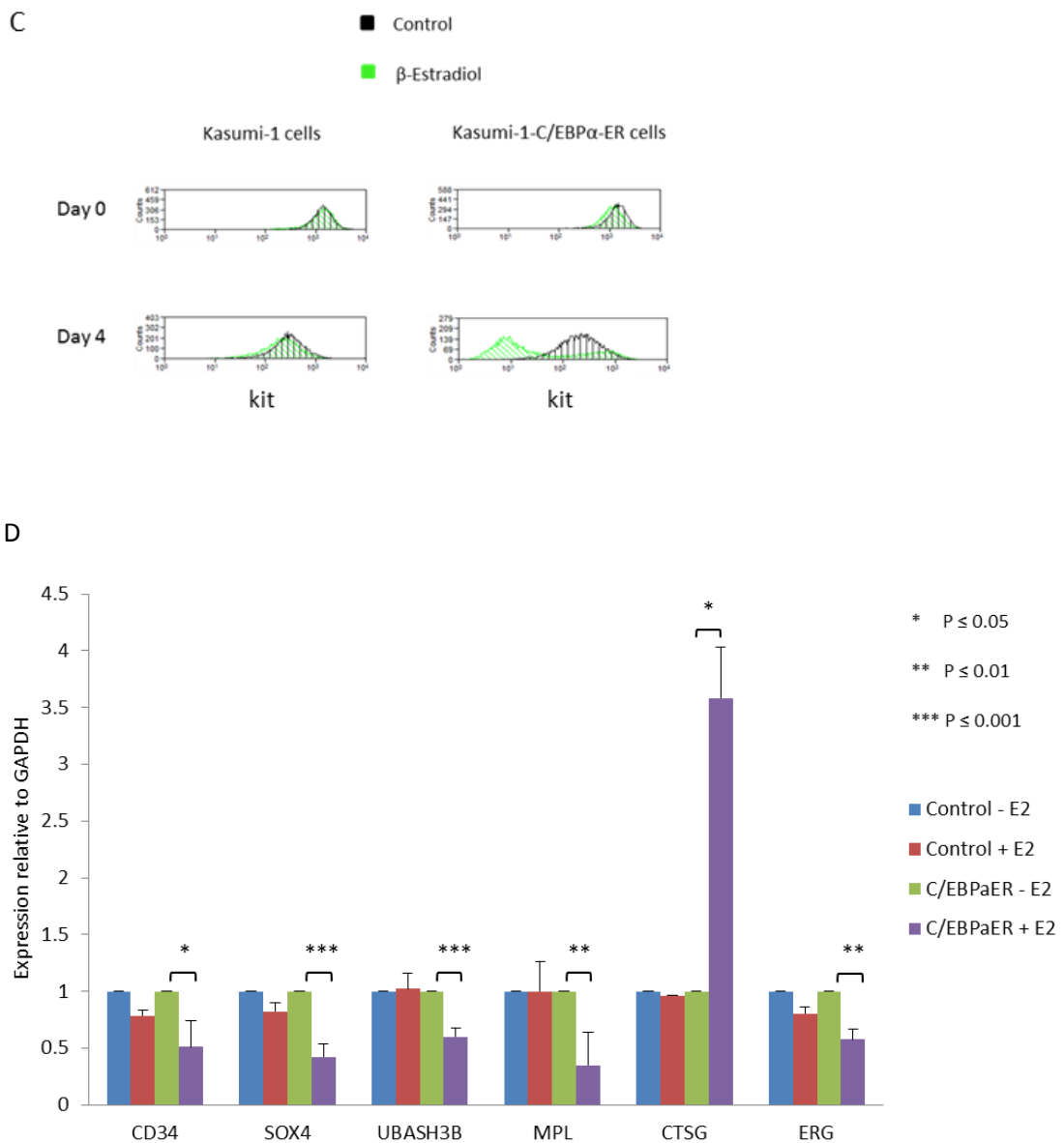
### **3.3.2 Activation of a $\beta$ -Estradiol inducible form of C/EBP $\alpha$ in t(8;21) AML cells alleviates differentiation block**

Given that the alleviation of *CEBPA* from RUNX1/ETO mediated repression was critical for the full differentiation response, we tested if overexpression of C/EBP $\alpha$  is sufficient to override the differentiation block. C/EBP $\alpha$  was overexpressed using a 17 $\beta$ -Estradiol inducible system (290). Kasumi-1 cells were transduced with a retrovirus from which a C/EBP $\alpha$  – estrogen receptor fusion protein is constitutively expressed (see figure 3-8 A and methods section 1.9) (290). However, only upon 17 $\beta$ - Estradiol addition can the fusion protein translocate into the nucleus and have transcriptional activation activity (figure 3-8 A). Induction of C/EBP $\alpha$  over a time course of 8 days lead to a gradual increase in myeloid differentiation, demonstrated by an increase in the surface expression of the myeloid marker CD11b (figure 3-8 B). This was accompanied by a decrease in the surface expression of c-Kit (figure 3-8 C).

C/EBP $\alpha$  is known to both drive the expression of myeloid genes and suppress stem cell genes (202). This was also true in this study as manual qPCR gene expression analysis showed that C/EBP $\alpha$  induction led to an increase in the expression of myeloid genes such as *CTSG* and a down regulation of the stem cell genes *MPL*, *SOX4*, *ERG* and *CD34* (figure 3-8 D).







**Figure 3-8: Activation of a  $\beta$ -Estradiol inducible form of C/EBP $\alpha$  in t(8;21) AML cells alleviated differentiation block with a reduction in self renewal**

A) Kasumi-1 cells were transduced with a construct from which a C/EBP $\alpha$  – estrogen receptor fusion protein is constitutively expressed. Upon 17 $\beta$ - Estradiol addition, the fusion protein can dissociate from Hsp90 and translocate into the nucleus. B) C/EBP $\alpha$  activation in Kasumi-1 cells stably transduced with C/EBP $\alpha$ -ER fusion leads to an increase in CD11b cell surface expression. The flow cytometry plots show a gradual increase in CD11b expression over 8 days of 17 $\beta$ -Estradiol mediated C/EBP $\alpha$  activation. C) C/EBP $\alpha$  activation in Kasumi-1 cells stably transduced with C/EBP $\alpha$ -ER fusion leads to a decrease in Kit surface expression after 4 days. D) C/EBP $\alpha$  activation in Kasumi-1 cells stable transduced with C/EBP $\alpha$ -ER fusion leads to a decrease in the expression of selected self-renewal genes and an increase in the expression of selected myeloid genes. Mock transduced Kasumi-1 cells were used as a control (Kasumi-1). Results shown are the average of three biological replicates and error bars indicate the standard deviation. ‘ctrl’ is no 17 $\beta$ -Estradiol ‘B-Estradiol’ is four days 17 $\beta$ -Estradiol treatment

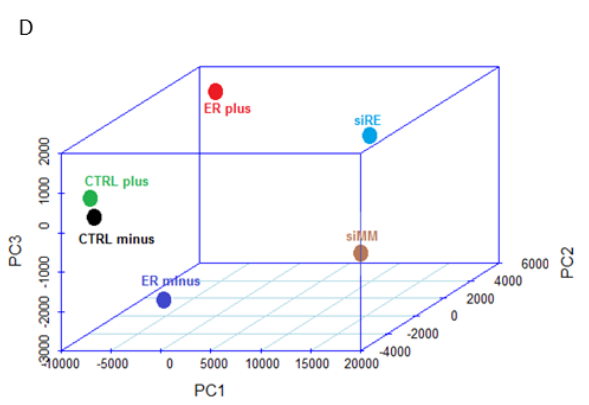
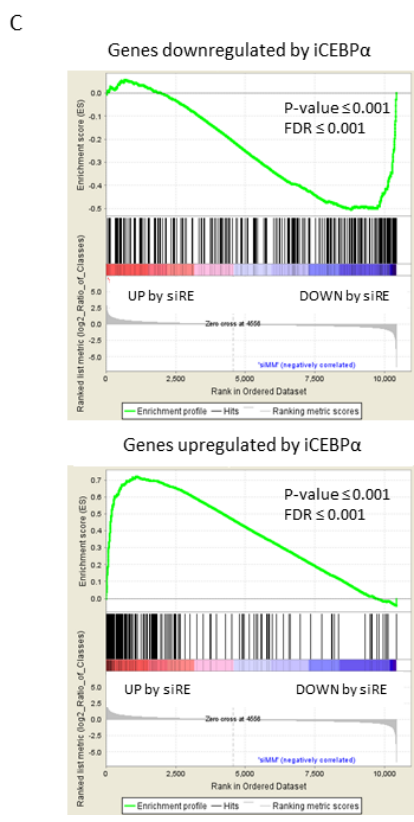
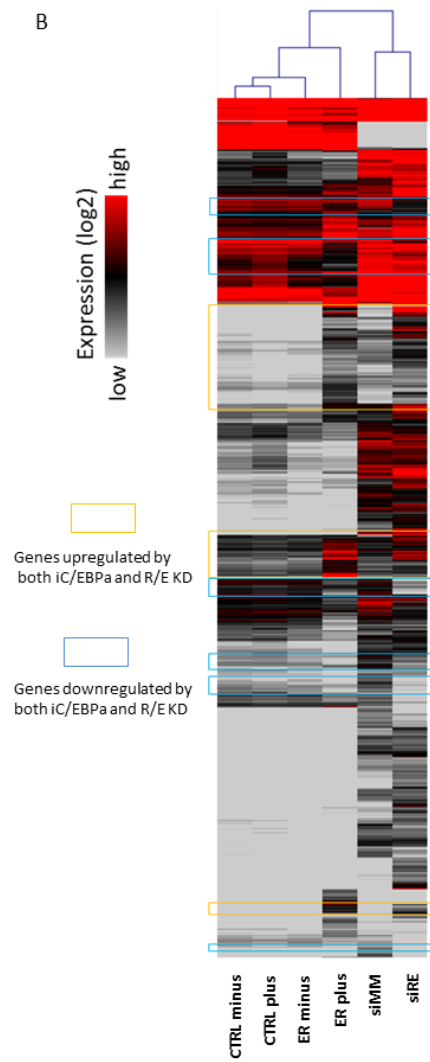
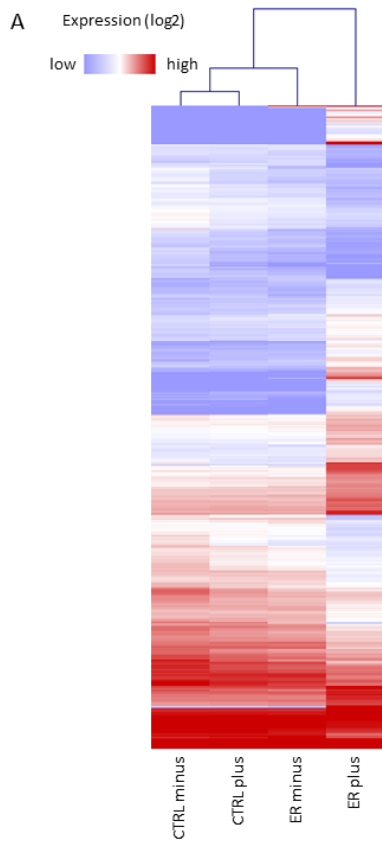
### **3.3.3 Genes that are differentially expressed after RUNX1/ETO knockdown are correlated with genes that are differentially expressed by C/EBP $\alpha$ induction**

Our experiments so far suggest that C/EBP $\alpha$  overexpression may be sufficient to override the RUNX1/ETO mediated differentiation block and direct cells back on course for myeloid differentiation. Such a process entails that the changes in gene expression patterns observed during myelopoiesis would be restored. To test this theory, and directly compare the effect of RUNX1/ETO knockdown with C/EBP $\alpha$  over expression, we conducted a genome wide mRNA expression analysis. RNA-seq libraries were prepared from Kasumi-1 cells with and without C/EBP $\alpha$  induction, as well as mock transduced cells with and without 17 $\beta$ - Estradiol treatment as a control. These data were then compared to existing RNA-seq data sets generated from Kasumi-1 cells with and without RUNX1/ETO knockdown, which were generated in Olaf Heidenreich's lab (199).

Expression data from genes which changed expression by at least two fold, in response to either C/EBP $\alpha$  or RUNX1/ETO knockdown, were subject to unsupervised hierarchical clustering and presented using a heat map (figure 3-9 A). The data show very few changes in gene expression in control cells whereas the induction of C/EBP $\alpha$  led to a dramatic alteration in the gene expression pattern. We then identified genes which changed expression by more than two-fold after 4 days of C/EBP $\alpha$  induction, as well as after 4 days of RUNX1/ETO knock-down. Figure 3-9 B displays an unsupervised clustering of fold-change of gene expression, indicating common clusters of genes which were up- or downregulated by both RUNX1/ETO knockdown and C/EBP $\alpha$  overexpression.

To examine the significance of the association between the genes differentially expressed by RUNX1/ETO knockdown and those differentially expressed by C/EBP $\alpha$  induction, we performed Gene Set Enrichment Analysis (GSEA) (figure 3-9 C). GSEA analysis determines the enrichment of genes differentially expressed by C/EBP $\alpha$  in the set of genes differentially expressed by RUNX1/ETO knockdown. Furthermore, it determines whether the genes respond in the same way to each manipulation. The plots demonstrate that many of the differentially expressed genes are shared between the two gene sets. A p-value of <0.001 indicates that the genes differentially expressed by induction of C/EBP $\alpha$  represent a statistically significant proportion of the set of genes differentially expressed by RUNX1/ETO depletion. Furthermore, it shows that those genes upregulated by C/EBP $\alpha$  are often also upregulated by RUNX1/ETO. The same is true for genes which are down regulated by each condition. Examples of shared, upregulated genes include the myeloid genes *CTSG*, *LAPTM5*, *LCP1*, *MS4A3*, *NKG7* and *RNASE2*. Examples of shared down regulated genes include *DUSP6* (a regulator of cell signalling) (291) and the stem cell gene CD34.

To visually compare the data sets in a genome wide fashion, principal component analysis was conducted with all expressed genes (figure 3-9 D). The three largest components were used to separate the gene sets. The expression of C/EBP $\alpha$ -ER (via the addition of 17 $\beta$ - Estradiol) led to changes in the gene expression profile of Kasumi-1- C/EBP $\alpha$ ER cells, leading to a separation of 'ER plus and ER minus' data sets by principal components two and three. In contrast, 'CTRL plus and CTRL minus' are positioned directly adjacent to each other, demonstrating the lack of gene expression change following doxycycline addition. The siMM and siRE gene sets are spread in the same direction by the same principal components (two and three), suggesting similar gene expression changes occur following RUNX1/ETO knockdown and C/EBP $\alpha$  overexpression. However, principal component one separates the RUNX1/ETO knockdown and C/EBP $\alpha$  over expression gene sets, suggesting there are some differences in the gene expression profiles.



**Figure 3-9: Genes that are differentially expressed by RUNX1/ETO knockdown were correlated with genes that were differentially expressed by C/EBP $\alpha$  induction**

A) Hierarchical clustering of genes differentially expressed in response to 4 days CEBP $\alpha$  induction. The genes included in the heat map are those which are differentially expressed between 'ER minus' and 'ER plus' datasets i.e. genes that change expression following C/EBP $\alpha$  induction with 17 $\beta$ -Estradiol. The expression level of these genes in mock transduced Kasumi-1 cells with and without 17 $\beta$ -Estradiol is shown in lanes 1 and 2 respectively). B) Hierarchical clustering of genes differentially expressed, by at least twofold, in response to 4 days RUNX1/ETO knockdown and 4 days CEBP $\alpha$  induction in Kasumi-1 cells. The heat map plots the RNA log<sub>2</sub> raw expression derived from RNA-seq data. Dark red indicates highly upregulated genes and grey indicates highly downregulated genes. Note how there are subsets of genes which are upregulated/downregulated by both C/EBP $\alpha$  induction and RUNX1/ETO knockdown. These represent a significant proportion of the heat map. These are highlight by the yellow and light blue boxes. 'ER plus' and 'ER minus' are data sets from Kasumi-1 cells transduced with the CEBP $\alpha$ -ER construct, with and without 17 $\beta$ -Estradiol treatment respectively. siMM and siRE are data sets from Kasumi-1 cells transfected with mismatch siRNA and RUNX1/ETO specific siRNA respectively. C) Gene Set Enrichment Analysis (GSEA) of genes differentially expressed following CEBP $\alpha$  induction vs genes differentially expressed following RUNX1/ETO knockdown. The upper section of the plot (green line) indicates the running enrichment score for the gene set as the analysis goes along the ranked list of genes (gene differentially expressed siMM vs siRE). The middle section of the plots (black vertical lines) indicates where genes of the gene set appear in the ranked list of genes. The lower portion is the ranking metric of the ranked gene list (siMM vs siRE genes). The iCEBP $\alpha$  and siMM vs siRE gene expression data were obtained via RNA-seq. D) 3D Principal component analysis using the three largest principal components to separate the gene expression data set, which were derived from RNA-seq experiments. 'CTRL plus' is data from mock transduced Kasumi-1 cells treated with 17 $\beta$ -Estradiol. 'CTRL minus' is data from mock transduced Kasumi-1 cells without 17 $\beta$ -Estradiol treatment. See 'B' for explanation of point labels.

### **3.3.4 Almost half of the genes differentially expressed by RUNX1/ETO knockdown are also differentially expressed by C/EBP $\alpha$ induction**

The results so far suggest that RUNX1/ETO knockdown and C/EBP $\alpha$  overexpression have similar effects on the gene expression profile of Kasumi-1 cells. To directly quantify the extent to which C/EBP $\alpha$  induction mimics RUNX1/ETO knockdown, we overlapped the genes upregulated by RUNX1/ETO knockdown with genes upregulated by induction of C/EBP $\alpha$ , and presented the data in a Venn diagram (figure 3-10). The same analysis was performed for down-regulated genes. Almost half the genes upregulated by C/EBP $\alpha$  induction were also upregulated by RUNX1/ETO knockdown. This was also the case for down regulated genes (figure 3-10 A) (See the appendices for a list of differentially expressed genes that are shared between RUNX1/ETO knockdown and C/EBP $\alpha$  overexpression).

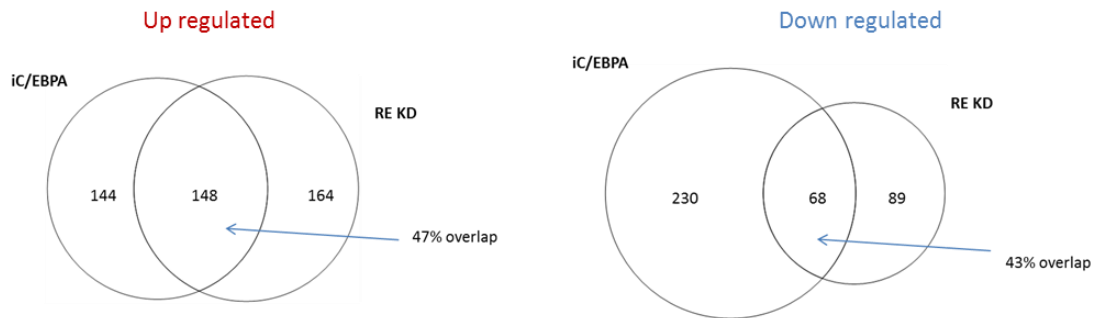
To try and determine if the effects seen by C/EBP $\alpha$  overexpression are directly due to C/EBP $\alpha$  binding, we calculated the proportion of genes differentially expressed by C/EBP $\alpha$  overexpression that are C/EBP $\alpha$  targets. We used existing C/EBP $\alpha$  ChIP-seq data to assign C/EBP $\alpha$  target genes (219). We found that the majority of upregulated genes are C/EBP $\alpha$  targets, however only 20% of down regulated genes are normally bound by C/EBP $\alpha$  (figure 3-10 B).

We next investigated whether the genes differentially expressed by C/EBP $\alpha$  induction were also bound by RUNX1/ETO, to examine whether C/EBP $\alpha$  up-regulation could directly override the repression mediated by RUNX1/ETO. The results suggest this could be the case; the majority of differentially expressed genes were RUNX1/ETO targets (figure 3-10 B).

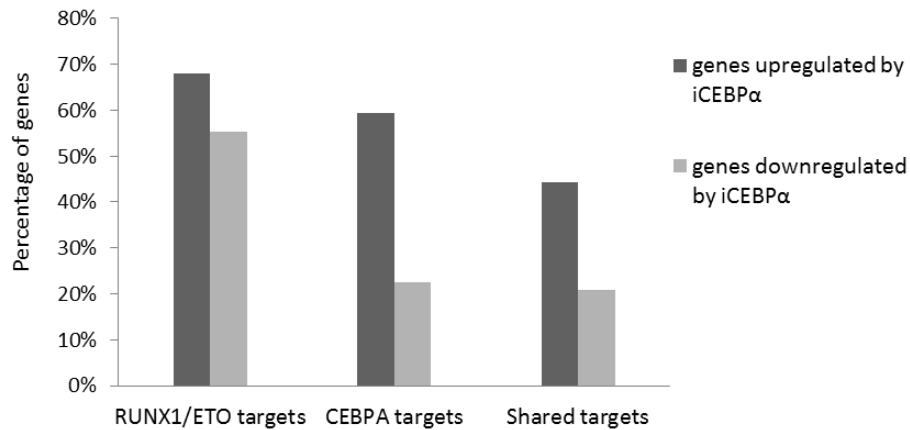
Taken together, our data indicated that C/EBP $\alpha$  is necessary for the differentiation response seen after RUNX1/ETO knockdown, and C/EBP $\alpha$  overexpression was sufficient to override the RUNX1/ETO mediated blockade of differentiation. This demonstrated that C/EBP $\alpha$  is an essential part of the pathway repressed by RUNX1/ETO.



A

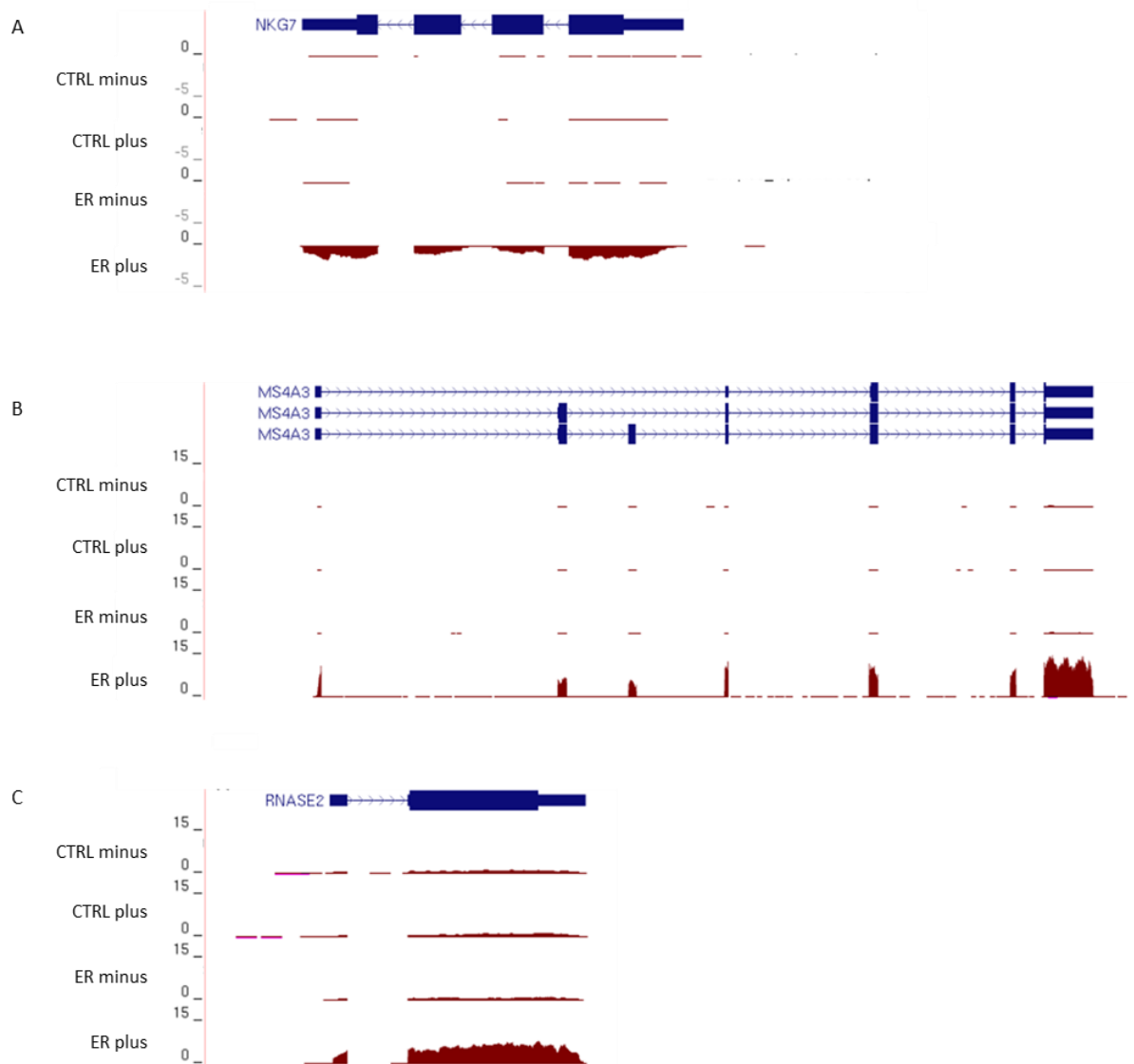


B



**Figure 3-10: Almost half of the genes differentially expressed by RUNX1/ETO knockdown were also differentially expressed by C/EBPα induction**

A) Left Venn diagram demonstrates the overlap between genes upregulated by induction of C/EBPα and those upregulated by siRE. Right Venn diagram demonstrates the overlap between genes downregulated by induction of C/EBPα and those downregulated by siRE. Data was obtained via RNA-seq and only genes with a greater than two fold expression change were included. B) The bar graph presents the proportion of genes upregulated (dark grey)/down regulated (light grey) by C/EBPα that are bound by RUNX1/ETO or C/EBPα or both. Binding was determined by ChIP-seq.



**Figure 3-11: RUNX1/ETO target genes that were inhibited by C/EBP $\alpha$  knockdown were upregulated by induction of C/EBP $\alpha$**

UCSC genome browser screen shot of RNA-seq data. Tracks represent, from top to bottom: mock transduced Kasumi-1 cells, mock transduced Kasumi-1 cells with 17 $\beta$ -Estradiol, Kasumi1-C/EBP $\alpha$ ER cells and finally Kasumi1-C/EBP $\alpha$ ER cells with 17 $\beta$ -Estradiol. Y-Axis represents normalised RPKM values for the expression of A) *NKG7* B) *MS4A3* and C) *RNASE2*. Data shown is from one of two biological replicates with similar results.

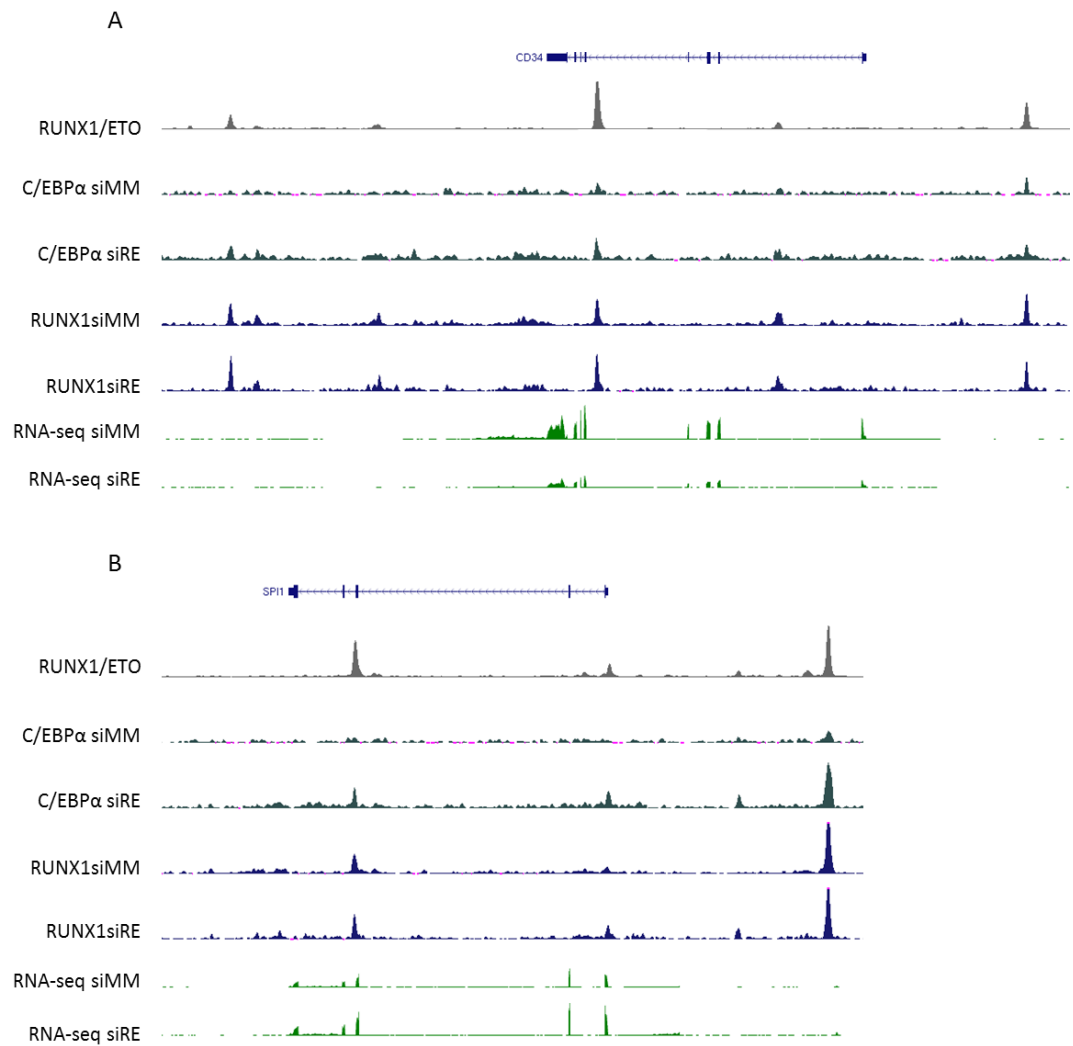
### **3.4 The effect of RUNX1/ETO knockdown on selected promoter-enhancer interactions**

It is well established that alterations in gene expression can be reflected by changes in DNA looping interactions (63, 118). There is an increasing body of evidence that transcription factors initiate and maintain these DNA loops (63, 279-281). We therefore hypothesised that the gene expression changes and alterations to the transcription factor binding profile after RUNX1/ETO knockdown may be accompanied by alterations in specific promoter enhancer interactions. To this end we used 4C-seq before and after RUNX1/ETO depletion to look at the contact intensity of specific DNA interactions.

4C-seq is a method used to assess all the interactions taking place with a specific region of the genome, termed the 'viewpoint'. The technique follows the fundamental principles of chromosome conformation captures; DNA is crosslinked, then fragmented and re-ligated to generate chimeric DNA fragments which represent regions of the genome that interact *in vivo*. These are then characterised by high throughput sequencing.

We used 4C-seq to determine interactions at the *CD34* locus and *SPI1/PU.1* locus. *CD34* gene was of interest as *CD34* expression is associated with the immature status of hematopoietic cells; expression is rapidly lost after hematopoietic cells differentiate. Moreover, its expression is also down-regulated after RUNX1/ETO depletion (199). Furthermore, at this locus we see significant alterations in the enrichment of transcription factors, namely the myeloid transcription factor C/EBP $\alpha$  (219). PU.1 is a myeloid transcription factor encoded by the *SPI1* gene. It has a crucial role in haematopoietic differentiation. *SPI1* also exhibits changes in C/EBP $\alpha$  enrichment after RUNX1/ETO depletion (219) (figure 3-12).

*SPI1* and *CD34* gene expression is regulated by previously characterised enhancer regions. These enhancers physically loop to their respective promoters to regulate gene expression (292-294). Work from the Tenen lab has demonstrated that RUNX1 binding sites are crucial for the interaction between the promoter and enhancer of both the *PU.1* and *CD34* loci (295) (63). As RUNX1 and RUNX1/ETO both contain a RUNT domain, and both bind to these genes (199), it is possible that removal of RUNX1/ETO could therefore have an influence on the interactions taking place at these loci.



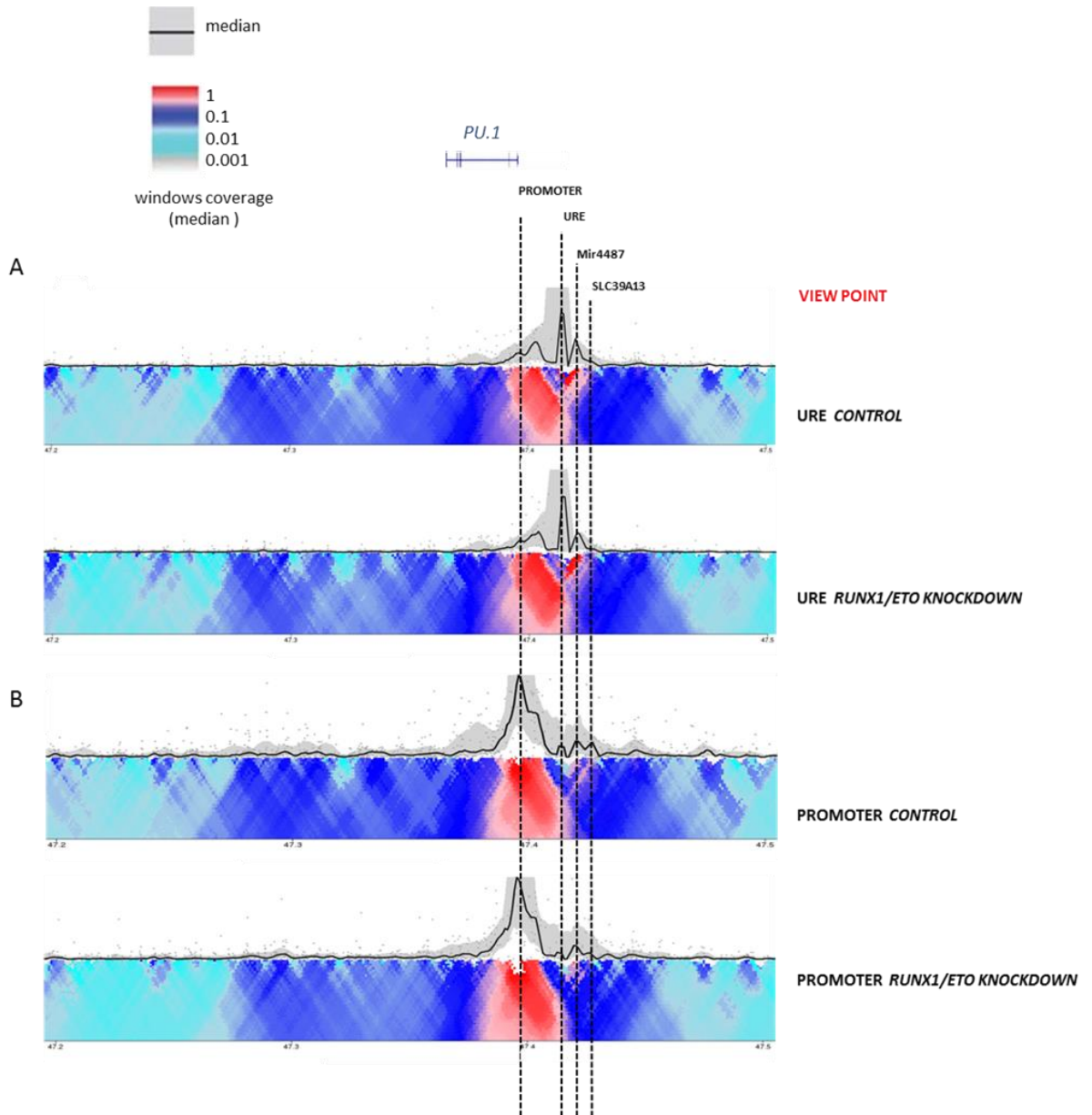
**Figure 3-12: *SPI1* and *CD34* loci were selected for 4C experiments as they are bound by RUNX1/ETO, RUNX1 and C/EBPα and are differentially expressed after RUNX1/ETO knockdown**

Transcription factor binding at the A) *SPI1* and B) *CD34* loci. Top track (grey) shows RUNX1/ETO ChIP-seq peaks in Kasumi-1 cells. Tracks 2 and 3 are C/EBPα ChIP-seq peaks in Kasumi-1 cells transfected with siMM and siRE respectively. Tracks 4 and 5 are RUNX1 ChIP-seq peaks in Kasumi-1 cells transfected with siMM and siRE respectively. Tracks 6 and 7 are RNA-seq data from Kasumi-1 cells transfected with siMM and siRE respectively.

### **3.4.1 A reciprocal interaction between the *SPI1* promoter and its upstream regulatory element (URE) was detected**

In order to determine the effect of RUNX1/ETO expression on selected cis-regulatory element interactions, RUNX1/ETO was depleted from Kasumi-1 cells via a four-day siRNA mediated knockdown. Chromatin from two biological replicate experiments was harvested. The knockdown was efficient and consistent between the two replicates. We then assessed DNA interactions with the *SPI1* and *CD34* loci using 4C-seq.

PU.1 expression is regulated by an enhancer element 14Kb upstream of the *SPI1* promoter (URE)(293). In agreement with previous findings (293) we detected a strong interaction between the two *cis*-elements which was indicated by a peak in the median contact intensity trend line at the *SPI1* promoter, with a corresponding red region in the contact intensity heat map (figure 3-13 A). The interaction was confirmed by a reciprocal 4C experiment in which the promoter was used as a viewpoint (figure 3-13 B). The reciprocal 4C revealed another interaction -22kb from the *SPI1* promoter which was consistently associated with both the *SPI1* promoter and URE (figures 3-13).



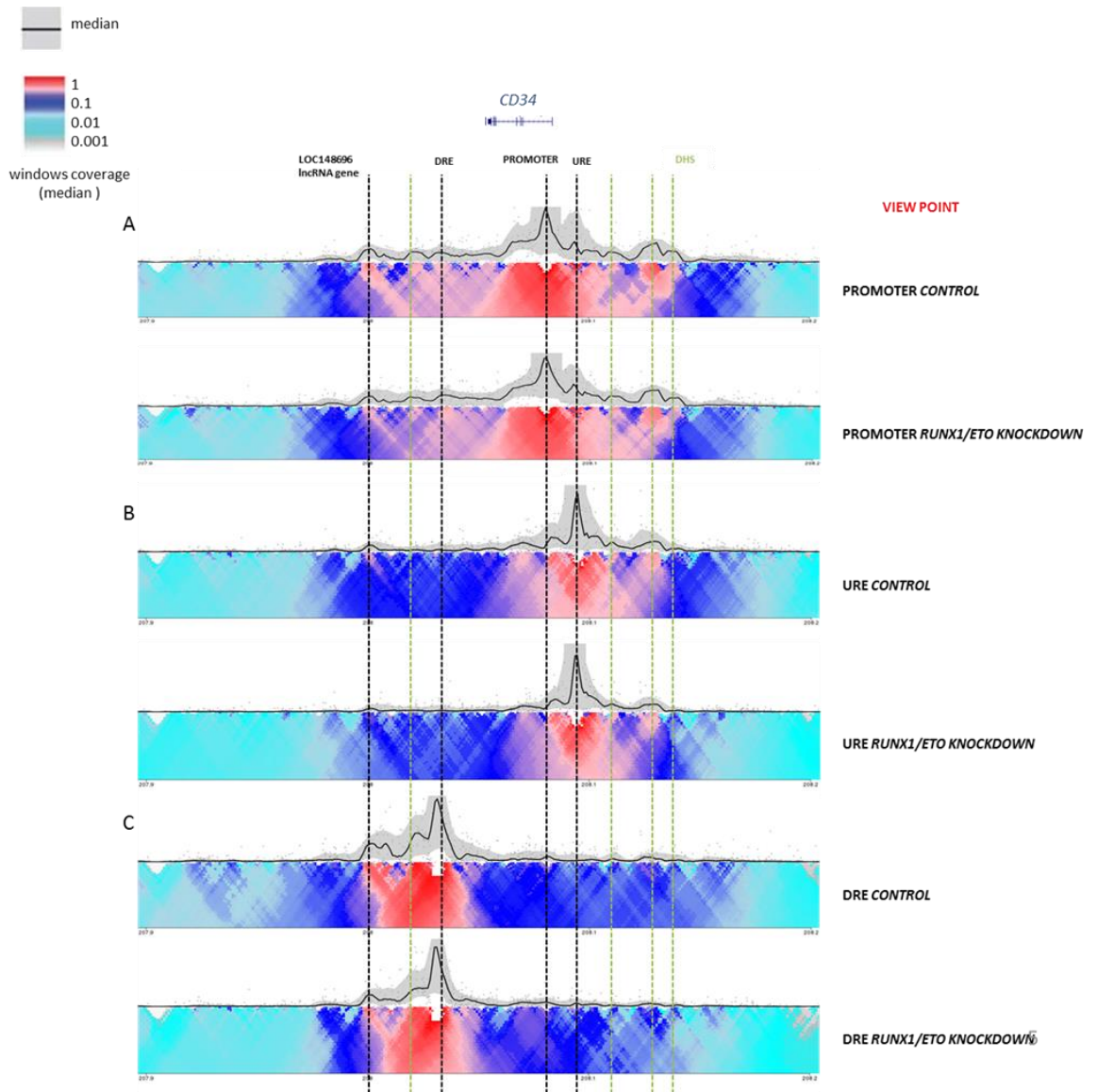
**Figure 3-13: 4C-seq detected a reciprocal interaction between the *SPI1* promoter and URE**  
 Local 4C contact profile with a viewpoint from the *PU.1* upstream regulatory element (URE) (A) and *PU.1* promoter (B), using chromatin from Kasumi-1 cells transfected with control siRNA (siMM) RUNX1/ETO specific siRNA (siRE). In the top panel (main trend), the contact intensity (black line) is calculated using a running median analysis of normalised read counts with a 3 kb sliding window. The 20th and 80th percentile are visualised as a grey trend graph. In the bottom panel, contact intensities are computed using linearly increasing sliding windows (scaled 2–50 kb) and displayed as a colour-coded heat map of positive 4C signal (maximum interaction set to 1). Local colour changes are log-scaled to indicate changes of statistical enrichment of captured sequences, corresponding to the enhancer – promoter interaction. The results presented here are an average of two biological replicates.

### 3.4.2 A reciprocal interaction between the CD34 promoter and its URE was detected

Experiments in transgenic mice have shown that *CD34* expression in long-term haematopoietic stem cells is affected by the physical interaction between the promoter and a -19kb downstream regulatory element (DRE) (63, 292) . With the *CD34* promoter as a viewpoint, an interaction between this DRE and the *CD34* promoter was detected in our human t(8;21) AML model. We also detected an interaction between the *CD34* promoter and an element +12kb upstream (URE) (figure 3-14 A). This interaction was confirmed via reciprocal 4C with the URE as the viewpoint (figure 3-14 B)

Figure 3-14 also demonstrates that the neighbouring gene, LOC148696, interacts with the *CD34* promoter, URE and the DRE. In addition, we detected several other intervening interactions. These correspond to DNaseI hypersensitive sites (data not shown), suggesting that these interactions may have a regulatory role.





**Figure 3-14: 4C-seq detected specific interactions at the *CD34* locus**

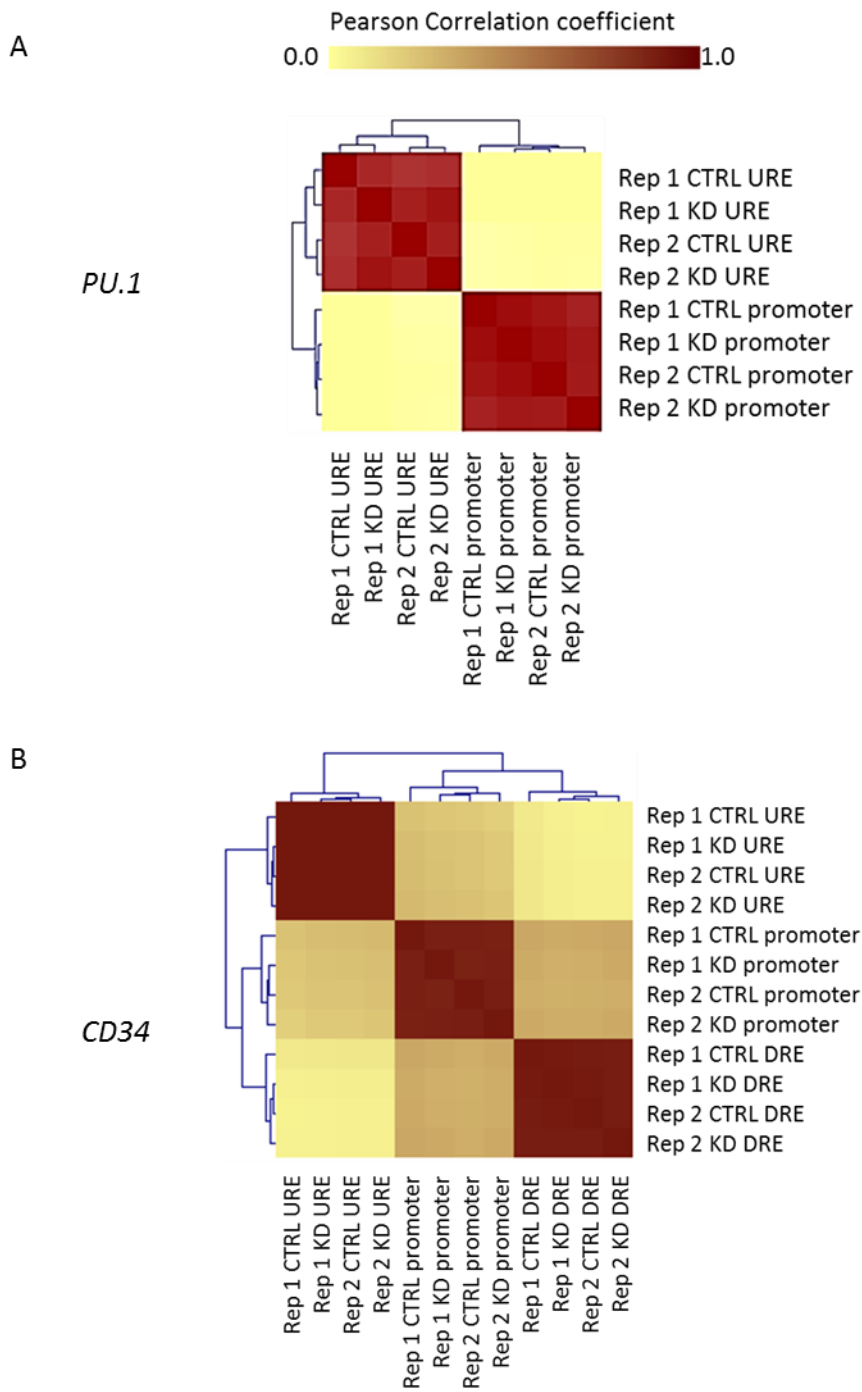
Local contact profile analysis with a viewpoint from the *CD34* promoter (A), *CD34* upstream regulatory element (URE) (B) and *CD34* downstream regulatory element (DRE) (C) using chromatin from Kasumi-1 cells transfected with mismatch siRNA (siMM) RUNX1/ETO specific siRNA (siRE). In the top panel (main trend), the contact intensity (black line) is calculated using a running median analysis of normalised read counts with a 3 kb sliding window. The 20th and 80th percentile are visualised as a grey trend graph. In the bottom panel, contact intensities are computed using linearly increasing sliding windows (scaled 2–50 kb) and displayed as a colour-coded heatmap of positive 4C signal (maximum interaction set to 1). Local colour changes are log-scaled to indicate changes of statistical enrichment of captured sequences, corresponding to the enhancer – promoter interaction. The results presented here are an average of two biological replicates.

### **3.4.3 4C-seq data was highly reproducible**

Hierarchical clustering on pairwise correlation coefficients was used to determine the correlation between 4C-seq data from all samples (figure 3-15). The heat map demonstrates a strong correlation between biological replicates. This indicates the experiment was highly reproducible. There is also a strong correlation between samples of the same viewpoint but transfected with a different siRNA. There is no correlation between samples with different viewpoints.

### **3.4.4 Genomic proximity cannot always be used as a predictor of enhancer function**

It is often assumed that *cis*-elements will preferentially target the nearest gene. However, Figure 3-13 shows that the *SPI1* URE forms an interaction with the *SPI1* promoter, rather than a gene 3 Kb closer (*SLC9A13*). A strong interaction is detected between the *SPI1* promoter and -14kb URE, but not between the -14kb URE and *SLC39A13* promoter. This is despite the open chromatin conformation of *SLC9A13* and the fact that its promoter is involved in an interaction with the *SPI1* promoter.



**Figure 3-15: 4C-seq data was highly reproducible**

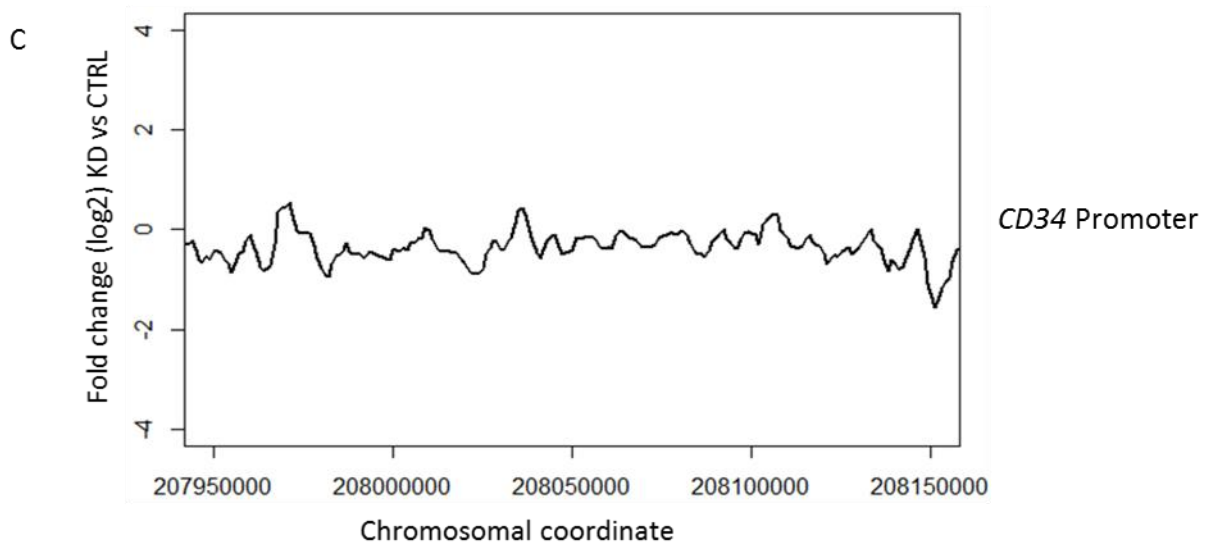
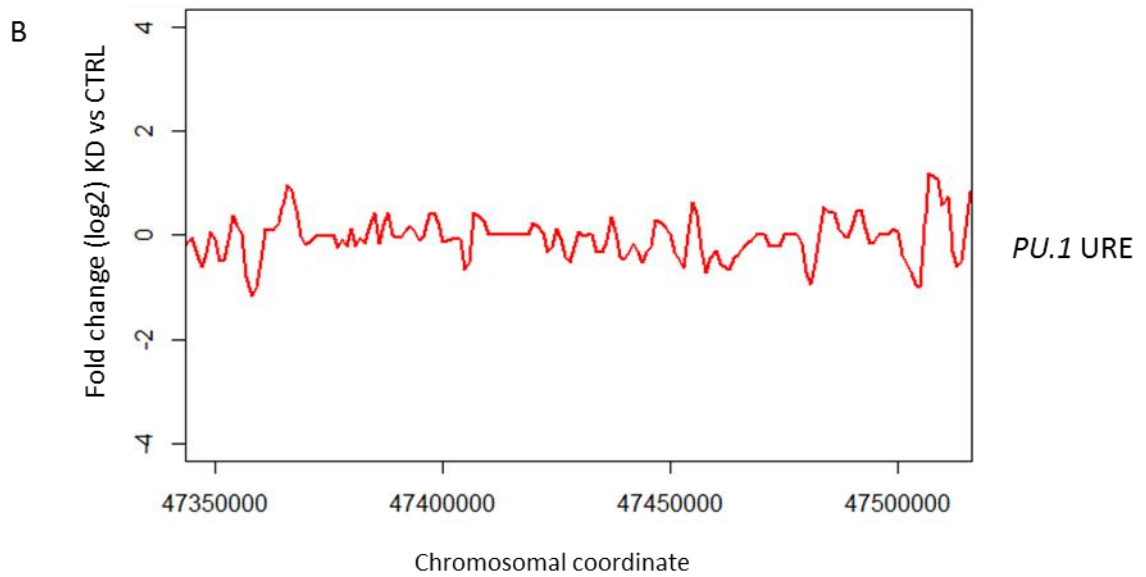
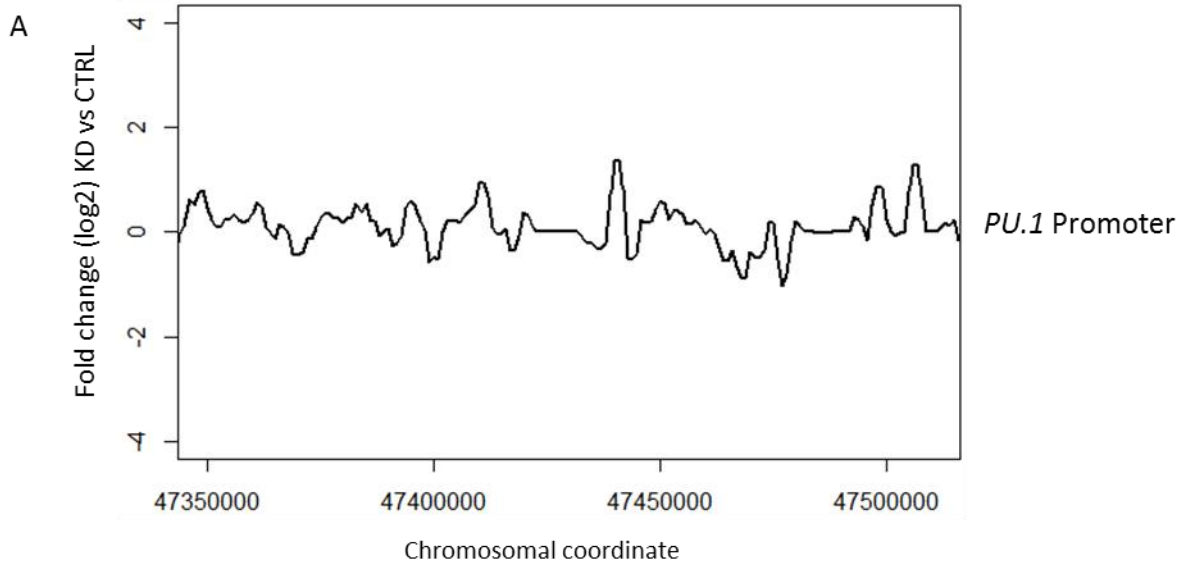
Hierarchical clustering on pairwise correlation coefficients of 4C-seq data from all samples with viewpoints at the A) *PU.1* locus and B) *CD34* locus. The heat map demonstrates a strong correlation between biological replicates. There is also a strong correlation between samples of the same viewpoint but transfected with a different siRNA. There is no correlation between samples with different viewpoints.

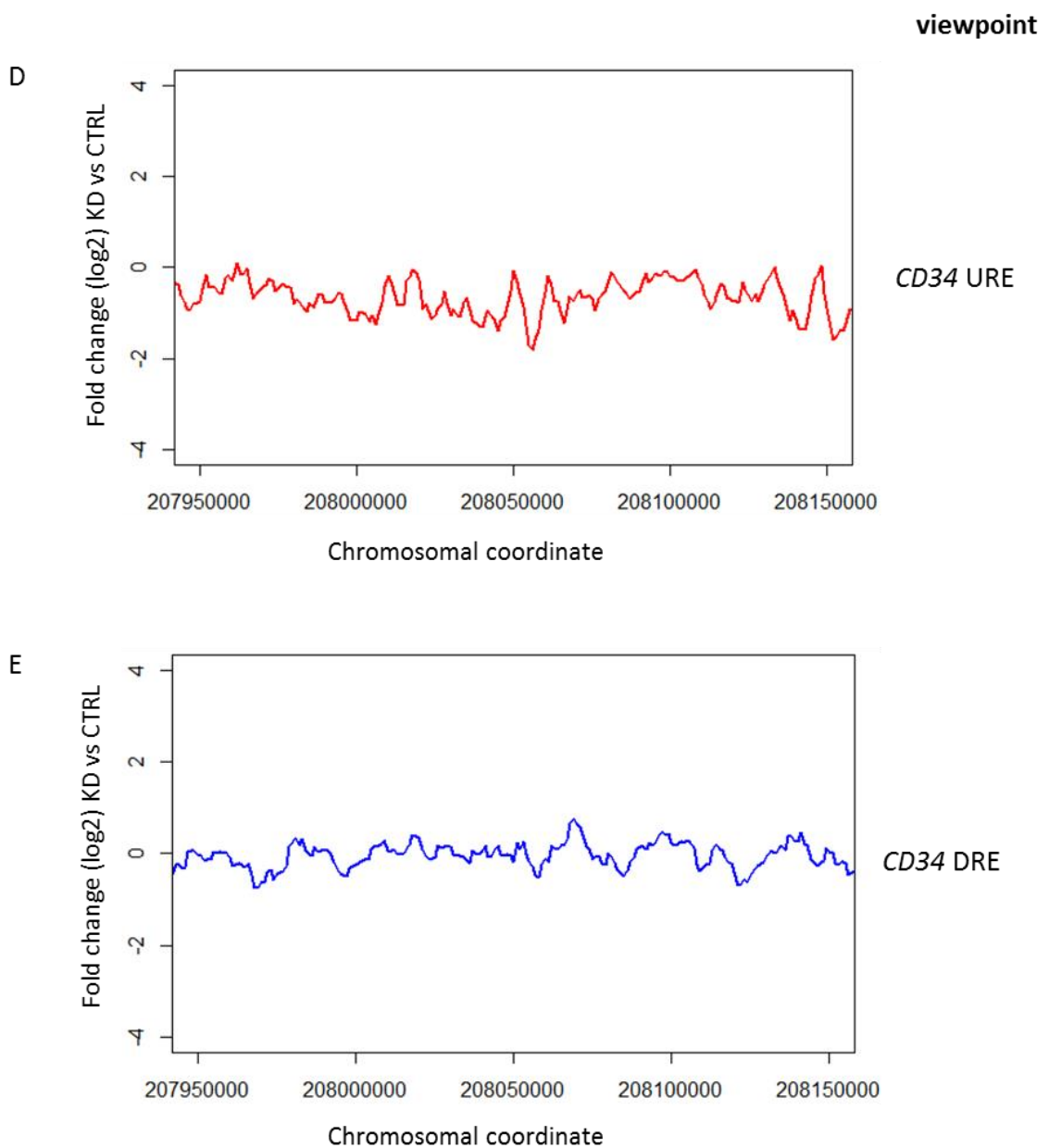
### **3.4.5 RUNX1/ETO depletion has no significant effect on the interactions between cis-regulatory elements at *SPI1* and *CD34* loci**

We sought to determine whether RUNX1/ETO knockdown had an effect on the interactions detected in the *SPI1* and *CD34* loci. Differential analysis was conducted with the R package 'DESeq2', to give a log<sub>2</sub> fold change between the RUNX1/ETO knockdown and control data, at 1kb resolution. The log<sub>2</sub> fold change for each genomic co-ordinate was then plotted (Figure 3-16). We saw no significant changes in interaction frequencies at the regions surrounding the viewpoint.

In conclusion, although RUNX1 binding site elimination inhibits promoter-enhancer interactions at the PU.1 and CD34 loci (63, 295), RUNX1/ETO knockdown had no significant effect on the interactions taking place at these loci. This is despite the fact these genes are differentially expressed after RUNX1/ETO knockdown.

Viewpoint





**Figure 3-16: Differential analysis revealed no significant difference in interaction frequency after RUNX1/ETO depletion**

R package DESeq2 was used to calculate the log<sub>2</sub> fold change between the median interaction frequency at each restriction fragment of the control (CTRL) and RUNX1/ETO knockdown (KD) 4C-seq data, with viewpoints A) PU.1 promoter B) PU.1 URE C) CD34 promoter D) CD34 URE and E) CD34 DRE. The y axis represents the log<sub>2</sub> fold change (RUNX1/ETO knockdown vs control) and the x-axis shows the genomic coordinates surrounding the viewpoint. A fold change >2 was not detected at any restriction fragment of the genomic region analysed.

### **3.5 The effect of RUNX1/ETO knockdown on genome wide *cis*-regulatory element interactions**

RUNX1/ETO knockdown had no effect on the DNA interactions taking place at the PU.1 and CD34 loci, despite the differential expression of these genes after knockdown. However we could not conclude that RUNX1/ETO had no influence on DNA interactions as we only interrogated two regions. There are hundreds of genes which alter their expression after knockdown; perhaps differences would be seen here. Preparing independent 4C libraries to assess all of these genes would be prohibitively expensive, as well as time consuming. Furthermore, this technique is limited by the ability to design specific and efficient primers to the region of interest.

We therefore decided to use a genome wide approach so that we could look at thousands of interactions in just one experiment. The original genome wide interaction mapping method is called HiC. This technique is a true genome wide approach and the experimenter can therefore look at all ligation junctions. However, due to the massively complex nature of the resultant library, the level of signal for any given point in the genome is very low. This requires the use of an infrequently cutting enzyme and large windows of analysis, so that signal can be pooled in order to gain informative signal. As a result, resolution is low and this method is more suitable for mapping large scale domain structures of the genome, rather than detailed, specific promoter to cis-element interactions.

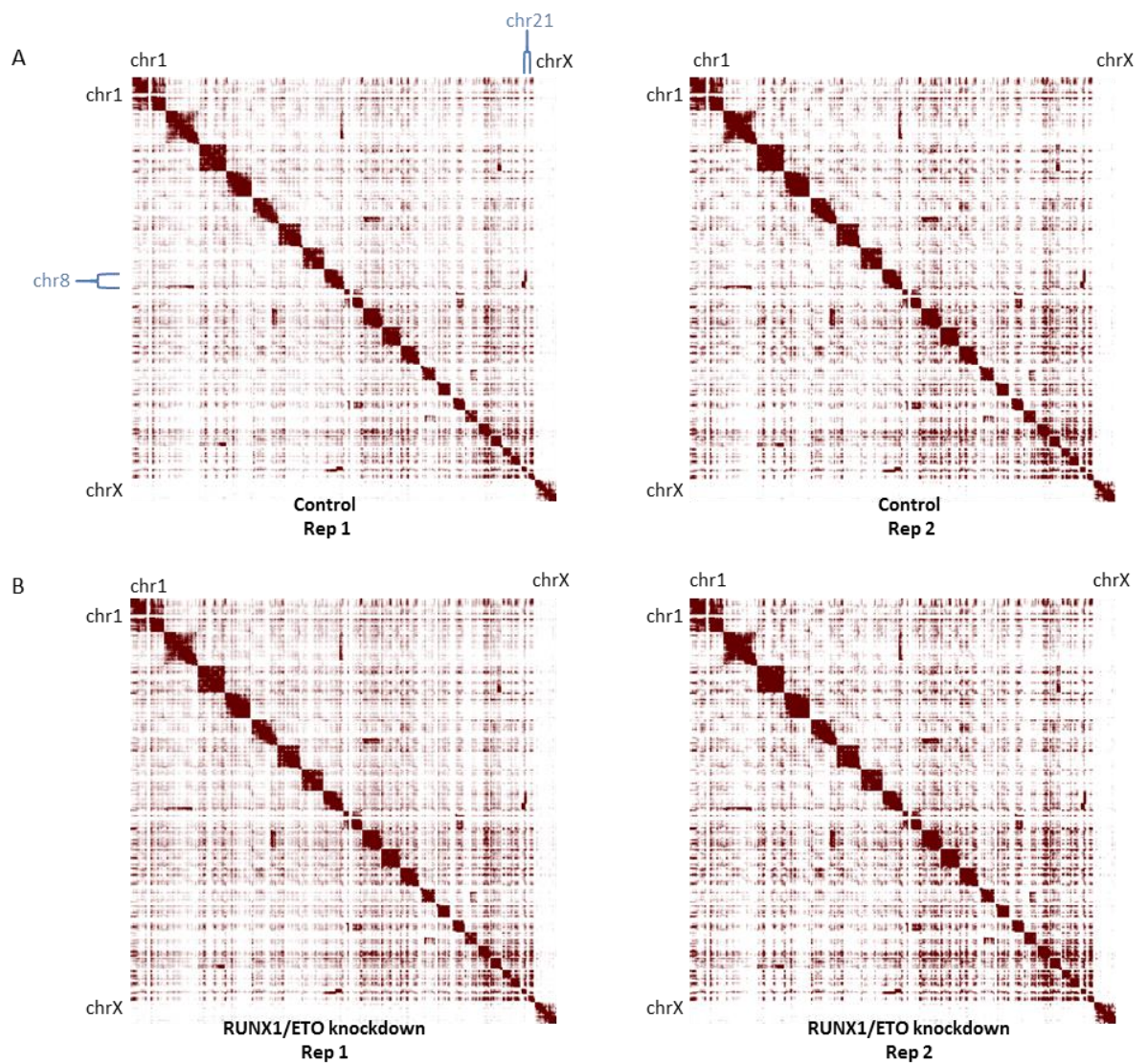
To circumvent this issue a technique was developed called Capture HiC (77, 141). In this method, traditional HiC library preparation is combined with solution hybridization selection to enrich HiC libraries for the long-range contacts of the 22,000 promoters to which the hybridization array is designed. The required breadth of sequencing coverage is therefore reduced, so that the resolution and signal increases, enabling us to detect specific DNA interactions.

### **3.5.1 The majority of DNA interactions in Kasumi-1 cells are intrachromosomal**

To map the genome wide DNA interactions we performed Capture HiC with the t(8;21) AML cell line, Kasumi-1. We wished to determine whether RUNX1/ETO had an effect on DNA interactions, so libraries were prepared using cells with and without 4 day siRNA mediated RUNX1/ETO knockdown. Two biological replicate samples were prepared.

As a means of quality controlling our HiC data, we first looked for well-established features of genome organisation, such as chromosome territories (131). To do this we plotted a matrix of genome wide interactions at a 10 Mb resolution (figure 3-17). We found that the majority of chromosomal interactions occurred in *cis* i.e. within the same chromosome. This is clearly illustrated by the blocks of high contact intensity centred along the diagonal of the heat map. This trend should always occur, regardless of cell type or species (92, 99). The depiction of chromosome territories offered reassurance that the experiment was successful. Figure 3-17 also demonstrated the reproducibility of the Capture HiC assay. The genome wide interactions matrices from biological replicate 1 and biological replicate 2 were almost identical.





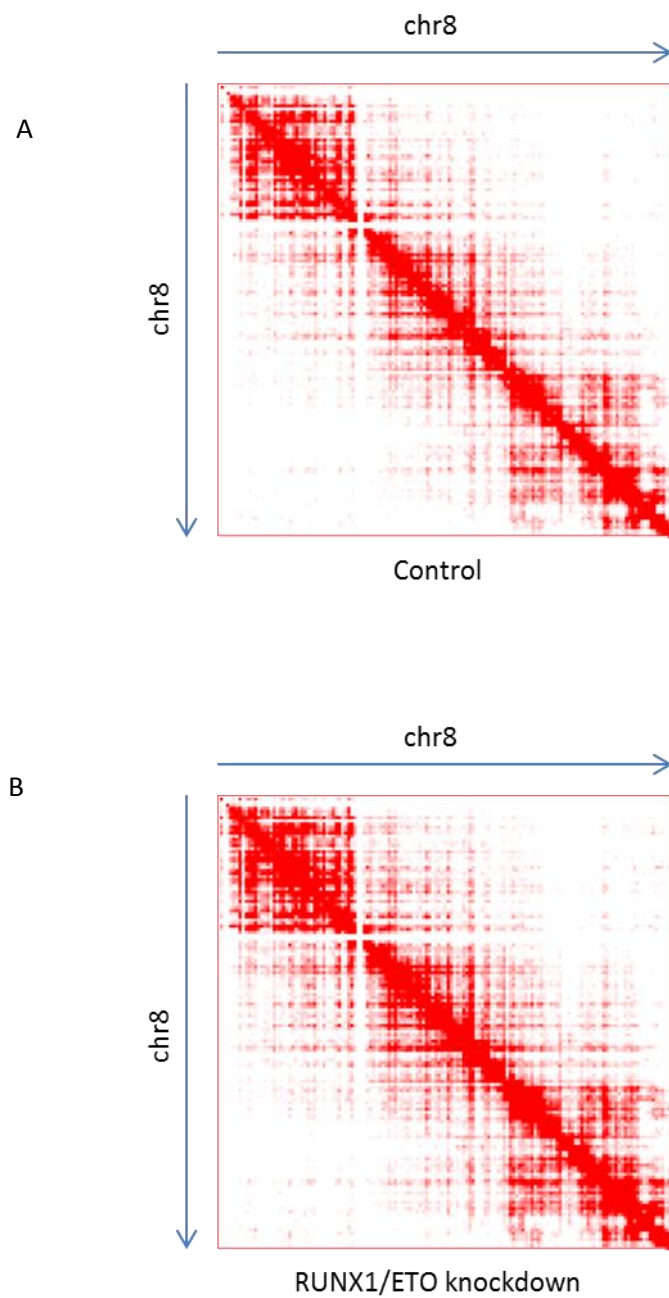
**Figure 3-17: 10 Mb resolution genome wide contact matrix showed the majority of DNA interactions are intrachromosomal**

Contact matrix across the whole genome. Each pixel represents a 10 Mb section of the genome. Colour intensity represents interaction frequency. The blocks of high contact intensity centred along the diagonal demonstrate the fact the majority of interactions occur intrachromosomally. The t(8;21) chromosomal translocation is detected. A) Interaction matrix generated with data from kasumi-1 cells transfected with siMM for four days. B) Interaction matrix generated with data from Kasumi-1 cells transfected with siRE for four days. Left hand plots are from biological replicate 1 and right hand plots are from biological replicate 2. Interaction data was generated by Capture HiC.

To further validate our data, we looked for evidence of the t(8;21) translocation in our interaction matrix. The ability to detect translocations with chromosome conformation capture analysis has been demonstrated previously (296, 297). Encouragingly, we saw a block of high contact intensity representing apparent interactions between chromosomes 8 and 21, which signifies the t(8;21) translocation.

### **3.5.2 Interaction frequency decreases with genomic distance**

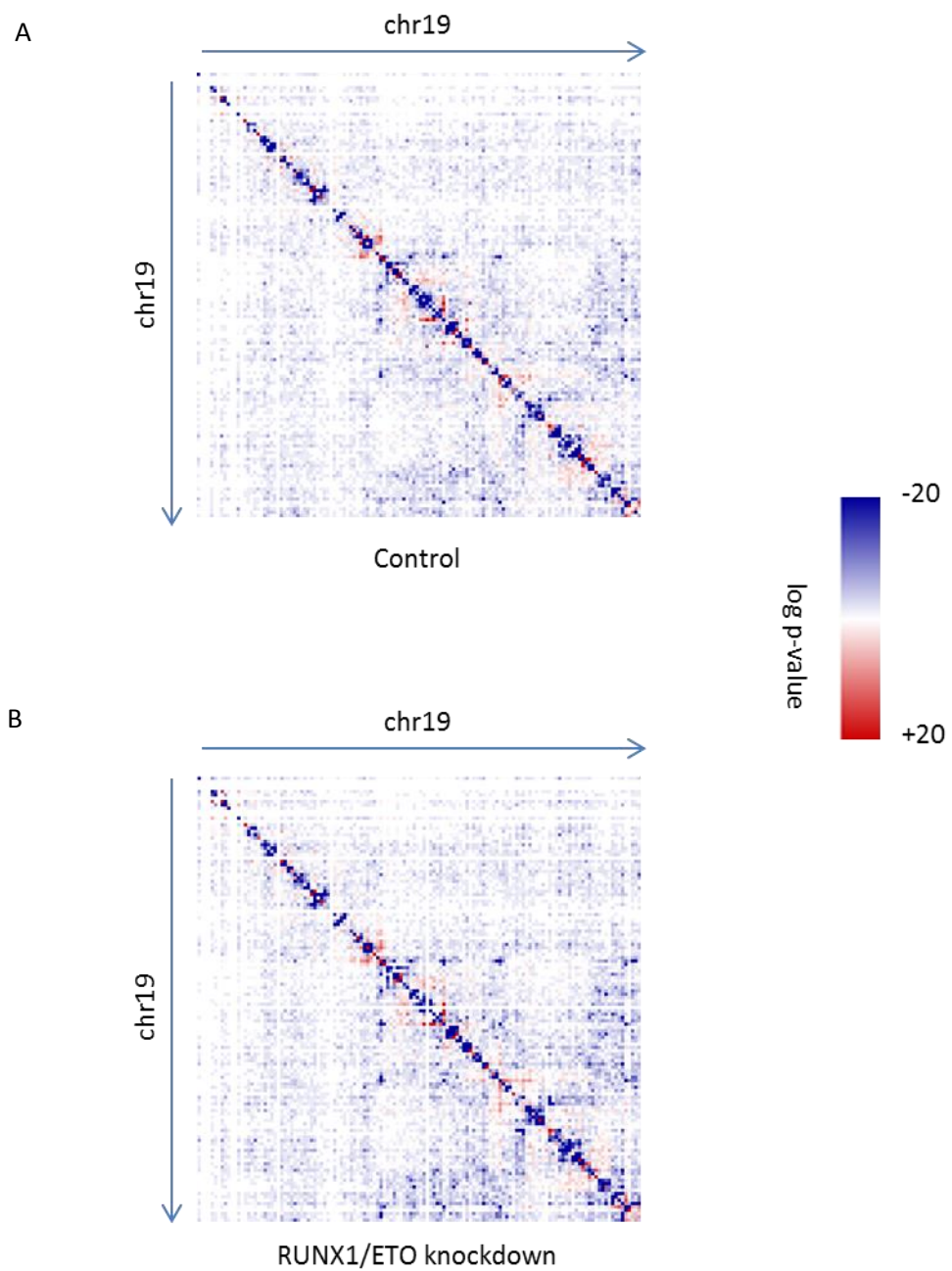
To generate figure 3-18 we focused on a single chromosome, chromosome 8, at higher resolution (1 Mb). Again we saw a strong diagonal of high contact intensity. This represents the well-established notion that the probability of DNA interaction decreases exponentially with genomic distance (92, 134). Based on this concept, a plot of expected contact intensity would decay to almost no signal after a relatively short distance. However, in the matrix constructed with our data, we can see regions of significant interaction at large genomic distances. This suggests that we have detected specific long range interactions.



**Figure 3-18: The 1 Mb resolution contact matrix of chromosome 8 showed a trend of decreased interaction frequency with genomic distance**

Contact matrix across chromosome 8, depicted in a heat map. Each pixel represents a 1 Mb section of the genome. Colour intensity represents interaction frequency. A) Interaction matrix generated with data from Kasumi-1 cells transfected with siMM for four days. B) Interaction matrix generated with data from Kasumi-1 cells transfected with siRE for four days. Interaction data was generated by Capture HiC. Data from replicate 1 one is shown here.

In summary, our promoter-capture HiC assays reproducibly detected interactions and they follow the expected trend. At the resolution of analysis used thus far it was not possible to define specific promoter-enhancer interactions; a higher resolution of analysis is required. Before doing so, it was necessary to determine which interactions are statistically significant. To this end, HOMER software was used to determine statistically significant interactions, taking into account the p-value and false discovery rate (FDR) relative to a background model. The heat map present in figure 3-19 shows the results of statistical analyses of all interactions taking place in chromosome 19.



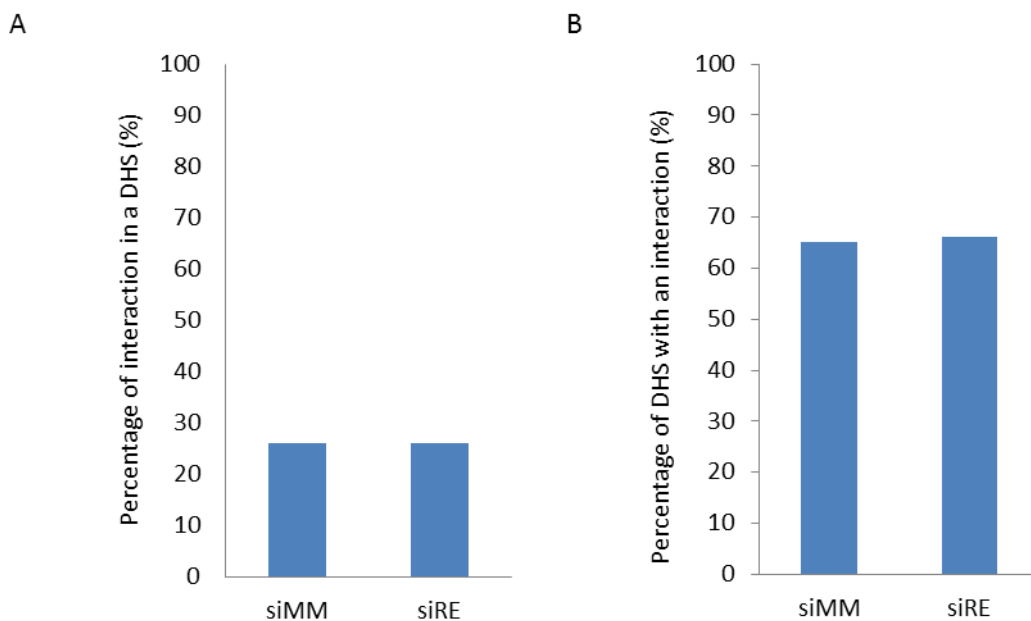
**Figure 3-19: Capture HiC interaction data was analysed to extract statistically significant interactions.** Contact matrix across chromosome 19, depicted in a heat map. Each pixel represents a 50 kb section of the genome. The colour represents the p-value of the interaction. The more blue the interaction, the more statistically significant it is and the redder, the less statistically significant. A) Interaction matrix made with data from Kasumi-1 cells transfected with siMM for four days. B) Interaction matrix made with data from Kasumi-1 cells transfected with siRE for four days. Interaction data was generated by Capture HiC. To determine statistically significant interactions, reads from replicate 1 and 2 were merged.

### **3.5.3 RUNX1/ETO knockdown led to alterations in specific promoter-*cis* element interactions**

We next analysed our CHiC data to examine whether the knock-down of RUNX1/ETO led to changes in the interactions between *cis*-regulatory elements. At low resolution (50Kb), no obvious differences were visible between the genome wide interaction matrices of control and RUNX1/ETO depleted cells (Figure 3-17). This result suggests that RUNX1/ETO depletion has no effect on the overall large-scale genomic organisation in t(8;21) cells.

However, differences between cell types are more likely to lie in specific promoter-*cis* – regulatory element interactions (92). We therefore analysed our data at higher resolution to help us determine what was driving the gene expression changes triggered by RUNX1/ETO knockdown. We were particularly interested in identifying the transcription factors that could be driving changes in *cis*-regulatory element interactions. To this end, using HOMER bioinformatics software, we performed differential analysis between the Capture HiC interactions in Kasumi-1 cells with and without RUNX1/ETO knockdown. A p-value 0.1 was considered to be a significant differential interaction.

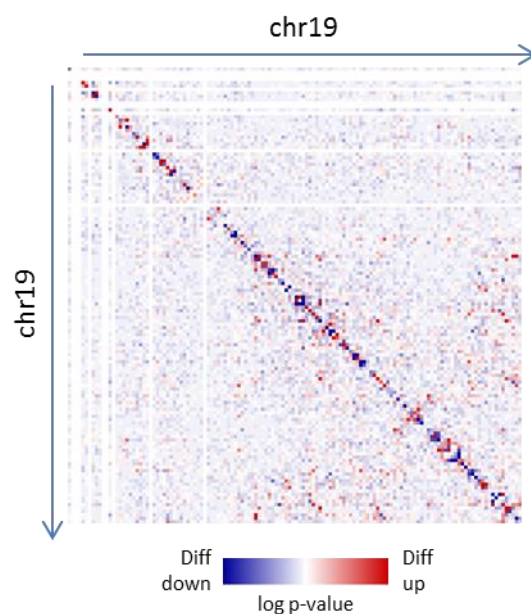
To identify transcription factor binding sites within interacting *cis*-elements, all significant interactions were filtered against our DHS data, such that only those ligated HindIII fragments overlapping with DHS were included in the analysis. This enabled us to direct our analysis to the *cis*-regulatory elements that were likely to be controlling gene expression. To this end, we used previously published DNaseI data from control and RUNX1/ETO depleted Kasumi-1 cells (219). Figure 3-20 A shows that approximately 30% of the significant interactions detected were DNaseI hypersensitive, confirming that many of them are of a structural nature (77, 99). In contrast, the majority of the DNaseI hypersensitive sites in the genome are taking part in Capture HiC interactions (Figure 3-20 B).



**Figure 3-20: Capture HiC interactions were filtered using DNaseI-seq data**

A) Percentage of Capture HiC interactions which are DNaseI hypersensitive B) Percentage of DNaseI hypersensitive sites involved in a Capture HiC interactions in Kasumi-1 cells transfected with mismatch (siMM) and RUNX1/ETO specific siRNA (siRE).

Figure 3-21 presents a comparison of all of the statistically significant interactions taking place in chromosome 19 between control and RUNX1/ETO depleted cells, with reduced interactions being plotted in blue and increased interactions plotted in red. This analysis shows that there were interactions which significantly change in intensity following RUNX1/ETO knockdown. We found that 917 interactions were significantly upregulated and 1457 were significantly downregulated after RUNX1/ETO knockdown. This result suggests that the removal of RUNX1/ETO has an effect on promoter-*cis* element interactions.

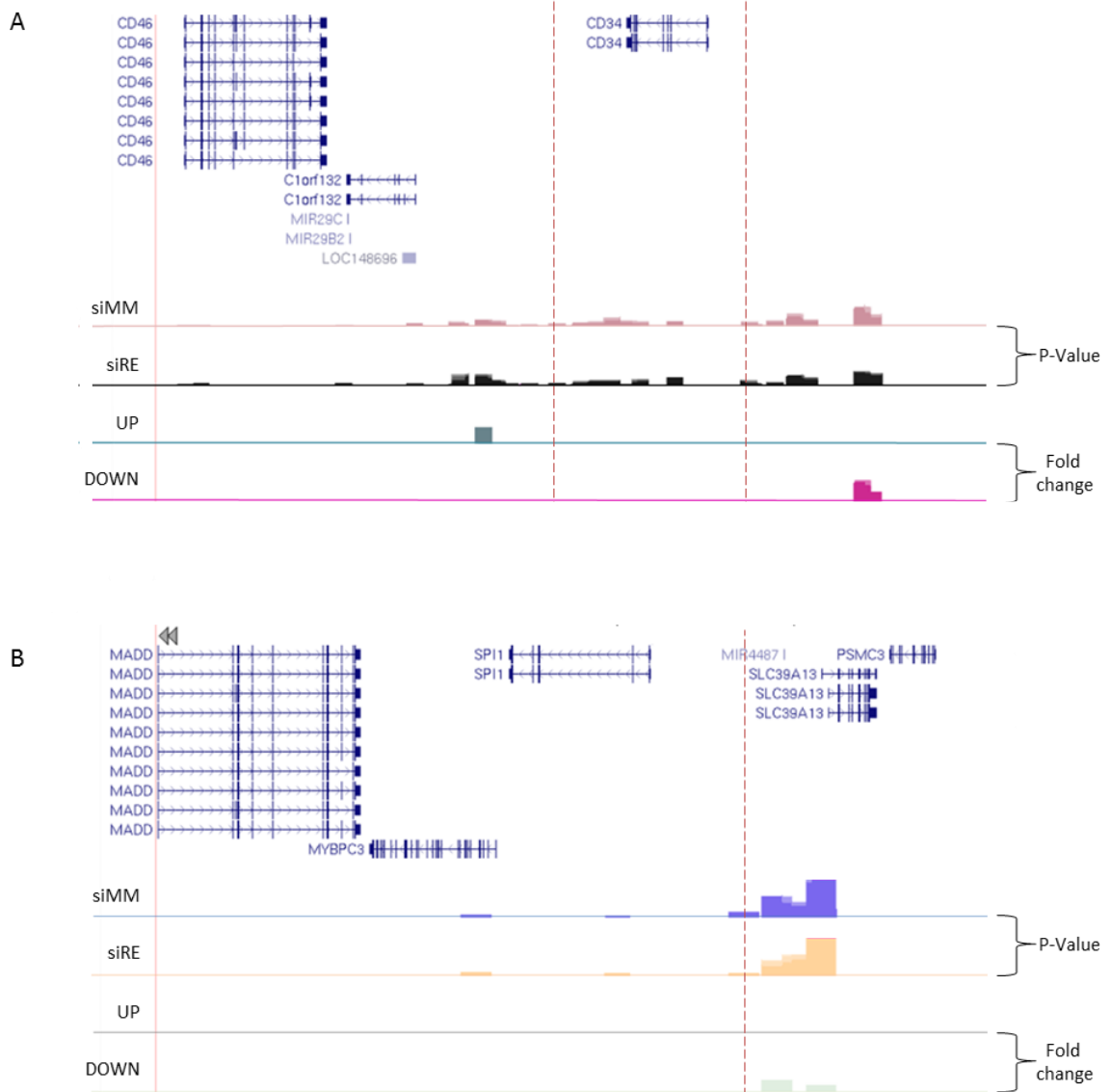


**Figure 3-21: RUNX1/ETO knockdown led to statistically significant differences in interaction strength.** This heat map presents the differential interactions in chromosome 19 as determined by Capture HiC. Each pixel represents a 50 kb section of the genome. The colour intensity represents the significance of the difference between the interactions strength in control and RUNX1/ETO knockdown ( $p$ -value). Blue interactions significantly decrease after RUNX1/ETO knockdown. Red interactions significantly increase after RUNX1/ETO knockdown.



#### **3.5.4 Capture HiC data support the 4C-seq data at the *CD34* and *SPI1* loci**

In order to validate our CHiC results, we extracted the interaction profiles from the *Spi1* (*Pu.1*) and *CD34* genes, visualised them on the UCSC genome browser and compared the results to our 4C data (Figure 3-22). This analysis revealed a significant interaction between the *CD34* promoter and both the known upstream and downstream regulatory elements. We also found that the *Spi.1/Pu.1* promoter was involved in a significant interaction with the already characterised -14kb enhancer region. These interactions did not change significantly after RUNX1/ETO knockdown, as demonstrated by the lack of a peak in the 'up' or 'down' track. Our 4C-seq experiments, which demonstrate that RUNX1/ETO knockdown has no effect on these interactions (figures 3-13 and 3-14), were therefore in concordance with the data from Capture HiC.

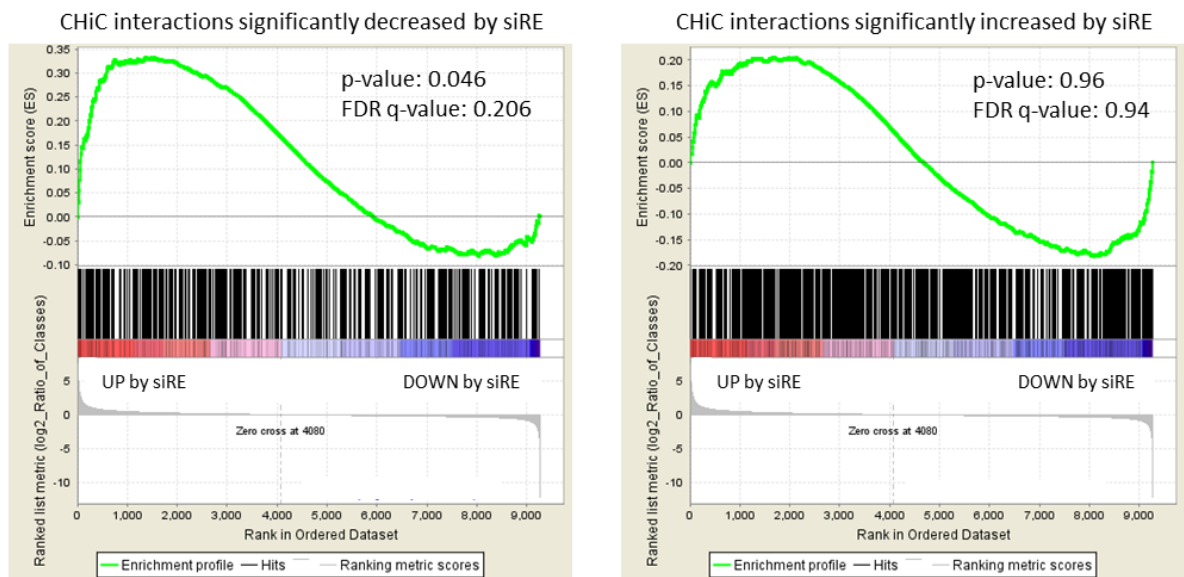


**Figure 3-22: Capture HiC data agreed with 4C data**

Tracks 1 and 2 show 5 kb sections of the genome taking part in a significant interaction with the A) *CD34* promoter or B) *PU.1* promoter, in cells transfected for 4 days with siMM and siRE respectively. The height of the block presents the  $-\log p$ -value. Tracks 3 and 4 mark interactions which change significantly after RUNX1/ETO knockdown. The height of the blocks represents the fold change. A) Red dashed lines mark the upstream and downstream regulatory elements. B) Red dashed line represents the -14 kb enhancer.

### **3.5.5 The gene expression changes after RUNX1/ETO knockdown are associated with alterations in cis-regulatory element interactions but are not directly correlated**

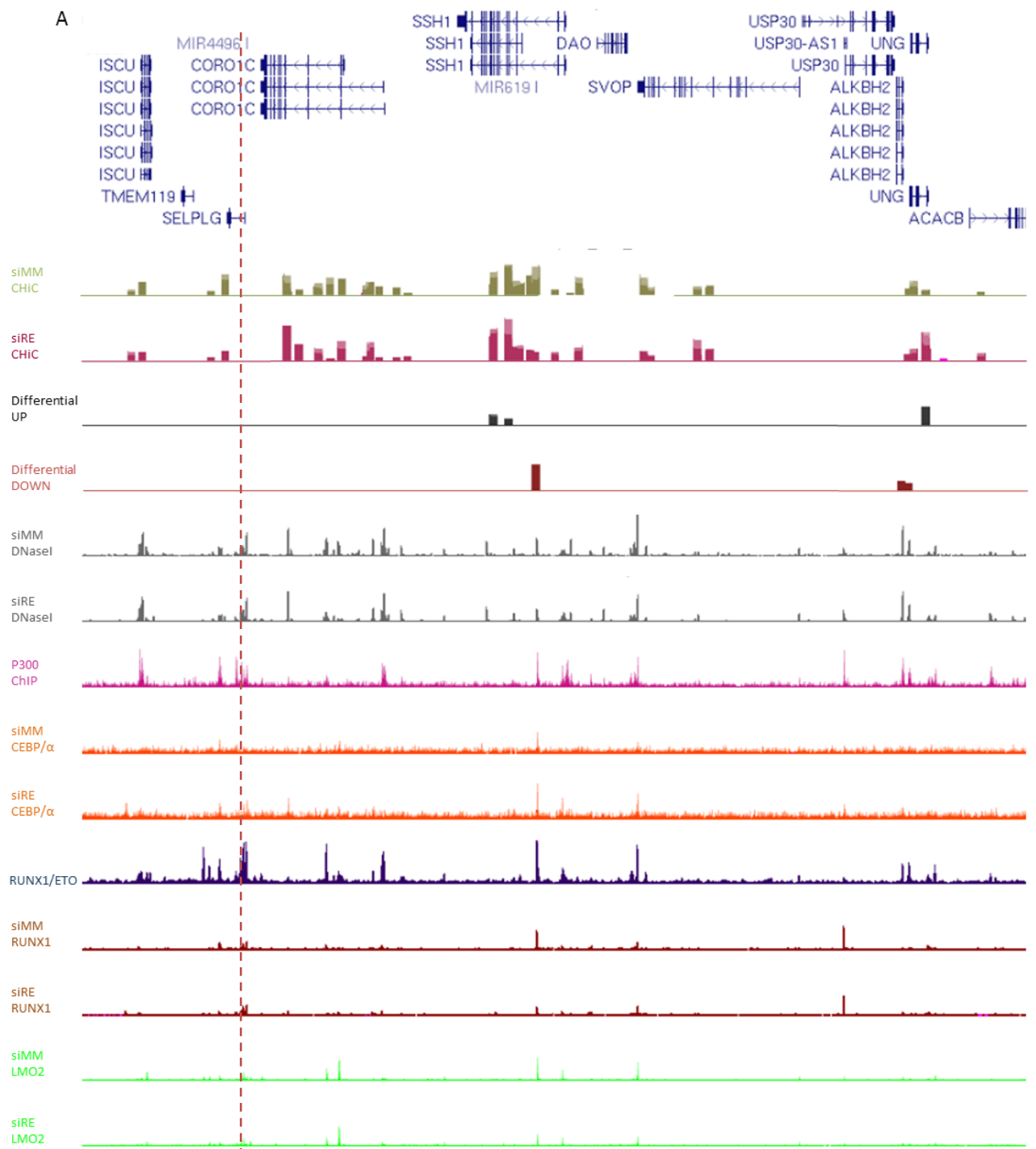
Alterations in gene expression can be reflected by changes in specific *cis*-regulatory element interactions (63, 118). This was true for gene expression changes after RUNX1/ETO knockdown. To examine whether differential interactions were directly correlated with the changes in gene expression before and after RUNX1/ETO knock-down, i.e. whether reduced interactions led to a loss and increased interaction led to an increase in gene expression, we conducted a Gene Set Enrichment analysis (GSEA). Figure 3-23 shows that there was no direct correlation, meaning that genes which were upregulated after RUNX1/ETO knockdown were associated with both increased and decreased DNA *cis*-regulatory element interactions. The same was true for down regulated genes, indicating a complex relationship between gene expression control and *cis*-regulatory element interactions.

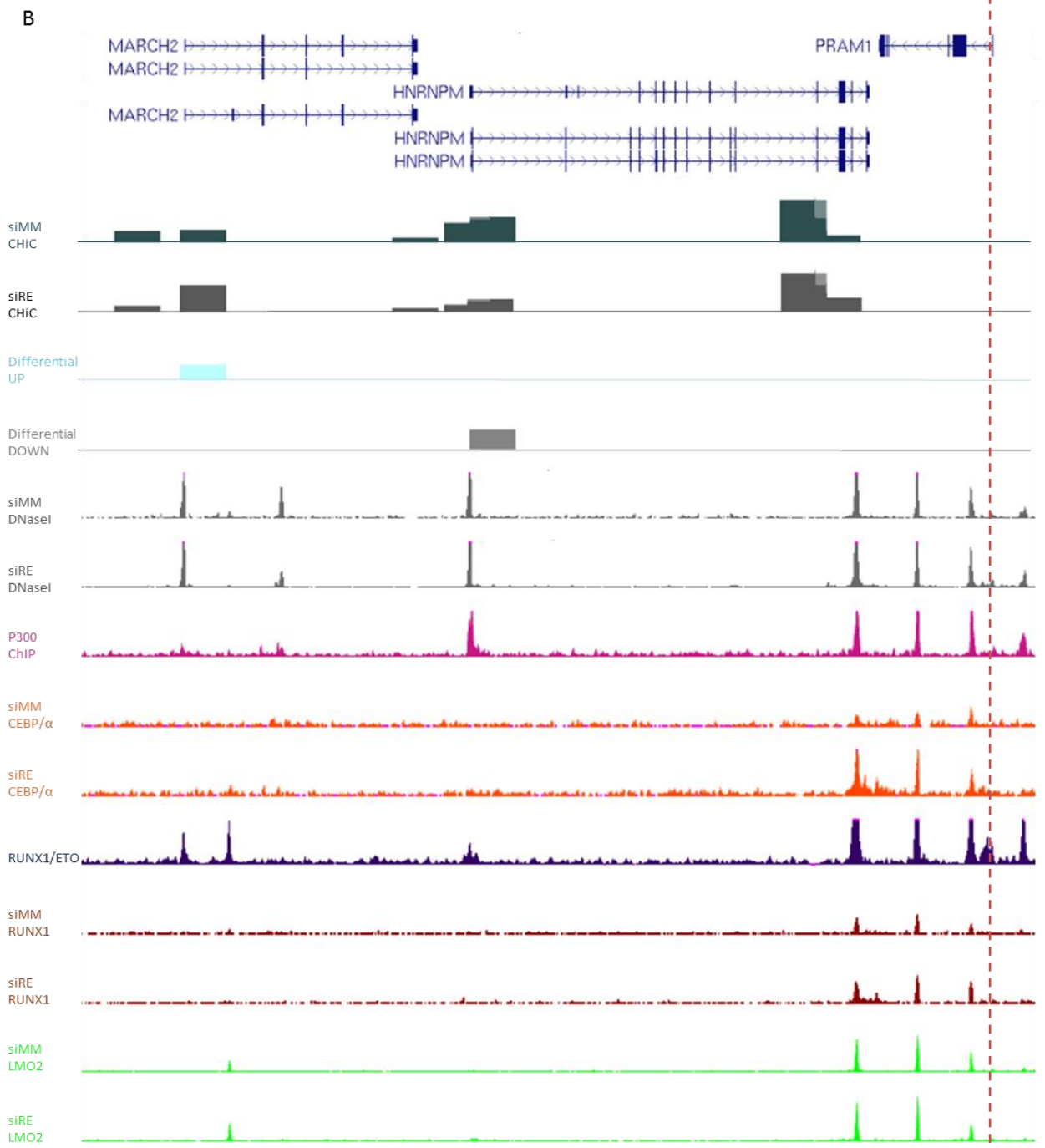


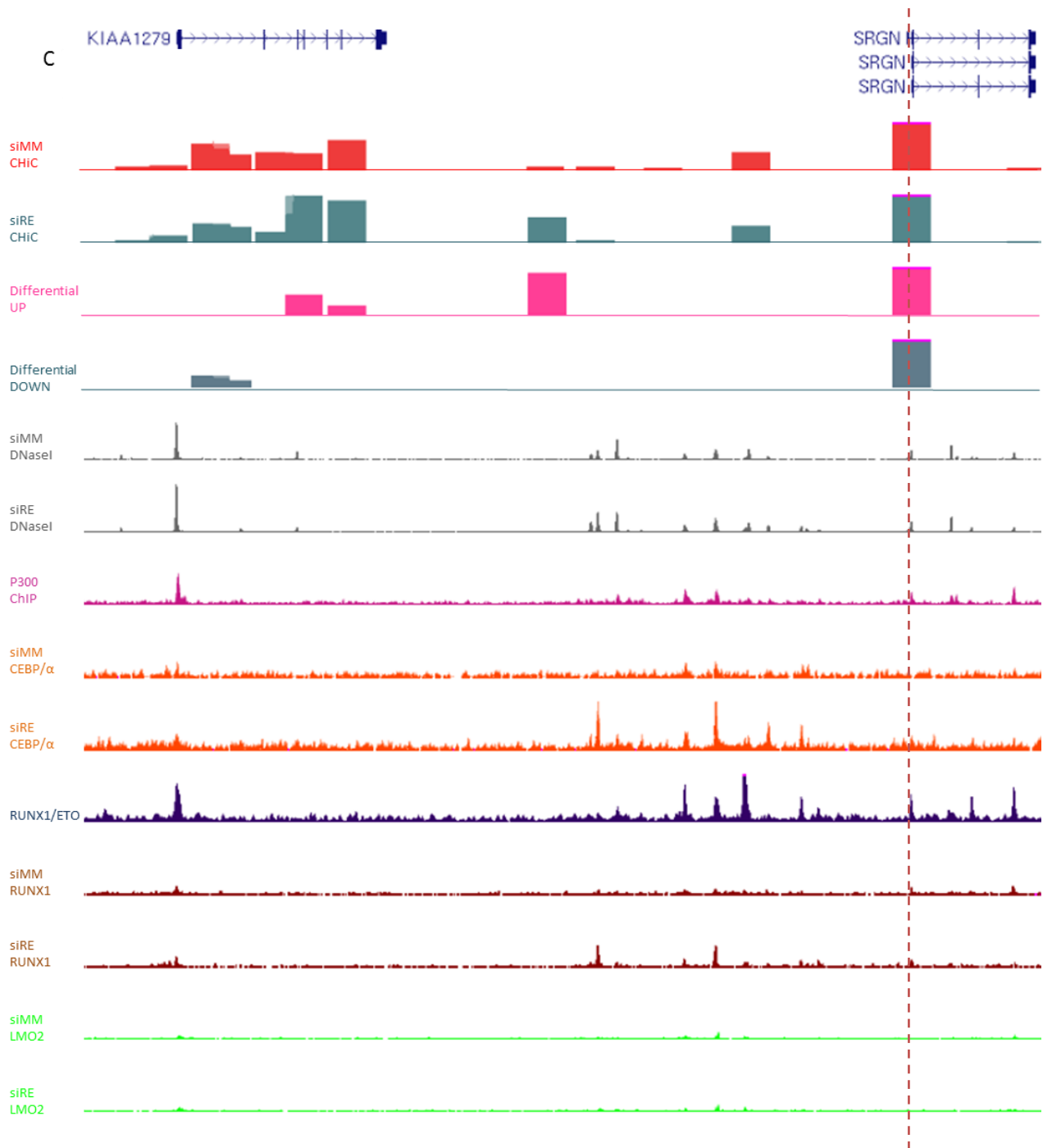
**Figure 3-23: Differential interactions were not directly correlated with differential gene expression.** Gene Set Enrichment Analysis (GSEA) based on interactions differentially increasing (left) or decreasing (right) by RUNX1/ETO knockdown. The upper section of the plot (green line) indicates the running enrichment score for the genes corresponding to the differential interaction in the ranked list of genes below (gene differentially expressed by RUNX1/ETO knockdown). The middle section of the plots (black vertical lines) indicates where genes corresponding to the differential interaction appear in the ranked list of genes. The lower portion is the ranking metric of the ranked gene list (gene differentially expressed siMM vs siRE). The siMM vs siRE gene expression data was obtained via RNA-seq.

The most likely reason for this observation is that, within a given gene locus, there is not a binary on/off interaction with a single enhancer, rather expression control involves multiple *cis*-regulatory elements with net positive or negative effects. After RUNX1/ETO knockdown gene promoters alter the *cis*-regulatory elements they interact with; increasing interactions with some elements and decreasing interactions with others. This was the case at several differentially expressed genes. Figure 3-24 presents examples of significantly upregulated haematopoietic genes (*PRAM1*, *SELPG* and *SRGN*) which have both increasing and decreasing interactions with their promoters. The same is true for the down regulated stem cell gene *CD34* (figure 3-22 A). *PRAM1* encodes PML-RAR Regulated Adaptor protein which is associated with granulocytic maturation (298). *SELPLG* encodes a glycoprotein expressed on the surface of myeloid cells (299). *SRGN* encodes a proteoglycan granule protein and is primarily expressed in haematopoietic cells (300).

To visualise interactions between *cis*-regulatory elements of specific genes, we uploaded the Capture HiC interaction data to the UCSC genome browser together with other genome-wide data sets generated by Dr Anetta Ptasińska (199, 219). For example, in figure 3-24 the Capture HiC interactions taking place with the *PRAM1*, *SELPG* and *SRGN* promoter are presented along with RUNX1/ETO, C/EBP $\alpha$  and RNA Polymerase II ChIP-seq data. In addition, DNaseI-seq data is presented.











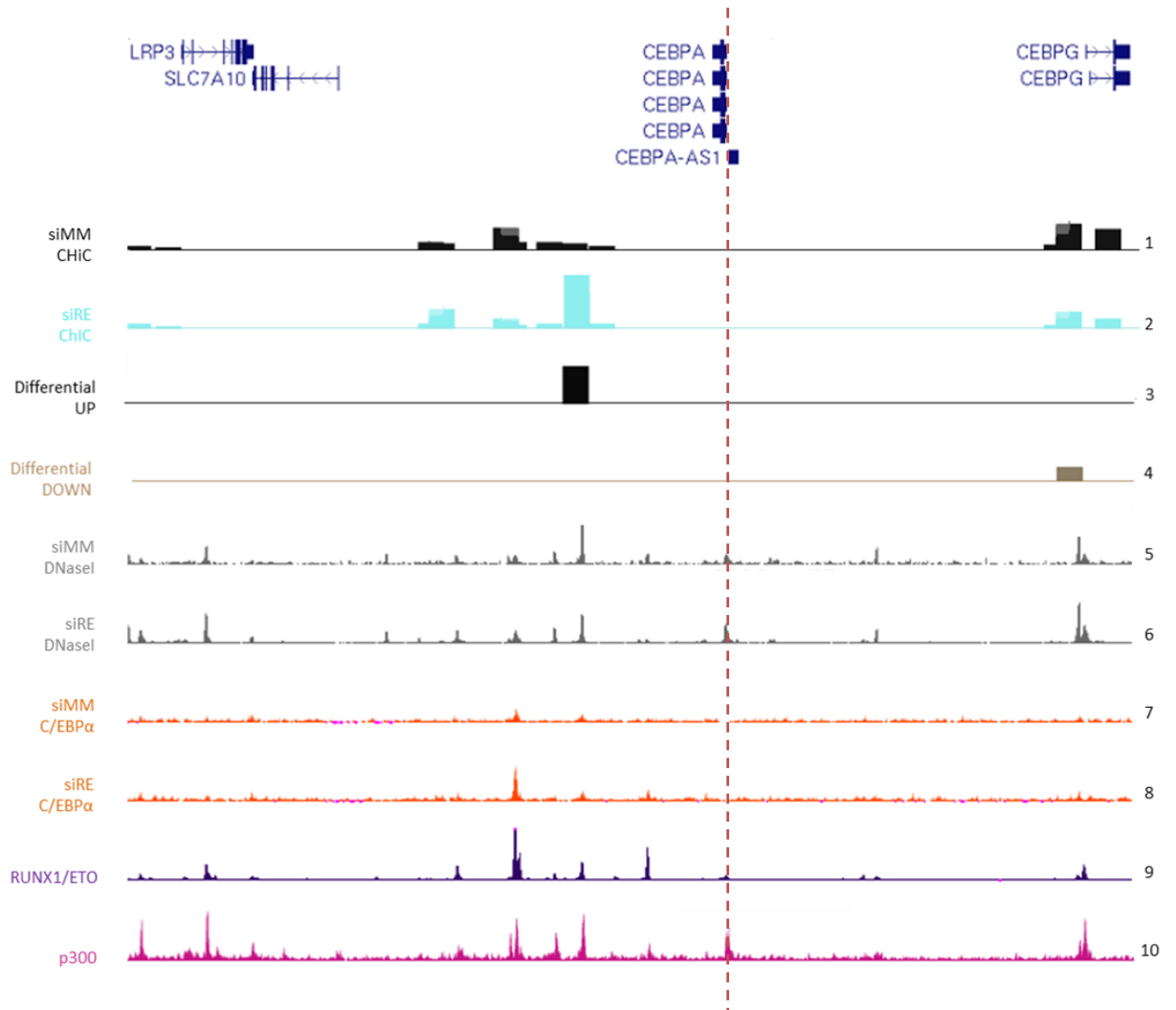
**Figure 3-24: Differentially expressed haematopoietic gene promoters, such as *SELPLG*, *SRGN* and *PRAM1*, had both increased and decreased *cis*-element interactions after RUNX1/ETO knockdown**  
Dashed red line represents the ‘viewpoint’ of analysis i.e. the promoter from which the interactions are being assessed. Tracks 1 and 2 show 5 kb sections of the genome taking part in a significant interaction with the A) *SELPLG* B) *PRAM1* and C) *SRGN* promoters (genes significantly upregulated after RUNX1/ETO knockdown), in cells transfected for 4 days with siMM and siRE respectively and measured via Capture HiC. The height of the block presents the  $-\log$  p-value. Tracks 3 and 4 mark interactions which change significantly after RUNX1/ETO knockdown. The height of the blocks represents the fold change. Track 3 marks interactions that increase after RUNX1/ETO knockdown. Track 4 marks interactions with decrease after RUNX1/ETO knockdown. Tracks 5 and 6 show DNaseI-seq data from siMM and siRE cells respectively. Tracks 7-10 show the binding patterns of P300, C/EBP $\alpha$ , RUNX1/ETO, RUNX1 and LMO2 p300 in siMM and siRE cells. D) UCSC genome browser screen shots of RNA-SEQ data demonstrating the upregulation of, from top to bottom, *SELPLG*, *PRAM1* and *SRGN* after RUNX1/ETO knockdown. siMM data is from Kasumi-1 cells transfected with control siRNA for four days. siRE data is from Kasumi-1 cells transfected with RUNX1/ETO specific siRNA

### 3.5.6 RUNX1/ETO depletion led to differential interactions at the *CEBPA* locus

This study has convincingly demonstrated that maintaining low levels of C/EBP $\alpha$  expression is critical for the maintenance of t(8;21) leukaemia. We therefore examined the interactions made by the *CEBPA* promoter in detail to gain more insight into how C/EBP $\alpha$  expression levels were regulated. Figure 3-25 illustrates the local interactions made by the *CEBPA* promoter. Interestingly, this promoter interacts with the neighbouring downstream gene *CEBPG* which encodes for another member of the C/EBP family, and with several upstream hypersensitive sites. One of these interactions corresponds to the already defined +40 kb enhancer (200). The interaction strength with this enhancer was not significantly altered by RUNX1/ETO knockdown. However, an interaction at +29 kb significantly increased after RUNX1/ETO knockdown. This site is hypersensitive and enriched for p300 binding. It is also bound by both C/EBP $\alpha$  and RUNX1/ETO.

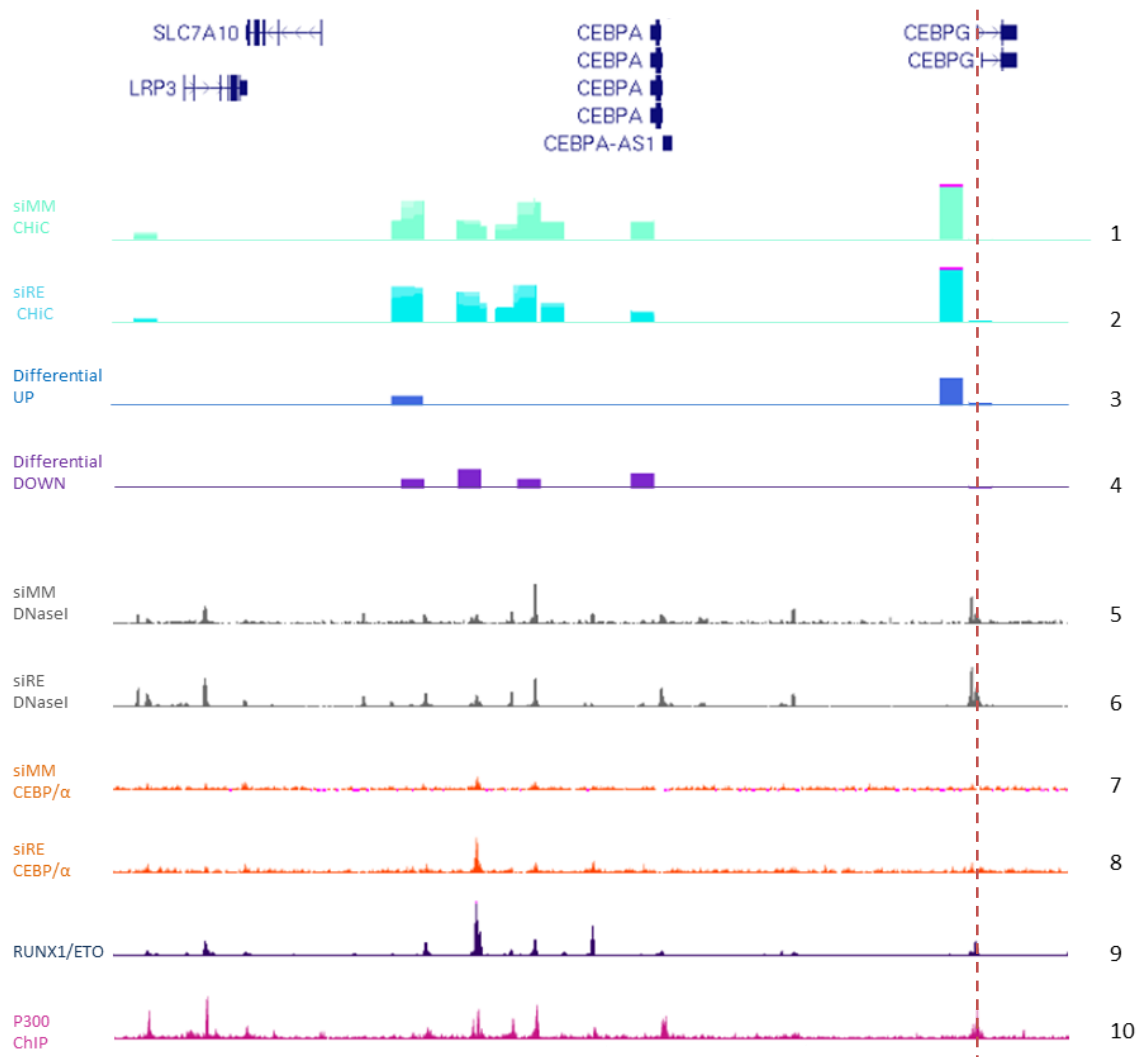
To further assess the reproducibility of our data, we checked to see whether the promoter-promoter interaction between the *CEBPA* promoter and the *CEBPG* promoter were detected when either promoter was used as the 'view-point'. Reassuringly this was the case. Figure 3-26 demonstrates that the interaction detected between the *CEBPA* promoter and the *CEBPG* promoter (figure 3-25) is detected with *CEBPG* promoter as the 'viewpoint'. In both cases this interaction decreases following RUNX1/ETO depletion.

In conclusion RUNX1/ETO knockdown had no effect on large-scale, domain structures of the genome. However, the presence of RUNX1/ETO does appear to influence the interaction profile of specific gene promoters.



**Figure 3-25: The C/EBPA promoter engaged in a strong interaction with a putative enhancer at +29 kb following RUNX1/ETO knockdown**

Dashed red line represents the ‘viewpoint’ of analysis i.e. the promoter from which the interactions are being assessed. Tracks 1 and 2 show 5 kb sections of the genome taking part in a significant interaction with the *C/EBPA* promoter, in cells transfected for 4 days with siMM and siRE respectively. The height of the block presents the  $-\log$  p-value. Tracks 3 and 4 mark interactions which change significantly after RUNX1/ETO knockdown. The height of the blocks represents the fold change. Tracks 5 and 6 show DNaseI hypersensitivity in siMM and siRE cells respectively. Tracks 7-10 show the binding patterns of C/EBP $\alpha$ , RUNX1/ETO and p300.



**Figure 3-26: The *CEBPG-CEBPA* promoter-promoter interaction was detected with either *CEBPA* or *CEBPG* promoters as the ‘viewpoint’**

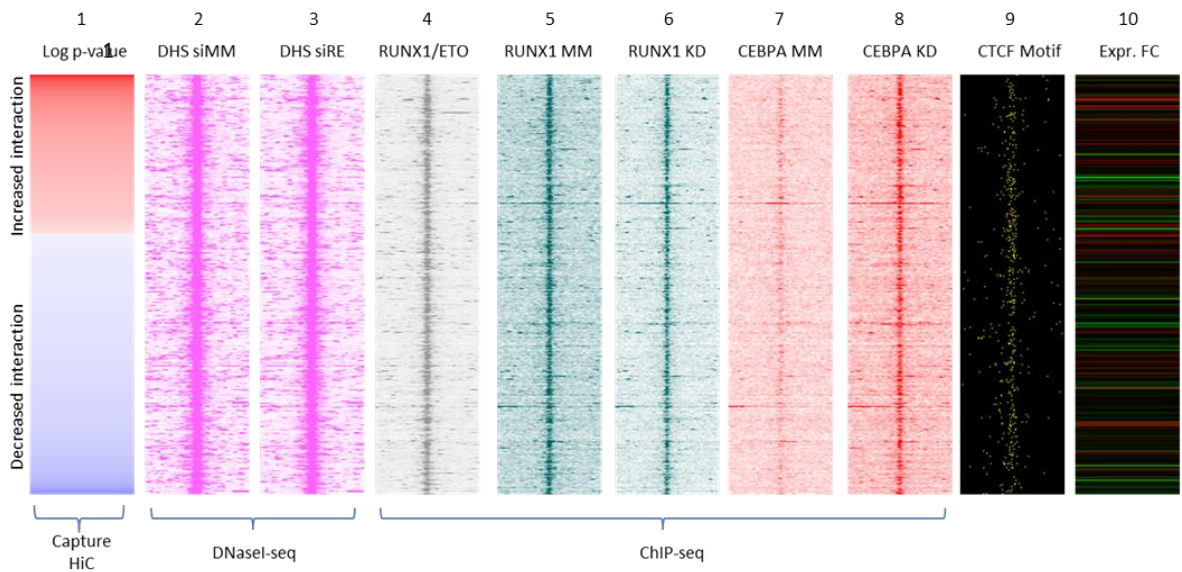
Dashed red line represents the ‘viewpoint’ of analysis i.e. the promoter from which the interactions are being assessed. Tracks 1 and 2 show 5 kb sections of the genome taking part in a significant interaction with the *CEBPG* promoter, in cells transfected for 4 days with siMM and siRE respectively. The height of the block presents the  $-\log p$ -value. Tracks 3 and 4 mark interactions which change significantly after RUNX1/ETO knockdown. The height of the blocks represents the fold change. Tracks 5 and 6 show DNaseI hypersensitivity in siMM and siRE cells respectively. Tracks 7-10 show the binding patterns of C/EBP $\alpha$ , RUNX1/ETO and p300.

### **3.5.7 C/EBP $\alpha$ and CTCF play a major role in mediating differential cis-regulatory element interactions before and after RUNX1/ETO knock-down**

To gain more insight into the mechanisms mediating the differential interactions, we investigated whether the change in interaction strength was associated with the level of DNaseI hypersensitivity, transcription factor occupancy and changes in gene expression. To this end, differential interactions were ranked by fold-change in p-value (figure 3-27). Fold change represents the difference in interaction strength between control and RUNX1/ETO knockdown. Associated DNaseI hypersensitive sites involved in differential interactions were plotted alongside. C/EBP $\alpha$ , RUNX1 and RUNX1/ETO ChIP-seq reads were then plotted around the summit of these DNaseI hypersensitive sites. In order to see whether CTCF was involved in changing interactions, we also examined whether its binding sites changed. We did not have CTCF ChIP-seq data, so CTCF motif enrichment was used instead. CTCF was included in this analyses as it has a principal role in the global organisation of chromatin architecture (301).

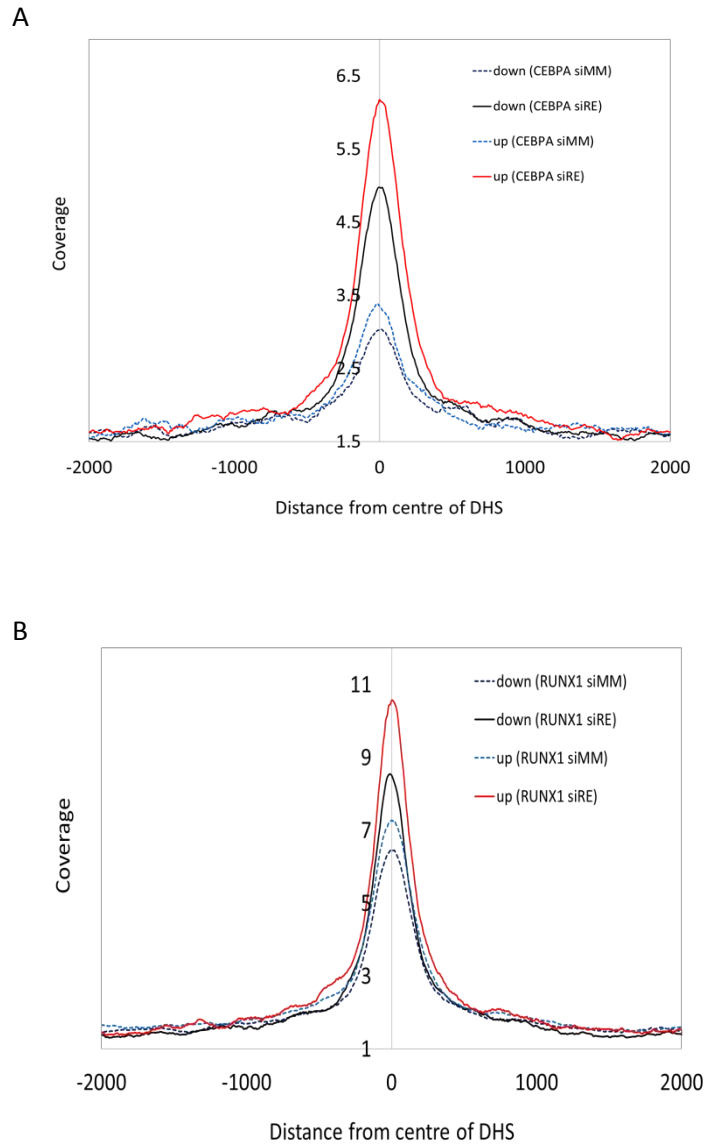
The resultant heat map, presented in figure 3-27, again supports the lack of direct correlation between the differential interaction fold change and gene expression change (figure 3-23). It also demonstrated that some genes with differential interactions are not differentially expressed (column 10). We also saw that differential interactions did not reproducibly correspond to alterations in the DNaseI hypersensitivity of the *cis*-regulatory elements involved (columns 2 and 3).

Figure 3-27 shows that the DHSs mediating the differential interactions were bound by RUNX1/ETO, RUNX1 and C/EBP $\alpha$ . The differential interactions exhibited an increase in C/EBP $\alpha$  binding. The interactions which went up after RUNX1/ETO knockdown appeared to have a greater increase in C/EBP $\alpha$  enrichment, relative to those interactions that went down. We plotted the average C/EBP $\alpha$  enrichment at all the differentially interacting DNaseI sites, to see whether this was indeed true. The average profile presented in figure 3-28 A offers a clear depiction of increased C/EBP $\alpha$  binding at differential interactions, particularly those that increased in interaction after RUNX1/ETO knockdown. This suggests that C/EBP $\alpha$  may play a role in driving the changes to promoter-*cis*-regulatory element interactions which follow RUNX1/ETO knockdown. The data suggest that it does so by either mediating interactions itself or by displacing other factors involved in interactions prior to knockdown. The average profile for RUNX1 enrichment, presented in figure 3-28 B, shows that that RUNX1 binding also increases at differential interactions, but to a lesser extent.



**Figure 3-27: Differential interactions were bound by RUNX1/ETO, RUNX1 and C/EBP $\alpha$ , and contained the CTCF motif**

The first column shows the fold change (control vs RUNX1/ETO knockdown) of Capture HiC interactions. Red represents an increase in interaction strength and blue represents a decrease. Columns 2 and 3 represent the DNase-seq reads from the corresponding interacting region. Column 4 shows RUNX1/ETO ChIP-seq reads from control Kasumi-1 cells aligned to the summit of the DHSs. Columns 5 and 6 show RUNX1 ChIP-seq reads from Kasumi-1 cells, transfected with siMM and siRE respectively. For columns 2 – 8 colour intensity corresponds to the number of reads. Columns 7 and 8 show C/EBP $\alpha$  ChIP-seq reads from Kasumi-1 cells, transfected with siMM and siRE respectively. Column 9 shows the occurrence of CTCF motifs in the interacting DNase-seq sites. Yellow dots represent motif enrichment. Column 10 shows gene expression fold change (siMM vs siRE) of the gene promoters which correspond to the differential interactions.












**Figure 3-28: Differential interactions exhibited increased C/EBP $\alpha$  and RUNX1 binding after RUNX1/ETO knockdown, particularly those that increased in interaction strength.**

A) This plot shows C/EBP $\alpha$  enrichment, before and after RUNX1/ETO knockdown, at DNaseI sites which are involved in a differential interaction. The x-axis represents the distance from the centre of the DNaseI site corresponding to a differential Capture HiC interaction. Peaks represent the coverage of C/EBP $\alpha$  ChIP-seq reads. Dashed lines show ChIP-seq reads from control kasumi-1 cells (transfected with siMM). Full lines show ChIP-seq reads from Kasumi-1 cells with RUNX1/ETO knockdown (transfected with siRE). ChIP-seq data was generated by Dr Anetta Ptasińska (219)

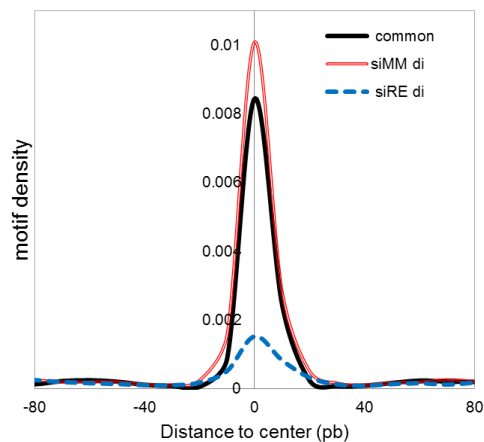
B) This plot is the generated by the same method as (A) but using RUNX1 enrichment obtained via RUNX1 ChIP-seq.



We found that the CTCF motif was present in the differentially interacting DNaseI hypersensitive sites (figure 3-27). We therefore conducted a *de novo* motif search in these sites, to verify the degree of CTCF motif enrichment. In support of figure 3-27, we found that the CTCF motif was significantly enriched in both interactions that increased and decreased following RUNX1/ETO knockdown (figure 3-29).

Interactions increased by RUNX1/ETO knockdown				Interactions decreased by RUNX1/ETO knockdown			
Motif	match	P-value	% targets	Motif	match	P-value	% targets
	CTCF	1e-83	12.80		CTCF	1e-49	15.28
	ETS	1e-53	44.36		ETS	1e-46	52.69
	SP1	1e-28	26.35		SP4	1e-21	29.99
	NFY	1e-24	15.36		AP1	1e-16	20.82
	RUNX	1e-22	15.64				

**Figure 3-29: Differentially interacting DNaseI hypersensitive sites were enriched for CTCF motifs**  
 This table is the result of *de novo* motif search of differentially interacting DNaseI hypersensitive sites, using HOMER bioinformatics software (285).



**Figure 3-30: The density of CTCF motifs in DNaseI footprints decreased following RUNX1/ETO knockdown**

DNaseI footprints unique to RUNX1/ETO depleted cells (siRE, blue), unique to control transfected cells (siMM, red) and shared between the two (common, black) were assed for CTCF enrichment with HOMER software(285). The peak height represents the density of the CTCF motif in the DNaseI footprints.

Since the CTCF motif is enriched at differential interactions we hypothesised that CTCF may play a role in orchestrating the changes in DNA interactions triggered by RUNX1/ETO knockdown. In order to investigate this, we needed to see if these CTCF motifs were bound by CTCF and whether this was affected by RUNX1/ETO knockdown. As we did not have CTCF ChIP-seq data we re-ran libraries used to generate existing DNaseI data (219) at high sequencing depth. We then used the Wellington digital footprinting algorithm (302) to identify regions that were protected from DNaseI digestions, as a surrogate measure of factor occupancy, and then searched for the CTCF motif in the DNaseI footprints. We examined footprints shared between control and RUNX1/ETO knockdown; unique to the control data set and unique to RUNX1/ETO knockdown. Figure 3-30 shows that there were many more footprinted CTCF motifs in control cells indicating that CTCF binding changes after RUNX1/ETO knockdown. This finding, in combination with the enrichment of the CTCF motif at differential interactions, suggested that changes in DNA interaction could be mediated by a reorganisation of CTCF binding following RUNX1/ETO depletion.

## Chapter 4. DISCUSSION

### 4.1 CBF complex inhibition in t(8;21) AML cells

Transcription factors have been considered an undruggable target (272, 273) until pioneering work from the Bushweller lab demonstrated the successful targeting of the CBF complex. They designed small molecule, allosteric inhibitors to block the RUNT domain to CBF $\beta$  interaction by inducing a conformation change in CBF $\beta$  (275, 278). We collaborated with the Bushweller lab to determine whether these inhibitor compounds could target RUNX1 or RUNX1/ETO in our t(8;21) AML model.

#### 4.1.1 Treatment of HL60 and Kasumi-1 cells with a CBF complex inhibitor induces apoptosis in a dose dependant manner

We first tested the effect of the CBF complex inhibitor on the viability of Kasumi-1 and HL60 cells. The compound induced apoptosis in Kasumi-1 cells, in a dose dependant manner (figure 3-1). 'RUNX1 dependence' has been reported to be unique to AML cells with genetic abnormalities that affect the CBF complex (265, 266). However, we found that HL60 cell, which have no CBF abnormality, also die after inhibitor application. RUNX1 knockdown in HL60 cells has been shown to have no reported effect on the cell viability (21). This suggests that the compound may induce cell death by additional mechanisms to RUNX1 inhibition, or that in these cells another RUNX family member compensates for the depletion of RUNX1. Furthermore, although published RUNX1 knockdown in Kasumi-1

cells does lead to increased apoptosis, it is not to the extent that we see after application of our inhibitors (265). However, it must be noted that knockdown and inhibitor experiments are not directly comparable; knockdown reduces RUNX1 protein levels whereas inhibitor treated cells retain protein expression.

#### **4.1.2 The CBF complex inhibitor has no significant effect on transcription factor binding**

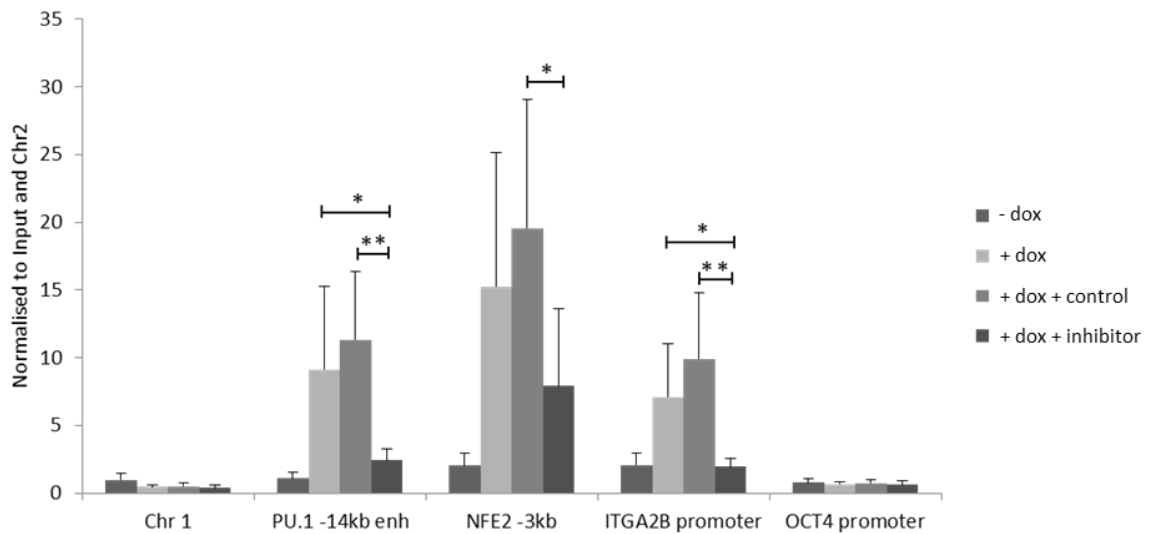
The interaction between RUNX1 and CBF $\beta$  is known to dramatically enhance DNA binding (237, 269). CHIP experiments were conducted to determine whether the inhibitor could affect binding of RUNX1 and RUNX1/ETO to DNA. Allosteric CBF complex inhibitors have resulted in a reduction in the association of RUNX1 to its DNA binding motif *in vitro* and also in embryonic stem cells carrying an inducible version of RUNX1 (275, 278). In contrast, our CHIP experiments at 6 hours showed there was no inhibition of DNA association upon inhibitor treatment (figure 3-2). This was despite the clear death response seen at 48 hours after inhibitor application, which suggests RUNX1 inhibition may be taking place.

Kasumi-1 cells undergo cell division approximately every 48 hours. During the process of mitosis, existing CBF complexes are likely to be dissociated from the DNA during DNA replication. New CBF complexes must then be formed and assembled on the DNA of daughter cells. We therefore hypothesised that the inhibitor may not be able to disrupt CBF complexes that are already assembled and bound to DNA. However, the compound might prevent the formation of new CBF complexes after mitosis and thus block RUNX1 DNA binding. Without the expression of RUNX1 target genes, the cells die at this point.

Unfortunately by this stage the inhibition of DNA binding is undetectable due to the inability to do accurate CHIP experiments on dead cells.

The above theory presumes that RUNX1 DNA binding and association with CBF $\beta$  is stable, unlike the highly dynamic behaviour exhibited by other transcription factors (303, 304). However, the majority of studies into transcription factor binding kinetics are conducted *in vitro*. It is likely that the kinetics will differ *in vivo*, due to the fact that chromosomal DNA is cluttered with other proteins and provides non-specific DNA sequences to which transcription factors can bind (305).

A recent publication from our group suggests that the compound may indeed be specific to de novo binding (278). The CBF complex inhibitor was tested in an embryonic stem (ES) cell line with inducible RUNX1. Upon treatment with doxycycline, ES cells express a HA tagged RUNX1 which can bind to RUNX1 target genes and drive their expression. In this system there will be many new HA-RUNX1-CBF $\beta$  complexes generated following induction and no complexes at the DNA prior to induction. When doxycycline is added in the presence of the inhibitor compound, HA-RUNX1 DNA binding is inhibited (Figure 4-1). The inhibition of RUNX1 binding may be detectable in this case as the compound is effective at blocking the formation of the new, induced RUNX1-CBF $\beta$  complexes. This is supported by co-immunoprecipitation demonstrating an inhibition of RUNX1 and CBF $\beta$  interaction (278). This system is preferable for the detection of RUNX1 inhibition as cell are not dependent on RUNX1 for survival, thus it is possible to conduct accurate CHIP experiments.



**Figure 4-1: CBF complex inhibitors block *de novo* RUNX1 binding domain**

This bar graph, published by Illendula *et al.*, presents enrichment of HA-Runx1 at target loci, following either no RUNX induction (dox -), Runx1 induction by dox (dox +), RUNX induction with control compound (AI-4 88) and RUNX1 induction with the inhibitor compound (AI-14-91). This data was generated using ChIP with an anti-HA antibody recognizing HA tagged Runx1. Manual qPCR was used to detect enrichment at specific loci. Enrichment is normalised to input and a negative control (Chr2). Error bars represent standard deviation between 5 biological replicates. A one-way ANOVA test was used to analyse variance in HA enrichment values between the inhibitor and control compound, or between inhibitor and +dox treatment, \* symbolises  $p < 0.05$  and \*\* symbolises  $p < 0.01$  (adapted from (278)).

Another explanation for the observation of cell death but lack of DNA binding inhibition is that the compound is hindering the transcription of RUNX1 genes, but not overtly blocking DNA binding. This could, for example, be due to the inhibition of the recruitment of co-activators additional to CBF $\beta$ .

#### **4.2 RUNX1/ETO depletion led to the up regulation of genes involved in myeloid differentiation**

The siRNA mediated knockdown of RUNX1/ETO has been used in several high impact publications (199, 219, 257) that investigate the effect of RUNX1/ETO on the gene expression and epigenetic profile of t(8;21) AML cells. We wished to manually validate the gene expression changes and assess a new, shRNA mediated knockdown system - SKNO-1 cells with an inducible shRNA specific to RUNX1/ETO - and compare results with previously published data. This inducible system will facilitate experiments in which conditions with and without RUNX1/ETO are required. It also avoids having to repeatedly transfect cells with siRNA.

RUNX1/ETO was successfully depleted from both Kasumi-1 and SKNO-1 R/E cells (Figure 3-3), with no significant effect on RUNX1 expression. There was a slight reduction of RUNX1 expression in SKNO-1 R/E cells after 72 hours, which could be attributed to differentiation of the cells in response to RUNX1/ETO depletion (figure 3-3 B) (199, 306). With both systems, there was an upregulation of genes involved in myeloid differentiation, which is in accordance our previously published data where a depletion method with the same siRNA sequence was used (199, 257). The shRNA depletion system gave comparable results to the siRNA system, allowing us to confidently use this model in



future experimentation. This system will provide prolonged RUNX1/ETO depletion, relative to siRNA depletion, which will allow more accurate time course analysis. Furthermore, it will streamline future experiments regarding RUNX1/ETO function, which could require analysis in the presence and absence of RUNX1/ETO.

#### **4.2.1 RUNX1/ETO knockdown has no significant effect on SP1 DNA binding**

We hypothesised that Sp1 may play a role in RUNX1/ETO mediated pathogenesis. This hypothesis was based on publications demonstrating a physical interaction between the two proteins and co-occupancy of RUNX1/ETO and Sp1 at the DNA. Furthermore, we observed that the Sp1 motif has increased protection after RUNX1/ETO depletion (figure 3-4).

This increased protection of the Sp1 motif suggests an increase in Sp1 binding after RUNX1/ETO knockdown. However, we actually saw a loss in the total number Sp1 ChIP-seq peaks after RUNX1/ETO knockdown. This was not due to a decrease in Sp1 expression, as mRNA and protein levels did not change (figure 3-5). Sp1 is known to exhibit positive autoregulation on its own promoter (307). As RUNX1/ETO is known to antagonise Sp1 transactivation activity, it is sensible to think that RUNX1/ETO knockdown should lead to an increase in Sp1 expression, however this was not the case (262).

The loss of Sp1 DNA binding sites after RUNX1/ETO knockdown (figure 3-6A) could be because Sp1 is recruited to its binding sites by RUNX1/ETO; the two proteins have been shown to physically interact. However, it should be noted that the published interaction between Sp1 and RUNX1/ETO was detected via co-immunoprecipitation with proteins from 293T cells transfected with plasmids expressing AML1-ETO and HA-tagged Sp1 (262). This experiment has inherent limitations as often highly expressed proteins bind non-specifically. When another group determined the constituents of the RUNX1/ETO complex via mass spectrometry, Sp1 was not detected (256).

Our analyses showed that Sp1 and RUNX1/ETO bind to distinct genomic sites (figure 3-6 B). This contradicts the work of another group, who showed RUNX1/ETO and Sp1 bind to the same sites (247). The reason for this could be the different cell models used. Maiques-Diaz *et al.* used human stem/progenitor hematopoietic cells (HSPCs), from human CD34+ umbilical cord blood samples, stably transduced with retroviruses expressing HA-tagged RUNX1/ETO. The HA tag is very short linear recognition sequence so is unlikely to effect the properties of RUNX1/ETO. However, this cannot be ruled out completely. One possible explanation for the discrepancy between our results is that the HSPCs expressed other transcription factors at different levels, thus RUNX1/ETO forms different complexes in this context. Furthermore, we performed genome wide Sp1 DNA binding analysis, whereas their conclusions were drawn from ChIP-Chip experiments which predominantly look at sequences around promoters, together with qPCR of specific loci. Perhaps the 7 genes they analysed are indeed bound by both RUNX1/ETO and Sp1, whereas at most other loci their binding is unrelated. In other words, their findings may be accurate at the 7 loci analysed, but not a genome wide phenomenon.

### **4.3 The importance of C/EBP $\alpha$ expression levels in t(8;21) AML**

#### **4.3.1 C/EBP $\alpha$ is required for the full upregulation of myeloid genes after RUNX1/ETO depletion**

There is a large body of evidence suggesting that the down regulation of C/EBP $\alpha$  expression in t(8;21) AML is a critical mechanism for leukaemia maintenance (see section 1.8.1). Our previously published ChIP-seq data suggested that alleviation of RUNX1/ETO mediated suppression of C/EBP $\alpha$  may be central to the differentiation response seen after RUNX1/ETO depletion, as we saw a considerable increase in C/EBP $\alpha$  expression and genome wide DNA binding following RUNX1/ETO knockdown (figure 1-16), a phenomenon not observed with other transcription factors. In this study we performed experiments in which RUNX1/ETO and C/EBP $\alpha$  were depleted together in order to directly test the importance of C/EBP $\alpha$  to the response to RUNX1/ETO knockdown.

RUNX1/ETO depletion led to a rapid upregulation of C/EBP $\alpha$  expression (figure 3-7 A) which is in keeping with several studies providing evidence for direct repression of C/EBP $\alpha$  by RUNX1/ETO. A physical interaction between the two proteins has been reported, which inhibits C/EBP $\alpha$  positive autoregulation (198). In addition, RUNX1/ETO binds to and represses both the C/EBP $\alpha$  promoter and enhancer (199, 200).

In our study, several myeloid genes were significantly upregulated after RUNX1/ETO knockdown (figure 3-7 C), which supports previously published gene expression data from our group (199). The effect of RUNX1/ETO depletion on gene expression is perhaps

unsurprising considering the results of gain of function studies, in which RUNX1/ETO is ectopically expressed in haematopoietic cells (308-310).

When the upregulation of C/EBP $\alpha$  was blocked by C/EBP $\alpha$  siRNA, the induction of select myeloid genes was significantly inhibited (figure 3-7 C). We found that the same genes were upregulated in response to C/EBP $\alpha$  overexpression, which validates that C/EBP $\alpha$  inhibition is responsible for this effect (see section 1.3.3 and figure 3-11). These results clearly demonstrate the essentiality of C/EBP $\alpha$  upregulation to the differentiation response following RUNX1/ETO depletion. This is consistent with the known role of C/EBP $\alpha$  in driving myeloid differentiation. For example, C/EBP $\alpha$  deficient mice display impaired myelopoiesis (197).

Not all myeloid C/EBP $\alpha$  target genes exhibited diminished upregulation following concomitant RUNX1/ETO and C/EBP $\alpha$  knockdown. This could be due to the compensatory effect of other C/EBP family members; C/EBP $\delta$  and C/EBP $\epsilon$  are also upregulated upon RUNX1/ETO knockdown. These CEBP proteins may be able to drive the expression of myeloid genes which are not specifically dependant on C/EBP $\alpha$ . For example, C/EBP $\epsilon$  drives the expression of genes involved in macrophage development and is primarily expressed in myeloid cells (311). Furthermore, gene replacement studies have shown that C/EBP $\alpha$  can be substituted for C/EBP $\beta$  in haematopoiesis (312). It is conceivable that the other C/EBP family members (C/EBP $\epsilon$  and C/EBP $\delta$ ) could exhibit similar redundancies and act synergistically with C/EBP $\alpha$ .

#### **4.3.2 Activation of a $\beta$ -Estradiol inducible form of C/EBP $\alpha$ in t(8;21) AML cells alleviates the RUNX1/ETO mediated differentiation block**

It has become clear that down-regulating C/EBP $\alpha$  repression is key to t(8;21) leukemia maintenance. When RUNX1/ETO is removed, thus alleviating repression, C/EBP $\alpha$  levels increase and cells undergo myeloid differentiation. To see whether the upregulation of C/EBP $\alpha$  is sufficient to override the RUNX1/ETO mediated differentiation block, we over-expressed an inducible form of C/EBP $\alpha$  (C/EBP $\alpha$  – ER) in t(8;21) AML cells. We found that, to a large extent, C/EBP $\alpha$  overexpression phenocopied RUNX1/ETO depletion.

This result is supported by a previous publication from Pabst et.al who found that C/EBP $\alpha$  levels were undetectable in RUNX1/ETO expressing cells and ectopic C/EBP $\alpha$  expression triggered terminal neutrophilic differentiation; the RUNX1/ETO mediated block in differentiation was overcome (198). Pabst *et al.* demonstrated the cells undergo myeloid differentiation, but no further transcriptional or epigenetic analysis was conducted and it was not clear which gene expression patterns were induced. We therefore performed RNA-seq, genome wide gene expression analysis, on Kasumi-1 cells with and without CEBP $\alpha$  over-expression.

We found that C/EBP $\alpha$  over expression led to both upregulation and down regulation of genes. A recently published study shows C/EBP $\alpha$  binds to and directly suppresses stem cell genes, such as SOX4 (202). Perhaps in our study C/EBP $\alpha$  is exhibiting a dual role, directly activating and repressing gene. However, we found that only a subset of the genes that are downregulated by C/EBP $\alpha$  overexpression are actual C/EBP $\alpha$  targets. This

suggests that C/EBP $\alpha$  may not be directly repressing these genes and that down-regulation is an indirect effect of a shift in the transcriptional network towards the differentiated state. It must be noted that the ChIP-seq data used to classify a gene as a 'C/EBP $\alpha$  target' was generated from Kasumi-1 cells (219) but not the very same cell line used in this study. Experiments examining the binding of the C/EBP $\alpha$ -ER fusion protein are therefore currently underway.

Correlation and Principal Component Analysis (PCA) of gene expression patterns demonstrate that induction of C/EBP $\alpha$  has similar effects to RUNX1/ETO knockdown. However, 'principle component one' highlights that there are small differences in the gene expression profiles of the RUNX1/ETO knockdown and C/EBP $\alpha$  over expression gene sets, even at baseline, despite Kasumi-1 cells being used in both cases. This is likely to be an artefact of the experimental methods used. For example, the RUNX1/ETO knockdown data set was obtained from cells which have been electroporated with siRNA. The non-specific effects of siRNA are well recognised and have been proven experimentally (313).

We found that C/EBP $\alpha$  overexpression can mimic many of the transcriptional effects of RUNX1/ETO knockdown; 40% of all genes changing expression after RUNX1/ETO knockdown also change expression by C/EBP $\alpha$  overexpression, despite RUNX1/ETO still being in the nucleus. Examples of shared, upregulated myeloid genes include *NKG7*, *MS4A3*, *RNASE2* and *LCP1*. There were also many shared down regulated genes, including *DUSP6* and the stem cell gene *CD34*. *DUSP6* gene encodes a signalling molecule which is involved in regulating proliferation and differentiation (314).

The significant overlap in gene expression change, between RUNX1/ETO knockdown and C/EBP $\alpha$  overexpression, is particularly impressive considering the majority of genes upregulated by C/EBP $\alpha$  are directly bound by RUNX1/ETO. This suggests that C/EBP $\alpha$  can override the repressive epigenetic effects of RUNX1/ETO. One possibility is that C/EBP $\alpha$  can displace the RUNX1/ETO complex from its target genes and/or transcriptional co-regulators. C/EBP $\alpha$  has been shown to bind to PU.1 and displace the PU.1 co-activator JUN (315). Perhaps C/EBP $\alpha$  could interfere with the RUNX1/ETO complex in a similar manner.

C/EBP $\alpha$  overexpression also has effects distinct from those of RUNX1/ETO knockdown. These are likely to be due to differences in total C/EBP $\alpha$  expression level in each experiment. C/EBP $\alpha$  levels in the overexpression experiments are likely to be much higher than those induced by RUNX1/ETO knockdown. The levels seen after 17 $\beta$ -Estradiol induction could be sufficient for C/EBP $\alpha$  to bind to relatively more accessible motifs, thus altering the expression of additional genes. There are publications suggesting that C/EBP $\alpha$  may act as a pioneer factor, thus enhancing its ability to manipulate the gene expression profile (316, 317). Furthermore, we found an enrichment of CEBP motif in the DNaseI sites which appear after RUNX1/ETO knockdown (figure 1-18). Perhaps CEBP $\alpha$  is acting as a pioneer factor here, driving the formation of these new DNaseI hypersensitive sites.

#### **4.4 The effect of RUNX1/ETO knockdown on specific cis-regulatory element interactions**

Since the advent of chromosome conformation capture methods, the notion that gene expression changes can be mediated by changes in DNA looping is well established (63, 118). There is now an increasing body of evidence that transcription factors initiate and maintain these DNA loops (63, 279-281). We therefore hypothesised that the gene expression changes seen after RUNX1/ETO knockdown may be driven by alterations in promoter-*cis* element interactions, and these changes to DNA interactions could be mediated by alterations in the transcription factor binding profile. We tested this hypothesis by performing chromatin conformation capture experiments on t(8;21) AML cells with and without RUNX1/ETO expression.

##### **4.4.1 RUNX1/ETO knockdown had no effect on specific cis-regulatory element interactions at the *SPI1* and *CD34* loci**

Several studies have established the effect of RUNX1/ETO knockdown on the transcriptomic and epigenetic profile of t(8;21) AML cells. However, the effect of these changes on chromosome conformation was not investigated. Here the 4C-seq method was used as a means of interrogating specific cis-regulatory element interactions, in this case within the *SPI1* and *CD34* loci. These genes were of particular interest as they exhibit changes in their transcription factor binding profile and gene expression after RUNX1/ETO knockdown, and the level of their expression level has a clear role in haematopoietic differentiation (293).



We successfully detected reciprocal interactions between the *SPI1* and *CD34* promoters and their already characterised regulatory elements, from both selected viewpoints. Reassuringly, the assay appears to be specific as regions of high contact intensity correspond to transcription start sites, putative *cis*-elements and regions of DNase I hypersensitivity (figure 3-13 and 3-14). Furthermore the agreement between biological replicates indicates the experiment was highly reproducible (figure 3-15).

Contrary to already published data, generated with manual 3C, we did not detect an interaction between the *CD34* promoter and DRE (292) (63) (figure 3-14). This could be due to use of a different method, and/or the genomic distance of the selected view-point fragment from the defined DRE (282). Here specific 4C-seq primers were designed to a restriction fragment close to a RUNX1/ETO ChIP-seq peak at this element. The interaction may have been detectable if the viewpoint was shifted closer to the defined enhancer region (between +18.8 to +19.6 kb). A previous study, which used 4C-seq to detect interactions with the Oct4 promoter, demonstrated that minor variations in viewpoint positioning can significantly influence the resulting interaction profile (282).

We had hypothesised that the reorganisation of transcription factor binding after RUNX1/ETO knock-down, within pre-existing open chromatin, could lead to alterations in specific DNA interactions. However, we saw that RUNX1/ETO depletion had no effect on interactions at the *SPI1* and *CD34* loci, despite the marked increase in CEBP $\alpha$  binding and change in gene expression (figure 3-16), indicating that changes in transcription factor occupancy did not impact on intra-nuclear interactions. Subsequent Capture HiC experiments confirmed the lack of differential interaction at these loci. It is therefore

likely that the expression changes are not due to differences in enhancer - promoter looping. Perhaps the alleviation of RUNX1/ETO mediated repression of the promoter via the recruitment of co-repressors is accountable for the increase in PU.1 expression (318, 319). The down regulation of CD34 could be due to the increase in C/EBP $\alpha$  binding at the promoter. C/EBP $\alpha$  is known to repress stem cell genes by binding to gene promoters (202).

The fact that we did not detect alterations in promoter-enhancer interaction, despite changes in gene expression and transcription factor binding profile, has also been observed by other researchers. For example, Eileen Furlong's group investigated the interactions made by 103 loci, which were selected based on their dynamic changes in gene expression and transcription factor occupancy between two drosophila developmental stages. They found that the majority of these differences were not reflected in alterations in promoter-enhancer DNA interactions (320).

However, from these limited data we could not conclude that RUNX1/ETO depletion has no effect on specific DNA interactions. Moreover, we only investigated loci which do not exhibit overt changes in DNaseI hypersensitivity after RUNX1/ETO knockdown. Although the majority of transcription factor binding alterations occur within pre-existing DNaseI sites, RUNX1/ETO knockdown does lead to the generation of many new DNaseI hypersensitivity sites at distal regions (figure 1-17). Interestingly, these regions are significantly enriched for C/EBP $\alpha$  motifs (figure 1-18). They may represent de novo enhancer regions, only engaging in promoter interaction after RUNX1/ETO depletion and the concomitant increase in CEBP $\alpha$  expression.

#### 4.5 The effect of RUNX1/ETO knockdown on genome wide DNA interactions

In Kasumi-1 cells 1,396 genes significantly change their expression after RUNX1/ETO depletion, and over 3000 DNaseI hypersensitivity sites form that are unique to RUNX1/ETO depleted cells (219). Using 4C-seq to interrogate all of these regions would be very time consuming and prohibitively expensive. The logical next step was therefore an assessment of the conformation of the entire genome using other methods such as HiC, which looks at all interactions within the genome. However, the vast complexity of HiC libraries requires extremely deep sequencing in order to provide sufficient resolution to define specific DNA contacts. To circumvent this issue, a collaborator of ours recently developed a method which combines HiC with solution hybridization selection (“Capture HiC”), to enrich HiC libraries for promoter regions, thus allowing us to assess the interactions made by 22,000 promoters with their distal elements in a single experiment (figure 2-2) (77).

Our results highlight the importance of genome wide interaction mapping. The *SPI1* URE enhancer interacts with the *SPI1* promoter and not with the promoter of the closest gene, *SLC39A13*. This supports already published data which demonstrates that not all distal *cis*-elements interact with the promoter of the nearest gene (78). Therefore *cis*-element interactions must be mapped experimentally as genomic distance should not be used as a predictor of DNA interaction and enhancer function.

#### **4.5.1 The majority of DNA interactions are intrachromosomal with a trend of decreased interaction frequency with genomic distance**

The aim of the ambitious Capture HiC experiments was to gain more insight into the global mechanisms governing the chromatin changes seen after RUNX1/ETO depletion, shining more light onto the molecular basis of the RUNX1/ETO mediated differentiation block. We therefore prepared Capture HiC libraries from Kasumi-1 cells, with and without siRNA mediated RUNX1/ETO knockdown.

As a means of quality controlling our Capture HiC data, we first looked for well-established features of genome organisation. We found that the majority of chromosomal interactions occur in *cis* i.e. within the same chromosome (figure 3-19). This should always be the case, regardless of cell type or species (92, 99). This trend is due, primarily, to the presence of chromosome territories which separate individual chromosomes (131).

To further validate our data, we looked for evidence of the t(8;21) translocation. The ability to detect translocations with chromosome conformation capture analysis has been demonstrated previously (296, 297). Encouragingly, we saw a block of high contact intensity representing apparent interactions between chromosome 8 and 21, which signifies the t(8;21) translocation.

We then focused on a single chromosome, chromosome 8, at higher resolutions (1 Mb) (figure 3-18) . We were able to display the well-established notion that the probability of DNA interaction decreases exponentially with genomic distance (92, 134). The same trends were seen in both biological replicates. This offered further reassurance that the experiment had been conducted effectively, and reproducibly, and we could continue to extract interaction data with confidence.

The statistically significant interactions were determined and were then filtered against our DNaseI-seq data, so that only interactions involving DNaseI hypersensitive restriction fragments were included. This was in order to direct our analysis to the active *cis*-elements which were likely to be regulating gene expression. Only approximately one third of the interactions were hypersensitive. This is in accordance with Mifsud, B *et al.* who generated Capture HiC libraries following the same protocol and found that interactions were often not enriched for DNaseI hypersensitivity, and this usually corresponded to a gene with low level expression (77). Some of these non-DNaseI hypersensitive interactions may, for example, be involved in polycomb mediated gene repression. Polycomb complexes have been shown to act by regulating promoter-promoter interactions (321, 322). It would be interesting to see if these non-hypersensitive interacting regions are predominantly promoter fragments. The non-hypersensitive interactions may also serve a more indirect, structural role in gene regulation, by helping to bringing active elements together.

#### **4.5.2 RUNX1/ETO knockdown led to alterations in specific promoter-*cis*-element interactions that did not correlate with gene expression change**

We found that the large-scale genomic structures were unaltered by RUNX1/ETO knockdown (figure 3-19). This was expected given that features such as the topologically associated domains are conserved between different cell types and different species (92). However, when HOMER software was used to detect the interactions that were differential between control and RUNX1/ETO knockdown, we found that hundreds of promoter-*cis* regulatory element interactions changed significantly. These differential interactions were associated with differentially expressed gene promoters; however the change in interactions strength was not correlated with the direction of gene expression change (figure 3-23). One explanation is that when the genes alter their expression, they both increase and decrease interactions within their profile of enhancers. This was indeed the case for several differentially expressed genes in this study (figure 3-24). This is in accordance with several studies, such as Choukrallah, MA *et al.* who demonstrated that, during B cell differentiation, enhancer repertoires are dynamically reorganised (323).

#### **4.5.3 DNA interactions at the *CEBPA* locus**

We have shown that low C/EBP $\alpha$  expression levels are critical for the maintenance of t(8;21) leukaemia. We investigated the interactions made by the *CEBPA* promoter to gain more insight into how *CEBPA* expression levels are regulated (figure 3-25). The *CEBPA* promoter is known to interact with a recently characterised +42 kb enhancer element (200). The interaction with this enhancer is needed for myeloid-lineage priming and to

drive sufficient *CEBPA* expression for neutrophilic differentiation (200). Prior to this study, we observed that RUNX1/ETO binding is highly enriched at this region. It would therefore be reasonable to propose that the rapid upregulation of *CEBPA* expression following RUNX1/ETO knockdown is due to the release of this enhancer from RUNX1/ETO mediated repression. However, our Capture HiC analysis showed that the interaction strength between the *CEBPA* promoter and this enhancer is not significantly altered by RUNX1/ETO knockdown.

In contrast, an interaction at +29kb significantly increases after RUNX1/ETO knockdown. This element is highly DNaseI hypersensitive and enriched for p300. Avellino, R *et al.* also detected a +29kb C/EBP $\alpha$  promoter interaction, but only in monocytes. They found it was enriched for H3K27ac, thus suggesting enhancer function. Using our ChIP-seq data we can see that this region is also bound by RUNX1/ETO, so perhaps release of this element from RUNX1/ETO mediated repression is responsible for the rapid increase in *CEBPA* expression after RUNX1/ETO knockdown. This result is concordant with the finding from our published metagene analysis which showed that after RUNX1/ETO knock-down, the gene expression pattern of Kasumi-1 cell changes into a pattern resembling that of monocytes (199).

We detected a strong interaction between the *CEBPA* promoter and the *CEBPG* promoter (figures 3-25 and 3-26). Alberich-Jordà, M *et al.* found that *CEBPG* is overexpressed in a subset of AMLs with silenced *CEBPA*, and proposed a model in which C/EBP $\alpha$  suppresses *CEBPG* expression (324). Perhaps the physical interaction between the two promoters facilitates regulation of this *CEBPA-CEBPG* axis.

#### **4.5.4 CTCF and C/EBP $\alpha$ may play a role in reshaping the promoter-*cis*-element interaction profile following RUNX1/ETO knockdown**

A *de novo* motif search of the differential DNA interactions revealed that these interactions were significantly enriched for CTCF motifs (figure 3-29). Furthermore, CTCF binding appears to be altered by RUNX1/ETO knockdown (figure 3-30). Taken together these data suggest that CTCF may play a role in reshaping the promoter-*cis*-element interaction profile following RUNX1/ETO knockdown.

CTCF binding was long considered to be largely invariant, apart from during the imprinting process during which methylation of CTCF motifs disrupts CTCF binding (325). However, recently John Stamatoyannopoulos's group found that CTCF binding is considerably different between different types of somatic cells and that these differences are strongly associated with DNA methylation (326). As methylation does not change significantly after RUNX1/ETO knockdown (data not published), methylation is unlikely to be a driver of CTCF binding alterations in our experiments. Another mechanism for CTCF binding modification is the eviction of CTCF from its binding site, via transcription and nucleosome repositioning, followed by the binding of another transcription factor in its place (327). This mechanism is a more likely explanation for the differential CTCF binding that we found.

We found a considerable increase in C/EBP $\alpha$  binding at the differentially interacting DNaseI hypersensitive sites, particularly at those which increase in interaction following RUNX1/ETO knockdown (figure 3-28). Perhaps C/EBP $\alpha$  is involved in forming these new promoter-*cis*-element interactions, and this is one of the mechanisms C/EBP $\alpha$  employs to



drive the differentiation response which follows RUNX1/ETO depletion. There is currently no published data proposing a role for C/EBP $\alpha$  in mediating DNA looping interactions. However, C/EBP $\alpha$  has been shown to co-associate with cohesin at the DNA (328), a protein known to be involved in maintaining promoter-enhancer interactions (329, 330). Furthermore, following RUNX1/ETO depletion, C/EBP $\alpha$  joins a transcription factor complex involving LMO2 which, in combination with its partner LBD1, is involved in DNA looping (331).

In summary, this work demonstrates that the presence or absence of RUNX1/ETO has a profound impact on the intra-nuclear organisation of t(8;21) cells. It also gives a first indication of which transcription factors are relevant for driving these changes.

## 4.6 Future work

### 4.6.1 Targeting RUNX1 in t(8;21) AML

1. We have shown that CBF complex inhibition leads to the death of t(8;21) AML cells, suggesting that RUNX1 activity was inhibited. The compound was designed to block RUNX1 DNA binding. However inhibition of DNA binding was not detectable in our system (section 3.1). We hypothesised that the inhibitory effect was specific to *de novo* CBF complex formation. We could test this hypothesis by measuring the effect of the compound on the binding of RUNX1 to its *de novo* binding sites, which occur following RUNX1/ETO depletion (figure 1-16). In these experiments we could use the shRNA knockdown system that was validated in this study (figure 3-3).

### 4.6.2 The role of C/EBP $\alpha$ in t(8;21) AML

1. We have convincingly demonstrated that low C/EBP $\alpha$  expression level are critical to t(8;21) AML maintenance, with the induction of C/EBP $\alpha$  largely phenocopying the effect of RUNX1/ETO depletion (section 3.3). However, the mechanism by which C/EBP $\alpha$  drives this response is unclear. To answer this question, we will measure C/EBP $\alpha$ -ER binding by ChIP-seq before and after induction, and see how many of the differentially expressed genes are directly bound by CEBP $\alpha$ ER. We will also measure the effect of C/EBP $\alpha$ -ER induction on the genome-wide binding of RUNX1/ETO to test whether C/EBP $\alpha$  overexpression can override the RUNX1/ETO mediated differentiation block and displace RUNX1/ETO from its binding sites.

2. Previous analysis revealed that the CEBP motif is enriched in DNaseI sites that are specific to RUNX1/ETO depleted Kasumi-1 cells (figure 1-18). It is therefore possible that C/EBP $\alpha$  is driving the formation of these *de novo* DNaseI sites. To see if this is the case, we could perform DNaseI-seq before and after C/EBP $\alpha$  induction. This experiment will enable us to gain more insight into how C/EBP $\alpha$  drives the differentiation response that follows RUNX1/ETO knockdown.

3. In this study we show that upregulation of C/EBP $\alpha$  by RUNX1/ETO knockdown is necessary to drive the full differentiation response. However, it is unclear how important other C/EBP proteins are to this process. *CEBPG* and *CEBPE* are also significantly upregulated after RUNX1/ETO knockdown; it is possible that these proteins contribute to the differentiation response. To test if this is the case, we could block all CEBP protein activity with a dominant negative CEBP leucine-zipper peptide (332), in conjunction with RUNX1/ETO knockdown. We could then compare the results to those obtained by co-depletion of C/EBP $\alpha$  and RUNX1/ETO (section 3.3.2).

#### **4.6.3 Promoter-*cis*-element interactions in t(8;21)**

1. We have successfully mapped the genome-wide promoter-distal-element interactions in a t(8;21) AML cell line (section 3.5). These results will be used to annotate genes to their respective *cis*-elements, which will enable us to accurately construct gene regulatory networks. Before doing so, we could determine which interacting elements are likely to be enhancers with our p300 and H3K9Ac ChIP-seq data. Additionally, we could perform H3K27Ac and H3K4Me1 ChIP-seq, as these modifications are indicators of potential

enhancer activity (333, 334). We could also assess the function of elements of interest, with luciferase reporter assays, such as the region +29kb of the C/EBP $\alpha$  promoter, which we found to strongly interact with the C/EBP $\alpha$  promoter, specifically in RUNX1/ETO depleted cells.

2. This work demonstrates that RUNX1/ETO knockdown has a profound impact on the intra-nuclear organisation of t(8;21) cells. One of the most intriguing results from this study was that CTCF could play a role in reshaping the DNA interaction profile following RUNX1/ETO knockdown. This hypothesis was based on the enrichment of differential interactions for the CTCF motif and the differential protection of the CTCF motif following RUNX1/ETO depletion (figures 3-29 and 3-30). We will now perform CTCF ChIP-seq, before and after RUNX1/ETO knockdown, to directly measure CTCF binding alterations and see if they correlate with differential promoter-distal-element interactions. It would also be useful to perform ChIP-seq with members of the Cohesion complex, such as RAD21, given that the Cohesin complex is found at most CTCF sites (335). These data could then be integrated with our existing ChIP-seq data sets to analyse the interplay of other factors with CTCF and the Cohesin complex.

3. In order to determine whether our data is relevant to disease, we will correlate the Kasumi-1 interaction profile with that of a t(8;21) AML patient. We have already prepared a Capture HiC library with a t(8;21) AML patient sample, which is now ready for analysis. In addition, we could compare the t(8;21) AML interaction profile with other AML subtypes, and normal samples of the same maturation stage. This will allow us to identify disease specific DNA interactions. The role of these interactions in disease progression could then be assessed with functional assays.

## REFERENCES

1. Cockerill PN. Structure and function of active chromatin and DNase I hypersensitive sites. *FEBS J.* 2011;278(13):2182-210.
2. Davies HG, Small JV. Structural units in chromatin and their orientation on membranes. *Nature.* 1968;217(5134):1122-5.
3. Ris H, Kubai DF. Chromosome structure. *Annu Rev Genet.* 1970;4:263-94.
4. Schalch T, Duda S, Sargent DF, Richmond TJ. X-ray structure of a tetranucleosome and its implications for the chromatin fibre. *Nature.* 2005;436(7047):138-41.
5. Thoma F, Koller T, Klug A. Involvement of histone H1 in the organization of the nucleosome and of the salt-dependent superstructures of chromatin. *J Cell Biol.* 1979;83(2 Pt 1):403-27.
6. Weischet WO, Allen JR, Riedel G, Van Holde KE. The effects of salt concentration and H-1 depletion on the digestion of calf thymus chromatin by micrococcal nuclease. *Nucleic Acids Res.* 1979;6(5):1843-62.
7. Xiao B, Freedman BS, Miller KE, Heald R, Marko JF. Histone H1 compacts DNA under force and during chromatin assembly. *Molecular Biology of the Cell.* 2012;23(24):4864-71.
8. Quénet D, McNally JG, Dalal Y. Through thick and thin: the conundrum of chromatin fibre folding in vivo. *EMBO Reports.* 2012;13(11):943-4.
9. Albig W, Doenecke D. The human histone gene cluster at the D6S105 locus. *Hum Genet.* 1997;101(3):284-94.
10. Luk E, Ranjan A, Fitzgerald PC, Mizuguchi G, Huang Y, Wei D, et al. Stepwise histone replacement by SWR1 requires dual activation with histone H2A.Z and canonical nucleosome. *Cell.* 2010;143(5):725-36.
11. Tagami H, Ray-Gallet D, Almouzni G, Nakatani Y. Histone H3.1 and H3.3 complexes mediate nucleosome assembly pathways dependent or independent of DNA synthesis. *Cell.* 2004;116(1):51-61.
12. Heo K, Kim H, Choi SH, Choi J, Kim K, Gu J, et al. FACT-mediated exchange of histone variant H2AX regulated by phosphorylation of H2AX and ADP-ribosylation of Spt16. *Mol Cell.* 2008;30(1):86-97.
13. Foltz DR, Jansen LE, Bailey AO, Yates JR, 3rd, Bassett EA, Wood S, et al. Centromere-specific assembly of CENP-a nucleosomes is mediated by HJURP. *Cell.* 2009;137(3):472-84.
14. Thakar A, Gupta P, Ishibashi T, Finn R, Silva-Moreno B, Uchiyama S, et al. H2A.Z and H3.3 histone variants affect nucleosome structure: biochemical and biophysical studies. *Biochemistry.* 2009;48(46):10852-7.
15. Chakravarthy S, Luger K. The histone variant macro-H2A preferentially forms "hybrid nucleosomes". *J Biol Chem.* 2006;281(35):25522-31.
16. Volle C, Dalal Y. Histone variants: the tricksters of the chromatin world. *Curr Opin Genet Dev.* 2014;25:8-14,138.
17. Cairns BR. The logic of chromatin architecture and remodelling at promoters. *Nature.* 2009;461(7261):193-8.

18. Deindl S, Hwang William L, Hota Swetansu K, Blosser Timothy R, Prasad P, Bartholomew B, et al. ISWI Remodelers Slide Nucleosomes with Coordinated Multi-Base-Pair Entry Steps and Single-Base-Pair Exit Steps. *Cell*.152(3):442-52.
19. Cote J, Quinn J, Workman JL, Peterson CL. Stimulation of GAL4 derivative binding to nucleosomal DNA by the yeast SWI/SNF complex. *Science*. 1994;265(5168):53-60.
20. Mizuguchi G, Shen X, Landry J, Wu WH, Sen S, Wu C. ATP-driven exchange of histone H2AZ variant catalyzed by SWR1 chromatin remodeling complex. *Science*. 2004;303(5656):343-8.
21. Mizzen C. Signaling to chromatin through histone modifications: how clear is the signal? *Cold Spring Harb Symp Quant Biol*. 1998;63:469-81.
22. Turner BM. Decoding the nucleosome. *Cell*. 1993;75:5-8.
23. Lopez-Rodas G. Histone deacetylase. A key enzyme for the binding of regulatory proteins to chromatin. *FEBS Lett*. 1993;317:175-80.
24. Loidl P. Histone acetylation: facts and questions. *Chromosoma*. 1994;103:441-9.
25. Strahl BD, Allis CD. The language of covalent histone modifications. *Nature*. 2000;403(6765):41-5.
26. Kolasinska-Zwierz P, Down T, Latorre I, Liu T, Liu XS, Ahringer J. Differential chromatin marking of introns and expressed exons by H3K36me3. *Nat Genet*. 2009;41(3):376-81.
27. Heintzman ND, Hon GC, Hawkins RD, Kheradpour P, Stark A, Harp LF, et al. Histone modifications at human enhancers reflect global cell-type-specific gene expression. *Nature*. 2009;459(7243):108-12.
28. Sterner DE, Berger SL. Acetylation of histones and transcription-related factors. *Microbiol Mol Biol Rev*. 2000;64(2):435-59.
29. Ropero S, Esteller M. The role of histone deacetylases (HDACs) in human cancer. *Mol Oncol*. 2007;1(1):19-25.
30. Reinke H, Horz W. Histones are first hyperacetylated and then lose contact with the activated PHO5 promoter. *Mol Cell*. 2003;11(6):1599-607.
31. Zhao Y, Lu J, Sun H, Chen X, Huang W, Tao D, et al. Histone acetylation regulates both transcription initiation and elongation of hsp22 gene in *Drosophila*. *Biochem Biophys Res Commun*. 2005;326(4):811-6.
32. Chandy M, Gutierrez JL, Prochasson P, Workman JL. SWI/SNF displaces SAGA-acetylated nucleosomes. *Eukaryot Cell*. 2006;5(10):1738-47.
33. Ito T, Ikehara T, Nakagawa T, Kraus WL, Muramatsu M. p300-mediated acetylation facilitates the transfer of histone H2A-H2B dimers from nucleosomes to a histone chaperone. *Genes Dev*. 2000;14(15):1899-907.
34. Zeng L, Zhou MM. Bromodomain: an acetyl-lysine binding domain. *FEBS Lett*. 2002;513(1):124-8.
35. Sanchez R, Zhou M-M. The role of human bromodomains in chromatin biology and gene transcription. *Current opinion in drug discovery & development*. 2009;12(5):659-65.
36. Simon JA, Kingston RE. Mechanisms of Polycomb gene silencing: knowns and unknowns. *Nat Rev Mol Cell Biol*. 2009;10(10):697-708.
37. Landgrave-Gómez J, Mercado-Gómez OF, GUEVARA-GUZMAN R. Epigenetic mechanisms in neurological and neurodegenerative diseases. *Frontiers in Cellular Neuroscience*. 2015;9.

38. Roeder RG. Transcriptional regulation and the role of diverse coactivators in animal cells. *FEBS Lett.* 2005;579(4):909-15.
39. Rasmussen EB, Lis JT. In vivo transcriptional pausing and cap formation on three *Drosophila* heat shock genes. *Proc Natl Acad Sci U S A.* 1993;90(17):7923-7.
40. Bai L, Shundrovsky A, Wang MD. Sequence-dependent kinetic model for transcription elongation by RNA polymerase. *J Mol Biol.* 2004;344(2):335-49.
41. Izban MG, Luse DS. Transcription on nucleosomal templates by RNA polymerase II in vitro: inhibition of elongation with enhancement of sequence-specific pausing. *Genes Dev.* 1991;5(4):683-96.
42. Ring BZ, Yarnell WS, Roberts JW. Function of *E. coli* RNA polymerase sigma factor sigma 70 in promoter-proximal pausing. *Cell.* 1996;86(3):485-93.
43. Kwak H, Lis JT. Control of Transcriptional Elongation. *Annual review of genetics.* 2013;47:483-508.
44. Price DH. P-TEFb, a cyclin-dependent kinase controlling elongation by RNA polymerase II. *Mol Cell Biol.* 2000;20(8):2629-34.
45. Ramanathan Y, Rajpara SM, Reza SM, Lees E, Shuman S, Mathews MB, et al. Three RNA polymerase II carboxyl-terminal domain kinases display distinct substrate preferences. *J Biol Chem.* 2001;276(14):10913-20.
46. Fujinaga K, Irwin D, Huang Y, Taube R, Kurosu T, Peterlin BM. Dynamics of human immunodeficiency virus transcription: P-TEFb phosphorylates RD and dissociates negative effectors from the transactivation response element. *Mol Cell Biol.* 2004;24(2):787-95.
47. Jang MK, Mochizuki K, Zhou M, Jeong HS, Brady JN, Ozato K. The bromodomain protein Brd4 is a positive regulatory component of P-TEFb and stimulates RNA polymerase II-dependent transcription. *Mol Cell.* 2005;19(4):523-34.
48. Yang Z, Yik JH, Chen R, He N, Jang MK, Ozato K, et al. Recruitment of P-TEFb for stimulation of transcriptional elongation by the bromodomain protein Brd4. *Mol Cell.* 2005;19(4):535-45.
49. Saunders A, Core LJ, Lis JT. Breaking barriers to transcription elongation. *Nat Rev Mol Cell Biol.* 2006;7(8):557-67.
50. Belotserkovskaya R, Oh S, Bondarenko VA, Orphanides G, Studitsky VM, Reinberg D. FACT facilitates transcription-dependent nucleosome alteration. *Science.* 2003;301(5636):1090-3.
51. Bortvin A, Winston F. Evidence that Spt6p controls chromatin structure by a direct interaction with histones. *Science.* 1996;272(5267):1473-6.
52. Endoh M, Zhu W, Hasegawa J, Watanabe H, Kim DK, Aida M, et al. Human Spt6 stimulates transcription elongation by RNA polymerase II in vitro. *Mol Cell Biol.* 2004;24(8):3324-36.
53. Ardehali MB, Yao J, Adelman K, Fuda NJ, Petesch SJ, Webb WW, et al. Spt6 enhances the elongation rate of RNA polymerase II in vivo. *EMBO J.* 2009;28(8):1067-77.
54. Core LJ, Waterfall JJ, Lis JT. Nascent RNA sequencing reveals widespread pausing and divergent initiation at human promoters. *Science.* 2008;322(5909):1845-8.
55. Hahn S. Structure and mechanism of the RNA polymerase II transcription machinery. *Nat Struct Mol Biol.* 2004;11(5):394-403.
56. Spitalny P, Thomm M. A polymerase III-like reinitiation mechanism is operating in regulation of histone expression in archaea. *Mol Microbiol.* 2008;67(5):958-70.



57. Lamas-Maceiras M, Singh BN, Hampsey M, Freire-Picos MA. Promoter-Terminator Gene Loops Affect Alternative 3'-End Processing in Yeast. *J Biol Chem.* 2016;291(17):8960-8.
58. Maston GA, Evans SK, Green MR. Transcriptional regulatory elements in the human genome. *Annu Rev Genomics Hum Genet.* 2006;7:29-59.
59. Ptashne M, Gann A. Transcriptional activation by recruitment. *Nature.* 1997;386(6625):569-77.
60. Orphanides G, Lagrange T, Reinberg D. The general transcription factors of RNA polymerase II. *Genes Dev.* 1996;10(21):2657-83.
61. Jonkers I, Lis JT. Getting up to speed with transcription elongation by RNA polymerase II. *Nat Rev Mol Cell Biol.* 2015;16(3):167-77.
62. de la Serna IL, Ohkawa Y, Berkes CA, Bergstrom DA, Dacwag CS, Tapscott SJ, et al. MyoD targets chromatin remodeling complexes to the myogenin locus prior to forming a stable DNA-bound complex. *Mol Cell Biol.* 2005;25(10):3997-4009.
63. Levantini E, Lee S, Radomska HS, Hetherington CJ, Alberich-Jorda M, Amabile G, et al. RUNX1 regulates the CD34 gene in haematopoietic stem cells by mediating interactions with a distal regulatory element. *The EMBO journal.* 2011;30(19):4059-70.
64. Chen Y, Bates DL, Dey R, Chen PH, Machado AC, Laird-Offringa IA, et al. DNA binding by GATA transcription factor suggests mechanisms of DNA looping and long-range gene regulation. *Cell Rep.* 2012;2(5):1197-206.
65. Lee TI, Young RA. Transcription of eukaryotic protein-coding genes. *Annu Rev Genet.* 2000;34:77-137.
66. Fishburn J, Mohibullah N, Hahn S. Function of a eukaryotic transcription activator during the transcription cycle. *Mol Cell.* 2005;18(3):369-78.
67. Sauer F, Tjian R. Mechanisms of transcriptional activation: differences and similarities between yeast, *Drosophila*, and man. *Curr Opin Genet Dev.* 1997;7(2):176-81.
68. Gill G, Pascal E, Tseng ZH, Tjian R. A glutamine-rich hydrophobic patch in transcription factor Sp1 contacts the dTAFII110 component of the *Drosophila* TFIID complex and mediates transcriptional activation. *Proc Natl Acad Sci U S A.* 1994;91(1):192-6.
69. Rabenstein MD, Zhou S, Lis JT, Tjian R. TATA box-binding protein (TBP)-related factor 2 (TRF2), a third member of the TBP family. *Proceedings of the National Academy of Sciences of the United States of America.* 1999;96(9):4791-6.
70. Soldatov A, Nabirochkina E, Georgieva S, Belenkaja T, Georgiev P. TAF(II)40 Protein Is Encoded by the e(y)1 Gene: Biological Consequences of Mutations. *Molecular and Cellular Biology.* 1999;19(5):3769-78.
71. Pham AD, Sauer F. Ubiquitin-activating/conjugating activity of TAFII250, a mediator of activation of gene expression in *Drosophila*. *Science.* 2000;289(5488):2357-60.
72. Brownell JE, Zhou J, Ranalli T, Kobayashi R, Edmondson DG, Roth SY, et al. Tetrahymena histone acetyltransferase A: a homolog to yeast Gcn5p linking histone acetylation to gene activation. *Cell.* 1996;84(6):843-51.
73. Goodman RH, Smolik S. CBP/p300 in cell growth, transformation, and development. *Genes Dev.* 2000;14(13):1553-77.

74. Mizzen CA, Yang XJ, Kokubo T, Brownell JE, Bannister AJ, Owen-Hughes T, et al. The TAF(II)250 subunit of TFIID has histone acetyltransferase activity. *Cell*. 1996;87(7):1261-70.
75. Morris JR, Petrov DA, Lee AM, Wu CT. Enhancer choice in cis and in trans in *Drosophila melanogaster*: role of the promoter. *Genetics*. 2004;167(4):1739-47.
76. de Villiers J, Schaffner W. A small segment of polyoma virus DNA enhances the expression of a cloned beta-globin gene over a distance of 1400 base pairs. *Nucleic Acids Res*. 1981;9(23):6251-64.
77. Mifsud B, Tavares-Cadete F, Young AN, Sugar R, Schoenfelder S, Ferreira L, et al. Mapping long-range promoter contacts in human cells with high-resolution capture Hi-C. *Nature genetics*. 2015.
78. Sanyal A, Lajoie BR, Jain G, Dekker J. The long-range interaction landscape of gene promoters. *Nature*. 2012;489(7414):109-13.
79. Marinic M, Aktas T, Ruf S, Spitz F. An integrated holo-enhancer unit defines tissue and gene specificity of the Fgf8 regulatory landscape. *Dev Cell*. 2013;24(5):530-42.
80. Privalsky ML. The role of corepressors in transcriptional regulation by nuclear hormone receptors. *Annu Rev Physiol*. 2004;66:315-60.
81. Chen JD, Evans RM. A transcriptional co-repressor that interacts with nuclear hormone receptors. *Nature*. 1995;377(6548):454-7.
82. Sande S, Privalsky ML. Identification of TRACs (T3 receptor-associating cofactors), a family of cofactors that associate with, and modulate the activity of, nuclear hormone receptors. *Mol Endocrinol*. 1996;10(7):813-25.
83. Dhordain P, Albagli O, Lin RJ, Ansieau S, Quief S, Leutz A, et al. Corepressor SMRT binds the BTB/POZ repressing domain of the LAZ3/BCL6 oncoprotein. *Proc Natl Acad Sci U S A*. 1997;94(20):10762-7.
84. Gelmetti V, Zhang J, Fanelli M, Minucci S, Pelicci PG, Lazar MA. Aberrant recruitment of the nuclear receptor corepressor-histone deacetylase complex by the acute myeloid leukemia fusion partner ETO. *Mol Cell Biol*. 1998;18(12):7185-91.
85. Guidez F, Ivins S, Zhu J, Soderstrom M, Waxman S, Zelent A. Reduced retinoic acid-sensitivities of nuclear receptor corepressor binding to PML- and PLZF-RARalpha underlie molecular pathogenesis and treatment of acute promyelocytic leukemia. *Blood*. 1998;91(8):2634-42.
86. Harris MB, Mostecky J, Rothman PB. Repression of an interleukin-4-responsive promoter requires cooperative BCL-6 function. *J Biol Chem*. 2005;280(13):13114-21.
87. Srinivasan L, Atchison ML. YY1 DNA binding and PcG recruitment requires CtBP. *Genes Dev*. 2004;18(21):2596-601.
88. Chen L, Widom J. Mechanism of transcriptional silencing in yeast. *Cell*. 2005;120(1):37-48.
89. Tagoh H, Ingram R, Wilson N, Salvagiotto G, Warren AJ, Clarke D, et al. The mechanism of repression of the myeloid-specific c-fms gene by Pax5 during B lineage restriction. *EMBO J*. 2006;25(5):1070-80.
90. Recillas-Targa F, Pikaart MJ, Burgess-Beusse B, Bell AC, Litt MD, West AG, et al. Position-effect protection and enhancer blocking by the chicken beta-globin insulator are separable activities. *Proc Natl Acad Sci U S A*. 2002;99(10):6883-8.

91. Ong C-T, Corces VG. CTCF: An Architectural Protein Bridging Genome Topology and Function. *Nature reviews Genetics*. 2014;15(4):234-46.
92. Dixon JR, Selvaraj S, Yue F, Kim A, Li Y, Shen Y, et al. Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature*. 2012;485(7398):376-80.
93. Guelen L, Pagie L, Brasset E, Meuleman W, Faza MB, Talhout W, et al. Domain organization of human chromosomes revealed by mapping of nuclear lamina interactions. *Nature*. 2008;453(7197):948-51.
94. Handoko L, Xu H, Li G, Ngan CY, Chew E, Schnapp M, et al. CTCF-mediated functional chromatin interactome in pluripotent cells. *Nat Genet*. 2011;43(7):630-8.
95. Faure AJ, Schmidt D, Watt S, Schwalie PC, Wilson MD, Xu H, et al. Cohesin regulates tissue-specific expression by stabilizing highly occupied cis-regulatory modules. *Genome Res*. 2012;22(11):2163-75.
96. Bonora G, Plath K, Denholtz M. A mechanistic link between gene regulation and genome architecture in mammalian development. *Curr Opin Genet Dev*. 2014;27:92-101.
97. de Wit E, Bouwman BA, Zhu Y, Klous P, Splinter E, Verstegen MJ, et al. The pluripotent genome in three dimensions is shaped around pluripotency factors. *Nature*. 2013;501(7466):227-31.
98. Cheutin T, Cavalli G. Progressive polycomb assembly on H3K27me3 compartments generates polycomb bodies with developmentally regulated motion. *PLoS Genet*. 2012;8(1):e1002465.
99. Lieberman-Aiden E, van Berkum NL, Williams L, Imakaev M, Ragozcy T, Telling A, et al. Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science*. 2009;326(5950):289-93.
100. Dixon JR, Jung I, Selvaraj S, Shen Y, Antosiewicz-Bourget JE, Lee AY, et al. Chromatin architecture reorganization during stem cell differentiation. *Nature*. 2015;518(7539):331-6.
101. Gibcus JH, Dekker J. The hierarchy of the 3D genome. *Mol Cell*. 2013;49(5):773-82.
102. Pederson T. The spatial organization of the genome in mammalian cells. *Curr Opin Genet Dev*. 2004;14(2):203-9.
103. Bolzer A, Kreth G, Solovei I, Koehler D, Saracoglu K, Fauth C, et al. Three-dimensional maps of all chromosomes in human male fibroblast nuclei and prometaphase rosettes. *PLoS Biol*. 2005;3(5):e157.
104. Cremer T, Cremer M. Chromosome territories. *Cold Spring Harb Perspect Biol*. 2010;2(3):a003889.
105. Yokota H, van den Engh G, Hearst JE, Sachs RK, Trask BJ. Evidence for the organization of chromatin in megabase pair-sized loops arranged along a random walk path in the human G0/G1 interphase nucleus. *J Cell Biol*. 1995;130(6):1239-49.
106. Lemaitre JM, Danis E, Pasero P, Vassetzky Y, Mechali M. Mitotic remodeling of the replicon and chromosome structure. *Cell*. 2005;123(5):787-801.
107. Dekker J, Rippe K, Dekker M, Kleckner N. Capturing chromosome conformation. *Science*. 2002;295(5558):1306-11.
108. Lomvardas S, Barnea G, Pisapia DJ, Mendelsohn M, Kirkland J, Axel R. Interchromosomal interactions and olfactory receptor choice. *Cell*. 2006;126(2):403-13.

109. Brown JM, Leach J, Reittie JE, Atzberger A, Lee-Prudhoe J, Wood WG, et al. Coregulated human globin genes are frequently in spatial proximity when active. *J Cell Biol.* 2006;172(2):177-87.
110. Splinter E, Grosveld F, de Laat W. 3C technology: analyzing the spatial organization of genomic loci in vivo. *Methods Enzymol.* 2004;375:493-507.
111. Rippe K, von Hippel PH, Langowski J. Action at a distance: DNA-looping and initiation of transcription. *Trends Biochem Sci.* 1995;20(12):500-6.
112. Ling JQ, Li T, Hu JF, Vu TH, Chen HL, Qiu XW, et al. CTCF mediates interchromosomal colocalization between *Igf2/H19* and *Wsb1/Nf1*. *Science.* 2006;312(5771):269-72.
113. Jackson DA, Hassan AB, Errington RJ, Cook PR. Visualization of focal sites of transcription within human nuclei. *EMBO J.* 1993;12(3):1059-65.
114. Sutherland H, Bickmore WA. Transcription factories: gene expression in unions? *Nat Rev Genet.* 2009;10(7):457-66.
115. Martin S, Pombo A. Transcription factories: quantitative studies of nanostructures in the mammalian nucleus. *Chromosome Res.* 2003;11(5):461-70.
116. Schoenfelder S, Sexton T, Chakalova L, Cope NF, Horton A, Andrews S, et al. Preferential associations between co-regulated genes reveal a transcriptional interactome in erythroid cells. *Nat Genet.* 2010;42(1):53-61.
117. Antoniou M, deBoer E, Habets G, Grosveld F. The human beta-globin gene contains multiple regulatory regions: identification of one promoter and two downstream enhancers. *The EMBO Journal.* 1988;7(2):377-84.
118. Ragozy T, Bender MA, Telling A, Byron R, Groudine M. The locus control region is required for association of the murine beta-globin locus with engaged transcription factories during erythroid maturation. *Genes Dev.* 2006;20(11):1447-57.
119. Noordermeer D, Branco MR, Splinter E, Klous P, van Ijcken W, Swagemakers S, et al. Transcription and chromatin organization of a housekeeping gene cluster containing an integrated beta-globin locus control region. *PLoS Genet.* 2008;4(3):e1000016.
120. Tolhuis B, Palstra RJ, Splinter E, Grosveld F, de Laat W. Looping and interaction between hypersensitive sites in the active beta-globin locus. *Mol Cell.* 2002;10(6):1453-65.
121. Noordermeer D, Duboule D. Chromatin architectures and Hox gene collinearity. *Curr Top Dev Biol.* 2013;104:113-48.
122. Andrey G, Montavon T, Mascrez B, Gonzalez F, Noordermeer D, Leleu M, et al. A Switch Between Topological Domains Underlies HoxD Genes Collinearity in Mouse Limbs. *Science.* 2013;340(6137).
123. Montavon T, Soshnikova N, Mascrez B, Joye E, Thevenet L, Splinter E, et al. A regulatory archipelago controls Hox genes transcription in digits. *Cell.* 2011;147(5):1132-45.
124. Beccari L, Yakushiji-Kaminatsui N, Woltering JM, Necsulea A, Lonfat N, Rodriguez-Carballo E, et al. A role for HOX13 proteins in the regulatory switch between TADs at the HoxD locus. *Genes Dev.* 2016;30(10):1172-86.
125. Boveri T. Die Blastomerenkerne von *Ascaris megalocephala* und die Theorie der Chromosomenindividualität. *Arch Zellforsch.* 1909(3):181-268.

126. Zorn C, Cremer T, Cremer C, Zimmer J. Laser UV microirradiation of interphase nuclei and post-treatment with caffeine. A new approach to establish the arrangement of interphase chromosomes. *Hum Genet.* 1976;35(1):83-9.
127. Zorn C, Cremer C, Cremer T, Zimmer J. Unscheduled DNA synthesis after partial UV irradiation of the cell nucleus. Distribution in interphase and metaphase. *Exp Cell Res.* 1979;124(1):111-9.
128. Cremer T, Cremer C, Baumann H, Luedtke EK, Sperling K, Teuber V, et al. Rabl's model of the interphase chromosome arrangement tested in Chinese hamster cells by premature chromosome condensation and laser-UV-microbeam experiments. *Hum Genet.* 1982;60(1):46-56.
129. Cremer T, Cremer C, Schneider T, Baumann H, Hens L, Kirsch-Volders M. Analysis of chromosome positions in the interphase nucleus of Chinese hamster cells by laser-UV-microirradiation experiments. *Hum Genet.* 1982;62(3):201-9.
130. Cremer M, Grasser F, Lanctot C, Muller S, Neusser M, Zinner R, et al. Multicolor 3D fluorescence in situ hybridization for imaging interphase chromosomes. *Methods Mol Biol.* 2008;463:205-39.
131. Cremer T, Cremer C. Chromosome territories, nuclear architecture and gene regulation in mammalian cells. *Nat Rev Genet.* 2001;2(4):292-301.
132. Amano T, Sagai T, Tanabe H, Mizushina Y, Nakazawa H, Shiroishi T. Chromosomal dynamics at the *Shh* locus: limb bud-specific differential regulation of competence and active transcription. *Dev Cell.* 2009;16(1):47-57.
133. Williamson I, Eskeland R, Lettice LA, Hill AE, Boyle S, Grimes GR, et al. Anterior-posterior differences in *HoxD* chromatin topology in limb development. *Development.* 2012;139(17):3157-67.
134. Dekker J, Marti-Renom MA, Mirny LA. Exploring the three-dimensional organization of genomes: interpreting chromatin interaction data. *Nat Rev Genet.* 2013;14(6):390-403.
135. Simonis M, Klous P, Splinter E, Moshkin Y, Willemsen R, de Wit E, et al. Nuclear organization of active and inactive chromatin domains uncovered by chromosome conformation capture-on-chip (4C). *Nat Genet.* 2006;38(11):1348-54.
136. Zhao Z, Tavoosidana G, Sjolinder M, Gondor A, Mariano P, Wang S, et al. Circular chromosome conformation capture (4C) uncovers extensive networks of epigenetically regulated intra- and interchromosomal interactions. *Nat Genet.* 2006;38(11):1341-7.
137. Nora EP, Lajoie BR, Schulz EG, Giorgetti L, Okamoto I, Servant N, et al. Spatial partitioning of the regulatory landscape of the X-inactivation centre. *Nature.* 2012;485(7398):381-5.
138. Phillips-Cremins JE, Sauria ME, Sanyal A, Gerasimova TI, Lajoie BR, Bell JS, et al. Architectural protein subclasses shape 3D organization of genomes during lineage commitment. *Cell.* 2013;153(6):1281-95.
139. Bau D, Sanyal A, Lajoie BR, Capriotti E, Byron M, Lawrence JB, et al. The three-dimensional folding of the alpha-globin gene domain reveals formation of chromatin globules. *Nat Struct Mol Biol.* 2011;18(1):107-14.
140. van Berkum NL, Lieberman-Aiden E, Williams L, Imakaev M, Gnirke A, Mirny LA, et al. Hi-C: a method to study the three-dimensional architecture of genomes. *J Vis Exp.* 2010(39).

141. Hughes JR, Roberts N, McGowan S, Hay D, Giannoulatou E, Lynch M, et al. Analysis of hundreds of cis-regulatory landscapes at high resolution in a single, high-throughput experiment. *Nat Genet.* 2014;46(2):205-12.
142. Muller-Sieburg CE, Whitlock CA, Weissman IL. Isolation of two early B lymphocyte progenitors from mouse marrow: a committed pre-pre-B cell and a clonogenic Thy-1-lo hematopoietic stem cell. *Cell.* 1986;44(4):653-62.
143. Spangrude GJ, Heimfeld S, Weissman IL. Purification and characterization of mouse hematopoietic stem cells. *Science (New York, NY).* 1988;241(4861):58-62.
144. Laiosa CV, Stadtfeld M, Graf T. Determinants of lymphoid-myeloid lineage diversification. *Annual review of immunology.* 2006;24:705-38.
145. Orkin SH, Zon LI. Hematopoiesis: an evolving paradigm for stem cell biology. *Cell.* 2008;132(4):631-44.
146. Umeda K, Heike T, Yoshimoto M, Shiota M, Suemori H, Luo HY, et al. Development of primitive and definitive hematopoiesis from nonhuman primate embryonic stem cells in vitro. *Development (Cambridge, England).* 2004;131(8):1869-79.
147. Boisset JC, van Cappellen W, Andrieu-Soler C, Galjart N, Dzierzak E, Robin C. In vivo imaging of haematopoietic cells emerging from the mouse aortic endothelium. *Nature.* 2010;464(7285):116-20.
148. Bertrand JY, Chi NC, Santoso B, Teng S, Stainier DY, Traver D. Haematopoietic stem cells derive directly from aortic endothelium during development. *Nature.* 2010;464(7285):108-11.
149. Rebel VI, Miller CL, Eaves CJ, Lansdorp PM. The repopulation potential of fetal liver hematopoietic stem cells in mice exceeds that of their liver adult bone marrow counterparts. *Blood.* 1996;87(8):3500-7.
150. Rebel VI, Miller CL, Thornbury GR, Dragowska WH, Eaves CJ, Lansdorp PM. A comparison of long-term repopulating hematopoietic stem cells in fetal liver and adult bone marrow from the mouse. *Experimental hematology.* 1996;24(5):638-48.
151. Zanjani ED, Ascensao JL, Tavassoli M. Liver-derived fetal hematopoietic stem cells selectively and preferentially home to the fetal bone marrow. *Blood.* 1993;81(2):399-404.
152. Harrison DE, Zhong RK, Jordan CT, Lemischka IR, Aistle CM. Relative to adult marrow, fetal liver repopulates nearly five times more effectively long-term than short-term. *Exp Hematol.* 1997;25(4):293-7.
153. Sharma MB, Limaye LS, Kale VP. Mimicking the functional hematopoietic stem cell niche in vitro: recapitulation of marrow physiology by hydrogel-based three-dimensional cultures of mesenchymal stromal cells. *Haematologica.* 2012;97(5):651-60.
154. Yin T, Li L. The stem cell niches in bone. *The Journal of clinical investigation.* 2006;116(5):1195-201.
155. Kondo M, Weissman IL, Akashi K. Identification of clonogenic common lymphoid progenitors in mouse bone marrow. *Cell.* 1997;91(5):661-72.
156. Akashi K, Traver D, Miyamoto T, Weissman IL. A clonogenic common myeloid progenitor that gives rise to all myeloid lineages. *Nature.* 2000;404(6774):193-7.
157. Christensen JL, Weissman IL. Flk-2 is a marker in hematopoietic stem cell differentiation: a simple method to isolate long-term stem cells. *Proceedings of the National Academy of Sciences of the United States of America.* 2001;98(25):14541-6.

158. Morrison SJ, Wandycz AM, Hemmati HD, Wright DE, Weissman IL. Identification of a lineage of multipotent hematopoietic progenitors. *Development*. 1997;124(10):1929-39.
159. Arinobu Y, Iwasaki H, Gurish MF, Mizuno S, Shigematsu H, Ozawa H, et al. Developmental checkpoints of the basophil/mast cell lineages in adult murine hematopoiesis. *Proceedings of the National Academy of Sciences of the United States of America*. 2005;102(50):18105-10.
160. Fogg DK, Sibon C, Miled C, Jung S, Aucouturier P, Littman DR, et al. A clonogenic bone marrow progenitor specific for macrophages and dendritic cells. *Science (New York, NY)*. 2006;311(5757):83-7.
161. Ohmori K, Luo Y, Jia Y, Nishida J, Wang Z, Bunting KD, et al. IL-3 induces basophil expansion in vivo by directing granulocyte-monocyte progenitors to differentiate into basophil lineage-restricted progenitors in the bone marrow and by increasing the number of basophil/mast cell progenitors in the spleen. *Journal of immunology (Baltimore, Md : 1950)*. 2009;182(5):2835-41.
162. Miyamoto T, Weissman IL, Akashi K. AML1/ETO-expressing nonleukemic stem cells in acute myelogenous leukemia with 8;21 chromosomal translocation. *Proc Natl Acad Sci U S A*. 2000;97(13):7521-6.
163. Adolfsson J, Borge OJ, Bryder D, Theilgaard-Monch K, Astrand-Grundstrom I, Sitnicka E, et al. Upregulation of Flt3 expression within the bone marrow Lin(-)Sca1(+)-kit(+) stem cell compartment is accompanied by loss of self-renewal capacity. *Immunity*. 2001;15(4):659-69.
164. Forsberg EC, Serwold T, Kogan S, Weissman IL, Passegue E. New evidence supporting megakaryocyte-erythrocyte potential of flk2/flt3+ multipotent hematopoietic progenitors. *Cell*. 2006;126(2):415-26.
165. Arinobu Y, Mizuno S, Chong Y, Shigematsu H, Iino T, Iwasaki H, et al. Reciprocal activation of GATA-1 and PU.1 marks initial specification of hematopoietic stem cells into myeloerythroid and myelolymphoid lineages. *Cell Stem Cell*. 2007;1(4):416-27.
166. Paul F, Arkin Y, Giladi A, Jaitin DA, Kenigsberg E, Keren-Shaul H, et al. Transcriptional Heterogeneity and Lineage Commitment in Myeloid Progenitors. *Cell*. 2015;163(7):1663-77.
167. Notta F, Zandi S, Takayama N, Dobson S, Gan OI, Wilson G, et al. Distinct routes of lineage development reshape the human blood hierarchy across ontogeny. *Science*. 2016;351(6269):aab2116.
168. Drissen R, Buza-Vidas N, Woll P, Thongjuea S, Gambardella A, Giustacchini A, et al. Distinct myeloid progenitor-differentiation pathways identified through single-cell RNA sequencing. *Nat Immunol*. 2016;17(6):666-76.
169. Scott EW, Simon MC, Anastasi J, Singh H. Requirement of transcription factor PU.1 in the development of multiple hematopoietic lineages. *Science*. 1994;265(5178):1573-7.
170. Orkin SH, Shivdasani RA, Fujiwara Y, McDevitt MA. Transcription factor GATA-1 in megakaryocyte development. *Stem Cells*. 1998;16 Suppl 2:79-83.
171. Zhang P, Zhang X, Iwama A, Yu C, Smith KA, Mueller BU, et al. PU.1 inhibits GATA-1 function and erythroid differentiation by blocking GATA-1 DNA binding. *Blood*. 2000;96(8):2641-8.

172. Rao G, Rekhtman N, Cheng G, Krasikov T, Skoultschi AI. Deregulated expression of the PU.1 transcription factor blocks murine erythroleukemia cell terminal differentiation. *Oncogene*. 1997;14(1):123-31.
173. Rekhtman N, Radparvar F, Evans T, Skoultschi AI. Direct interaction of hematopoietic transcription factors PU.1 and GATA-1: functional antagonism in erythroid cells. *Genes & Development*. 1999;13(11):1398-411.
174. Hoppe PS, Schwarzfischer M, Loeffler D, Kokkaliaris KD, Hilsenbeck O, Moritz N, et al. Early myeloid lineage choice is not initiated by random PU.1 to GATA1 protein ratios. *Nature*. 2016;535(7611):299-302.
175. Starck J, Cohet N, Gonnet C, Sarrazin S, Doubeikovskaia Z, Doubeikovski A, et al. Functional cross-antagonism between transcription factors FLI-1 and EKLF. *Mol Cell Biol*. 2003;23(4):1390-402.
176. DeKoter RP, Singh H. Regulation of B lymphocyte and macrophage development by graded expression of PU.1. *Science*. 2000;288(5470):1439-41.
177. Leddin M, Perrod C, Hoogenkamp M, Ghani S, Assi S, Heinz S, et al. Two distinct auto-regulatory loops operate at the PU.1 locus in B cells and myeloid cells. *Blood*. 2011;117(10):2827-38.
178. Koschmieder S, Agrawal S, Radomska HS, Huettner CS, Tenen DG, Ottmann OG, et al. Decitabine and vitamin D3 differentially affect hematopoietic transcription factors to induce monocytic differentiation. *Int J Oncol*. 2007;30(2):349-55.
179. Zhang P, Iwasaki-Arai J, Iwasaki H, Fenyus ML, Dayaram T, Owens BM, et al. Enhancement of hematopoietic stem cell repopulating capacity and self-renewal in the absence of the transcription factor C/EBP alpha. *Immunity*. 2004;21(6):853-63.
180. Laslo P, Pongubala JM, Lancki DW, Singh H. Gene regulatory networks directing myeloid and lymphoid cell fates within the immune system. *Semin Immunol*. 2008;20(4):228-35.
181. Tamura T, Nagamura-Inoue T, Shmeltzer Z, Kuwata T, Ozato K. ICSBP directs bipotential myeloid progenitor cells to differentiate into mature macrophages. *Immunity*. 2000;13(2):155-65.
182. Kumar CC. Genetic abnormalities and challenges in the treatment of acute myeloid leukemia. *Genes Cancer*. 2011;2(2):95-107.
183. Vardiman JW, Thiele J, Arber DA, Brunning RD, Borowitz MJ, Porwit A, et al. The 2008 revision of the World Health Organization (WHO) classification of myeloid neoplasms and acute leukemia: rationale and important changes. *Blood*. 2009;114(5):937-51.
184. Osato M. Point mutations in the RUNX1/AML1 gene: another actor in RUNX leukemia. *Oncogene*. 2004;23(24):4284-96.
185. Liu P, Tarle SA, Hajra A, Claxton DF, Marlton P, Freedman M, et al. Fusion between transcription factor CBF beta/PEBP2 beta and a myosin heavy chain in acute myeloid leukemia. *Science*. 1993;261(5124):1041-4.
186. Pabst T, Mueller BU, Zhang P, Radomska HS, Narravula S, Schnittger S, et al. Dominant-negative mutations of CEBPA, encoding CCAAT/enhancer binding protein-alpha (C/EBPalpha), in acute myeloid leukemia. *Nat Genet*. 2001;27(3):263-70.



187. Calligaris R, Bottardi S, Cogoi S, Apezteguia I, Santoro C. Alternative translation initiation site usage results in two functionally distinct forms of the GATA-1 transcription factor. *Proc Natl Acad Sci U S A.* 1995;92(25):11598-602.
188. Li Z, Godinho FJ, Klusmann JH, Garriga-Canut M, Yu C, Orkin SH. Developmental stage-selective effect of somatically mutated leukemogenic transcription factor GATA1. *Nat Genet.* 2005;37(6):613-9.
189. Holtschke T, Lohler J, Kanno Y, Fehr T, Giese N, Rosenbauer F, et al. Immunodeficiency and chronic myelogenous leukemia-like syndrome in mice with a targeted mutation of the ICSBP gene. *Cell.* 1996;87(2):307-17.
190. Schwieger M, Lohler J, Friel J, Scheller M, Horak I, Stocking C. AML1-ETO inhibits maturation of multiple lymphohematopoietic lineages and induces myeloblast transformation in synergy with ICSBP deficiency. *J Exp Med.* 2002;196(9):1227-40.
191. Rosenbauer F, Wagner K, Kutok JL, Iwasaki H, Le Beau MM, Okuno Y, et al. Acute myeloid leukemia induced by graded reduction of a lineage-specific transcription factor, PU.1. *Nat Genet.* 2004;36(6):624-30.
192. Steidl U, Steidl C, Ebralidze A, Chapuy B, Han HJ, Will B, et al. A distal single nucleotide polymorphism alters long-range regulation of the PU.1 gene in acute myeloid leukemia. *J Clin Invest.* 2007;117(9):2611-20.
193. Antonson P, Xanthopoulos KG. Molecular cloning, sequence, and expression patterns of the human gene encoding CCAAT/enhancer binding protein alpha (C/EBP alpha). *Biochem Biophys Res Commun.* 1995;215(1):106-13.
194. Lin FT, MacDougald OA, Diehl AM, Lane MD. A 30-kDa alternative translation product of the CCAAT/enhancer binding protein alpha message: transcriptional activator lacking antimitotic activity. *Proceedings of the National Academy of Sciences of the United States of America.* 1993;90(20):9606-10.
195. Calkhoven CF, Muller C, Leutz A. Translational control of C/EBPalpha and C/EBPbeta isoform expression. *Genes Dev.* 2000;14(15):1920-32.
196. Mueller BU, Pabst T. C/EBPalpha and the pathophysiology of acute myeloid leukemia. *Curr Opin Hematol.* 2006;13(1):7-14.
197. Zhang DE, Zhang P, Wang ND, Hetherington CJ, Darlington GJ, Tenen DG. Absence of granulocyte colony-stimulating factor signaling and neutrophil development in CCAAT enhancer binding protein alpha-deficient mice. *Proc Natl Acad Sci U S A.* 1997;94(2):569-74.
198. Pabst T, Mueller BU, Harakawa N, Schoch C, Haferlach T, Behre G, et al. AML1-ETO downregulates the granulocytic differentiation factor C/EBPalpha in t(8;21) myeloid leukemia. *Nat Med.* 2001;7(4):444-51.
199. Ptasinska A, Assi SA, Mannari D, James SR, Williamson D, Dunne J, et al. Depletion of RUNX1/ETO in t(8;21) AML cells leads to genome-wide changes in chromatin structure and transcription factor binding. *Leukemia.* 2012;26(8):1829-41.
200. Avellino R, Havermans M, Erpelinck C, Sanders MA, Hoogenboezem R, van de Werken HJ, et al. An autonomous CEBPA enhancer specific for myeloid-lineage priming and neutrophilic differentiation. *Blood.* 2016;127(24):2991-3003.
201. Radomska HS, Basseres DS, Zheng R, Zhang P, Dayaram T, Yamamoto Y, et al. Block of C/EBP alpha function by phosphorylation in acute myeloid leukemia with FLT3 activating mutations. *J Exp Med.* 2006;203(2):371-81.

202. Zhang H, Alberich-Jorda M, Amabile G, Yang H, Staber PB, Di Ruscio A, et al. Sox4 is a key oncogenic target in C/EBPalpha mutant acute myeloid leukemia. *Cancer cell*. 2013;24(5):575-88.
203. Kreso A, Dick John E. Evolution of the Cancer Stem Cell Model. *Cell Stem Cell*.14(3):275-91.
204. Martens JH, Mandoli A, Simmer F, Wierenga BJ, Saeed S, Singh AA, et al. ERG and FLI1 binding sites demarcate targets for aberrant epigenetic regulation by AML1-ETO in acute myeloid leukemia. *Blood*. 2012;120(19):4038-48.
205. Rowley JD. Identificaton of a translocation with quinacrine fluorescence in a patient with acute leukemia. *Ann Genet*. 1973;16(2):109-12.
206. Follows GA, Tagoh H, Lefevre P, Hodge D, Morgan GJ, Bonifer C. Epigenetic consequences of AML1-ETO action at the human c-FMS locus. *The EMBO journal*. 2003;22(11):2798-809.
207. Mrózek K, Marcucci G, Paschka P, Bloomfield CD. Advances in molecular genetics and treatment of core-binding factor acute myeloid leukemia. *Current opinion in oncology*. 2008;20(6):711-8.
208. Wang J, Hoshino T, Redner RL, Kajigaya S, Liu JM. ETO, fusion partner in t(8;21) acute myeloid leukemia, represses transcription by interaction with the human N-CoR/mSin3/HDAC1 complex. *Proceedings of the National Academy of Sciences of the United States of America*. 1998;95(18):10860-5.
209. Nimer SD, Moore MA. Effects of the leukemia-associated AML1-ETO protein on hematopoietic stem and progenitor cells. *Oncogene*. 2004;23(24):4249-54.
210. Moldenhauer A, Frank RC, Pinilla-Ibarz J, Holland G, Bocconi P, Scheinberg DA, et al. Histone deacetylase inhibition improves dendritic cell differentiation of leukemic blasts with AML1-containing fusion proteins. *J Leukoc Biol*. 2004;76(3):623-33.
211. Wiemels JL, Xiao Z, Buffler PA, Maia AT, Ma X, Dicks BM, et al. In utero origin of t(8;21) AML1-ETO translocations in childhood acute myeloid leukemia. *Blood*. 2002;99(10):3801-5.
212. Rhoades KL, Hetherington CJ, Harakawa N, Yergeau DA, Zhou L, Liu LQ, et al. Analysis of the role of AML1-ETO in leukemogenesis, using an inducible transgenic mouse model. *Blood*. 2000;96(6):2108-15.
213. Yuan Y, Zhou L, Miyamoto T, Iwasaki H, Harakawa N, Hetherington CJ, et al. AML1-ETO expression is directly involved in the development of acute myeloid leukemia in the presence of additional mutations. *Proc Natl Acad Sci U S A*. 2001;98(18):10398-403.
214. Cabezas-Wallscheid N, Eichwald V, de Graaf J, Lower M, Lehr HA, Kreft A, et al. Instruction of haematopoietic lineage choices, evolution of transcriptional landscapes and cancer stem cell hierarchies derived from an AML1-ETO mouse model. *EMBO Mol Med*. 2013;5(12):1804-20.
215. Sakurai M, Oshimura M, Kakati S, Sandberg AA. Letter: 8-21 translocation and missing sex chromosomes in acute leukaemia. *Lancet*. 1974;2(7874):227-8.
216. Peterson LF, Boyapati A, Ahn EY, Biggs JR, Okumura AJ, Lo MC, et al. Acute myeloid leukemia with the 8q22;21q22 translocation: secondary mutational events and alternative t(8;21) transcripts. *Blood*. 2007;110(3):799-805.
217. Hatlen MA, Wang L, Nimer SD. AML1-ETO driven acute leukemia: insights into pathogenesis and potential therapeutic approaches. *Front Med*. 2012;6(3):248-62.

218. Miyoshi H, Kozu T, Shimizu K, Enomoto K, Maseki N, Kaneko Y, et al. The t(8;21) translocation in acute myeloid leukemia results in production of an AML1-MTG8 fusion transcript. *The EMBO Journal*. 1993;12(7):2715-21.
219. Ptasińska A, Assi SA, Martínez-Soria N, Imperato MR, Piper J, Cauchy P, et al. Identification of a dynamic core transcriptional network in t(8;21) AML that regulates differentiation block and self-renewal. *Cell reports*. 2014;8(6):1974-88.
220. Kagoshima H, Shigesada K, Satake M, Ito Y, Miyoshi H, Ohki M, et al. The Runt domain identifies a new family of heteromeric transcriptional regulators. *Trends Genet*. 1993;9(10):338-41.
221. Wang S, Wang Q, Crute BE, Melnikova IN, Keller SR, Speck NA. Cloning and characterization of subunits of the T-cell receptor and murine leukemia virus enhancer core-binding factor. *Mol Cell Biol*. 1993;13(6):3324-39.
222. North T, Gu TL, Stacy T, Wang Q, Howard L, Binder M, et al. Cbfa2 is required for the formation of intra-aortic hematopoietic clusters. *Development*. 1999;126(11):2563-75.
223. Kitabayashi I, Yokoyama A, Shimizu K, Ohki M. Interaction and functional cooperation of the leukemia-associated factors AML1 and p300 in myeloid cell differentiation. *EMBO J*. 1998;17(11):2994-3004.
224. Imai Y, Kurokawa M, Yamaguchi Y, Izutsu K, Nitta E, Mitani K, et al. The corepressor mSin3A regulates phosphorylation-induced activation, intranuclear location, and stability of AML1. *Mol Cell Biol*. 2004;24(3):1033-43.
225. Medvinsky A, Dzierzak E. Definitive hematopoiesis is autonomously initiated by the AGM region. *Cell*. 1996;86(6):897-906.
226. Okuda T, van Deursen J, Hiebert SW, Grosveld G, Downing JR. AML1, the target of multiple chromosomal translocations in human leukemia, is essential for normal fetal liver hematopoiesis. *Cell*. 1996;84(2):321-30.
227. Chen MJ, Yokomizo T, Zeigler BM, Dzierzak E, Speck NA. Runx1 is required for the endothelial to haematopoietic cell transition but not thereafter. *Nature*. 2009;457(7231):887-91.
228. Gowney JD, Shigematsu H, Li Z, Lee BH, Adelsperger J, Rowan R, et al. Loss of Runx1 perturbs adult hematopoiesis and is associated with a myeloproliferative phenotype. *Blood*. 2005;106(2):494-504.
229. Huang G, Zhang P, Hirai H, Elf S, Yan X, Chen Z, et al. PU.1 is a major downstream target of AML1 (RUNX1) in adult mouse hematopoiesis. *Nature genetics*. 2008;40(1):51-60.
230. McKercher SR, Torbett BE, Anderson KL, Henkel GW, Vestal DJ, Baribault H, et al. Targeted disruption of the PU.1 gene results in multiple hematopoietic abnormalities. *The EMBO journal*. 1996;15(20):5647-58.
231. Meyers S, Downing JR, Hiebert SW. Identification of AML-1 and the (8;21) translocation protein (AML-1/ETO) as sequence-specific DNA-binding proteins: the runt homology domain is required for DNA binding and protein-protein interactions. *Molecular and cellular biology*. 1993;13(10):6336-45.
232. Yergeau DA, Hetherington CJ, Wang Q, Zhang P, Sharpe AH, Binder M, et al. Embryonic lethality and impairment of haematopoiesis in mice heterozygous for an AML1-ETO fusion gene. *Nature genetics*. 1997;15(3):303-6.

233. Okuda T, Cai Z, Yang S, Lenny N, Lyu CJ, van Deursen JM, et al. Expression of a knocked-in AML1-ETO leukemia gene inhibits the establishment of normal definitive hematopoiesis and directly generates dysplastic hematopoietic progenitors. *Blood*. 1998;91(9):3134-43.
234. Hildebrand D, Tiefenbach J, Heinzl T, Grez M, Maurer AB. Multiple regions of ETO cooperate in transcriptional repression. *The Journal of biological chemistry*. 2001;276(13):9889-95.
235. Wang J, Sauntharajah Y, Redner RL, Liu JM. Inhibitors of histone deacetylase relieve ETO-mediated repression and induce differentiation of AML1-ETO leukemia cells. *Cancer research*. 1999;59(12):2766-9.
236. Shimada H, Ichikawa H, Nakamura S, Katsu R, Iwasa M, Kitabayashi I, et al. Analysis of genes under the downstream control of the t(8;21) fusion protein AML1-MTG8: overexpression of the TIS11b (ERF-1, cMG1) gene induces myeloid cell proliferation in response to G-CSF. *Blood*. 2000;96(2):655-63.
237. Yan M, Burel SA, Peterson LF, Kanbe E, Iwasaki H, Boyapati A, et al. Deletion of an AML1-ETO C-terminal NcoR/SMRT-interacting region strongly induces leukemia development. *Proceedings of the National Academy of Sciences of the United States of America*. 2004;101(49):17186-91.
238. Link KA, Lin S, Shrestha M, Bowman M, Wunderlich M, Bloomfield CD, et al. Supraphysiologic levels of the AML1-ETO isoform AE9a are essential for transformation. *Proc Natl Acad Sci U S A*. 2016;113(32):9075-80.
239. Amann JM, Nip J, Strom DK, Lutterbach B, Harada H, Lenny N, et al. ETO, a target of t(8;21) in acute leukemia, makes distinct contacts with multiple histone deacetylases and binds mSin3A through its oligomerization domain. *Mol Cell Biol*. 2001;21(19):6470-83.
240. Grisolan JL, O'Neal J, Cain J, Tomasson MH. An activated receptor tyrosine kinase, TEL/PDGFBetaR, cooperates with AML1/ETO to induce acute myeloid leukemia in mice. *Proceedings of the National Academy of Sciences of the United States of America*. 2003;100(16):9506-11.
241. Vangala RK, Heiss-Neumann MS, Rangatia JS, Singh SM, Schoch C, Tenen DG, et al. The myeloid master regulator transcription factor PU.1 is inactivated by AML1-ETO in t(8;21) myeloid leukemia. *Blood*. 2003;101(1):270-7.
242. Pabst T, Mueller BU, Harakawa N, Schoch C, Haferlach T, Behre G, et al. AML1-ETO downregulates the granulocytic differentiation factor C/EBPalpha in t(8;21) myeloid leukemia. *Nature medicine*. 2001;7(4):444-51.
243. Choi Y, Elagib KE, Delehanty LL, Goldfarb AN. Erythroid inhibition by the leukemic fusion AML1-ETO is associated with impaired acetylation of the major erythroid transcription factor GATA-1. *Cancer research*. 2006;66(6):2990-6.
244. Zhang J, Kalkum M, Yamamura S, Chait BT, Roeder RG. E protein silencing by the leukemogenic AML1-ETO fusion protein. *Science (New York, NY)*. 2004;305(5688):1286-9.
245. Saeed S, Logie C, Francoijs KJ, Frige G, Romanenghi M, Nielsen FG, et al. Chromatin accessibility, p300, and histone acetylation define PML-RARalpha and AML1-ETO binding sites in acute myeloid leukemia. *Blood*. 2012;120(15):3058-68.

246. Wang L, Gural A, Sun XJ, Zhao X, Perna F, Huang G, et al. The leukemogenicity of AML1-ETO is dependent on site-specific lysine acetylation. *Science (New York, NY)*. 2011;333(6043):765-9.
247. Maiques-Diaz A, Chou FS, Wunderlich M, Gomez-Lopez G, Jacinto FV, Rodriguez-Perales S, et al. Chromatin modifications induced by the AML1-ETO fusion protein reversibly silence its genomic targets through AML1 and Sp1 binding motifs. *Leukemia*. 2012;26(6):1329-37.
248. Minucci S, Maccarana M, Ciocce M, De Luca P, Gelmetti V, Segalla S, et al. Oligomerization of RAR and AML1 transcription factors as a novel mechanism of oncogenic activation. *Mol Cell*. 2000;5(5):811-20.
249. Liu Y, Cheney MD, Gaudet JJ, Chruszcz M, Lukasik SM, Sugiyama D, et al. The tetramer structure of the Nery homology two domain, NHR2, is critical for AML1/ETO's activity. *Cancer Cell*. 2006;9(4):249-60.
250. Kwok C, Zeisig BB, Qiu J, Dong S, So CW. Transforming activity of AML1-ETO is independent of CBFbeta and ETO interaction but requires formation of homo-oligomeric complexes. *Proc Natl Acad Sci U S A*. 2009;106(8):2853-8.
251. Yan M, Ahn EY, Hiebert SW, Zhang DE. RUNX1/AML1 DNA-binding domain and ETO/MTG8 NHR2-dimerization domain are critical to AML1-ETO9a leukemogenesis. *Blood*. 2009;113(4):883-6.
252. Okumura AJ, Peterson LF, Okumura F, Boyapati A, Zhang DE. t(8;21)(q22;q22) Fusion proteins preferentially bind to duplicated AML1/RUNX1 DNA-binding sequences to differentially regulate gene expression. *Blood*. 2008;112(4):1392-401.
253. Elagib KE, Racke FK, Mogass M, Khetawat R, Delehanty LL, Goldfarb AN. RUNX1 and GATA-1 coexpression and cooperation in megakaryocytic differentiation. *Blood*. 2003;101(11):4333-41.
254. Zhang J, Kalkum M, Yamamura S, Chait BT, Roeder RG. E protein silencing by the leukemogenic AML1-ETO fusion protein. *Science*. 2004;305(5688):1286-9.
255. Guo C, Hu Q, Yan C, Zhang J. Multivalent binding of the ETO corepressor to E proteins facilitates dual repression controls targeting chromatin and the basal transcription machinery. *Mol Cell Biol*. 2009;29(10):2644-57.
256. Sun XJ, Wang Z, Wang L, Jiang Y, Kost N, Soong TD, et al. A stable transcription factor complex nucleated by oligomeric AML1-ETO controls leukaemogenesis. *Nature*. 2013;500(7460):93-7.
257. Dunne J, Cullmann C, Ritter M, Soria NM, Drescher B, Debernardi S, et al. siRNA-mediated AML1/MTG8 depletion affects differentiation and proliferation-associated gene expression in t(8;21)-positive cell lines and primary AML blasts. *Oncogene*. 2006;25(45):6067-78.
258. Martinez Soria N, Tussiwand R, Ziegler P, Manz MG, Heidenreich O. Transient depletion of RUNX1/RUNX1T1 by RNA interference delays tumour formation in vivo. *Leukemia*. 2009;23(1):188-90.
259. Wilson NK, Foster SD, Wang X, Knezevic K, Schutte J, Kaimakis P, et al. Combinatorial transcriptional control in blood stem/progenitor cells: genome-wide analysis of ten major transcriptional regulators. *Cell stem cell*. 2010;7(4):532-44.

260. Gilmour J, Assi SA, Jaegle U, Kulu D, van de Werken H, Clarke D, et al. A crucial role for the ubiquitously expressed transcription factor Sp1 at early stages of hematopoietic specification. *Development*. 2014;141(12):2391-401.
261. Novershtern N, Subramanian A, Lawton LN, Mak RH, Haining WN, McConkey ME, et al. Densely interconnected transcriptional circuits control cell states in human hematopoiesis. *Cell*. 2011;144(2):296-309.
262. Wei H, Liu X, Xiong X, Wang Y, Rao Q, Wang M, et al. AML1-ETO interacts with Sp1 and antagonizes Sp1 transactivity through RUNT domain. *FEBS Lett*. 2008;582(15):2167-72.
263. Mahon FX, Rea D, Guilhot J, Guilhot F, Huguette F, Nicolini F, et al. Discontinuation of imatinib in patients with chronic myeloid leukaemia who have maintained complete molecular remission for at least 2 years: the prospective, multicentre Stop Imatinib (STIM) trial. *Lancet Oncol*. 2010;11(11):1029-35.
264. Maiques-Diaz A, Hernando M, Sanchez-Lopez A, Rio-Machin A, Shrestha M, Mulloy JC, et al. MAPK8-mediated stabilization of SP1 is essential for RUNX1-RUNX1T1 - driven leukaemia. *Br J Haematol*. 2016;172(5):807-10.
265. Ben-Ami O, Friedman D, Leshkowitz D, Goldenberg D, Orlovsky K, Pencovich N, et al. Addiction of t(8;21) and inv(16) acute myeloid leukemia to native RUNX1. *Cell reports*. 2013;4(6):1131-43.
266. Goyama S, Schibler J, Cunningham L, Zhang Y, Rao Y, Nishimoto N, et al. Transcription factor RUNX1 promotes survival of acute myeloid leukemia cells. *J Clin Invest*. 2013;123(9):3876-88.
267. Schnittger S, Dicker F, Kern W, Wendland N, Sundermann J, Alpermann T, et al. RUNX1 mutations are frequent in de novo AML with noncomplex karyotype and confer an unfavorable prognosis. *Blood*. 2011;117(8):2348-57.
268. Yan J, Liu Y, Lukasik SM, Speck NA, Bushweller JH. CBFbeta allosterically regulates the Runx1 Runt domain via a dynamic conformational equilibrium. *Nat Struct Mol Biol*. 2004;11(9):901-6.
269. Barton JL, Bunka DH, Knowling SE, Lefevre P, Warren AJ, Bonifer C, et al. Characterization of RNA aptamers that disrupt the RUNX1-CBFbeta/DNA complex. *Nucleic Acids Res*. 2009;37(20):6818-30.
270. Wang Q, Stacy T, Miller JD, Lewis AF, Gu TL, Huang X, et al. The CBFbeta subunit is essential for CBFalpha2 (AML1) function in vivo. *Cell*. 1996;87(4):697-708.
271. Roudaia L, Cheney MD, Manuylova E, Chen W, Morrow M, Park S, et al. CBFbeta is critical for AML1-ETO and TEL-AML1 activity. *Blood*. 2009;113(13):3070-9.
272. Yan C, Higgins PJ. Drugging the undruggable: transcription therapy for cancer. *Biochim Biophys Acta*. 2013;1835(1):76-85.
273. Johnston SJ, Carroll JS. Transcription factors and chromatin proteins as therapeutic targets in cancer. *Biochim Biophys Acta*. 2015;1855(2):183-92.
274. Zhang L, Li Z, Yan J, Pradhan P, Corpora T, Cheney MD, et al. Mutagenesis of the Runt domain defines two energetic hot spots for heterodimerization with the core binding factor beta subunit. *J Biol Chem*. 2003;278(35):33097-104.
275. Gorczynski MJ, Grembecka J, Zhou Y, Kong Y, Roudaia L, Douvas MG, et al. Allosteric inhibition of the protein-protein interaction between the leukemia-associated proteins Runx1 and CBFbeta. *Chem Biol*. 2007;14(10):1186-97.

276. Illendula A, Pulikkan JA, Zong H, Grembecka J, Xue L, Sen S, et al. A small-molecule inhibitor of the aberrant transcription factor CBF $\beta$ -SMMHC delays leukemia in mice. *Science*. 2015;347(6223):779-84.
277. Bhagwat AS, Vakoc CR. Targeting Transcription Factors in Cancer. *Trends Cancer*. 2015;1(1):53-65.
278. Illendula A, Gilmour J, Grembecka J, Tirumala VSS, Boulton A, Kuntimaddi A, et al. Small Molecule Inhibitor of CBF $\beta$ -RUNX Binding for RUNX Transcription Factor Driven Cancers. *EBioMedicine*.8:117-31.
279. Palstra RJ, Tolhuis B, Splinter E, Nijmeijer R, Grosveld F, de Laat W. The beta-globin nuclear compartment in development and erythroid differentiation. *Nature genetics*. 2003;35(2):190-4.
280. Tolhuis B, Palstra RJ, Splinter E, Grosveld F, de Laat W. Looping and interaction between hypersensitive sites in the active beta-globin locus. *Molecular cell*. 2002;10(6):1453-65.
281. Deng W, Lee J, Wang H, Miller J, Reik A, Gregory PD, et al. Controlling long-range genomic interactions at a native locus by targeted tethering of a looping factor. *Cell*. 2012;149(6):1233-44.
282. van de Werken HJ, Landan G, Holwerda SJ, Hoichman M, Klous P, Chachik R, et al. Robust 4C-seq data analysis to screen for regulatory DNA interactions. *Nature methods*. 2012;9(10):969-72.
283. Wingett S, Ewels P, Furlan-Magaril M, Nagano T, Schoenfelder S, Fraser P, et al. HiCUP: pipeline for mapping and processing Hi-C data. *F1000Research*. 2015;4:1310.
284. Mifsud B, Martincorena I, Darbo E, Sugar R, Schoenfelder S, Fraser P, et al. GOTHIC, a simple probabilistic model to resolve complex biases and to identify real interactions in Hi-C data. *bioRxiv*. 2015.
285. Heinz S, Benner C, Spann N, Bertolino E, Lin YC, Laslo P, et al. Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol Cell*. 2010;38(4):576-89.
286. Shivdasani RA, Rosenblatt MF, Zucker-Franklin D, Jackson CW, Hunt P, Saris CJ, et al. Transcription factor NF-E2 is required for platelet formation independent of the actions of thrombopoietin/MGDF in megakaryocyte development. *Cell*. 1995;81(5):695-704.
287. Stanley ER, Berg KL, Einstein DB, Lee PS, Pixley FJ, Wang Y, et al. Biology and action of colony-stimulating factor-1. *Mol Reprod Dev*. 1997;46(1):4-10.
288. Heesch S, Schlee C, Neumann M, Stroux A, Kuhn A, Schwartz S, et al. BAALC-associated gene expression profiles define IGFBP7 as a novel molecular marker in acute leukemia. *Leukemia*. 2010;24(8):1429-36.
289. Wierstra I. Sp1: emerging roles--beyond constitutive activation of TATA-less housekeeping genes. *Biochem Biophys Res Commun*. 2008;372(1):1-13.
290. Rapino F, Robles EF, Richter-Larrea JA, Kallin EM, Martinez-Climent JA, Graf T. C/EBP $\alpha$  induces highly efficient macrophage transdifferentiation of B lymphoma and leukemia cell lines and impairs their tumorigenicity. *Cell Rep*. 2013;3(4):1153-63.
291. Arora D, Kothe S, van den Eijnden M, Hoof van Huijsduijnen R, Heidel F, Fischer T, et al. Expression of protein-tyrosine phosphatases in Acute Myeloid Leukemia cells: FLT3 ITD sustains high levels of DUSP6 expression. *Cell Commun Signal*. 2012;10(1):19.

292. Okuno Y, Huettner CS, Radomska HS, Petkova V, Iwasaki H, Akashi K, et al. Distal elements are critical for human CD34 expression in vivo. *Blood*. 2002;100(13):4420-6.
293. Ebralidze AK, Guibal FC, Steidl U, Zhang P, Lee S, Bartholdy B, et al. PU.1 expression is modulated by the balance of functional sense and antisense RNAs regulated by a shared cis-regulatory element. *Genes & development*. 2008;22(15):2085-92.
294. Li Y, Okuno Y, Zhang P, Radomska HS, Chen H, Iwasaki H, et al. Regulation of the PU.1 gene by distal elements. *Blood*. 2001;98(10):2958-65.
295. Staber PB, Zhang P, Ye M, Welner RS, Levantini E, Di Ruscio A, et al. The Runx-PU.1 pathway preserves normal and AML/ETO9a leukemic stem cells. *Blood*. 2014;124(15):2391-9.
296. Rickman DS, Soong TD, Moss B, Mosquera JM, Dlabal J, Terry S, et al. Oncogene-mediated alterations in chromatin conformation. *Proceedings of the National Academy of Sciences of the United States of America*. 2012;109(23):9083-8.
297. Zhang Y, McCord RP, Ho Y-J, Lajoie BR, Hildebrand DG, Simon AC, et al. Chromosomal translocations are guided by the spatial organization of the genome. *Cell*. 2012;148(5):908-21.
298. Moog-Lutz C, Peterson EJ, Lutz PG, Eliason S, Cave-Riant F, Singer A, et al. PRAM-1 is a novel adaptor protein regulated by retinoic acid (RA) and promyelocytic leukemia (PML)-RA receptor alpha in acute promyelocytic leukemia cells. *J Biol Chem*. 2001;276(25):22375-81.
299. Sako D, Comess KM, Barone KM, Camphausen RT, Cumming DA, Shaw GD. A sulfated peptide segment at the amino terminus of PSGL-1 is critical for P-selectin binding. *Cell*. 1995;83(2):323-31.
300. Raja SM, Wang B, Dantuluri M, Desai UR, Demeler B, Spiegel K, et al. Cytotoxic cell granule-mediated apoptosis. Characterization of the macromolecular complex of granzyme B with serglycin. *J Biol Chem*. 2002;277(51):49523-30.
301. Phillips JE, Corces VG. CTCF: master weaver of the genome. *Cell*. 2009;137(7):1194-211.
302. Piper J, Elze MC, Cauchy P, Cockerill PN, Bonifer C, Ott S. Wellington: a novel method for the accurate identification of digital genomic footprints from DNase-seq data. *Nucleic Acids Res*. 2013;41(21):e201.
303. Bosisio D, Marazzi I, Agresti A, Shimizu N, Bianchi ME, Natoli G. A hyper-dynamic equilibrium between promoter-bound and nucleoplasmic dimers controls NF-kappaB-dependent gene activity. *EMBO J*. 2006;25(4):798-810.
304. Karpova TS, Kim MJ, Spriet C, Nalley K, Stasevich TJ, Kherrouche Z, et al. Concurrent fast and slow cycling of a transcriptional activator at an endogenous promoter. *Science*. 2008;319(5862):466-9.
305. Chen J, Zhang Z, Li L, Chen BC, Revyakin A, Hajj B, et al. Single-molecule dynamics of enhanceosome assembly in embryonic stem cells. *Cell*. 2014;156(6):1274-85.
306. Lorsbach RB, Moore J, Ang SO, Sun W, Lenny N, Downing JR. Role of RUNX1 in adult hematopoiesis: analysis of RUNX1-IRES-GFP knock-in mice reveals differential lineage expression. *Blood*. 2004;103(7):2522-9.
307. Kanai M, Wei D, Li Q, Jia Z, Ajani J, Le X, et al. Loss of Krüppel-Like Factor 4 Expression Contributes to Sp1 Overexpression and Human Gastric Cancer Development and Progression. *American Association for Cancer Research*. 2006;12(21):6395-402.



308. Tonks A, Pearn L, Musson M, Gilkes A, Mills KI, Burnett AK, et al. Transcriptional dysregulation mediated by RUNX1-RUNX1T1 in normal human progenitor cells and in acute myeloid leukaemia. *Leukemia*. 2007;21(12):2495-505.
309. Alcalay M, Meani N, Gelmetti V, Fantozzi A, Fagioli M, Orleth A, et al. Acute myeloid leukemia fusion proteins deregulate genes involved in stem cell maintenance and DNA repair. *J Clin Invest*. 2003;112(11):1751-61.
310. Mulloy JC, Cammenga J, Berguido FJ, Wu K, Zhou P, Comenzo RL, et al. Maintaining the self-renewal and differentiation potential of human CD34+ hematopoietic cells using a single genetic element. *Blood*. 2003;102(13):4369-76.
311. Williams SC, Du Y, Schwartz RC, Weiler SR, Ortiz M, Keller JR, et al. C/EBPepsilon is a myeloid-specific activator of cytokine, chemokine, and macrophage-colony-stimulating factor receptor genes. *J Biol Chem*. 1998;273(22):13493-501.
312. Jones LC, Lin ML, Chen SS, Krug U, Hofmann WK, Lee S, et al. Expression of C/EBPbeta from the C/ebpalpha gene locus is sufficient for normal hematopoiesis in vivo. *Blood*. 2002;99(6):2032-6.
313. Jackson AL, Burchard J, Schelter J, Chau BN, Cleary M, Lim L, et al. Widespread siRNA "off-target" transcript silencing mediated by seed region sequence complementarity. *RNA*. 2006;12(7):1179-87.
314. Dickinson RJ, Keyse SM. Diverse physiological functions for dual-specificity MAP kinase phosphatases. *J Cell Sci*. 2006;119(Pt 22):4607-15.
315. Reddy VA, Iwama A, Iotzova G, Schulz M, Elsasser A, Vangala RK, et al. Granulocyte inducer C/EBPalpha inactivates the myeloid master regulator PU.1: possible role in lineage commitment decisions. *Blood*. 2002;100(2):483-90.
316. Ohlsson E, Hasemann MS, Willer A, Lauridsen FKB, Rapin N, Jendholm J, et al. Initiation of MLL-rearranged AML is dependent on C/EBPalpha. *The Journal of Experimental Medicine*. 2014;211(1):5-13.
317. Pedersen TA, Kowenz-Leutz E, Leutz A, Nerlov C. Cooperation between C/EBPalpha TBP/TFIIB and SWI/SNF recruiting domains is required for adipocyte differentiation. *Genes Dev*. 2001;15(23):3208-16.
318. Heinzl T, Lavinsky RM, Mullen TM, Soderstrom M, Laherty CD, Torchia J, et al. A complex containing N-CoR, mSin3 and histone deacetylase mediates transcriptional repression. *Nature*. 1997;387(6628):43-8.
319. Fazi F, Racanicchi S, Zardo G, Starnes LM, Mancini M, Travaglini L, et al. Epigenetic silencing of the myelopoiesis regulator microRNA-223 by the AML1/ETO oncoprotein. *Cancer cell*. 2007;12(5):457-66.
320. Ghavi-Helm Y, Klein FA, Pakozdi T, Ciglar L, Noordermeer D, Huber W, et al. Enhancer loops appear stable during development and are associated with paused polymerase. *Nature*. 2014;512(7512):96-100.
321. Kondo T, Isono K, Kondo K, Endo TA, Itohara S, Vidal M, et al. Polycomb potentiates meis2 activation in midbrain by mediating interaction of the promoter with a tissue-specific enhancer. *Dev Cell*. 2014;28(1):94-101.
322. Schoenfelder S, Sugar R, Dimond A, Javierre BM, Armstrong H, Mifsud B, et al. Polycomb repressive complex PRC1 spatially constrains the mouse embryonic stem cell genome. *Nat Genet*. 2015;47(10):1179-86.

323. Choukrallah MA, Song S, Rolink AG, Burger L, Matthias P. Enhancer repertoires are reshaped independently of early priming and heterochromatin dynamics during B cell differentiation. *Nat Commun.* 2015;6:8324.
324. Alberich-Jorda M, Wouters B, Balastik M, Shapiro-Koss C, Zhang H, Di Ruscio A, et al. C/EBPgamma deregulation results in differentiation arrest in acute myeloid leukemia. *J Clin Invest.* 2012;122(12):4490-504.
325. Bell AC, Felsenfeld G. Methylation of a CTCF-dependent boundary controls imprinted expression of the Igf2 gene. *Nature.* 2000;405(6785):482-5.
326. Wang H, Maurano MT, Qu H, Varley KE, Gertz J, Pauli F, et al. Widespread plasticity in CTCF occupancy linked to DNA methylation. *Genome Research.* 2012;22(9):1680-8.
327. Lefevre P, Witham J, Lacroix CE, Cockerill PN, Bonifer C. The LPS-induced transcriptional upregulation of the chicken lysozyme locus involves CTCF eviction and noncoding RNA transcription. *Mol Cell.* 2008;32(1):129-39.
328. Schmidt D, Schwalie PC, Ross-Innes CS, Hurtado A, Brown GD, Carroll JS, et al. A CTCF-independent role for cohesin in tissue-specific transcription. *Genome Res.* 2010;20(5):578-88.
329. Rollins RA, Korom M, Aulner N, Martens A, Dorsett D. Drosophila nipped-B protein supports sister chromatid cohesion and opposes the stromalin/Scs3 cohesion factor to facilitate long-range activation of the cut gene. *Mol Cell Biol.* 2004;24(8):3100-11.
330. Dorsett D, Eissenberg JC, Misulovin Z, Martens A, Redding B, McKim K. Effects of sister chromatid cohesion proteins on cut gene expression during wing development in Drosophila. *Development.* 2005;132(21):4743-53.
331. Song SH, Hou C, Dean A. A positive role for NLI/Ldb1 in long-range beta-globin locus control region function. *Mol Cell.* 2007;28(5):810-22.
332. Krylov D, Olive M, Vinson C. Extending dimerization interfaces: the bZIP basic region can form a coiled coil. *EMBO J.* 1995;14(21):5329-37.
333. Creighton MP, Cheng AW, Welstead GG, Kooistra T, Carey BW, Steine EJ, et al. Histone H3K27ac separates active from poised enhancers and predicts developmental state. *Proc Natl Acad Sci U S A.* 2010;107(50):21931-6.
334. Bogdanovic O, Fernandez-Minan A, Tena JJ, de la Calle-Mustienes E, Hidalgo C, van Kruysbergen I, et al. Dynamics of enhancer chromatin signatures mark the transition from pluripotency to cell specification during embryogenesis. *Genome Res.* 2012;22(10):2043-53.
335. Parelho V, Hadjur S, Spivakov M, Leleu M, Sauer S, Gregson HC, et al. Cohesins functionally associate with CTCF on mammalian chromosome arms. *Cell.* 2008;132(3):422-33.

## APPENDICES

**Table 1: Genes differentially expressed by both C/EBP $\alpha$  overexpression and RUNX1/ETO knockdown:**

Gene Name	Control Kasumi-1 minus E2 (fpkm)	Control Kasumi-1 plus E2 (fpkm)	Kasumi-1-C/EBP $\alpha$ ER minus E2 (fpkm)	Kasumi-1-C/EBP $\alpha$ ER plus E2 (fpkm)	Kasumi-1 siMM (fpkm)	Kasumi-1 siRE (fpkm)
ATF3	3.38	3.00	3.00	1.88	5.64	2.02
GUCY1A3	10.89	9.00	12.00	5.64	5.75	1.02
MYO1B	6.14	5.00	5.00	2.90	5.83	3.27
IRF1	5.39	5.00	5.00	2.40	5.86	2.99
TRIM47	2.70	3.00	2.00	1.29	6.15	2.31
ZNF436	7.43	6.00	5.00	3.35	6.24	3.17
ETV5	6.74	6.00	5.00	2.42	6.89	2.86
PLOD2	3.98	3.00	5.00	1.61	7.33	2.68
PIK3C2B	3.14	3.00	3.00	2.10	7.94	3.31
CCND1	4.30	3.00	4.00	2.16	8.06	1.44
SVOPL	3.34	2.00	3.00	1.61	8.93	2.76
MOV10	7.38	7.00	7.00	4.57	9.25	3.26
GPRC5C	7.33	8.00	5.00	0.52	9.65	1.68
FBXW9	3.13	4.00	2.00	1.90	10.01	5.73
CHAC1	8.38	13.00	7.00	1.91	10.41	5.34
NCKIPSD	5.40	6.00	4.00	2.28	11.02	6.49
NAV1	4.08	4.00	3.00	2.33	11.65	3.22
ZFP36L1	2.86	3.00	2.00	1.20	11.91	3.72
CA8	6.52	6.00	5.00	2.62	12.98	5.50
SLMO2-ATP5E	21.94	22.00	17.00	9.77	13.33	5.56
HDAC7	5.11	5.00	4.00	3.32	13.87	6.15
STC2	9.00	10.00	6.00	2.66	14.22	7.55
APOBEC3G	13.62	10.00	15.00	3.14	14.45	3.09
GATA2	4.88	5.00	4.00	2.57	15.50	3.70
SLC45A3	5.43	6.00	6.00	3.45	18.05	2.80
ANGPT1	20.06	20.00	21.00	10.35	20.01	2.04
DHRS3	20.53	21.00	17.00	7.65	20.21	6.64
JUN	4.25	5.00	2.00	1.52	21.30	10.39
DDIT4	29.65	35.00	10.00	2.53	21.55	11.69
ST18	14.29	13.00	15.00	5.84	21.64	2.36

Gene Name	Control Kasumi-1 minus E2 (fpkm)	Control Kasumi-1 plus E2 (fpkm)	Kasumi-1-C/EBP $\alpha$ ER minus E2 (fpkm)	Kasumi-1-C/EBP $\alpha$ ER plus E2 (fpkm)	Kasumi-1 siMM (fpkm)	Kasumi-1 siRE (fpkm)
MDFI	11.94	11.00	9.00	4.69	23.50	2.35
CD96	19.28	17.00	21.00	9.08	24.24	12.84
ABHD4	4.14	4.00	5.00	1.95	24.78	13.51
UBE2L6	17.56	17.00	18.00	9.28	27.23	13.74
ISYNA1	3.95	4.00	3.00	2.78	29.14	12.65
CD300A	32.26	32.00	22.00	5.12	30.00	16.72
SLC2A3	27.38	26.00	24.00	9.27	30.92	1.95
BIN1	17.72	21.00	18.00	10.80	33.99	12.88
ASNS	81.44	88.00	66.00	16.58	34.97	20.17
CD52	35.08	31.00	28.00	10.18	41.36	16.17
RPTOR	15.01	19.00	12.00	8.15	43.29	19.03
CD69	44.75	43.00	39.00	17.30	43.49	12.57
IFI16	39.31	33.00	37.00	19.26	46.21	14.53
TMIGD2	23.83	29.00	22.00	13.80	51.87	28.47
TRIM8	14.60	15.00	15.00	7.94	53.44	22.55
TBX1	6.00	6.00	6.00	3.02	57.68	22.60
DUSP6	74.95	69.00	78.00	33.12	81.98	20.26
NRN1	61.65	60.00	57.00	22.90	92.06	50.25
JUP	24.98	32.00	23.00	14.74	100.29	48.22
FSCN1	21.74	29.00	18.00	15.02	141.91	59.04
LOC101927497	42.52	37.00	49.00	18.87	152.00	78.68
CKB	28.24	38.00	23.00	11.52	170.00	44.72
PTPRCAP	28.74	43.00	26.00	9.24	206.68	112.12
EGFL7	49.79	83.00	41.00	39.59	223.04	59.09
CD34	114.50	103.00	120.00	61.58	405.03	99.18
TBC1D16	2.16	2.86	1.97	1.22	5.23	2.65
RAB38	2.41	2.61	1.81	0.96	5.43	2.40
FGF11	2.42	2.47	2.57	0.52	5.60	3.11
FGF16	2.06	1.82	2.19	0.88	5.87	2.79
RLTPR	2.47	2.89	2.63	1.09	7.45	3.58
AK4	3.21	2.98	2.28	0.66	8.65	4.78
DDN	2.60	2.83	2.27	1.65	8.91	4.13
VWCE	2.08	3.01	1.87	1.50	9.92	5.25
MEX3A	2.17	2.02	2.22	1.14	10.09	5.09
MTSS1	1.97	1.67	2.25	0.99	11.20	0.74
ZNF467	1.74	2.24	1.69	1.04	13.90	7.60
SHANK3	1.74	2.37	1.61	0.71	16.89	5.75