

Gene regulation by RUNX1 in the absence of consensus sequences

Alexandra Morgan Woodworth

BSc (Hons), GradDipL&T

Submitted in fulfilment of the requirements for the degree of Doctor of Philosophy

University of Tasmania

School of Medicine

Hobart

November 2019



**UNIVERSITY^{of}
TASMANIA**

Table of Contents

Declaration.....	iv
Statement of Co-Authorship.....	v
Acknowledgements.....	vi
List of Figures.....	vii
List of Tables.....	ix
Abbreviations.....	x
Abstract.....	xiii
1 Introduction.....	1
1.1 Transcription factors.....	1
1.2 RUNX Family Transcription Factors.....	2
1.3 RUNX1.....	4
1.3.1 Structure of RUNX1.....	4
1.3.2 Function of RUNX1.....	7
1.4 Leukaemia.....	8
1.4.1 Disruption of RUNX1 in leukaemia.....	9
1.4.2 RUNX1-ETO.....	13
1.4.3 Other RUNX1 translocations.....	14
1.5 Regulation of gene expression by RUNX1.....	15
1.5.1 RUNX1 regulation in the context of chromatin.....	17
1.6 A genome-wide view of transcription factor localisation.....	20
1.6.1 A more complex model of transcription factor binding.....	21
1.7 Research aims.....	23
2 Materials and Methods.....	26
2.1 Cell Culture.....	26
2.1.1 Culture conditions.....	26
2.2 Cloning of candidate gene promoters.....	27
2.2.1 Primer design.....	27
2.2.2 PCR amplification of candidate promoters.....	30
2.2.3 Ligation of PCR products into pXPG.....	30
2.2.4 Plasmid transformation and isolation.....	31
2.2.5 DNA sequencing.....	31
2.2.6 Generation of enhancer plasmid constructs.....	32
2.3 Luciferase reporter assays.....	35
2.3.1 Plasmid preparation.....	35

2.3.2	Transfection of myeloid cell lines	35
2.3.3	Quantitation of protein by Bradford assay	36
2.3.4	Quantification of reporter construct activity by luciferase assay	36
2.4	Analysis of mRNA expression by reverse transcription quantitative PCR	37
2.4.1	RNA extraction	37
2.4.2	cDNA synthesis.....	37
2.4.3	Reverse transcription quantitative PCR	38
2.5	Chromatin conformation capture	40
2.5.1	3C primer design	40
2.5.2	Fixation of cells with formaldehyde.....	43
2.5.3	Restriction enzyme digestion of chromatin	43
2.5.4	Ligation and purification of digested chromatin.....	43
2.5.5	Analysis of 3C products by qPCR.....	44
3	Analysis of genome-wide RUNX1 localisation	46
3.1	Introduction	46
3.2	Results.....	48
3.2.1	Genome-wide analysis of RUNX1 ChIP-Seq data	48
3.2.2	Analysis of RUNX1 binding at promoter regions.....	68
3.3	Discussion.....	81
4	Regulation of candidate gene promoters by RUNX1	88
4.1	Introduction	88
4.2	Results.....	90
4.2.1	Selection of candidate genes	90
4.2.2	Analysis of candidate promoters which bind RUNX1 in the presence of a consensus sequence	92
4.2.3	Analysis of candidate promoters which recruit RUNX1 in the absence of a consensus sequence	99
4.2.4	Response of candidate gene promoters to RUNX1-ETO.....	108
4.2.5	FLI-1 as a regulatory partner in target gene regulation.....	111
4.2.6	Epigenetic markers at candidate promoters	113
4.2.7	Expression in epigenetically modified cells.....	117
4.3	Discussion.....	120
5	Potential role of enhancers in RUNX1 gene regulation	126
5.1	Introduction	126
5.2	Results.....	130
5.2.1	Identification of RUNX1 associated enhancer-promoter pairs.....	130

5.2.2	Analysis of candidate promoter-enhance interactions.....	135
5.2.3	Enhancer reporter assays.....	143
5.2.4	Validation of enhancers by chromatin conformation capture	145
5.3	Discussion.....	149
6	Final discussion and future directions	154
7	References	162

Declaration

Declaration of originality

This thesis contains no material that has been accepted for a degree or diploma by the University of Tasmania or any other institution, except by way of background information and duly acknowledged in the thesis. To the best of my knowledge and belief, this thesis contains no material previously published or written by another person except where due acknowledgement is made in the text.

Authority of access

This thesis may be made available for loan and limited copying in accordance with the *Copyright Act 1968*.

Statement of ethical conduct

The research associated with this thesis abides by the international and Australian codes on human and animal experimentation, the guidelines by the Australian Government's Office of Gene Technology Regulator and the rulings of the Safety, Ethics and Institutional Biosafety Committees of the University.

Funding

Support for this project was provided by the David Collins Leukaemia Foundation of Tasmania Elite Research Scholarship.

This thesis contains fewer than 100,000 words in length, excluding tables, figure legends, and bibliographies.

Alexandra Morgan Woodworth, BSc(Hons), GradDipL&T.

Statement of Co-Authorship

No data presented in this thesis has been published at the time of writing.

Contribution of Work

I assess my contribution to the work described in each chapter to be the following:

- Chapter 3 80%
 - Annotation of ChIP-seq data was performed by Kris Hardy at the University of Canberra. Further analysis and generation of figures was undertaken by the candidate
- Chapter 4 95%
 - TSA and decitabine treatment of cells and subsequent RNA extraction was conducted by Jessica Phillips.
- Chapter 5 100%

We, the undersigned, endorse the above stated contribution of work undertaken for each of the persons contributing to this thesis:

Signed:

Alexandra Woodworth
Candidate
School of Medicine
University of Tasmania
29th November 2019

Associate Professor Adele Holloway
Primary Supervisor
School of Medicine
University of Tasmania
29th November 2019

Professor Lisa Foa
Acting Head of School
School of Medicine
University of Tasmania
29th November 2019

Acknowledgements

Firstly, I would like to thank my supervisor Adele Holloway, for your support and encouragement over the course of both my honours and PhD. Thank you for responding to my desperate email more than four years ago and saving me from a career teaching high school maths. Despite my pessimism and complaining, I have enjoyed working with you and appreciate all you have done for me. Now take a holiday, you deserve it.

Secondly, thanks to my co-supervisors Jo Dickinson and Kate Brettingham-Moore. Your guidance and feedback were always welcome and helpful throughout the project. Talking through problems with you at lab meetings always helped me sort out what was going wrong (usually everything).

Additional thanks to Kris Hardy from the University of Canberra for the initial annotation work, Jac Charlesworth for help with IPA, and Phillippa Taberlay for assistance planning the 3C assay. I'd also like to thank members of the Gene Regulation and Cancer Genetics groups for your support, both in the lab and out. Special thanks to Jess, Emma, and Aparna for your help and advice, and for listening to me vent about cloning for four years straight.

Thank you to the David Collins Leukaemia Foundation for the scholarship which has allowed me to focus on research for the duration of my PhD, as well as travel to present my findings at both national and international conferences. I appreciated the opportunity to meet my scholarship donor and was saddened to hear of his recent passing.

Thanks to all my fellow students and friends in Room 502 and beyond. I won't say it's been fun, but it would have been much harder without your company.

Thanks to my family. To Mum and Dad, for your constant support and borderline embarrassing pride in my work, and to Kate for reminding me why I love science with your spontaneous inquiries about how things work. I'm still not quite sure why your pegs are degrading at different rates based on their colour, but I'll think of some experiments to find out.

Finally, thanks to Nic for "like, cooking and stuff." More than anyone, I couldn't have gotten through the last four years without you. You don't owe me dinner this time, it's my turn.

List of Figures

Figure 1.1. RUNX family members.	3
Figure 1.2. RUNX1 isoforms.	6
Figure 1.3. Locations of RUNX1 mutations.	11
Figure 1.4. RUNX1 recruits both activating and repressive cofactors.	19
Figure 3.1. Genomic distribution of RUNX1 and RUNX1-ETO ChIP binding in Kasumi-1 cells.	50
Figure 3.2. Presence of RUNX consensus sequences at ChIP peaks.	53
Figure 3.3. Presence of RUNX1 consensus sequence associated with a ChIP peak by genomic region.	55
Figure 3.4. The genomic distribution of RUNX1 ChIP peaks categorised by presence of a consensus sequence.	57
Figure 3.5. The genomic distribution of RUNX1-ETO ChIP peaks categorised by presence of a consensus sequence.	58
Figure 3.6. Percentage CpG content categorised by presence of consensus sequence.	60
Figure 3.7. Percentage GC content categorised by presence of consensus sequence.	61
Figure 3.8. Distance to the nearest p300 ChIP peak by genomic context of RUNX1 peak.	63
Figure 3.9. Distance to the nearest p300 ChIP peak by genomic context of RUNX1-ETO peak.	64
Figure 3.10. Distance to the nearest N-CoR ChIP peak by genomic context of RUNX1 peak.	66
Figure 3.11. Distance to the nearest N-CoR ChIP peak by genomic context of RUNX1-ETO peak.	67
Figure 3.12. Diagram of the Aryl Hydrocarbon Signalling Pathway.	77
Figure 3.13. Diagram of the HMGB1 Signalling Pathway.	78
Figure 3.14. Diagram of the RhoA Signalling Pathway.	79
Figure 3.15. Diagram of the HIPPO Signalling Pathway.	80
Figure 4.1. Expression of candidate ARHGAP family genes in leukaemic cell lines.	93
Figure 4.2. Promoter constructs generated from genes with RUNX1 ChIP peaks associated with a consensus sequence.	95
Figure 4.3. Consensus sequence present candidate promoter response to RUNX1 overexpression.	96
Figure 4.4. Expression of candidate HAT family genes in leukaemic cell lines.	100
Figure 4.5. Consensus sequence absent promoter constructs.	103
Figure 4.6. Consensus sequence absent candidate promoter response to RUNX1 overexpression.	104
Figure 4.7. KAT7 construct response to RUNX1 overexpression.	106
Figure 4.8. KAT7-HAT1 construct response to RUNX1 overexpression.	107
Figure 4.9. Candidate construct response to RUNX1-ETO overexpression.	110
Figure 4.10. Candidate construct response to RUNX1 and FLI1 overexpression.	112
Figure 4.11. Histone modifications at consensus sequence present candidate RUNX1 ChIP sites.	115
Figure 4.12. Histone modifications at candidate RUNX1 ChIP sites.	116
Figure 4.13. Expression of candidate genes in leukaemic cell lines treated with epigenetic modifiers.	119
Figure 5.1. Formation of enhancer-promoter loops.	128
Figure 5.2. Representative map of RUNX1 promoter-intergenic/intronic ChIP-PET linkages.	134
Figure 5.3. RUNX1 ChIP peaks linked to consensus sequence present promoters.	137
Figure 5.4. RUNX1 ChIP peaks linked to consensus sequence absent promoters.	138
Figure 5.5. Features of candidate linked regions.	140

Figure 5.6. Response of enhancer constructs to RUNX1.	144
Figure 5.7. Map of 3C primers used to analyse interaction with the HAT1 promoter.	146
Figure 5.8. Regions linked to the HAT1 promoter through 3C.	148
Figure 6.1. Models of regulation by RUNX1.....	157

List of Tables

Table 2.1. Primers for amplification of HAT promoters.....	28
Table 2.2. Primers for amplification of ARHGAP promoters.....	29
Table 2.3. Primers for amplification of enhancer regions.	34
Table 2.4. Primers for RT-qPCR analysis.	39
Table 2.5. Primers for 3C analysis of HAT1+190k region.	41
Table 2.6. Primers for 3C analysis of HAT1-31k region.....	42
Table 3.1. RUNX Consensus Sequences.	52
Table 3.2. Motifs enriched in RUNX1 “consensus absent” promoter ChIP peaks as determined by Homer motif analysis.	70
Table 3.3. Homer <i>de novo</i> motifs enriched in RUNX1 “consensus absent” promoter ChIP peaks.	71
Table 3.4. Biological pathways enriched for genes under both RUNX1 binding modalities.	73
Table 3.5. Biological pathways enriched for genes with RUNX1 promoter ChIP peaks.	75
Table 4.1. Genes clustered within RHOGAP and HAT nodes of the RhoA and HMGB1 signalling pathways.....	91
Table 4.2. Consensus sequences present in candidate promoter regions.	98
Table 5.1. Genes with promoter and intergenic/intronic RUNX1 ChIP peaks.	132
Table 5.2. Features of candidate linked regions.....	142

Abbreviations

°C	degrees Celsius
3'	3 prime
3C	chromatin conformation capture
5'	5 prime
A	adenine
AHR	aryl hydrocarbon receptor
ALL	acute lymphocytic leukaemia
AML	acute myeloid leukaemia
AP3M2	adaptor related protein complex 3 subunit mu 2
ARHGAP	Rho GTPase activating protein
ARNT	AHR nuclear translocator
ATAC-seq	assay for transposable-accessible chromatin using sequencing
BLAST	basic local alignment search tool
bp	base pair
BSA	bovine serum albumin
C	cytosine
CBF	core binding factor
CDC42	cell division cycle 42
ChIA-PET	chromatin interaction analysis with paired-end tag sequencing
ChIP	chromatin immunoprecipitation
ChIP-chip	chromatin immunoprecipitation microarray
ChIP-seq	chromatin immunoprecipitation sequencing
CML	chronic lymphocytic leukaemia
CML	chronic myeloid leukaemia
CMV	cytomegalovirus
CpG	cytosine-guanine dinucleotide
DBD	DNA binding domain
DLX2	distal-less homeobox 2
DNA	dioxyribonucleic acid
<i>E. coli</i>	<i>Escherichia coli</i>
EAP	Epstein-Barr associated protein
E-box	enhancer box
EDTA	ethylenediaminetetraacetic acid
EIF	eukaryotic translation initiation factor
ELF1	E74 like ETS transcription factor 1
EMSA	electrophoretic mobility shift assay
ERG	ETS related gene
ERK	extracellular signal-regulated kinase
ETO	eight twenty-one
ETS	E-twenty-six
ETV6	ETS variant transcription factor 6
EVI1	ecotropic viral integration site 1
FBS	foetal bovine serum
FLI-1	friend leukaemia integration 1
FOXM1	forkhead box M1

g	acceleration due to gravity
G	guanine
GAP	GTPase activating protein
GAPDH	glyceraldehyde 3-phosphat dehydrogenase
GATA2	GATA binding protein 2
gDNA	genomic DNA
GDP	guanine diphosphate
GLI	glioma associated oncogene
GMAT	genome-wide mapping technique
GM-CSF	granulocyte macrophage colony-stimulating factor
GNAT	GCN5-related N-acetyltransferase
GTP	guanosine triphosphate
GTPase	guanosine triphosphatase
H3K27Ac	histone 3 lysine 27 acetylation
H3K27Me3	histone 3 lysine 27 trimethylation
H3K4Me1	histone 3 lysine 4 monomethylation
H3K4Me3	histone 3 lysine 4 trimethylation
HAT	histone acetyltransferase
HBO1	histone acetyltransferase bound to ORC
HDAC	histone deacetylase
HEPES	hydroxyethyl piperazineethanesulfonic acid
HGNG	HUGO Gene Nomenclature Committee
HIPK2	homeodomain interacting protein kinase 2
HMGB1	high mobility group box-1
HOMER	hypergeometric optimisation of motif enrichment
HSC	haematopoietic stem cell
HT-SELEX	high-throughput systematic evolution of ligands by exponential enrichment
Ig	immunoglobulin
IPA	Ingenuity Pathway Analysis
ITGB4	integrin β 4
KAT	lysine acetyltransferase
kb	kilobase
L-broth	lysogeny broth
LMO2	LIM domain only 2
LYL1	lymphoblastic leukaemia derived sequence 1
M-CSF	macrophage colony-stimulating factor
MDS1	myelodysplasia syndrome 1
miRNA	micro RNA
mL	millilitre
MORF	MOZ-related factor
MOZ	monocytic leukaemia zinc finger protein
mRNA	messenger ribonucleic acid
mSin3A	mammalian Sin3 protein A
MTG16	myeloid translocation gene on chromosome 16
N-CoR	nuclear receptor co-repressor
NFAT	nuclear factor of activated T-cells
NF- κ B	nuclear factor kappa B
ng	nanogram

NHR	Nervy homology region
NK cell	natural killer cell
nm	nanometre
ORC	origin recognition complex
ORGANIC	assay for occupied regions of genomes from affinity-purified naturally isolated chromatin
PCR	polymerase chain reaction
PML	promyelocytic leukaemia
PRDM16	PR domain containing 16
qPCR	quantitative PCR
RAC	Ras-related C3 botulinum toxin substrate
RHD	Runt homology domain
RhoA	Ras homolog family member A
RNA	ribonucleic acid
RNA Pol II	ribonucleic acid polymerase II
RNase	ribonuclease
RPMI	Roswell Park Memorial Institute
RT-qPCR	reverse transcription quantitative PCR
RUNX	Runt-related transcription factor
SAGE	serial analysis of gene expression
SCL	stem cell leukaemia
SDS	sodium dodecyl sulphate
SMRT	silencing mediator of retinoic acid and thyroid hormone receptor
SPI1	SFFV proviral integration oncogene
STAGE	sequence tag analysis of genomic enrichment
STAT1	signal transducer and activator of transcription 1
SUMO	small ubiquitin-like modifier
SWI/SNF	switch/sucrose non-fermentable
T	thymine
TAD	topologically associated domain
TAZ	tafazzin
TLE	transducin-like enhancer protein
TNF α	tumour necrosis factor alpha
TRC α	T-cell receptor alpha
Tris	tris(hydroxymethyl)aminomethane
TSA	trichostatin A
TSS	transcription start site
TTS	transcription termination site
UCSC	University of California Santa Cruz
UTR	untranslated region
V	volts
XAP2	HBV X-associated protein
YAP1	yes-associated protein 1
μ F	micro-Faraday
μ g	microgram
μ L	microlitre
μ M	micromolar

Abstract

Runx-related transcription factor 1 (RUNX1) is a transcription factor that has an important role in haematopoietic cell development and function and is frequently disrupted in leukaemia. RUNX1 is commonly described as a sequence-specific DNA binding factor which recognises the consensus sequence TG(T/C)GGT in the promoter and enhancer regions of its target genes to affect changes in gene expression. However, the advent of techniques to study DNA-protein interactions on a genome-wide scale has provided the opportunity to re-assess RUNX1 localisation and function, and this analysis suggests that the classical model of RUNX1 function is incomplete. By fully understanding the mechanisms by which RUNX1 maintains its target gene expression profiles under normal cellular conditions, insights into disrupted function can be gained and interventions can be developed.

Analysis of publicly available RUNX1 ChIP-Seq data determined that the majority of RUNX1 binding in haematopoietic cells occurs outside of gene promoter regions, in intergenic or intronic regions. Furthermore, approximately one fifth of all RUNX1-DNA binding sites on a genome-wide scale were not associated with a canonical consensus sequence, and this was particularly prevalent in promoter regions, with almost half of RUNX1 binding in promoter regions occurring in the absence of consensus sequences. This suggests that recruitment of RUNX1 to gene targets occurs through multiple mechanisms and raises the possibility that it may function differently depending on its location and mode of recruitment. Similar results were obtained for localisation of the RUNX1 fusion protein RUNX1-ETO, the product of a common chromosomal translocation in leukaemia.

This data set was used to investigate the different modes of binding and action of RUNX1 with the aim of establishing whether binding in the absence of consensus sequences constitutes a

novel mechanism of RUNX1 binding and if genes regulated in this way respond differently to RUNX1 disruption. This study identified biological pathways enriched for genes with RUNX1 binding in their promoters, and in which only one type of binding (in the presence or absence of a consensus sequence) occurred. Clusters of functionally related RUNX1-bound genes were selected from two of these pathways; the RhoA and HMGB1 signalling pathways and used as models to investigate regulation by RUNX1.

The RhoA signalling pathway genes, ARHGAP1, ARHGAP4 and ARHGAP12, all bound RUNX1 at their promoters, in association with a consensus sequence. However, while ARHGAP4 and ARHGAP12 responded to RUNX1 in reporter assays, ARHGAP1 did not, suggesting that a consensus sequence alone is not sufficient for a promoter to respond to RUNX1. In contrast, a group of histone acetyl transferase genes from the HMGB1 signalling pathway bound RUNX1 at their promoters in the absence of a consensus sequence. Both the KAT6B and KAT2B promoters were activated by RUNX1 in the absence of a consensus sequence, suggesting a mechanism whereby RUNX1 recruitment to DNA does not require the canonical consequence sequence and may rely on recruitment by additional transcription factors or distal chromatin elements. Interestingly, evidence presented here indicates that while RUNX1-ETO inhibits RUNX1 activity at classically regulated promoters which contain RUNX consensus sequences, it cannot inhibit RUNX1 activity in the absence of canonical consensus sequences. This suggests that genes regulated by RUNX1 through different transcriptional mechanisms may respond differently to disruption of RUNX1 in diseases such as leukaemia.

The majority of RUNX1 binding occurs outside of promoters, with one potential explanation for intergenic or intronic RUNX1 binding being that it functions at gene enhancers. Data presented here identified two such potential enhancers of the HAT1 gene. While HAT1 bound RUNX1 at

its promoter, although in the absence of a consensus sequence, it does not respond directly to RUNX1 in reporter assays. Potential enhancers were identified approximately 31 kb upstream and 190 kb downstream of the promoter. While interactions with these regions were confirmed, neither were found to behave as an enhancer in reporter assays. However, the data presented is consistent with the +190 kb region representing a promoter-promoter interaction which may be related to the formation of large interconnected transcriptional complexes. Such promoter-promoter interactions could at least partly explain the prevalence of RUNX1 binding at promoters in the absence of a consensus sequence.

This study has expanded our understanding of the mechanisms by which RUNX1 affects transcription of its target genes and has characterised distinct modes of operation at candidate genes: either directly through a consensus sequence; directly in the absence of a consensus sequence; or distally through an interacting regulatory element.

1 Introduction

1.1 Transcription factors

Transcription factors are a class of proteins which share a function in regulating and facilitating the expression of genes by modulating their transcription. Broadly, transcription factors fall into two categories: general factors, which are components of the transcription machinery such as RNA polymerase II; and sequence-specific transcription factors. The latter category of proteins contain a DNA binding domain, which interacts with specific DNA elements throughout the genome (Kadonaga 2004). Through interaction with sequence elements, transcription factors determine the expression of genes, which in turn regulate all complex biological processes in the cell (Prange, Singh et al. 2014). Genes coding for transcription factors comprise approximately 10% of all genes in the human genome (Levine and Tjian 2003). Transcription factors are classified based on the type of DNA binding domain they contain. Major DNA binding domain classes include homeodomain, basic helix-loop-helix, leucine zipper and zinc finger transcription factors (Wingender, Chen et al. 2001). To regulate transcription, these factors localise to regulatory elements of DNA through amino acid motifs in their DNA binding domains and facilitate recruitment of the transcription machinery. Generally, the DNA binding domains of transcription factors recognise a specific sequence of nucleotides, usually a 5-6 base pair motif referred to as its consensus sequence. However, the sequences recognised by an individual transcription factor can be highly variable, with corresponding changes in binding affinity. Regulatory elements such as promoters and enhancers are often characterised by the presence of many consensus sequences and are bound by numerous transcription factors simultaneously. Regulation of a single element by multiple transcription factors working in a co-ordinated manner is one way in which precise control of transcription is achieved (Morgunova and Taipale 2017).

1.2 RUNX Family Transcription Factors

The RUNX family of transcription factors, named for their binding domain's homology to the *Drosophila* Runt protein, regulate various cellular processes involved in normal development and cell fate determination (Kagoshima, Shigesada et al. 1993). RUNX proteins are DNA-binding factors which form heterodimers with the ubiquitously expressed non-DNA binding Core Binding Factor β (CBF β), to form the Core Binding Factor (CBF). The heterodimer has increased DNA binding affinity compared to the RUNX protein alone (Wang, Stacy et al. 1996). There are three RUNX transcription factors in mammals (Figure 1.1). RUNX1 is a master regulator of haematopoiesis and is vital for both the development of haematopoietic stem cells and the normal function of adult haematopoiesis (Okuda, Van Deursen et al. 1996). RUNX2 is essential for the formation of bone, and is crucial for the differentiation of osteocytes and chondrocytes (Komori, Yagi et al. 1997, Otto, Thornell et al. 1997, Inada, Yasui et al. 1999). RUNX3 is involved in neurogenesis, and when knocked out in mouse models leads to severe limb ataxia due to disrupted synaptic connectivity (Levanon, Bettoun et al. 2002). In addition to these essential functions, all three RUNX proteins have roles in other organs and systems (de Bruijn and Dzierzak 2017). All three are frequently disrupted in various cancers, demonstrating their importance in maintaining normal cellular development and function (Chuang, Ito et al. 2013).

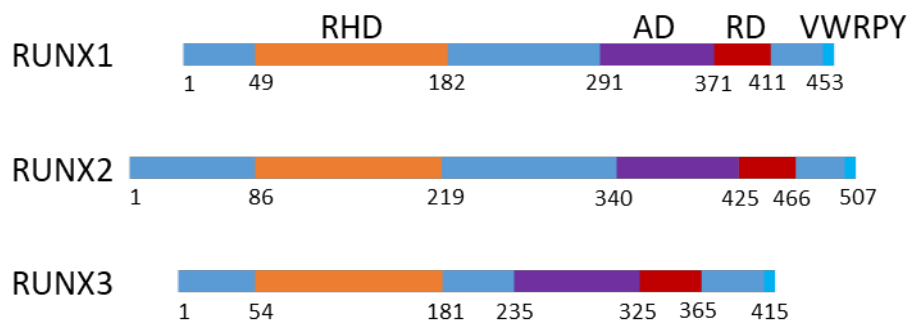


Figure 1.1. RUNX family members.

Schematic representation of human RUNX protein family members. The Runt homology domain (orange), activation domain (purple), repression domain (red) and VWRPY motif (blue) are shown. Relevant amino acid positions are indicated below. Only the RUNX1B isoform is shown. Adapted from Chuang, Ito et al. 2013.

1.3 RUNX1

The Runt-related transcription factor 1 (RUNX1) gene was originally identified at the breakpoint of the t(8;21) fusion gene in acute myeloid leukaemia (AML) cells (Rowley 1973, Miyoshi, Shimizu et al. 1991). The protein was identified as a sequence-specific binding factor at enhancers within the Moloney murine leukaemia virus and polyomavirus genome (Speck and Baltimore 1987, Kamachi, Ogawa et al. 1990). RUNX1 has since been shown to be a master regulator of haematopoiesis.

1.3.1 Structure of RUNX1

The structure and function of RUNX1 is well described, as it has been thoroughly investigated as a major regulator of haematopoiesis. RUNX1 consists of a DNA binding runt homology domain, as well as a transactivation and repression domain, both located at the C-terminal end of the polypeptide and an N-terminal VWRPY motif which directs nuclear localisation (Figure 1.1). This runt homology domain shares 92% sequence homology with the drosophila runt gene (Berardi, Sun et al. 1999), which is involved in sex determination, segmentation and neurogenesis (Kagoshima, Shigesada et al. 1993). There are three major isoforms of RUNX1 resulting from transcription from two promoters and alternative splicing (Figure 1.2). The proximal promoter P2 directs transcription of the isoforms 1A and 1B, of which the first 241 amino acids are identical but differ in alternative splicing of the carboxyl terminus. After 241 amino acids the isoforms diverge. The 1A isoform ends after a further 9 amino acids while the 1B isoform is much longer, containing a total of 453 amino acids. Isoform 1C is transcribed from the distal P1 promoter and differs from 1B only in the first 32 amino acids (Miyoshi, Ohira et al. 1995). Isoform 1C is expressed only during early development of definitive haematopoietic stem cells, while 1A and 1B are expressed throughout haematopoietic differentiation (Challen and Goodell 2010). Isoform 1A is more prevalent than 1B in CD34⁺

progenitor cells in human cord blood, and its relative overexpression is linked to increased cellular self-renewal (Tsuzuki, Hong et al. 2007). All transcripts encode the Runt DNA binding domain, which forms an s-type immunoglobulin fold similar to those found in the transcription factors NF- κ B, NFAT, p53, STAT1, and the T-domain (Berardi, Sun et al. 1999). Unlike these similar domains, the Runt domain is unique in that it recognises DNA through loops formed between β -sheets at both ends of the Ig fold. The Runt domain in RUNX1 undergoes allosteric change in three looping regions (L11, 9 and 5) when bound to DNA (Bäckström, Wolf-Watz et al. 2002), and is stabilised by heterodimerisation with CBF β (Bartfeld, Shimon et al. 2002). The DNA target sequence of the Runt domain is predisposed to adopt a unique bent conformation prior to protein binding.

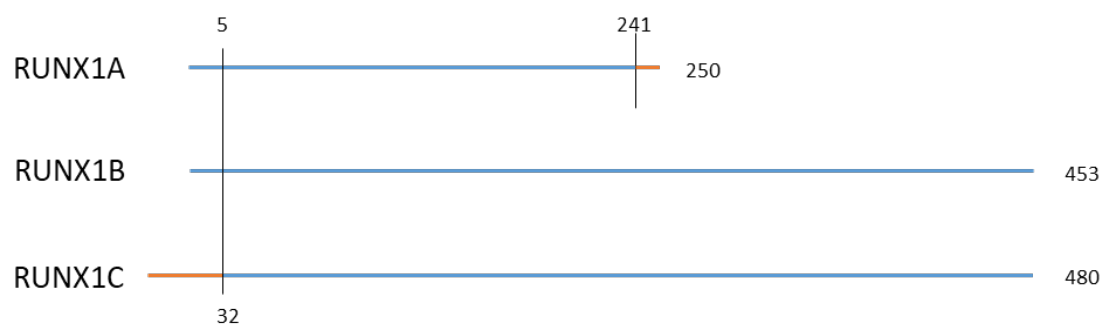


Figure 1.2. RUNX1 isoforms.

Schematic protein representation of RUNX1 isoforms A, B, and C. Amino acids which differ from the RUNX1B isoform are indicated in orange. Adapted from Sood, Kamikubo et al. 2017.

1.3.2 Function of RUNX1

RUNX1 is required for the establishment of haematopoietic stem cells (HSCs) during embryogenesis and subsequent differentiation of HSCs into mature myeloid, lymphoid, and megakaryocyte lineages (Friedman 2009). Two stages of haematopoiesis occur, primitive, involving differentiation of primitive macrophages and early erythrocytes from progenitors in the yolk sac for rapid development of the embryo (Palis, Robertson et al. 1999) and definitive, which involves generation of mature haematopoietic cells of various lineages from a common definitive HSC. RUNX1 is present at the earliest stages of HSC development during embryogenesis (Lam and Zhang 2012).

Evidence for the function of RUNX1 has been obtained through knockout mouse studies, which disrupt the protein at the Runt homology domain and therefore prevent the function of all Runx1 isoforms. Runx1 heterozygous knockout mice are healthy, but double knockout mice die between 12.5- and 13.5-days post-conception from haemorrhaging of the central nervous system (Okuda, Van Deursen et al. 1996). This embryonic lethality has been further linked to defects in angiogenesis caused by lack of angiopoietin-1 expression (Takakura, Watanabe et al. 2000). Analysis of Runx1 double knockout mouse embryos at 11.5-days post-conception identified no significant differences in the yolk sac between wild type and knockout mice, but the Runx1 negative embryos completely lacked haematopoiesis in the liver. Runx1-null mice have nucleated primary erythrocytes, consistent with yolk sac derivation. As the normal murine developmental cycle shifts the site of haematopoiesis from the yolk sac to the foetal liver at this stage, this suggests that the mice have successful primitive haematopoiesis, but lack definitive haematopoiesis and that RUNX1 is only necessary for the latter. More recent evidence indicates that RUNX1 also has a role in primary haematopoiesis, as the primitive

erythrocytes developing in Runx1-null mice exhibit disrupted phenotypes (Yokomizo, Hasegawa et al. 2008).

It is clear that RUNX1 is vital for embryonic development of haematopoietic programs, and a role in adult haematopoiesis has also been elucidated through analysis of conditional Runx1 knockout mice. These models demonstrate that while RUNX1 is not required for adult haematopoiesis in the same way as it is in the embryo, lack of Runx1 nevertheless results in haematopoietic abnormalities. In particular, conditional knockout of Runx1 in haematopoietic cells, induced using a Cre-LoxP mouse model, resulted in inefficient platelet and common lymphocyte progenitor production, inhibited B-and T-cell maturation, and increased proliferation of myeloid progenitor populations and peripheral blood neutrophils (Gowney, Shigematsu et al. 2005). This supports evidence from Runx1 haploinsufficient mice, which have increased number of HSCs, but a lower number of functional long-term HSCs (Sun and Downing 2004). Maintenance of HSCs is impaired in Runx1 deficient cells, leading to exhaustion of the stem cells after the initial expansion and an eventual decline in HSC numbers (Jacob, Osato et al. 2010). RUNX1 impaired cells in general exhibit increased proliferative capacity but decreased functionality.

1.4 Leukaemia

RUNX1 is critical in establishing gene expression profiles during haematopoietic cell development and function, and therefore its disruption is commonly associated with the development of haematopoietic malignancies such as leukaemia. An estimated 4000 leukaemia cases were diagnosed in 2018 in Australia, making it the 8th most commonly diagnosed cancer. Almost 2000 deaths were attributed to various leukaemias in the same year. While survival rates have improved steadily, 5-year survival rates remain around 60% (2010-

2014; Australian Institute of Health and Welfare 2019). RUNX1 mutations in AML patients are associated with poorer prognosis, exhibiting shorter disease-free periods as well as lower complete remission and overall survival rates (Tang, Hou et al. 2009). The incidence of RUNX1 mutations in AML is estimated to be between 5 and 16% and it has been proposed that AML with RUNX1 mutation forms a distinct clinical subtype of the disease (Jalili, Yaghmaie et al. 2018). RUNX1 is required at several stages for maturation of haematopoietic stem cells into mature and specialised lineages, as well as negatively regulating renewal of immature cells.

1.4.1 Disruption of RUNX1 in leukaemia

Genetic mutations causing leukaemia have been historically thought to fall into two broad categories: class I mutations which lead to uncontrolled proliferation and inhibition of apoptosis, and class II mutations which block cellular maturation (Gilliland 2002). More recent studies have added complexity to this model, identifying a third class of mutations which facilitate epigenetic and chromatin state change, and elucidating the interdependence of different classes (Naoe and Kiyoi 2013). RUNX1 mutations, which generally cause loss of function, have been historically assigned to class II. Loss of RUNX1 function commonly occurs through point mutations, deletions, or translocations (Ichikawa, Yoshimi et al. 2013), with disruption of RUNX1 leading to myelodysplastic syndrome, acute myeloid leukaemia, acute lymphoblastic leukaemia and chronic myelomonocytic leukaemia. Leukaemias with translocations occurring at RUNX1 or its binding partner CBF β , including the RUNX1-ETO translocation, are referred to as core binding factor (CBF) leukaemia, which is more prevalent in younger patients and generally associated with a better prognosis. However, leukaemias with RUNX1 somatic mutations tend to occur in older patients and have a poorer prognosis (Sood, Kamikubo et al. 2017). Haploinsufficiency of RUNX1 has been linked to the inheritance of familial platelet disorder, a disease which predisposes its sufferers to the development of

AML as the second copy of the gene is vulnerable to inactivating mutations leading to full knockout (Song, Sullivan et al. 1999, Preudhomme, Renneville et al. 2009).

Alterations which increase RUNX1 copy number can also lead to disease. RUNX1 gene duplications resulting from hyperdiploidy occurred in 50% of subjects in a study of Mexican children with ALL (Rosales-Rodríguez, Fernández-Ramírez et al. 2016). Point mutations, in which a single nucleotide is altered from the wild type sequence, can lead to dominant-negative forms of RUNX1, particularly when they occur within the DNA-binding Runt domain or in the C-terminal region of the protein (Imai, Kurokawa et al. 2000, Michaud, Wu et al. 2002, Churpek, Garcia et al. 2010). One study of 470 *de novo* AML patients identified 63 distinct mutations, as summarised in Figure 1.3 (Tang, Hou et al. 2009).

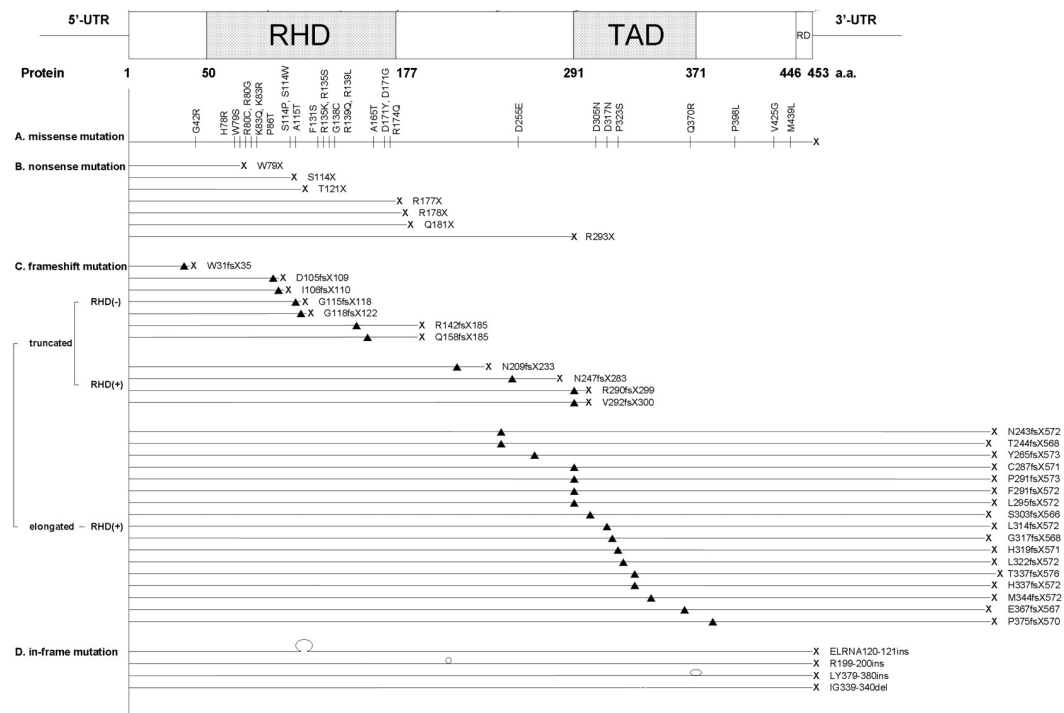


Figure 1.3. Locations of RUNX1 mutations.

Figure taken from Tang, Hou et al 2009. Four kinds of mutations were identified in a study of 470 individuals with *de novo* AML, missense, nonsense, frameshift, and in-frame mutations. ▲ and | represent the sites of mutation; X, the site of a stop codon; Ω, the site of insertion; the site of deletion. RD indicates repression domain; TAD, transactivation domain; UTR, untranslated region; and a.a., amino acid.

Mutations within the Runt domain that have been identified in individuals with leukaemia include silent mutations; where the change in DNA has no effect on the protein sequence as the resulting codon is synonymous with the wild type, though the mutation may still influence the gene by altering splice sites or disrupting chromatin features; missense mutations, in which the altered base changes the coded amino acid; and nonsense or frameshift mutations which commonly lead to premature translational termination. An early study of individuals with AML and CML identified RUNX1 point mutations in six out of 117 cases (Osato, Asou et al. 1999). Mutations identified had various effects on RUNX1-DNA interactions as assayed by electrophoretic mobility shift assays (EMSA). Of these mutations, H58N occurs on the periphery of the Runt domain and had no discernible effect on DNA binding. At the other end of the Runt domain, mutation or premature termination at residue 177 prevented RUNX1-DNA binding, but this was restored on heterodimerisation with CBF β due to induced conformational change. No DNA binding was detected for missense mutations at residues R80 and K83, but the mutant protein is still able to form heterodimers. Both DNA binding and heterodimerisation capacity was removed by a 4 base pair insertion at C72 and premature termination at S114 (Osato, Asou et al. 1999). Mutations also disrupted the localisation of RUNX1 to the nucleus. Since this study, many more mutations in RUNX1 have been identified in large cohorts of individuals with leukaemia (Preudhomme, Warot-Loze et al. 2000, Tang, Hou et al. 2009, Gaidzik, Bullinger et al. 2011, Schnittger, Dicker et al. 2011). The majority of RUNX1 mutations have been detected in the Runt homology domain of the protein, affecting its ability to bind DNA targets but not its ability to heterodimerise with CBF β . This suggests a potential mechanism for the dominant-negative effects of mutant RUNX1 where sequestration of CBF β , makes it unavailable for interaction with the wild type protein and in turn impairing its DNA binding activity (Hyde, Liu et al. 2017).

In addition to mutations within the gene, RUNX1 is a common target for chromosomal translocations in cancer cells. The translocation resulting in the fusion of the RUNX1 gene with the RUNX1T1 locus, which encodes the Eight-Twenty-One (ETO) gene was the first translocation event to be identified in relation to cancer (Rowley 1973). Since then, approximately 55 different translocation partners have been identified for RUNX1, though the partner gene involved has only been identified in 21 of these cases (De Braekeleer, Douet-Guilbert et al. 2011). Fusion proteins which retain the Runt domain but lose the transactivation domain are thought to promote leukaemogenesis by inhibiting wild type RUNX1 function. Most identified translocations are rare, having been only identified in a few cases, but there have been seven recurrent RUNX1 translocations described (De Braekeleer, Ferec et al. 2009).

1.4.2 RUNX1-ETO

One of the most commonly recurring translocations, t(8;21), in which the first five exons of the RUNX1 gene are fused to the Eight-Twenty-One (ETO) gene at intron 1, is present in approximately 12% of all AML cases (Lam and Zhang 2012). The resulting fusion protein, RUNX1-ETO, contains the N-terminal 177 amino acids of RUNX1, which includes the Runt homology domain, fused to the majority (575 amino acids) of the ETO protein. The Runt domain in this fusion protein is functional, allowing for both DNA binding and heterodimerisation with CBF β , but replacement of the transactivation domain with the ETO protein disrupts usual RUNX1 transcriptional regulation (Lin, Mulloy et al. 2017). The ETO protein contains four evolutionarily conserved Nervy homology regions (NHR) and a nuclear localisation signal, which serves to direct the protein to the nucleus (Davis, McGhee et al. 2003). The second of these NHRs is crucial for the leukaemic potential of the chimeric protein, as it is the site of oligomerisation to form a tetramer (Liu, Cheney et al. 2006). This homo-

oligomeric formation and subsequent leukaemogenesis is independent of CBF β (Kwok, Zeisig et al. 2009). As well as oligomerisation, the fused ETO protein acts as a transcriptional repressor by recruiting other factors, NCOR, SMRT, HDACs 1-3, and mSin3A, to form a corepressor complex (Gelmetti, Zhang et al. 1998, Wang, Hoshino et al. 1998, Amann, Nip et al. 2001). The formation of this complex has been demonstrated in overexpression studies, however the interaction between components is only weak *in vivo*, suggesting that the formation may be dynamic and context-dependent (Sun, Wang et al. 2013).

1.4.3 Other RUNX1 translocations

Other recurring translocations of the RUNX1 gene include t(16;21), t(3;21), t(1;21), and t(12;21) (De Braekeleer, Ferec et al. 2009). The t(16;21) translocation fuses RUNX1 to MTG16 (Myeloid translocation gene on chromosome 16), which is an ETO family gene. The resulting fusion protein has properties similar to those of RUNX1-ETO in that it can repress RUNX1 targets by recruiting corepressor complexes (Gamou, Kitamura et al. 1998). Three different translocation partner genes have been identified in the heterogeneous t(3;21) translocation: EAP, EVI1 and MDS1. RUNX1-EAP translocation is out of phase, introducing a premature truncation and the fusion protein is proposed to act as a repressor of normal RUNX1 (Nucifora, Begy et al. 1993, Sacchi, Nisson et al. 1994). Fusion with EVI1 attaches two zinc-finger domains to the Runt domain of RUNX1, which have a negative effect on normal RUNX1 expression as well as encouraging proliferation and impeding apoptosis (Mitani 2004). EVI1 also exists as a fusion protein with MDS1, caused by alternative splicing at the MDS1-EVI1 complex locus (Tanaka, Oshikawa et al. 2017). The RUNX1-MDS1-EVI1 fusion protein has similar effects to the RUNX1-EVI1 fusion (Sood, Talwar-Trikha et al. 1999). Also similar in effect to RUNX1-EVI1, the t(1;21) translocation fuses the Runt homology domain to almost the entire PRDM16 protein, which shares significant homology to EVI1 (Stevens-Kroef, Schoenmakers et al. 2006). Most

commonly observed in paediatric acute lymphoblastic leukaemia, the t(12;21) is distinct from other translocations in that it retains the RUNX1 transactivation domain, fusing the 3' region of RUNX1 and the 5' end of ETV6. The ETV6 protein contains a repressive domain, allowing the fusion protein to bind to RUNX1 targets and recruit HDACs (Romana, Mauchauffe et al. 1995, Zelent, Greaves et al. 2004).

1.5 Regulation of gene expression by RUNX1

In vitro, the RUNX1 RHD binds to the consensus DNA sequence 5'-TG(T/C)GGT-3' (Meyers, Downing et al. 1993), considered to be the canonical RUNX1 binding sequence. Electrophoretic mobility shift assays and X-ray crystallography studies of the RUNX1 RHD identified three critical arginine residues which bind to specific guanines in the DNA sequence TG(T/C)GGT (Tahirov, Inoue-Bungo et al. 2001). Understanding of the function of RUNX1 as a transcription factor has been largely driven by investigation of target genes, which have been mostly discovered by locating consensus sequences in gene promoters or identifying differential expression patterns through mRNA screens (Otto, Lübbert et al. 2003). For example, Interleukin 3 was identified as a potential RUNX1 target gene in T cells through the presence of both a canonical RUNX1 consensus sequence (TGTGGT) and a variant sequence (TGTGGG) in its promoter. Co-transfection, gel shift and mutagenesis experiments demonstrate the response of these elements to RUNX1, though the variant sequence showed a lower affinity for the transcription factor and its mutation did not disrupt promoter responsiveness (Uchida, Zhang et al. 1997). Similarly, in myeloid cells, RUNX1 has been shown to up-regulate the M-CSF (macrophage colony-stimulating factor) receptor (Zhang, Fujioka et al. 1994), complement receptor type 1 (Kim, Lee et al. 1999) and CD36 (Armesilla, Calvo et al. 1996) genes through interaction with the consensus sequence in gene promoters.

On its own, RUNX1 is a relatively weak transcription factor. Heterodimerisation with CBF β increases affinity for DNA binding more than 40-fold over RUNX1 alone (Gu, Goetz et al. 2000), but activation or repression of RUNX1 target genes often relies on cooperation with other transcription factors, forming large multiprotein complexes (Yoshida and Kitabayashi 2008). One such co-factor is yes-associated protein (YAP1), a component of the Hippo signalling pathway. RUNX1 binds to target promoters, such as *Itch*, and recruits YAP1 through interactions between the RUNX1 PY motif, located within the RUNX1 activation domain and the YAP1 WW domain (Yagi, Chen et al. 1999, Levy, Reuven et al. 2008). This interaction is altered in the event of DNA damage, during which YAP1 becomes phosphorylated at tyrosine residues and reduces its binding preference for RUNX proteins and instead binds p73 and activates pro-apoptotic targets (Levy, Adamovich et al. 2008). This example highlights the complexity of regulation of RUNX1 targets and the co-operative mechanism by which transcription factors affect changes in the cellular environment. Another group of commonly associated transcription factors are from the ETS family. One such factor, PU.1, which is encoded by the *SPI1* gene, has been shown to act synergistically with RUNX1 to activate the M-CSF receptor promoter (Petrovick, Hiebert et al. 1998). Co-operative binding between the two factors is driven by RUNX-ETS hybrid consensus sequences located at target regulatory sites (Zhao, Osipovich et al. 2017).

RUNX1 activity is also modulated by post-translational modifications. Phosphorylation of RUNX1 by the ERK1/2 kinase within its transactivation domain increases its transactivation ability by driving association with the co-activator p300 (Tanaka, Kurokawa et al. 1996, Wang, Huang et al. 2009).

1.5.1 RUNX1 regulation in the context of chromatin

Though many investigations into the function of RUNX1 consider its target regulatory regions in isolation, in reality transcription factors must operate in the context of chromatin. Broadly speaking, chromatin exists on a dynamic spectrum between two opposing states: open and accessible euchromatin and densely compacted heterochromatin, in which the DNA is tightly wound around histone proteins. Access to the DNA by transcription factors and transcription machinery is restricted by heterochromatin, thus repressing gene expression. Conversion between the two states is vital for precise control of gene expression programs and is influenced by a number of factors, including DNA methylation, histone modifications or variants, and recruitment of chromatin remodelling proteins such as the SWI/SNF complex (Kassabov, Zhang et al. 2003, Choy, Wei et al. 2010, Gardner, Allis et al. 2011).

RUNX1 is capable of influencing the formation of a transcriptionally active euchromatin state through interaction with chromatin modifying co-factors such as p300/CBP, KAT6A, PML, PRMT1, and HIPK2 (Yoshida and Kitabayashi 2008, Brettingham-Moore, Taberlay et al. 2015). Additionally, while RUNX1 was initially characterised as a transcriptional enhancer, it has since become clear that it also serves a repressive function in some circumstances (Durst and Hiebert 2004). This is demonstrated at the CD4 promoter in mice (Taniuchi, Osato et al. 2002, Wheeler, VanderZwan et al. 2002) and the p21 promoter in human cell lines (Lutterbach, Westendorf et al. 2000). Repression occurs through recruitment of corepressors mSin3 or TLE/Groucho through the repressive domain of the RUNX1 protein and also through repressive chromatin modifications caused by recruitment of the histone methyltransferases SUV39H1, PRMT4, and PRMT6, and subsequent histone deacetylase binding (Aronson, Fisher et al. 1997, Lutterbach, Westendorf et al. 2000, Reed-Inderbitzin, Moreno-Miralles et al. 2006, Brettingham-Moore, Taberlay et al. 2015). The effect of RUNX1 on transcription is therefore

dependent on the context in which it interacts with DNA and the composition of the transcriptional complex in which it is acting, having influences on both transcriptional activity and epigenetic state (Figure 1.4) (Brettingham-Moore, Taberlay et al. 2015).

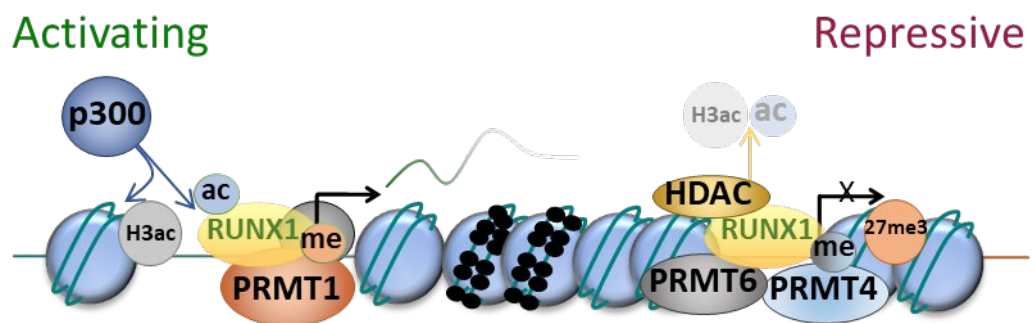


Figure 1.4. RUNX1 recruits both activating and repressive cofactors.

RUNX1 can act as a transcriptional activator or repressor dependent on the balance of coactivators/corepressors associated with it at a particular time. RUNX1 can recruit coactivators or corepressors as well as epigenetic modifiers to affect transcriptional activation and chromatin state. Figure from Brettingham-Moore, Taberlay et al. 2015.

1.6 A genome-wide view of transcription factor localisation

The behaviour of transcription factors has been extensively studied using various *in vitro* technologies, such as protein binding microarrays, high-throughput systematic evolution of ligands by exponential enrichment (HT-SELEX), and DNA immunoprecipitation paired with either microarray or sequencing analysis (Levo and Segal 2014). These techniques provide precise quantitative data on affinities of a wide range of transcription factors for DNA and have proven invaluable in identifying transcription factor binding sites. Additionally, a range of technologies have been employed to examine transcription factor function and binding context *in vivo*, including chromatin immunoprecipitation paired with microarrays or sequencing (ChIP-chip or ChIP-seq), assay for transposase-accessible chromatin using sequencing (ATAC-seq), and assays for occupied regions of genomes from affinity-purified naturally isolated chromatin (ORGANIC) (Levo and Segal 2014). While there is often considerable agreement between transcription factor occupancy predicted based on *in vitro* binding affinities and that observed through *in vivo* assays, these assays have also frequently revealed unexpected localisation of transcription factors. The difference between transcription factor behaviour *in vitro* and *in vivo* highlights the complexity of transcription factor function and represents a gap in current understanding of transcription factor regulatory mechanisms.

Genome-wide technologies such as ChIP-seq allow us to obtain a more wholistic view of transcription factor localisation than was possible through analysis of individual genes in isolation. Chromatin immunoprecipitation techniques involve cross-linking proteins such as transcription factors to DNA using formaldehyde followed by isolation and sonication of genomic DNA to generate protein-bound fragments. These fragments are precipitated with protein-specific antibodies and the cross-links are reversed. Originally, ChIP analyses identified genomic regions by PCR with primers for regions of interest, comparing experimental products

to those generated from precipitation with unrelated antibody or IgG. While these studies are useful in confirming enrichment of protein binding to specific candidate regions, they are limited in their scope for identifying novel binding sites. To overcome this limitation, more recent techniques pair ChIP assays with detection technology such as micro-arrays (ChIP-chip) or various Sanger sequencing processes (ChIP-SAGE, ChIP-STAGE and ChIP-GMAT), reviewed in Hoffman & Jones, 2009 (Hoffman and Jones 2009). Pairing chromatin immunoprecipitation with flow cell sequencing (ChIP-Seq) delivers improved coverage, provides more detailed information, and is the highest resolution ChIP technique currently in use, allowing analysis of genome-wide protein binding in an unbiased manner.

1.6.1 A more complex model of transcription factor binding

A genome wide perspective has provided new insights into the mechanism of gene regulation by another transcription factor, Myc. Myc comprises one half of the Myc-Max heterodimer, which has been shown to bind to E-box sequence elements (CACGTG) *in vitro* (Blackwell, Kretzner et al. 1990). However, a study by Guo and colleagues interrogating the binding of Myc using ChIP-seq data identified that a large proportion of binding occurs in the absence of the canonical E-box sequence (Guo, Li et al. 2014). In fact, the occupancy of Myc on a genome-wide scale correlates far better with loci bound by RNA Polymerase II (Pol II) than sites of E-box elements, being found on average 100 bp upstream of promoter proximal paused Pol II. The relaxed sequence specificity of the Myc-Max heterodimer was demonstrated using electrophoretic mobility shift assays (EMSA) on 26 bp fragments of DNA containing either the E-box, related, or unrelated sequence at the centre, and through protein binding microarrays with all possible 8 bp sequences. Even sequences with high binding affinity, which may represent possible novel consensus sequences, were less well correlated with Myc occupancy than Pol II. The study concluded that expected sequence specificity is insufficient to account

for the presence of Myc across the genome. The authors propose a mechanism by which Myc is recruited by “Mediator”, another factor which associates with Pol II, and then binds to DNA in a non-sequence-specific manner at promoters.

In addition to Myc, there is evidence that other transcription factors also have mechanisms of transcriptional regulation that do not require their accepted consensus sequences. NF- κ B has been shown to bind active enhancers, more than half of which do not contain its canonical consensus sequence (Kolovos, Georgomanolis et al. 2016). Direct binding of FOXM1 to DNA through its DNA binding domain to sequences not containing the canonical forkhead consensus sequence has been demonstrated by ChIP (Sanders, Gormally et al. 2015). Finally, low-affinity variant consensus sequences for the GLI transcription factor have been shown to be responsive to moderate concentrations of GLI, demonstrating the functionality of non-consensus sequence-containing promoters (Winklmayr, Schmid et al. 2010).

In light of the analysis described above, it is apparent that there is a gap in understanding of RUNX1 binding and function. While algorithms designed to locate consensus sequences and transcription studies to examine differential expression profiles are valuable methods for identifying consensus sequence mediated RUNX1 gene targets, they have not captured the whole range of RUNX1 action. This is because reliance on promoter consensus sequence identification overlooks targets which differ substantially from the canonical TG(T/C)GGT sequence, and misses uncharacterised distal regulatory elements such as enhancers. Additionally, expression analysis relies on dramatic changes in expression levels which may not be detectable under all conditions, depending on cell type.

Recent genome-wide studies of RUNX1 indicate that the recruitment of RUNX1 to its consensus sequence in gene promoters represents only a small proportion of global RUNX1 binding throughout the genome (Ptasinska, Assi et al. 2014, Trombly, Whitfield et al. 2015). These studies investigated the co-localisation of the leukaemia associated fusion protein RUNX1-ETO with a variety of other transcription factors, such as RUNX1 itself, p300 and N-CoR on a genome-wide scale, and in doing so uncovered new information about the occupancy of RUNX1. In a manner reminiscent of Myc, genome-wide studies revealed that the majority of RUNX1 binding occurs outside of gene promoters, largely within genes and in intergenic regions. Further, many binding sites were entirely devoid of RUNX consensus sequences. Findings from these genome-wide studies require a reassessment of the accepted mechanism of RUNX1 transcriptional action.

1.7 Research aims

RUNX1 is an important transcription factor which governs many of the processes involved in haematopoiesis. Disruption of RUNX1 regulatory networks frequently contributes to the development of leukaemia. It is therefore vital to understand the mechanism by which RUNX1 operates to maintain these networks. ChIP-seq databases which detail RUNX1 binding on a genome-wide scale provide a valuable resource for identifying novel regulatory mechanisms not detected in single target gene studies. Analysis of ChIP-seq data has identified RUNX1 binding sites which are not associated with canonical consensus sequences. Hence, the hypothesis of this study is as follows:

The function of RUNX1 is determined by the manner in which it is recruited to target genes, which occurs either directly through its consensus sequence or indirectly through interaction with other transcriptional proteins or transcription complexes.

Aims: Overall objective: To characterise the different modes of action of RUNX1 in myeloid cells

- Aim 1: To analyse the genomic and sequence context of RUNX1 localisation through analysis of ChIP binding sites in publicly available data
- Aim 2: To characterise the different modes of action of RUNX1 at gene promoters
- Aim 3: To characterise the role of RUNX1 at sites of distal chromatin interactions

2 Materials and Methods

2.1 Cell Culture

Commercially available leukaemic cell lines were used to determine expression of candidate genes, response to RUNX1 overexpression and chromatin interactions involving candidate promoters. K562 cells are derived from a CML patient in terminal blast crisis. They are granulocytic and highly undifferentiated. KG1a cells are derived from KG-1 cells, which are from the bone marrow of an AML patient, and are less mature than the parent lineage. Kasumi-1 cells are myeloblasts with the t(8;21) translocation, resulting in the production of the RUNX1-ETO fusion protein.

2.1.1 Culture conditions

K562, KG1a and Kasumi-1 cells were obtained from American Type Culture Collection and were cultured in Gibco® Roswell Park Memorial Institute (RPMI) 1640 medium (Life Technologies, USA), containing D-glucose (2 g/L), glutathione (1 mg/L), hydroxyethyl piperazineethanesulfonic acid (HEPES; 6 g/L), and sodium bicarbonate (2 g/L, added on preparation). Media was supplemented with heat-inactivated foetal bovine serum (FBS; 10% for K562 and KG1a cells, 20% for Kasumi-1 cells; Life Technologies, USA) and penicillin (100 U/mL)/streptomycin (100 µg/mL; Sigma-Aldrich, USA). K562 and KG1a cells were subcultured every 2-3 days to maintain a concentration between 1×10^5 and 1×10^6 cells/mL. Kasumi-1 cells were subcultured weekly to maintain a similar cellular concentration. All cell lines were cultured in a humidified incubator at 37°C and 5% CO₂.

2.2 Cloning of candidate gene promoters

Promoters of candidate genes were cloned into the pXPG plasmid (Bert, Burrows et al. 2000), which contains a luciferase gene, for use in reporter assays to determine the effects of RUNX1 overexpression on these regions.

2.2.1 Primer design

The promoter regions of candidate genes were cloned into the pXPG plasmid. Primers (Tables 2.1 and 2.2) were designed and checked for specificity for each promoter using the Primer-BLAST program (Ye, Coulouris et al. 2012) to amplify a region of approximately 500-1000 bp in size incorporating both the TSS and the site of a RUNX1 ChIP peak (Trombly, Whitfield et al. 2015). Additional primers (also described in Table 2.1) were designed to amplify a region containing only the proximal ChIP peak in the KAT7 promoter. Recognition sites for *XhoI* and *HindIII* restriction enzymes were included at the 5' end of forward and reverse primers respectively, as well as five random nucleotides for increased restriction enzyme digestion efficiency (Tables 2.1 and 2.2).

Table 2.1. Primers for amplification of HAT promoters.

The primers used to amplify promoter regions of consensus sequence absent candidates are detailed. Restriction enzyme recognition sites are underlined, and the expected size of the PCR fragment is listed.

Promoter	Primer Sequences	Fragment Size
HAT1	Forward: 5'-TGCT <u>ACTCGAGT</u> TCTCCCGCAACTGAAACTC-3'	500 bp
	Reverse: 5'-TGCTAAAGCTTATTTCCGAGCTACGATCACC-3'	
KAT6A	Forward: 5'-TGCT <u>ACTCGAGG</u> GGTTAAGACGGGTGCTTCT-3'	765 bp
	Reverse: 5'-TGCTAAAGCTTGGGAAGGACAGCCGAGATCC-3'	
KAT6B	Forward: 5'-TGCT <u>ACTCGAGA</u> AAGGAGCTGGCGACTGAG-3'	504 bp
	Reverse: 5'-TGCTAAAGCTTGGTCATCTCGTTCCACATAA-3'	
KAT7	Forward: 5'-TGCT <u>ACTCGAGG</u> CCTCGTAACTCTCTGCTCA-3'	889 bp
	Reverse: 5'-TGCTAAAGCTTTTTTCCTTCACTTCCGGCTG-3'	
KAT7 Deletion	Forward: 5'-TGCT <u>ACTCGAGC</u> CCCGCGACATCAACTAACTC-3'	568 bp
	Reverse: 5'-TGCTAAAGCTTTTTTCCTTCACTTCCGGCTG-3'	
KAT2A	Forward: 5'-TGCT <u>ACTCGAGC</u> ACCACACTACTGCAGAACTG-3'	596 bp
	Reverse: 5'-TGCTAAAGCTTGGGGTGGGAGTCGGAATC-3'	
KAT2B	Forward: 5'-TGCT <u>ACTCGAGC</u> CTCCTCCGGCAGATTTGA-3'	560 bp
	Reverse: 5'-TGCTAAAGCTTCTCGGGACTGCGGACTAG-3'	

Table 2.2. Primers for amplification of ARHGAP promoters.

The primers used to amplify promoter regions of consensus sequence present candidates are detailed. Restriction enzyme recognition sites are underlined, and the expected size of the PCR fragment is listed.

Promoter	Primer Sequences	Fragment Size
ARHGAP1	Forward: 5'-TGCTA <u>CTCGAG</u> TGAGGAGAAAGCTGACCACA-3'	814 bp
	Reverse: 5'-TGCTAAAGCTTCTCCCGCTCTTCTATCCC-3'	
ARHGAP4	Forward: 5'-TGCTA <u>CTCGAG</u> CCTCCCACCAAATTGAAGCC-3'	964 bp
	Reverse: 5'-TGCTAAAGCTTGTGGAACCCCACTGCTCC-3'	
ARHGAP9	Forward: 5'-TGCTA <u>CTCGAG</u> ACAGTAGGTACCAAAGGCCC-3'	487 bp
	Reverse: 5'-TGCTAAAGCTTCTCACTTCCTGTTGCCAACTT-3'	
ARHGAP12	Forward: 5'-TGCTA <u>CTCGAG</u> AATGGACTCAGTGCCCCG-3'	500 bp
	Reverse: 5'-TGCTAAAGCTTGTAAGTTACAGGGCACGTCG-3'	

2.2.2 PCR amplification of candidate promoters

Genomic DNA (gDNA) was isolated from K562 cells using the AllPrep DNA/RNA Mini Kit (Qiagen, USA) and used as template DNA for the PCR amplification of all candidate promoters. A 25 µL reaction containing 1X Phusion® High-Fidelity PCR Master Mix with GC Buffer (Thermo Fisher Scientific, USA) and 0.5 µM primers (Tables 2.1 and 2.2) was used to amplify the gDNA as per manufacturer's instructions. The PCR reaction was carried out using a Verti® 96 Well Thermal Cycler (Applied Biosystems™, USA) under the following cycling conditions: initial denaturation at 98°C for 30 seconds, followed by 35 cycles consisting of 30 seconds at 98°C, 30 seconds at 68-72°C (depending on primer set) and 30 seconds at 72°C. After the final cycle, a further 72°C extension step for 7 minutes was included. PCR products were visualised by electrophoresis on 2% agarose gels and amplified promoter regions were excised from the gels. Products were extracted from excised gel bands using the Illustra GFX DNA and Gel Band Purification kit (GE Healthcare, USA) as per manufacturer's instructions. Promoter DNA was eluted in 20 µL Elution Buffer Type 4.

2.2.3 Ligation of PCR products into pXPG

Purified PCR products and pXPG plasmid were digested with *XhoI* and *HindIII* restriction enzymes (New England Biolabs, USA) for 3 hours at 37°C. Digested DNA was purified from the reaction mixture using the Illustra GFX DNA and Gel Band Purification kit (GE Healthcare, USA) as per manufacturer's instructions and eluted in 20 µL Elution Buffer Type 4. Purified digested fragments and plasmid were quantified using a NanoDrop® ND-1000 spectrophotometer (NanoDrop Technologies, USA). Promoter fragments (insert) and pXPG plasmid (vector) were ligated in a 10 µL reaction containing 1 µL T4 DNA Ligase, 1X T4 DNA Ligase Reaction Buffer (New England Biolabs, USA), 20-50 ng digested plasmid DNA and sufficient volume of digested

promoter fragment DNA to ensure a 1:3 vector to insert ratio. Reactions were incubated at room temperature overnight.

2.2.4 Plasmid transformation and isolation

Ligated plasmid vector DNA (5 µL) was transformed into StrataClone SoloPack Competent Cells (25 µL; Agilent, USA) as per manufacturer's instructions. Transformed cells (100 µL) were plated on agar plates containing 100 µg/mL ampicillin (Sigma-Aldrich, USA) and incubated overnight at 37°C. Colonies were screened for successfully ligated vector by PCR using GoTaq Green Master Mix (Promega, USA) as per manufacturer's instructions, with cycling conditions as described in Section 2.2.2. Plasmid DNA was isolated from 1 mL overnight cultures of positive PCR colonies using the QIAprep Spin Miniprep kit (Qiagen, USA) as per manufacturer's instructions and eluted in 20 µL Buffer EB. Plasmid DNA (2 µL) was analysed by restriction enzyme digest and gel electrophoresis to confirm recombinants. Glycerol stocks were prepared by addition of 850 µL glycerol (Sigma-Aldrich, USA) to 150 µL bacterial culture and stored at -80°C.

2.2.5 DNA sequencing

Recombinant plasmids, as confirmed by PCR and restriction enzyme digest, were sequenced using the BigDye® Terminator v3.1 Cycle Sequencing Kit (Applied Biosystems, USA) to verify the identity of the insert and ensure that no mutations had occurred in the promoter. A 10 µL reaction containing 0.25 µL BigDye® Terminator, 1.75 µL BigDye® Terminator sequencing buffer, 0.528 µM of either the forward or reverse PCR primer used to amplify the initial fragment (Tables 2.1 and 2.2), and approximately 200 ng of purified plasmid DNA.

Alternatively, inserts were amplified from plasmid DNA using Phusion® High-Fidelity PCR Master Mix with GC Buffer (Thermo Fisher Scientific, USA) as described in Section 2.2.2,

purified by agarose gel electrophoresis and extracted from gel bands using the Illustra GFX DNA and Gel Band Purification kit (GE Healthcare, USA). The sequencing reaction was prepared as above, using 20 ng PCR amplified insert. The reaction was conducted using a Verti® 96 Well Thermal Cycler (Applied Biosystems™, USA) under the following cycling conditions: initial denaturation at 96°C for 1 minute, followed by 25 cycles consisting of 10 seconds at 96°C, 5 seconds at 50°C and 4 minutes at 60°C. Sequencing products were purified using the Agencourt® CleanSEQ® kit (Agencourt Bioscience Corporation, USA) and eluted in 40 µL water. Sequencing was conducted using an ABI Prism® 310 Genetic Analyser (Applied Biosystems, USA) and data was collected using 310 Data Collection and Sequence Analysis Software (Applied Biosystems, USA). Sequencing files were aligned to the reference promoter sequence and analysed for mutations using the Sequencher 4.10.1 program (Gene Codes Corporation, USA).

2.2.6 Generation of enhancer plasmid constructs

Potential enhancer regions were cloned into promoter plasmids upstream of the promoter insert in both the forward and reverse orientation. Primers were designed to encompass regions of interest at 31 kb upstream and 190 kb downstream of the HAT1 promoter, 100 kb upstream of the KAT6A promoter, as well as the distal RUNX1 ChIP peak of the KAT7 promoter as described in Section 2.2.1 (Table 2.3). Specifically, primers were designed to include either a XhoI restriction site on each primer (HAT1+190 kb and KAT7 distal) or as two pairs of forward and reverse primers, with a XhoI site on either the forward or reverse primer and a BamHI site on its pair (HAT1-31 kb and KAT6A-100 kb). Fragments and promoter plasmids (HAT1-pXPG for HAT1 enhancer fragments and KAT7 distal site) were digested with the appropriate restriction enzymes and ligated, either non-specifically (HAT1+190 kb and KAT7 distal) or in forward and reverse direction (HAT1-31 kb and KAT6A-100 kb) as described in Section 2.2.3. Successfully

ligated plasmids were identified, and the inserts were confirmed for both identity and insert direction as described in Sections 2.2.4-5.

Table 2.3. Primers for amplification of enhancer regions.

The primers used to amplify candidate enhancer regions are detailed. Restriction enzyme recognition sites are underlined, and the expected size of the PCR fragment is listed.

Enhancer	Primer Sequences	Fragment Size
KAT7 distal site	5'-TGCT <u>ACTCGAGG</u> CCTCGTAACTCTCTGCTCA-3'	234 bp
	5'-TGCT <u>ACTCGAGG</u> AGTTAGTTGATGTCGCGGG-3'	
HAT1+190k	5'-TGCT <u>ACTCGAGC</u> CTTGCCCAGCGCGA-3'	534 bp
	5'-TGCT <u>ACTCGAGT</u> CTGGGCACCGATGTACTTT-3'	
HAT1-31k Fwd	5'-TGCT <u>AGGATCCT</u> GACGATATGCAAAACAGTAAATGA-3'	524 bp
	5'-TGCT <u>ACTCGAGC</u> ATGTGGGACTCAGTGGGAC-3'	
HAT1-31k Rev	5'-TGCT <u>AGGATCCC</u> ATGTGGGACTCAGTGGGAC-3'	524 bp
	5'-TGCT <u>ACTCGAGT</u> GACGATATGCAAAACAGTAAATGA-3'	
KAT6A-100k Fwd	5'-TGCT <u>AGGATCCT</u> AAGTGTGGAGCGGGTCA-3'	744 bp
	5'-TGCT <u>ACTCGAGG</u> GGCGCTGGAGTATGTG-3'	
KAT6A-100k Rev	5'-TGCT <u>AGGATCCG</u> GGCGCTGGAGTATGTG-3'	744 bp
	5'-TGCT <u>ACTCGAG</u> AGTAAGTGTGGAGCGGGTTCA-3'	

2.3 Luciferase reporter assays

Luciferase reporter assays were conducted to determine the response of constructed reporter plasmids to RUNX1 (pCMV5-AML1B; Addgene, USA; Meyers and Hiebert, 1995) RUNX1-ETO (pCMV5-AML1-ETO; Addgene, USA, Meyers and Hiebert, 1995), FLI-1 (Addgene, USA), and combinations thereof compared to CMV control (RcCMV; Invitrogen, USA) in K562 cells.

2.3.1 Plasmid preparation

Starter cultures of reporter plasmids were prepared by inoculating 1 mL lysogeny broth (L-broth) supplemented with 100 µg/mL ampicillin (Sigma-Aldrich, USA) from glycerol cultures. Cultures were incubated at 37°C with shaking for 6 hours, then added to 99 mL L-broth (with ampicillin) and incubated for a further 16 hours overnight. The plasmid DNA was extracted and purified using the Qiagen Plasmid Plus Maxi Kit (Qiagen, USA) as per manufacturer's instructions and eluted in 100 µL Buffer EB. Purified plasmid DNA was quantified using a NanoDrop® ND-1000 spectrophotometer (NanoDrop Technologies, USA) and diluted to 1000 ng/µL with water.

2.3.2 Transfection of myeloid cell lines

K562 cells were subcultured 24 hours before transfection to produce a growing culture at approximately 5×10^5 cells/mL. Cells (4.5×10^6 cells in 300 µL RPMI with 20% FBS) were transfected in duplicate, with 10-15 µg plasmid (reporter with or without transcription factor plasmid) added to each, by electroporation in a Gene Pulser® electroporation cuvette at 270 V, 950 µF, and infinite resistance with a Bio-Rad Gene Pulser® XCell™ (Bio-Rad, USA). For some transfections, plasmids were supplemented with RcCMV plasmid to ensure an equal amount of plasmid in each transfection. Transfected cells were supplemented with 1 mL RPMI (10% FBS)

and allowed to recover for 10 minutes. Duplicate transfections were combined and made up to 10 mL with RPMI (10% FBS). Cells were incubated at 37°C and 5% CO₂ for 24 hours post-transfection. Cells were then harvested by centrifugation at 500 g for 5 minutes, washed with 1 x phosphate buffered saline (PBS; Thermo Fisher Scientific, USA) and recentrifuged. Cell pellets were resuspended in 100 µL 1 x Lysis buffer (Promega, USA) and stored at -20 °C. Protein was isolated from cell lysates by centrifugation at 13,000 g for 5 minutes and retention of the supernatant, which was stored at -20°C.

2.3.3 Quantitation of protein by Bradford assay

Protein concentration of cell lysates was determined by Bradford assay. Aliquots of bovine serum albumin (BSA, New England Biolabs, USA) at concentrations of 0.1, 0.2, 0.3, 0.4 and 0.5 mg/mL were prepared for a standard curve. Cell lysates were diluted 1 in 25. Protein Assay Dye Reagent Concentrate (BioRad, USA) was diluted 1 in 5 and 990 µL added to 10 µL diluted lysate or standard. Protein absorbance was measured at 595 nm after 5 minutes using a SpectaMax[®] Plus³⁸⁴ Microplate Reader Spectrophotometer (Molecular Devices, USA). A standard curve was constructed from BSA standard absorbances using Microsoft[®] Excel and protein concentration of lysates was determined.

2.3.4 Quantification of reporter construct activity by luciferase assay

The activity of reporter constructs post-transfection was analysed using the Luciferase Reporter System (Promega, USA). An aliquot of 10 µg of protein was added to alternating wells of a 96 well plate (Greiner Bio-one, Germany), leaving adjacent wells vacant, and made up to 100 µL with 1X Lysis Buffer (Promega, USA). Luciferase reagent (100 µL) was added to each well and the luminescence was measured using a 2.5 second integration time on a Veritas[™] Microplate Luminometer (Turner Biosystems, USA).

2.4 Analysis of mRNA expression by reverse transcription quantitative PCR

Gene expression was analysed in leukaemic cell lines by reverse transcription quantitative PCR (RT-qPCR) from extracted mRNA.

2.4.1 RNA extraction

RNA was extracted from cells using TRI Reagent® (Sigma-Aldrich, USA) as per manufacturer's instructions. Briefly, cell pellets (approximately 4×10^6 cells) were lysed by addition of TRI Reagent® (500 µL) and incubated at room temperature for 5 minutes. RNA was isolated from lysed samples by solvent extraction. Chloroform (100 µL) was added and samples were shaken vigorously for 15 seconds before resting at room temperature for 10 minutes. Samples were centrifuged at 13,000 g for 15 minutes at 4°C and the top aqueous phase containing RNA was retained. RNA was precipitated overnight at -20°C by addition of isopropanol (250 µL). Precipitated RNA was collected by centrifugation, washed with 70% ethanol and resuspended in diethylpyrocarbonate treated water. RNA was quantified using a NanoDrop® ND-1000 spectrophotometer (NanoDrop Technologies, USA).

2.4.2 cDNA synthesis

cDNA was synthesised from RNA samples using the iScript™ cDNA synthesis kit (Bio-Rad, USA) as per manufacturer's instructions. Briefly, a 20 µL reaction consisting of 1 x iScript Reaction Mix, iScript Reverse Transcriptase (1 µL) and RNA (up to 1 µg) was prepared. The reaction was incubated at 25°C for 5 minutes, then 42°C for 30 minutes and 85°C for 5 minutes.

2.4.3 Reverse transcription quantitative PCR

RT-qPCR primers (Table 2.5) were designed for each candidate gene to amplify between 50 and 150 bp of mRNA, ensuring that the transcript crosses an exon boundary by locating each primer in adjacent exons, using the Primer-BLAST program (Ye, Coulouris et al. 2012).

Synthesised cDNA (100 ng) was amplified using the SensiFAST™ SYBR® No-ROX kit (Bioline, UK) in a 25 µL reaction containing 1 x SensiFAST™ Master Mix, DMSO (0.75 µL) and forward and reverse primers (1.25 nmol). Reactions were conducted using a Corbett RotorGene 6000 (Qiagen, USA) with the following conditions: initial denaturation at 95°C for 3 minutes followed by 40 cycles consisting of 95°C for 10 seconds, 60°C for 10 seconds, and 72°C for 20 seconds, with a final extension at 72°C for 3 minutes. The amplification of human Glyceraldehyde 3-phosphate dehydrogenase (GAPDH) or β 2-Microglobulin mRNA transcripts was conducted in parallel with candidate genes to normalise for differences in synthesis or input. To confirm that the qPCR reaction amplified a single product, melt curves were conducted for each primer pair over the range 60-95°C and analysed for the presence of a single peak. Products were visualised by gel electrophoresis and serially diluted to generate standard curves for determination of copy numbers from cycle times.

Table 2.4. Primers for RT-qPCR analysis.

Primers designed to detect levels of transcripts for candidate and control genes and the expected size of the fragment are detailed.

Gene	Primer Sequences	Fragment Size
HAT1	5'-CGGGAAAAATTAATGCAAGG-3'	138
	5'-TTGTTCCGGCATCACTCATGT-3'	
KAT6A	5'-ACTCCACCACCTACGAATGC-3'	91
	5'-AAGCTTTGCCATGTGATCCT-3'	
KAT6B	5'-TTGCTTCAAGAATGCTGACC-3'	85
	5'-TTCGGGATTGTCTTTACTGC-3'	
KAT7	5'-GGGAAAACACCTAGTTTTAAAGAGA-3'	53
	5'-TGGCTATCCACTCATCAATCA-3'	
KAT2A	5'-GCCCTGACTACTACGAGGC-3'	59
	5'-CGCTCAGTCATGGTCTTCAG-3'	
KAT2B	5'-TAAGGTTCCCATGGATCTG-3'	50
	5'-CCTATTCTTGAGGCGTTCCTC-3'	
ARHGAP1	5'-TTGTGTTTAGTGCCTGTCGAA-3'	89
	5'-GTACTGGTCCAGGGTGTGCT-3'	
ARHGAP4	5'-ACACGTGGAGGTGGATAAGG-3'	52
	5'-CTAAACACAGAGTCCATGTTCTG-3'	
ARHGAP9	5'-AACAGCCAGGACAAGAAGGT-3'	72
	5'-CTCCGGTGACCACATGAATG-3'	
ARHGAP12	5'-CATTCGCTTCATCAACCTCA-3'	101
	5'-CTCTCTCGAAGGCATCACG-3'	
ARHGAP6	5'-GCAACAGGACTGTGACCTGA-3'	51
	5'-TGTCATCTTTGGGAATGGT-3'	
GAPDH	5'-AAATATGATGACATCAAGAAGG-3'	67
	5'-AGCCCAGGATGCCCTTGAGGG-3'	
B2-Microglobulin	5'-ACTGAATTCACCCACTGA-3'	114
	5'-CCTCCATGATGCTGCTTACA-3'	

2.5 Chromatin conformation capture

Spatial proximity of the HAT1 promoter region to potential enhancers was analysed in KG1a cells using chromatin conformation capture (3C).

2.5.1 3C primer design

Digestion sites for the selected restriction enzyme (Mbol) were mapped across the region from the ChIP peak in the HAT1 promoter to potential enhancers at 31 kb upstream and 190 kb downstream of the TSS. Regions of DNA between restriction sites were selected for analysis to give an even spread across the region (Figure 5.8). One primer was designed in the promoter region within 50 bp of the Mbol restriction site, and two primers were designed for each selected region using the Primer-BLAST program (Ye, Coulouris et al. 2012) to generate products crossing the restriction site boundaries (Tables 2.6 and 2.7).

Table 2.5. Primers for 3C analysis of HAT1+190k region.

Primers designed for each end of selected fragments generated by restriction enzyme digest for 3C. The location of the fragment relative to the HAT1 promoter ChIP peak and the expected size of the product are detailed.

Fragment location	Primer Sequence	Product Size
Promoter	5'-CGGGTTGATTCGTCCTTCCT-3'	
7,957-8,597 bp	5'-CAAGCAATTGGAAGAGAAAATGTTT-3'	132
	5'-TTGTCCATTGGTATCTTCAGGGATT-3'	69
11,801-12,368 bp	5'-TACTCCTGTATTATTTCCCCCTTC-3'	66
	5'-AGATGAAAGTGTGCCACAAAGAA-3'	153
21,729-22,505 bp	5'-TGGACGGAAAGGGACTGCT-3'	116
	5'-GGCACATAATGTTGTTCTGTCC-3'	106
45,168-45,592 bp	5'-CATAAATATTCCTTGGGCCACTTT-3'	70
	5'-TGGCAACCACATAAATAGGCTAGT-3'	139
79,157-79,764 bp	5'-GCCACTTCCTGCATATAAGTATGA-3'	124
	5'-GTCCAGCAAATAGTGCCGGG-3'	106
105,763-106,162 bp	5'-CCTACTGACTCAAATCCATATGTTC-3'	134
	5'-GTTGGCCAGGATGGTCTTGA-3'	69
155,933-156,333 bp	5'-TCAGTCTTCAGTGATTCAGAAGC-3'	65
	5'-TTTGGTCTCTAAACCTGTTGTGGT-3'	67
173,760-174,499 bp	5'-CCAACGCACTACCCTCCA-3'	91
	5'-ATCCGTCCGCTGTCCTCATT-3'	120
188,436-189,983 bp (Region of Interest)	5'-CCGAAAGAGGATGCGACCAGA-3'	82
	5'-TGAAGCACCTTTCCTAAAGACAT-3'	128

Table 2.6. Primers for 3C analysis of HAT1-31k region.

Primers designed for each end of selected fragments generated by restriction enzyme digest for 3C. The location of the fragment relative to the HAT1 promoter ChIP peak and the expected size of the product are detailed.

Fragment location	Primer Sequence	Product Size
Promoter	5'-CTCTCTGGCCCTAGGAAAGTTTC-3'	
4,354-5,005 bp	5'-TTCAACATGAGCCGGTCCATTT-3'	139
	5'-GGGGGAATTTAGAGATGGGGATTA-3'	131
7,683-8,181 bp	5'-GACCTTCACGGGTGCACAAG-3'	62
	5'-GGCGGAAGGTGAGTAAGCAA-3'	78
12,087-12,420 bp	5'-TGGAGCAAGGTAAGAATCTCCTG-3'	83
	5'-GATTGCAGATAACGGGAAAAGTCA-3'	109
13,632-13,987 bp	5'-TTGGTATAATGCGGAGGCCCC-3'	122
	5'-CCACACATAGTCCCACACCTATC-3'	140
17,818-18,291 bp	5'-TAACACTTATCTCAGTGCCTGCT-3'	98
	5'-TCATCTACTGGGCAGTTGCTTC-3'	68
18,920-19,146 bp	5'-GAATTGGGATATGCCCAGGAA-3'	106
	5'-GCAATCCAAGGACACAATGACC-3'	71
26,381-26,892 bp	5'-CTTCCTCACAGCCTGACTCTAA-3'	94
	5'-TGAAAACATACACAGGGTCGGG-3'	114
30,278-31,093 bp (Region of Interest 1)	5'-AGACTTTTGAATGGGAAGGAAC-3'	51
	5'-CTGTCCCTAGCTCTTGACAACTTT-3'	122
31,268-31,626 bp (Region of Interest 2)	5'-GTTGTCAAAGAATCGGCCAGG-3'	113
	5'-GCCTAACGCAGCAAGAAAGA-3'	129

2.5.2 Fixation of cells with formaldehyde

KG1a cells were cultured to a concentration of 5×10^5 cells/mL and aliquots of 5 million cells in 25 mL RPMI medium were prepared. Cells were incubated with 1% formaldehyde at room temperature for 10 minutes. The reaction was quenched by addition of 2.1 mL of 1M glycine and incubation on ice for 5 minutes. To extract fixed nuclei, cells were washed and resuspended in 1 mL nuclei buffer (consisting of 10 mM Tris, pH 7.5, 10 mM NaCl, 3 mM MgCl₂, 0.1 mM EDTA, pH 8, 0.5% Igepal, 56 ng/mL spermine and 28 ng/mL spermidine) with 1 µL/mL Complete EDTA-Free Protease Inhibitor Cocktail (Roche Applied Science, Switzerland) and incubated on ice for 15 minutes with occasional agitation.

2.5.3 Restriction enzyme digestion of chromatin

Nuclei were washed and resuspended in CutSmart Buffer (358 µL; New England Biolabs, USA). SDS (10%, 11 µL) was added and the solution was incubated at 37°C for 1 hour with constant agitation. The reaction was returned to ice and TritonX-100 (10%, 75µL) was added and carefully mixed. Mbol (500 U) was added and the reaction was incubated at 37°C overnight.

2.5.4 Ligation and purification of digested chromatin

Digested chromatin was placed on ice and ligated in-nucleus by addition of 10 x Ligation Buffer (820 µL; New England Biolabs, USA) and BSA (82 µL, 10 mg/mL) in 6.71 mL Nuclease-free water. Blunt/TA ligase master mix (250 µL; New England Biolabs) was added and the mixture was incubated at room temperature for 4 hours. Ligated chromatin was treated with Proteinase K (50 µL, 10 mg/mL, incubated overnight at 65°C followed by another addition of 50 µL and 2 hours further incubation) and RNase A (12.5 µL, incubated at 37°C for 30 minutes) and purified by phenol:chloroform extraction. Extracted DNA was ethanol precipitated

overnight at -20°C, collected by centrifugation, washed with 70% ethanol and resuspended in 1 x TE Buffer. Samples were stored at -20°C.

2.5.5 Analysis of 3C products by qPCR

Ligated chromatin was analysed by qPCR for the presence of chromatin regions ligated to candidate promoters. SYBR Premix Ex Taq II (Tli RNase H Plus) (Takara Bio Inc., USA) was used for analysis as per manufacturer's instructions. Briefly, a 25 µL reaction containing 1 x SYBR Premix Ex Taq II, 1 x ROX reference Dye, 0.4 µM forward and reverse primers and 50 ng template DNA was prepared. PCR was conducted using a QuantStudio 3 Real-Time PCR System (Applied Biosystems, USA) with the following conditions: initial denaturation at 95°C for 30 seconds, followed by 40 cycles consisting of 95°C for 5 seconds, then 60°C for 30 seconds. To confirm that successful qPCR reactions amplified a single product, melt curves were conducted for each primer pair over the range 60-95°C and analysed for the presence of a single peak. Products were visualised by gel electrophoresis and serially diluted to generate standard curves for determination of copy numbers from cycle times.

3 Analysis of genome-wide RUNX1 localisation

3.1 Introduction

Results generated from a genome-wide approach to determining transcription factor localisation, together with single gene studies, as outlined in Chapter 1, suggests that the current understanding of RUNX1 recruitment and function at DNA is incomplete. Pairing ChIP assays with sequencing technology provides a snapshot of the genomic localisation of a transcription factor within a cell at a particular time. One such ChIP-Seq study of Runx1 binding in murine embryonic stem cells identified its consensus sequence TG(T/C)GGT in only just over half of all Runx1-bound regions, suggesting that binding in the remaining regions may be due to either novel consensus sequences or recruitment by other factors (Tanaka, Joshi et al. 2012). A novel sequence was proposed, which comprises a longer consensus sequence and contains a TGTAGT motif in the 3' region. This sequence is only a single base mismatch to the canonical RUNX1 consensus sequence and was found to be enriched in Runx1-bound regions that did not contain the canonical consensus sequence. However, binding through this variant sequence still only accounts for a small proportion of the binding within these regions.

A gene-specific investigation of RUNX1 regulation of the gene encoding the integrin $\beta 4$ reinforces this finding. Reporter studies of the ITGB4 gene indicate that the gene is regulated by RUNX1 and sequential deletion constructs of the ITGB4 promoter and subsequent reductions in responsiveness narrow the responsive region to between 175 and 58 bp upstream from the TSS (Phillips, Taberlay et al. 2018). Though this region contains putative RUNX1 consensus sequences, deletion and mutation of these sequences had no effect on responsiveness, indicating that regulation of the ITGB4 gene is not dependent on consensus sequence specificity for RUNX1 activation, at least in this context.

Two ChIP-seq databases have recently been compiled which describe RUNX1 binding in Kasumi-1 cells. Kasumi-1 cells are a leukaemic cell line which expresses both RUNX1 and the fusion protein RUNX1-ETO. Firstly, a study by Trombly et al. examined the co-localisation of RUNX1 and the fusion protein RUNX1-ETO with the nuclear co-repressor protein (N-CoR) and co-activator p300 (Trombly, Whitfield et al. 2015). This study also generated ChIP-Seq libraries for histone modifications H3K4me3 and H3K27me3, associated with gene activation and repression, respectively. Secondly, a study by Ptasinska et al. investigated the dynamics of RUNX1/RUNX1-ETO regulation, demonstrating that the two complexes compete for the same genomic sites (Ptasinska, Assi et al. 2014). Additionally, this study determined the response of the transcriptional regulatory system after knock-down of RUNX1 or RUNX1-ETO, leading to impairment of cellular viability and loss of self-renewal respectively. While both of these studies generated whole-genome localisation data for RUNX1 and RUNX1-ETO, neither examined the genomic context of either in any great detail. Nor did they investigate the mechanisms of recruitment of the transcription factor to regulatory regions, beyond identifying that RUNX1 and RUNX1-ETO bound regions display enrichment of consensus sequences. Of relevance, ChIP-seq studies conducted in mouse cells found that only a minority of RUNX1 binding occurred in promoter regions, with the rest located largely in intergenic and intronic regions (Tanaka, Joshi et al. 2012, Zang, Luyten et al. 2016).

The model of RUNX1 gene regulation in which it recognises and binds canonical TG(T/C)GGT consensus sequences in the promoter regions of its target genes therefore does not fully describe the genome-wide localisation of RUNX1, leading to the hypothesis that alternative mechanisms are responsible for RUNX1 binding outside of promoter regions and in the absence of its consensus sequence. Herein, the genomic and sequence context of RUNX1 and

RUNX1-ETO binding has been examined within a publicly available ChIP-seq dataset with the aim of characterising atypical RUNX1 binding events and using results from this analysis to identify RUNX1-bound genes of potential biological relevance for use as candidate genes in functional studies.

3.2 Results

3.2.1 Genome-wide analysis of RUNX1 ChIP-Seq data

Publicly available data from the human RUNX1 ChIP-seq study conducted by Trombly et al., including called ChIP peaks, was accessed from the Gene Expression Omnibus (Trombly, Whitfield et al. 2015; GSE62847). The RUNX1 and RUNX1-ETO peaks from a second study (Ptasinska, Assi et al. 2014; GSE29225) were used for confirmation and cross-referencing. Both studies were conducted in the Kasumi-1 cell line, using different antibodies, but both recognising the RUNX1 C-terminal domain, which is not shared by RUNX1-ETO. To detect RUNX1-ETO, an N-terminal antibody was used, which would also be expected to bind to the wild-type ETO protein.

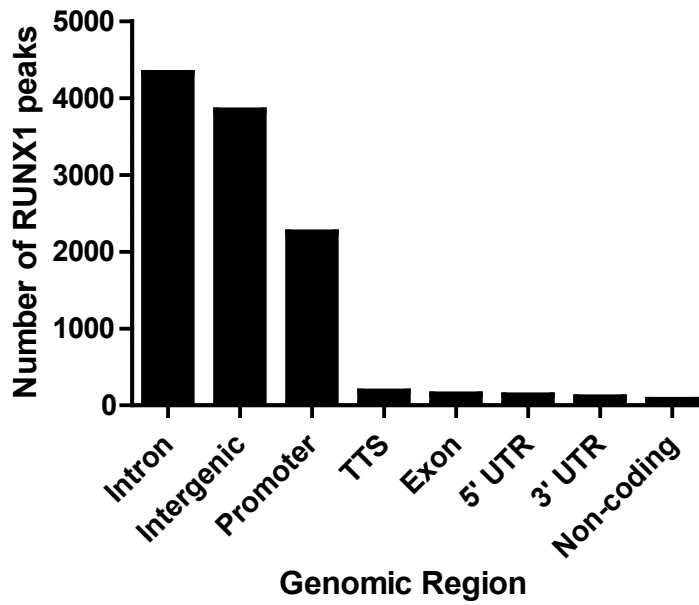
3.2.1.1 *Genomic context of RUNX1 binding*

The ChIP peaks from the Trombly dataset were annotated to the hg19 reference genome using HOMER (Heinz, Benner et al. 2010, <http://homer.ucsd.edu/homer/index.html>), which determines the centre of each peak and provides information about the surrounding genomic context. Each peak was defined by a 100 bp region centred on the middle of the ChIP peak and assigned a unique identifier. The genomic context was also determined by HOMER, using the default classification cut-off points and a hierarchical system in case of multiple applicable region classifications. Regions were designated transcription start sites (TSS, redefined as “promoter” in this data, if located from -1 kb to +100 bp of a TSS), transcription termination

site (TTS, from -100 bp to +1 kb of the end of a gene), exonic, 5' or 3' UTR, non-coding, intronic or intergenic. HOMER uses the RefSeq database to determine the distance from the centre of the peak to the nearest TSS. The nearest gene was identified and gene identifiers from Unigene, RefSeq and Ensembl databases were retrieved. Available data on gene type (i.e. protein coding, non-coding RNA, etc.) and a brief description of the gene according to RefSeq were also recorded.

Analysis of the genomic context of RUNX1 binding in the dataset (Trombly, Whitfield et al. 2015) is depicted in Figure 3.1a and revealed that the majority of RUNX1 binding occurred in intronic (4513 sites, 38.6%) or intergenic (3989 sites, 34.1%) regions. Promoter binding was the next most common, with 2511 sites (21.5%). Other annotated regions; TTS, 5' and 3'UTRs, exons, and non-coding regions, accounted for only a small proportion of all binding (Figure 3.1a). This mirrors what was described in previous studies of Runx1 binding and genomic localisation in mouse megakaryocytes and haematopoietic precursor cells, which reported only 5-12% binding in promoter regions, with intergenic and intronic regions making up 49-50% and 37-44% respectively (Wilson, Foster et al. 2010, Zang, Luyten et al. 2016). Distribution of genomic context for the RUNX1-ETO ChIP peaks was similar overall (Figure 3.1b), with 7109 intronic (41.4%), 5860 intergenic (34.2%) and 2859 promoter (17.5%) peaks with relatively little binding in other regions. This was as expected, as a 71% overlap between RUNX1 and RUNX1-ETO ChIP peaks was previously reported from this dataset (Trombly, Whitfield et al. 2015).

A



B

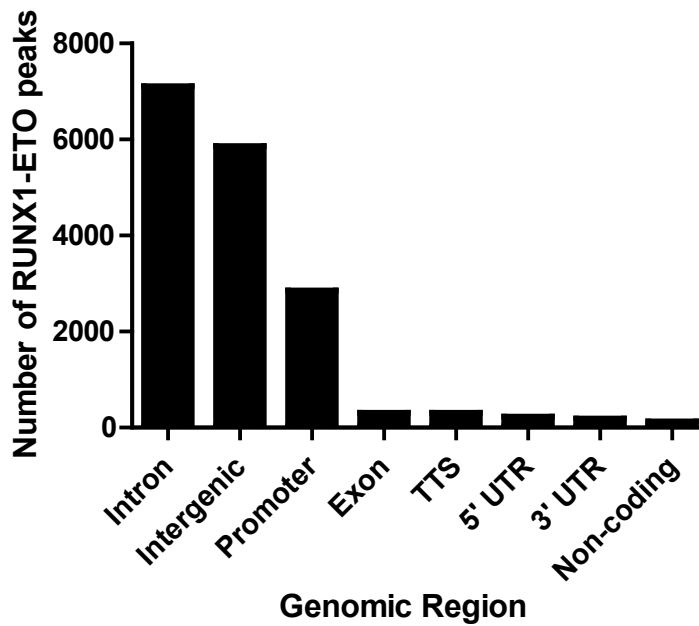


Figure 3.1. Genomic distribution of RUNX1 and RUNX1-ETO ChIP binding in Kasumi-1 cells.






Number of ChIP peaks in each genomic region as categorised by HOMER analysis of data from Trombly, Whitfield et al., 2015 et. al for **A)** RUNX1 or **B)** RUNX1-ETO is shown.

3.2.1.2 *Presence of consensus sequences at RUNX1 binding sites*

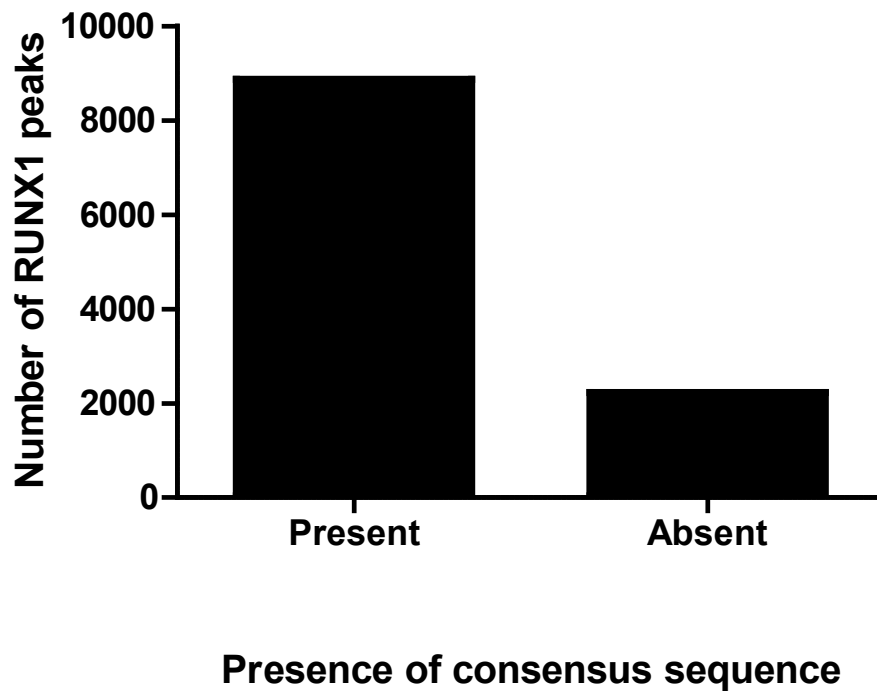
Using HOMER's known motif function, each peak was assessed for the presence of any of four RUNX family consensus sequences, plus an ETS:RUNX hybrid sequence, within 500 base pairs up or down stream (Heinz, Benner et al. 2010). HOMER's known motif function uses sequences experimentally determined from previous studies (Barski, Jothi et al. 2009, Hollenhorst, Chandler et al. 2009, Wilson, Foster et al. 2010, Little, Noushmehr et al. 2011, Sanda, Lawton et al. 2012) (Table 3.1). A peak with at least one of these sequences present within this window was classified as "consensus present" and a peak with no consensus sequences within 500 bp was classified as "consensus absent". "Consensus present" peaks accounted for the majority of identified RUNX1 (79.8%) and RUNX1-ETO (75%) binding (Figures 3.2a, 3.2b). However, "consensus absent" ChIP peaks accounted for 1 in 5 binding sites for RUNX1 and 1 in 4 for RUNX1-ETO.

Table 3.1. RUNX Consensus Sequences.

Position weight matrices used by HOMER motif analysis to determine presence or absence of consensus sequences in the ChIP dataset generated by Trombly, Whitfield, et al., 2015. Motifs were determined from previous experimental studies (Barski et al., 2009; Hollenhorst et al., 2009; Little et al., 2011; Sanda et al., 2012; Wilson et al., 2010).

Motif	Experimental model
	RUNX-AML(Runt)/CD4+-PolIII-ChIP-Seq
	RUNX1(Runt)/Jurkat-RUNX1-ChIP-Seq
	RUNX(Runt)/HPC7-Runx1-ChIP-Seq
	RUNX2(Runt)/PCa-RUNX2-ChIP-Seq
	ETS:RUNX(ETS,Runt)/Jurkat-RUNX1-ChIP-Seq

A



B

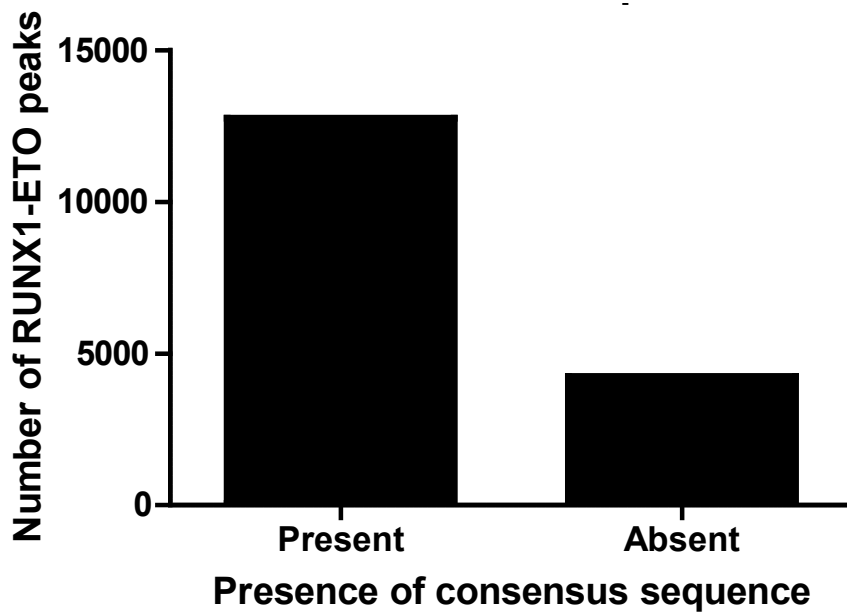
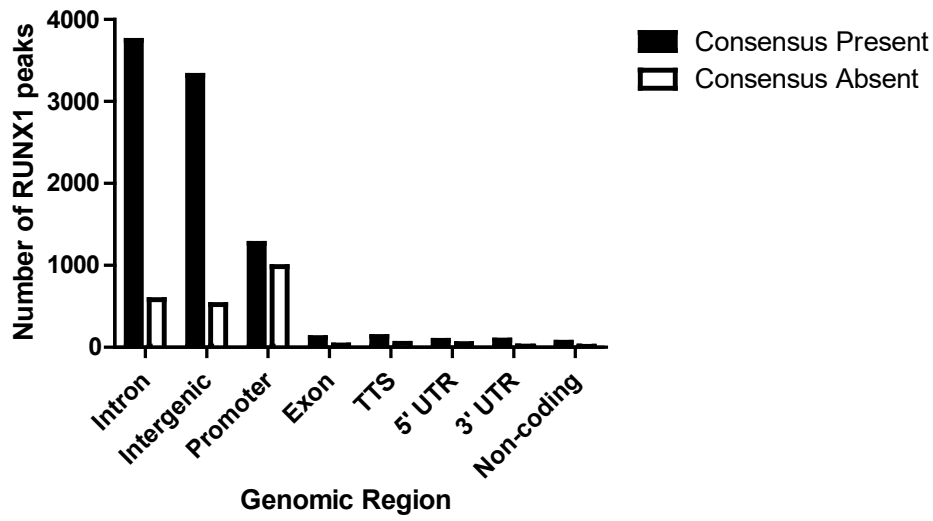


Figure 3.2. Presence of RUNX consensus sequences at ChIP peaks.

Number of ChIP peaks within the Trombly dataset (Trombly, Whitfield et al. 2015) with or without a RUNX consensus sequence within 1000 bp (± 500 bp) for **A)** RUNX1 or **B)** RUNX1-ETO.

To determine if peaks occurring in the absence of a consensus sequence are distributed consistently throughout the genome, the genomic context of the “consensus present” and “consensus absent” peaks was determined. Considered as a percentage of total number of RUNX1 ChIP peaks occurring in each genomic region, a consensus sequence was associated with the majority (81-88%) of peaks identified in intergenic regions, introns, exons, non-coding regions and 3’ UTR (Figure 3.3a), comparable to the genome-wide average of 79.8% (Figure 3.2a). However, peaks annotated to TTS, 5’ UTRs, and promoter regions had a comparatively higher proportion of peaks found to lack a consensus sequence (26%, 35% and 44% respectively; Figure 3.3a). There is therefore a disparity in localisation of “consensus absent” RUNX1 ChIP peaks. When all peaks are considered, RUNX1 binding is most commonly distributed in intergenic and intronic regions with about 20% at promoter regions (Figure 3.4a). This distribution is similar if only the “consensus present” peaks are considered (Figure 3.4b). In contrast, in promoter regions 1080 peaks were identified that were not associated with a consensus sequence (Figure 3.4c), almost twice as many as observed for any other region, despite promoter ChIP peaks only being the third most common binding location overall.

A



B

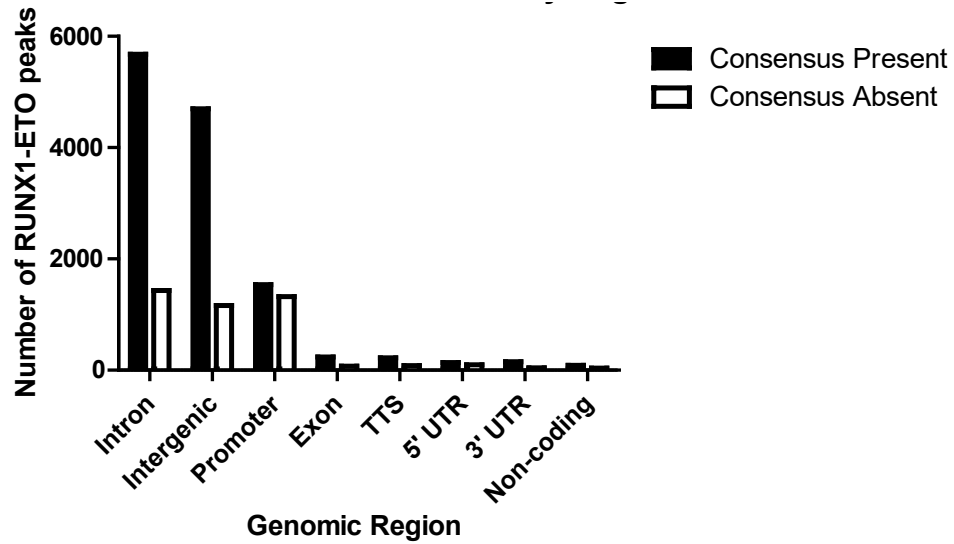
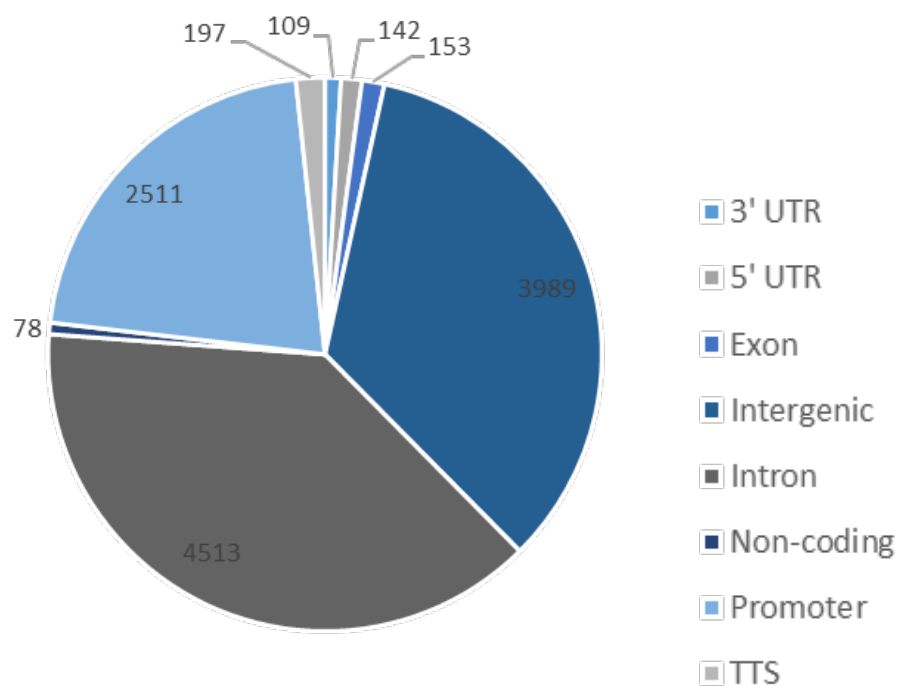


Figure 3.3. Presence of RUNX1 consensus sequence associated with a ChIP peak by genomic region.

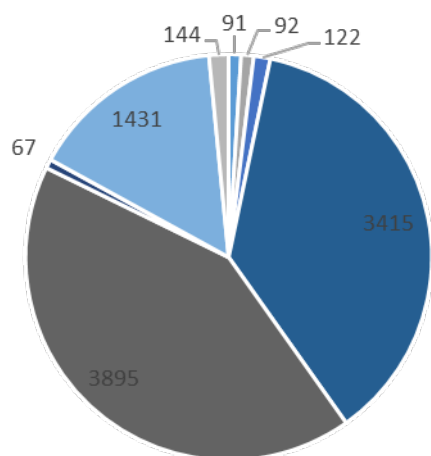
Number of ChIP peaks with or without a RUNX consensus sequence within 1000 bp (± 500 bp) per genomic region as determined by HOMER for **A)** RUNX1 or **B)** RUNX1-ETO.

The similarity in genomic distribution of RUNX1-ETO to RUNX1 peaks also held for the presence of consensus sequences within these peaks (Figures 3.3b, 3.5a and 3.5b). However, the RUNX1-ETO peaks that were not associated with a consensus sequence show a slightly different distribution pattern to RUNX1 (Figure 3.5c) as while promoter regions have a comparably higher percentage of peaks lacking a consensus sequence (almost 50%), localisation to promoters make up only approximately one third of total RUNX1-ETO binding in the absence of a consensus sequence (Figure 3.5c) compared to nearly half for RUNX1 (Figure 3.4c). This reflects higher numbers of RUNX1-ETO CHIP peaks within intergenic and intronic regions that did not contain consensus sequences compared to RUNX1, with approximately 20% of the RUNX1-ETO peaks in these regions lacking consensus sequences. Despite this, RUNX1-ETO binding shows a similar enrichment of “consensus absent” binding in promoter regions and 5' UTRs (46% and 41%, respectively) though the increase in TTS regions is not observed as strongly with only 27% “consensus absent” binding.

A



B



C

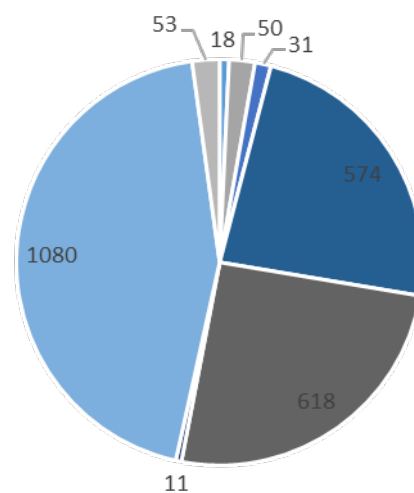
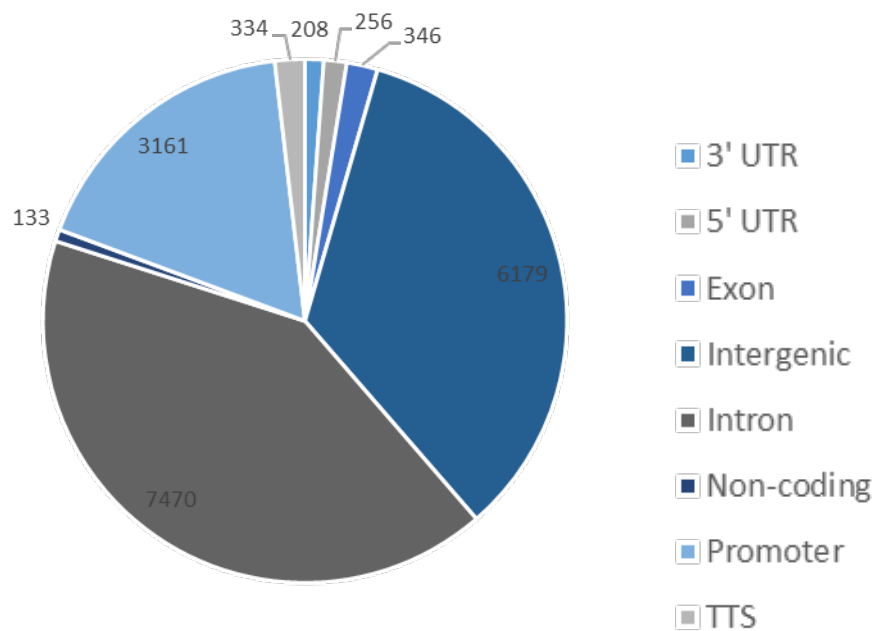


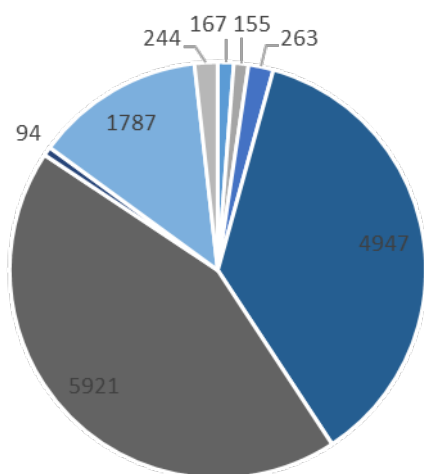
Figure 3.4. The genomic distribution of RUNX1 ChIP peaks categorised by presence of a consensus sequence.

Breakdown of genomic context as determined by HOMER annotation of **A)** all RUNX1 ChIP peaks **B)** ChIP peaks containing consensus sequences and **C)** ChIP peaks in which a consensus sequence was not identified. For legend, see **A**.

A



B



C

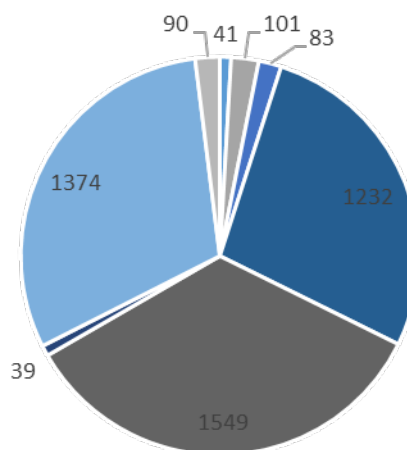


Figure 3.5. The genomic distribution of RUNX1-ETO ChIP peaks categorised by presence of a consensus sequence.

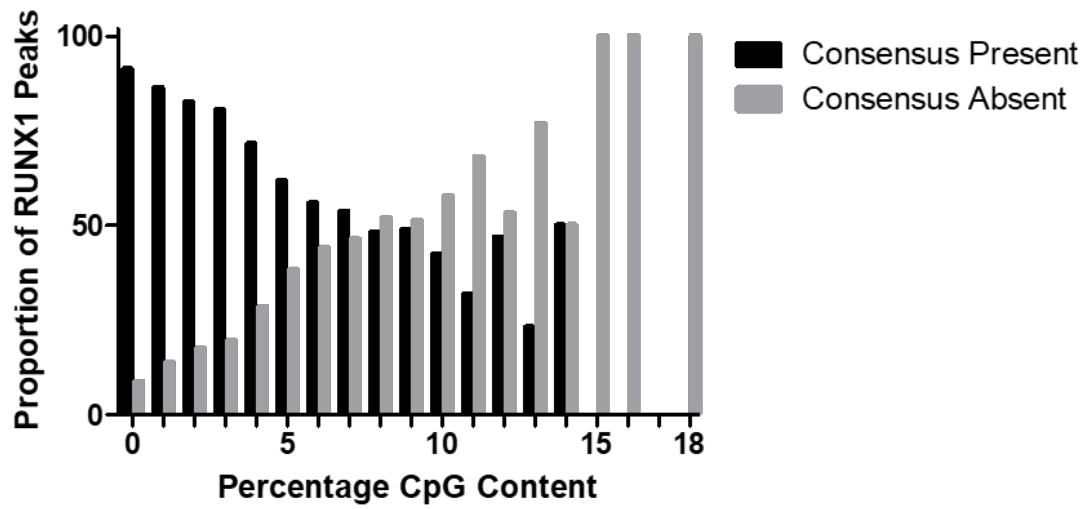
Breakdown of genomic context as determined by HOMER annotation of **A)** all RUNX1-ETO ChIP peaks **B)** ChIP peaks containing consensus sequences and **C)** ChIP peaks in which a consensus sequence was not identified. For legend, see **A**.

To summarise, promoter regions comprise only a relatively small proportion of RUNX1 and RUNX1-ETO occupancy, with the majority of localisation occurring in intergenic or intronic regions. Additionally, RUNX1 and RUNX1-ETO ChIP peaks have been identified in the absence of recognised RUNX consensus sequences across the genome, but the proportion of ChIP peaks which are not associated with a consensus sequence is higher in promoters than other regions, accounting for almost half of observed promoter RUNX1 ChIP peaks.

3.2.1.3 Epigenetic context of RUNX1 binding

The epigenetic environment of a DNA locus influences the ability and mechanism by which transcription factors interact with the genome. In particular, CpG dinucleotides are susceptible to epigenetic modification by methylation, which can lead to transcriptional silencing when accumulation of this modification occurs in CpG islands in gene promoters (Jones 2012). Therefore, the GC context of RUNX1 ChIP peaks in promoter regions was considered in order to assess the potential impact of DNA methylation on the regulation of RUNX1 target genes. Specifically, the CpG context of each promoter ChIP peak was analysed using HOMER to determine percentages of CpG sites and GC content. The percentage of CpG sites ranged from 0 to 18% for annotated ChIP peaks, meaning that for each 100 bp region up to 9 out of 50 dinucleotides were CpGs. Interestingly, a clear trend emerged for both RUNX1 and RUNX1-ETO ChIP peaks in which higher CpG percentages were observed for peaks without consensus sequences than for those in which consensus sequences were present (Figure 3.6). This trend was likewise evident in the GC% of the ChIP peak regions, calculated from the number of guanine and cytosine residues in the 1000 bp surrounding the centre of the ChIP peak, which ranged from 25 to 82% GC content with higher percentages apparent for “consensus absent” peaks (Figure 3.7). This potentially demonstrates a preference for RUNX1 recruitment to CpG enriched promoters in the absence of its consensus sequence.

A



B

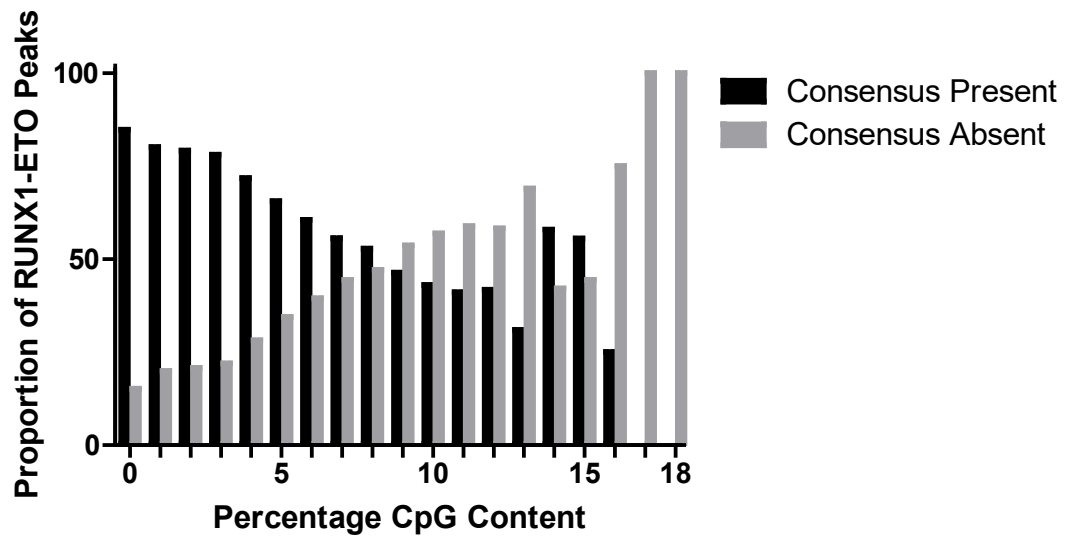
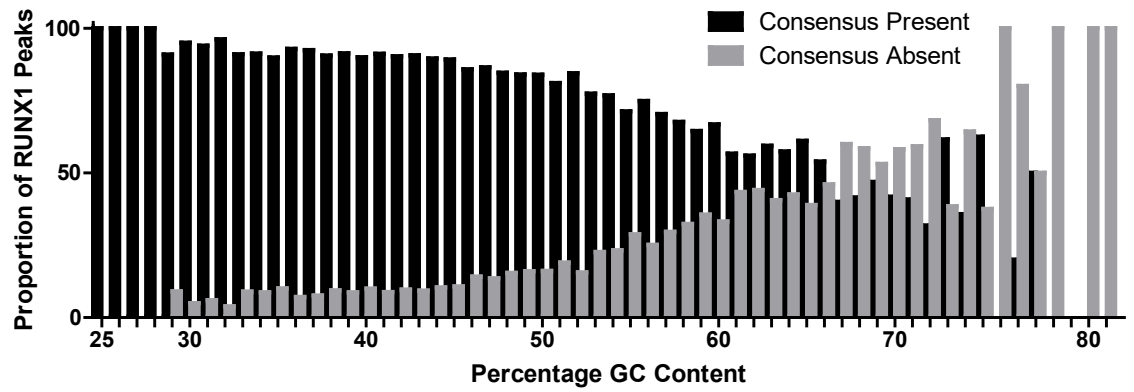


Figure 3.6. Percentage CpG content categorised by presence of consensus sequence.

Proportion of consensus present or absent **A)** RUNX1 and **B)** RUNX1-ETO promoter ChIP peaks separated by percentage of CpG sites in the 100 bp region surrounding each peak (to the nearest 0.01%).

A



B

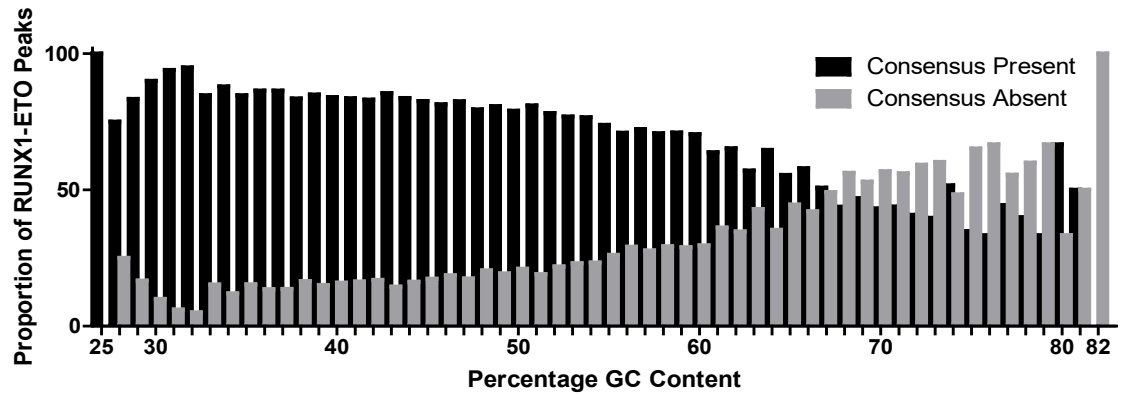


Figure 3.7. Percentage GC content categorised by presence of consensus sequence.

Proportion of consensus present or absent **A)** RUNX1 and **B)** RUNX1-ETO ChIP peaks separated by percentage of GC nucleotides in the 1000 bp region surrounding each peak (to the nearest 1%).

3.2.1.4 *Co-localisation of RUNX1 with co-factors*

Information on co-factor binding, specifically p300 and N-CoR (nuclear receptor corepressor), was also available from ChIP-seq data generated by Trombly and colleagues (Trombly, Whitfield et al. 2015). These factors were chosen in the original study as they are known to interact with RUNX1 and RUNX1-ETO respectively. p300 is typically associated with gene activation and interacts with numerous nuclear proteins as a coactivator (Eckner, Ewen et al. 1994). Conversely, N-CoR is a repressive cofactor which facilitates a reduction in target gene expression (Hörlein, Näär et al. 1995). Distance to the nearest co-factor ChIP peak was determined for RUNX1 and RUNX1-ETO peaks in intronic, intergenic and promoter regions, as these represent the majority of the dataset. The distance from the RUNX1 peak was arbitrarily categorised as <500 bp, <10,000 bp, intervals of 10,000 bp up to 50,000 bp, and >50,000 bp. For RUNX1/RUNX1-ETO ChIP peaks within intergenic and intronic regions, the nearest p300 peak was most commonly located either within 500 base pairs or further than 50,000 bp, with more distant peaks more prevalent (Figure 3.8). For RUNX1/RUNX1-ETO ChIP peaks in promoter regions the nearest p300 peak was more frequently within 500 base pairs, which is to be expected as p300 itself tends to localise to promoters, as well as active enhancers (Holmqvist and Mannervik 2013). For both RUNX1 and RUNX1-ETO ChIP peaks, an enrichment for p300 peaks within 500 base pairs was observed for peaks that did not contain a consensus sequence, compared to peaks where a consensus sequence was present. This was apparent across all genomic regions, but particularly evident in promoter regions (Figures 3.8, 3.9).

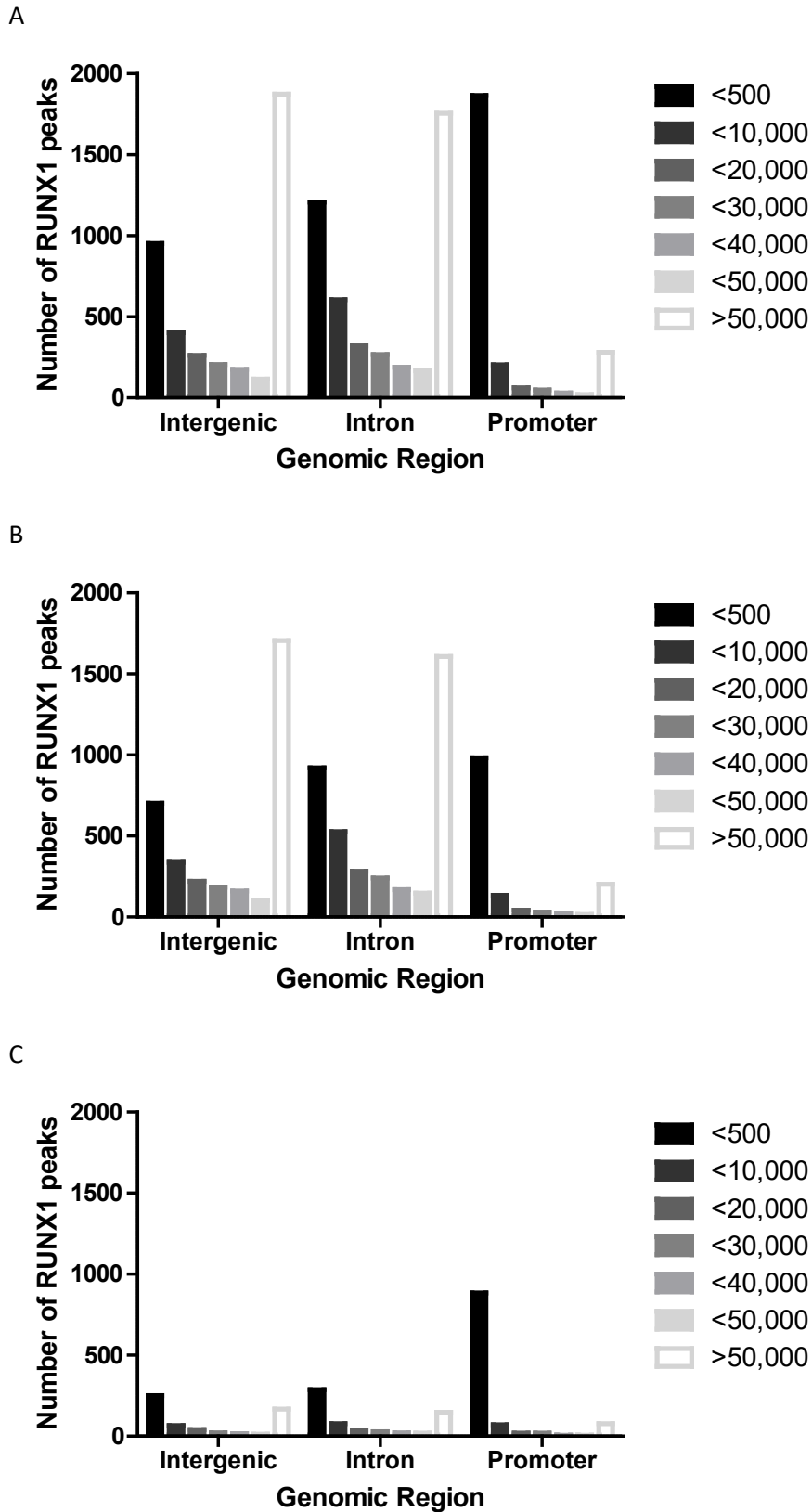


Figure 3.8. Distance to the nearest p300 ChIP peak by genomic context of RUNX1 peak.

Number of RUNX1 intergenic, intronic and promoter ChIP peaks categorised by the distance to the nearest p300 ChIP peak for **A)** all peaks, **B)** ChIP peaks containing consensus sequences, and **C)** ChIP peaks in which a consensus sequence was not identified.

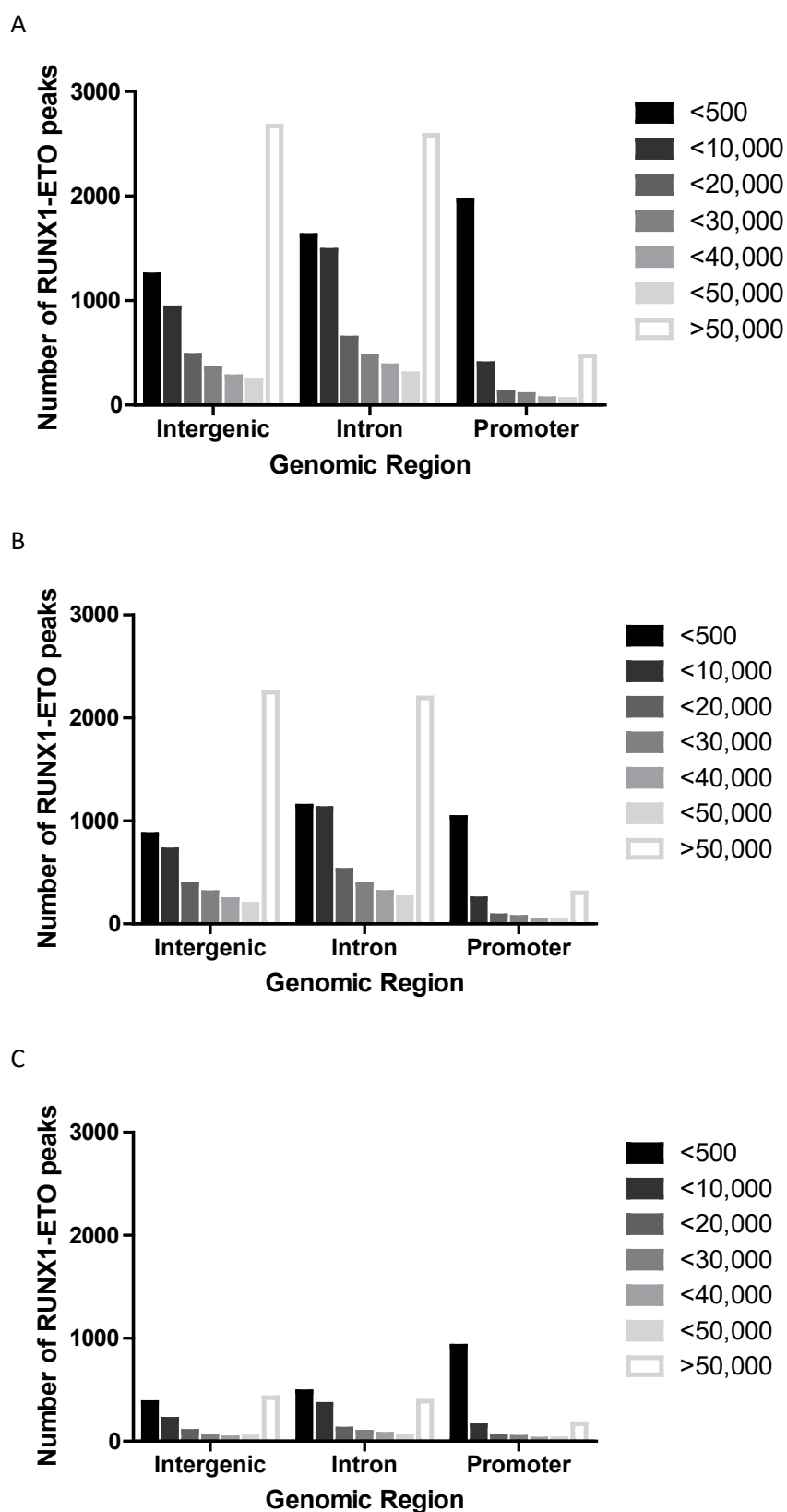


Figure 3.9. Distance to the nearest p300 ChIP peak by genomic context of RUNX1-ETO peak.

Number of RUNX1-ETO intergenic, intronic and promoter ChIP peaks categorised by the distance to the nearest p300 ChIP peak for **A)** all peaks, **B)** ChIP peaks containing consensus sequences, and **C)** ChIP peaks in which a consensus sequence was not identified.

N-CoR peaks were located within 500 base pairs of the RUNX1 or RUNX1-ETO ChIP peak in more than half of all intergenic and intronic cases and in almost half of promoter cases (Figures 3.10a, 3.11a). The pattern of binding of N-CoR relative to RUNX1 / RUNX1-ETO is similar when only “consensus present” or “consensus absent” peaks are considered (Figures 3.10b-c, 3.11b-c), indicating that N-CoR localisation was not influenced by whether RUNX1 or RUNX1-ETO was recruited in the presence or absence of consensus sequences. This analysis demonstrates an increased localisation of p300, but not N-CoR in the vicinity of RUNX1 and RUNX1-ETO when they are bound in the absence of a consensus sequence, which may imply a functional relevance for the co-factor in consensus sequence independent RUNX1 recruitment or function.

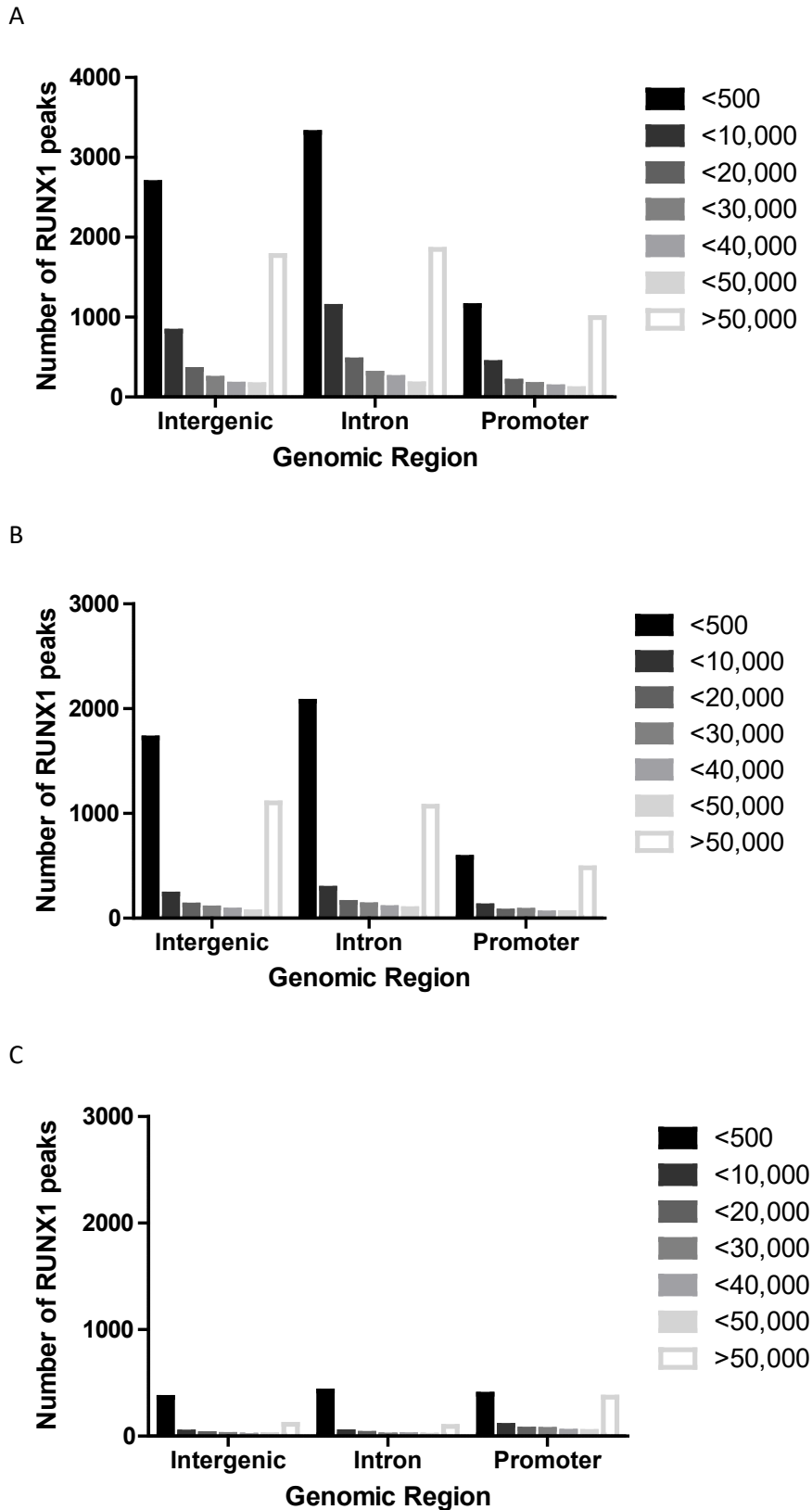


Figure 3.10. Distance to the nearest N-CoR ChIP peak by genomic context of RUNX1 peak.

Number of RUNX1 intergenic, intronic and promoter ChIP peaks categorised by the distance to the nearest N-CoR ChIP peak for **A)** all peaks, **B)** ChIP peaks containing consensus sequences, and **C)** ChIP peaks in which a consensus sequence was not identified.

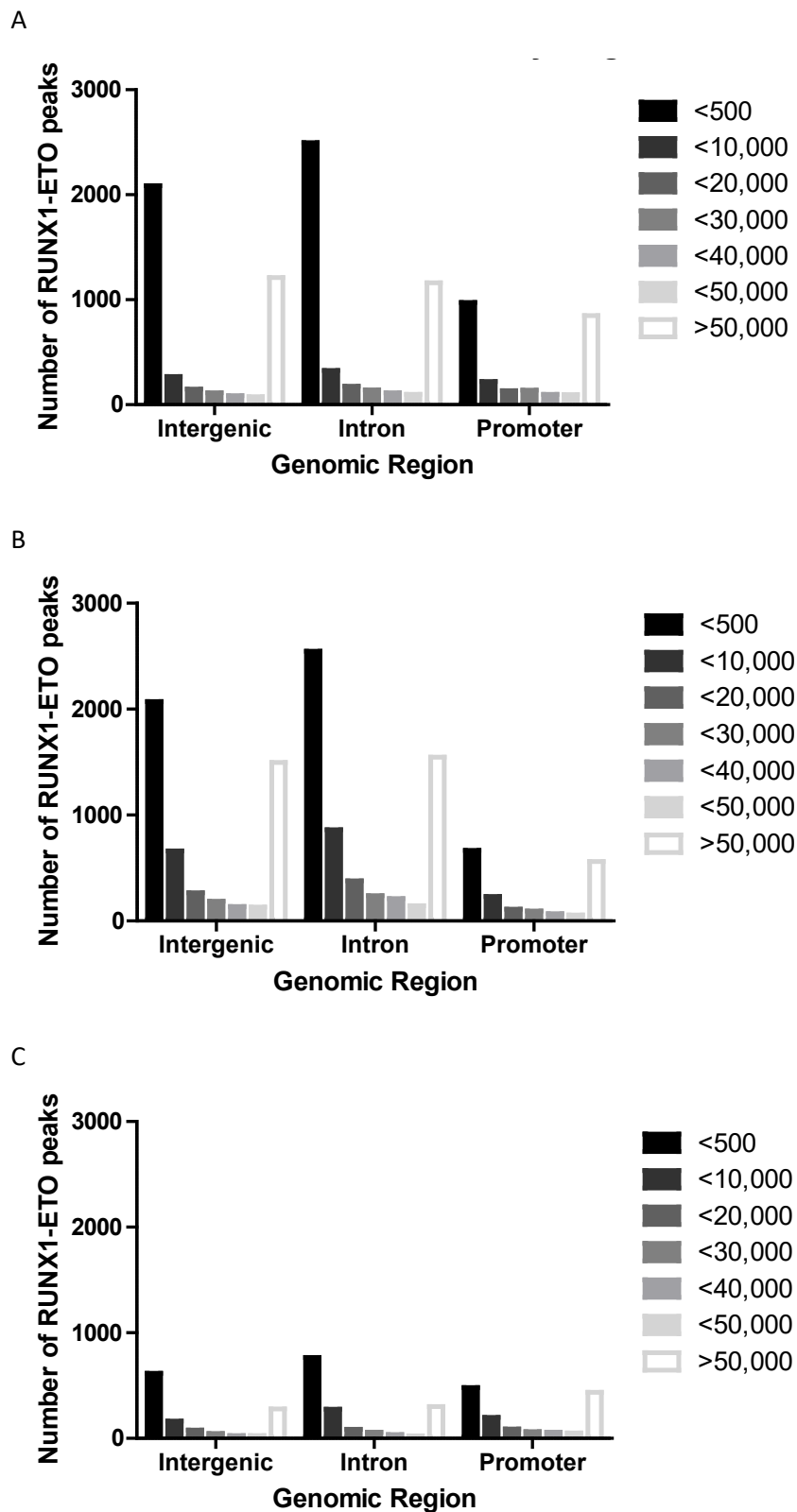


Figure 3.11. Distance to the nearest N-CoR ChIP peak by genomic context of RUNX1-ETO peak.

Number of RUNX1-ETO intergenic, intronic and promoter ChIP peaks categorised by the distance to the nearest N-CoR ChIP peak for **A)** all peaks, **B)** ChIP peaks containing consensus sequences, and **C)** ChIP peaks in which a consensus sequence was not identified.

3.2.2 Analysis of RUNX1 binding at promoter regions

3.2.2.1 *Identification of alternative consensus sequences for RUNX1 recruitment*

As RUNX1 binding in the absence of a recognised consensus sequence is more common in gene promoters than other regions of the genome, this subset of RUNX1 ChIP peaks was selected for further investigation in order to understand the mechanism of RUNX1 recruitment in the absence of consensus sequence binding. Approximately 50% of promoter RUNX1 peaks were not associated with a RUNX consensus sequence, and therefore HOMER was used to identify any common sequence in these promoters that may be responsible for RUNX1 recruitment. The genomic positions of each “consensus absent” promoter peak were compiled and entered into the HOMER findMotifsGenome.pl program (Heinz, Benner et al. 2010, <http://homer.ucsd.edu/homer/index.html>). Both known and de novo motifs enriched in regions in the input file were identified. The top eleven results by p-value for known motifs were variations of the ETS consensus sequence, followed by the E-box sequence (Table 3.2). The most significant result, the FLI-1 (ETS) motif was identified at the location of 48.15% of this subset of promoter RUNX1 ChIP peaks, compared to only 15.48% of assessed background regions, which were randomly selected regions matched on GC content. Interestingly ETS family members are well described as binding partners for RUNX1, and a RUNX1 complex containing FLI-1 was identified through a similar analysis of murine Runx1 (Wilson, Foster et al. 2010). De novo motif discovery characterised 17 enriched motifs not specifically associated with particular transcription factors (Table 3.3). The most significant of these was a sequence similar to the EHS ETS consensus sequence, which was identified in 48% of “consensus absent” promoter regions (compared to 14.5% of background sequences). Interestingly, sequences similar to RUNX2 and RUNX consensus binding sequences were identified in 24% and 18% of analysed regions respectively. This indicates that some of the promoters categorised as “consensus absent” may in fact be recruiting RUNX1 through non-canonical variant consensus

sequences, though this does not fully explain the presence of “consensus absent” promoter binding.

Table 3.2. Motifs enriched in RUNX1 “consensus absent” promoter ChIP peaks as determined by Homer motif analysis.

Images of the 20 most significantly enriched sequences within 200 bp of promoter RUNX1 ChIP peaks that do not contain canonical RUNX1 consensus sequences. The motif image indicates the proportion of each nucleotide at a given position by relative size. The name and type of each motif is drawn from Homer databases. The percentage of target sequences with each motif is listed, as well as the percentage of randomly selected background sequences.



































	Name	Type	% of Input Sequences with Motif	% Background Sequences with Motif
	Fli1	ETS	48.15%	15.48%
	ETV1	ETS	48.44%	16.81%
	ETS1	ETS	39.98%	12.34%
	ERG	ETS	47.37%	17.47%
	Etv2	ETS	35.05%	10.11%
	GABPA	ETS	37.18%	11.79%
	EHF	ETS	36.28%	12.39%
	EWS:FLI1-fusion	ETS	24.65%	6.58%
	Elk4	ETS	30.73%	10.17%
	Elk1	ETS	30.28%	10.10%
	EWS:ERG-fusion	ETS	21.65%	5.40%
	E-box	Basic Helix-loop-helix	10.44%	1.11%
	PU.1	ETS	18.94%	4.51%
	ELF5	ETS	24.28%	7.31%
	ELF1	ETS	26.54%	9.21%
	USF1	Basic Helix-loop-helix	18.61%	4.79%
	TFE3	Basic Helix-loop-helix	8.50%	0.83%
	Ets1-distal	ETS	12.49%	2.26%
	CLOCK	Basic Helix-loop-helix	19.15%	5.39%
	USF2	Basic Helix-loop-helix	15.16%	3.52%

Table 3.3. Homer *de novo* motifs enriched in RUNX1 “consensus absent” promoter ChIP peaks.

Images of sequences enriched in promoter RUNX1 ChIP peaks that do not contain canonical RUNX1 consensus sequences, which do not correspond to known transcription factor consensus sequences. The motif image indicates the proportion of each nucleotide at a given position by relative size. The most similar transcription factor sequence in the Homer database is identified. The percentage of target sequences with each motif is listed, as well as the percentage of randomly selected background sequences.

	Best Match	% of Input Sequences with Motif	% Background Sequences with Motif
	EHF	47.99%	14.51%
	USF1	14.50%	2.05%
	RUNX1	23.91%	6.83%
	RUNX	17.91%	5.75%
	NFY	13.15%	4.15%
	CEBPB	20.25%	10.43%
	SP1	27.28%	15.95%
	GFY	6.08%	1.58%
	AP-1	13.64%	6.92%
	SPI1	1.19%	0.06%
	CRE	8.59%	3.87%
	PB0033.1_Irf3_1	19.39%	12.66%
	NRF	7.27%	3.54%
	YY2	8.26%	4.26%
	MYB	5.18%	2.18%
	GFX	1.56%	0.31%
	Elk1	3.94%	1.69%

3.2.2.2 *Biological pathways enriched for RUNX1-bound promoters*

To determine whether the two categories of RUNX1 binding (i.e. in the presence or absence of a consensus sequence) were responsible for regulating functionally distinct genes and biological processes, the biological networks and processes these genes are involved in were investigated. The nearest gene for each promoter peak in the two categories were determined and Ingenuity Pathway Analysis (IPA) was used to perform a core analysis. Analysis was conducted using the flexible format and genes identified through human gene symbols in association with the HUGO Gene Nomenclature Committee (HGNC) and Entrez Gene guidelines, to identify the biological pathways enriched for RUNX1 bound genes.

The most significant biological pathway for both sets of genes was EIF2 signalling, which contained 28 genes with RUNX1 promoter binding in the presence of a consensus sequence and 25 without consensus sequences (Table 3.4). This pathway is involved in stress response, reducing global translation and allowing cells to conserve resources or, alternatively, inducing apoptosis (Wek, Jiang et al. 2006). Other highly enriched pathways include phagosome maturation, regulation of eIF4 and p70S6K signalling, and unfolded protein response.

Table 3.4. Biological pathways enriched for genes under both RUNX1 binding modalities.

Pathways highly enriched for genes exhibiting both RUNX1 binding modalities as determined by IPA analysis. The number of genes with RUNX1 promoter ChIP peaks in each pathway was determined from IPA.

Ingenuity Canonical Pathways	Consensus sequence present		Consensus sequence absent	
	-log(p-value	No. Genes	-log(p-value	No. Genes
EIF2 Signalling	5.37	28	5.88	25
Phagosome maturation	3.14	17	2.89	14
Regulation of eIF4 and p70S6K Signalling	2.59	18	2.51	15
Unfolded protein response	1.86	8	3.80	10

Interestingly, a third of all enriched pathways contained only one category of genes, indicating that there are biological pathways containing genes with promoters comprising RUNX1 ChIP peaks that fell within only the “consensus present” or “consensus absent” category. A proportion of these would occur by chance, and these are likely to be those smaller pathways containing only one or two RUNX1 bound genes. However, a number of pathways contained upwards of ten RUNX1 bound genes, which is perhaps indicative of likely functionally distinct RUNX1 activity within these pathways. A functional distinction in RUNX1 binding modality between different pathways raises the possibility of a broader and more flexible role for RUNX1 in regulating target genes. Different binding mechanisms could potentially respond differently to RUNX1 modulation and to its disruption in diseases such as leukaemia.

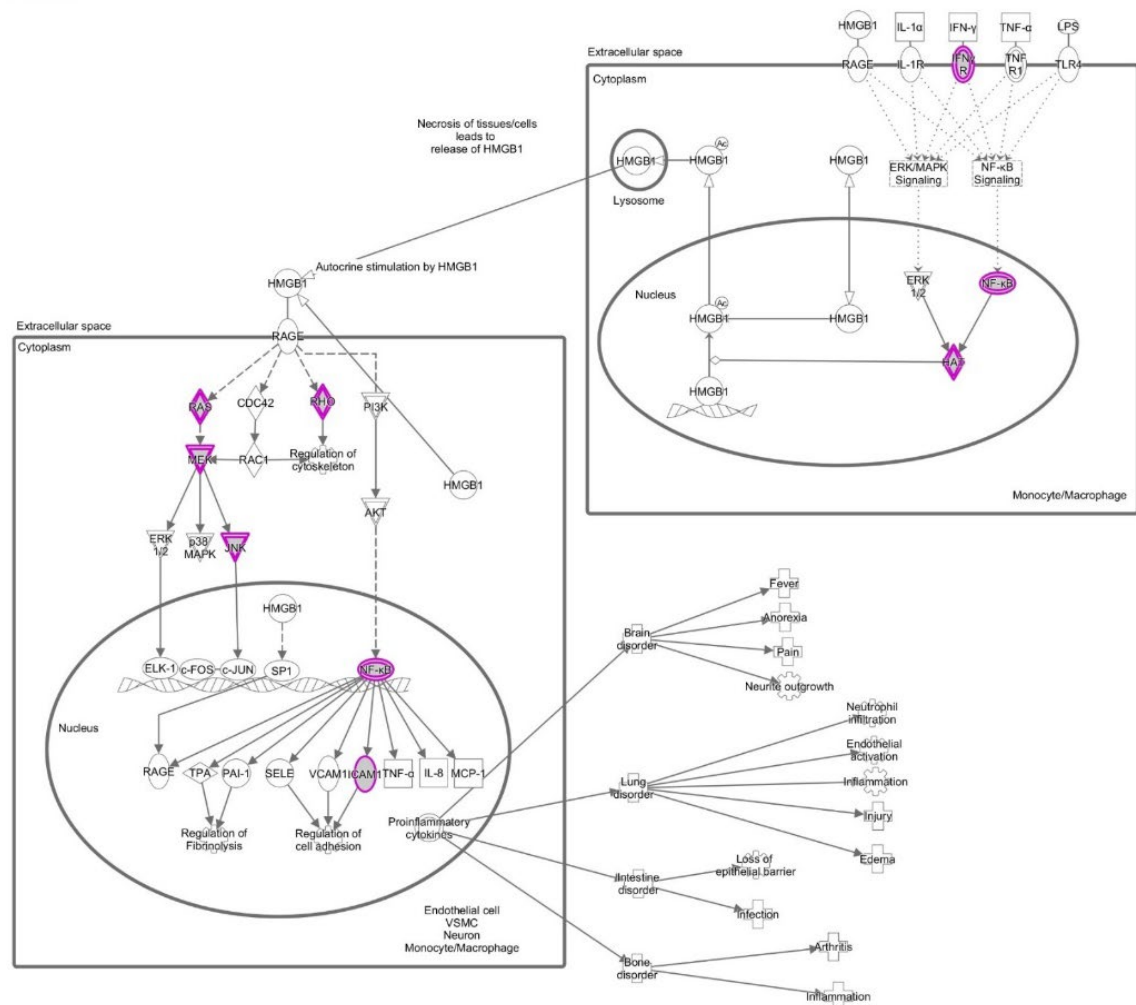
To further investigate potential functional distinctions between RUNX1 binding modalities, a filtering process was developed to rank biological pathways of interest. The IPA analysis generates a P-value for each pathway identified, indicating the enrichment of RUNX1 bound genes in the pathway above what is predicted to occur by chance. Pathways were ranked by P-value and the top 100 pathways considered for further analysis. These top 100 pathways were categorised as containing genes identified as “consensus present”, “consensus absent” or both. Pathways identified as being enriched for genes with only one binding modality were considered further (Table 3.5). Pathways containing fewer than ten RUNX1 bound genes were excluded on the basis that these are more likely to contain genes from only one binding category purely by chance, thus selecting only those which may represent functionally distinct gene regulation.

Table 3.5. Biological pathways enriched for genes with RUNX1 promoter ChIP peaks.

Pathways containing genes exhibiting only one RUNX1 binding modality from the top 100 significant pathways enriched for RUNX1-bound genes as determined by IPA analysis. The number of genes with RUNX1 promoter ChIP peaks in each pathway was determined from IPA.

Ingenuity Canonical Pathways	Consensus Sequence	-log(p-value)	No. Genes
Mechanisms of Viral Exit from Host Cells	N	4.07	9
Aryl Hydrocarbon Receptor Signaling	N	3.60	17
RhoA Signaling	Y	2.63	16
Cell Cycle Control of Chromosomal Replication	Y	2.36	6
Superpathway of Cholesterol Biosynthesis	Y	2.28	6
Methylglyoxal Degradation I	N	2.21	2
Glutathione-mediated Detoxification	N	2.01	5
N-acetylglucosamine Degradation I	Y	1.99	2
Ascorbate Recycling (Cytosolic)	N	1.92	2
HIPPO signalling	Y	1.90	11
EGF Signaling	N	1.85	7
Pentose Phosphate Pathway	Y	1.74	3
Role of JAK2 in Hormone-like Cytokine Signaling	N	1.72	5
STAT3 Pathway	N	1.72	8
2-ketoglutarate Dehydrogenase Complex	Y	1.71	2
N-acetylglucosamine Degradation II	Y	1.71	2
Serine Biosynthesis	N	1.71	2
Glycogen Degradation III	N	1.70	3
HMGB1 Signaling	N	1.64	11
Role of JAK1, JAK2 and TYK2 in Interferon Signaling	N	1.64	4
3-phosphoinositide Biosynthesis	Y	1.63	16
PDGF Signaling	N	1.60	8
Xenobiotic Metabolism Signaling	N	1.56	20
Oxidative Phosphorylation	Y	1.55	12
dTMP De Novo Biosynthesis	Y	1.51	2
Ephrin Receptor Signaling	N	1.51	14
Trans, trans-farnesyl Diphosphate Biosynthesis	Y	1.51	2
CD27 Signaling in Lymphocytes	N	1.50	6
Role of PKR in Interferon Induction and Antiviral Response	N	1.45	5
Cholesterol Biosynthesis I	Y	1.42	3
Cholesterol Biosynthesis II (via 24,25-dihydrolanosterol)	Y	1.42	3
Cholesterol Biosynthesis III (via Desmosterol)	Y	1.42	3
GDP-glucose Biosynthesis	N	1.42	2
Agrin Interactions at Neuromuscular Junction	N	1.40	7
Pentose Phosphate Pathway (Non-oxidative Branch)	Y	1.35	2
Zymosterol Biosynthesis	Y	1.35	2
Adenine and Adenosine Salvage VI	N	1.34	1

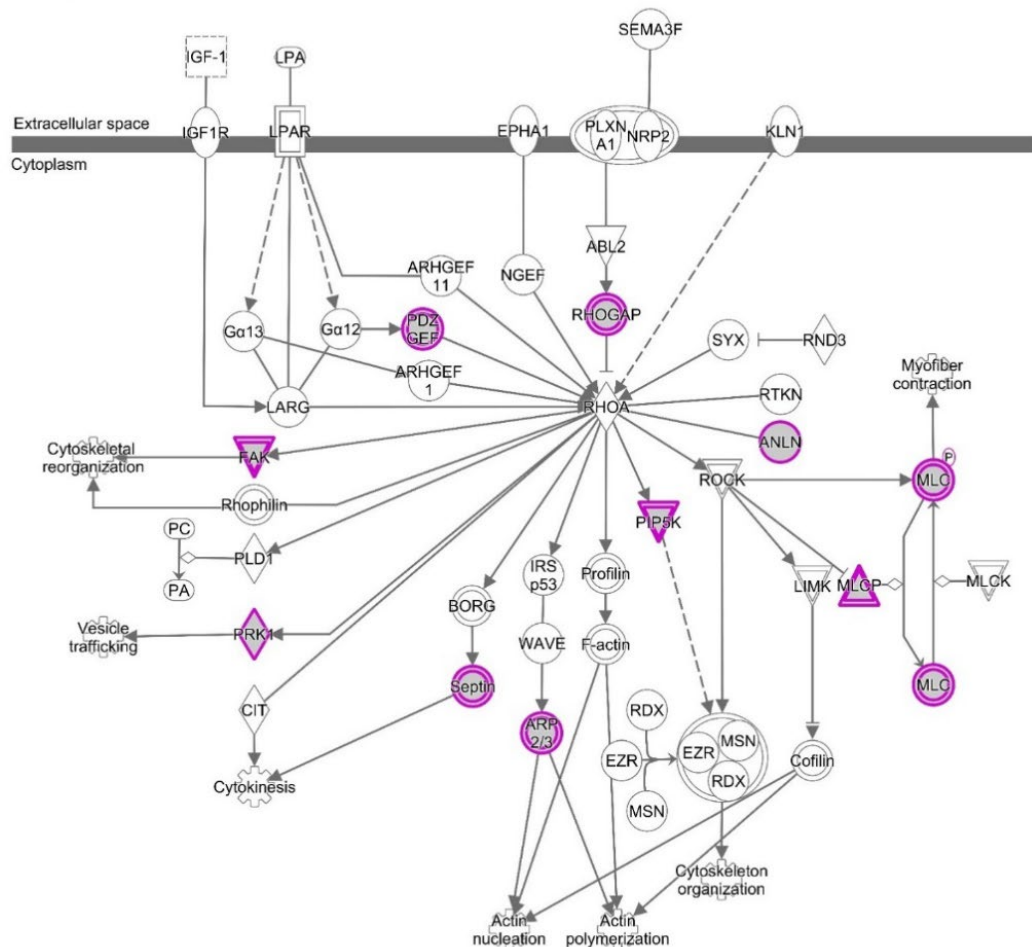
This filtering strategy enabled the identification of pathways of potential biological interest as well as groups of genes that may be instructive in characterising the different modalities of RUNX1 recruitment. For example, the Aryl hydrocarbon receptor and HMGB1 signalling pathways, (Figures 3.12, 3.13) contained genes with promoters that bound RUNX1, but in the absence of consensus sequences. In addition, some of these genes were focused to nodes within the pathway. Similarly, the RhoA and HIPPO signalling pathways (Figures 3.14, 3.15), contained genes with promoters containing RUNX1 consensus sequences.



00-2016 QIAGEN. All rights reserved.

Figure 3.13. Diagram of the HMGB1 Signalling Pathway.

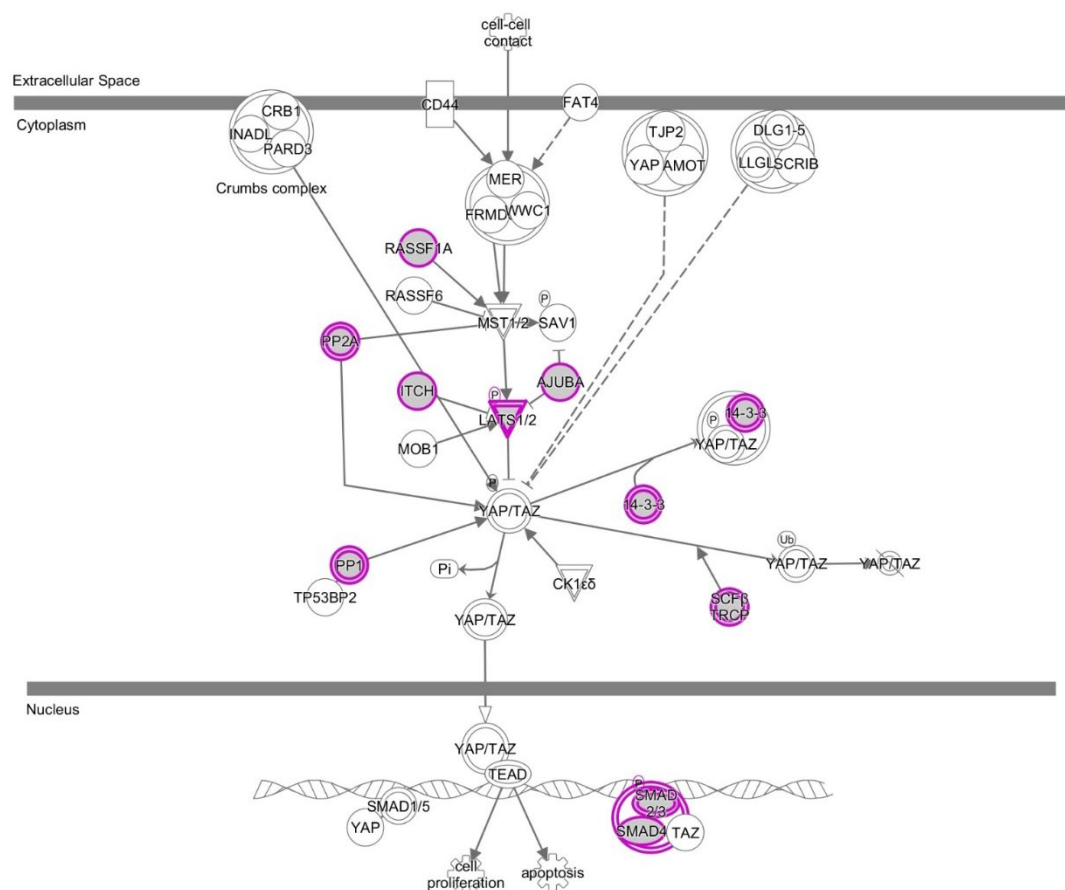
Map of protein interactions contributing to HMGB1 signalling generated by IPA. All identified RUNX1-bound promoters of genes within this pathway lacked consensus sequences. Pathway elements with RUNX1 promoter ChIP peaks are highlighted in pink. Double outlines around highlighted nodes indicate that a number of genes contribute to that portion of the pathway, at least one of which contains a promoter RUNX1 ChIP peak. Pathway element shapes represent functional classes of gene products; rectangles with solid lines for cytokines, rectangles with dotted lines for growth factors, triangles for phosphatases, diamonds for enzymes, and ovals for transcriptional regulators or modulators.



© 2000-2016 QIAGEN. All rights reserved.

Figure 3.14. Diagram of the RhoA Signalling Pathway.

Map of protein interactions contributing to RhoA signalling generated by IPA. All identified RUNX1-bound promoters of genes within this pathway contained consensus sequences. Pathway elements with RUNX1 promoter ChIP peaks are highlighted in pink. Double outlines around highlighted nodes indicate that a number of genes contribute to that portion of the pathway, at least one of which contains a promoter RUNX1 ChIP peak. Pathway element shapes represent functional classes of gene products; rectangles with solid lines for cytokines, rectangles with dotted lines for growth factors, triangles for phosphatases, diamonds for enzymes, and ovals for transcriptional regulators or modulators.



© 2000-2016 QIAGEN. All rights reserved.

Figure 3.15. Diagram of the HIPPO Signalling Pathway.

Map of protein interactions contributing to HIPPO signalling generated by IPA. All identified RUNX1-bound promoters of genes within this pathway contained consensus sequences. Pathway elements with RUNX1 promoter ChIP peaks are highlighted in pink. Double outlines around highlighted nodes indicate that a number of genes contribute to that portion of the pathway, at least one of which contains a promoter RUNX1 ChIP peak. Pathway element shapes represent functional classes of gene products; rectangles with solid lines for cytokines, rectangles with dotted lines for growth factors, triangles for phosphatases, diamonds for enzymes, and ovals for transcriptional regulators or modulators.

3.3 Discussion

The classical model of transcription factor binding describes the transcription factor being recruited by a consensus sequence at promoter elements and affecting transcriptional change directly by recruiting co-factors and transcriptional machinery (Levine and Tjian 2003). This is also a model that has been well described for the RUNX1 transcription factor, for example in regulating M-CSF and CD36 (Zhang, Fujioka et al. 1994, Armesilla, Calvo et al. 1996). However, the data outlined here suggest that regulation of genes by RUNX1 is far more complex than the simple model of consensus sequence mediated transcription factor binding. Analysis of ChIP-seq data challenges the idea that RUNX1 modulates transcription solely via the established RUNX1 binding mechanism in two main ways. Firstly, analysis of genome wide distribution of RUNX1 indicates that binding in promoter regions represents only a minority of RUNX1 localisation and that binding is more prevalent in intergenic and intronic regions. Secondly, binding in the absence of a recognised consensus sequence is relatively common, occurring in 1 in 5 RUNX1 binding events across the genome, and is even more prevalent in promoter regions (Figures 3.2, 3.3a). Taken together, this indicates that RUNX1 is binding to DNA in a manner which is occurring at loci outside of the expected genomic context and sequence specificity and suggests that additional mechanisms govern RUNX1 recruitment to DNA and transcriptional regulation by RUNX1.

Genome-wide technologies such as ChIP-seq are increasingly uncovering this kind of unexpected binding behaviour of transcription factors, as the increased scope of the analysis generates data from regions which were not priority targets for case-by-case studies as was the case for Myc (Guo, Li et al. 2014). In the past, it has been difficult to gain a holistic understanding of transcription factor binding across the genome due to the challenge of identifying and functionally characterising transcription factor binding outside promoter

regions. This resulted in an increased focus on promoter regions, as these are the best understood transcription factor bound regulatory regions and the most straightforward to locate. The focus on promoters has potentially left other genomic regions under-investigated. This is particularly true for transcription factors such as RUNX1 with well-characterised consensus sequences. These sequences are easily identifiable bioinformatically and therefore have directed the selection of targets for further study. However, the specifically targeted nature of these studies means that while they provide the opportunity to investigate regulation of individual promoters in detail, they do not necessarily provide a true holistic map of transcription factor activity and function.

Genome-wide ChIP-seq data has identified regions of DNA to which RUNX1 binds in a manner independent of recognised consensus sequences. It is possible that this finding has no functional relevance for RUNX1 and is merely an artefact of the ChIP process, the cross-linking of RUNX1 to DNA and subsequent immunoprecipitation a result of coincidental proximity and not functional binding. While possible, the similarity of results between the Trombly and Ptasinska datasets (Ptasinska, Assi et al. 2014, Trombly, Whitfield et al. 2015), and the comparability to data from mouse studies (Wilson, Foster et al. 2010, Tanaka, Joshi et al. 2012) indicates that the “*consensus sequence absent*” binding identified is not just experimental artefact. It is also possible that while the binding of RUNX1 to DNA observed in these studies is specific, RUNX1 has no function at regions where it is recruited in the absence of a consensus sequence. Given that protein-DNA binding is an energetically unfavourable reaction, it seems inefficient for this to occur without purpose and is an unlikely explanation for the phenomenon. If this is the case, non-functional RUNX1 binding should be identified relatively easily through expression studies of such RUNX1-associated genes.

Localisation of RUNX1 independently of the recognised consensus sequence suggests that there is an additional mechanism by which RUNX1 locates its targets. There are a number of possibilities for recruitment mechanisms. Firstly, it is possible that RUNX1 is acting in a completely sequence independent manner, being recruited indiscriminately through affinity for DNA. However, this seems unlikely in the case of RUNX1, given the increase in non-consensus sequence dependent binding frequency observed in promoter regions compared to intergenic or intronic regions. Indiscriminate binding would likely result in an even distribution of this type of binding throughout the genome, which is evidently not the case for RUNX1. This suggests that there is a secondary mechanism involved in targeting RUNX1 to the genome.

It is more likely that another transcription factor is directly interacting with the DNA and that RUNX1 is recruited by this protein. This has been identified as the mechanism for other transcription factors once thought to depend on consensus sequence binding, such as Myc (Guo, Li et al. 2014). A study of the Myc ChIP-seq data identified binding in the absence of the canonical E-box sequence (CACGTG) and demonstrated that the consensus sequence (or other possible 8-mers) insufficiently accounts for its genome-wide occupancy. Analysis of Myc localisation detected co-localisation with the transcriptional machinery itself and suggested that Myc recruitment in non-E-box regions was driven by interactions between the transcription factor and the transcriptional machinery rather than by recruitment to DNA. Some possible co-factors for RUNX1 have been identified in studies of transcription factor co-localisation in haematopoietic cells, including SCL, LYL1, LMO2, GATA2, ERG, and FLI-1 (Wilson, Foster et al. 2010, Diffner, Beck et al. 2013). In addition to evidence that these transcription factors commonly occupy the same genomic positions, direct protein-protein interactions have been demonstrated between RUNX1 and GATA2, SCL and ERG (Wilson, Foster et al. 2010). This mechanism is supported by evidence of frequent ETS consensus sequences present in RUNX1-

bound promoters which lack a canonical RUNX1 consensus sequence, particularly that of FLI1 which is present at almost half of these sites (Figure 3.12). Data from p300 occupancy also supports this mechanism, with p300 binding occurring closer to RUNX1 ChIP peaks in promoters without consensus sequences than those with consensus sequences (Figure 3.8). This may suggest that RUNX1 binding in the absence of a consensus sequence is associated with active transcriptional complexes, including p300, and that these complexes may be involved in the recruitment of RUNX1.

Another possible explanation for RUNX1 recruitment at promoters in the absence of a recognised consensus sequence is that these genes are being regulated by RUNX1 through distal regulatory enhancers which form long-range DNA loops to bring RUNX1 into the proximity of the target promoters. The ability of RUNX1 to act at enhancer elements containing its consensus sequence has been well described, for example in regulating the human granulocyte macrophage colony-stimulating factor (GM-CSF) gene (Bowers, Calero-Nieto et al. 2010). The presence of RUNX1 at distal regulatory elements such as enhancers may also explain the unexpected prevalence of RUNX1 in intergenic and intronic genomic regions.

Pathways enriched for genes with RUNX1 binding in their promoters provide an insight into the biological relevance of RUNX1 regulation. Of particular interest are the pathways which contain genes regulated by RUNX1 in only one of its binding modalities (“consensus present” or “consensus absent”). If these differential pathways represent real clustering of genes regulated by RUNX1 through different mechanisms, this has implications for the nature of RUNX1 regulatory networks. If there are two distinct regulatory mechanisms for RUNX1 which affect different gene networks and potentially serve different purposes, this raises several questions:

- What is the biological purpose of the two binding modalities?
- Do changes in RUNX1 expression or function affect these pathways differently?
- Are the genes in pathways where RUNX1 is recruited independently of consensus sequence subject to additional control mechanisms in the form of other transcription factors?

It is known that genes regulated through the well described consensus sequence dependent mechanism are disrupted by expression of the fusion protein RUNX1-ETO (Lam and Zhang 2012), which contains the same DNA binding and recognition domain as the wild type protein but has the transactivation domain replaced with a repressive domain. Analysis of ChIP-seq data demonstrates a similar binding distribution for RUNX1-ETO as for RUNX1, with binding occurring both largely in non-promoter regions and in contexts in which the recognised consensus sequence is absent. It is worth noting, however, that the antibody used to detect RUNX1-ETO is unable to distinguish between the RUNX1-ETO fusion protein and the wild type ETO protein. Little is known about the function of the ETO protein on its own, including its DNA binding ability, though studies indicate that it may have a role in gut development (Calabi, Pannell et al. 2001). As the protein is not highly expressed in haematopoietic cells, it is not likely to significantly confound the results of the RUNX1-ETO ChIP-seq data analysed here (Barseguian, Lutterbach et al. 2002), although this could partly explain the incomplete overlap of RUNX1 and RUNX1-ETO binding. The observed binding of RUNX1-ETO in the absence of a consensus sequence raises the question of whether genes regulated in this manner respond similarly to RUNX1-ETO as those with consensus sequences.

If individual genes are affected differently by RUNX1 disruption and these genes form parts of distinct biological pathways, it stands to reason that these pathways may be differently susceptible to dysregulation, for example during the development of cancer.

Analysis of genome-wide ChIP data suggests an important role for RUNX1 at non-promoter regions and recruitment of the protein in the absence of a recognised consensus sequence. Identification of genes that fall into these categories, and furthermore, pathways over-represented by different modalities of RUNX1 binding provides a platform to explore questions arising from this data. Several such pathways were identified, which provide examples of distinct binding modalities operating on different systems.

4 Regulation of candidate gene promoters by RUNX1

4.1 Introduction

Bioinformatic analysis of genome-wide RUNX1 localisation has highlighted its recruitment to DNA in the absence of canonical consensus sequences. While this occurs throughout the genome, it is particularly prevalent in promoter regions. ChIP-seq analysis is a useful tool for identifying transcription factor binding, however, the detection of a protein at a particular DNA locus by ChIP does not necessarily indicate that the protein has a functional effect at that locus. ChIP analysis usually relies on formaldehyde cross-linking of proteins and DNA. This method captures a snapshot of protein in the vicinity of DNA at the point of cross-linking, which may include proteins which have only a transient and/or non-functional interaction with the DNA. Normalisation against background readings and statistical analysis are employed to reduce detection of random binding, but the functional effect of the protein at a particular locus must still be experimentally confirmed.

IPA analysis described in Chapter 3 identified biological pathways enriched for RUNX1 binding at gene promoters, and further, identified pathways in which binding was exclusively associated with either the presence or the absence of a recognised consensus sequence (Table 3.4). Some of these pathways contain 'nodes' enriched for these different categories of genes which provide an ideal opportunity to explore the functional relevance of RUNX1 recruitment to DNA in the absence of its consensus sequence and whether the modality of binding influences RUNX1 function.

For example, the Rho signalling pathway which regulates cell adhesion molecules as well as cell migration and polarity contained 16 genes which were found to bind RUNX1 in their promoters

in association with a RUNX1 consensus sequence (Table 3.14). Of these, 4 ARHGAP genes appear in a functional node consisting of 9 related genes. These genes encode Ras homology GTPase-activating proteins (GAPs), which are important in regulating the ratio of GTP bound to GDP bound RhoA, affecting its role as a molecular switch. GAPs function by acting as a catalyst for Rho GTPase, accelerating the hydrolysis of GTP to GDP and inactivating Rho molecules (Haga and Ridley 2016). The human genome contains approximately 80 GAPs, but fewer than half have been extensively studied. Rho GAPs have been commonly associated with restricting cell migration and invasion, and are frequently downregulated in cancer (Kandpal 2006).

In contrast, the HMGB1 signalling pathway, which is involved in proinflammatory cytokine response, contained 11 genes which were found to bind RUNX1 in their promoters (Table 3.13), however none of these promoters contained canonical RUNX1 consensus sequences. Of these 11 genes, 4 appeared in a functional node of 7 histone modifying genes. These genes encode proteins which transfer acetyl groups to lysine residues on target proteins. Acetylated proteins are involved in a diverse range of processes, such as cell cycle control, metabolism and stress response (Drazic, Myklebust et al. 2016).

It is tempting to speculate that these two clusters of genes that demonstrate RUNX1 binding at their promoters are regulated by distinct RUNX1-dependent mechanisms, determined by different mechanism of recruitment, and this supports a hypothesis that RUNX1 can regulate gene expression in a manner that is independent of its canonical consensus sequence. The aims of this chapter were therefore to explore the regulation of these two categories of promoters by RUNX1 and whether they respond differently to the RUNX1-ETO fusion protein, and further to investigate whether there are contextual or structural features that characterise these different sites of RUNX1 binding.

4.2 Results

4.2.1 Selection of candidate genes

Biological pathways from IPA analysis were inspected to determine the distribution of RUNX1-bound genes within the pathway. A number of RUNX1-bound genes were identified within clusters of related genes at network “nodes” (Table 4.1). RUNX1 genes within these nodes were selected as candidates for functional studies. Specifically, the genes HAT1, KAT6A, KAT6B and KAT7 were selected from a cluster of histone acetyl transferase genes in the HMGB1 signalling pathway. These 4 genes fall within a ‘node’ of 7 genes within the HMGB1 signalling pathway, were all found to bind RUNX1 at their promoters as determined by ChIP-seq (Chapter 3), but in the absence of a recognised consensus sequence. Further examination of genes in this node identified a RUNX1-ETO binding peak in the promoter of the KAT2B gene, which was therefore also included as a candidate for functional analysis. In contrast, ARHGAP 1, 4, 9, and 12 are part of a 9 gene ‘node’ within the RhoA signalling pathway, all genes contain RUNX1 consensus sequences in their promoters and were shown to bind RUNX1 in ChIP-seq studies.

Table 4.1. Genes clustered within RHOGAP and HAT nodes of the RhoA and HMGB1 signalling pathways.

Genes identified by IPA analysis as having RUNX1 ChIP peaks in their promoter regions, as well as genes without peaks from the same node within the pathway. Genes with RUNX1 ChIP peaks were selected as candidate genes for further analysis. *RUNX1 binding was not identified in the KAT2B promoter in this analysis, but the promoter contains a RUNX1-ETO ChIP peak.

RhoA Signalling		HMGB1 Signalling	
RUNX1 Genes	Non-RUNX1 Genes	RUNX1 Genes	Non-RUNX1 Genes
ARHGAP1	ARHGAP5	HAT1	KAT2A
ARHGAP4	ARHGAP6	KAT6A	KAT2B*
ARHAGP9	ARHGAP8	KAT6B	RBBP7
ARHGAP12	ARHGAP35	KAT7	
	DCL1		

4.2.2 Analysis of candidate promoters which bind RUNX1 in the presence of a consensus sequence

To determine whether the selected ARHGAP family genes are expressed in haematopoietic cells and therefore whether RUNX1 is likely to contribute to their regulation in these cells, their expression in the leukaemic cell lines K562, KG1a, and Kasumi-1 was assessed. These three cell lines are derived from leukaemic cells and represent different stages of leukaemogenesis (see Chapter 2.1). mRNA was extracted from each cell line, analysed by RT-qPCR, and expression of each gene normalised to GAPDH expression. Expression of the RhoA signalling pathway genes, ARHGAP1, ARHGAP4, ARHGAP9 and ARHGAP12 was examined. All genes were detected in all cell lines (Figure 4.1), however, there was variation in the expression levels between cell lines. For each gene, expression tended to be lowest in K562, with higher expression in Kasumi-1 and KG1a cells, although for ARHGAP1 this difference was not significant for Kasumi-1 cells. Mean relative expression of ARHGAP4 in K562 cells was approximately a third of that observed in KG1a and Kasumi-1 cells, although this difference was not significant. Expression of ARHGAP 1, 4, and 12 were similar in KG1a and Kasumi-1 cells, while ARHGAP9 expression was significantly higher in Kasumi-1 cells.

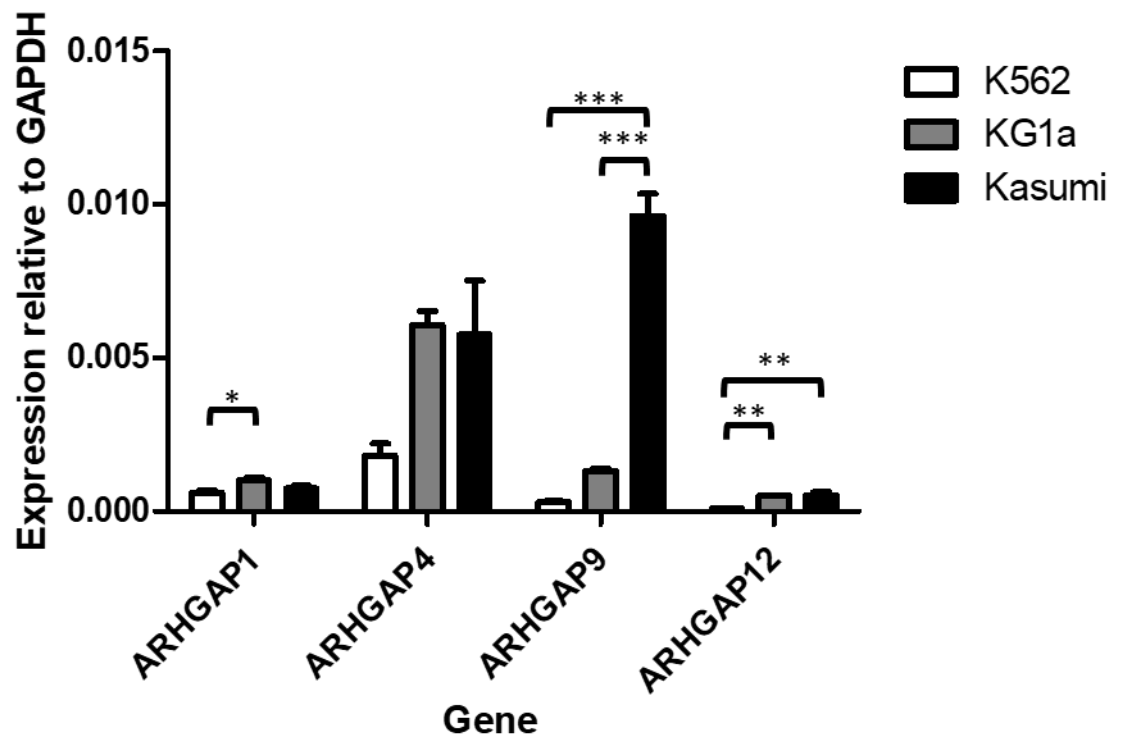


Figure 4.1. Expression of candidate ARHGAP family genes in leukaemic cell lines.

Total mRNA was isolated from K562, KG1a and Kasumi cells, reverse transcribed and expression levels of candidate genes (as noted) determined by qRT-PCR, normalised to GAPDH. Values are expressed as mean \pm SEM (n=3). Statistical significance was determined by one-way ANOVA with Tukey's multiple comparisons post-test. *p<0.05, **p<0.01, ***p<0.001.

To validate the selected candidate genes as functional RUNX1 targets, luciferase reporter constructs were generated incorporating their promoter regions. Constructs comprising a region of approximately 500-800 bp upstream of the TSS, as necessary to include the site to which the RUNX1 ChIP peak was mapped, were generated using the pXPG luciferase reporter plasmid (Figure 4.2). For ARHGAP1, ARHGAP4 and ARHGAP 12, a RUNX1 ChIP peak was present in both sets of ChIP-seq data (Ptasinska, Assi et al. 2014, Trombly, Whitfield et al. 2015). In addition, RUNX1-ETO binding was detected to these promoters (Trombly, Whitfield et al. 2015). In contrast the ARHGAP9 ChIP peak was not replicated in the Ptasinska dataset (Ptasinska, Assi et al. 2014) and on further analysis the associated consensus sequence was located downstream of the TSS and this gene was therefore not included in the promoter analysis.

Activity of the ARHGAP1, ARHGAP4, and ARHGAP12 promoter constructs was assessed following their transfection into K562 cells. Basal activity of all promoters was detected in K562 cells (Figure 4.3a). Following overexpression of RUNX1, increased activity of between 5 and 7-fold compared to CMV control transfected cells was observed for the ARHGAP4 and ARHGAP12 promoters (Figure 4.3b). In contrast, no change in ARHGAP1 promoter activity was detected upon RUNX1 overexpression.

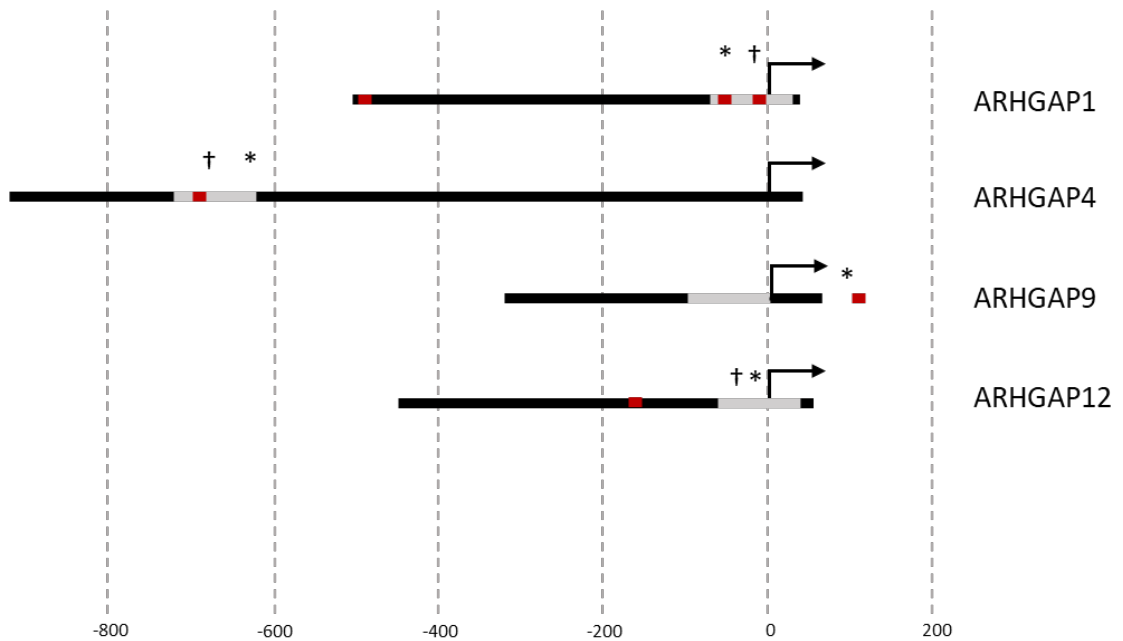
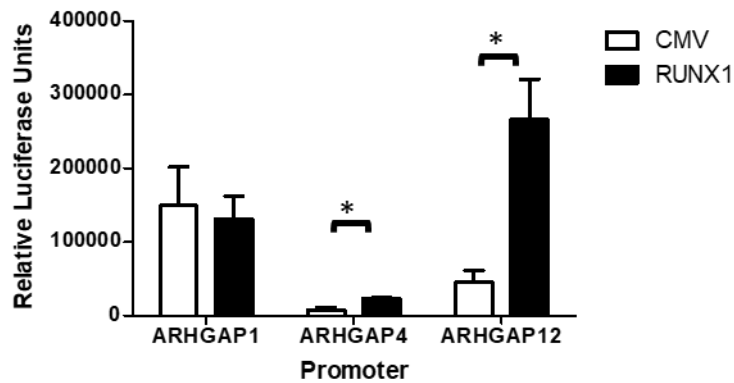


Figure 4.2. Promoter constructs generated from genes with RUNX1 ChIP peaks associated with a consensus sequence.

Schematic representation of candidate gene promoter regions analysed in reporter assays. RUNX1 ChIP peaks from Trombly, Whitfield et al. (2015) are represented by grey bars. ChIP peaks replicated in data from Ptasinska, Assi et al. (2014) are marked by †, sites of RUNX1-ETO ChIP peaks from Trombly, Whitfield et al. (2015) are marked by *. RUNX1 consensus sequences are represented by red bars (details in Table 4.2). Scale indicates distance in base pairs from the TSS, represented by arrows. Regions were amplified by PCR and ligated into the pXPG plasmid for reporter assays.

A



B

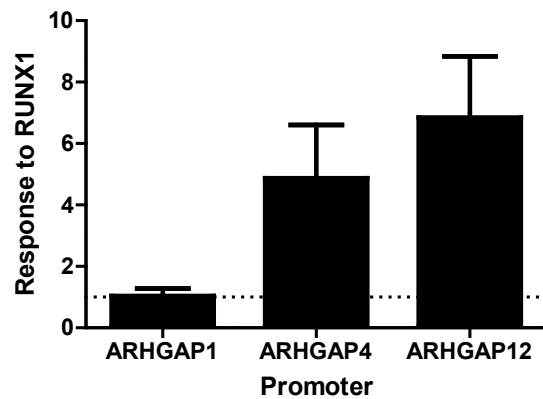


Figure 4.3. Consensus sequence present candidate promoter response to RUNX1 overexpression.

K562 cells were transfected with promoter-pXPG constructs as well as a RUNX1 expression or CMV control vector. Protein was harvested 24 hours post-transfection and luciferase activity was measured. Values are expressed as mean values \pm SEM (n=3-4) of: **A**) raw luminescence values for CMV and RUNX1 transfections, as indicated and **B**) luminescence of RUNX1 overexpression transfections relative to CMV. Statistical analysis was conducted using Student's two-tailed t-test with paired values. *p<0.05.

In conclusion, the promoter regions of ARHGAP4 and ARHGAP12 respond as expected for a classical RUNX1 target gene, with a promoter containing a recognised RUNX1 consensus sequence, but the ARHGAP1 promoter does not respond directly to RUNX1 overexpression in K562 cells. The consensus sequences identified in these promoters were therefore examined (Table 4.2) to determine whether differences in these sequences may explain their disparate responses to RUNX1. Interestingly, despite containing two ETS:RUNX consensus sequences, to which the observed ChIP peak was mapped (identical to those in the ARHGAP4 promoter) and a third more distal RUNX canonical consensus sequence (similar to that in ARGHAP12), the ARHGAP1 promoter was not responsive to RUNX1. This suggests that the RUNX1 consensus sequence alone is not sufficient to support activation of a promoter by RUNX1 in a reporter assay and suggests that other factors or sequences contribute to this activity.

Table 4.2. Consensus sequences present in candidate promoter regions.

DNA sequences containing RUNX consensus sequences within candidate promoters. Consensus sequence types obtained from HOMER (<http://homer.ucsd.edu/homer/>). Core bases are underlined.

Promoter	Consensus Sequence Type	Sequence
ARHGAP1	ETS:RUNX (2x, within peak)	<u>CCCACTTCC</u> GGC, <u>CCCACTTCCT</u> GC
	RUNX/RUNX1/RUNX2/RUNX-AML (Runt) (overlapping, distal)	CTCTGTGGT <u>C</u> AG
ARHGAP4	ETS:RUNX	<u>ACCACTTCCT</u> GC
ARHGAP12	RUNX (Runt)	CTGCGGT <u>CA</u> A

4.2.3 Analysis of candidate promoters which recruit RUNX1 in the absence of a consensus sequence

The HMGB1 pathway candidate genes, HAT1, KAT6A, KAT6B, KAT7 and KAT2B, represent a functional cluster of genes and that all had RUNX1 binding detected at their promoters in the absence of a consensus sequence. Expression of all genes was detected by RT-qPCR in all three myeloid cell lines, K562, KG1a and Kasumi-1, although the pattern of expression was different for each gene (Figure 4.4). HAT1 expression was significantly lower in KG1a cells compared to both K562 and Kasumi-1 cells. Conversely, KAT6A displayed significantly higher expression in KG1a cells than in either of the other cell lines. KAT6B exhibited low expression in K562 and KG1a cells, but significantly higher Kasumi-1 expression. Both KAT7 and KAT2B showed comparable expression across all three cell lines assayed. Notably, the expression of KAT6A and KAT6B across all cell lines was lower than that of other HAT genes by a factor of 5-10. In conclusion, all candidate genes were expressed in the haemopoietic cell lines, but their expression was variable across cell lines. While the HMGB1 genes were generally more highly expressed than the RhoA pathway genes the pattern of expression for individual genes was not consistent between cell types.

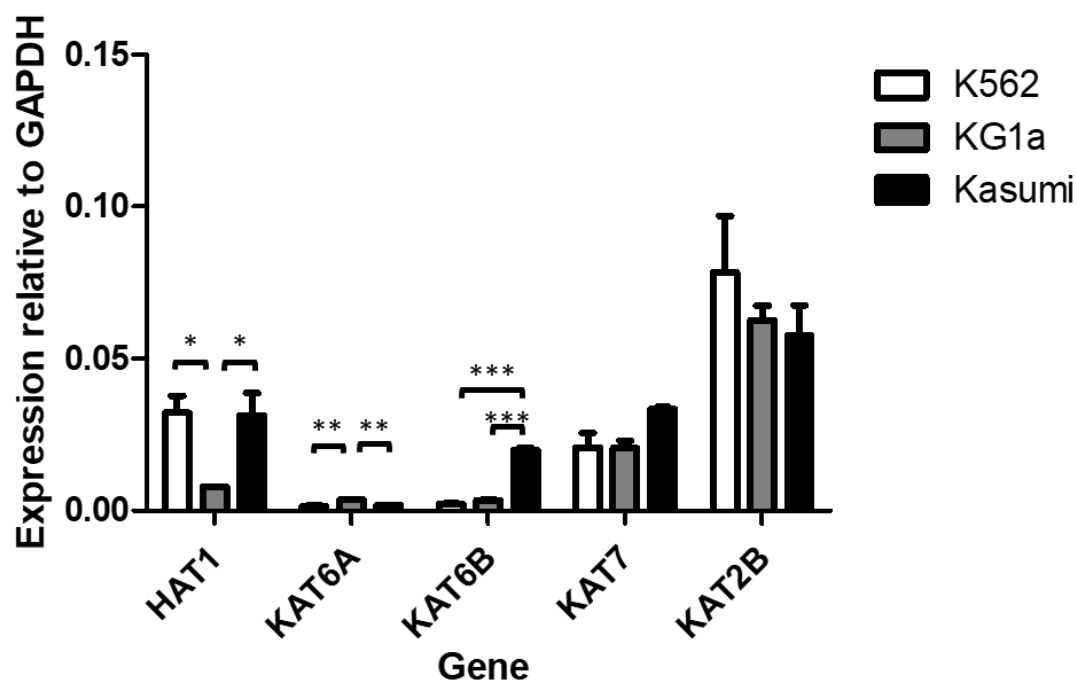


Figure 4.4. Expression of candidate HAT family genes in leukaemic cell lines.

Total mRNA was isolated from K562, KG1a and Kasumi cells, reverse transcribed and expression levels of candidate genes (as noted) determined by qRT-PCR normalised to GAPDH. Values are expressed as mean \pm SEM (n=3). Statistical significance was determined by one-way ANOVA with Tukey's multiple comparisons post-test. *p<0.05, **p<0.01, ***p<0.001.

Luciferase reporter constructs comprising a region of approximately 500-1000 bp upstream of the TSS of the HMGB1 signalling pathway gene promoters, including the region to which the RUNX1 ChIP peak was mapped, were generated using the pXPG luciferase reporter plasmid (Figure 4.5). For HAT1, KAT6A, KAT6B and KAT7 the RUNX1 ChIP peak was present in both sets of ChIP-seq data (Ptasinska, Assi et al. 2014, Trombly, Whitfield et al. 2015). The KAT2B RUNX1 ChIP peak was only detected in the Ptasinska data set (Ptasinska, Assi et al. 2014). RUNX1-ETO peaks were also mapped to the HAT1, KAT6B and KAT2B regions. The KAT7 promoter contained two RUNX1 ChIP peaks, one near the TSS and another approximately 550 bp upstream. A RUNX1-ETO peak was also detected at the KAT7 promoter, but only at the site of the distal RUNX1 ChIP peak.

Activity of the gene promoters was then assayed following transfection of the promoter constructs into K562 cells, with basal luciferase activity detected for all promoters, although the levels were relatively low for both KAT6A and KAT2B (Figure 4.6a). The response of the promoters following RUNX1 overexpression was variable (Figure 4.6b). Both the KAT6B and KAT2B promoters responded to RUNX1 with an approximately 3-fold increase in promoter activity. This demonstrates responsiveness of both these promoters to RUNX1 in the absence of a canonical consensus sequence and supports the hypothesis that RUNX1 binding at these promoters has a functional outcome. In contrast, HAT1 and KAT6A promoter constructs did not respond to RUNX1 overexpression, despite the ChIP evidence of RUNX1 localisation at these promoters. This suggests that for these promoters, RUNX1 cannot bind and/or transcriptionally regulate the promoter in a reporter context or that RUNX1 binding alone is not sufficient to regulate these genes. Distinct from both of these responses, the construct containing the region -764 to +37 bp of the KAT7 promoter showed decreased activity in

response to RUNX1, although this result was not significant and was likely impacted by the high degree of variability in the CMV control transfections for this particular construct.

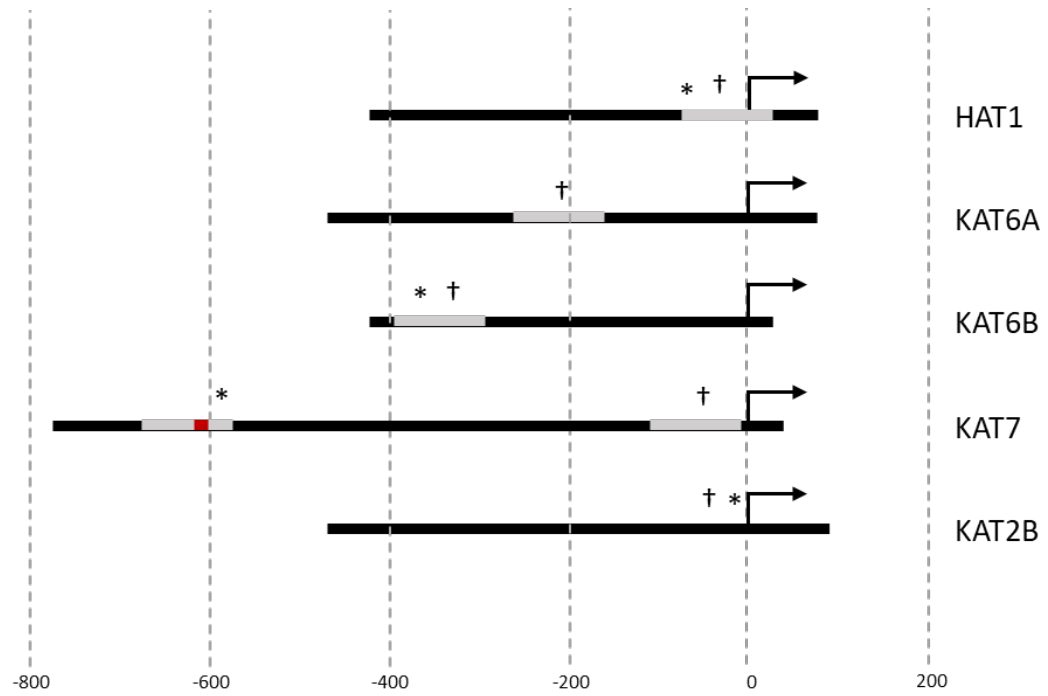
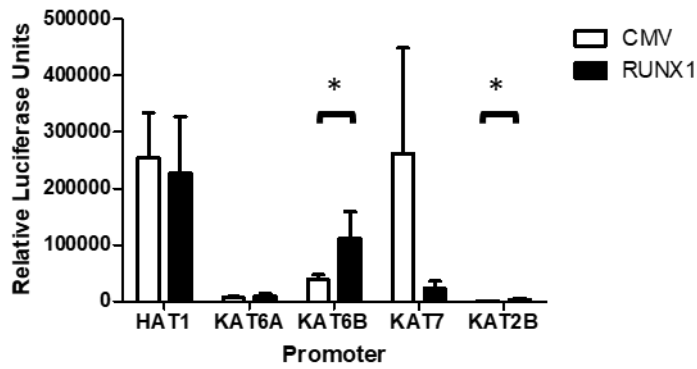


Figure 4.5. Consensus sequence absent promoter constructs.

Schematic representation of candidate gene promoter regions analysed in reporter assays. RUNX1 ChIP peaks from Trombly, Whitfield et al. (2015) are represented by grey bars. ChIP peaks replicated in data from Ptasinska, Assi et al. (2014) are marked by †, sites of RUNX1-ETO ChIP peaks from Trombly, Whitfield et al (2015) are marked by *. RUNX1 consensus sequences are represented by red bars. Scale indicates distance in base pairs from the TSS, represented by arrows. Regions were amplified by PCR and ligated into the pXPG plasmid for reporter assays.

A



B

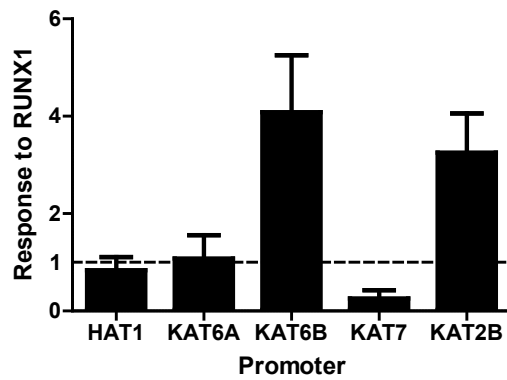


Figure 4.6. Consensus sequence absent candidate promoter response to RUNX1 overexpression.

K562 cells were transfected with promoter-pXPG constructs as well as a RUNX1 expression or CMV control vector. Protein was harvested 24 hours post-transfection and luciferase activity was measured. Values are expressed as mean values \pm SEM (n=3-4) of: **A)** raw luminescence values for CMV and RUNX1 transfections and **B)** luminescence of RUNX1 overexpression transfections relative to control. Statistical analysis was conducted using Student's two-tailed t-test with paired values. *p<0.05.

The KAT7 promoter construct encompassed two identified RUNX1 ChIP peaks, a distal site at -675 to -575 bp and a proximal site at -109 to -9 bp, with the distal site containing a RUNX1 consensus sequence. To determine the precise region responsible for transcriptional repression in response to RUNX1, analysis of a deletion construct lacking the region from -764 to -530 bp and removing the distal RUNX1 ChIP site was generated. This deletion relieved the repressive effect of RUNX1, with RUNX1 overexpression having no effect on the deletion construct, indicating that the repressive effect of RUNX1 was due to the distal site containing the consensus sequence (Figure 4.7). To further investigate RUNX1 function at this site, the distal region (-764 to -530 bp) was cloned upstream of the HAT1 promoter in the HAT1-pXPG plasmid. To test the potential of this element to act as an enhancer, it was inserted in both forward and reverse orientation adjacent to the HAT1 promoter. A characteristic of enhancers is that they are able to function in either direction due to the flexibility of chromatin loops (Zentner and Scacheri 2012). The resulting constructs were transfected into K562 cells alongside the HAT1 promoter construct alone. Results show increased expression for the KAT7-HAT1 construct in both orientations, though the forward oriented construct was more variable (Figure 4.8). While this contrasts with the original repressive function observed of this element in combination with the KAT7 promoter (Figure 4.6), it suggests that this element is acting as a transcriptional enhancer. The disparity in activity when this element is linked to its cognate promoter compared to HAT1, suggests that while this region is responsive to RUNX1, the nature of this response is variable. It is possible that inserting this small DNA fragment into an artificial construct divorces it from the broader genomic context which determines its response to RUNX1, whether it be activation or repression.

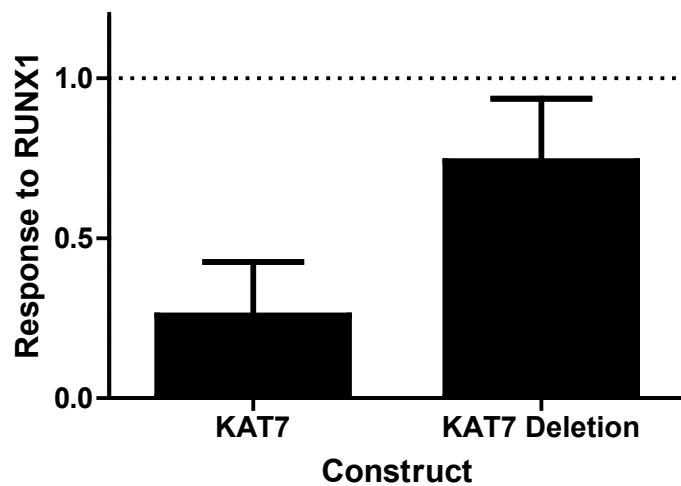


Figure 4.7. KAT7 construct response to RUNX1 overexpression.

K562 cells were transfected with KAT7 promoter-pXPG constructs with and without the distal ChIP site, as well as a RUNX1 expression or CMV control vector. Protein was harvested 24 hours post-transfection and luciferase activity was measured. Values are expressed as mean values \pm SEM (n=4) of luminescence of RUNX1 overexpression transfections relative to control. Statistical analysis was conducted using Student's two-tailed t-test with paired values. No statistically significant difference was detected.

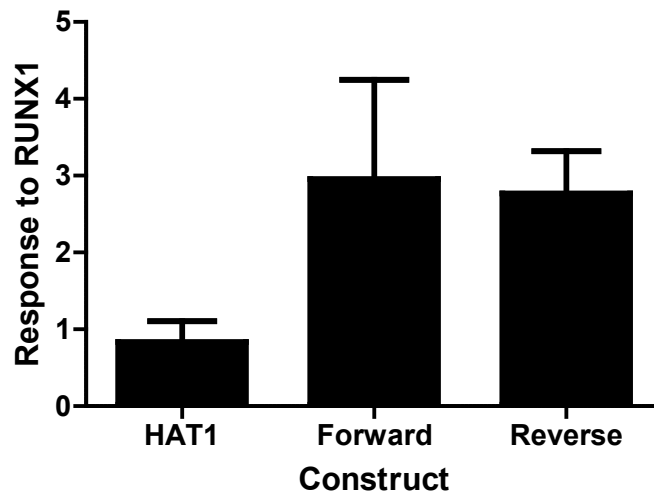


Figure 4.8. KAT7-HAT1 construct response to RUNX1 overexpression.

K562 cells were transfected with KAT7 distal site-HAT1 promoter constructs in the forward and reverse orientation, as well as a RUNX1 expression or CMV control vector. Protein was harvested 24 hours post-transfection and luciferase activity was measured. Values are expressed as mean values \pm SEM (n=4) of luminescence of RUNX1 overexpression transfections relative to control. Statistical analysis was conducted using one-way ANOVA with Tukey's multiple comparisons post-test. No statistically significant difference between transfections was detected.

These experiments have therefore demonstrated that RUNX1 can function as a transcriptional regulator at a target gene promoter in the absence of a consensus binding sequence, as demonstrated by the KAT6B and KAT2B promoters. However, in other instances while RUNX1 binding was demonstrated by ChIP, RUNX1 could not regulate these promoters in isolation, as found for HAT1 and KAT6A. The mechanisms involved in both cases warrant further investigation.

4.2.4 Response of candidate gene promoters to RUNX1-ETO

In addition to binding RUNX1, the majority of candidate gene promoters examined also bound RUNX1-ETO (Figures 4.2 and 4.5), as determined by largely overlapping RUNX1-ETO ChIP peaks with the RUNX1 peaks. RUNX1-ETO is generally repressive of genes that are activated by RUNX1 and has been reported to compete with RUNX1 for DNA binding (Lam and Zhang 2012). To investigate whether promoters that respond to RUNX1 in the absence of a consensus sequence are similarly affected by the RUNX1-ETO fusion protein, candidates which both have RUNX1-ETO ChIP peaks in their promoters and show response to RUNX1 on its own (KAT2B and KAT6B) were co-transfected with combinations of CMV control, RUNX1 and RUNX1-ETO plasmids. For comparison, the ARHGAP4 promoter which contains a consensus RUNX1 sequence and was found to be RUNX1 responsive in reporter assays was also examined. As expected, the ARHGAP4 promoter was activated by RUNX1 and repressed by RUNX1-ETO (Figure 4.9). Furthermore, activation of the ARHGAP4 promoter by RUNX1 was inhibited by co-transfection with RUNX1-ETO. Activation of KAT2B and KAT6B by RUNX1 was observed as previously, but their response to RUNX1-ETO differed from ARHGAP4. RUNX1-ETO had no effect on the KAT2B and KAT6B promoters and furthermore did not inhibit RUNX1 activation of the promoter when co-transfected with RUNX1 (Figure 4.9). This suggests that, while the ChIP-

seq data indicates that RUNX1-ETO can be recruited to these promoters similarly to RUNX1, it does not appear to be capable of transcriptionally regulating these promoters.

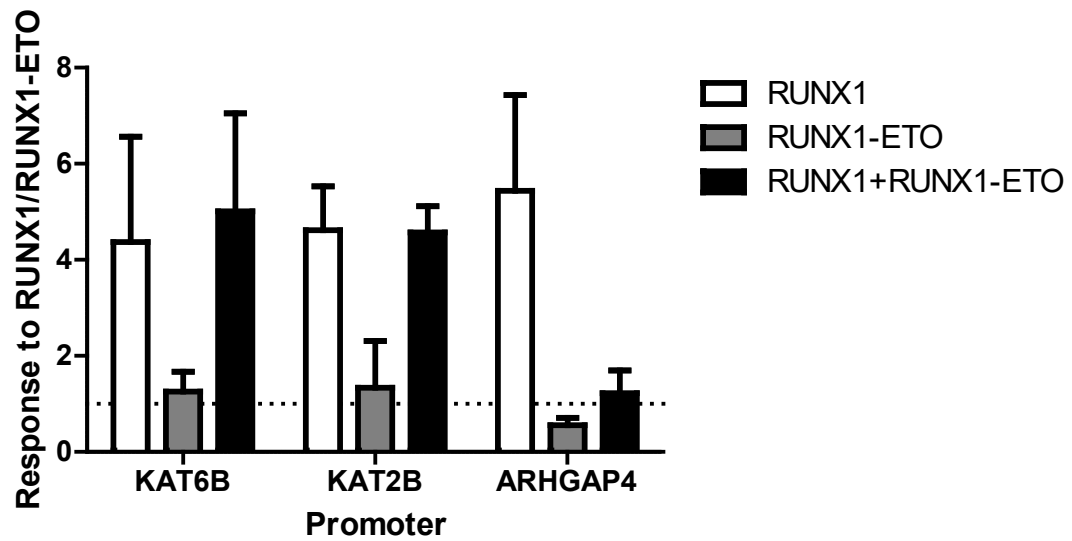


Figure 4.9. Candidate construct response to RUNX1-ETO overexpression.

K562 cells were transfected with promoter-pXPG constructs, as well as combinations of RUNX1, RUNX1-ETO and CMV control vector. Protein was harvested 24 hours post-transfection and luciferase activity was measured. Values are expressed as mean values \pm SEM (n=3) of luminescence of RUNX1 overexpression transfections relative to control. Statistical analysis was conducted using one-way ANOVA with Tukey's multiple comparisons post-test. No statistically significant difference between transfections was detected.

4.2.5 FLI-1 as a regulatory partner in target gene regulation

While RUNX1 binding was detected at the ARGHAP1 and HAT1 promoters by ChIP, these promoters failed to respond to RUNX1 in reporter assays. While the ARGHAP1 promoter contains a RUNX1 consensus sequence, the HAT1 promoter does not. In both cases it is therefore possible that RUNX1 recruitment to these promoters requires additional factors. One possible co-factor is FLI-1, which was the most significant result of HOMER known motif discovery in promoters which did not contain RUNX1 consensus sequences (Table 3.4) and has previously been shown to co-localise and interact with RUNX1 (Huang, Yu et al. 2009, Zang, Luyten et al. 2016). Both these candidate genes also have an E-box element in their promoter regions, which is the DNA sequence recognised by FLI-1 (Table 3.2, Figure 4.10a). FLI-1 expression is also low in the K562 cell line in which the transfection experiments were undertaken (Shia, Okumura et al. 2012). To investigate whether FLI-1 may be required for regulation of these promoters by RUNX1, transfections were conducted as before with combinations of RUNX1, FLI-1 and CMV control plasmids with ARGHAP1 and HAT1 promoter reporters. The results show that while both ARGHAP1 and HAT1 reporter constructs respond to FLI-1, co-operation with RUNX1 was not evident (Figure 4.10b). Like RUNX1, FLI-1 is able to act as both an activator and repressor of transcription, depending on context (Li, Luo et al. 2015), and for both these promoters, FLI-1 acted as a repressor. However, this result indicates that FLI-1 does not have a combinatorial role in regulation of these genes by RUNX1, and acts as a repressor of these promoters.

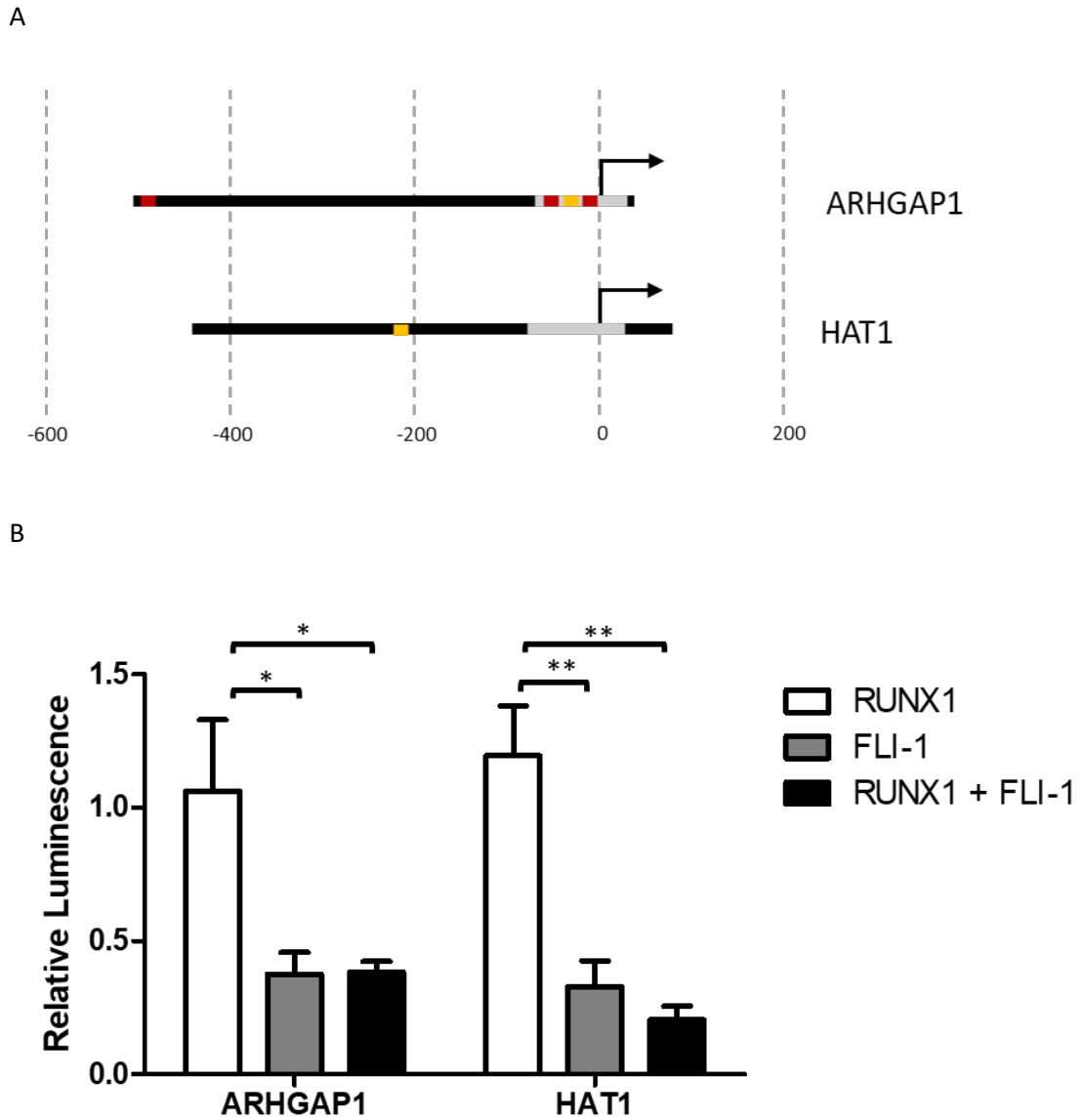


Figure 4.10. Candidate construct response to RUNX1 and FLI1 overexpression.

A) Schematic of ARHGAP1 and HAT1 promoters. RUNX1 ChIP peaks are indicated in grey, RUNX consensus sequences in red and FLI-1 sequences (CAGGAAA) in yellow. **B)** K562 cells were transfected with promoter-pXPG constructs, as well as combinations of RUNX1, FLI1 and CMV control vector. Protein was harvested 24 hours post-transfection and luciferase activity was measured. Values are expressed as mean values \pm SEM (n=3) of luminescence of RUNX1 overexpression transfections relative to control. Statistical analysis was conducted using one-way ANOVA with Tukey's multiple comparisons post-test. *p<0.05, **p<0.01.

4.2.6 Epigenetic markers at candidate promoters

To investigate whether particular epigenetic characteristics are associated with the ability of RUNX1 to regulate a promoter in the presence or absence of a consensus sequence, the epigenetic environment of the candidate gene promoters was considered. As discussed in Chapter 3, there is evidence to suggest that RUNX1 binding in the absence of consensus sequences is associated with increased CpG content at binding sites. CpG islands were identified in all candidate promoters, indicating a potential for epigenetic regulation by DNA methylation. The presence of CpG islands does not confirm DNA methylation at these sites and provides no information on potential methylation patterns or differences between RUNX1 binding modalities, which need to be determined experimentally for each cell type.

Information on the presence of histone modifications in K562 cells was sourced from ENCODE (ENCODE Project Consortium 2007) and interrogated at the sites of RUNX1 ChIP peaks in candidate promoters. Three types of modifications were considered: mono- and tri-methylation at lysine 4 on histone 3 (H3K4Me1 and H3K4Me3), and acetylation of lysine 27 on histone 3 (H3K27Ac). H3K4Me3 is generally considered a marker of transcriptionally active chromatin and is usually found near promoters. Monomethylation of the same lysine residue is often found near regulatory elements such as enhancers, while H3K27Ac marks are found in the same regions and indicate likely active regulatory elements. The combination of histone modifications at a given locus can provide information about the likely regulatory function of that region (Kimura 2013). All candidate ChIP peaks were located in the vicinity of K27 acetylation, suggestive of active promoters (Figures 4.11 and 4.12) which is supported by the expression data (Figures 4.1 and 4.4). Given that candidate genes were selected based on the presence of a RUNX1 ChIP peaks in their promoter regions, it is unsurprising that K4 trimethylation is present at the majority of candidate peaks, as this histone modification is also associated with active promoters. H3K4 monomethylation was largely absent from most candidate promoters, as expected for a marker usually associated with enhancer elements.

This modification was present at low levels at the ARHGAP4 promoter and the site of the distal KAT7 ChIP peak, suggesting that these regions may serve as enhancers. This supports the reporter assay results indicating that the distal KAT7 site is able to act as an enhancer element increasing activity of the HAT1 promoter (Figure 4.7). Importantly, the pattern of histone marks at candidate ChIP peaks is largely consistent with the expected function of these sites and does not appear to vary between promoters that contain consensus RUNX1 sequences and those that do not. This suggests that those histone modifications examined here are unlikely to influence RUNX1 binding modality.

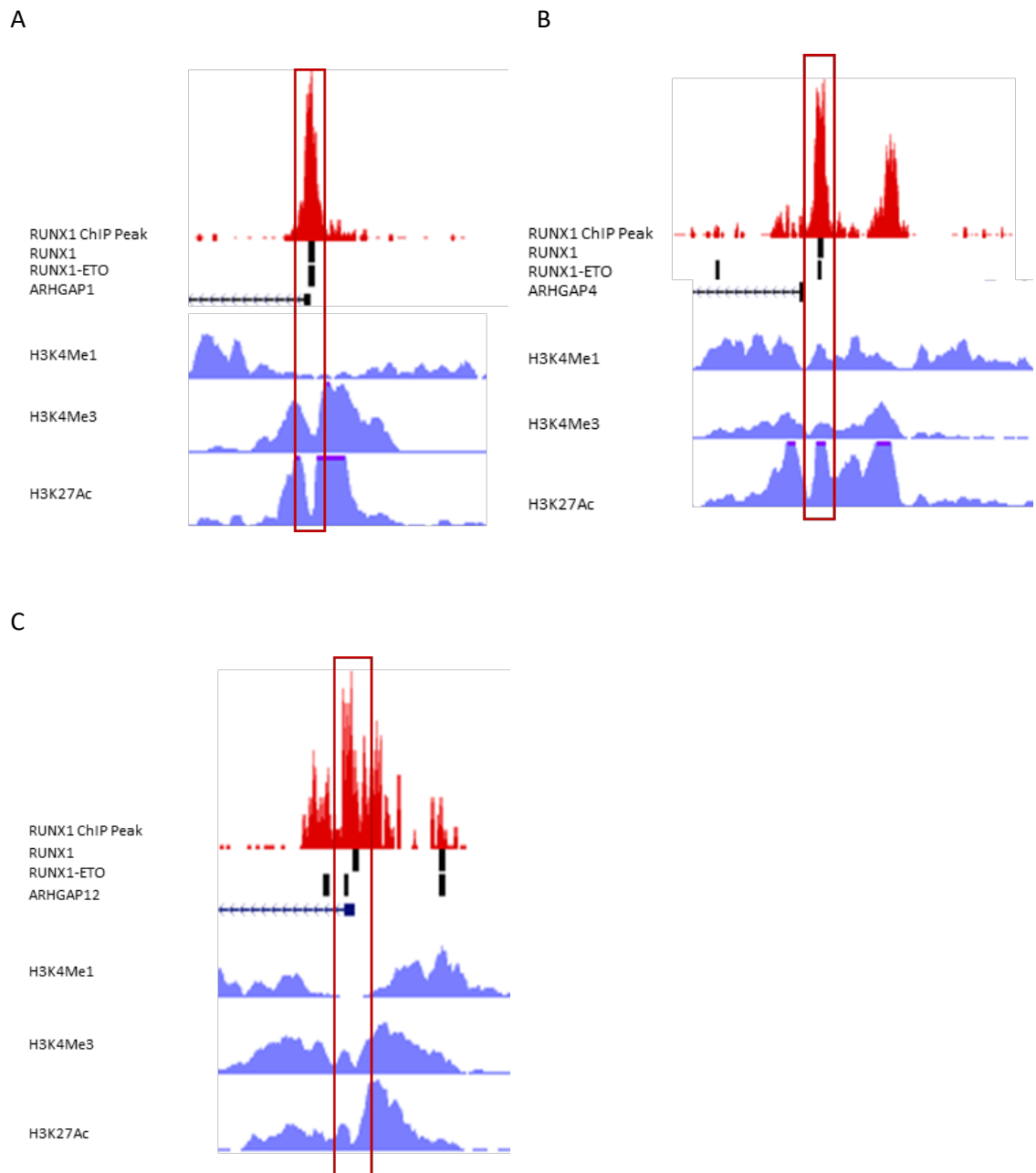


Figure 4.11. Histone modifications at consensus sequence present candidate RUNX1 ChIP sites.

A) ARHGAP1, B) ARHGAP4 and C) ARHGAP12. Screenshots from UCSC Genome browser showing RUNX1 raw (Ptasinska, Assi et al. 2014) and peak called (Trombly, Whitfield et al. 2015) ChIP peaks, TSS of candidate gene and ChIP peaks for histone modifications in K562 cells (Ram, Goren et al. 2011) (<https://www.encodeproject.org/>, GSM733656, GSM733680, GSM733692).

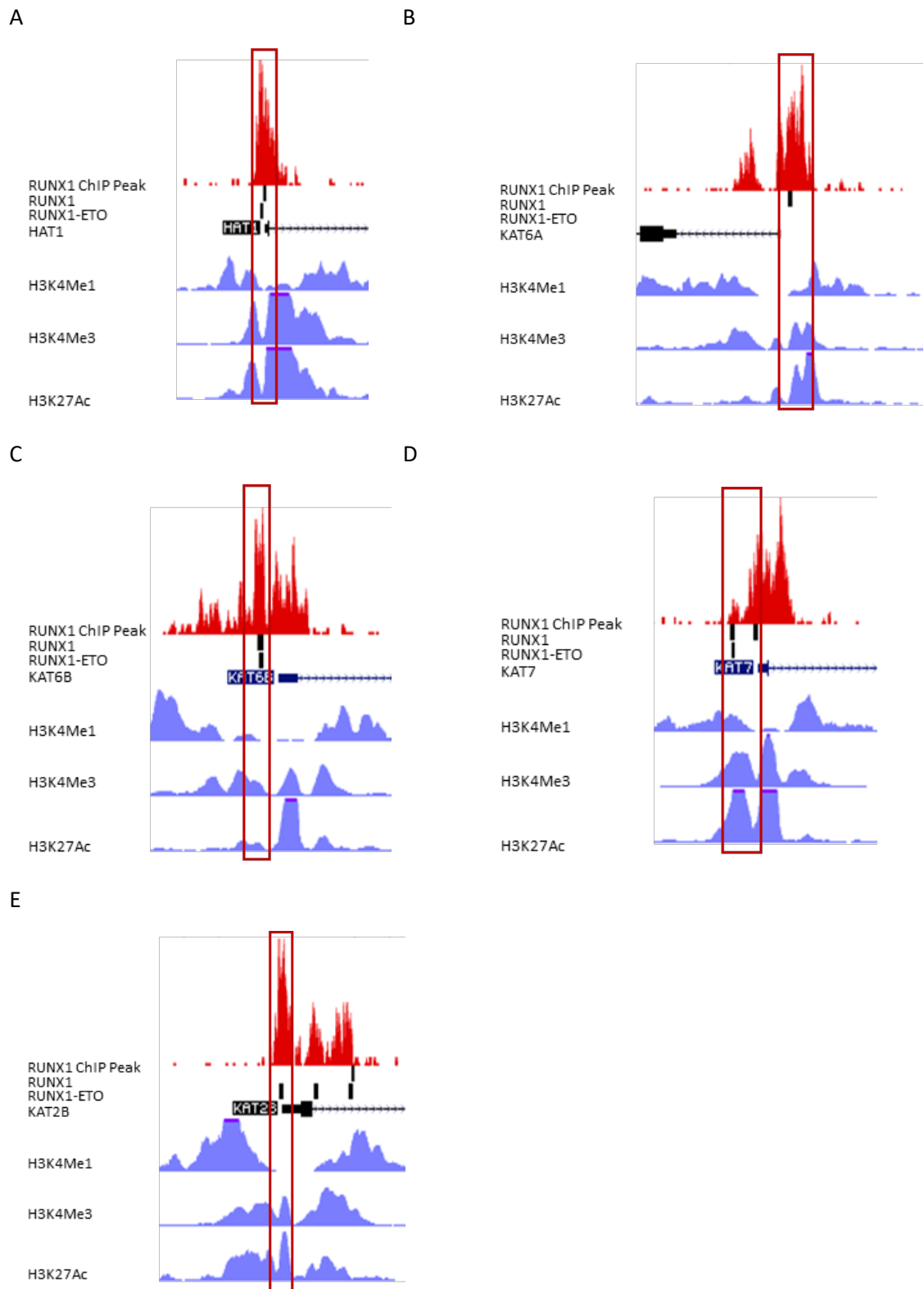


Figure 4.12. Histone modifications at candidate RUNX1 ChIP sites.

A) HAT1, B) KAT6A, C) KAT6B, D) KAT7 and E) KAT2B. Screenshots from UCSC Genome browser showing RUNX1 raw (Ptasinska, Assi et al. 2014) and peak called (Trombly, Whitfield et al. 2015) ChIP peaks, TSS of candidate gene and ChIP peaks for histone modifications in K562 cells (Ram, Goren et al. 2011) (<https://www.encodeproject.org/>, GSM733656, GSM733680, GSM733692).

4.2.7 Expression in epigenetically modified cells

To determine whether modifying the epigenetic environment influences activity of the endogenous candidate genes where RUNX1 binds in the absence of a consensus sequence, K562, KG1a and Kasumi-1 cells were treated with either 5-aza-2'-deoxycytidine (decitabine), trichostatin A (TSA) or both and the expression of HMGB1 pathway candidate genes was assessed by RT-qPCR. Decitabine is an analogue of cytidine which inhibits DNA methylation by incorporating into DNA during division and inhibiting DNA methyltransferases (Hackanson and Daskalakis 2014). TSA is an inhibitor of class I and II histone deacetylase enzymes, which alters gene expression by preventing the removal of acetyl groups from histone tails, causing the decompaction of chromatin (Wood, Rymarchyk et al. 2018). The expression of HAT1, KAT6A and KAT6B was determined in K562, KG1a and Kasumi-1 cells (Figure 4.13). These candidates were selected for analysis due to the presence of a CpG island in their promoter regions and their relatively low baseline expression in one or all of the experimental cell lines. Low expression levels may result from epigenetic repression, which could be relieved by treatment with decitabine if they are the result of DNA methylation at the promoter CpG island or by TSA treatment if the cause is compaction into heterochromatin. Expression of HAT1 did not change following decitabine or TSA treatment in either K562 or Kasumi-1 cells, which is as expected given the relatively high baseline expression in these cell lines. In KG1a cells, HAT1 expression increased after TSA treatment compared to untreated cells but decreased in both the decitabine treated and combination azacytidine and TSA treated cells. KAT6A and KAT6B expression was similarly reduced after treatment with decitabine in both KG1a and Kasumi-1 cells, though neither showed significant changes in K562 cells. KAT6A expression was unchanged by TSA treatment in all cell lines, and the combined treatment showed the same reduction in expression compared to untreated cells as the decitabine treatment alone. Treatment of KAT6B with TSA resulted in decreased expression in both KG1a and Kasumi-1 cells, with combined TSA and decitabine treatment producing a cumulative reduction in

expression. These data indicate that demethylation has a negative effect on expression of HAT1, KAT6A and KAT6B in KG1a and Kasumi-1 cells, which runs counter to expectations, as CpG island methylation is generally considered a repressive modification but may be due to indirect effects on these genes.

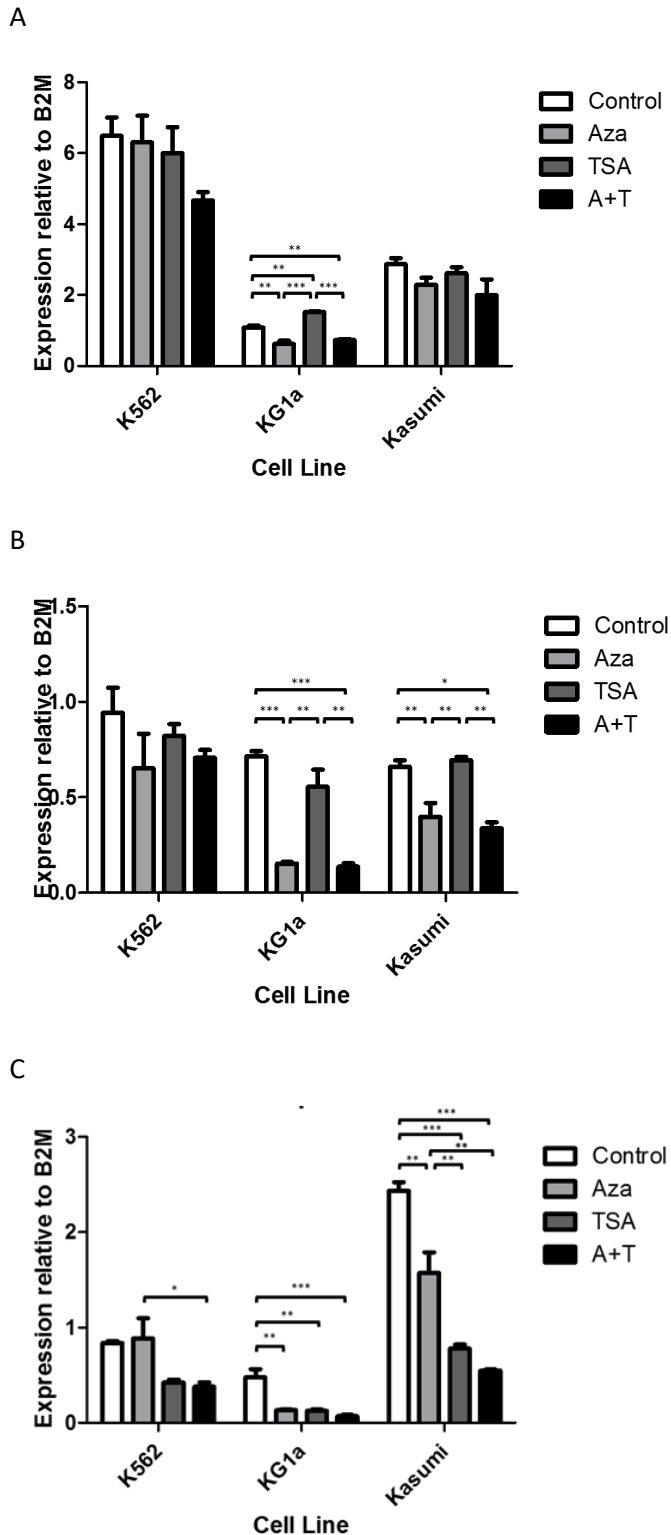


Figure 4.13. Expression of candidate genes in leukaemic cell lines treated with epigenetic modifiers.

Total mRNA was isolated from K562, KG1a and Kasumi cells treated with either decitabine, TSA, or both, reverse transcribed and expression measured using qPCR with primers for **A)** HAT1, **B)** KAT6A, and **C)** KAT6B, then normalised to β 2-microglobulin. Values are expressed as mean \pm SEM (n=3). Statistical significance determined by one-way ANOVA with Tukey's multiple comparisons post-test. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

4.3 Discussion

The data presented in this chapter reveal the complexity of regulation of genes by RUNX1. Though all candidate genes were expressed in leukaemic cell lines, their expression varied by cell type and the variation did not follow a consistent pattern across the candidate genes.

The candidate genes ARHGAP4 and ARHGAP12 demonstrate the archetypical mechanism of RUNX1 gene regulation, in which RUNX1 acts on promoter regions in the vicinity of its consensus sequence. In contrast ARHGAP1 did not respond to RUNX1 in a reporter context, despite the presence of a RUNX1 consensus sequence in its promoter.

The lack of response to RUNX1 observed for the ARHGAP1 promoter indicates that, in this case at least, the presence of a RUNX consensus sequence is not sufficient to drive transcription. It is possible that the lack of response is caused by variations in the consensus sequence itself. However, the RUNX1 ChIP peak detected at the ARHGAP1 promoter contains two copies of the extended ETS:RUNX sequence (Table 4.1), the major binding residues of which are identical to those in the ARHGAP4 promoter, which is RUNX1 responsive. This indicates that it is not the sequence itself preventing a RUNX1 response, and that something else is missing from the promoter construct. It may be that the surrounding sequence context is important for the RUNX1 response or that an additional co-factor is required in this instance. The ARHGAP1 promoter construct also contained a canonical RUNX consensus sequence at the distal end of the cloned fragment, which is identical to the sequence found in the ARHGAP12 promoter. This element was also not sufficient to support a RUNX1 response in a reporter assay and no RUNX1 binding was detected at this site by ChIP, suggesting again that the broader context of the sequence is important both in a reporter assay but also in the endogenous environment, and in that case additional chromatin elements may influence RUNX1 recruitment.

The candidate gene promoters KAT6B and KAT2B, which do not contain consensus sequences, respond directly to RUNX1 overexpression in reporter assays. This supports the hypothesis of this study that RUNX1 can regulate its target genes through a mechanism which does not require a consensus sequence. The question, then, is exactly what the mechanism is for this non-consensus regulation. It is possible that RUNX1 is being recruited to lower affinity DNA sites directly, or to a novel consensus sequence, though this does not explain the greater propensity of such binding observed in promoter regions compared to the rest of the genome. A more likely explanation is that the response is mediated by another factor such, as is postulated for Myc (Guo, Li et al. 2014).

The activity of the KAT7 promoter is repressed by RUNX1 overexpression. The promoter region contains two RUNX1 ChIP peaks, one of which is associated with a consensus sequence. Deletion of the consensus sequence containing site largely relieved the repression, suggesting that this site is responsible for the observed repression and that the KAT7 promoter is responsive to RUNX1 in a consensus sequence-dependent manner.

Despite evidence that promoters can respond directly to RUNX1 in the absence of a consensus sequence, not all such promoters behave the same. Candidate promoters HAT1 and KAT6A did not respond to RUNX1 overexpression. This lack of response may occur for the same reasons discussed above for ARHGAP1, or it may be that the RUNX1 ChIP peak observed at these loci is the result of binding elsewhere that has been brought into proximity of the promoter at the time of chromatin fixation.

FLI-1 is a good candidate for the 'missing factor' for candidate promoters not responding to RUNX1, as FLI-1 consensus sequences are enriched in regions of non-consensus RUNX1 ChIP binding. Additionally, direct interactions between RUNX1 and FLI-1 have been demonstrated in a haematopoietic model previously (Wilson, Foster et al. 2010). It is possible that the FLI-1 consensus sequence is responsible for directing FLI-1 to its targets where it subsequently recruits RUNX1 to affect transcription. However, this hypothesis was not supported by the evidence from co-transfection assays of RUNX1 and FLI-1. The HAT1 and ARHGAP1 promoters were both repressed by overexpression of FLI-1, and this was unaffected by the presence of RUNX1, indicating that there is no combinatorial effect of RUNX1 and FLI-1 regulation on these promoters. In order to identify potential co-regulators, co-transfections could be conducted using a high-throughput method to test a wider range of transcription factors. Alternatively, a "pull-down assay" could be conducted using the promoter region as bait and paired with mass spectrometry to identify other factors bound at these promoters to narrow down the focus for functional studies.

RUNX1 disruption is a common occurrence in the development of haematological malignancies such as leukaemia. Therefore, the question of whether genes in which RUNX1 binds in the absence of a consensus sequence respond in the same manner as those with consensus sequences to this disruption was investigated. The translocation resulting in the fusion protein RUNX1-ETO is one of the most common and well-studied disruptions and therefore a suitable starting point to investigate non-consensus binding in disrupted systems. RUNX1-ETO contains the DNA binding domain but lacks the transactivation domain of wild-type RUNX1 and is therefore localises to RUNX1 targets by binding to the RUNX1 consensus sequence (Lam and Zhang 2012). Additionally, RUNX1-ETO recruits co-repressive factors such as N-CoR and therefore represses genes which are normally activated by RUNX1 (Lin, Mulloy et al. 2017).

RUNX1 and RUNX1-ETO compete for binding sites in heterozygotes, resulting in impaired activation by RUNX1 combined with repression by RUNX1-ETO (Lam and Zhang 2012). Combinatorial transfection assays demonstrated this dominant negative effect of RUNX1-ETO on RUNX1 at the ARHGAP4 promoter, which is behaving as an archetypical RUNX1 target. The promoter was activated by RUNX1 and repressed by RUNX1-ETO, with RUNX1-ETO repressing activation of the ARHGAP4 promoter by RUNX1. In contrast, the candidate promoters that did not contain RUNX1 consensus sequences, KAT6B and KAT2B, did not respond to RUNX1-ETO. Further, RUNX1-ETO had no effect on activation of the promoters by RUNX1 overexpression. This is interesting, as it demonstrates a difference in the ability of RUNX1-ETO to disrupt RUNX1 function depending on the mechanism by which RUNX1 is regulating its target genes. RUNX1-ETO has been demonstrated to bind to each of these promoters in ChIP-seq experiments, so the difference in response appears to be one of regulation rather than recruitment (Trombly, Whitfield et al. 2015). As the DBD is identical between RUNX1 and RUNX1-ETO, it is clear that this difference occurs in the replaced transactivation domain. It is possible that this domain is somehow required for RUNX1 to regulate targets in the absence of a consensus sequence, potentially through interactions with other co-factors.

The histone context of candidate genes in K562 cells is consistent with their roles as active promoters, which is supported by the presence of p300 at all promoter sites. The reduction in expression of candidate HATs after treatment with decitabine is unusual, in that removal of DNA methylation is generally considered an activating modification (Estey 2013). It is possible that due to the role these genes play in regulating epigenetic processes, specifically histone acetylation, they are subject to the interplay of multiple epigenetic mechanisms rather than just methylation. Results from Chapter 3 indicate that RUNX1 binding in the absence of consensus sequences is associated with higher levels of both G and C nucleotides and CpG

dinucleotides. However, the two categories of candidate promoters were similar in terms of both GC% and CpG%. GC% ranged from 44-72% for “consensus absent” promoters and 57-72% for “consensus present” promoters. Consensus sequence absent CpG% ranged from 1.1-8.8%, while consensus sequence present promoters ranged from 1.2-12.3%. In both cases, no candidate was at either extreme of the range described in the genome-wide data (Figures 3.6 and 3.7).

Biological pathways have been identified that could be key targets of RUNX1, and which may be regulated by RUNX1 through mechanistically distinct means. Elements of these pathways with RUNX1 binding in their promoters are biologically relevant in the development of AML. Repression of KAT7 in AML patients reduces acetylation of histone 4 at lysine 5 and is associated with poorer patient outcomes (Sauer, Arteaga et al. 2015). Both KAT6A and KAT6B have been identified as targets of chromosomal fusions and rearrangements in leukaemia (Liang, Prouty et al. 1998, Chaffanet, Gressin et al. 2000, Panagopoulos, Fioretos et al. 2001). Disruption of RUNX1 therefore has the potential to result in not only dysregulation of its direct targets, but also disruption of a wider body of genes, changing context dependent expression.

5 Potential role of enhancers in RUNX1 gene regulation

5.1 Introduction

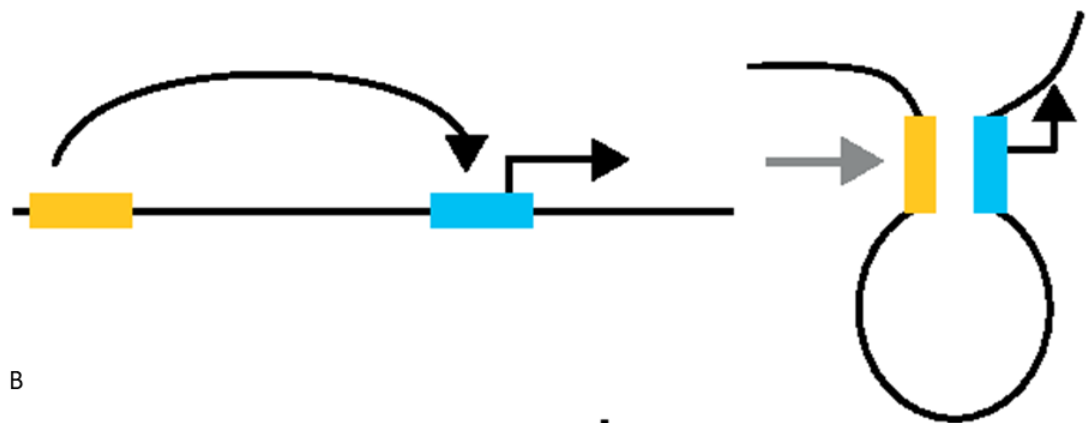
While much of the functional analysis of RUNX1 has focused on its association with promoters, the majority of RUNX1 binding, as determined through ChIP-seq analysis and described in Chapter 3, occurs in either intergenic or intronic regions. The functional outcome of RUNX1 binding in these regions is likely more complex than at promoter regions, but it is an obvious possibility that the transcription factor is binding to these regions because they serve as transcriptional enhancers. Enhancers are regulatory regions which facilitate transcription of their associated genes (Schoenfelder and Fraser 2019), and while the mechanisms through which they regulate gene expression are relatively well understood identifying and demonstrating the functionality of enhancer regions is a challenging endeavour, for several reasons. They can be located a great distance from the transcription start sites of the genes that they regulate and can be either upstream or downstream of the target gene. Furthermore, enhancers do not necessarily regulate the next closest promoter, a promoter can be regulated by multiple enhancers and a single enhancer can regulate multiple promoters (Mohrs, Blankespoor et al. 2001). No general sequence characteristic has been described that distinguishes enhancers, which limits the usefulness of bioinformatic tools in their identification and analysis. Additionally, the dynamic nature of enhancers means that interactions can be transient, and also cell type specific.

To affect transcription at distant promoters, enhancers need to be brought into close proximity to their targets. For regions where the chromatin between the enhancer and promoter is highly compacted, the distance between regulatory regions can be small enough for diffusion of factors from enhancer to promoter to affect transcription. Diffusion directed by proteins scanning the intervening chromatin between enhancer and promoter has been

described at both the *lac* repressor in *E. coli*. and for eukaryotic genes, such as binding of C/EBP, HNF-1, and HNF-3 at the upstream enhancer of HNF-4 α (Hatzis and Talianidis 2002, Hammar, Leroy et al. 2012).

Diffusion is only a plausible mechanism for enhancer action when the distances involved are very small (200-400 nm) (Pennacchio, Bickmore et al. 2013). For enhancers located a greater distance from their targets, proximity is achieved through the formation of chromatin loops which serve to deliver proteins bound to DNA at the enhancer to the target promoter (Figure 5.1a). Enhancer loops are generally confined within regions of interacting DNA bounded by insulator proteins called topologically associated domains (TADs, Figure 5.1b). TAD boundaries are conserved between different cell types, though the enhancer-promoter interactions within a TAD may vary greatly (Dixon, Selvaraj et al. 2012). However, as with many mechanisms of epigenetic regulation, TADs can be disrupted in cancer. The formation of new, smaller TADs interrupts normal enhancer-promoter interactions and leads to further dysregulation of genes (Achinger-Kawecka, Taberlay et al. 2016). This form of transcriptional disruption is particular to gene regulation by enhancers, as transcription factors in gene promoters are already in close proximity to their target TSS and are therefore not directly affected.

A



B

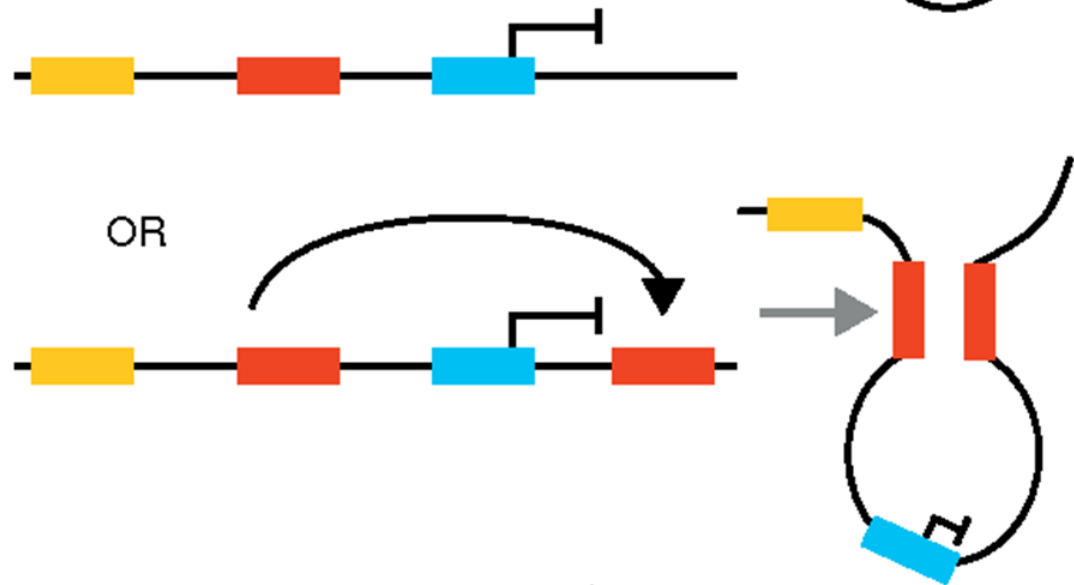


Figure 5.1. Formation of enhancer-promoter loops.

A) Chromatin loop formation brings enhancer elements into contact with their associated promoters to drive transcription. **B)** The presence of insulator elements interferes with this interaction either by interposing between the enhancer and promoter elements or by forming loops which exclude the enhancer. Adapted from Krivega and Dean (2012).

A variety of techniques have been developed to investigate the linkage between enhancers and their associated promoters (Han, Zhang et al. 2018). These techniques commonly involve identifying interacting regions by cross-linking chromatin and analysing regions of DNA which have been linked together. One such method is chromatin interaction analysis with paired-end tag (ChIA-PET) sequencing. Similar to ChIP, ChIA-PET involves cross-linking DNA and proteins, followed by chromatin fragmentation and immunoprecipitation of specific proteins, commonly transcription factors. Interacting DNA is tagged, ligated and sequenced to detect separate fragments bound to the same protein molecule. This is an effective method for identifying interacting DNA loci, with the caveat that detection relies on the presence of the chosen protein at the site of interaction (Fullwood, Liu et al. 2009). Other methods which do not have this limitation include chromatin conformation capture (3C) and its derivatives, 4C, 5C, Hi-C, and others, which directly investigate cross-linked chromatin (Han, Zhang et al. 2018).

RUNX1, as a transcription factor, has the capacity to regulate gene transcription through localisation to enhancer DNA, and this has been previously described in a number of single gene studies. In particular, it has been shown to bind to overlapping RUNX1 consensus sequences as a dimer to act on an enhancer of the GM-CSF gene (Bowers, Calero-Nieto et al. 2010) and to co-localise with transcription factor Ets1 to regulate the TRC α enhancer (Kasahara, Shiina et al. 2017). Genome-wide ChIP-seq analysis has shown that the majority of RUNX1 binding occurs outside of promoter regions (Chapter 3), leading to the hypothesis that this represents RUNX1 binding at enhancers which form long range DNA loops and functional interactions with the promoter of target genes. This scenario could also explain the lack of response of some candidate gene promoters to RUNX1 in reporter assays, despite demonstration of RUNX1 recruitment in ChIP-seq analysis (Chapter 4). While it is possible that such gene promoters may require additional factors not provided within the context of a

reporter assay, an alternative possibility is that the RUNX1 binding observed at the promoter in ChIP-seq analysis may be the result of direct binding at a distal regulatory region which has been cross-linked to the promoter due to proximity during the assay. Therefore, the potential for DNA loop formation between regions of RUNX1-bound DNA was investigated here. Interactions between intergenic/intronic RUNX1 ChIP peaks and promoter peaks were mapped through ChIA-PET linkage.

5.2 Results

5.2.1 Identification of RUNX1 associated enhancer-promoter pairs

To address this, several different approaches were used to identify potential RUNX1-bound enhancers and their likely target promoter. Initially, a “top down” approach was used, starting with the entirety of intergenic and intronic RUNX1 binding observed through ChIP-seq to identify those likely to represent potential enhancers. Firstly, the genes with intergenic or intronic ChIP peaks annotated to them were considered. While all RUNX1 ChIP peaks were annotated to the nearest gene (Chapter 3), this does not necessarily indicate that this is the gene that is being regulated by RUNX1 in each instance. In fact it is estimated that only 7% of enhancers control the nearest gene promoter (Snetkova and Skok 2018). However, this was used as a starting point for interrogating the data for potential enhancers. Therefore, the ChIP-Seq data was searched for genes with at least two separate RUNX1 peaks annotated to them: one in a promoter region and one in intergenic or intronic regions. This provided a subset of genes that might be regulated by RUNX1 associated promoter-enhancer links for further examination.

Annotation of intergenic/intronic and promoter ChIP peaks to the nearest gene produced 5515 and 2461 unique genes, respectively. Of these 635 genes were common between the two sets, having RUNX1 ChIP peaks annotated to both their promoter and nearby intergenic or intronic regions (Table 5.1). Genes identified in this manner were then examined in the UCSC genome browser for interactions through interrogation of publicly available RNA Pol II ChIA-PET data for K562 cells, the rationale being that this ChIP-PET data set will contain regions of DNA that interact with RNA PolII bound promoters, at least in this cell line. Promoters containing RUNX1 ChIP peaks were thus examined for evidence of interaction with distal regions also containing RUNX1 peaks using RNA Pol II ChIA-PET data (Fullwood, Han et al. 2010). In some cases, this identified an interaction with an intergenic or intronic peak originally annotated to the same gene. In others, the interaction identified was to another region containing a more distal ChIP peak. In both cases, the linked RUNX1 bound loci were further evaluated to determine their likelihood of functioning as enhancers. A particular combination of histone modifications, specifically the presence of monomethylation of histone 3 lysine 4 (H3K4Me1) and acetylation of lysine 27 (H3K27Ac), coupled with the absence of trimethylation at lysine 4 (H3K4Me3) is considered characteristic of an active enhancer (Bulger and Groudine 2011). Using pooled data from the ENCODE project from various cell types, including K562 (ENCODE Project Consortium 2007), linked distal regions were evaluated for the presence of these histone modifications and genes with potential RUNX1-bound enhancers were thus identified (Table 5.1). Of the 635 genes initially annotated to both promoter and intergenic/intronic regions, 96 were linked through ChIA-PET data and a further 37 of these had linked regions which displayed chromatin features characteristic of enhancers. Notably, the candidate genes ARHGAP1, HAT1 and KAT6B examined in Chapter 4 were identified in this data set as being linked to distal RUNX1-bound regions.

Table 5.1. Genes with promoter and intergenic/intronic RUNX1 ChIP peaks.

List of genes with RUNX1 ChIP peaks in their promoters and nearby intergenic or intronic regions. Genes with links through ChIA-PET data from their promoter peak to another ChIP peak are in bold, genes where the linked peak has the histone marks associated with an enhancer (H3K4Me1, H3K27Ac, lack of H3K4Me3) are in italics.

AAK1	ATG13	CAMK1D	COLGALT2	DYNLRB2	FKBP15	HPS3	LATS1	MAPK14	NAA35	PACSLN2	PSMA1	RP56KA1	SLC39A13	THOC1	UTP11L
ABHD2	ATG7	CAMK2D	COMMD1	DYRK1A	FMNL1	<i>HUS1</i>	LBR	MAPKAPK3	NAA40	PAFAH2	PSMC5	RP59	SLC39A8	THUMP3	VAC14
ABI1	ATP2B4	CAP1	COMMD8	E2F3	FNDC3A	ICAM2	LCP2	MARCH8	NABP1	PARK7	PTGS2	RRAGB	SLCO4C1	TIPARP	VASP
ACAA1	ATP2C1	CAPNS1	COMMD9	EEF1A1	FNDC3B	ICT1	LDLRAD4	MAT2B	NARF	PARVG	PTK2B	RRAGC	SMAD2	TLE4	VIM-AS1
ACAD9	ATP6V1C1	CASC11	COPB1	EGR2	FNIP1	IDE	LEMD2	MBIP	NAT10	PCDH7	PTPN4	RSL24D1	SMPD3	TM9SF2	VP518
										PCED1B-AS1	QTRT1	RSU1	SNHG8	TMED7-TICAM2	VRK1
ACTG1	ATP9B	CASC20	CORO1C	EHD1	FOXN2	IER5	LGALS12	MCC	NCF4	PCF11	RAB29	RTFDC1	SOC5	TMEM131	WAS
ACTR1A	AVL9	CASP10	CREB3	EHD3	FTO	IFNGR2	LIMD2	MDN1	NCR3LG1					TMEM161B	
ADARB2-AS1	B3GALNT2	CASP8	CRY1	EID2	FZD1	IGF2BP3	LINC00339	ME3	NDFIP2-AS1	PDCD6IP	RAB7A	RTN4	SPAG9	TMEM168	WASL
ADCY7	B3GNT2	CAT	CSF1R	EIF4A2	GADD45B	IGFBP7	LINC00856	MEF2C	NDUFA12	PDE4A	RAC2	RUNX1	SPN	TMEM168	WBP2
AGAP2	BANK1	CBFA2T3	CSGALNACT2	EIF4G3	GBGT1	IL1OR8-AS1	LINC00861	MEGF9	NDUFA4	PFKFB3	RAD23B	SAE1	SPRY2	TMEM173	WDR36
AGTRAP	BATF	CBLL1	CST3	ELL	GCLC	IL1RN	LINC00877	MEIS1	NDUF86	PGS1	RAD52	SAP30BP	SPTY2D1	TMEM209	WDR4
AHCTF1	BCAS3	CBR4	CTBP2	ENSA	GFI1	IL7R	LINC00926	MFNG	NDUF89	PHF2	RALGAP2	SAR1A	SRGN	TMEM39B	WDR66
AK2	BCL6	CBX1	CTDP1	ENTPD7	GLRX3	INCENP	LINC01003	MFSD2A	NEAT1	PHF20L1	RALGDS	SATB1	SRSF5	TMEM60	WDR70
AKR1A1	BDH1	CBX8	CTNNB1	EPDR1	GNA15	INPP5B	LINC01132	MICU1	NEDD9	PHLPP1	RAP1GAP2	SBF2-AS1	SSBP2	TMEM68	WDR81
ALDH4A1	BHLHE40	CCDC124	CTNNBIP1	ERLIN2	GNPDA1	INPP5F	LINC01353	MICU2	NEK6	PI4KA	RAP2B	SCFD2	SSR1	TNNI2	WNK1
ANKRD13D	BICD2	CCDC6	CTR9	ERO1LB	GOLPH3	INSIG2	LMBR1	MIR1281	NFATC3	PICALM	RASA1	SCOC	STAT1	TNRC18	XPO1
ANO10	BIN2	CCNG2	CTSD	ESF1	GPR114	IQCG	LMNB1	MIR23A	NFKBIA	PIGV	RASAL3	SDE2	STAT3	TNRC6B	YKT6
ANTXR2	BIRC5	CCNI	CYP51A1	ESRRA	GPR132	IQGAP2	LMTK2	MIR3687-1	NINJ2	PJA2	RASSF5	SEC14L1	STBD1	TPD52L2	ZBED5-AS1
APBB1IP	BLM	CCPG1	CYTIP	ESYT2	GPR68	IRF5	LOC100506023	MIR4437	NKG7	PLAUR	RBM43	SELPLG	STEAP4	TPK1	ZC3H3
APTX	BMP4	CCSER2	DACH1	ETFA	GPR84	ITGA4	LOC100507316	MIR4505	NLRP12	PLEC	RBM5	SENCR	STIP1	TRAF3IP3	ZMIZ1-AS1
ARAP1	BPI	CD164	DARS	EVL	GRK6	ITGAM	LOC100630918	MIR4706	NLRP3	PLEK	RBPJ	SEPT9	STMN1	TREML2	ZMYM4
ARF3	BRIP1	CD200R1	DCTN4	EYA3	GSK3A	ITGB2-AS1	LOC100996654	MIR548AN	NOL4L	PLEKHH3	RCAN1	SERPINB1	STOM	TRIP4	ZNF251
ARHGAP1	BRMS1L	CD244	DCUN1D4	FADS2	GSN	ITPKB-IT1	LOC101927780	MIR6785	NOL8	POLE	RELL1	SETX	STT3B	TRPC2	ZNF318
ARHGAP12	BRWD1	CDCA7	DENND4A	FAM102B	GSTA4	ITPRIP	LOC101928737	MKL1	NOTCH2NL	PPCDC	RFT1	SF3B1	STX6	TRPM2	ZNF385A
ARHGAP15	BTD	CDK5RAP2	DENND6A	FAM129B	GYPC	KAT6B	LOC101929551	MNT	NR1D1	PPP1R15A	RFX8	SFSWAP	SUB1	TSC22D4	ZNF397
ARHGAP22	BUD13	CDKN2AIP	DERL1	FAM179B	H2AFV	KCNIP2	LOC102724467	MOB3A	NR1D2	PPP1R9B	RG51	SFT2D1	SUMF1	TSEN54	ZNF497
ARHGAP25	AS1	CEP44	DGUOK	FAM192A	HAT1	KDM3A	LOC146880	MOSPD3	NRBF2	PPP2R5A	RG518	SFXN5	SURF4	TSN	ZNF507
ARHGAP30	C10orf128	CEP63	DHRS3	FAM208A	HBP1	KDM4B	LOC439933	MPO	NREP	PPP6R1	RG519	SGTA	SYPL1	TSNARE1	ZNF608
ARHGAP9	C11orf24	CEP68	DHX36	FAM214A	HCG27	KHDRBS1	LPXN	MRPL1	NSUN6	PRDM10	RHOB	SH2B3	SYTL3	TSNAX-DISC1	ZNF696
ARHGFE2	C11orf58	CERS6	DIAPH1	FAM219A	HCK	KIAA1919	LRGUK	MRPS22	NTMT1	PRMT3	RHOF	SH3BP2	TAB1	TTC33	ZNF787
ARL4A	C12orf66	CFL1	DIEF	FAM63A	HERPUD1	KIAA2018	LRRC40	MRPS23	NUCB2	PRMT7	RILPL2	SHB	TAOK3	TLL12	ZNRF2
ARL6IP5	C16orf72	CFLAR	DLST	FAR1	HEXA-AS1	KIF18A	LRRC6	MSRA	NUDCD1	PROSER2-AS1	RIN3	SIGLEC12	TBC1D10C	TXNL1	ZRANB3
ARNT	C18orf25	CHMP6	DLX2	FBXO15	HIF1AN	KIF2C	LRRC8D	MTDH	NUDCD3	PRPF18	RIOK1	SIRPB2	TBC1D22A	UBAC1	ZSCAN12
ARRB1	C18orf8	CHRA1	DMWD	FBXO32	HK3	KLF6	LUZP6	MTF2	OGFOD2	PRPF40A	RNF166	SLA	TBL1X	UBAP2	ZSCAN2
ARSG	C1orf220	CHSY1	DMXL1	FBXO38	HMG20A	KLLN	LY6E	MTG2	OSBP18	PRPSAP2	RNF4	SLA2	TBPL1	UBC	ZZZ3
ASXL1	C20orf197	CIPC	DOK2	FCGRT	HNRNP	KRTCAP2	LYL1	MXI1	OSM	PSAP	RNU6-2	SLC14A	TCERG1	UBE2S	
ASXL2	C21orf33	CKS2	DOLPP1	FCHSD2	HNRNP	LACE1	MALAT1	MYADM	OTUD7B	PSAT1	ROBO1	SLC25A19	TCF12	UPF2	
ATAD1	C3orf58	CLIC4	DROSHA	FERMT3	HNRNP	LACTB2	MAP2K1	MYCBP2	OTULIN	PSD3	RPL31	SLC35B1	TCFEX1D1	UPP1	
ATAD2B	C3orf67	CLN6	DUSP7	FGGY	HNRNP	LAI1	MAP3K3	MYH9	OXR1	PSD4	RPL38	SLC36A1	TEC	USP32	
ATAD3A	C4orf27	CNDP2	DYNC2H1	FHIT	HPS1	LAMTOR1	MAP3K8	MYL12B	OXS1	PSEN1	RPLP2	SLC39A11	TFRC	USP4	

While this method identified a set of potential enhancer linkages, it is time consuming and dependent on annotations to the nearest gene and therefore prone to excluding enhancers which operate on more distal genes. Therefore, an alternate method was used in parallel to bioinformatically identify promoter-enhancer pairs in the entire dataset in an unbiased manner. This method considered all intergenic/intronic RUNX1 peaks and linked them to the transcription start sites of RUNX1 promoter bound genes through the same ChIA-PET dataset used previously (Fullwood, Han et al. 2010). This analysis identified approximately 35,000 links between RUNX1 ChIP-Seq peaks, which were annotated to 9953 unique genes. Of the 635 genes with potential enhancers identified in the previous analysis, 404 were also identified using this method. Additional information was gathered on the linkage, including the chromatin state of the linked peak (as annotated by HOMER, described in Chapter 3), the score of the ChIA-PET linkage, which is a measure of confidence in the relevance of the detected interaction, and the presence of a recognised RUNX consensus sequence at either end. This dataset was then filtered to only include peaks identified as being in either weak or strong enhancer regions according to chromatin state. Circos plots were generated from the filtered data for each chromosome using the RCircos package in R (Zhang, Meltzer et al. 2013) to visualise enhancer-TSS links. An example Circos plot for chromosome 2 is depicted in Figure 5.2. The majority of the linkages identified occur over short distances, however some long-distance interactions were also identified. As enhancer linkage is uncommon between Topologically Associated Domains (TADs), the chromatin structure around these linkages requires further examination, as demonstration of interactions occurring within TADs would impart greater confidence that they are likely functional interactions.

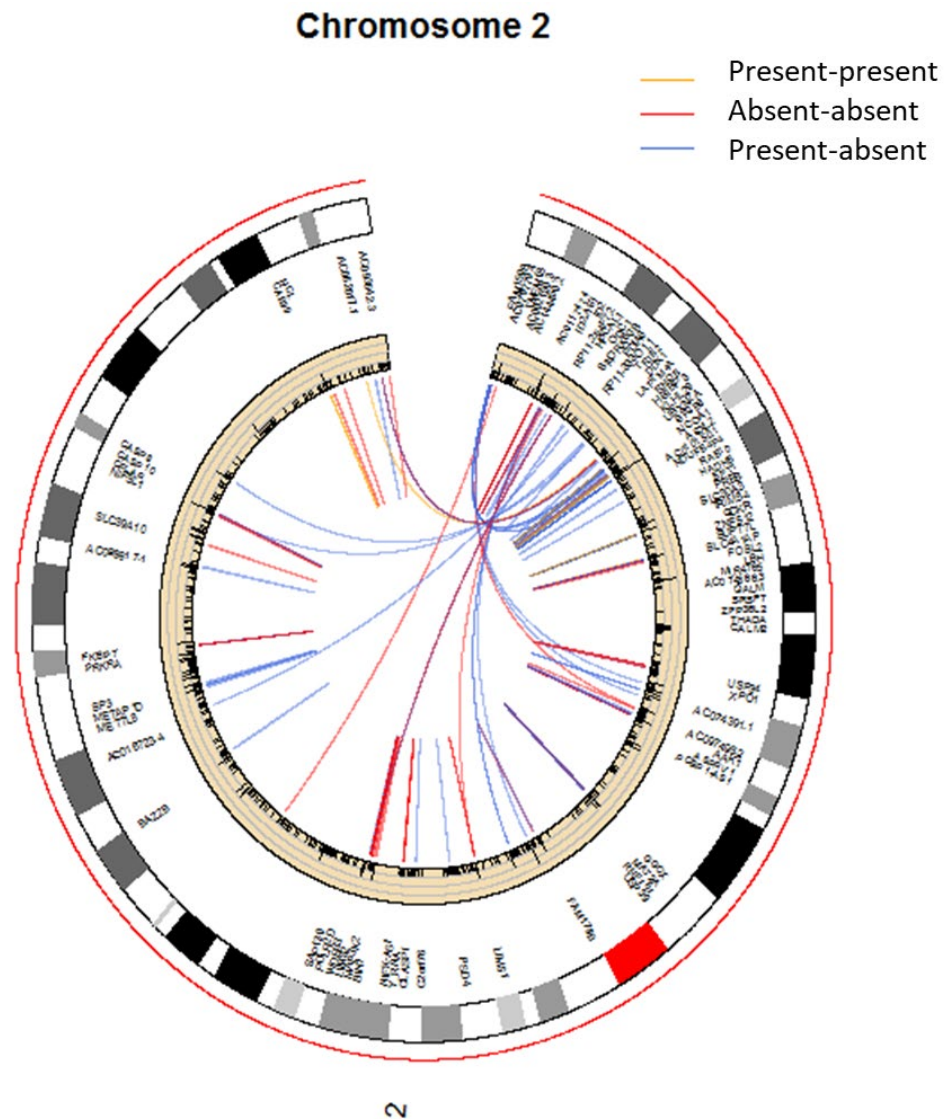


Figure 5.2. Representative map of RUNX1 promoter-intergenic/intronic ChIP-PET linkages.

Generated using the Rcirco package in R. The plot shows, from outside to inside, a map of Chromosome 2, names of identified linking genes and RUNX1 binding sites. The central area of the plot shows links between RUNX1-bound promoters and RUNX1-bound intergenic loci considered likely to be enhancers. Links represent regions of DNA which interact in the cell through RNA Polymerase II, as determined by ChIA-PET assay (Fullwood, Liu et al. 2009). Line colour corresponds to the presence of RUNX1 consensus sequences at the interacting sites: orange represents present-present interactions, red represents absent-absent and blue represents present-absent interactions.

5.2.2 Analysis of candidate promoter-enhance interactions

Bioinformatic analysis of publicly available datasets was therefore utilised to map intergenic or intronic RUNX1 ChIP peaks to their potential target promoters, but their function as enhancers and their ability to regulate the linked promoters remains to be tested. This analysis was therefore applied to the candidate RUNX1-bound promoters examined earlier. Of these genes, ARHGAP1, ARHGAP9, ARHGAP12, HAT1, and KAT6B were identified as genes with annotated promoter and intergenic/intronic ChIP peaks, and ARHGAP1, ARHGAP9, HAT1, KAT6B, KAT7, and KAT2B were included in the genome-wide ChIA-PET interaction database. Interestingly, despite being identified as RUNX1-bound promoters through ChIP-seq analysis (Chapter 3), both the ARHGAP1 and HAT1 promoters were found to be unresponsive to RUNX1 in reporter assays (Chapter 4).

Candidate gene promoters were inspected for strong ChIA-PET links (those with a score of 10 or above) between the RUNX1 ChIP peak site and distal regions. These distal regions were of particular interest if they contained another RUNX1 ChIP peak, with or without a consensus sequence, or were classified as an enhancer in the Functional Annotation of the Mammalian Genome 5 (FANTOM5) database, which comprises a map of transcription start sites and active enhancers derived from cap analysis of gene expression (CAGE) with single-molecule sequencing experimental data (Andersson, Gebhard et al. 2014). This analysis identified a number of potentially relevant chromatin interactions for candidate promoters (Figures 5.3 and 5.4). Interestingly, more interactions were identified for the set of promoters which bound RUNX1 in the absence of a recognised consensus sequence, which could be evidence that, RUNX1 binding at promoters in the absence of a consensus sequence is a result of indirect recruitment through a linked region of DNA. Of these identified interactions, those occurring in candidate genes which did not respond directly to RUNX1 (HAT1 and KAT6A) were prioritised

for further study, as this indicates that the cloned promoter fragment alone is insufficient for RUNX1 action at these genes.

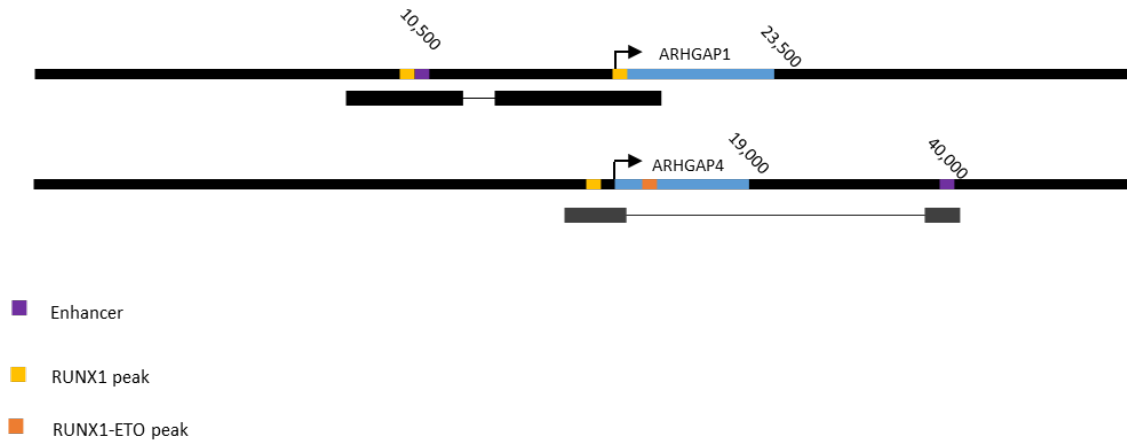


Figure 5.3. RUNX1 ChIP peaks linked to consensus sequence present promoters.

Schematic representation of consensus sequence present candidate genes and surrounding DNA. Genes are marked in blue, with transcription start sites indicated by arrows. RUNX1 ChIP peak sites are marked in yellow and RUNX1-ETO sites in orange. Enhancers, as identified by the FANTOM database (Andersson, Gebhard et al. 2014) are in purple. Linked regions identified in ChIA-PET data are indicated by connected bars, the colour of which indicates the strength of the ChIA-PET evidence, black for very strong and dark grey for moderately strong evidence (Fullwood, Liu et al. 2009). No evidence of linked peaks was identified for either ARHGAP9 or ARHGAP12.

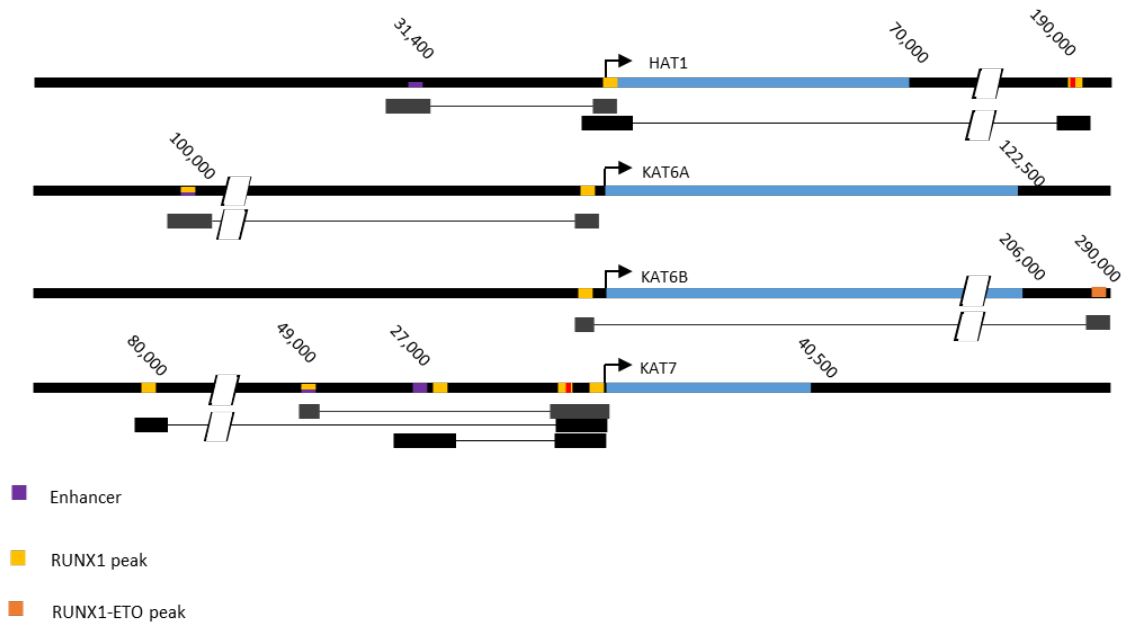


Figure 5.4. RUNX1 ChIP peaks linked to consensus sequence absent promoters.

Schematic representation of consensus sequence present candidate genes and surrounding DNA. Genes are marked in blue, with transcription start sites indicated by arrows. RUNX1 ChIP peak sites are marked in yellow and RUNX1-ETO sites in orange. Enhancers, as identified by the FANTOM5 database (Andersson, Gebhard et al. 2014) are in purple. Linked regions identified in ChIA-PET data are indicated by connected bars, the colour of which indicates the strength of the ChIA-PET evidence, black for very strong and dark grey for moderately strong evidence (Fullwood, Liu et al. 2009).

Two regions linked through ChIA-PET data were identified for HAT1 (Figure 5.5a, Table 5.2), one approximately 31 kb upstream of the HAT1 TSS (HAT1-31k) and one 190 kb downstream (HAT1+190k). The -31 kb region does not have a RUNX1 peak in the Trombly dataset which was used for this analysis (Trombly, Whitfield et al. 2015), but review of the Ptasinska RUNX1 ChIP-seq dataset (Ptasinska, Assi et al. 2014), revealed an uncalled potential RUNX1 ChIP peak. This region is also identified as an active enhancer by FANTOM5. The +190 kb region contains a RUNX1 ChIP peak in the Trombly dataset (Trombly, Whitfield et al. 2015) which also contains a recognised RUNX1 consensus sequence. One linked region was identified for KAT6A approximately 100 kb upstream of the TSS (KAT6A-100k, Figure 5.5b, Table 5.2). This region contained a RUNX1-ETO ChIP peak but not RUNX1 in the Trombly dataset (Trombly, Whitfield et al. 2015) and a probable uncalled RUNX1 peak in the Ptasinska data (Ptasinska, Assi et al. 2014). Interestingly, the interacting region for KAT6A was not detected in the earlier analyses. This is because the linked region does not have a called RUNX1 ChIP peak and was therefore not considered as a potential interacting region.

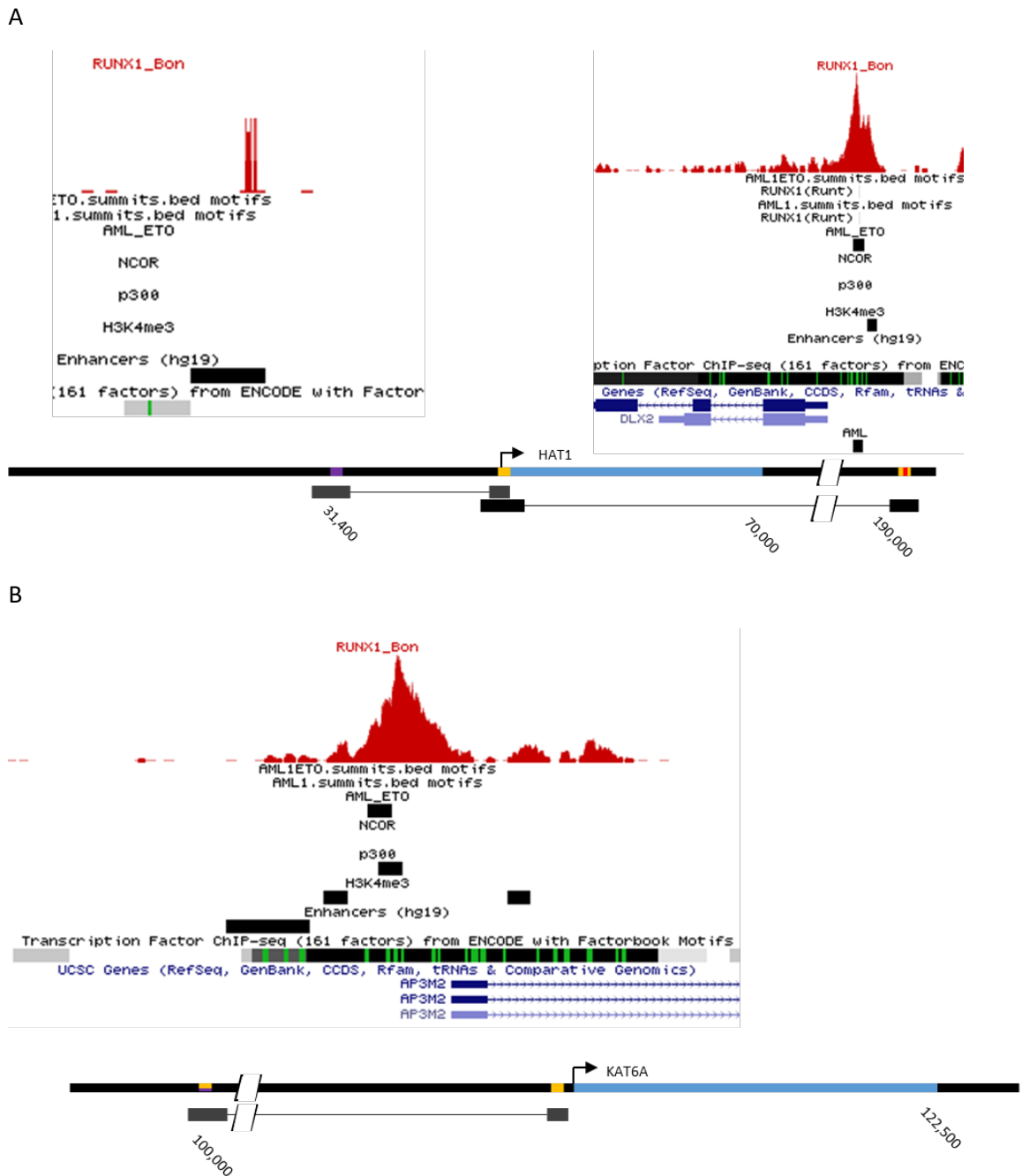


Figure 5.5. Features of candidate linked regions.

Schematic representation of consensus sequence absent candidate genes **A) HAT1** and **B) KAT6A** and surrounding DNA. Genes are marked in blue, with transcription start sites indicated by arrows. RUNX1 ChIP peak sites are marked in yellow and RUNX1-ETO sites in orange. Enhancers, as identified by the FANTOM database (Andersson, Gebhard et al. 2014) are in purple. Linked regions identified in ChIA-PET data are indicated by connected bars, the colour of which indicates the strength of the ChIA-PET evidence, black for very strong and dark grey for moderately strong evidence. Screenshots from UCSC genome browser show raw RUNX1 ChIP peaks in red (Ptasinska, Assi et al. 2014) and peak called sites for RUNX1 (labelled AML in browser), RUNX1-ETO (labelled AML_ETO), N-CoR, p300, and H3K4 methylation designated by black bars (Trombly, Whitfield et al. 2015) at the linked site, as well as nearby genes, RUNX consensus sequences and enhancer locations.

5.2.2.1 *Epigenetic marks at selected regions*

Histone modifications in the vicinity of the interacting regions were investigated for evidence of enhancer activity using ENCODE data as before, specifically the presence of H3K4Me1 and H3K27Ac, and the absence of H3K4Me3 (Table 5.2). The HAT1-31k site is located in an intron of the SLC25A12 gene and has no evidence of any of the three histone modifications. The HAT1+190k site shows evidence of all three marks, consistent with an active regulatory region, but more indicative of promoter function. This would be consistent with its location adjacent to the DLX2 gene and it is likely that it is a promoter element involved in regulating the nearby DLX2 gene. The KAT6A-100k site likewise has evidence of association with all three histone marks and may be associated with the promoter of the AP3M2 gene, which is 350 bp away from the RUNX1-ETO ChIP peak. However, the activity of enhancers is cell type specific, and the ENCODE data is sourced from ChIP-seq studies from seven cell lines. When considering only histone modifications data from K562 cells, the HAT1+190k site remains unchanged, but the presence of all three modifications is less pronounced at the KAT6A-100k site. While this analysis does not provide strong support for these elements functioning as enhancers it is suggestive of them having regulatory function.

Table 5.2. Features of candidate linked regions.

	HAT1+190k	HAT1-31k	KAT6A-100k
RUNX1 ChIP peak	✓	✗	✗
RUNX1-ETO ChIP peak	✓	✗	✓
Uncalled RUNX1 ChIP peak	✓	✓	✓
H3K4Me1	✓	✗	✓
H3K27Ac	✓	✗	✓
H3K4Me3	✓	✗	✓
FANTOM5 Enhancer	✗	✓	✗
Nearby genes	DLX2 (promoter)	SLC25A12 (intron)	AP3M2 (promoter)

5.2.3 Enhancer reporter assays

Potential enhancer regions were isolated and cloned into the pXpG luciferase reporter plasmid, adjacent to the relevant promoter and in both forward and reverse orientation. Constructs were transfected into K562 cells in the presence of RUNX1 overexpression, and luciferase activity measured to determine whether there was any change in activity, in response to RUNX1, caused by the enhancer compared to the promoter construct alone. The KAT6A enhancer construct (in either orientation) showed no change in activity compared to the promoter alone, indicating that the +100 kb region is not functioning as an enhancer, at least in this context (Figure 5.6a). The HAT1+190 kb linked region showed increased activity in response to RUNX1, in the reverse direction only (Figure 5.6b). This indicates that this region may function as an enhancer of HAT1 expression. The +190 kb region contains a canonical RUNX consensus sequence, so its response to RUNX1 may be expected, though epigenetic marks at the locus do not clearly indicate that it is an enhancer. Further, it increased activity of the promoter only when it was in the reverse orientation, suggesting activity more consistent with a promoter element. Interestingly, this is the orientation in which this element is positioned adjacent to the DLX2 gene. The HAT1-31 kb region showed no significant response to RUNX1 in either orientation, suggesting that, as for KAT6A-100k, this region is not acting as an enhancer in this assay (Figure 5.6c).

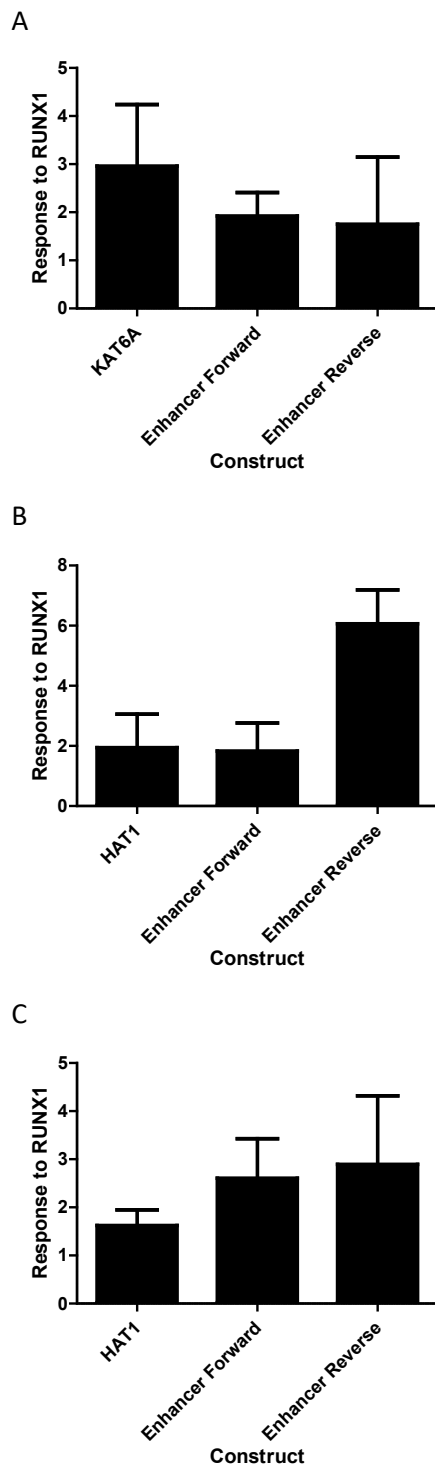


Figure 5.6. Response of enhancer constructs to RUNX1.

K562 cells were transfected with promoter-pXPG or enhancer-promoter-pXPG constructs as well as a RUNX1 expression or CMV control vector. Protein was harvested 24 hours post-transfection and luciferase activity was measured. Values are expressed as mean values \pm SEM (n=3) of luminescence of RUNX1 overexpression transfections relative to control for: **A)** KAT6A-100k; **B)** HAT1+190k and **C)** HAT1-31k. Statistical analysis was conducted using one-way ANOVA with Tukey's multiple comparisons post-test. A significant difference in means was detected for HAT1+190k ($p < 0.05$) but the post-test did not identify any difference in column comparisons.

5.2.4 Validation of enhancers by chromatin conformation capture

The region 190 kb downstream of the HAT1 promoter increased response to RUNX1 above the HAT1 promoter alone *in vitro*, suggesting that it may be able to function as an enhancer of HAT1 expression. This region, and the region 31 kb upstream were selected for *in vivo* analysis of enhancer-promoter linkage. The two regions that were analysed as potential enhancers of HAT1 were suggested to be linked to the HAT1 promoter through RNA PolII ChIA-PET data. To validate these interactions in another experimental model a chromatin conformation capture (3C) assay was conducted. Unlike the ChIA-PET analysis, this tests DNA interactions directly and does not rely on the presence of a particular protein, such as RNA PolII. 3C utilises formaldehyde fixation to form crosslinks between nearby chromatin regions, followed by restriction enzyme digestion and ligation of digested ends. This method generates fragments of DNA consisting of two interacting regions ligated together at the restriction site. Using primers designed to each fragment component, the presence of specific interacting regions can be detected by qPCR. Primer sets were designed to interrogate regions 40 kb upstream and 200kb downstream of the promoter (Figure 5.7) and used to interrogate DNA fragments isolated from KG1a cells. This cell line was chosen to complement the data generated by ChIA-PET in the K562 cells.

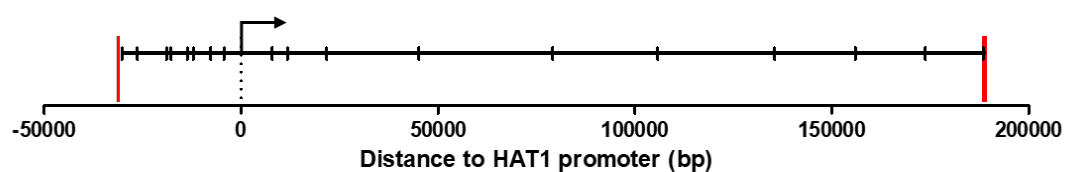
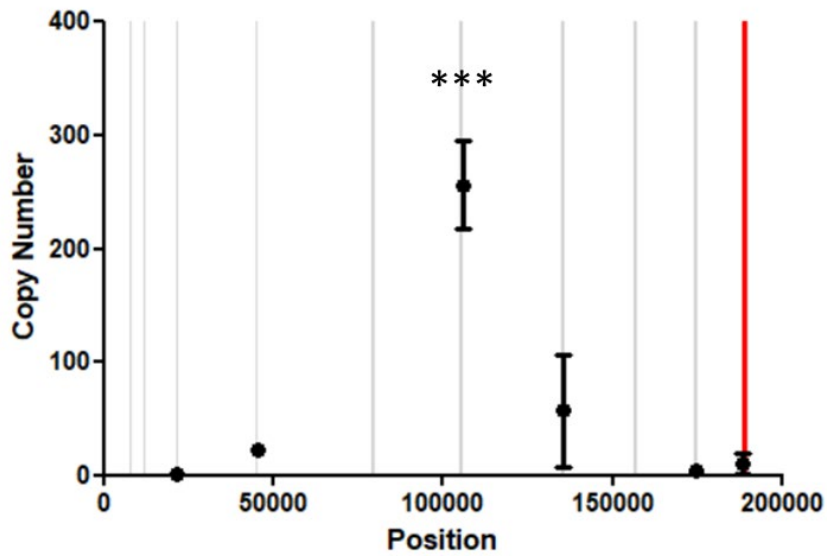


Figure 5.7. Map of 3C primers used to analyse interaction with the HAT1 promoter.

Schematic diagram of the regions analysed by 3C assay. Horizontal lines represent fragments analysed by qPCR, red bars mark the interacting regions of interest. The HAT1 TSS is marked by the arrow.

The 3C assay detected interactions between both the +190k and -31k regions and the HAT1 promoter anchor fragment. Amplicons representing interactions with both of these regions were detected at higher levels than the surrounding regions, although the +190 kb region was not detected as strongly as the -31kb region and was not statistically significant (Figure 5.8). Taken together with the ChIA-PET data from K562 cells, this provides strong evidence of interaction between these regions. The evidence for the HAT1+190 kb region is less clear, and the promoter-distal region linkage is not as pronounced. While it is possible that this region exerts some influence on the HAT1 promoter as an enhancer, it is more probable that it functions as the promoter of a different gene. An additional region at approximately 100 kb downstream of the HAT1 promoter was identified as highly significant in this assay and may represent another potential enhancer of HAT1.

A



B

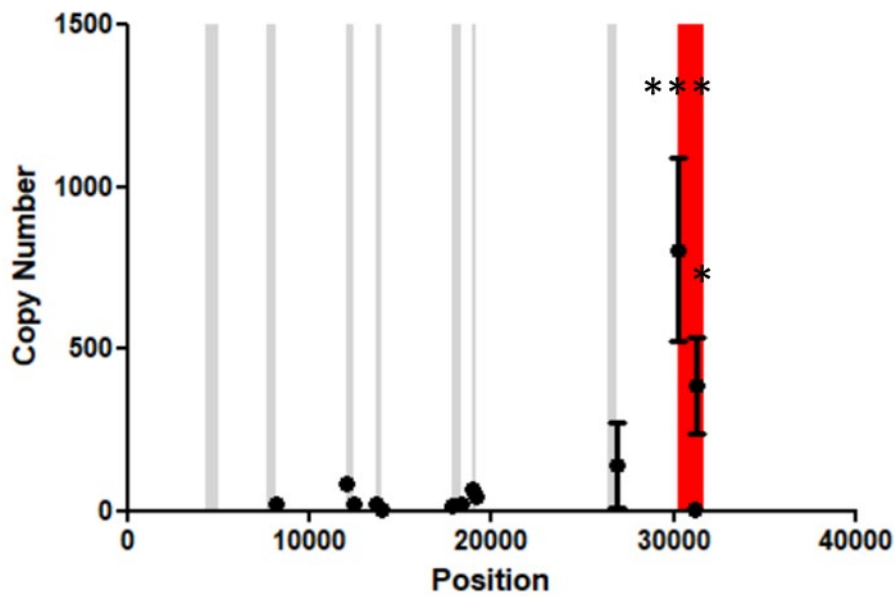


Figure 5.8. Regions linked to the HAT1 promoter through 3C.

Copy numbers of regions ligated to the HAT1 promoter as a result of 3C assay as determined by qPCR for **A)** 190 kb upstream and **B)** 31 kb downstream of the TSS. Position of each region is relative to the HAT1 promoter RUNX1 ChIP peak. Values are expressed as mean \pm SEM (n=3). Fragments assayed by qPCR are represented by grey bands and the interacting region of interest is represented by the red bar. Statistical analysis was conducted using one-way ANOVA with Dunnett's multiple comparisons post-test against a negative control. *p<0.05, ***p<0.001.

5.3 Discussion

The majority of RUNX1 binding identified in Kasumi-1 cells was found within intergenic or intronic regions. As RUNX1 is known to regulate some genes through their enhancers, it is reasonable to infer that gene regulation through enhancers is the functional outcome of RUNX1 binding in at least some of these regions. However, it is difficult to determine which binding sites are functional enhancers from genomic location alone, and even more challenging to ascertain the gene being regulated by each site. The methods discussed above attempts to use publicly available experimental data to elucidate the role of these intergenic/intronic binding sites, firstly by starting with promoter binding sites and inspecting for nearby intergenic sites, and secondly by starting with the intergenic/intronic sites themselves and linking to promoter regions. Both methods are valuable, and the database of linkages provides a genome-wide set of possible RUNX1 enhancer interactions which can be interrogated for genes of interest.

The available ChIA-PET data used in this analysis is not ideal, as the assay was conducted in K562 cells (granulocytic and undifferentiated) while the ChIP-seq data is from Kasumi-1 cells (myeloblasts which contain the t(8;21) translocation). While these cell lines are both derived from leukaemic cells, it is not certain that chromatin interactions observed in one cell type are applicable in the other. Further influencing this is the 3C assay, which was conducted in KG1a cells (derived from AML bone marrow) because of the baseline activity of the candidate genes. While it is reasonable to expect some overlap in enhancer activity between these cell lines given their similarity, enhancer function is highly dependent on cell type (Jin, Li et al. 2011) and therefore a lack of evidence of enhancer function in one cell type does not conclusively rule out the possibility that the region acts as an enhancer in other cells.

Given that a number of gene promoters had been identified in Chapter 3 as binding RUNX1 through ChIP-seq analysis, but did not respond to RUNX1 in reporter assays, these promoters represented ideal candidates for analysis of potential enhancer interactions. If these genes are genuine RUNX1 targets and the promoter itself is not sufficient for RUNX1 response, it is reasonable to infer that regulation is dependent on an additional response element, such as an enhancer. Potential enhancer regions were identified for the HAT1 and KAT6A genes based on linkage through ChIA-PET data. While HAT1 was identified as a gene with a potential enhancer in the linkage databases, KAT6A was not. This highlights the limitations of the filtering strategy employed. In addition to identifying binding that may not represent functional enhancers, potential enhancers which do not meet the selection criteria may be missed. In this case, as the linked region did not contain a RUNX1 ChIP peak in the Trombly dataset (Trombly, Whitfield et al. 2015), it was not considered in either bioinformatic analysis. However, the lack of response observed in the promoter reporter assays prompted a manual search for linked regions. This identified a linked RUNX1-ETO ChIP peak and potential uncalled RUNX1 ChIP peak from the Ptasinska data (Ptasinska, Assi et al. 2014).

The KAT6A-100k region did not show a response to RUNX1 in reporter assays. This suggests that the region is not acting as an enhancer of the KAT6A gene. It is possible, as the linked region identified is in the promoter of another gene (AP3M2) that this is a true enhancer interaction, but that the interaction goes in the opposite direction, with the KAT6A promoter RUNX1 site acting as an enhancer for the promoter at the other end. A reporter assay with the KAT6A-100k region and the promoter region of AP3M2 could be conducted to investigate this possibility.

The orientation specific activity of the +190 kb region is curious, as enhancer regions should function irrespective of direction. The linked region is, like the KAT6A region, in the promoter of another gene, in this case DLX2. The orientation of the DLX2 gene is such that the reversed HAT1+190 kb region is essentially the DLX2 promoter region in the correct orientation for transcription. It is possible that the reverse direction specific activity of the cloned region is the result of it being a RUNX1-responsive promoter instead of an enhancer. If so, this is not a particularly surprising result as the region contains a RUNX1 consensus sequence and its regulation by RUNX1 is to be expected. The 3C assay identified some evidence of interaction between the HAT1 promoter ChIP peak and the HAT1+190k region, though it was not the strongest interaction. The assay indicated another potential linked region at approximately 105 kb downstream on the HAT1 TSS. This interaction was not detected in ChIA-PET linkage and the linked region falls within the first intron of the METAP1D gene. There is some evidence of H3K4Me1 marks in K562 cells at this locus, suggesting that it may have an active regulatory role, though the nearest RUNX1 ChIP peak is 20 kb away at the METAP1D promoter.

While the evidence is not particularly strong, there is some indication that the HAT1-31k region is acting as a RUNX1-responsive enhancer of the HAT1 gene. This illustrates a model of RUNX1 activity which involves binding at distal regulatory elements which interact with gene promoters in the absence of consensus sequences at either site. Having demonstrated the feasibility of this type of RUNX1 interaction, the question of the mechanism occurring in these cases remains.

Promoter-promoter interactions, such as those observed for HAT1 and KAT6A, have been observed in approximately 42% of genome-wide ChIA-PET links (Li, Ruan et al. 2012). These interactions have been proposed to be a result of the formation of multi-gene complexes

associated with the transcription machinery. It is further suggested that these promoters may have regulatory function on other linked promoters, acting as both promoters and enhancers simultaneously. This kind of chromatin interaction may account for a proportion of the observed RUNX1 binding in the absence of consensus sequences for ChIP-seq data. RUNX1 recruited to a promoter in a consensus sequence-dependent manner may be detected at other promoter sequences through cross-linking assays if these promoters are components of the same transcription complex.

6 Final discussion and future directions

The precise maintenance of complex regulatory networks is essential for regular cellular functions such as development and differentiation. The systems by which regulatory instructions are encoded and subsequently read by transcription factors are therefore a major focus for research. These systems are intricate intersections of many processes which must be both tightly controlled and highly responsive to changes in the cellular environment. It is therefore not straightforward to decipher their mechanisms of action. As new techniques to interrogate regulatory networks and study transcription factor function on a genome-wide scale are developed, our ability to understand the complex interaction between transcription factors and their targets, and therefore the cellular processes that they regulate, improves.

RUNX1 is a master regulator of haematopoiesis, and a frequent target for mutations and translocations in leukaemia. A complete understanding of the mechanisms of RUNX1 action is necessary in order to fully comprehend its effects on regulatory networks and the impact of its disruption. The mechanism of RUNX1 action has always been understood in the context of its consensus sequence, which acts as both a marker for the transcription factor's intended targets and the site of protein-DNA binding. However, evidence has emerged from both single gene and genome-wide studies which indicates that this view of RUNX1 action is incomplete. Recent studies of other transcription factors have brought into question the paradigm of transcription factors operating exclusively through interaction with specific DNA consensus sequences.

One of the first and best described transcription factors to be implicated in binding in the absence of consensus sequences, Myc, was shown to have better genome-wide occupancy

correlation with transcription start sites, and therefore the transcriptional machinery, than its canonical E-box consensus sequences (Guo, Li et al. 2014). DNA sequences preferentially bound by the MYC/MAX heterodimer were subsequently investigated through protein binding microarrays and electrophoretic gel shift assays for all possible hexamers (Allevato, Bolotin et al. 2017). At least one of the top 21 bound sequences was detected at 87% of binding sites genome-wide, though binding to low-affinity sites was modulated by MYC dosage. Another recent study into the feasibility of the two proposed mechanisms of Myc action, E-box driven or otherwise, used an inducible Myc model system and mathematical modelling of binding affinities (Lorenzin, Benary et al. 2016). The study concluded that direct DNA binding does not fully explain the occupancy of Myc at promoters and proposes interactions with other proteins such as WDR5 and transcription machinery components P-TEFb, TBP and TRRAP.

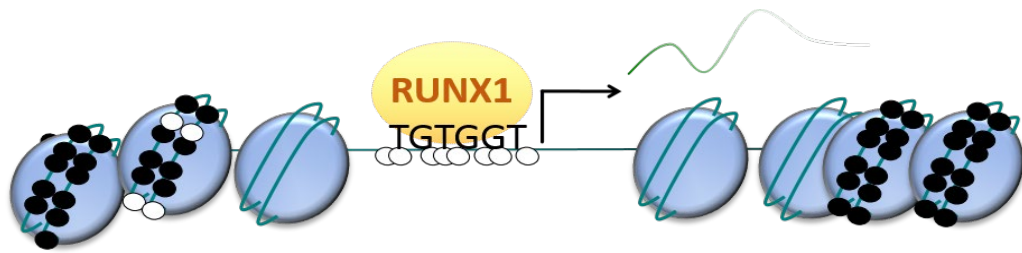
The data presented here from analysis of ChIP-seq experiments in human leukaemic cell lines supports findings from similar studies in mouse models that question the classically held view of RUNX1 binding and function (Tanaka, Joshi et al. 2012). Specifically, these analyses have demonstrated that the majority of RUNX1 binding occurs outside of promoter regions, and that some of the binding occurs in the absence of canonical consensus sequences. Further analysis of this data has begun to build a picture of the context of RUNX1 binding events in the absence of consensus sequences. Firstly, these binding events occur more frequently in promoter regions than elsewhere in the genome (almost half in promoters compared to 21% genome-wide). Secondly, binding in the absence of consensus sequences is associated with a higher CG percentage and more CpG dinucleotides. Thirdly, binding in the absence of consensus sequences was more closely associated with p300 binding than RUNX1 binding in the presence of a consensus sequence. Finally, the binding sites are enriched for a range of other transcription factor consensus sequences, including many ETS family member motifs.

Taken together, this suggests that in the absence of consensus sequences RUNX1 is recruited to its targets through interaction with additional factors which accumulate in the promoter regions of genes and are potentially dependent on DNA methylation status and the co-factor p300.

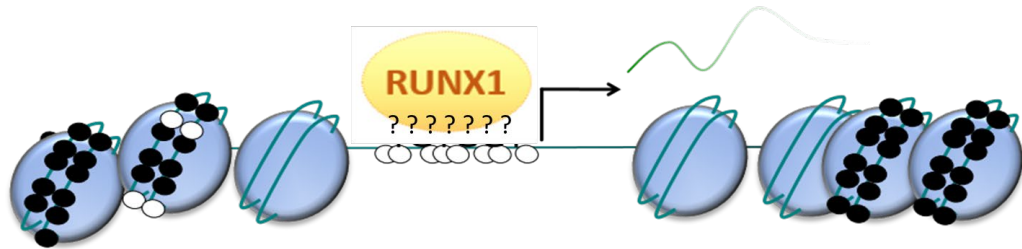
Genes bound by RUNX1 in different modalities were demonstrated to cluster within biological pathways. The prevalence of pathways containing more than 10 genes with RUNX1-bound promoters under only one binding modality suggests that RUNX1 may control different biological processes through different mechanisms. Further, it is possible that these mechanisms respond differently to changes in the cellular environment and allow the two subsets of RUNX1 targets to operate independently of each other.

Analysis of candidate genes selected from these different pathways by expression, reporter, and linkage assays has begun to elucidate the different modalities of RUNX1 action. Data from this study suggests three distinct models of RUNX1 transcriptional regulation. The first of these is the classical RUNX1 mechanism, in which RUNX1 is recruited directly through its consensus sequence in gene promoters (Figure 6.1a). This is illustrated by the “consensus present” candidate genes ARHGAP4 and ARHGAP12, which were shown to respond to RUNX1 in reporter assays. This mechanism is well understood in the literature and represents the majority of known RUNX1 activity. Further RUNX1 activation of these genes is inhibited by the RUNX1-ETO fusion protein.

A



B



C

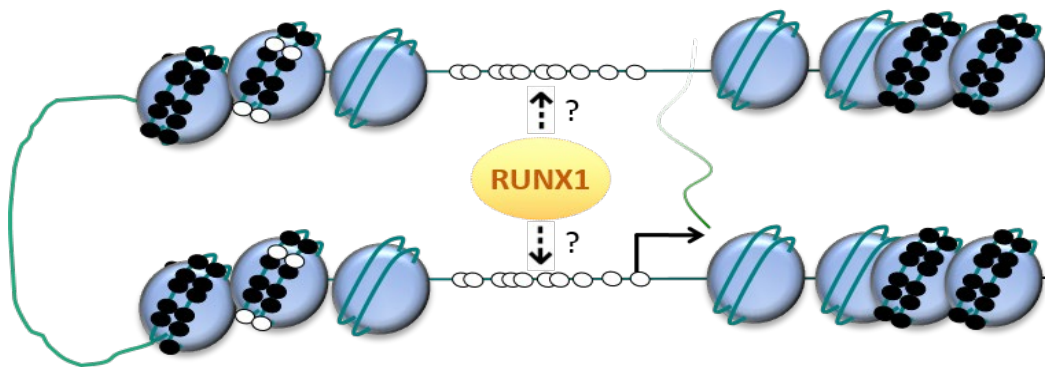


Figure 6.1. Models of regulation by RUNX1.

Depiction of RUNX1 action through three mechanisms: **A)** directly through its consensus sequence; **B)** directly in the absence of its consensus sequence; and **C)** indirectly, through interaction with distal regulatory elements and subsequent loop formation. DNA is represented by blue lines wrapped around nucleosomes (blue spheres) and CpG sites are represented by white (unmethylated) or black (methylated) circles on the DNA. Transcription start sites are represented by arrows.

The second proposed model is one in which RUNX1 is recruited directly to gene promoters through a mechanism other than its consensus sequence (Figure 6.1b). This is illustrated by the case of KAT6B and KAT2B, which responded to RUNX1 in reporter assays. While data from this study provides evidence that RUNX1 can regulate promoter activity in the absence of its consensus sequence, the exact mechanism involved remains to be elucidated. This manner of RUNX1 binding and function has been previously described at the promoter of the ITGB4 gene, where it is hypothesised to bind as part of a complex of hematopoietic transcription factors (Phillips, Taberlay et al. 2018). It is possible that RUNX1 is able to bind directly to lower affinity DNA sites which do not contain consensus sequences, though in this case what is directing the transcription factor specifically to its targets is unclear. There is no evidence from the literature that RUNX1 functions in a non-specific manner, and such a scenario is not supported by the distribution of RUNX1 binding observed from ChIP-seq data. If the binding was occurring in a completely sequence-independent manner it is unlikely that the proportion of “consensus absent” binding would be higher in promoter regions than across the rest of the genome. This suggests that an alternative mechanism is driving RUNX1 recruitment in the absence of its consensus sequence. A likely candidate, FLI-1, was ruled out as a cooperative binding factor in the cases of HAT1 and ARHGAP1, as while the promoter region was responsive to FLI-1, combinatorial transfections showed no cumulative response for RUNX1 and FLI-1 together. Importantly, while RUNX1-ETO was also shown to bind to these promoters in vivo, it did not inhibit RUNX1 activity at these promoters. However, while some promoters without RUNX consensus sequences responded to RUNX1, this was not the case for all promoters investigated. Specifically, the HAT1 and KAT6A promoters did not respond in reporter assays. This suggests that direct action of RUNX1 at promoters does not account for all ChIP binding identified in the absence of consensus sequences.

The third model of RUNX1 action involves direct recruitment to enhancer regions and subsequent interaction with promoters through chromatin proximity (Figure 6.1c). This model incorporates the evidence from ChIP-seq that RUNX1 binding is common at regions other than promoters, many of which may be serving as enhancers. This is demonstrated, to some degree, in the case of HAT1 for the region 31 kb upstream of the TSS through 3C analysis, though the region did not respond in reporter assays. This particular example does not contain a RUNX consensus sequence at either end of the interaction, and therefore leads to the same questions as the second model regarding RUNX1 recruitment to either loci. The region 190 kb downstream of the HAT1 promoter shows some evidence of interaction with the promoter from 3C data and responds to RUNX1 in reporter assays, but only in the reverse orientation. These results, coupled with the proximity of the interacting region to the DLX2 TSS, suggests that the RUNX1 ChIP site at the HAT1 promoter is involved in promoter-promoter interactions. Such interactions are common, and have been detected in genome-wide ChIA-PET and HiC studies (Li, Ruan et al. 2012, Schoenfelder, Javierre et al. 2018). Promoter-promoter interactions occur largely, but not exclusively, within TADs and are likely the result of simultaneous interactions with transcriptional machinery, forming large “transcription factories” driving expression of genes regulated by the same transcription factors (Li, Ruan et al. 2012). Like promoter-enhancer interactions, these associations are dynamic and dependent of cell type. Alterations in cell type can lead to loss or gain of promoter-promoter interactions, as demonstrated in a mouse embryonic stem cell model (Joshi, Wang et al. 2015).

There are limitations on the conclusions which can be drawn regarding transcription factor action from reporter assays. As the region of interest is amplified and ligated into plasmid DNA, it becomes divorced from its original chromatin context. Therefore, while these assays are useful in determining if a particular region is responsive to RUNX1, a negative result does

not definitively rule out the region as a RUNX1 target. Indeed, as discussed previously, there is some evidence for the involvement of epigenetic factors such as DNA methylation in the mechanism of RUNX1 action in the absence of its consensus sequence.

In addition to the DNA sequence of the binding site, the specific structure of the site may influence the specificity of transcription factor binding. Factors such as helical twist and minor groove width contribute to the ability of DNA binding proteins to recognise binding sites. The “shape motif” of binding sites is not entirely independent of DNA sequence but provide an extra layer of specificity. A tool has been developed to determine DNA shapes enriched at binding sites, which has been utilised to demonstrate that MYC-MAX heterodimer has a different shape preference than MAX alone (Samee, Bruneau et al. 2019).

The ChIP-seq analysis conducted in this study provides a useful resource for further investigation of RUNX1 binding in both a non-promoter and non-consensus sequence manner. The difficulty comes in finding functionally relevant RUNX1 binding targets within this large body of data. The current study utilised a candidate gene approach to begin mining the data for evidence of RUNX1 functionality, which resulted in the selection of clusters of genes from the RhoA and HMGB1 signalling pathways. These, and other pathways which contained genes bound by RUNX1 exclusively in the presence or absence of consensus sequences identified in the analysis contain more genes than were selected for analysis here. Analysis of RUNX1 function at these additional gene promoters would provide additional evidence of the conditions under which RUNX1 functions in the absence of consensus sequences. It is difficult with current evidence to draw strong conclusions about the necessary context for RUNX1 binding in the three models proposed above from the few clear examples of each in the selected candidate genes. With more data from candidate studies, it may be possible to

identify commonalities between genes regulated through different modalities which drive the different mechanisms of RUNX1 action. Commonalities between genes regulated by RUNX1 in the absence of consensus sequences would provide both tools with which to interrogate the dataset to identify more candidates, and also insight into the mechanisms occurring.

To elucidate the composition of any transcription factor complex interacting with RUNX1, pull-down assays potentially coupled with mass spectroscopy could be used. This method has previously been utilised to identify proteins which interact with T-box transcription factor 18 (TBX18) and Forkhead box M1 (FOXM1) (Xie, Jia et al. 2017, Rivera-Reyes, Kleppa et al. 2018). This analysis would provide an indication of other proteins which commonly interact with RUNX1 and could be used as a basis for investigating additional factors involved in RUNX1 binding identified in the absence of consensus sequences.

It is clear that the mechanisms by which RUNX1 regulates transcription in leukaemic cells is more complex than previously thought. This study has identified novel RUNX1 binding modalities and demonstrated their functional relevance. These findings expand our understanding of the role of RUNX1 in gene regulation and open new avenues of investigation into the transcriptional mechanisms involved in transcription factor regulation by “master regulators” such as RUNX1 in both healthy cells in haematopoiesis and disease states such as leukaemia.

7 References

- Achinger-Kawecka, J., P. C. Taberlay and S. J. Clark (2016). Alterations in three-dimensional organization of the cancer genome and epigenome. Cold Spring Harbor symposia on quantitative biology, Cold Spring Harbor Laboratory Press.
- Allevato, M., E. Bolotin, M. Grossman, D. Mane-Padros, F. M. Sladek and E. Martinez (2017). "Sequence-specific DNA binding by MYC/MAX to low-affinity non-E-box motifs." PloS one **12**(7): e0180147.
- Amann, J. M., J. Nip, D. K. Strom, B. Lutterbach, H. Harada, N. Lenny, J. R. Downing, S. Meyers and S. W. Hiebert (2001). "ETO, a target of t (8; 21) in acute leukemia, makes distinct contacts with multiple histone deacetylases and binds mSin3A through its oligomerization domain." Molecular and cellular biology **21**(19): 6470-6483.
- Andersson, R., C. Gebhard, I. Miguel-Escalada, I. Hoof, J. Bornholdt, M. Boyd, Y. Chen, X. Zhao, C. Schmidl and T. Suzuki (2014). "An atlas of active enhancers across human cell types and tissues." Nature **507**(7493): 455.
- Armesilla, A. L., D. Calvo and M. A. Vega (1996). "Structural and Functional Characterization of the Human CD36 Gene Promoter IDENTIFICATION OF A PROXIMAL PEBP2/CBF SITE." Journal of Biological Chemistry **271**(13): 7781-7787.
- Aronson, B. D., A. L. Fisher, K. Blechman, M. Caudy and J. P. Gergen (1997). "Groucho-dependent and-independent repression activities of Runt domain proteins." Molecular and cellular biology **17**(9): 5581-5587.
- Australian Institute of Health and Welfare (2019). Cancer data in Australia. Canberra, AIHW.
- Bäckström, S., M. Wolf-Watz, C. Grundström, T. Härd, T. Grundström and U. H. Sauer (2002). "The RUNX1 Runt domain at 1.25 Å resolution: a structural switch and specifically bound chloride ions modulate DNA binding." Journal of molecular biology **322**(2): 259-272.
- Barseguian, K., B. Lutterbach, S. W. Hiebert, J. Nickerson, J. B. Lian, J. L. Stein, A. J. Van Wijnen and G. S. Stein (2002). "Multiple subnuclear targeting signals of the leukemia-related AML1/ETO and ETO repressor proteins." Proceedings of the National Academy of Sciences **99**(24): 15434-15439.
- Barski, A., R. Jothi, S. Cuddapah, K. Cui, T.-Y. Roh, D. E. Schones and K. Zhao (2009). "Chromatin poises miRNA-and protein-coding genes for expression." Genome research **19**(10): 1742-1751.
- Bartfeld, D., L. Shimon, G. C. Couture, D. Rabinovich, F. Frolow, D. Levanon, Y. Groner and Z. Shakked (2002). "DNA recognition by the RUNX1 transcription factor is mediated by an allosteric transition in the RUNT domain and by DNA bending." Structure **10**(10): 1395-1407.
- Berardi, M. J., C. Sun, M. Zehr, F. Abildgaard, J. Peng, N. A. Speck and J. H. Bushweller (1999). "The Ig fold of the core binding factor α Runt domain is a member of a family of structurally and functionally related Ig-fold DNA-binding domains." Structure **7**(10): 1247-1256.

- Bert, A. G., J. Burrows, C. S. Osborne and P. N. Cockerill (2000). "Generation of an improved luciferase reporter gene plasmid that employs a novel mechanism for high-copy replication." Plasmid **44**(2): 173-182.
- Blackwell, T. K., L. Kretzner, E. M. Blackwood, R. N. Eisenman and H. Weintraub (1990). "Sequence-specific DNA binding by the c-Myc protein." Science **250**(4984): 1149-1151.
- Bowers, S. R., F. J. Calero-Nieto, S. Valeaux, N. Fernandez-Fuentes and P. N. Cockerill (2010). "Runx1 binds as a dimeric complex to overlapping Runx1 sites within a palindromic element in the human GM-CSF enhancer." Nucleic acids research **38**(18): 6124-6134.
- Brettingham-Moore, K. H., P. C. Taberlay and A. F. Holloway (2015). "Interplay between Transcription Factors and the Epigenome: Insight from the Role of RUNX1 in Leukemia." Frontiers in immunology **6**: 499.
- Bulger, M. and M. Groudine (2011). "Functional and mechanistic diversity of distal transcription enhancers." Cell **144**(3): 327-339.
- Calabi, F., R. Pannell and G. Pavloska (2001). "Gene Targeting Reveals a Crucial Role for MTG8 in the Gut." Molecular and cellular biology **21**(16): 5658-5666.
- Chaffanet, M., L. Gressin, C. Preudhomme, V. Soenen - Cornu, D. Birnbaum and M. J. Pébusque (2000). "MOZ is fused to p300 in an acute monocytic leukemia with t (8; 22)." Genes, Chromosomes and Cancer **28**(2): 138-144.
- Challen, G. A. and M. A. Goodell (2010). "Runx1 isoforms show differential expression patterns during hematopoietic development but have similar functional effects in adult hematopoietic stem cells." Experimental hematology **38**(5): 403-416.
- Choy, J. S., S. Wei, J. Y. Lee, S. Tan, S. Chu and T.-H. Lee (2010). "DNA methylation increases nucleosome compaction and rigidity." Journal of the American Chemical Society **132**(6): 1782-1783.
- Churpek, J. E., J. S. Garcia, J. Madzo, S. A. Jackson, K. Onel and L. A. Godley (2010). "Identification and molecular characterization of a novel 3' mutation in RUNX1 in a family with familial platelet disorder." Leukemia & lymphoma **51**(10): 1931-1935.
- Davis, J. N., L. McGhee and S. Meyers (2003). "The ETO (MTG8) gene family." Gene **303**: 1-10.
- De Braekeleer, E., N. Douet-Guilbert, F. Morel, M.-J. Le Bris, C. Férec and M. De Braekeleer (2011). "RUNX1 translocations and fusion genes in malignant hemopathies." Future Oncology **7**(1): 77-91.
- De Braekeleer, E., C. Férec and M. De Braekeleer (2009). "RUNX1 translocations in malignant hemopathies." Anticancer research **29**(4): 1031-1037.
- de Bruijn, M. and E. Dzierzak (2017). "Runx transcription factors in the development and function of the definitive hematopoietic system." Blood **129**(15): 2061-2069.

Diffner, E., D. Beck, E. Gudgin, J. A. Thoms, K. Knezevic, C. Pridans, S. Foster, D. Goode, W. K. Lim and L. Boelen (2013). "Activity of a heptad of transcription factors is associated with stem cell programs and clinical outcome in acute myeloid leukemia." Blood **121**(12): 2289-2300.

Dixon, J. R., S. Selvaraj, F. Yue, A. Kim, Y. Li, Y. Shen, M. Hu, J. S. Liu and B. Ren (2012). "Topological Domains in Mammalian Genomes Identified by Analysis of Chromatin Interactions." Nature **485**(7398): 376-380.

Drazic, A., L. M. Myklebust, R. Ree and T. Arnesen (2016). "The world of protein acetylation." Biochimica et Biophysica Acta (BBA)-Proteins and Proteomics **1864**(10): 1372-1401.

Durst, K. L. and S. W. Hiebert (2004). "Role of RUNX family members in transcriptional repression and gene silencing." Oncogene **23**(24): 4220.

Eckner, R., M. E. Ewen, D. Newsome, M. Gerdes, J. A. DeCaprio, J. B. Lawrence and D. M. Livingston (1994). "Molecular cloning and functional analysis of the adenovirus E1A-associated 300-kD protein (p300) reveals a protein with properties of a transcriptional adaptor." Genes & development **8**(8): 869-884.

ENCODE Project Consortium (2007). "Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project." Nature **447**(7146): 799.

Estey, E. (2013). "Epigenetics in clinical practice: the examples of azacitidine and decitabine in myelodysplasia and acute myeloid leukemia." Leukemia **27**(9): 1803.

Friedman, A. D. (2009). "Cell cycle and developmental control of hematopoiesis by Runx1." Journal of cellular physiology **219**(3): 520-524.

Fullwood, M. J., Y. Han, C. L. Wei, X. Ruan and Y. Ruan (2010). "Chromatin interaction analysis using paired - end tag sequencing." Current protocols in molecular biology **89**(1): 21.15. 21-21.15. 25.

Fullwood, M. J., M. H. Liu, Y. F. Pan, J. Liu, H. Xu, Y. B. Mohamed, Y. L. Orlov, S. Velkov, A. Ho and P. H. Mei (2009). "An oestrogen-receptor- α -bound human chromatin interactome." Nature **462**(7269): 58.

Gaidzik, V. I., L. Bullinger, R. F. Schlenk, A. S. Zimmermann, J. Röck, P. Paschka, A. Corbacioglu, J. Krauter, B. Schlegelberger and A. Ganser (2011). "RUNX1 mutations in acute myeloid leukemia: results from a comprehensive genetic and clinical analysis from the AML study group." Journal of Clinical Oncology **29**(10): 1364-1372.

Gamou, T., E. Kitamura, F. Hosoda, K. Shimizu, K. Shinohara, Y. Hayashi, T. Nagase, Y. Yokoyama and M. Ohki (1998). "The partner gene of AML1 in t (16; 21) myeloid malignancies is a novel member of the MTG8 (ETO) family." Blood **91**(11): 4028-4037.

Gardner, K. E., C. D. Allis and B. D. Strahl (2011). "Operating on chromatin, a colorful language where context matters." Journal of molecular biology **409**(1): 36-46.

Gelmetti, V., J. Zhang, M. Fanelli, S. Minucci, P. G. Pelicci and M. A. Lazar (1998). "Aberrant recruitment of the nuclear receptor corepressor-histone deacetylase complex by the acute myeloid leukemia fusion partner ETO." Molecular and cellular biology **18**(12): 7185-7191.

Gilliland, D. G. (2002). Molecular genetics of human leukemias: new insights into therapy. Seminars in hematology, Elsevier.

Growney, J. D., H. Shigematsu, Z. Li, B. H. Lee, J. Adelsperger, R. Rowan, D. P. Curley, J. L. Kutok, K. Akashi and I. R. Williams (2005). "Loss of Runx1 perturbs adult hematopoiesis and is associated with a myeloproliferative phenotype." Blood **106**(2): 494-504.

Gu, T.-L., T. L. Goetz, B. J. Graves and N. A. Speck (2000). "Auto-inhibition and partner proteins, core-binding factor β (CBF β) and Ets-1, modulate DNA binding by CBF α 2 (AML1)." Molecular and cellular biology **20**(1): 91-103.

Guo, J., T. Li, J. Schipper, K. A. Nilson, F. K. Fordjour, J. J. Cooper, R. Gordân and D. H. Price (2014). "Sequence specificity incompletely defines the genome-wide occupancy of Myc." Genome biology **15**(10): 482.

Hackanson, B. and M. Daskalakis (2014). Decitabine. Small Molecules in Oncology, Springer: 269-297.

Haga, R. B. and A. J. Ridley (2016). "Rho GTPases: Regulation and roles in cancer cell biology." Small GTPases **7**(4): 207-221.

Hammar, P., P. Leroy, A. Mahmutovic, E. G. Marklund, O. G. Berg and J. Elf (2012). "The lac repressor displays facilitated diffusion in living cells." Science **336**(6088): 1595-1598.

Han, J., Z. Zhang and K. Wang (2018). "3C and 3C-based techniques: the powerful tools for spatial genome organization deciphering." Molecular Cytogenetics **11**(1): 21.

Hatzis, P. and I. Talianidis (2002). "Dynamics of enhancer-promoter communication during differentiation-induced gene activation." Molecular cell **10**(6): 1467-1477.

Heinz, S., C. Benner, N. Spann, E. Bertolino, Y. C. Lin, P. Laslo, J. X. Cheng, C. Murre, H. Singh and C. K. Glass (2010). "Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities." Molecular cell **38**(4): 576-589.

Hoffman, B. G. and S. J. Jones (2009). "Genome-wide identification of DNA-protein interactions using chromatin immunoprecipitation coupled with flow cell sequencing." Journal of Endocrinology **201**(1): 1.

Hollenhorst, P. C., K. J. Chandler, R. L. Poulsen, W. E. Johnson, N. A. Speck and B. J. Graves (2009). "DNA specificity determinants associate with distinct transcription factor functions." PLoS genetics **5**(12): e1000778.

Holmqvist, P.-H. and M. Mannervik (2013). "Genomic occupancy of the transcriptional co-activators p300 and CBP." Transcription **4**(1): 18-23.

Hörlein, A. J., A. M. Näär, T. Heinzl, J. Torchia, B. Gloss, R. Kurokawa, A. Ryan, Y. Kamei, M. Söderström and C. K. Glass (1995). "Ligand-independent repression by the thyroid hormone receptor mediated by a nuclear receptor co-repressor." Nature **377**(6548): 397.

Huang, H., M. Yu, T. E. Akie, T. B. Moran, A. J. Woo, N. Tu, Z. Waldon, Y. Y. Lin, H. Steen and A. B. Cantor (2009). "Differentiation-dependent interactions between RUNX-1 and FLI-1 during megakaryocyte development." Molecular and cellular biology **29**(15): 4103-4115.

Hyde, R. K., P. Liu and A. D. Friedman (2017). RUNX1 and CBF β mutations and activities of their wild-type alleles in AML. RUNX Proteins in Development and Cancer, Springer: 265-282.

Ichikawa, M., A. Yoshimi, M. Nakagawa, N. Nishimoto, N. Watanabe-Okochi and M. Kurokawa (2013). "A role for RUNX1 in hematopoiesis and myeloid leukemia." International journal of hematology **97**(6): 726-734.

Imai, Y., M. Kurokawa, K. Izutsu, A. Hangaishi, K. Takeuchi, K. Maki, S. Ogawa, S. Chiba, K. Mitani and H. Hirai (2000). "Mutations of the AML1 gene in myelodysplastic syndrome and their functional implications in leukemogenesis." Blood **96**(9): 3154-3160.

Inada, M., T. Yasui, S. Nomura, S. Miyake, K. Deguchi, M. Himeno, M. Sato, H. Yamagiwa, T. Kimura and N. Yasui (1999). "Maturational disturbance of chondrocytes in Cbfa1 - deficient mice." Developmental dynamics: an official publication of the American Association of Anatomists **214**(4): 279-290.

Jacob, B., M. Osato, N. Yamashita, C. Q. Wang, I. Taniuchi, D. R. Littman, N. Asou and Y. Ito (2010). "Stem cell exhaustion due to Runx1 deficiency is prevented by Evi5 activation in leukemogenesis." Blood **115**(8): 1610-1620.

Jalili, M., M. Yaghmaie, M. Ahmadvand, K. Alimoghaddam, S. A. Mousavi, M. Vaezi and A. Ghavamzadeh (2018). "Prognostic Value of RUNX1 Mutations in AML: A Meta-Analysis." Asian Pacific journal of cancer prevention: APJCP **19**(2): 325.

Jin, F., Y. Li, B. Ren and R. Natarajan (2011). "Enhancers: multi-dimensional signal integrators." Transcription **2**(5): 226-230.

Jones, P. A. (2012). "Functions of DNA methylation: islands, start sites, gene bodies and beyond." Nature Reviews Genetics **13**(7): 484.

Joshi, O., S.-Y. Wang, T. Kuznetsova, Y. Atlasi, T. Peng, P. J. Fabre, E. Habibi, J. Shaik, S. Saeed and L. Handoko (2015). "Dynamic reorganization of extremely long-range promoter-promoter interactions between two states of pluripotency." Cell stem cell **17**(6): 748-757.

Kadonaga, J. T. (2004). "Regulation of RNA polymerase II transcription by sequence-specific DNA binding factors." Cell **116**(2): 247-257.

Kagoshima, H., K. Shigesada, M. Satake, Y. Ito, H. Miyoshi, M. Ohki, M. Pepling and P. Gergen (1993). "The runt domain identifies a new family of heteromeric transcriptional regulators." Trends in Genetics **9**(10): 338-341.

Kamachi, Y., E. Ogawa, M. Asano, S. Ishida, Y. Murakami, M. Satake, Y. Ito and K. Shigesada (1990). "Purification of a mouse nuclear factor that binds to both the A and B cores of the polyomavirus enhancer." Journal of virology **64**(10): 4808-4819.

Kandpal, R. (2006). "Rho GTPase activating proteins in cancer phenotypes." Current Protein and Peptide Science **7**(4): 355-365.

Kasahara, K., M. Shiina, I. Fukuda, K. Ogata and H. Nakamura (2017). "Molecular mechanisms of cooperative binding of transcription factors Runx1–CBF β –Ets1 on the TCR α gene enhancer." PloS one **12**(2): e0172654.

Kassabov, S. R., B. Zhang, J. Persinger and B. Bartholomew (2003). "SWI/SNF unwraps, slides, and rewaps the nucleosome." Molecular cell **11**(2): 391-403.

Kim, J. H., S. Lee, J. K. Rho and S. Y. Choe (1999). "AML1, the target of chromosomal rearrangements in human leukemia, regulates the expression of human complement receptor type 1 (CR1) gene." The international journal of biochemistry & cell biology **31**(9): 933-940.

Kimura, H. (2013). "Histone modifications for human epigenome analysis." Journal of human genetics **58**(7): 439.

Kolovos, P., T. Georgomanolis, A. Koeferle, J. D. Larkin, L. Brant, M. Nikolicć, E. G. Gusmao, A. Zirkel, T. A. Knoch and W. F. van Ijcken (2016). "Binding of nuclear factor κ B to noncanonical consensus sites reveals its multimodal role during the early inflammatory response." Genome research **26**(11): 1478-1489.

Komori, T., H. Yagi, S. Nomura, A. Yamaguchi, K. Sasaki, K. Deguchi, Y. Shimizu, R. Bronson, Y.-H. Gao and M. Inada (1997). "Targeted disruption of Cbfa1 results in a complete lack of bone formation owing to maturational arrest of osteoblasts." cell **89**(5): 755-764.

Krivega, I. and A. Dean (2012). "Enhancer and promoter interactions—long distance calls." Current opinion in genetics & development **22**(2): 79-85.

Kwok, C., B. B. Zeisig, J. Qiu, S. Dong and C. W. E. So (2009). "Transforming activity of AML1-ETO is independent of CBF β and ETO interaction but requires formation of homo-oligomeric complexes." Proceedings of the National Academy of Sciences **106**(8): 2853-2858.

Lam, K. and D.-E. Zhang (2012). "RUNX1 and RUNX1-ETO: roles in hematopoiesis and leukemogenesis." Frontiers in bioscience: a journal and virtual library **17**: 1120.

Levanon, D., D. Bettoun, C. Harris - Cerruti, E. Woolf, V. Negreanu, R. Eilam, Y. Bernstein, D. Goldenberg, C. Xiao and M. Fliegau (2002). "The Runx3 transcription factor regulates development and survival of TrkC dorsal root ganglia neurons." The EMBO journal **21**(13): 3454-3463.

Levine, M. and R. Tjian (2003). "Transcription regulation and animal diversity." Nature **424**(6945): 147.

Levo, M. and E. Segal (2014). "In pursuit of design principles of regulatory sequences." Nature Reviews Genetics **15**(7): 453-468.

Levy, D., Y. Adamovich, N. Reuven and Y. Shaul (2008). "Yap1 phosphorylation by c-Abl is a critical step in selective activation of proapoptotic genes in response to DNA damage." Molecular cell **29**(3): 350-361.

Levy, D., N. Reuven and Y. Shaul (2008). "A regulatory circuit controlling Itch-mediated p73 degradation by Runx." Journal of Biological Chemistry **283**(41): 27462-27468.

- Li, G., X. Ruan, R. K. Auerbach, K. S. Sandhu, M. Zheng, P. Wang, H. M. Poh, Y. Goh, J. Lim and J. Zhang (2012). "Extensive promoter-centered chromatin interactions provide a topological basis for transcription regulation." Cell **148**(1-2): 84-98.
- Li, Y., H. Luo, T. Liu, E. Zacksenhaus and Y. Ben-David (2015). "The ets transcription factor Fli-1 in development, cancer and disease." Oncogene **34**(16): 2022.
- Liang, J., L. Prouty, B. J. Williams, M. A. Dayton and K. L. Blanchard (1998). "Acute mixed lineage leukemia with an inv (8)(p11q13) resulting in fusion of the genes for MOZ and TIF2." Blood **92**(6): 2118-2122.
- Lin, S., J. C. Mulloy and S. Goyama (2017). RUNX1-ETO Leukemia. RUNX proteins in development and cancer, Springer: 151-173.
- Little, G. H., H. Noushmehr, S. K. Baniwal, B. P. Berman, G. A. Coetzee and B. Frenkel (2011). "Genome-wide Runx2 occupancy in prostate cancer cells suggests a role in regulating secretion." Nucleic acids research **40**(8): 3538-3547.
- Liu, Y., M. D. Cheney, J. J. Gaudet, M. Chruszcz, S. M. Lukasik, D. Sugiyama, J. Lary, J. Cole, Z. Dauter and W. Minor (2006). "The tetramer structure of the Nrvy homology two domain, NHR2, is critical for AML1/ETO's activity." Cancer cell **9**(4): 249-260.
- Lorenzin, F., U. Benary, A. Baluapuri, S. Walz, L. A. Jung, B. von Eyss, C. Kisker, J. Wolf, M. Eilers and E. Wolf (2016). "Different promoter affinities account for specificity in MYC-dependent gene regulation." Elife **5**: e15161.
- Lutterbach, B., J. J. Westendorf, B. Linggi, S. Isaac, E. Seto and S. W. Hiebert (2000). "A mechanism of repression by acute myeloid leukemia-1, the target of multiple chromosomal translocations in acute leukemia." Journal of Biological Chemistry **275**(1): 651-656.
- Meyers, S., J. Downing and S. Hiebert (1993). "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 **13**(10): 6336-6345.
- Meyers, S. and S. W. Hiebert (1995). "Indirect and direct disruption of transcriptional regulation in cancer: E2F and AML-1." Critical Reviews™ in Eukaryotic Gene Expression **5**(3-4).
- Michaud, J., F. Wu, M. Osato, G. M. Cottles, M. Yanagida, N. Asou, K. Shigesada, Y. Ito, K. F. Benson and W. H. Raskind (2002). "In vitro analyses of known and novel RUNX1/AML1 mutations in dominant familial platelet disorder with predisposition to acute myelogenous leukemia: implications for mechanisms of pathogenesis." Blood **99**(4): 1364-1372.
- Mitani, K. (2004). "Molecular mechanisms of leukemogenesis by AML1/EVI-1." Oncogene **23**(24): 4263.
- Miyoshi, H., M. Ohira, K. Shimizu, K. Mitani, H. Hirai, T. Imai, K. Yokoyama, E. Soceda and M. Ohki (1995). "Alternative splicing and genomic structure of the AML1 gene involved in acute myeloid leukemia." Nucleic acids research **23**(14): 2762-2769.

- Miyoshi, H., K. Shimizu, T. Kozu, N. Maseki, Y. Kaneko and M. Ohki (1991). "t (8; 21) breakpoints on chromosome 21 in acute myeloid leukemia are clustered within a limited region of a single gene, AML1." Proceedings of the National Academy of Sciences **88**(23): 10431-10434.
- Mohrs, M., C. M. Blankespoor, Z.-E. Wang, G. G. Loots, V. Afzal, H. Hadeiba, K. Shinkai, E. M. Rubin and R. M. Locksley (2001). "Deletion of a coordinate regulator of type 2 cytokine expression in mice." Nature immunology **2**(9): 842.
- Morgunova, E. and J. Taipale (2017). "Structural perspective of cooperative transcription factor binding." Current opinion in structural biology **47**: 1-8.
- Naoe, T. and H. Kiyoi (2013). "Gene mutations of acute myeloid leukemia in the genome era." International journal of hematology **97**(2): 165-174.
- Nucifora, G., C. R. Begy, P. Erickson, H. A. Drabkin and J. D. Rowley (1993). "The 3; 21 translocation in myelodysplasia results in a fusion transcript between the AML1 gene and the gene for EAP, a highly conserved protein associated with the Epstein-Barr virus small RNA EBER 1." Proceedings of the National Academy of Sciences **90**(16): 7784-7788.
- Okuda, T., J. Van Deursen, S. W. Hiebert, G. Grosveld and J. R. Downing (1996). "AML1, the target of multiple chromosomal translocations in human leukemia, is essential for normal fetal liver hematopoiesis." Cell **84**(2): 321-330.
- Osato, M., N. Asou, E. Abdalla, K. Hoshino, H. Yamasaki, T. Okubo, H. Suzushima, K. Takatsuki, T. Kanno and K. Shigesada (1999). "Biallelic and Heterozygous Point Mutations in the Runt Domain of the AML1/PEBP2 α B Gene Associated With Myeloblastic Leukemias." Blood **93**(6): 1817-1824.
- Otto, F., M. Lübbert and M. Stock (2003). "Upstream and downstream targets of RUNX proteins." Journal of cellular biochemistry **89**(1): 9-18.
- Otto, F., A. P. Thornell, T. Crompton, A. Denzel, K. C. Gilmour, I. R. Rosewell, G. W. Stamp, R. S. Beddington, S. Mundlos and B. R. Olsen (1997). "Cbfa1, a candidate gene for cleidocranial dysplasia syndrome, is essential for osteoblast differentiation and bone development." Cell **89**(5): 765-771.
- Palis, J., S. Robertson, M. Kennedy, C. Wall and G. Keller (1999). "Development of erythroid and myeloid progenitors in the yolk sac and embryo proper of the mouse." Development **126**(22): 5073-5084.
- Panagopoulos, I., T. Fioretos, M. Isaksson, U. Samuelsson, R. Billström, B. Strömbeck, F. Mitelman and B. Johansson (2001). "Fusion of the MORF and CBP genes in acute myeloid leukemia with the t (10; 16)(q22; p13)." Human Molecular Genetics **10**(4): 395-404.
- Pennacchio, L. A., W. Bickmore, A. Dean, M. A. Nobrega and G. Bejerano (2013). "Enhancers: five essential questions." Nature Reviews Genetics **14**(4): 288.
- Petrovick, M. S., S. W. Hiebert, A. D. Friedman, C. J. Hetherington, D. G. Tenen and D.-E. Zhang (1998). "Multiple functional domains of AML1: PU. 1 and C/EBP α synergize with different regions of AML1." Molecular and cellular biology **18**(7): 3915-3925.

- Phillips, J. L., P. C. Taberlay, A. M. Woodworth, K. Hardy, K. H. Brettingham - Moore, J. L. Dickinson and A. F. Holloway (2018). "Distinct mechanisms of regulation of the ITGA6 and ITGB4 genes by RUNX1 in myeloid cells." Journal of cellular physiology **233**(4): 3439-3453.
- Prange, K. H., A. A. Singh and J. H. Martens (2014). "The genome-wide molecular signature of transcription factors in leukemia." Experimental hematology **42**(8): 637-650.
- Preudhomme, C., A. Renneville, V. Bourdon, N. Philippe, C. Roche-Lestienne, N. Boissel, N. Dhedin, J.-M. André, P. Cornillet-Lefebvre and A. Baruchel (2009). "High frequency of RUNX1 biallelic alteration in acute myeloid leukemia secondary to familial platelet disorder." Blood **113**(22): 5583-5587.
- Preudhomme, C., D. Warot-Loze, C. Roumier, N. Grardel-Duflos, R. Garand, J. L. Lai, N. Dastugue, E. Macintyre, C. Denis and F. Bauters (2000). "High incidence of biallelic point mutations in the Runt domain of the AML1/PEBP2 α B gene in Mo acute myeloid leukemia and in myeloid malignancies with acquired trisomy 21." Blood **96**(8): 2862-2869.
- Ptasinska, A., S. A. Assi, N. Martinez-Soria, M. R. Imperato, J. Piper, P. Cauchy, A. Pickin, S. R. James, M. Hoogenkamp and D. Williamson (2014). "Identification of a dynamic core transcriptional network in t (8; 21) AML that regulates differentiation block and self-renewal." Cell reports **8**(6): 1974-1988.
- Ram, O., A. Goren, I. Amit, N. Shores, N. Yosef, J. Ernst, M. Kellis, M. Gymrek, R. Issner and M. Coyne (2011). "Combinatorial patterning of chromatin regulators uncovered by genome-wide location analysis in human cells." Cell **147**(7): 1628-1639.
- Reed-Inderbitzin, E., I. Moreno-Miralles, S. Vanden-Eynden, J. Xie, B. Lutterbach, K. Durst-Goodwin, K. Luce, B. Irvin, M. Cleary and S. Brandt (2006). "RUNX1 associates with histone deacetylases and SUV39H1 to repress transcription." Oncogene **25**(42): 5777.
- Rivera-Reyes, R., M.-J. Kleppa and A. Kispert (2018). "Proteomic analysis identifies transcriptional cofactors and homeobox transcription factors as TBX18 binding proteins." PloS one **13**(8): e0200964.
- Romana, S., M. Mauchauffe, M. Le Coniat, I. Chumakov, D. Le Paslier, R. Berger and O. Bernard (1995). "The t (12; 21) of acute lymphoblastic leukemia results in a tel-AML1 gene fusion." Blood **85**(12): 3662-3670.
- Rosales-Rodríguez, B., F. Fernández-Ramírez, J. C. Núñez-Enríquez, A. C. Velázquez-Wong, A. Medina-Sansón, E. Jiménez-Hernández, J. Flores-Lujano, J. G. Peñaloza-González, R. M. Espinosa-Elizondo and M. L. Pérez-Saldívar (2016). "Copy Number Alterations Associated with Acute Lymphoblastic Leukemia in Mexican Children. A report from The Mexican Inter-Institutional Group for the identification of the causes of childhood leukemia." Archives of medical research **47**(8): 706-711.
- Rowley, J. D. (1973). Identificaton of a translocation with quinacrine fluorescence in a patient with acute leukemia. Annales de genetique.
- Sacchi, N., P. E. Nilsson, P. C. Watkins, F. Faustinella, J. Wijsman and A. Hagemeijer (1994). "AML1 fusion transcripts in t (3; 21) positive leukemia: Evidence of molecular heterogeneity and usage

of splicing sites frequently involved in the generation of normal AML1 transcripts." Genes, Chromosomes and Cancer **11**(4): 226-236.

Samee, M. A. H., B. G. Bruneau and K. S. Pollard (2019). "A de novo shape motif discovery algorithm reveals preferences of transcription factors for DNA shape beyond sequence motifs." Cell systems **8**(1): 27-42. e26.

Sanda, T., L. N. Lawton, M. I. Barrasa, Z. P. Fan, H. Kohlhammer, A. Gutierrez, W. Ma, J. Tatarek, Y. Ahn and M. A. Kelliher (2012). "Core transcriptional regulatory circuit controlled by the TAL1 complex in human T cell acute lymphoblastic leukemia." Cancer cell **22**(2): 209-221.

Sanders, D. A., M. V. Gormally, G. Marsico, D. Beraldi, D. Tannahill and S. Balasubramanian (2015). "FOXN1 binds directly to non-consensus sequences in the human genome." Genome biology **16**(1): 130.

Sauer, T., M. F. Arteaga, F. Isken, C. Rohde, K. Hebestreit, J.-H. Mikesch, M. Stelljes, C. Cui, F. Zhou and S. Göllner (2015). "MYST2 acetyltransferase expression and Histone H4 Lysine acetylation are suppressed in AML." Experimental hematology **43**(9): 794-802. e794.

Schnittger, S., F. Dicker, W. Kern, N. Wendland, J. Sundermann, T. Alpermann, C. Haferlach and T. Haferlach (2011). "RUNX1 mutations are frequent in de novo AML with noncomplex karyotype and confer an unfavorable prognosis." Blood **117**(8): 2348-2357.

Schoenfelder, S. and P. Fraser (2019). "Long-range enhancer–promoter contacts in gene expression control." Nature Reviews Genetics: 1.

Schoenfelder, S., B.-M. Javierre, M. Furlan-Magaril, S. W. Wingett and P. Fraser (2018). "Promoter capture Hi-C: high-resolution, genome-wide profiling of promoter interactions." JoVE (Journal of Visualized Experiments)(136): e57320.

Shia, W.-J., A. J. Okumura, M. Yan, A. Sarkeshik, M.-C. Lo, S. Matsuura, Y. Komeno, X. Zhao, S. D. Nimer and J. R. Yates (2012). "PRMT1 interacts with AML1-ETO to promote its transcriptional activation and progenitor cell proliferative potential." Blood **119**(21): 4953-4962.

Snetkova, V. and J. A. Skok (2018). "Enhancer talk." Epigenomics **10**(4): 483-498.

Song, W.-J., M. G. Sullivan, R. D. Legare, S. Hutchings, X. Tan, D. Kufrin, J. Ratajczak, I. C. Resende, C. Haworth and R. Hock (1999). "Haploinsufficiency of CBFA2 causes familial thrombocytopenia with propensity to develop acute myelogenous leukaemia." Nature genetics **23**(2): 166.

Sood, R., Y. Kamikubo and P. Liu (2017). "Role of RUNX1 in hematological malignancies." Blood **129**(15): 2070-2082.

Sood, R., A. Talwar-Trikha, S. Chakrabarti and G. Nucifora (1999). "MDS1/EVI1 enhances TGF- β 1 signaling and strengthens its growth-inhibitory effect, but the leukemia-associated fusion protein AML1/MDS1/EVI1, product of the t (3; 21), abrogates growth-inhibition in response to TGF- β 1." Leukemia **13**(3): 348.

Speck, N. A. and D. Baltimore (1987). "Six distinct nuclear factors interact with the 75-base-pair repeat of the Moloney murine leukemia virus enhancer." Molecular and Cellular Biology **7**(3): 1101-1110.

Stevens-Kroef, M. P., E. Schoenmakers, M. van Kraaij, E. Huys, S. Vermeulen, B. van der Reijden and A. G. van Kessel (2006). "Identification of truncated RUNX1 and RUNX1-PRDM16 fusion transcripts in a case of t (1; 21)(p36; q22)-positive therapy-related AML." Leukemia **20**(6): 1187.

Sun, W. and J. R. Downing (2004). "Haploinsufficiency of AML1 results in a decrease in the number of LTR-HSCs while simultaneously inducing an increase in more mature progenitors." Blood **104**(12): 3565-3572.

Sun, X.-J., Z. Wang, L. Wang, Y. Jiang, N. Kost, T. D. Soong, W.-Y. Chen, Z. Tang, T. Nakadai and O. Elemento (2013). "A stable transcription factor complex nucleated by oligomeric AML1-ETO controls leukaemogenesis." Nature **500**(7460): 93.

Tahirov, T. H., T. Inoue-Bungo, H. Morii, A. Fujikawa, M. Sasaki, K. Kimura, M. Shiina, K. Sato, T. Kumasaka and M. Yamamoto (2001). "Structural analyses of DNA recognition by the AML1/Runx-1 Runt domain and its allosteric control by CBF β ." Cell **104**(5): 755-767.

Takakura, N., T. Watanabe, S. Suenobu, Y. Yamada, T. Noda, Y. Ito, M. Satake and T. Suda (2000). "A role for hematopoietic stem cells in promoting angiogenesis." Cell **102**(2): 199-209.

Tanaka, K., G. Oshikawa, H. Akiyama, S. Ishida, T. Nagao, M. Yamamoto and O. Miura (2017). "Acute myeloid leukemia with t (3; 21)(q26. 2; q22) developing following low-dose methotrexate therapy for rheumatoid arthritis and expressing two AML1/MDS1/EVI1 fusion proteins: A case report." Oncology letters **14**(1): 97-102.

Tanaka, T., M. Kurokawa, K. Ueki, K. Tanaka, Y. Imai, K. Mitani, K. Okazaki, N. Sagata, Y. Yazaki and Y. Shibata (1996). "The extracellular signal-regulated kinase pathway phosphorylates AML1, an acute myeloid leukemia gene product, and potentially regulates its transactivation ability." Molecular and cellular biology **16**(7): 3967-3979.

Tanaka, Y., A. Joshi, N. K. Wilson, S. Kinston, S. Nishikawa and B. Göttgens (2012). "The transcriptional programme controlled by Runx1 during early embryonic blood development." Developmental biology **366**(2): 404-419.

Tang, J.-L., H.-A. Hou, C.-Y. Chen, C.-Y. Liu, W.-C. Chou, M.-H. Tseng, C.-F. Huang, F.-Y. Lee, M.-C. Liu and M. Yao (2009). "AML1/RUNX1 mutations in 470 adult patients with de novo acute myeloid leukemia: prognostic implication and interaction with other gene alterations." Blood **114**(26): 5352-5361.

Taniuchi, I., M. Osato, T. Egawa, M. J. Sunshine, S.-C. Bae, T. Komori, Y. Ito and D. R. Littman (2002). "Differential requirements for Runx proteins in CD4 repression and epigenetic silencing during T lymphocyte development." Cell **111**(5): 621-633.

Trombly, D. J., T. W. Whitfield, S. Padmanabhan, J. A. Gordon, J. B. Lian, A. J. Van Wijnen, S. K. Zaidi, J. L. Stein and G. S. Stein (2015). "Genome-wide co-occupancy of AML1-ETO and N-CoR defines the t (8; 21) AML signature in leukemic cells." BMC genomics **16**(1): 309.

Tsuzuki, S., D. Hong, R. Gupta, K. Matsuo, M. Seto and T. Enver (2007). "Isoform-specific potentiation of stem and progenitor cell engraftment by AML1/RUNX1." PLoS medicine **4**(5): e172.

- Uchida, H., J. Zhang and S. D. Nimer (1997). "AML1A and AML1B can transactivate the human IL-3 promoter." The Journal of Immunology **158**(5): 2251-2258.
- Wang, J., T. Hoshino, R. L. Redner, S. Kajigaya and J. M. Liu (1998). "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 **95**(18): 10860-10865.
- Wang, L., G. Huang, X. Zhao, M. A. Hatlen, L. Vu, F. Liu and S. D. Nimer (2009). "Post-translational modifications of Runx1 regulate its activity in the cell." Blood Cells, Molecules, and Diseases **43**(1): 30-34.
- Wang, Q., T. Stacy, J. D. Miller, A. F. Lewis, T.-L. Gu, X. Huang, J. H. Bushweller, J.-C. Bories, F. W. Alt and G. Ryan (1996). "The CBF β subunit is essential for CBF α 2 (AML1) function in vivo." Cell **87**(4): 697-708.
- Wek, R., H.-Y. Jiang and T. Anthony (2006). Coping with stress: eIF2 kinases and translational control, Portland Press Ltd.
- Wheeler, J. C., C. VanderZwan, X. Xu, D. Swantek, W. D. Tracey and J. P. Gergen (2002). "Distinct in vivo requirements for establishment versus maintenance of transcriptional repression." Nature genetics **32**(1): 206-210.
- Wilson, N. K., S. D. Foster, X. Wang, K. Knezevic, J. Schütte, P. Kaimakis, P. M. Chilarska, S. Kinston, W. H. Ouwehand and E. Dzierzak (2010). "Combinatorial transcriptional control in blood stem/progenitor cells: genome-wide analysis of ten major transcriptional regulators." Cell stem cell **7**(4): 532-544.
- Wingender, E., X. Chen, E. Fricke, R. Geffers, R. Hehl, I. Liebich, M. Krull, V. Matys, H. Michael and R. Ohnhäuser (2001). "The TRANSFAC system on gene expression regulation." Nucleic acids research **29**(1): 281-283.
- Winklmayr, M., C. Schmid, S. Laner-Plamberger, A. Kaser, F. Aberger, T. Eichberger and A.-M. Frischauf (2010). "Non-consensus GLI binding sites in Hedgehog target gene regulation." BMC molecular biology **11**(1): 2.
- Wood, M., S. Rymarchyk, S. Zheng and Y. Cen (2018). "Trichostatin A inhibits deacetylation of histone H3 and p53 by SIRT6." Archives of biochemistry and biophysics **638**: 8-17.
- Xie, Z., Y. Jia and H. Li (2017). Studying Protein–Protein Interactions by Biotin AP-Tagged Pulldown and LTQ-Orbitrap Mass Spectrometry. Proteomics for Drug Discovery, Springer: 129-138.
- Yagi, R., L. F. Chen, K. Shigesada, Y. Murakami and Y. Ito (1999). "A WW domain - containing yes - associated protein (YAP) is a novel transcriptional co - activator." The EMBO journal **18**(9): 2551-2562.
- Ye, J., G. Coulouris, I. Zaretskaya, I. Cutcutache, S. Rozen and T. L. Madden (2012). "Primer-BLAST: a tool to design target-specific primers for polymerase chain reaction." BMC bioinformatics **13**(1): 134.

Yokomizo, T., K. Hasegawa, H. Ishitobi, M. Osato, M. Ema, Y. Ito, M. Yamamoto and S. Takahashi (2008). "Runx1 is involved in primitive erythropoiesis in the mouse." Blood **111**(8): 4075-4080.

Yoshida, H. and I. Kitabayashi (2008). "Chromatin regulation by AML1 complex." International journal of hematology **87**(1): 19-24.

Zang, C., A. Luyten, J. Chen, X. S. Liu and R. A. Shivdasani (2016). "NF-E2, FLI1 and RUNX1 collaborate at areas of dynamic chromatin to activate transcription in mature mouse megakaryocytes." Scientific reports **6**: 30255.

Zelent, A., M. Greaves and T. Enver (2004). "Role of the TEL-AML1 fusion gene in the molecular pathogenesis of childhood acute lymphoblastic leukaemia." Oncogene **23**(24): 4275.

Zentner, G. E. and P. C. Scacheri (2012). "The chromatin fingerprint of gene enhancer elements." Journal of Biological Chemistry **287**(37): 30888-30896.

Zhang, D.-E., K.-i. Fujioka, C. J. Hetherington, L. H. Shapiro, H.-M. Chen, A. T. Look and D. G. Tenen (1994). "Identification of a region which directs the monocytic activity of the colony-stimulating factor 1 (macrophage colony-stimulating factor) receptor promoter and binds PEBP2/CBF (AML1)." Molecular and Cellular Biology **14**(12): 8085-8095.

Zhang, H., P. Meltzer and S. Davis (2013). "RCircos: an R package for Circos 2D track plots." BMC bioinformatics **14**(1): 244.

Zhao, J.-y., O. Osipovich, O. I. Koues, K. Majumder and E. M. Oltz (2017). "Activation of Mouse Tcrb: Uncoupling RUNX1 Function from Its Cooperative Binding with ETS1." The Journal of Immunology **199**(3): 1131-1141.

8 Appendix: Figure 3.12 Enlarged

