BIOCHEMICAL CHARACTERIZATION AND REGULATION OF TRANSCRIPTION OF POLYCOMB GROUP RING FINGER 5

by

Christopher Larson Cochrane

B.Sc., The University of British Columbia, 2004

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

in

THE FACULTY OF GRADUATE AND POSTDOCTORAL STUDIES

(Experimental Medicine)

THE UNIVERSITY OF BRITISH COLUMBIA

(Vancouver)

August 2013

© Christopher Larson Cochrane, 2013
Abstract

The Polycomb Group (PcG) is a highly conserved group of genes which serve to repress transcription via specific modifications of histones in chromatin. The PcG has well-established roles in development and is involved, by mutation or dysregulation, in many human diseases including cancer. This study identifies the gene PCGF5, which is a paralogue of the oncogene Bmi1, as a transcriptional target of Notch signalling in T cell acute lymphoblastic leukemia (T-ALL). Evidence suggests that this regulation is direct and that the Notch transactivation complex binds DNA at several regions near the PCGF5 gene. PCGF5 is found to be expressed at a higher level in T-ALL than other hematopoietic malignancies. PCGF5 is found to associate with the PcG proteins RING1A and RING1B and its overexpression results in increased ubiquitylation of histone H2A, suggesting it shares functional similarity to Bmi1. Despite their similarities, Bmi1 and PCGF5 have a different spectrum of binding partners and are targeted to different locations in the genome. Overexpression of PCGF5 does not significantly alter hematopoietic development in vivo; however, enforced expression of PCGF5 in bone marrow progenitors results in the generation of fewer colonies in a myeloid colony forming assay. This study suggests that PCGF5 may have as yet unappreciated roles in PcG biology and merits further study into its effects on development and hematopoietic neoplasia.
### Table of Contents

Abstract .................................................................................................................................................. ii

List of Tables ......................................................................................................................................... viii

List of Figures ...................................................................................................................................... ix

Abbreviations ....................................................................................................................................... xiii

Acknowledgments ................................................................................................................................. xx

Dedication ............................................................................................................................................. xxi

Chapter 1 Introduction ............................................................................................................................. 1

1.1 Rationale .......................................................................................................................................... 1

1.2 The Polycomb Group of genes ......................................................................................................... 2

1.2.1 Genetics ...................................................................................................................................... 2

1.2.2 Polycomb genes in mammals ..................................................................................................... 3

1.2.3 Defining the PcG complex .......................................................................................................... 5

1.2.4 Histone modification ................................................................................................................... 9

1.2.5 Targeting and recruitment .......................................................................................................... 11

1.3 Epigenetics and the histone code .................................................................................................... 12

1.3.1 The epigenetic code .................................................................................................................. 12

1.3.2 Histone modifications ............................................................................................................... 13

1.4 The Notch signalling pathway ....................................................................................................... 17

1.4.1 Notch genetics .......................................................................................................................... 17
1.4.2 The Notch transactivation complex ................................................................. 18
1.4.3 Mechanism of signal transduction ................................................................. 18
1.4.4 Notch and T cell Acute Lymphoblastic Leukemia ......................................... 19
1.4.5 Hypotheses ........................................................................................................ 20

Chapter 2  Materials and methods ........................................................................... 23

Chapter 3  Notch signalling regulates expression of PCGF5 in T cell Acute Lymphoblastic Leukemia cells

3.1 Introduction .......................................................................................................... 33
3.2 Results .................................................................................................................. 35
3.2.1 Identification of PCGF5 as a target of Notch .............................................. 35
3.2.2 Notch activation of PCGF5 expression does not require de novo protein translation ...... 42
3.2.3 Notch binds upstream of the PCGF5 locus ................................................. 42
3.2.4 Notch1 and CSL binding of PCGF5 locus is significantly higher than expected by chance ...... 50
3.2.5 PCGF5 expression in T-ALL cases ............................................................... 52
3.2.6 Tissue distribution .......................................................................................... 55
3.2.7 Splice variants and genomic organization .................................................... 58
3.3 Discussion ............................................................................................................. 59
3.3.1 Notch transcriptional regulation of PCGF5 ............................................... 59
3.3.2 Notch and CSL binding of the PCGF5 locus ............................................ 67
3.3.3 Notch expression in cancer ........................................................................... 69
3.3.4 PCGF5 and Notch linked in normal development ....................................... 70
Chapter 4  PCGF5 overexpression does not alter steady state hematopoiesis but results in reduced myeloid colony output by bone marrow progenitors .............................................................................................................. 72

4.1 Introduction .......................................................................................................................................................... 72

4.1.1 The role of Notch in lymphopoiesis .............................................................................................................. 72

4.1.2 Notch and T-cell Acute Lymphoblastic Leukemia ......................................................................................... 73

4.1.3 The role of Bmi1 and the Polycomb Group in lymphopoiesis ....................................................................... 74

4.1.4 Experimental approaches .............................................................................................................................. 76

4.1.5 Hypotheses ...................................................................................................................................................... 77

4.2 Results ............................................................................................................................................................... 77

4.2.1 Overexpression of neither PCGF5 nor Bmi1 causes significant alterations to hematopoiesis in bone marrow transplant recipients .............................................................................................................. 77

4.2.2 PCGF5 does not alter lymphoid progenitor performance in OP9 coculture assay .................................. 89

4.2.3 PCGF5 causes reduced myeloid CFC output in bone marrow progenitors .................................................. 97

4.3 Discussion .......................................................................................................................................................... 101

4.3.1 Comment on Experimental Design .............................................................................................................. 101

4.3.2 PCGF5 overexpression in bone marrow does not alter engraftment or multilineage reconstitution in transplant recipients ...................................................................................................................... 102

4.3.3 Recipients of PCGF5 overexpressing bone marrow have an unaltered distribution of hematopoietic lineages .............................................................................................................................................. 103

4.3.4 PCGF5 overexpressing progenitors show no difference to controls in an in vitro model of lymphoid development ................................................................................................................................. 104
4.3.5 Enforced expression of PCGF5 and Bmi1 reduce clonogenic activity of bone marrow progenitors in a colony forming assay ............................................................................................................. 107

4.3.6 Enforced expression of PCGF5 does not replicate constitutive Notch signalling ........... 108

4.3.7 Physiologic effects of altered gene dose of the Polycomb Group and tissue-specific phenotypes ........................................................................................................................................... 109

4.3.8 Open question: does PCGF5 contribute to leukemogenicity of Notch .......................................... 110

Chapter 5 PCGF5 is a member of the Polycomb Group and a paralogue of Bmi1 ................. 112

5.1 Introduction ................................................................................................................................................................................. 112

5.2 Results ........................................................................................................................................................................................... 114

5.2.1 PCGF5 contains a RING domain with homology to Bmi1 ................................................................. 114

5.2.2 Sequence divergence outside the RING domain of PCGF family members ......................... 117

5.2.3 PCGF5 shares some, but not all, putative post-translational modification sites with Bmi1 ... 119

5.2.4 PCGF5 associates with RING1B in competition with Bmi1 ............................................................... 120

5.2.5 PCGF5 overexpression induces increased total H2AK119 ubiquitylation .............................. 124

5.2.6 Regulation of gene expression by PCGF5 ......................................................................................... 126

5.3 Discussion.................................................................................................................................................................................. 130

5.3.1 Protein structure and interactions ................................................................................................. 131

5.3.2 Evidence for PCGF5-mediated ubiquitylation of H2AK119 ......................................................... 133

5.3.3 Implications of PCGF competition for RING1B interaction ..................................................... 136

5.3.4 PCGF impacts on gene expression ............................................................................................... 137

Chapter 6 Conclusions and perspectives ......................................................................................... 140
6.1 Polycomb complexes: shuffling the deck................................................................. 140

6.2 Transcriptional regulation of the Polycomb Group ............................................. 144

6.3 PCGF5 and maintenance of repression................................................................. 146

6.4 Mechanism of repression...................................................................................... 147

6.5 In vivo future directions......................................................................................... 148

6.6 Conclusion ............................................................................................................. 149

References .................................................................................................................. 151

Appendix ...................................................................................................................... 164

A1 PCGF5 binds chromatin at locations distinct from Bmi1......................................... 164

A2 PCGF5 SERs are associated with increased H2AK119ub when RING1B is also present .......... 168
List of Tables

Table 1.2.1 Mouse and human homologues of *Drosophila* PcG.......................................................... 7

Table 2.1 – Details of public microarray data ......................................................................................... 28

Table 4.2.1 Antibodies used for flow cytometry and the cell types they identify................................. 76

Table 6.1.1 Phenotype of loss-of-function for PcG genes..................................................................... 138
List of Figures

Figure 1.1 – Schematic of the core components and activities of the Polycomb Repressive Complexes 1 and 2 (PRC1 and PRC2) ............................................................................................................. 6

Figure 3.2.1 – Notch inhibition causes changes in gene expression in five T-ALL cell lines .................. 36

Figure 3.2.2 – Regulation of PCGF5 across five T-ALL cell lines ........................................................................ 38

Figure 3.2.3 – Compound E treatment reduces PCGF5 mRNA in human and mouse T-ALL cell lines..... 39

Figure 3.2.4 – Effect of alternative methods of Notch inhibition on PCGF5 expression ......................... 41

Figure 3.2.5 – Notch regulation of PCGF5 does not require de novo translation ................................. 43

Figure 3.2.6 – Exon structure of human PCGF5 transcripts ........................................................................ 44

Figure 3.2.7 – Notch binds near the PCGF5 locus ....................................................................................... 45

Figure 3.2.8 – Notch1 and CSL enrich regions at or near the PCGF5 promoters ................................. 46

Figure 3.2.9 - Notch and CSL bind enhancer upstream of PCGF5 in mouse ............................................. 47

Figure 3.2.10 – Notch1 enriches the 3′ PCGF5 promoter in mouse ............................................................. 48

Figure 3.2.11 – SERs for Notch1 and CSL are abundant near the PCGF5 locus ........................................ 49

Figure 3.2.12 – PCGF5 expression among various cytogenetic subtypes of T-ALL ............................... 51

Figure 3.2.13 – PCGF5 expression does not correlate with disease progression in T-ALL ................. 52

Figure 3.2.14 – PCGF5 expression is higher in T-ALL than in other hematopoietic malignancies .... 53

Figure 3.2.15 – PCGF5 expression in mouse hematopoietic cells .............................................................. 54
Figure 3.2.16 – PCGF5 expression in mouse thymic subsets .......................................................... 55

Figure 3.2.17 – PCGF5 expression in human thymic and peripheral T cell subsets ............................... 56

Figure 3.2.18 – Detection of PCGF5 splice variants in T-ALL cells by RT-PCR ................................. 57

Figure 3.3.1 – Samples cluster by cell line not Notch signalling status ........................................... 59

Figure 3.3.2 – Paired comparison of 5 cell lines treated with Cmpd E or vehicle control ..................... 61

Figure 3.3.3 – Paired comparison of 10 cell lines treated with Cmpd E or vehicle control ................. 63

Figure 4.2.1 – GFP-positive cells are present in the peripheral blood and bone marrow of transplant recipients ........................................................................................................................................ 77

Figure 4.2.2 – PCGF5 and Bmi1 do not alter T cell frequency in the peripheral blood ...................... 78

Figure 4.2.3 – No evidence of ectopic T cell development in bone marrow of PCGF5 or Bmi1 recipient mice ........................................................................................................................................ 79

Figure 4.2.4 – PCGF5 and Bmi1 do not alter B cell frequencies in transplant recipient peripheral blood 80

Figure 4.2.5 – PCGF5 and Bmi1 do not alter B cell frequencies in transplant recipient bone marrow ... 81

Figure 4.2.6 – PCGF5 and Bmi1 do not alter granulocyte frequencies in peripheral blood ............... 82

Figure 4.2.7 – PCGF5 and Bmi1 do not alter granulocyte frequencies in bone marrow ..................... 83

Figure 4.2.8 – Summary of lineage frequencies in the peripheral blood and bone marrow of transplant recipients ........................................................................................................................................ 85

Figure 4.2.9 – OP9-DL1 cells give rise to DN T cells over monocytes and B cells ......................... 87
Figure 4.2.10 – T lymphoid progression of fetal liver cells on OP9-DL1 ........................................... 88

Figure 4.2.11 – No effect of Bmi1 or PCGF5 on DN subsets on OP9-DL1 ........................................... 89

Figure 4.2.12 – DP cell generation on OP9-DL1 ................................................................................. 90

Figure 4.2.13 – PCGF5-expressing DN thymocytes rapidly become DP on OP9-DL1 ......................... 91

Figure 4.2.14 – PCGF5 and Bmi1 do not mimic Notch stimulation...................................................... 92

Figure 4.2.15 – Overexpression of Bmi1 and PCGF5 decrease myeloid colony forming cell frequency. 94

Figure 4.2.16 – Bmi1 and PCGF5 overexpression result in reduced proportion of CD11b bright cells ... 95

Figure 4.2.17 – Bmi1 and PCGF5 do not affect GMP frequency in bone marrow ............................... 96

Figure 5.2.1 – Domain structure schematic of the PCGF family ....................................................... 111

Figure 5.2.2 – Important features of Bmi1’s RING domain are conserved in PCGF5 ....................... 112

Figure 5.2.3 – Homology between PCGF family members ............................................................... 113

Figure 5.2.4 – Relative similarity of PCGF family members ........................................................... 115

Figure 5.2.5 – The sequence around the ubiquitylation site on Lys81 of Bmi1 is conserved in PCGF5 .. 116

Figure 5.2.6 – PCGF5 interacts with RING1B.................................................................................. 118

Figure 5.2.7 – Bmi1 and PCGF5 compete for association with RING1B............................................. 119

Figure 5.2.8 – Bmi1 and PCGF5 overexpression correlate with increased H2AK119ub ....................... 121

Figure 5.2.9 – PCGF5 and Bmi1 modulate expression of a subset of Hox genes in MEFs .................... 124
Figure 5.2.10 – Transduced cells express HA-tagged Bmi1 and PCGF5 protein........................................125

Figure 5.2.11 – PCGF5 has variable effects on Ink4a expression.................................................................126

Figure 6.1.1 – PCGF family members within PRC1 each interact with a different subset of binding partners.............................................................................................................................140

Figure A1 PCGF5 has a different target repertoire than Bmi1.................................................................161

Figure A2 PCGF5 SERs correlate with increased H2AK119ub signal........................................................162

Figure A3 – Sites co-occupied by Bmi1 do not account for all of PCGF5-associated H2AK119ub..........164

Figure A4 – PCGF5 sites co-occupied by RING1B have high H2AK119ub.................................................165

Figure A5 – PCGF5 SERs locations are less biased towards TSSs than Bmi1 and RING1B SERs...........166

Figure A6 Distribution of Bmi1 and PCGF5 SERs on selected Bmi1 target loci in human......................167
Abbreviations

5-FU: 5-fluorouracil

ΔE: Notch allele with deletion of the extracellular portion

ADAM: A Disintegrin And Metalloproteinase

AML: Acute Myelogenous Leukemia

Arg: arginine

Asp: aspartate

ATP: Adenosine triphosphate

B-ALL: B cell Acute Lymphoblastic Leukemia

Bmi1: B cell-specific Maloney murine leukemia virus integration site 1

CADASIL: Cerebral autosomal dominant arteriopathy with subcortical infarcts and leukoencephalopathy

CBX: Chromobox homologue

cDNA: complementary DNA

CFC: Colony forming cell

ChIP: Chromatin Immunoprecipitation

ChIP-seq: Chromatin Immunoprecipitation followed by sequencing

CHX: Cyclohexamide

CMP: Common myeloid progenitors
Cmpd E: γ-secretase inhibitor XXI; Compound E

COMPASS: Complex Proteins Associated with Set1

CSL: protein named for orthologues CBF-1, Su(H), lag-1

Dl1: Delta-like-1

DNA: Deoxyribonucleic acid

DNMAML1: Dominant-negative MamL1

DN: Double negative

DP: Double positive

DTT: Dithiothreitol

Eμ: Immunoglobulin M heavy chain enhancer

E(z): Enhancer of Zeste

EDTA: Ethylene diamine tetraacetic acid

Eed: Embryonic ectoderm development

EGF: Epidermal growth factor

Esc: Extra sex combs

EZH1 and 2: Enhancer of Zeste homologue 1 and 2

FACS: Fluorescence activated cell sorting

FBS: Fetal Bovine Serum
FSC: Forward Scatter

GEO: Gene expression omnibus

GFP: Green fluorescent protein

GMP: Granulocyte monocyte progenitor

Glu: glutamate

GSI: γ-secretase inhibitor

H2A: Histone H2A

H2AK119ub: Histone H2A ubiquitylated on lysine 119

H3K4me1: Histone H3 monomethylated on lysine 4

H3K4me3: Histone H3 trimethylated on lysine 4

H3K27me3: Histone H3 trimethylated on lysine 27

HAT: Histone acetyltransferase

HCl: Hydrochloric acid

HDAC: Histone Deacetylase

HEPES: 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid

His: histidine

Hox: Homeobox

HPH: Human Polyhomeotic
ICN: Intracellular Notch

Ig: Immunoglobulin

IL3 and 6: Interleukin-3 and -6

IRES: Internal ribosome entry site

kb: kilobase

kDa: kilodaltons

Lys: lysine

MACS: Model-based analysis of ChIP-seq

Maml1: Mastermind-like-1

MBLR: Mel-18 and Bmi1-like RING finger protein

M-CSF: Macrophage colony stimulating factor

MDa: Megadaltons

MEF: Mouse embryonic fibroblast

Mel-18: Melanoma nuclear protein 18

MIG: MSCV-IRES-GFP

Mll: Mixed lineage leukemia

mRNA: messenger RNA

MSCV: Mouse stem cell virus
MW: Molecular weight

NCBI: National Center for Biotechnology Information

NK: Natural Killer

NotchΔE: a mutant of Notch which lacks most of the extracellular domain but contains the transmembrane region and is GSI sensitive

NsPc1: Nervous system Polycomb-1

OP: Osteopetrotic

PBS: Phosphate Buffered Saline

Pc: Polycomb

Pcl: Polycomb-like

Ph: Polyhomeotic

Psc: Posterior sex combs

PcG: Polycomb Group

PCGF: Polycomb Group RING Finger

PCR: Polymerase Chain Reaction

PEST: Sequence motif rich in Proline (P), Glutamate (E), Serine (S) and Threonine (T)

PMSF: Phenyl methyl sulfonyl fluoride

PRC1 and 2: Polycomb Repressive Complex 1 and 2
PRE: Polycomb Repressive Element
PVDF: Polyvinylidene fluoride
RAG: Recombination-activating gene
RNA: Ribonucleic acid
RING: Really Interesting New Gene
rh: recombinant human
rm: recombinant mouse
RT-PCR: Reverse Transcription followed by Polymerase Chain Reaction
UDG: Uracil-DNA glycosylase
SER: Significantly enriched region
SET: methyltransferase protein family named for Su(var)3-9, Enhancer of Zeste, Trithorax proteins
SMRT: silencing mediator for retinoid and thyroid hormone receptors
SP: Single positive
SSC: Side scatter
Su(H): Suppressor of hairless
Suz12: Suppressor of Zeste 12
T: Threonine
T-ALL: T cell Acute Lymphoblastic Leukemia
TAN-1: Translocation-associated Notch protein-1

TBS: Tris Buffered Saline

TCA: Trichloroacetate

TCR: T cell receptor

UCSC: University of California, Santa Cruz

Val: valine

Y: Tyrosine

YY1: Yin Yang1
Acknowledgments

Firstly, I would like to extend my gratitude to Dr. Andrew Weng for giving me the opportunity to study in his laboratory and to make and learn from my mistakes. The members of my supervisory committee, Dr. Ninan Abraham, Dr. Keith Humphries, and Dr. Rob Kay provided invaluable criticism and advice during the execution of the work described in this thesis, as well as in the preparation of the dissertation itself. Their time and effort is greatly appreciated. I would like to deeply thank my lab mates Sam Gusscott, Sonya Lam, Arla Yost, Vincenzo Giambra, Christopher Jenkins, and Helen Shevchuck for enriching my graduate experience and for providing lively discussion. I thank Dr. Muriel David for my early technical training which was invaluable throughout my Ph.D. I would like to acknowledge Carol Wai for performing the expression profiling experiment wherein PCGF5 was identified and Dr. Hind Medyouf for supplying the murine leukemia cells used to measure Ink4a/ARF expression. Dr. Maarten Van Lohuizen kindly supplied Bmi1-/- fibroblasts, and Dr. Danny Reinberg generously shared analyzed ChIP-seq data for the PCGF family.
Dedication

I humbly dedicate this thesis to Terra, my brilliant partner and best friend, from whom I derive my inspiration, and to my parents for their unstinting support.
Chapter 1 Introduction

1.1 Rationale

The Notch pathway is involved in many aspects of biology but from a clinical perspective it is of particular importance in the context of T-cell acute lymphoblastic leukemia (T-ALL). Notch mutations are common in T-ALL [5] and experiments in mice support a critical role for Notch in the pathogenesis of this disease [6, 7]. Attempts to target the Notch pathway itself as a means of therapy have so far met with limited success due to the critical role the Notch pathway plays in homeostasis of many other tissues, notably in gut epithelium [8]. Identification of Notch target genes which mediate Notch’s oncogenicity may yield novel therapeutic targets for T-ALL, particularly those which are relatively tissue-specific, that may achieve much of the therapeutic benefit of Notch inhibition with fewer side-effects in other tissues and organs.

Further, study of the function of Notch target genes in normal hematopoiesis will improve our understanding of the mechanism by which Notch mediates development of blood cells, particularly of T lymphocytes, for which an intact Notch1 pathway is critical [9].

This study will focus on the gene Polycomb Group Ring Finger 5 (PCGF5) which was identified as a transcriptional target of Notch. PCGF5 is predicted by sequence homology to comprise part of the Polycomb Group of genes which act as epigenetic repressors of gene transcription. The PCGF5 protein is homologous to Bmi1 which, by mediating ubiquitylation of histones with its binding partner RING1, causes repression of a number of genes including those of the Hox family [10] and the tumour suppressors Ink4a and ARF [11]. It’s repression of Ink4a/ARF, in particular, makes Bmi1 a potent oncogene.
Given that no function or activity has yet been ascribed to PCGF5 in the literature, it was important to investigate its function in comparison to Bmi1. Bmi1 has been described to have various functional effects in hematopoietic development and is required for efficient T cell development. Investigation of PCGF5 in these contexts might reveal whether PCGF5 has a Bmi1-like oncogenic activity, whether it might antagonize Bmi1 by competition for its binding partners, or whether its function is less closely related.

Answering these questions could shed light on how Notch promotes T cell development and may have implications for the treatment of T-ALL.

1.2 The Polycomb Group of genes

1.2.1 Genetics

Early genetic work studying anatomic development of Drosophila identified the homeotic genes as critical for the determination and maintenance of cell fate. Gain or loss of function mutations affecting homeotic genes resulted in developmental abnormalities. One such mutant, referred to as Sex combs reduced, resulted in loss of sex comb bristles on the first leg pair in loss of function, and conversely caused bristle development on all three pairs of legs in gain of function [12, 13], a phenotype called extra sex combs. The extra sex combs phenotype had been previously identified twice in the 1940s when Slifer and P Lewis discovered the Extra sex combs (Esc) and Polycomb (Pc) mutants, respectively [14, 15]. Thus, it was clear that homeotic genes were powerful regulators of cell fate and that their specific and stable expression was required for appropriate development. Left unanswered was the question of how homeotic genes were regulated and how their stable expression patterns were maintained.
A study conducted by Gerd Jurgens group and published in Nature in 1985 sought to identify mutants that led to an extra sex combs trait. In so doing, he discovered what he called the Polycomb Group (PcG) [16] which was represented by the already known Pc and Esc genes, as well as Polycomb-like (Pcl), Sex combs on midleg (Scm), Additional sex combs (Asl), and Posterior sex combs (Psc). Jurgens concluded that these genes represented trans-regulatory elements of the homeotic gene BX-C. Further work identified more members of the PcG including Polyhomeotic (Ph), Enhancer of Zeste (E(z)), Sex combs extra (Sce), super sex combs (Ssc), pleiohomeotic and multi sex combs [17-22].

Prior studies of the Pc mutation also suggested a repressor role for the PcG. Indeed, Pc mutant embryos underwent abnormal segmentation resulting in posterior-type structures occurring anteriorly; however, the type of posterior structures formed and the regions affected depended on the genotype of several homeotic loci [23]. Lewis thus concluded that the Pc locus encoded a repressor that regulated several homeotic genes. This meant that crosses of Pc with different homeotic mutants each gave different morphologic phenotypes due to the derepression of a different repertoire of homeotic genes. Lewis’ hypothesis was subsequently confirmed by molecular studies of homeotic gene expression in PcG mutant flies [24-26]. Thus, it was becoming apparent that Jurgens’ Polycomb Group was a critical regulator of gene expression in the developing fly but from the identification of the Esc and Pc mutants in the 1940s it would be nearly 50 years until the PcG came to the attention of mammalian biologists.

1.2.2 Polycomb genes in mammals

The study of PcG genes in mammals began with a retroviral mutagenesis screen for insertions that would collaborate to cause lymphoma in the Eμ-myc background. One such gene was a locus termed Bmi1 for B cell-specific Mo-MLV integration site 1 which, upon viral insertion, resulted in overexpression of the Bmi1 gene product [27, 28]. Sequence analysis of this novel gene revealed sequence similarity with Polycomb genes, in particular Drosophila Psc [29]. In the course of this study, the authors also
identified another Psc orthologue, Mel-18, whose cDNA was cloned from mouse melanoma cells where it was found to be overexpressed [30].

While the oncogenic effects of Bmi1 overexpression were indeed intriguing, it remained to be seen whether the emerging mammalian PcG had analogous functional effects to its Drosophila counterpart. The first evidence for functional conservation arose from the targeted deletion of Bmi1. Mice derived from Bmi1-/- embryonic stem cells, while viable, displayed hematopoietic and neurological defects characterized by hypocellularity in both cases. Importantly, these mice also suffered from posterior transformation defects along the anteroposterior axis of the skeleton [31]. These defects include an extra vertebral bone rostral of C1, additional ribs from the C7 vertebra, transformation of the T7 vertebra to T8 such that only 6 (instead of 7) ribs connect to the sternum, absence of ribs connecting to T13 and the association of ilial bones with L6 instead of S1. In general, the skeletal transformations were characterized by posterior characteristics shifting anteriorly. This effect on embryonic patterning in the anteroposterior axis was reminiscent of the patterning defects observed in sex comb mutants in flies. As Drosophila Polycomb genes exert their effects on embryonic patterning by regulating expression Hox genes, the authors noted that the skeletal defects of the Bmi1 -/- mice were analogous to Hox mutants (reviewed [32, 33]. Following up from the Bmi1 null study van Lohuizen’s group generated a Bmi1 transgenic mouse driven by the immunoglobulin heavy chain enhancer. These mice displayed transformation of several vertebrae toward the adjacent anterior – exactly the opposite phenotype of the loss of function. The authors also describe a posterior shift in the boundary of HoxC5 expression [34].

Studies of a Mel-18 null mouse similarly revealed that vertebra often adopted features of its posterior neighbour [35], as seen in the Bmi1 null mouse. Akasaka’s study went further and assessed Hox gene expression. They found that expression boundaries of many, though not all, Hox genes were shifted one
segment anteriorly in concordance with the anatomical transformations. They concluded that Mel-18 plays a role in preventing inappropriate expression of Hox genes in the developing embryo in a manner analogous to Psc in Drosophila. Thus, Mel-18 deletion confers a phenotype strikingly similar to that of Bmi1 deletion offering support that the two genes so closely related by sequence are functionally similar as well.

Similar axial transformation and commensurate Hox gene dysregulation have been described in mice deficient for the Pc homologue M33 [36], the Ph homologue PHC1/Rae-28 [37] and RING1A [38].

To determine whether Bmi1 possessed the ability to repress transcription, mammalian cells were transfected with either dPsc or Bmi1 fused to the bacterial LexA protein which would target the fusion protein to a reporter gene via LexA binding sites [39]. These experiments revealed that both Bmi1 and dPsc repressed expression of targeted loci in the mammalian system. Similar studies using LexA fusions to RING1 and Pc demonstrated that these too acted as transcriptional repressors [40].

Thus, the role of the PcG in regulating gene expression during development was clear; however, little was known about the mechanism by which Polycomb proteins exerted their effects.

1.2.3 Defining the PcG complex

The genetic data amassed by the study of Drosophila PcG mutants did little to address whether the products of the various genes identified acted individually or as a complex. In 1992, Franke et al. described a multimeric complex of between 2 and 5 MDa immunoprecipitated with an antibody against Pc [41]. This complex also contained Ph, a result which strongly favoured the idea that PcG proteins acted as a complex.

Affinity purification from insect sf9 cells characterized the complex further, revealing that ectopically expressed Pc, Ph, Psc and Ring co-precipitated stoichiometrically. The authors concluded that these
proteins formed what they termed the Polycomb Repressive Complex-1 (PRC1) core complex [42]. At that time, the *Drosophila Ring* gene had not been genetically implicated as a PcG member, but subsequent studies determined *dRing* was encoded by *sex combs extra (sce)* [43].

Figure 1.1 – Schematic of the core components and activities of the Polycomb Repressive Complexes 1 and 2 (PRC1 and PRC2) Figure adapted from Otte, A.P. and T.H. Kwaks, *Gene repression by Polycomb group protein complexes: a distinct complex for every occasion?* Curr Opin Genet Dev, 2003. 13(5): p. 448-54.

Reconstitution of the PRC1 complex *in vitro* revealed it to be active in suppressing SWI/SNF-mediated nucleosome remodelling in a plasmid supercoiling assay [44], suggesting that Pc, Ph, Psc and Ring formed a core complex involved in transcriptional suppression. Curiously, the genetically linked PcG proteins E(z) and Pcl were found not to associate with the PRC1 complex and were not required for PRC1 activity. This finding, together with the observation that the mammalian homologue of E(z), EZH2 co-precipitated with the homologue of Esc (Eed) [45, 46] suggested that there may be a second, distinct PRC complex.

Subsequently, such a complex was purified from human 293 cells containing the homologues of EZH2, Eed and Suz12 (homologues of *E(z)*, Esc, and *Su(var)12*, respectively. Further, this complex was found to
possess histone methyltransferase activity specific for lysines at position 9 and 27 of histone H3 [47] when oligonucleosomes are provided as a substrate.

<table>
<thead>
<tr>
<th>Drosophila</th>
<th>Mouse</th>
<th>Human</th>
</tr>
</thead>
<tbody>
<tr>
<td>PRC1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pc</td>
<td>M33(CBX2), Pc2(CBX4), Pc3(CBX8)</td>
<td>M33(CBX2), PC2(CBX4), PC3(CBX8)</td>
</tr>
<tr>
<td>Ph</td>
<td>Rae-28</td>
<td>PHC1(HPH1,RAE28), PHC2(HPH2), PHC3(HPH3)</td>
</tr>
<tr>
<td>Psc</td>
<td>Bmi1, Mel-18, MBLR, Pcgf1, Pcgf3, Pcgf5</td>
<td>BMI1, MEL-18, MBLR, PCGF1, PCGF3, PCGF5</td>
</tr>
<tr>
<td>Ring(sce)</td>
<td>Ring1a, Ring1b</td>
<td>RING1A, RING1B</td>
</tr>
<tr>
<td>Pcl</td>
<td>Pcl1(Phf1)</td>
<td>PCL1(PHF1), PCL2(MTF2)</td>
</tr>
<tr>
<td>PRC2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Esc</td>
<td>Eed</td>
<td>EED</td>
</tr>
<tr>
<td>E(z)</td>
<td>Ezh1, Ezh2</td>
<td>EZH1, EZH2</td>
</tr>
<tr>
<td>Su(z)12</td>
<td>Suz12</td>
<td>SUZ12</td>
</tr>
</tbody>
</table>

Table 1.2.1 Mouse and human homologues of Drosophila PcG
These studies have led to a paradigm whereby members of the PcG form two complexes termed PRC1 and PRC2 (Table 1.2.1). In *Drosophila*, PRC1 is comprised of Ring, Psc, Pc and Ph, while PRC2 is comprised of E(z), Su(z)12 and Esc. This structure is conserved in mammals with RING1, Bmi1, HPH1 and M33 forming PRC1 and EZH2, Suz12 and Eed making up PRC2 (Figure 1.1).

In mammals, however, the picture is complicated by the presence in the genome of multiple paralogues of many PcG members as shown in Table 1.2.1. The mammalian homologues of Ph, PHC1 (Rae-28) and PHC2 both co-precipitate with Bmi1 and interestingly, also co-precipitate with each other [48]. Further, there are six predicted homologues of *Drosophila* Psc in mammalian genomes and many have been observed to co-precipitate with both RING1 and members of the Ph family [49-53]. Homologues of *Drosophila* Pc are present in five copies CBX2, 4, 6, 7, and 8. The composition of PRC2 seems more homogeneous, as only the E(z) homologues, EZH1 and EZH2, exist in multiple copies.

Further variability in PRC complex formation is evident by the co-precipitation, in some contexts, of PRC1 members with PRC2 members, as well as the presence of non-‗core‘ members whose association may be tissue- or stage-dependent [54]. One such protein is the transcription factor YY1 (homologue of *Drosophila* Pho) whose association with the PRC2 complex seems quite robust [55, 56]. While association of YY1 with PRC1 has been described [57] it has proved to be transient or present in substoichiometric quantities [56, 58].

Histone deacetylase complexes (HDACs) are another class of proteins that have been observed to associate with PRC2 in both *Drosophila* [59] and human [60] and HDAC activity is required for suppression of transcription by Polycomb proteins. However, this interaction, like that with YY1, does not seem to be constitutive as complexes have been purified that contain HDAC proteins while others seemed to be devoid of these [47].
Thus, the PcG could now be divided into two classes based on the complexes they comprise: RING, Pc, Ph and Psc for PRC1 and Eed, Suz12 and E(z) for PRC2. The question remained, however, what the activities of these complexes are and by what mechanism do they repress gene expression.

1.2.4 Histone modification

While it was known that PRC complexes were localized to repressed loci, the mechanism by which PcG proteins inhibited transcription was unclear. The association of HDACs with PRCs offered some mechanistic insight, but did little to explain the need for two distinct Polycomb complexes, and left unanswered the question regarding the fraction of PRC2 complexes which do not contain HDACs.

A large stride forward was made in the discovery that the PRC2 complex, specifically E(z)/EZH2 [47, 61] possessed histone methyltransferase activity specific for lysine position 27 of histone H3 (H3K27). Further, it was also demonstrated that the PRC1 complex is specifically recruited to this histone mark via the Pc protein [62].

An enzymatic activity has been ascribed to PRC1 as well. A 250-300kDa complex purified from HeLa cells containing Bmi1, RING1A, RING1B and PHC2 was found to possess in vitro E3-ubiquitin ligase activity for Lysine 119 of histone H2A [63]. The same study identified RING1B as the catalytic subunit of the complex and determined that its RING domain was required for E3 ligase activity. Further, they found that knock-down of RING1B results in decreased H2A ubiquitylation in vivo.

Crucially, while RING1B was found to provide the catalytic activity of the complex, free RING1B protein was found to be approximately 200-fold less active than equimolar quantities of RING1B-containing PRC1 complex [10]. Addition of Bmi1 and RING1A to the reaction did, indeed, enhance H2A ubiquitylation significantly. Thus, Bmi1 and RING1A serve as co-factors for the E3 ligase activity of RING1B. Subsequent analysis of Hox gene expression levels confirmed that Bmi1 and, to a lesser extent
RING1A, are required for efficient repression of their transcription. Mel-18, a paralogue with about 70% amino acid identity with Bmi1, is also capable of forming a stable complex with RING1A and RING1B and enhancing the E3 ligase activity of RING1B in a manner dependent on its phosphorylation [49].

All ubiquitin ligases rely on upstream E1 and E2 ligases as part of the catalytic cascade. For the Bmi1/RING1B dimer, the E2 ligases UbcH3a, b, and c as well as UbcH6 were found to be competent to promote Bmi1/RING1B E3 ligase activity whereas other E2 ligases were not [64]. Subsequently, the Bmi1/RING1B complex was crystalized along with UbcH5c. UbcH5c was found to interact directly with RING1B, but not Bmi1 [65]. The interaction between E2 and E3 ligases mediate the transfer of ubiquitin from the E2 ligase to the substrate.

Since the substrate for Bmi1/RING1B is nucleosomes, specifically histone H2A, it was of interest to determine whether the RING domain dimer had affinity for DNA. Bentley et al. indeed found that Bmi1/RING1B bound duplex DNA in a non-sequence-specific manner to a surface comprising both RING1B and Bmi1. Indeed, mutation of basic residues on this surface to alanines on either Bmi1 or RING1B abrogated both DNA binding and ligase activity to H2A [65].

RING1B null mice are embryonic lethal, implying that RING1A is unable to compensate for RING1B loss in the context of a developing embryo; however, RING1A has similar E3 ligase activity for H2A in vitro [64] and RING1A/B double knock-out MEFs show lower H2A ubiquitylation of the inactivated X-chromosome than the RING1B single knock-out [66].

The determination that the PRC1 and PRC2 complexes possessed enzymatic activity on histones as ubiquitin ligases and methyltransferases respectively was an enormous advance to the field of Polycomb biology. However, the question of how these complexes were targeted to the loci they repress remained unanswered.
1.2.5 Targeting and recruitment

In *Drosophila*, PRC complexes are recruited to Polycomb Repressive Elements (PREs) which are cis-elements that are necessary and sufficient for local repression of transcription by the Polycomb machinery [67, 68]. While PREs are discrete regions, defined to within a few hundred base pairs, PRC-associated histone marks are found to extend many kilobases from the location of PRC binding [69, 70]. This characteristic of highly localized chromatin binding and extended distribution of histone marks has been suggested to be consistent with a looping mechanism whereby nucleosomes distal from bound PRC complexes are brought in contact with the PRCs to facilitate their modification, possibly via the association of the chromodomain of Pc with that fraction of histones that had been previously modified [69].

Despite the identification of PREs in *Drosophila*, the mechanism by which PRC complexes are recruited to DNA is not entirely clear. Pleiohomeotic (Pho) and Pho-like (Phol) do bind DNA in a sequence specific manner [71, 72] and Pho consensus sequences are common in PREs [73]. However, PRCs do not seem to associate constitutively with either Pho or Phol. Further, while Pho is required for repression of some targets [74], loss of both Pho and Phol displaces Polycomb proteins from some, but not all PREs, suggesting there are alternate mechanism for Polycomb recruitment [71].

Indeed, while the PRC1 component Pc is able to bind the H3K27me3 mark deposited by PRC2 [74, 75], not all PRC1-like complexes contain Pc [52]. Further, while ES cells deficient for members of PRC2 show only mild phenotypes and are able to maintain an undifferentiated state [76], PRC1-deficient ES cells differentiate spontaneously and upregulate PcG target genes [77, 78]. In addition, Chromatin Immunoprecipitation (ChIP) analysis of PRC1 binding in ES cells lacking the PRC2 member Eed, which lack H3K27 trimethylation, demonstrated that the distribution and quantity of H2AK119ub was not
significantly altered and while RING1B binding was significantly lower, its distribution was similar to control Eed+/+ cells [79].

In mammals the situation is less clear as nothing resembling a PRE-equivalent has yet been discovered. Since none of the members of the core PRC1 and PRC2 complexes, nor Pho and Phol, can adequately explain the recruitment of Polycomb proteins to chromatin, additional mechanisms must exist. Indeed, one category of molecules whose role in Polycomb recruitment is emerging is that of non-coding RNAs [80, 81] which have been found to interact with PRC2 and other chromatin modifying complexes although much work remains to delineate how non-coding RNAs fit into the Polycomb puzzle.

1.3 Epigenetics and the histone code

1.3.1 The epigenetic code

The ubiquityl and methyl marks deposited by PRC1 and PRC2 respectively are but two of a large number of modifications made to histones and DNA which collectively form the epigenetic code. It is becoming increasingly clear that, beyond genetics, epigenetics comprises an integral means of regulating gene expression. Perturbed epigenetic function has been identified in cancer [82] and causes defective embryonic development [83].

Whereas the traditional transcription factor-mediated view of regulation of gene expression holds that transcription factors bind to DNA in a manner dependent on the presence of specific motifs in the primary sequence of bases, the substrate for epigenetic regulation is the higher order structure of DNA, chromatin. Chromatin is formed of nucleosomes, which consist of 147bp of helical DNA wrapped around a histone octamer containing two each of the histones H2A, H2B, H3 and H4. This repeating structure is the basis of the ‘beads on a string’ model of eukaryotic chromatin which allows for the efficient compaction of DNA in the nucleus. These structures are dynamic and the chromatin state, that
is whether the nucleosomes are tightly compacted as they are in heterochromatin, or in an open state as in euchromatin, is strongly correlated with the level of active transcription of genes in that region.

The post-translational modifications of chromatin that form the basis of epigenetics regulate the accessibility to DNA by the transcription machinery either through direct changes in the electrostatic characteristics of histones or by recruiting co-factors. These post-translational modifications include the methylation of cytosine residues in DNA and the acetylation, phosphorylation, mono-, di, and tri-methylation, ubiquitylation, sumoylation and proline isomerization of histones within the nucleosome. The bulk of the modifications occur near the N-termini of histones which protrude between the coiled DNA strands and are exposed outside of the globular histone core due to their highly basic amino acid composition [84]. Indeed, these protruding histone tails are able to contact neighbouring nucleosomes as well as chromatin remodelling factors which are able to move nucleosomes along DNA via the hydrolysis of ATP.

These marks serve either as activatory signals to facilitate local gene transcription, or as repressive marks which block transcription. Histone marks also serve to either inhibit or facilitate further modification, thus creating a multipartite code whereby the effect on gene regulation is dependent on the sum of several types of local modifications as reviewed by Margueron et al. [85]. The modifications and their associated substrates are summarized below alongside their effect on transcription. Particular emphasis will be placed on ubiquitylation and methylation, as these are the modifications catalyzed by the Polycomb Repressive Complexes-1 and -2 respectively.

1.3.2 Histone modifications

**Acetylation of histones**

Acetylation of histones generally serves to open up chromatin to facilitate transcription. Indeed, several lysine residues are found to be acetylated on histones around the promoters and 5’ end of the coding
region of transcriptionally active genes. Acetyl group are deposited by histone acetyl transferases (HATs) of the GNAT, MYST, and CBP/p300 families and serve to neutralize the positive charge on the modified lysines. This reduces the strength of the electrostatic interactions holding the DNA to the nucleosome facilitating access to DNA by the transcriptional machinery. Acetylation can also disrupt interactions between histones and has been shown to impair formation of 30nm fibres in vitro [86]. Acetylated histones are also able to recruit bromodomain-containing proteins to chromatin [87]. These modifications are dynamic as the activity of the HATs is opposed by that of the histone deacetylases (HDACs), which remove acetyl groups from lysines thereby restoring their positive charge and enhancing their association with DNA. Consistent with this, HDACs are generally associated with transcriptional repression [88]. HDAC inhibitors have been developed for use in cancer therapy [89].

**Phosphorylation of histones**

Histones are also subject to phosphorylation on serine, threonine, and tyrosine residues. The phosphoryl group instills a strong negative charge upon histones which disrupts interaction with DNA. Phosphoryl groups are deposited by a variety on kinases and, once again, this modification is dynamically reversed by phosphatases such as PP1 [90]. Phosphorylation of Serine 10 on histone H3 is perhaps the most well described and is associated both with active transcription and condensation of chromatin during mitosis.

**Methylation of histones**

Histones are modified by methylation of lysines by a variety of proteins that share in common a domain termed SET, for the Su(var)3-9, Enhancer of Zeste, Trithorax genes in which it is present [91-93]. Su(var)3-9 was found to possess methyltransferase activity for lysine 9 of histone H3 (H3K9)[94], Trx and its mammalian homologue Set1 methylate H3K4 as part of the COMPASS complex [95] and Enhancer of
Zeste (EZH2 in mammals) methylates H3K27 [61, 62] as part of PRC2. There are also non-SET domain lysine methyltransferases as well as arginine methyltransferases.

Thus, histones may be methylated on several different sites and further, each lysine may be mono-, di- or tri-methylated. This allows for histones to serve as the substrate for a complex methylation code. Methylation of histones serves as a means by which gene expression can be regulated epigenetically and may be activatory or repressive.

The human Mll gene is one such SET-containing lysine methyltransferase with specificity for H3K4 [96, 97]. Mll was identified via its involvement in a series of frequent translocations in human acute myelogenous leukemia (AML) [98] and was dubbed mixed-lineage leukemia. Mll was observed to bear homology to the yeast protein Set1 which prompted studies which identified the yeast COMPASS protein complex [99].

H3K4 trimethylation is robustly associated with gene transcription and histone acetylation and is enriched within the gene body as well as the promoter region of actively transcribed genes while the dimethylated form shows no such pattern [100, 101]. By contrast H3K9, H3K27 and H3K20 are strongly correlated with condensed chromatin and inactive transcription as reviewed [102].

Methylation was initially thought to be irreversible and serve as a permanent epigenetic record; however, LSD1 and coREST were found to remove mono- and di-methyl marks from H3K4 [103, 104] and di- and tri-methylated lysines were found to be demethylated by the jumanji class of proteins [104-111]. Interestingly, several demethylases have been described to interact with PRC1 complex members suggesting that H2A ubiquitylation and histone demethylation could occur in tandem [53, 108].
**Ubiquitylation of histones**

Like methylation, histone ubiquitylation can serve as both an activatory or repressive signal. For instance ubiquitylation of Lys123 on histone H2B appears to be required for subsequent deposition of the activatory marks H3K4 and H3K79 methylation [112]; however, methylation of H3K36 and activation of the *GAL1* gene in yeast required removal of this ubiquitin by Ubp-8 [113]. By contrast, ubiquitylation of Lys119 of histone H2A is associated with repression of transcription. Indeed, genes marked with H2AK119ub have been found to be repressed despite RNA Pol II being present at the promoter and loss of RING1B resulted in loss of H2AK119ub and subsequent derepression of target genes [78].

**Polycomb-mediated marks**

Of particular relevance to the study of the PcG are the tri-methylation of histone H3 at lysine 27 (H3K27) and the ubiquitylation of histone H2A at lysine 119 (H2AK119). The former mark is deposited by EZH2 as part of the PRC2 complex and the latter by the RING1B-containing PRC1. Both marks serve to repress gene transcription of targeted loci.

The methylation of H3K27 by PRC2 is thought to occur prior to H2AK119 ubiquitylation by PRC1 at a given locus. Mechanistically, this is thought to be mediated via the binding of trimethyl H3K27 by the chromodomain of the PRC1 complex member Pc [74, 75]. While many methyl-modified histones, such as H3K4, H3K36, and H3K79, are associated with activation or elongation of transcription, the mark deposited by PRC2 on H3K27 is inhibitory.

Once thought to be a static scaffold serving only to compact DNA, it has become clear that chromatin and the histones that comprise it form a dynamic system which, along with the epigenetic machinery, provides a layer of transcriptional control above that of the trans-acting transcription factor. The epigenetic modification of chromatin plays vital roles in development and enables the tissue- and
developmental stage-specific expression of genes required for proper physiology of all organisms from yeast to mammals.

### 1.4 The Notch signalling pathway

#### 1.4.1 Notch genetics

Another pathway first identified in *Drosophila* and found to be conserved and have important roles in mammalian biology is the Notch pathway. The Notch phenotype in *Drosophila melanogaster* was first described in 1914. Certain mutant flies were observed to have notched wings, inspiring the name [114]. Mutant alleles of Notch were also employed by Don Poulson whose work provided compelling evidence, for the first time, that genetics directly impacted embryonic development [115]. Poulson described the failure of Notch mutant embryos to develop mesoderm and endoderm and instead show hypertrophic neural development – a phenotype dubbed neurogenic.

The gene whose mutation caused the Notch phenotype was sequenced in 1985 and found to encode a 2703 amino acid polypeptide containing 36 EGF-like repeats [116]. The authors speculated from its hydrophobic properties that the Notch protein likely spanned the membrane due to the presence of what appeared to be a N-terminal signal peptide and a transmembrane-like region between amino acid residues 1720-1740. This was determined to be accurate as the development of antibodies against various regions of the Notch1 protein determined that Notch does indeed form a single-span protein with the N-terminal EGF repeats exposed extracellularly and the C-terminus in the cytoplasm [117].

Notch’s mechanism of action was uncovered thanks, in part, to the discovery of a translocation in a case of T cell acute lymphoblastic leukemia involving the human homologue of *Drosophila* Notch which yielded a truncation lacking the extracellular domain [118]. Inspired by this mutant allele, Coffman created a mutant form of the *Xenopus* Notch homologue, Xotch, which lacked its extracellular domain.
This mutant, called XotchΔE, proved to act as a dominant active and caused severe defects in embryogenesis.

1.4.2 The Notch transactivation complex

Notch was observed to regulate the subcellular localization of Drosophila Su(H), which had previously been found to interact genetically with Notch [119]. The mammalian homologue of Su(H), CSL (for CBF-1, Su(H), lag-1) was able to bind to the promoter of the HES1 gene, known to be activated by Notch, but was not, on its own, able to activate transcription. However, when the analogue of Coffman’s ΔE construct was expressed, Notch was detected in the nucleus. In a gel shift assay, extracts from 293T cells expressing alleles of Notch were able to cause shifts in CSL migration [120]. Pure preparations of Notch proteins, however, did not interact with CSL in the way that cell extracts did, suggesting there is another cellular component required for complex formation. This component was identified in a yeast two hybrid screen to be the protein Mastermind [121]. Combined, these studies demonstrated that Notch, a transmembrane protein, had a dual role both as a surface receptor and as a nuclear activator of transcription.

The ternary complex suggested by Patcherski and Kimble’s data was confirmed when the crystal structure for Notch, Mastermind-like-1 (Maml1) and CSL was solved in association with DNA [122, 123]. The formation of this complex on CSL displaces repressors [124, 125] and recruits activators such as p300 [126, 127]. This converts CSL from a repressor to an activator of transcription and induces the expression of target genes.

1.4.3 Mechanism of signal transduction

So how does Notch, a transmembrane protein, exert its effects in the nucleus? It was suggested that the intracellular domain of Notch, ICN, translocated to the nucleus to interact with CSL, but it was not clear how Notch was liberated from the membrane. Edman degradation sequencing of ICN derived
from membrane-tethered Notch revealed that all ICN polypeptides began at the same amino acid residue: valine 1744. It was observed that this species accumulated when the Notch ligand Jagged-1 was co-expressed. Further, mutation of this residue abrogated Notch activity in vitro [126] and in vivo [128]. It was later confirmed that the Presenilin complex is responsible for an intramembrane cleavage event which releases ICN from the membrane whereby it translocates to the nucleus [129-132].

It turns out that cleavage by Presenilin, or γ-secretase, is but the final in a series of three proteolytic cleavages of Notch. Full length Notch exists, in fact, as a heterodimer formed by non-covalent association of the N-terminal extracellular portion of Notch [133] with the C-terminal transmembrane portion after cleavage by the furin protease in the golgi [134]. Upon interaction with ligand at the membrane, a second cleavage is mediated by an ADAM-type metalloprotease [135, 136], permitting the final cleavage at valine 1744 by γ-secretase which frees Notch to translocate to the nucleus.

1.4.4 Notch and T cell Acute Lymphoblastic Leukemia

The human homologue of Notch was determined in 1991 to be identical to the TAN-1 gene which had been implicated by translocation in rare cases of T-ALL [118]. Since the t(7;9)(q34;q34.3) translocation placing TAN-1 under the control of the TCRβ promoter proved to be extremely rare, Notch was deemed to be only of peripheral importance to the pathology of T-ALL. This dogma was, however, turned on its head with the discovery in 2004 that point mutations in the Notch gene were present in over 50% of cases [5].

These mutations fell into two categories: either they disrupted the heterodimerization domain which held the N-terminal extracellular half of Notch to the C-terminal transmembrane half, or they were missense or nonsense mutations which resulted in truncated protein. The former class either increased the sensitivity of Notch to activation by ligand, or removed ligand dependence altogether [5], and the
latter resulted in a truncated protein lacking the C-terminal PEST domain responsible for mediating Notch’s degradation [5].

Notch is not known to play as critical a role in any other malignancy but its association with cancer is not limited to T-ALL. *Notch1* mutations have also been identified in two independent studies in chronic lymphocytic leukemia and these were found to be associated with lower survival in that disease [137-139]. Activating mutations within the PEST domain of *Notch1* were similarly found in mantle cell lymphoma and inhibition of the Notch pathway reduced proliferation and induced apoptosis in MCL cell lines [140]. Similar gain-of-function mutations were discovered in the *Notch2* gene in cases of marginal zone lymphoma, particularly in splenic cases (25%) compared with non-splenic (5%) [141]. Gain-of-function fusions of *Notch1*, which occur rarely in T-ALL, have also been observed in breast cancer and these cells showed sensitivity to γ-secretase inhibition [142].

While Notch is strongly implicated in the pathogenesis of T-ALL, it was recently reported by two groups that Notch signalling is low or absent in human samples of AML and that enforced Notch signalling has a tumour suppressor effect [143, 144]. Notch has also been described as a tumour suppressor in melanoma, as recently reviewed [145], and in several types of squamous cell cancer [146, 147].

Further studies have implicated the Notch pathway in neural development, the vascular disease CADASIL, bone maintenance as well as lymphopoietic and intestinal hemostasis. Thus, the study of Notch and its transcriptional targets has clear implications for human health and pathology.

### 1.4.5 Hypotheses

**Hypothesis #1**  *Notch drives the expression of transcriptional target genes in T-ALL which are candidate oncogenes in this disease.*

Chapter 3 posits that Notch, as a transcription factor, drives the expression of genes which contribute to leukemogenesis but which have not yet been identified and characterized. This investigation led to the
identification of PCGF5. The data show that PCGF5 expression is driven by Notch signalling and that PCGF5 seems to be a direct target of the Notch transactivation complex. PCGF5 is highly expressed in developing thymocytes, as well as mature T cells, and its expression correlates with Notch signalling activity.

_Hypothesis #2  PCGF5 has roles in hematopoietic development and enforced expression of PCGF5 during hematopoiesis will cause aberrant phenotypes._

Chapter 4 details experiments conducted to test the hypothesis that enforced expression of PCGF5 will confer alterations in hematopoietic development consistent with its putative role as a Notch-driven oncogene. No strong phenotypes associated with Notch gain-of-function in hematopoietic progenitors were observed by enforced expression of PCGF5.

_Hypothesis #3  PCGF5 associates physically with the PRC1 core subunit RING1B._

_Hypothesis #4  PCGF5:RING1B complexes have ubiquitylating activity towards K119 of histone H2A._

PCGF5 was so named due to its sequence homology to a subset of Polycomb Group genes but there remained no functional data to back up this association. Thus, Chapter 5 examines _in vitro_ data confirming PCGF5’s identity as a functional member of the Polycomb Group. PCGF5 protein was observed to associate with RING1B and this association seemed to occur in competition with Bmi1. Overexpression of PCGF5 in mouse fibroblasts was found to cause a global increase in the ubiquitylhistone H2A mark which is the hallmark of PRC1 activity. It was further observed that overexpression of PCGF5 or Bmi1 caused overlapping but non-identical changes in gene expression.

These data support the conclusion that PCGF5 is a Polycomb Group gene that takes part in a PRC1-like complex and is competent to promote PRC1 activity and regulate gene expression. PCGF5 expression is
transactivated in T-ALL cells by oncogenic Notch and merits further study to determine whether it may serve as a therapeutic target.
Chapter 2  Materials and methods

**Cell lines and reagents** – The human T-ALL cell lines used to identify PCGF5 as a target of Notch (All-SIL, DND 41, HPB-ALL, KopTK1, and T-ALL1) were grown in RPMI supplemented with 10% fetal bovine serum (FBS), 1mM Sodium Pyruvate, 2mM Glutamax (Invitrogen), Penicillin and Streptomycin (Stem Cell Technologies). Notch inhibition was accomplished using the Compound E (γ-secretase inhibitor XXI, EMD Chemicals). Bmi1 null and control immortalized mouse embryonic fibroblasts (MEFs) were a kind gift from Martaan Van Lohuizen. These fibroblasts were derived from mice bearing null alleles for Bmi1 and ARF [31, 148]. MEFs were cultured in Dulbecco’s Modified Eagle Medium, supplemented with 0.1mM 2-mercaptoethanol, 0.2M HEPES, 10% fetal bovine serum, 1mM Sodium Pyruvate, 2mM Glutamax (Invitrogen), Penicillin and Streptomycin (Stem Cell Technologies). PlatE cells [149] were used for packaging of ecotrophic virus and 293T cells used both for packaging and transient over expression were grown in Dulbecco’s Modified Eagle Media supplemented with 10% fetal bovine serum (FBS), Sodium Pyruvate, Glutamax, HEPES (20mM) and 2-Mercaptoethanol (0.1 mM), Penicillin and Streptomycin. The OP9-GFP and OP9-DL1 cells are derived from the OP mouse strain (Kodama S. 1994) and were grown in Modified Eagle Media with Alpha modification supplemented with 10% fetal bovine serum (FBS), Sodium Pyruvate, Glutamax, Penicillin and Streptomycin as above.

**Mice** – The mice employed for transplantation experiments were of the C57Bl/6J. The otherwise syngeneic Pep3b strain, which bears an alternatively modified form of the CD45 surface marker were also employed to track donor/recipient cells.

**Retroviral production and transduction** – MIG retroviruses are based on a Mouse Stem Cell Virus (MSCV) vector that expresses green fluorescent protein (GFP) from an Internal Ribosome Entry Site (IRES). Either calcium phosphate or Turbofect in vitro reagent (Fermentas) were used to transfect 293T cells with viral vectors and accessory plasmids or PlatE, which stably express gag, pol and env, cells with
viral vectors alone. Accessory plasmids were EcoPac and pCMV VSVg for ecotrophic virus and pKat Ampho and pCMV VSVg for amphotrophic virus. Virus-containing supernatant was harvested and day 2 and 3 post-transfection and pooled. The supernatant was filtered through 0.45μm filters then either used fresh or frozen for storage at -80 degrees.

Transduction was performed in the presence of 4μg/ml polybrene either at 37 degrees in the incubator for 4 hours to overnight or at 1800g and room temperature for two hours. After the transduction period, viral supernatant was replaced with appropriate media. Subsequent rounds of infection were performed similarly as required.

**Colony Forming Cell Assay** - Bone marrow transplants – Cells were harvested from mouse bone marrow with or without injection 4 days prior with 5-fluorouracil (5-FU). Erythrocytes were depleted from the suspension by ammonium chloride lysis then cells were put in culture with 10ng/ml each mIL-3, mIL6, Flt3L and m100ng/ml SCF (Stem Cell Technologies or Peprotech). The following day, cells were divided into wells of a 24-well plate for infection. Cells were centrifuged in the 24-well plate, most of the media was aspirated off, and then 750μL of virus-containing supernatant was added.

Infection efficiency was determined the following day by flow cytometry then cells expressing green fluorescent protein (GFP) were enriched by Fluorescence Activated Flow Cytometry (FACS) sorting. Sorted cells were counted and then resuspended at 10⁴ cells/ml for 5-FU treated mice then plated in 35mm dishes. CFC media contained methylcellulose, rm IL-3, rm IL-6, rm Stem Cell Factor (SCF), fetal bovine serum, rh Insulin, human transferrin (iron saturated), 2-Mercaptoethanol and Iscove’s Modified Dulbecco’s Media and supplements provided by the vendor (Stem Cell Technologies, Vancouver, BC). Colonies were counted after 10 days of culture.

**Bone marrow transplants** – Cells were harvested from mouse bone marrow with or without injection 4 days prior with 5-fluorouracil. Erythrocytes were depleted from the suspension by ammonium chloride
lysis then cells were put in culture with 10ng/ml each mouse IL-3, mouse IL6, Flt3L and m100ng/ml SCF (Stem Cell Technologies or Peprotech). The following day, cells were divided into wells of a 24-well plate for infection. Cells were centrifuged in the 24-well plate, most of the media was aspirated off, and then 750μL of virus-containing supernatant was added.

Infection efficiency was determined the following day by flow cytometry and then 50,000 cells/recipient were injected into lethally irradiated mice (840 rads, X-ray). Mice were given water supplemented with ciprofloxacin to decrease the risk of opportunistic infection. Engraftment was measured when required by flow cytometric confirmation of the presence of GFP-expressing cells in peripheral blood harvested from the tail vein.

Mice were sacrificed by carbon dioxide asphyxiation and dissected to collect the required tissues which included spleen, bone marrow, thymus and lymph nodes.

**OP9 Coculture** – Fetal livers were dissected from d13.5 embryos then disaggregated by mechanical pipetting. The cell suspension was passed through a 70μm strainer then counted. Cells were resuspended in media and incubated with anti-CD24 antibody then Low-Tox-M Complement (Cedarlane, Burlington, Ont) was added to lyse the CD24-bound cells. Viable cells are recovered by density centrifugation with Lympholyte-M (Cedarlane, Burlington, Ont). Cells were counted and then plated onto fresh non-confluent OP9-GFP or OP9-DL1 cells (a kind gift from JC Zuniga-Pflucker) at a density of 6 X 10^4. After 5-6 days the cells dissociated from the feeder layer by hard pipetting then filtration through a 70μm strainer. Cells were counted, aliquots taken for flow, and then replated onto fresh non-confluent feeders.

Analysis was conducted on a FACS Calibur. Cells were stained either with a Double Negative panel consisting of a lineage panel (CD4, CD8), CD44, CD25 and CD45 to distinguish hematopoietic cells from feeders. For measuring Double Positive cells CD4, CD8 and CD45 were employed.
**Flow Cytometry acquisition and analysis** – All data were collected on FACS Calibur (BD Biosciences) benchtop flow cytometers except for 6-7 colour analysis of bone marrow transplant recipient mice which was acquired on a FACS Canto II (BD Biosciences) benchtop instrument. Data was analysed using FlowJo (Tree Star, Ashland, OR).

**Antibodies** – Flow cytometry was performed with the following antibodies purchased from eBioscience, unless otherwise noted: CD3e (clone 145-2C11), CD4 (clone L3T4), CD8 (clone 53-6.7), Mac1/CD11b (clone M1/70), CD16/CD32 (clone 93), CD19 (clone 1D3), CD25 (clone BC96), CD44 (clone IM7), CD45.1 (clone A20) CD45.2 (clone 104), CD24 (clone M1/69), CD117/c-kit (clone 2B8), Sca1 (clone D7), Ly-6G/Gr-1 (clone RB6-8C5), CD45R/B220 (clone RA3-6B2), CD34 (clone RAM34), TER-119 (clone TER-119).

Immunoblots were performed with the following primary antibodies: Ub-H2A (clone E6C5, Millipore), histone H3 (rabbit polyclonal, Cell Signalling Technologies), FLAG (clone M2, Sigma-Aldrich; rabbit polyclonal, Rockland), HA.7 (Sigma), Bmi1 (H-99, Santa Cruz Biotechnology; clone 1.T.21, Abcam), RING1A (rabbit polyclonal, Abcam). Bound primary antibody was detected with HRP-coupled secondary antibodies from Amersham.

**Western Blotting and Immunoprecipitation** – SDS-PAGE experiments were performed using either Invitrogen Novex pre-cast gels and accompanying buffers or using home-made gels. Gels were cast using the BioRad Protean III apparatus with Bio-Rad Acrylamide/Bis solution 19:1, 0.1% ammonium persulfate, 0.1% sodium dodecyl sulfite (SDS), in a Trizma (Invitrogen) and glycine based buffer. Proteins lysates were mixed 2:1 with sample loading buffer containing 2-mercaptoethanol, SDS ethylene diamine tetraacetic acid (EDTA) and bromophenol blue in a Tris-based buffer and migrated alongside PageRuler molecular weight markers (Fermentas) to resolve protein size.

Histone extracts were generally resolved on 15% gels, whereas acrylamide percentages were chosen for other experiments to suit the size of the protein being probed for. Once proteins were resolved
electrophoretically to the desired degree, proteins were transferred from the gel to nitrocellulose membrane or 0.2μm PVDF (Bio-Rad) immersed in transfer buffer (20% methanol, 0.192M Glycine and 0.025M Tris) at either 70V for 90 minutes or 40V overnight. Membranes were blocked with 5% skim milk powder or SuperBlock solution (Pierce), then incubated with primary antibodies for between 1 hour and overnight, then washed in Tris-Buffered Saline + 0.05% Tween-20 (TBS-T), probed with secondary antibodies for 1 hour, and washed thoroughly for 30-45 minutes in TBS-T. The ubiquitin-H2A antibody specifically required PVDF and SuperBlock solution. Enhanced chemiluminescence reagents (Amersham ECL Plus) were used for detection and this signal was visualized by exposure to Kodak Biomax film.

**Expression Microarray** – Analysis and generation of heatmaps was performed using dChip software [150]. Expression data generated in the Dr. Andrew Weng’s lab by Carol Wai were normalized to the array of median brightness (CP060329H133P09 Kmyc M12) by the invariant set method [151]. The dChip program then calculated the model based expression values using the PMs and MMs. Probesets were filtered based on “presence” call > 40%, signal intensity > or = 50 in at least 50% of samples and variability constrained by 0.3 < Standard deviation / Mean < 10. Samples were divided into ‘Notch On’ (n=20) and ‘Notch Off’ (n=10) groups and the group means were calculated by dChip for each probeset in the filtered list. A supervised analysis was performed for probesets that were at least 1.2-fold higher and had an expression value 100 units higher in the ‘Notch On’ group than in the ‘Notch Off’ group assess by paired t-test. The False Discovery Rate was estimated by testing 100 permutations of paired data (with and without Cmpd E in the same cell line) by the same criteria as that used for the supervised analysis.

For publically available datasets, normalized datasets were downloaded from the GEO website as Series Matrix files, then loaded into dChip. GEO accession numbers, normalization methods, and annotation files from [http://www.affymetrix.com/support/support_result.affx](http://www.affymetrix.com/support/support_result.affx) are listed in Table 2.1.
**Table 2.1 – Details of public microarray data**

<table>
<thead>
<tr>
<th>Figure</th>
<th>GEO acc #</th>
<th>Normalization</th>
<th>Platform</th>
<th>Annotation</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.2.4A; 3.2.5</td>
<td>GSE29544</td>
<td>RMA package in R</td>
<td>GPL570: [HG-U133_Plus_2]</td>
<td>HG-U133_Plus_2 Annotations, CSV format, Release 32</td>
</tr>
<tr>
<td>3.2.4B</td>
<td>GSE18198</td>
<td>RMA with quantile normalization</td>
<td>GPL570: [HG-U133_Plus_2]</td>
<td>HG-U133_Plus_2 Annotations, CSV format, Release 32</td>
</tr>
<tr>
<td>3.2.12</td>
<td>GSE26713</td>
<td>MAS 5.0 with RMA</td>
<td>GPL570: [HG-U133_Plus_2]</td>
<td>HG-U133_Plus_2 Annotations, CSV format, Release 32</td>
</tr>
<tr>
<td>3.2.13A; 3.2.14B</td>
<td>GSE18497</td>
<td>RMA with quantile normalization</td>
<td>GPL570: [HG-U133_Plus_2]</td>
<td>HG-U133_Plus_2 Annotations, CSV format, Release 32</td>
</tr>
<tr>
<td>3.2.13B</td>
<td>GSE14618</td>
<td>MAS 5.0 with RMA</td>
<td>GPL570: [HG-U133_Plus_2]</td>
<td>HG-U133_Plus_2 Annotations, CSV format, Release 32</td>
</tr>
<tr>
<td>3.2.14A</td>
<td>GSE14479</td>
<td>RMA, log2 transformation</td>
<td>GPL570: [HG-U133_Plus_2]</td>
<td>HG-U133_Plus_2 Annotations, CSV format, Release 32</td>
</tr>
<tr>
<td>3.2.16</td>
<td>GSE10246</td>
<td>GC-RMA</td>
<td>GPL1261: [Mouse430_2]</td>
<td>Mouse430_2 Annotations, CSV format, Release 32</td>
</tr>
<tr>
<td>3.2.17</td>
<td>GSE31082</td>
<td>MAS 5.0</td>
<td>GPL1261: [Mouse430_2]</td>
<td>Mouse430_2 Annotations, CSV format, Release 32</td>
</tr>
<tr>
<td>3.2.17</td>
<td>GSE30631</td>
<td>MAS 5.0</td>
<td>GPL81: [MG_U74Av2]</td>
<td>MG_U74Av2 Annotations, CSV format, Release 32</td>
</tr>
<tr>
<td>3.2.18</td>
<td>GSE1460</td>
<td>RMA with quantile normalization</td>
<td>GPL97: [HG-U133B]</td>
<td>HG-U133B Annotations, CSV format, Release 32</td>
</tr>
<tr>
<td>3.3</td>
<td>GSE5716</td>
<td>MAS 5.0 + adjustment of interarray intensity</td>
<td>GPL570: [HG-U133_Plus_2]</td>
<td>HG-U133_Plus_2 Annotations, CSV format, Release 32</td>
</tr>
</tbody>
</table>

**Cloning** – Human PCGF5 cDNA was purchased in a pCMB-Sport6 vector from OpenBiosystems. PCGF5 was amplified by PCR with a 5’ primer containing a Kozak consensus and a BglII restriction site, and a 3’ primer containing the native stop codon and a XhoI site. The amplicon was digested with BglII/XhoI and cloned into MSCV-IRES-GFP (MigR1) to be driven by the LTR promoter. To allow for detection and immunoprecipitation, PCGF5 was amplified once again with the same 3’ primer, but with a 5’ primer containing an in-frame BglII site as before but starting from the second amino acid of PCGF5. This was cloned into both pcDNA3/nFLAG and pcDNA3/nHA where PCGF5 would be fused to an N-terminal FLAG or HA-tag, respectively with an Arg-Ser linker between the tag and the coding sequence of PCGF5 derived from the BglII site. All constructs were sequence verified.
Mouse Bmi1 was amplified from a bacterial expression vector, kindly shared by Dr. Titia Sixma of the Netherlands Cancer Institute, using primers containing a 5’ BglII site and a 3’ XhoI site. The amplicon was cloned into MigR1, pcDNA3/nFLAG and pcDNA3/nHA as above and clones were sequenced.

**Quantitative PCR** – Gene expression levels were determined by Quantitative Polymerase Chain Reaction of Reverse Transcribed mRNA (RT-PCR). Cells were lysed in Trizol reagent (Invitrogen) and protein was removed by extraction with chloroform. RNA was isolated in the aqueous phase then precipitated with 2-propanol, then washed in 70% ethanol and pelleted. The RNA pellet was air dried then dissolved in water treated with DEPC to remove RNase activity. The material was then quantified spectrophotometrically either using a benchtop spectrophotometer (Bio-Rad) or a NanoDrop device (Invitrogen). Up to 2µg of RNA was reverse transcribed using either Superscript II (Invitrogen) or GoScript (Fermentas) reverse transcriptase to yield cDNA.

RT-PCR reactions employed Platinum Quantitative PCR Supermix-UDG (Invitrogen) which contains the polymerase enzyme, deoxynucleotides, and Sybr green and appropriate buffers at 2x concentration. Also included is a Uracil DNA glycosylase to prevent carryover contamination. To this was added diluted cDNA, gene specific primers and water to make correct the volume such that the Supermix contents were at the appropriate concentrations. Reactions were performed in 96-well plates in a volume of 20µL in a Dyad Disciple thermocycler (Bio-Rad) and analyzed with the Gene Expression Analysis for iCycler iQ Real-Time PCR Detection System Excel macro (Bio-Rad). Each primer set was assayed at a range of temperatures and subsequent assays were performed at the temperature which gave the most efficient amplification of specific product as determined by melt curve analysis and gel electrophoresis.

**qPCR Primer Sets**

mouse Bmi1: GGA TGG ACT GAC GAA TGC TGG; CCA GAT GAA GTT GCT GAT GAC CC
mouse HoxA9: GAA CCA GAT CTT GAC CTG C; ACG GCA GGT ATA TGC GCT

mouse HoxA10: AGT TCC AAA GGC GAA AAT GC; GTG AGT TCT GGG GCA GAG GC

mouse HoxB4: TTC ACG TGA GCA CGG TAA AC; GTT GGG CAA CTT GTG GTC TT

mouse HoxB7: AAC CGA GTT CCT TCA ACA TG; CGA GTC AGG TAG CGA TTG TA

mouse HoxC8: CCA CGT CCA AGA CTT CTT CCA CCA CGG C; CAC TTC ATC CTT CGA TTC TGG AAC C

**Histone Ubiquitylation** – Cells were washed twice in PBS then resuspended in PBS with 0.2N HCl and incubated on ice for 15 minutes. Acid lysates were centrifuged at 4 degrees and maximum speed for 15 minutes to pellet debris. The supernatant was transferred to a clean tube, to which an equal volume of 50% trichloroacetate (TCA) in water was added to precipitate protein. The suspension was incubated at 4 degrees for between 15 minutes to overnight then centrifuged at 4 degrees and maximum speed for 15 minutes to pellet precipitated protein. The supernatant was then decanted the pellet was washed three times in ice cold acetone to remove the TCA. After the last wash, the acetone was removed and the pellet was air dried. Once dry, the pellet was dissolved in alkaline SDS-PAGE sample buffer (50mM Tris pH 8.8, 2% SDS, 10% glycerol, and 100mM DTT added immediately prior to use) at 95 degrees to aid dissolution. The dissolved protein could then be quantitated using the BCA Protein Assay Reagent (Pierce). The samples were then aliquoted for resolution by SDS-PAGE and diluted in an equal volume of RIPA buffer (50mM Tris, 150mM NaCl, 0.1% SDS, 0.5% sodium deoxycholate, 1% NP-40, 1mM PMSF) supplemented with 9% 2-mercaptoethanol and bromophenol blue and heated to 95 degrees again for 5’. Samples were then run at a 15% polyacrilimide gel and transferred to 0.2μm PVDF which were blocked with SuperBlock solution (Pierce). Membranes were cut between the 15 kDa and 25 kDa molecular weight markers (Fermentas). The higher MW membrane was probed for ub-H2A clone E6C5
(Millipore) at a dilution of 1:1000 and the lower MW membrane was probed with anti-histone H3 antibody at 1:1000 as a loading control.

The integrated density of the ub-H2A and H3 bands were determined using ImageJ software (Wayne Rasband, National Institute of Health) and the ub-H2A signal was normalized to H3 then reported as relative ub-H2A levels compared to control.

**ChIP-seq analysis** – ChIP-seq data for Notch and CSL were downloaded from NCBI GEO database (accession #GSE29600) [2]. This data had already been processed such that human sequence reads were mapped to hg18 and mouse sequence reads were mapped to mm9 using Illumina Genome Analyzer Pipeline allowing 2 mismatches. I analyzed this data, in BED file format, as 40bp intervals, using the Cistrome Galaxy [152] suite of bioinformatic software. Replicate datasets were processed using the Cistrome Galaxy intersect tool and only intervals appearing in both datasets were retained for further analysis. Significantly enriched regions (SERs) over control pre-immune rabbit IgG or non-immune mouse IgG input control were identified in the human genome build hg18 using the MACS utility using a p-value of $10^{-7}$ with a small/large fold enrichment for model building or 10/30 and small/large lambda of 1000/10,000 bp. The effective genome size was 2,770,000,000 bp. When duplicate tags were present at the exact same location only a single tag was kept for analysis. WIG files were generated by Cistrome Galaxy during this process and these were visualized with the UCSC Genome Browser hg18 assembly. Enrichment tracks show fold enrichment as the ratio between the ChIP-Seq tag count and $\lambda_{local}$ for every 10 bp. $\lambda_{local}$ is a model-based control [153] designed to reduce local biases.

The PCGF family dataset in the appendix was shared by Dr. Reinberg [52] as a list of filtered SERs with left and right boundaries, fold enrichment over input control, and p-value. From the Gao et al. manuscript:
Sequenced reads (36bp) from each ChIP-seq experiments were aligned to the human reference genome (assembly hg19) using Bowtie [154]. Duplicated reads were removed with samtools [155]. ChIP-seq read density files were generated using igvtools [156] by extending each reads to 200bp and were viewed in Integrative Genomics Viewer (IGV) [156] or UCSC genome browser (genome.ucsc.edu). Significantly (P<0.01) enriched regions (SER) for each ChIP-seq dataset were identified with QESEQ [157].

Average profiles of H2AK119ub enrichment over input control were calculated and rendered using Site-Pro on Cistrome over a window extending 5kb in each direction from the centre of the SER [158]. Lists of SERs co-enriched by multiple proteins were obtained using the Intersect tool set at 1 bp of overlap; non-overlapping SERs were obtained using the Subtract tool set at 1 bp overlap. Both tools are part of the Operate on Genomic Intervals function of Cistrome Galaxy [152].
Chapter 3 Notch signalling regulates expression of PCGF5 in T cell Acute Lymphoblastic Leukemia cells

Contribution

The expression profiling experiment featured in Figures 3.2.1-2 and 3.3.1-2 was performed in Dr. Andrew Weng’s lab by Carol Wai. I analyzed these data to identify genes that significantly correlated with Notch signalling status.

I performed the quantitative RT-PCR experiments in Figures 3.2.3 and 3.2.5.

For Figures 3.2.4, 3.2.5B, 3.2.12-18, and 3.3.3 I performed statistical analysis of probeset values for normalized microarray data downloaded from the NCBI Gene Expression Omnibus database.

The ChIPseq data in Figures 3.2.7-11 was downloaded from the NCBI Gene Expression Omnibus. I performed the normalization and peak calling to identify SERs and generate enrichment tracks.

3.1 Introduction

Notch1 is a potent oncogene in T-Cell Acute Lymphoblastic Leukemia (T-ALL) and the majority of T-ALL cases harbour activating mutations of Notch1. Activated Notch exerts its effects by translocating to the nucleus after release from the membrane. There it interacts with the DNA binding protein CSL in a ternary complex with MamL1 [122, 123]. Interaction with Notch and MamL1 displace proteins such as SMRT and convert CSL from a transcriptional repressor to an activator [124, 125]. This conversion depends on the recruitment of more transactivators such as CBP/p300 which mediates acetylation of histones and transcription of target genes [127].

Given that the outcome of Notch signalling is the transactivation of Notch target genes, the implication of the Notch pathway in the pathogenesis of T-ALL begs the questions:
1. Which genes does Notch1 regulate in T-ALL cells?

2. Which of these contribute to the genesis and pathology of T-ALL?

One such gene is the proto-oncogene myc [159, 160]. No doubt myc serves as a potent driver of Notch oncogenicity; however, there is conflicting evidence as to whether myc alone is sufficient to recapitulate the effects of Notch. In an experiment with cell lines which rely on Notch for their growth, transduction with c-myc rescued 3 of 5 human cell lines tested as well as the sole mouse line tested [160]. Another study reported c-myc-mediated rescue of Notch inhibition of mouse cell lines, although survival was the only parameter measured in this study [161]. It should be noted, however, that in the first study to describe frequent Notch mutations only 5 of 30 cell lines tested demonstrated sensitivity to Notch inhibition [5], suggesting that clonal selection for Notch-independence may be common in cultured T-ALL cells.

By contrast, an in vivo study employed a mouse model where both Notch and c-myc are overexpressed leading to T-ALL. Subsequent loss of exogenous Notch, while retaining the c-myc transgene, caused tumour regression. However, loss of exogenous c-myc was tolerated if the Notch signal persisted [7]. It is likely that in this case Notch driven c-myc is able to compensate for loss of exogenous myc. However, the insufficiency of myc alone, after loss of Notch, to sustain leukemic growth demonstrates that, in this context, Notch drives expression of genes, other than c-myc, that are required for T-ALL cells.

This study describes finds that expression of Polycomb Group Ring Finger 5 (PCGF5) was significantly reduced upon Notch inhibition and we show evidence that this is a result of direct transcriptional upregulation of the PCGF5 locus by Notch. Thus, PCGF5 is a transcriptional target of Notch and a candidate effector of Notch-driven pathogenesis in T-ALL.
3.2 Results

3.2.1 Identification of PCGF5 as a target of Notch

To identify which genes are regulated by Notch in human T-ALL, an expression profiling experiment was performed for five different human T-ALL cell lines that each require Notch for their growth [5]. In these cells, Notch signalling was disrupted by treatment for 12 hours with the γ-secretase inhibitor (GSI) Compound E (Cmpd E). Samples were also prepared from each cell line transduced to express either an activated mutant form of Notch (ICN) that is insensitive to GSI or the known Notch target gene, myc. The ICN-transduced cells were included to control for non-specific effects of Compound E. True targets of Notch should remain elevated in ICN-expressing cells and genes whose expression changes upon Compound E treatment despite ICN expression are likely being modulated by the drug in a Notch-independent manner. Since myc is known to modulate the expression of many genes, and c-myc itself is a target of Notch, the c-myc transduced samples serve to cull targets of myc from the target list to enrich for other direct targets of Notch transactivation. After treatment, cells were harvested to be assayed on an Affymetrix HG-U133 Plus 2.0 microarray.

Probesets were filtered based on having a variation across samples of $0.3 < \text{Standard deviation / Mean} < 10$; Presence call in $\geq 40\%$ of samples; expression level $\geq 50$ in $50\%$ of samples. This reduced the probeset list from 54675 to 6009 probesets that met the inclusion criteria.
The combined data from the five cell lines was subjected to a supervised analysis to discover genes that correlated with Notch signalling across all five isogenic systems. The comparison was set to find genes in which the expression values in the ‘Notch On’ group were at least 1.2-fold and at least 100 units greater than for the ‘Notch Off’. The student’s t-test was used to test the hypothesis that the mean expression value for a given probeset was higher in the ‘Notch On’ group than in the ‘Notch Off’ group. The null hypothesis was that there was no relationship between probeset expression values and Notch

Figure 3.2.1 Notch causes changes in gene expression in 5 T-ALL cell lines Heatmap of genes whose expression correlates with Notch signaling status. Five human T-ALL cell lines, each either left transduced, or transduced with either Mig ICN or Mig c-myc were treated with either 1 μM Cmpd E or vehicle control. Samples were assigned to either ‘Notch On’ (|) or ‘Notch Off’ (O) groups then subjected to a supervised analysis for genes which were significantly higher in the | group than the O group. The null hypothesis was tested by student’s t-test and genes found to significantly correlate (p < 0.05) are listed ordered by ascending p value. Expression values are normalized for each probeset across all samples with a mean of 0 and an s.d. of 1. Darkest blue indicates 3 s.d. below the mean; darkest red indicates 3 s.d. above the mean.
signalling status. **Figure 3.2.1** shows genes, ranked from smallest p-value to largest whose expression is significantly different (p < 0.05) between the ‘Notch On’ and ‘Notch Off’ groups. Each probeset is scaled across the entire row, all samples and all cell lines, such that the median sample is set to 0. The Heatmap shows variation from -3 to +3 standard deviations (s.d.).

The False Discovery Rate (FDR) was estimated by randomly permuting the samples 100 times into an A group of 20 samples and a B group of 10 samples and testing each permutation by the same criteria as that used for the supervised analysis. This test gave very high rate of Type I errors (median FDR=50%; 9 genes and the 90th percentile FDR = 1394%; 251 genes) when applied to the entire dataset due to high variability between cell lines. To counter this, a paired t-test using p < 0.05 was conducted for only the vehicle control and Cmpd E-treated untransduced cells. The remaining test parameters were left unchanged (change of 1.2-fold and expression level difference >= 100). This calculation estimated the median and 90th percentile FDR to be 0% and 20%, respectively.

Reassuringly, canonical Notch targets such as the Hes and Deltex families were both identified in the comparison in **Figure 3.2.1**. *IGF1R*, recently reported by Dr. Weng’s lab to be a Notch transcriptional target [162], also appeared on the list. The pre T cell antigen receptor alpha (*PTCRA*), another canonical target of Notch1 in T cells [163] was found to be regulated by Notch in some, but not all, cell lines and thus does not appear on the list.

One gene that was consistently regulated by Notch in all five cell lines was Polycomb Group Ring Finger 5, or *PCGF5* (**Figure 3.2.1**). In **Figure 3.2.2**, the heat map scaling is done independently for each cell line to better illustrate the directionality of changes attributed to the experimental interventions rather than differences between cell lines. *PCGF5* very closely followed the pattern predicted for a *bona fide* Notch target in all five cell lines, decreasing upon treatment with Compound E in parental and myc-transduced...
cells and maintaining high expression in the presence or absence of Compound E in ICN-transduced cells.

**Figure 3.2.2 – Regulation of PCGF5 across five T-ALL cell lines** Heatmap of PCGF5 expression normalized within each cell line for the four PCGF5 probesets that showed significant regulation by Notch by a supervised analysis. Legend below shows treatment with 1 μM Compound E for 12 hrs and transduction with either ICN- and myc-expressing retroviruses. Expression values are normalized for each probeset across samples within each cell line with a mean of 0 and an s.d. of 1. Darkest blue indicates 3 s.d. below the mean; darkest red indicates 3 s.d. above the mean.

While the gene array is a powerful high-throughput tool for identifying changes in gene expression, it does not always correlate quantitatively or qualitatively with a qPCR approach. Indeed, microarray data frequently underestimates fold-changes relative to qPCR [164]. Thus, to validate that Notch does in fact regulate PCGF5, mRNAs where prepared from cell lines treated with Cmpd E or DMSO vehicle control and these were reverse transcribed, then subjected to qPCR. Primers were designed to span the splice junction of exon 9 and 10 to eliminate the possibility of amplifying contaminating genomic DNA. PCGF5 mRNA was consistently found to decrease upon Notch inhibition in human T-ALL cell lines, including ALLSIL, DND-41 and KopTK-1 shown here (**Figure 3.2.3**).
To determine whether this finding is conserved between species, PCGF5 mRNA was measured in the mouse cell line T6E after mock treatment or treatment with Compound E (Figure 3.2.3). T6E cells were derived from a mouse leukemia induced by infection of bone marrow progenitors with an activated mutant of Notch. This mutant, NotchΔE lacks most of the extracellular domain but, unlike the ICN mutant which lacks the transmembrane region and is thus translated as a cytosolic protein, NotchΔE
remains tethered to the membrane and requires γ-secretase for its activity. PCGF5 was found to be downregulated upon Compound E treatment in T6E cells demonstrating that the regulatory link between Notch and PCGF5 is conserved between human and mouse.

Since γ-secretase is known to act on substrates other than Notch [165], it was possible GSI was having an effect on gene expression that was independent on Notch. To account for this, data was analyzed from a study that employed dominant-negative Mastermind-like-1 (DNMAML1), which forms a complex with Notch and CSL but is not competent to induce transcription. This provides a means of inhibiting Notch independent of γ-secretase inhibition which should eliminate non-specific drug effects from the analysis. In cells pre-treated for three days with Compound E, expression of DNMAML1 causes reduced and delayed induction of PCGF5 after the drug is removed (Figure 3.2.4A).

Another study utilized an alpha-helical stapled polypeptide, SAHM1, to inhibit Notch. SAHM1 is a synthetic polypeptide mimetic of residues 21 to 36 of human MAML1, a component of the Notch transactivation complex [4]. The SAHM1 helix contains hydrocarbon staples formed between non-natural amino acids incorporated substituted into two positions separated by three amino acids. In this way, the two substituted positions are separated by one turn of the helix and are thus brought into close proximity and subject to chemical cross-linking. The resulting ‘stapled’ polypeptide forms a more stable helix and possesses higher metabolic stability [166] than unmodified polypeptides.

SAHM1 is thought to function by forming a complex with Notch and CSL thereby excluding MAML1, thus rendering the complex inactive. In this case, PCGF5 was reported to be decreased by 3 of 4 probesets in KopTK-1 (Figure 3.2.4B). The downregulation was much less impressive, however, for HPB-ALL.

Collectively, these data support the conclusion that PCGF5 transcript levels correlate with Notch signaling in a variety of T lineage cell lines in both human and mouse.
Figure 3.2.4 – Effect of alternative methods of Notch inhibition on PCGF5 expression

A) CUTLL1 cells transduced with either empty vector or dominant negative Mastermind-like-1 (DNMAML1) were treated for 3 days with 1 μM Compound E to block Notch signaling. The Compound E was then washed out and samples were collected 2 hours (2h) and 4 hours (4h) after washout (dataset GSE29544 [2]). B) KopTK-1 and HPB-ALL cells were treated with the synthetic stapled polypeptide SAHM1 for 24 hours (dataset GSE18198 [4]). Samples were then subjected to expression profiling by Affymetrix array and data for PCGF5 probesets was visualized using dChip. Expression values are normalized for each probeset across all samples with a mean of 0 and an s.d. of 1. Darkest blue indicates 3 s.d. below the mean; darkest red indicates 3 s.d. above the mean.
3.2.2 Notch activation of PCGF5 expression does not require do novo protein translation

While the array experiment was designed to exclude genes whose regulation by Notch was indirect via c-myc it remained to be seen whether PCGF5 mRNA levels were being modulated by the action of another protein whose expression was dependent on Notch. To exclude this scenario, DND-41 cells treated for 24 hours with Compound E were released from Notch inhibition by a washout with media devoid of Compound E but with or without cyclohexamide (CHX), a potent inhibitor of protein translation. Cells treated with CHX are unable to synthesize new protein and thus, in this experiment, any genes induced by Notch are unable to exert effects via their protein product. PCGF5 was found to be induced after Compound E wash-out to equal measure in the presence or absence of CHX suggesting that its induction is a direct consequence of Notch1-mediated transactivation (Figure 3.2.5A).

This finding was further corroborated by expression profiling data, from dataset GSE29544 on GEO, of the CUTLL1 human T-ALL cell line [2]. In this study cells were treated with Compound E then the drug was withdrawn in the presence or absence of CHX. PCGF5 message was found to recover about 50% of the Compound E-induced diminution four hours after Compound E washout in the CHX treated sample, although the control recovered to baseline at this time point (Figure 3.2.5B).

3.2.3 Notch binds upstream of the PCGF5 locus

The CHX experiments were consistent with direct regulation of PCGF5 by Notch so it was of interest to determine whether components of the Notch transactivation complex bind the PCGF5 locus. Initiation of transcription of PCGF5 occurs at two alternative promoters in both human and mouse. These promoters produce different first exons, but splice to the same second exon which encodes the start codon of PCGF5. Therefore, both promoters produce an identical coding sequence. In human the two transcription start sites (TSSs) are separated by about 57 kb of genomic sequence (Figure 3.2.6).
**Figure 3.2.5 – Notch regulation of PCGF5 does not require de novo translation**

A) RT-qPCR for PCGF5 from cDNA of DND-41 cells cultured in media with Compound E for 24 hours then the inhibitor was washed out and replaced with either control media or media containing 20 μM cyclohexamide (CHX). Cells were collected at 8 hrs and 24 hrs post wash-out. PCR performed in triplicate and normalized to β-actin and shown relative to Cmpd E-treated sample. Error bars indicate the standard deviation. * indicates p < 0.05; n.s. indicates p > 0.05 as determined by student’s t test.

B) CUTLL1 cells were treated in triplicate as 3 independent cultures for 3 days with 1 μM Compound E then the drug was washed out and replaced with either control media or media containing 20 μM CHX. Samples were collected 2 hours (2h) and 4 hours (4h) after washout as indicated. Samples were then subjected to expression profiling by Affymetrix array and data for PCGF5 probesets was visualized using dChip. Expression profiling data from GSE29544 [2]. Expression values are normalized for each probeset across all samples with a mean of 0 and an s.d. of 1. Darkest blue indicates 3 s.d. below the mean; darkest red indicates 3 s.d. above the mean.
A previously reported ChIP-seq dataset for Notch1 and CSL from the T-ALL cell line CUTLL-1 was obtained (GSE29600) [2]. This enabled the identification of regions in the genome which with Notch1 and CSL associate. A region near PCGF5 was listed in the supplemental data of Wang et al.’s paper as being enriched by Notch1. In their analysis, they detected Notch1 binding 180kp upstream of the PCGF5 gene, although CSL was not detected at this site. The authors note that they observed many Notch1-only regions throughout the genome where CSL was not detected and suggest this might be due to inaccessibility of CSL by the antibody due to hindrance by the extended complex.

**Figure 3.2.6 – Exon structure of human and mouse PCGF5 transcripts**  Schematic of PCGF5 transcripts. Coding exons are depicted as filled boxes, non-coding exons are depicted as open boxes. Transcription of PCGF5 can be initiated at two different promoters giving rise to the same coding sequence. The asterisk (*) marks exon 9 which is omitted in variant mouse transcripts.
Since the published data only reported the nearest enriched site, the data was analyzed from the unprocessed aligned sequence files using the MACS (Model-based Analysis of ChIP-Seq) peak calling algorithm [153] to identify SERs with a p-value $< 10^{-7}$. This analysis detected the region listed by the authors of the study [2], indicated in Figure 3.2.7 by the arrow with the asterisk, as well as several other

**Figure 3.2.7 – Notch binds near the PCGF5 locus** Schematic of ChIP-seq enrichment for the immunoprecipitated targets indicated at left. Notch SERs are indicated by the arrows above. Grey arrows are Notch1 SERs only, black arrows regions significantly enriched by both Notch1 and CSL. The arrow with the * indicates the SER listed Wang et al. paper [2]. The PCGF5 gene is indicated below by the black line with TSS sites indicated by arrows, and neighbouring genes are indicated. Y-axes are scaled to top of highest peak for each track and represent tag fold enrichment over control. Dataset GSE29600; experiment performed on CUTLL1 cells.
Notch1 SERs in the region of *PCGF5*. CSL was also found to significantly enrich one SER as well as another region upstream. Consistent with these sites serving enhancer functions, H3K4me1 was observed to be abundant while H3K4me3 was low in this region (**Figure 3.2.7**).

**Figure 3.2.8 – Notch1 and CSL enrich regions at or near the PCGF5 promoters** ChIP-seq enrichment tracks for Notch1, CSL, H3K4me3, H3K4me1 and H3K27me3 are shown in a 200kb window around the PCGF5 gene. The grey arrows indicate regions enriched by Notch1. The cartoon below indicates the location of exons and the two promoters which are marked by heavy H3K4me3. Y-axes are scaled to top of highest peak for each track and represent tag fold enrichment over control. Dataset GSE29600; experiment performed on CUTLL1 cells.
One of the Notch1 SERs identified by the MACS analysis was located at the 5′ *PCGF5* promoter and another was approximately 15 kb to the 3′ direction of the TTS of *PCGF5* but neither location showed CSL enrichment that met the $10^{-7}$ cut-off (see grey arrows in Figure 3.2.8). Both *PCGF5* promoters showed high levels of H3K4me3 which is suggestive of active transcription.

A similar ChIP-seq experiment was conducted for the mouse T-ALL cell line T6E. Notch and CSL were both found to enrich a region approximately 124 kb upstream of the *PCGF5* transcription start site (Figure 3.2.9). The 3′ promoter was enriched by Notch1 ChIP in the mouse cells. This site was enriched by Notch1 but not by CSL (Figure 3.2.10).
3.2.4 Notch1 and CSL binding of PCGF5 locus is significantly higher than expected by chance

To determine in an unbiased fashion whether the distribution of Notch1 and CSL SERs was unusual in the genome or whether many genes were similarly enriched in the ChIP-seq experiment, we measured two parameters: the distance from PCGF5 to the nearest Notch1 and CSL SERs, and the density of SERs in the region of the PCGF5 gene. This analysis was completed using the Genomic Regions Enrichment of Annotations Tool [3].

First, the distance from the (TSS) to the nearest Notch or CSL site for every gene in the RefSeq list was calculated. The median distance from a genic TSS to the nearest Notch1 site was 65645 bp in the upstream direction and 63142 bp downstream. For CSL the median distances were 204938 bp upstream and 232032 bp downstream. The greater distances for CSL are likely due, in part, to the fact that there are fewer CSL SERs overall.
The TSS of PCGF5 for the RefSeq transcript (NM_032373), which initiates at the 3’ promoter, is 62495 bp from the nearest Notch1 site and 33586 bp from the nearest CSL site. Comparing this with the rest of the genome shows that the distance to the nearest Notch1 SER is near the median for all genes, and the distance to the nearest CSL site is closer than 91% of genes. For the transcripts that initiates at the 5’ promoter (NM_001257101.1), one Notch1 peak overlaps the TSS, centred 224 bp upstream, while the nearest CSL site is 23910 bp downstream. For this promoter, the proximity to its nearest Notch1 is nearer than 89% of genes, while that figure for CSL is also 89%.

Figure 3.2.11 – SERs for Notch1 and CSL are abundant near the PCGF5 locus Histogram showing 17,744 transcription starts sites (TSS) ordered from most SERs at left to least at right. Number of SERs for Notch1 (red line) and CSL (blue line) within 300kb of a RefSeq TSS are plotted. In this analysis, the PCGF5 locus had 7 SERs for Notch1 (†) and 5 for CSL (X). Analysis was completed using Genomic Regions Enrichment of Annotations Tool [3].

The TSS of PCGF5 for the RefSeq transcript (NM_032373), which initiates at the 3’ promoter, is 62495 bp from the nearest Notch1 site and 33586 bp from the nearest CSL site. Comparing this with the rest of the genome shows that the distance to the nearest Notch1 SER is near the median for all genes, and the distance to the nearest CSL site is closer than 91% of genes. For the transcripts that initiates at the 5’ promoter (NM_001257101.1), one Notch1 peak overlaps the TSS, centred 224 bp upstream, while the nearest CSL site is 23910 bp downstream. For this promoter, the proximity to its nearest Notch1 is nearer than 89% of genes, while that figure for CSL is also 89%.
Since this analysis addressed only the single nearest site, the density of Notch1 and CSL sites over the PCGF5 locus was compared to the rest of the genome. Notch1 and CSL sites were enumerated within 300 kb of the TSS for each RefSeq gene. The region around the TSS of PCGF5 was found to have 7 Notch1 SERs and 5 CSL SERs. When compared against the rest of the genome only 0.8% of RefSeq genes have more Notch1 sites, while 0.7% of genes have more CSL SERs (Figure 3.2.11).

Thus, both Notch1 and CSL bind near the PCGF5 locus in multiple locations including, for Notch1 at least, one of two PCGF5 TSSs, although no CSL SERs are located uncommonly close to either TSS. However, the abundance of Notch1 and CSL sites in the region around PCGF5 is much higher than expected by chance.

### 3.2.5 PCGF5 expression in T-ALL cases

Since the experiments to this point had been performed in cell lines, publically available datasets for human T-ALL patient samples were consulted to gain insight into the profile of PCGF5 expression in primary T-ALL. Expression microarray data published by Homminga et al. allowed the comparison of 117 cases of human T-ALL of various cytogenetic classes and also included 7 normal bone marrow samples [1]. PCGF5 expression was relatively consistent between the cytogenetic subtypes though the HoxA-translocated samples were slightly higher than the bone marrow control, the TAL1 subtype samples and the cases of unknown cytogenetic class (p-value < 0.05) (Figure 3.2.12).

One important question was whether PCGF5 correlated in any way with outcome or whether its expression was altered upon relapse. Gene expression microarray data was analyzed for 41 paired diagnosis and relapse samples from GEO dataset GSE18497 [167] to determine whether PCGF5 expression changes upon re-emergence of disease. While there was considerable variation between cases, there was no significant change between diagnosis and relapse samples (Figure 3.2.13A).
In another study, samples were taken at diagnosis from patients and these were later categorized into those which responded to therapy and those for which induction therapy failed \[168\]. Relapse samples were also taken from patients whose disease returned. Analysis of this dataset (GSE14618) revealed no significant differences between cases which responded to treatment and those that did not nor was there any change at relapse (Figure 3.2.13B).
So while PCGF5 mRNA levels did not seem to vary between subtypes or stages of T-ALL, it remained to be ascertained whether PCGF5 expression was specific for T-ALL compared with other hematopoietic malignancies. Comparisons to acute myelogenous leukemia (AML) and B cell acute lymphoblastic leukemia (B-ALL) were conducted for the gene expression microarray datasets GSE14479 and GSE18497 respectively [167, 169]. PCGF5 expression was found to be significantly higher (p-value < 0.0001) in T-ALL compared with AML and B-ALL.

**Figure 3.2.13 – PCGF5 expression does not correlate with disease progression in T-ALL**

PCGF5 signal intensity for probeset 226326_at from dataset GSE18497 (A) and GSE14618 (B). In the left panel, Diagnosis indicates samples collected at disease presentation and Relapse indicates samples acquired from patients who had undergone standard treatment but the disease had returned. In the right panel, categories represent samples taken at diagnosis from patients that achieved remission (Response; n=30) and those for whom induction therapy failed (Failure; n=6). Relapse samples (n=14) were taken from patients who had undergone standard treatment but the disease had returned. Values shown as arbitrary units (A.U.). Lower and upper limits of the boxes represent the 25th and 75th percentiles, the line through the box indicates the median and the whiskers indicate the minimum and maximum value within the dataset. Tested for significant differences by student’s t-test for A and by 1 way ANOVA followed by Neuman-Keuls pairwise comparison. No significant differences observed.
ALL compared with either AML (Figure 3.2.14A) or B-ALL (Figure 3.2.14B) consistent with its regulation by Notch and the high Notch activity in T-ALL cells.

Thus, PCGF5 is expressed at a higher level in T lineage leukemia cells than leukemias originating from other cell types, but does not appear to correlate with outcome.

3.2.6 Tissue distribution

To determine in which tissues PCGF5 is expressed, a large dataset of murine tissues was analyzed (NCBI GEO dataset GSE10246) [170]. Within the mouse hematopoietic compartment, PCGF5 mRNA is observed to be highest in HSCs and remains moderately high in common myeloid progenitors (CMP) and
granulocyte monocyte progenitors (GMP). In mature cells, PCGF5 is high in T lymphocytes but is very low in B cells, NK calls and Mac1, Gr1 positive granulocytes (Figure 3.2.15).

To delineate PCGF5’s regulation in finer resolution with respect to T lymphopoiesis additional datasets from the GEO database were consulted. Analysis of gene expression microarray data for populations from mouse thymus showed highest PCGF5 expression in double-negative cells and lowest expression in the double-positive population. Thymocytes that were single positive for CD4 or CD8 expressed levels intermediate between DN and DP [171] (Figure 3.2.16, left panel). To achieve yet greater resolution among thymic subsets, data published by Shi et al. [172] were consulted. In their study, thymocytes were sorted into DN1, DN2, DN3, preDP (which they defined as lin- CD24+ CD25- CD44lo) and DP

Figure 3.2.15 – PCGF5 expression in mouse hematopoietic cells Expression data for PCGF5 in hematopoietic stem cells (HSC), Granulocyte Macrophage Progenitors (GMP), Common Myeloid Progenitors (CMP), Megakaryocyte Erythroid Progenitors (MEP), Natural Killer cells (NK), B cells, Granulocytes (Gran) CD4 (CD4 T cell) and CD8 (CD8 T cell) single positive peripheral T cells. GEO dataset GSE10246. Expression values are normalized for each probeset across all samples with a mean of 0 and an s.d. of 1. Darkest blue indicates 3 s.d. below the mean; darkest red indicates 3 s.d. above the mean.
populations. This analysis revealed increasing PCGF5 expression from DN1 peaking at the preDP stage before declining sharply in DP cells (Figure 3.2.16, right panel).

![Figure 3.2.16 – PCGF5 expression in mouse thymic subsets](image)

In humans PCGF5 mRNA is high in intrathymic T progenitors and remains elevated in DP. Expression is observed to be lower in both CD4 SP thymocytes and peripheral adult CD4 T cells but was found to be high in cord blood CD4 T cells (Figure 3.2.17).

Thus, the identification of PCGF5 as a Notch target in T-ALL seems to be no coincidence as it is expressed predominantly in T-lymphoid tissues, both normal and leukemic, and its expression correlates well with known patterns of Notch activity.
3.2.7 Splice variants and genomic organization

The human PCGF5 gene maps to the long arm of chromosome ten and spans 63.7 kb. There are several transcript variants for PCGF5 described in human. Two of these use the canonical promoter and two use an alternative promoter within the first intron of the longer transcript. In mouse, there is an additional transcript which does not use exon 9, but instead splices exon 8 to exon 10 (Figure 3.2.6). To determine whether this variant was specific to mouse, primers were designed to amplify the coding region of PCGF5. PCR was conducted on cDNA from human T-ALL cells as well as mouse thymocytes and two bands were detected: one corresponding to the full-length species containing exon 9 and the other consistent with non-usage of exon 9 (Figure 3.2.18). The predicted sizes of these variants are 796 bp and 736 respectively. The PCR amplicons were eluted from the gel, cloned and sequenced which

![Figure 3.2.17 – PCGF5 expression in human thymic and peripheral T cell subsets](image)

Expression data for PCGF5 in intrathymic T progenitor cells (ITTP) DP, CD4 SP thymocytes and CD4 SP T cells from adult blood (Adult CD4 SP) and cord blood (Cord CD4 SP). Data from GSE1460. Expression values are normalized for each probeset across all samples with a mean of 0 and an s.d. of 1. Darkest blue indicates 3 s.d. below the mean; darkest red indicates 3 s.d. above the mean.
confirmed their identities. The skipping of this exon yields a size difference of 22 amino acids in the final protein.

3.3 Discussion

3.3.1 Notch transcriptional regulation of PCGF5

In an effort to identify transcriptional targets of the Notch pathway in T-ALL, expression profiling was performed on a panel on T-ALL cell lines that depend on Notch for their growth. This experiment was designed with the intention of discovering direct targets of Notch other than myc. Three strategies were employed in this study to enrich for direct targets:

1. Short duration of Compound E treatment
2. Compound E treatment of myc-transduced cells

Figure 3.2.18 – Detection of PCGF5 splice variants in T-ALL cells by RT-PCR At left, agarose gel of RT-PCR products from ALLSIL human T-ALL cells (Human) and mouse B6 thymocytes (Mouse) using primers specific for the coding region of PCGF5. The sizes of the upper amplicon is consistent with full length PCGF5 (predicted size 836 bp) and the lower is consistent with the Δexon9 transcript lacking exon 9 (predicted size 770 bp).
3. **Compound E treatment of ICN-transduced cells**

Firstly, the Compound E treatment was a relatively short 12 hours. Since the half-life of activated Notch is several hours, particularly for PEST deleted alleles [173], a 12 hour Compound E treatment results in a ‘Notch-off’ state of only a few hours. Notch is known to regulate the expression of many genes, some of which are transcription factors or repressors themselves, so it was hoped this short time course would limit change in the protein level of other transcriptional modulators secondary to Notch inhibition. In this way, this experiment would favour the detection of direct Notch targets, rather than genes whose expression is modulated secondarily by the action of a third gene whose expression requires Notch.

The second strategy employed to enrich for direct targets was to include cells transduced with retroviruses driving expression of myc treated with Compound E or vehicle control. Since c-myc itself is a target of Notch signalling and is a potent transcriptional activator, described to regulate as many as thousands of genes [174, 175], some of the changes observed upon Notch inhibition could be attributed to decreased myc levels. Comparison of the Compound E-treated samples in parental cells, in which myc expression would decrease, to those transduced to express myc constitutively would allow myc targets to be culled from the list of putative direct Notch targets.

The third strategy was the treatment with Compound E of cells transduced with retroviruses containing ICN. Since ICN contains no transmembrane domain and is translated as a cytosolic protein it is insensitive to γ-secretase inhibition. As such, comparison of parental cells treated with Compound E to similarly treated ICN-expressing cells would help discriminate true Notch targets from genes whose regulation is affected by administration of Compound E in a Notch-independent manner.
Figure 3.3.1  Samples cluster by cell line not Notch signalling status  Unsupervised clustering of the 30 samples over the 6009 probesets in the filtered list. Cell line identities are indicated at top. Expression level for each probeset is normalized across all samples. Expression values are normalized for each probeset across all samples with a mean of 0 and an s.d. of 1. Darkest blue indicates 3 s.d. below the mean; darkest red indicates 3 s.d. above the mean.
The critical consideration in this analysis was directionality. Since the five model cell lines each harbour different mutations of Notch and, therefore, have heterogeneous levels of endogenous Notch signalling [5] it was the directionality, namely that expression was higher in ‘Notch on’ states that in ‘Notch off’ states that was important. The supervised analysis was designed to identify genes that met that condition. We noted that unsupervised analysis to identify samples with a similar pattern of gene expression led to clustering by identity of the parent cell lines rather than by transgene or drug treatment (Figure 3.3.1).

As can be seen in Figure 3.3.1, magnitude of the differences in gene expression between cell lines is vastly greater than the differences between treatments within a single cell line. This meant that any random permutation of samples would be susceptible to a high FDR due to unequal distribution of the cell lines between group. By omitting the transduced samples and conducting a paired t-test we were able to mitigate this problem. Median and 90th percentile FDR assessed by this method were 0% and 20%, respectively. Importantly, the correlation between PCGF5 and Notch was retained in this analysis (Figure 3.3.2).

A study by Palomero et al. [159] encountered similar challenges with variation between cell lines. To compensate for these variations in the hopes of identifying transcriptional targets of the Notch pathway, they normalized the microarray expression values by the amount of active intracellular Notch as measured by western blot. Their method is described in their manuscript:

A metagene entry proportional to the concentration of activated NOTCH1 protein in the nucleus was constructed by using stepwise regression from quantitative measurements of ICN1 protein level in T-ALL cell lines, was integrated into the microarray gene expression profile measurements, and was used to build an ARACNe network for the NOTCH1 signaling hub...[159]
I subjected the Palamero et al. data [159] to a comparison of mock-treated samples to Cmpd E-treated samples to determine if PCGF5 was also on their list. The gene list was filtered for probesets was variation across samples (0.4 < Standard deviation / Mean < 10) and expression level (signal >= 100 in > 50% of samples). Genes that were 1.2-fold higher and had an expression value 100 units higher in the

**Figure 3.3.2 – Paired comparison of 5 cell lines treated with Cmpd E or vehicle control**

Untransduced samples from Figure 3.2.1 subject to a supervised paired analysis for genes that are higher in control (Cmpd E -) than in γ-secretase inhibited (Cmpd E +) cell lines (> 1.2-fold; change in expression level >= 100; p < 0.05). Expression values are normalized for each probeset across all samples with a mean of 0 and an s.d. of 1. Darkest blue indicates 3 s.d. below the mean; darkest red indicates 3 s.d. above the mean. A=ALLSIL; K=KopTK-1; D=DND-31; H=HPB-ALL; T=TALL-1
mock samples as compared to the Cmpd E samples and a paired t-test value < 0.05. The comparison identified 22 genes. Four probesets for PCGF5 were identified in that set (Figure 3.3.3). Testing 100 randomized permutations found the median and 90th percentile FDR were both 0%.

Crucially, all putative targets chosen for further study in our study were validated by single gene qPCR. Further, the fact that four probesets annotated to PCGF5 were identified in both our data set and the Palomero et al. dataset [159] suggests that the association between its expression and Notch signalling was robust.

This design of our microarray experiment enabled the identification of candidate Notch target genes in five isogenic models, a benefit not possible in the analysis of primary human or animal samples. As the unsupervised analysis showed (Figure 3.3.1), the variability between cell lines, caused by genotype of the patient from which it was derived, cell of origin from which the tumour was transformed, and history of in vitro selection, among other factors, exceeds the variability introduced by the experimental intervention. Thus, the genes identified in using the supervised clustering of "Notch On vs Off" specimen classes may have been obscured by factors outside of the experimental intervention due to varying genetic background of the 5 different cell lines. Accordingly, we proceeded to conventional single gene qRT-PCR methods to validate PCGF5 as a bona fide Notch target gene.
**Figure 3.3.3 – Paired comparison of 10 cell lines treated with Cmpd E or vehicle control**

Gene expression data from GSE5716 subject to a supervised paired analysis for genes that are higher in control (Cmpd E-) than in γ-secretase inhibited (Cmpd E+) cell lines (>1.2-fold; change in expression level >= 100; p < 0.05). Expression values are normalized for each probeset across all samples with a mean of 0 and an s.d. of 1. Darkest blue indicates 3 s.d. below the mean; darkest red indicates 3 s.d. above the mean.
The observation in the expression profiling experiment that PCGF5 mRNA decreases upon Compound E treatment in the parental cells and the myc-transduced cells, but not the ICN-transduced cells, leads to the following conclusions:

1. PCGF5 expression correlates with Notch signalling activity
2. PCGF5 expression is not affected non-specifically by Compound E
3. PCGF5 expression is not maintained by myc in the absence of active Notch signalling

PCGF5 expression was found to be linked to Notch signalling in a dataset from the T-ALL cell line, MOLT4, as well (GSE6495) [176].

It was subsequently observed that PCGF5 is induced upon release of Notch from γ-secretase inhibition in cells treated with CHX. Since CHX treatment blocks translation, mRNA whose transcription is induced via the activity of Notch cannot be translated into protein. Therefore, Notch is able to induce increased PCGF5 expression without the benefit of any of its target genes being translated into protein, suggesting a direct action of Notch on the PCGF5 locus. The CHX experiment, however, does not rule out mechanisms of regulation that do not depend on protein translation, such as microRNAs.

It should be noted that in the qPCR experiment in DND-41 cells no control was included to confirm CHX was active at inhibiting translation. For instance, a western blot for a known target of Notch, such as myc or Hes, could lend support that CHX was functioning as expected if increase in Notch target protein was observed in control samples but not CHX-treated samples. While this confirmation was not done, supervised analysis of the dataset showed large changes in gene expression upon CHX treatment suggesting, at least, that the drug is having an effect.
3.3.2 Notch and CSL binding of the PCGF5 locus

The ChIP-seq data published by Wang et al. identified a region 136 kb upstream of the PCGF5 locus that was enriched by Notch1 immunoprecipitation [2]. A more detailed analysis of this dataset was conducted, starting from the unprocessed list of reads aligned to the genome. This list contains all the aligned sequence reads detected without any statistical validation to suggest whether their presence reflects a significant enrichment of that location by the immunoprecipitation. The MACS algorithm [153] was employed to identify SERs over input controls provided in the that met a threshold p-value of $10^{-7}$. The authors used the QuEST statistical framework [177]. While both algorithms have some similar features – they each align forward and reverse strands separately to estimate the shift for each dataset do to sequence reads deriving from the 3’ end – subtle differences may account for some of the variation in interpretation.

While the published analysis did not determine the presence of CSL at the upstream Notch1 SER, a CSL SER was detected at this site in subsequent analysis. This analysis also detected another region enriched by both Notch1 and CSL approximately 80 kb upstream of the site listed by Want et al. If Notch1 were to regulate PCGF5 from this distance, it would be via an enhancer effect rather than direct transactivation of the promoter. Consistent with this function, high levels of H3K4me1 are observed extending out from the Notch1 site, while H3K4me3 is low in the region around this site (Figure 3.2.7). While histone acetylation and histone depletion occur at both promoters and enhancers, high levels monomethylated H3K4 along with low trimethylated H3K4 is strongly predictive of enhancers [178].

It was also noted that, while many of the Notch1 SERs were no closer to PCGF5 than neighbouring genes, the PCGF5 locus had low levels of H3K27me3 while neighbouring genes had high levels of this repressive mark. Further, the gene immediately 3’ of PCGF5 showed no evidence of H3K4me3 at its promoter, despite being flanked by two Notch1/CSL SERs within 50 kb of its TSS, further evidence its
transcription was not being activated by Notch. This suggests that, if the Notch and CSL sites are activating transcription in this region of the genome, their effects may be targeted towards PCGF5.

Given the difficulty in assigning Notch1 SERs to PCGF5 specifically, the distribution of Notch1 and CSL peaks around the PCGF5 locus was compared to the rest of the genome. While it was the case that many genes had Notch1 and CSL SERs nearby (the median distance for all RefSeq genes to the nearest SER was about 65 kb for Notch1 and about 200 kb for CSL) the density of Notch1 and CSL SERs in the region of PCGF5 was extremely uncommon—indeed, PCGF5 is in the 99th percentile of RefSeq genes ranked by abundance of SERs for each Notch1 and CSL within 300 kb. Other known targets of Notch, such as Notch1 itself, Hes1, Hes4, Hes5 and myc each had multiple Notch1 and CSL SERs within 300 kb or their promoters.

The presence of enhancer-like Notch1 sites near PCGF5 provides a mechanism for Notch’s regulation of PCGF5. Indeed, the authors correlated their ChIP-seq data with expression profiling data (GSE29544) in the same cells treated with Compound E to block Notch signalling. They found that genes which Notch bound as an enhancer were more highly regulated by Notch inhibition that those for which Notch bound at the promoter. That said, it is interesting to note that analysis of the Wang et al. data [2] shows SERs for Notch1 at the 3’ for PCGF5, although no SER for CSL was detected at this location (Figure 3.2.8).

The authors posit that their inability to detect CSL binding at all Notch1 SERs could be due to steric hindrance of the epitope by other members of the complex that occurs at some, but not all loci [2]. As Notch is not known to bind directly to DNA, this seems a likely explanation. Such steric hindrance might result in inability to precipitate some types of multimeric complexes, or may simply result in a global reduction in signal strength for CSL ChIP-seq peaks. Both mechanisms would result in fewer peaks detected overall, which is what is observed.
Comparison of this data with the mouse T6E cell line, which models a Notch-driven T-ALL, shows the presence of an upstream enhancer-like region, bound by both CSL and Notch. Since the authors did not include histone marks in their study of the mouse cells, it cannot be concluded that this site has the epigenetic characteristics of an enhancer, but it is otherwise analogous to the human case.

In any case, be it via a distal enhancer or direct binding of the promoter, the ChIP-seq data corroborates the Notch inhibition effect on PCGF5 mRNA levels and ties the Notch signalling pathway to the transcriptional regulation of PCGF5.

### 3.3.3 PCGF5 expression in cancer

Publically available expression profiling data were employed to determine whether PCGF5 expression correlated with leukemia subtype or outcome. In contrast with Bmi1 [179-181], PCGF5 expression does not seem to correlate with treatment success, nor is its expression altered during relapse. It does, however, show some specificity for T-ALL as its expression is significantly higher in that disease than in AML and B-ALL. This could simply be a consequence on its regulation by Notch as active Notch is a hallmark in T-ALL but not in AML or B-ALL. Whether its expression is important for the growth or survival of T-ALL cells remains to be determined.

While dysregulation of PcG proteins has been seen in cancer, particularly for Bmi1 [182], mutations of PcG genes have been observed as well. The PRC2 methyltransferase EZH2 was found to be commonly mutated in follicular lymphoma and diffuse large B cell lymphoma [183, 184]. Nearly all of these mutations involved the tyrosine at position 641. Mutations have been observed for many other PcG members, although they are rare, and identical mutations have not been observed in multiple cases, as they have for EZH2, suggesting these may not be important in disease pathology. To date, PCGF5 itself has not been observed to be mutated in cancer.
Had mutations of *PCGF5* been observed in T-ALL or had a bias in its expression been measured toward some cytogenetic types of T-ALL over others, tenable hypotheses might have been formed regarding its potential role in this disease. However, since Notch1, which is frequently activated by mutation across all types of T-ALL and is a critical driver of T-ALL, induces PCGF5 expression, it is perhaps natural that expression levels are similar in all T-ALL subtypes. Likewise, if PCGF5 were required for T-ALL growth, its expression is assured by the activation of Notch, so no mutation or ‘aberrant’ upregulation (gene amplification, translocation, etc.) would be required. Thus, further studies are required to determine if *PCGF5* has role in T-ALL biology that might be revealed by loss-of-function.

### 3.3.4 PCGF5 and Notch linked in normal development

Analysis of public gene expression datasets reveals that PCGF5 is highly expressed in cells of the T lymphoid lineage. This observation, too, ties *PCGF5* to Notch in the context of development. Notch drives lymphoid fate over myeloid and T lymphoid fate at the expense of B. Once in the thymus, Notch signalling is critical during the DN stage and loss of Notch results in a failure for cells to undergo β-selection. However, once β-selection is achieved the successfully rearranged TCRβ chain associates with pre-Tα providing the cell with a TCR signal, Notch1 is downregulated via the induction of the repressor Id3 [185].

In comparison, PCGF5 expression is moderate upon immigration to the thymus and increases throughout the development during the double-negative stages peaking at the stage immediately following β-selection. This roughly parallels the expression pattern for Notch1, albeit delayed slightly, as Notch1 peaks in the DN3 stage whereas PCGF5 peaks after DN3. As previously mentioned, Notch protein has a half-life of several hours [173] and if PCGF5 mRNA has a long half-life it may persist for long after transactivation of its promoter ceases which could explain why PCGF5 message peaks after
that of Notch1. Thus, the expression pattern of PCGF5 in developing thymocytes is consistent with its expression being primarily driven by Notch.

PCGF5 mRNA is also elevated in hematopoietic stem cells and CMPs and GMPs, although the role of Notch is less clear, particularly for CMPs and GMPs, in the physiology of these cells. The robust expression of PCGF5 in these immature cells is potentially of interest. Since many cancers display a transcriptional profile consistent with an immature, undifferentiated phenotype, it is possible elevated levels of PCGF5 mRNA driven by activated Notch could help confer characteristics associated with undifferentiated cells, such as self-renewal, to T-ALL.

3.3.5 Conservation between human and mouse

PCGF5 expression was found to depend on Notch signaling in both human and mouse. The amino acid sequence is similarly conserved between the two species. Only six amino acid substitutions are found between the full length human and mouse PCGF5 proteins. Four of these are encoded by the alternatively spliced exon 9, suggesting that this often excluded sequence is subject to less evolutionary constraint than the rest of the coding sequence.
Chapter 4  PCGF5 overexpression does not alter steady state hematopoiesis but results in reduced myeloid colony output by bone marrow progenitors

Contribution

I designed, executed, and analyzed the data for all experiments in this chapter. The mice were tended by the staff of the Animal Resource Centre at the BC Cancer Research Centre.

4.1 Introduction

4.1.1 The role of Notch in lymphopoiesis

The Notch signalling pathway has numerous roles in hematopoietic development and Notch1 is required for the generation of T cells in mice. Loss of Notch1 in thymocytes leads to a failure to develop beyond the CD4, CD8 double-negative (DN) stage as well as a substantial decrease in the absolute numbers of cells in the thymus [9]. Loss of Notch was also seen to result in an accumulation of B cells in the thymi of mice in which Notch1 had been deleted, while myeloid, erythroid and NK populations in the bone marrow were found to be unaffected.

Notch’s role in T cell development is manifold. First, Notch skews lymphoid progenitors toward a T fate at the expense of B [9, 186]. Notch is also required for β-selection, in part, due to its transactivation of the pre-TCRα gene [163]. This is not Notch’s only function in T cell development, however, as addition of a pre-rearranged TCRβ does not render Notch dispensable for T cell development. TCRβ-expressing RAG-/− cells differentiate into double-positive (DP) thymocytes in the presence of the Notch ligand delta-like 1 (DL1) but fail to differentiate, and indeed suffer a loss in cellularity, in the absence of a Notch signal [187, 188]. Notch also provides a metabolic stimulus to developing T cells via the PI3K-Akt pathway and
this function can be restored in the absence on Notch signalling by expressing constitutively active Akt [188].

The importance of Notch in T cell development was underscored by the gain-of-function experiment in which hematopoietic progenitors were transduced with retroviruses carrying a constitutively active mutant of Notch transplanted into recipient mice. Analysis of these mice revealed that Notch was driving T lymphoid differentiation in the marrow independent of the thymus – indeed over 10% of the mononuclear cells in the marrow of these mice were double positive immature T cells. Overexpression of constitutively activated Notch, in contrast, led to a block in B cell development [189]. Persistent expression of activated Notch in hematopoietic progenitors not only leads to a skewing of cell fate from B toward the T lineage, but to neoplastic transformation [6].

However, while Notch is critical for induction of thymic differentiation, Notch mRNA is strongly downregulated following β-selection [185] and enforced expression of activated Notch in thymocytes which prevents modulation of the Notch signal impairs differentiation of DP thymocytes into mature CD4 or CD8 single positive T cells [190].

### 4.1.2 Notch and T-cell Acute Lymphoblastic Leukemia

As mentioned above, retrovirally driven activated Notch in bone marrow progenitors results in fully penetrant T-ALL in mice with a latency of between 8 and 25 weeks, depending on the allele of Notch used [6]. Spontaneous mutations in the Notch1 gene commonly arise in other genetic models of T-ALL suggesting a critical role for the Notch pathway in the pathogenesis of T-ALL [191, 192].

In humans, translocations resulting in a constitutively active allele of Notch were observed to occur very rarely [118]. This observation prompted closer examination of the Notch1 gene which revealed point mutations in approximately 50% of cases [5] of human T-ALL. These mutations fell into two distinct
classes: point mutations in the heterodimerization (HD) domain that result in amino acid substitutions and frameshift or nonsense mutations that result in a truncated protein lacking the C-terminal PEST domain which mediates degradation of ICN. The former class results in ligand-independent, or ligand-sensitized, cleavage of Notch from the membrane resulting in activation of the Notch signal. The latter class results in an increased half-life of the activated ICN polypeptide resulting in a net increase in Notch signal, although PEST mutants are thought to depend on ligand for their activation. Indeed, in reporter assays, HD mutants transactivate Notch reporters to a significantly higher degree than wild-type Notch1 in the absence of ligand, and when HD and PEST deletions are expressed in cis the activation is higher still [5].

Additionally, Fbxw7, the ubiquitin ligase responsible for degradation of intracellular Notch protein has been described to be mutated or deleted in 8.6% of T-ALL cases examined [193], thus revealing another means by which the Notch pathway may be activated in leukemia. Collectively, these genetic data provide strong evidence for a central role for Notch in the pathogenesis of T-ALL.

Inhibition of Notch signalling by treatment with GSI results in growth arrest of several human T-ALL cell lines [194] underscoring the continued importance of Notch in established T-ALL cells. The authors also reported a similar growth arrest when a truncated form of the Mastermind-like-1 (MamL1) was overexpressed in T-ALL cell lines. MamL1, along with CSL, forms a transactivation complex with intracellular Notch on DNA [122, 123] but the truncation mutant is not competent to cause transactivation and thus acts as a dominant negative. Indeed, Notch inhibition has been the subject of several clinical trials with the hope of providing more specific and less toxic therapies for T-ALL [195].

4.1.3 The role of Bmi1 and the Polycomb Group in lymphopoiesis

*PCGF5* was identified as a Notch-regulated gene in T-ALL cells and was of particular interest given its homology to *Bmi1*. *Bmi1* is a potent oncogene in hematopoietic cells and was discovered as a common
integration site in an Εμ-myc model of B cell lymphoma [27]. Retroviral integrations near the Bmi1 locus that activated transcription of Bmi1 were found in 35% of the lymphomas that arose in the mice. These integrations resulted in an increase of Bmi1 protein and this protein was predominantly nuclear. These findings suggested that Bmi1 cooperated with the myc transgene to increase the penetrance and decrease the latency of lymphoma in this model. Subsequently, Bmi1 has been implicated in many other forms of cancer [196-199] and its expression is associated with poor prognosis [179, 200].

Bmi1 also plays a critical role in T cell development as demonstrated by the involution of thymi in Bmi1-deficient mice. These mice also presented small spleens and hypoplastic bone marrow. The thymi of these mice contained an extremely high fraction of DN cells, though total cellularity was reduced approximately 8-fold [31]. The increased DN fraction has been attributed to a block in the DN3 stage [201]. Bmi1-deficient mice also have impaired stem cell self-renewal [202].

Bmi1 gain-of-function, in the form of an Εμ-Bmi1 transgene, caused subtle perturbations in the thymus. While there was no detectable change in mature peripheral T cells, a subset of double-negative cells in the thymus were found to display an aberrant CD25 dim phenotype. The bone marrow of these mice was found to contain an expanded population of pre-B cells with lower B220 surface expression than controls [203].

More generally, defects in hematopoiesis, and specifically of T lymphopoiesis, manifest as phenotypes of several PcG loss-of-function models. Mice deficient for the Bmi1 homologue Mel-18 show hyporesponsiveness of their lymphocyte progenitors to IL-7 resulting in a severe reduction of lymphoid cells in the bone marrow and spleen [204]. Interestingly, hematopoietic stem cell self-renewal is elevated in Mel-18-deficient mice [205]. So, while loss of Mel-18 or Bmi1 has overlapping phenotypes in some tissues, they seem to have opposing effects on hematopoietic stem cells.
Rae28, a mouse homolog of *Drosophila Ph*, also displays thymic hypotrophy, B cell developmental arrest and impaired self-renewal of stem cells, in addition to many other developmental defects, when deleted in mice [37, 206, 207].

While much is yet to be understood relating to the mechanism by which the PcG contributes to hematopoietic physiology, in the case of Bmi1, much of its function can be attributed to its repression of the tumour suppressors *Ink4a* and *Arf* which comprise the *CDKN2a* locus. Crossing of Bmi1 null mice with CDKN2a mice ameliorates both the stem cell [11, 208, 209] and thymic defects [201].

In human cases of T-ALL there is strong selection for loss, either via deletion [210] or epigenetic silencing [211, 212] of these tumour suppressors, as reviewed in [213], suggesting that ablation of expression of these genes is critical for T-ALL pathogenesis. Thus, it seems likely that a Bmi1-like activity might promote leukemogenesis in cases were genetic deletion has not occurred. This prompted us to speculate whether PCGF5 might possess such an activity and play a similar role to Bmi1 in this context of normal development and malignant transformation.

Thus, PCGF5 could be posited to play a role in hematopoietic development and, in particular, T cell development both due to its regulation by Notch and by its affiliation with the PcG.

### 4.1.4 Experimental approaches

Insight into the function of PCGF5 was sought in the context of three complex biological systems. These include:

1. Syngeneic transplant of transduced bone marrow
2. *In vitro* T cell development assay with Notch-expressing feeder cells
3. *In vitro* colony forming assay for myeloid progenitors
4.1.5 Hypotheses

Each of these approaches was employed to investigate the effects of PCGF5 gain-of-function in hematopoiesis and to test the following hypothesis:

*PCGF5 is an oncogenic effector of Notch signalling and enforced expression of PCGF5 will result in perturbed hematopoiesis.*

If PCGF5 were the Notch effector responsible for some of the phenotypes observed in cells expressing activated Notch, PCGF5 overexpression might simulate a Notch gain-of-function. This conclusion would be supported if PCGF5 overexpressing cells underwent extrathymeric T cell development or displayed impaired ability to transition from DP to SP cells. Conversely, if PCGF5 acted in a negative feedback loop to inhibit the Notch pathway, this too could be revealed by a phenotype similar to Notch loss-of-function such as B cell fate choice at the expense of T, or arrest in T cell development at the DN stage.

PCGF5 overexpression was found not to be sufficient to replace Notch in driving T cell development, nor did PCGF5 show any evidence of negatively regulating Notch. PCGF5 overexpression did not cause any robust alterations in steady state hematopoiesis but did cause a reduction in colony output in the myeloid CFC assay.

4.2 Results

4.2.1 Overexpression of neither PCGF5 nor Bmi1 causes significant alterations to hematopoiesis in bone marrow transplant recipients

The first approach to discover a physiological effect for PCGF5 was to overexpress it in bone marrow progenitors by means of a retrovirus. The MSCV-IRES-GFP (MIG) retroviral backbone was employed to track transduced cells by their expression of GFP. Bone marrow from 5-FU-treated mice was harvested and transduced *in vitro* with retroviruses then transplanted into lethally irradiated recipient mice. In
addition to PCGF5-bearing retroviruses, MigR1, which expressed only GFP, was included as a negative control for artifacts related to transduction, and Mig Bmi1, which expressed the prototypical PCGF family member Bmi1, was included as a point of reference for any PCGF5-associated phenotypes observed. In total, 3 mice were transplanted with MigR1-transduced marrow, 4 with Bmi1, and 12 with PCGF5 expressing marrow, although not all parameters were measured for each recipient.

Peripheral blood was drawn after three to four weeks to confirm the presence of GFP-positive cells. Each recipient was successfully engrafted and, over the duration of the study, GFP percentages ranged from over 50% to less than 2%. Variation was likely due to differences between cohorts in the efficiency of retrovirus production and infection.

Mice were sacrificed for analysis 8-16 weeks after transplantation and the peripheral blood and bone marrow were assessed for reconstitution of several lineages by transduced cells. These cells were stained with antibodies targeting cell surface proteins for the discrimination of lineages by Fluorescence Activated Flow Cytometry (FACS). The antibodies used and the lineages they identify are listed in Table 4.2.1.

<table>
<thead>
<tr>
<th>Surface Antigens</th>
<th>Cell Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>CD4+ (CD8, CD11b, Gr1, CD19 negative)</td>
<td>CD4 T helper cells</td>
</tr>
<tr>
<td>CD8+ (CD4, CD11b, Gr1, CD19 negative)</td>
<td>CD8 Cytotoxic T cells</td>
</tr>
<tr>
<td>CD11b/Gr1 (CD4, CD8, CD19 negative)</td>
<td>Myeloid Cells</td>
</tr>
<tr>
<td>CD19 (CD4, CD8, Gr1, CD11b negative)</td>
<td>B cells</td>
</tr>
<tr>
<td>Lin- c-kit+ Sca1- FcyR+ CD34+</td>
<td>Granulocyte Macrophage Progenitors (GMP)</td>
</tr>
</tbody>
</table>

Table 4.2.1 – Antibodies used for flow cytometry and the cell types they identify

Each of the recipients was found to contain GFP-positive cells in the bone marrow and peripheral blood (Figure 4.2.1). GFP-positive cells were also detected in the spleens of the recipients at a similar level.
Figure 4.2.1 – GFP-positive cells are present in the peripheral blood and bone marrow of transplant recipients

A) Percent GFP positive cells in transplant recipient peripheral blood and bone marrow cells depleted of red blood cells as assessed by FACS. Each point represents one mouse, horizontal hash marks represent the mean within each cohort, error bars represent standard deviation. Data from an earlier set of PCGF5 recipients depicted in red.

B) Example flow plots for one mouse each of MigR1 control, Mig Bmi1 and Mig PCGF5. Numbers indicate percent falling within GFP positive gate.
Figure 4.2.2 – PCGF5 and Bmi1 do not alter T cell frequency in the peripheral blood

FACS analysis of CD8 SP and CD4 SP T cells in transplant recipient peripheral blood A) Scatter plots show percent CD4 and CD8 SP cells in transplant recipient peripheral blood depleted of red blood cells. Each point represents one mouse, horizontal hash marks represent the mean within each cohort, error bars represent standard deviation. B) Example flow plots for one mouse each of MigR1 control, Mig Bmi1 and Mig PCGF5. Numbers indicate percent falling within each quadrant.
Figure 4.2.3 – No evidence of ectopic T cell development in bone marrow of PCGF5 or Bmi1 recipient mice

FACS analysis of CD8 SP and CD4 SP T cells in transplant recipient bone marrow. A) Scatter plots show percent DP (left panel) and CD8 and CD4 SP (right panel) cells in transplant recipient bone marrow depleted of red blood cells. Each point represents one mouse, horizontal hash marks represent the mean within each cohort, error bars represent standard deviation. B) Example flow plots for one mouse each of MigR1 control, Mig Bmi1 and Mig PCGF5. Numbers indicate percent falling within the gate.
Figure 4.2.4 – PCGF5 and Bmi1 do not alter B cell frequencies in transplant recipient peripheral blood  

FACS analysis for the B cell marker CD19 in peripheral blood  A) Scatter plots show percent CD19 positive cells in transplant recipient peripheral blood depleted of red blood cells. Each point represents one mouse, horizontal hash marks represent the mean within each cohort, error bars represent standard deviation.  B) Example flow plots for one mouse each of MigR1 control, Mig Bmi1 and Mig PCGF5. Numbers indicate percent falling within the gate.
Figure 4.2.5 – PCGF5 and Bmi1 do not alter B cell frequencies in transplant recipient bone marrow

FACS analysis for the B cell marker CD19 in bone marrow  

A) Scatter plots show percent CD19 positive cells in transplant recipient bone marrow depleted of red blood cells. Each point represents one mouse, horizontal hash marks represent the mean within each cohort, error bars represent standard deviation.  

B) Example flow plots for one mouse each of MigR1 control, Mig Bmi1 and Mig PCGF5. Numbers indicate percent falling within the gate.
Figure 4.2.6 – PCGF5 and Bmi1 do not alter granulocyte frequencies in peripheral blood

FACS analysis of peripheral blood for the granulocyte marker Gr1.  
A) Scatter plots show percent Gr1 positive cells in transplant recipient peripheral blood depleted of red blood cells. Each point represents one mouse, horizontal hash marks represent the mean within each cohort, error bars represent standard deviation.  
B) Example flow plots for one mouse each of MigR1 control, Mig Bmi1 and Mig PCGF5. Numbers indicate percent falling within the gate.
Figure 4.2.7 – PCGF5 and Bmi1 do not alter granulocyte frequencies in bone marrow

FACS analysis of bone marrow for the granulocyte marker Gr1. A) Scatter plots show percent Gr1 positive cells in transplant recipient bone marrow depleted of red blood cells. Each point represents one mouse, horizontal hash marks represent the mean within each cohort, error bars represent standard deviation. B) Example flow plots for one mouse each of MigR1 control, Mig Bmi1 and Mig PCGF5. Numbers indicate percent falling within the gate.
Since PCGF5 is regulated by the Notch pathway, and Notch is critical for T cell development, cells from peripheral blood were stained with the T cell markers CD4 and CD8. The percentage of CD4 and CD8 T cells in red blood cell-depleted peripheral blood was determined for each mouse (Figure 4.2.2). CD8 SP cells made up between 2.3% and 8% of peripheral blood with no significant variation between control cells and Bmi1 or PCGF5-transduced cells. CD4 SP cells were present at between 5 and 14% and similarly showed no change between conditions. The frequencies of these cells were also similar to the untransduced population within these mice.

Since aberrant Notch signaling can induce extrathymic T cell development leading to the presence of CD4, CD8 DP cells in the bone marrow [189], it was of interest to determine whether enforced expression of the Notch target PCGF5 caused a similar phenotype. The percentage of CD4 and CD8 SP cells were not altered in the bone marrow of recipients of Bmi1- or PCGF5-transduced marrow compared to control. Further, there was no evidence of an aberrant DP population in the bone marrow in mice transplanted with PCGF5-expressing marrow (Figure 4.2.3).

The surface marker CD19, which is expressed by mature B cells and committed B cell progenitors in the bone marrow, was employed to detect B lineage cells. No significant differences in the percentage of CD19 expressing cells were observed between control and either Bmi1- or PCGF5-transduced cells in either peripheral blood (Figure 4.2.4) or bone marrow (Figure 4.2.5). The CD19 population was again similar between GFP positive and negative fractions.

Granulocytic cells were identified by their expression of the marker Gr1 and by their characteristic of high side scatter (SSC) by flow cytometry. The percentage of Gr1, SSC high cells in peripheral blood of recipients of Bmi1- and PCGF5-transduced marrow was similar to that controls (Figure 4.2.6). Similarly, there was no significant change in the Gr1, SSC high population in bone marrow (Figure 4.2.7).
Figure 4.2.8 – Summary of lineage frequencies in the peripheral blood and bone marrow of transplant recipients. Average lineage frequencies for peripheral blood (top panel) and bone marrow (bottom panel) of all recipients. T cells frequencies (sum of CD4 SP and CD8 SP) are in black, CD19-positive cell frequencies are in grey, and Gr1-positive frequencies are in white.
These findings are summarized in Figure 4.2.8 which shows no significant alterations in the proportion of Gr1-positive, CD19-positive or T cells (CD4 or CD8 positive) in the peripheral blood or bone marrow of transplant recipients.

4.2.2 PCGF5 does not alter lymphoid progenitor performance in OP9 coculture assay

Because of the importance of Notch in T cell development, and in light of the fact that PCGF5 was discovered as a Notch target in T lineage leukemia cells, the investigation into the effect of PCGF5 overexpression in T lymphoid development was extended beyond the bone marrow transplant strategy. For this purpose, the OP9-DL1 co-culture system was adopted. This system allows hematopoietic progenitors to undergo T cell development in vitro due to signals derived from the feeders and from the Notch ligand, Dl1, which they have been engineered to express [214]. Fetal livers were harvested from ed13.5 embryos, subjected to CD24-depletion and then plated on OP9-DL1 or OP9-GFP controls to induce their proliferation. At the first passage interval, three to six days after the initiation of culture, cells were infected with control, Bmi1, or PCGF5-expressing retroviruses and replated on fresh OP9-DL1 feeder cells. After an additional passage of six to eight days, cells were harvested for analysis.

Cells grown on OP9-GFP feeders failed to progress along the path toward T cell differentiation, instead predominantly adopting a monocytic or B lymphoid fate as indicated by their expression of CD11b and CD19 (Figure 4.2.9A), respectively. In contrast, cells cultured on OP9-DL1 expressed the T cell marker CD3ε (Figure 4.2.9A) and acquired CD25 expression consistent with progression to the DN2 stage. At this point cells become lineage restricted to T lymphoid fate with limited natural killer and dendritic cells potential [215, 216]. There was also evidence of loss of CD44 to adopt a DN3 phenotype indicating full commitment to T lymphoid fate (Figure 4.2.9B). Figure 4.2.10 shows progression of T cell development from day 6 of culture of fetal liver on OP9-DL1 to day 12, demonstrating that the system was successfully promoting T cell development.
Figure 4.2.9 – OP9-DL1 cells give rise to DN T cells over monocytes and B cells

A) Histograms for CD11b, CD19 and CD3 are shown to compare differentiation status of fetal liver cells grown for 14 days on OP9-DL1 (blue line) or OP9-GFP (red line).  B) Analysis of DN stage progression for cells grown for 14 days on OP9-GFP and OP9-DL1 as indicated.  Schematic at right shows the DN stage associated with each quadrant.  Numbers indicated percent of cells falling within each quadrant.  Similar results were obtained for 3 independent experiments.
Comparison of the DN fractions for Bmi1- or PCGF5-transduced cells against control-transduced cells showed no significant alterations. Each had a very similar distribution of DN subsets and each was similar to the untransduced GFP-negative cells in the same well, which served as a valuable internal control (Figure 4.2.11).

This system did not efficiently generate double positive T cell progenitors in the course of these studies, the possible reasons for which will be discussed in section 4.3. Within the GFP-positive fraction the control culture gave rise to 0.3% DP cells, while the Bmi1 and PCGF5 cultures gave rise to 0.4% and 1.05% respectively while the fraction of untransduced DP cells remained constant at 0.3-0.4% (Figure 4.2.12).
Since the culture of fetal liver progenitors on OP9-DL1 did not allow assessment of the transition from DN to DP in the studies described herein, DN cells were sorted from the thymi of 5 week old mice wild type B6 mice for input into the coculture assay. Cells were sorted by flow cytometry to deplete cells expressing CD4 or CD8. Contamination of the DN cells by DP and SP cells after sorting was routinely under 5% (Figure 4.2.13A). The sorted DN cells were then infected with either control virus or virus

Figure 4.2.11 – No effect of Bmi1 or PCGF5 on DN subsets on OP9-DL1 Flow Cytometry plots of double negative cells (CD4-CD8-CD19-CD11b-Gr1-NK1.1-)at day 14 of coculture on OP9-DL1 feeder cells. Cells were stained for CD44 and CD25 to reveal their stage of development. The fraction of cells lying within each quadrant is indicated. Cells were transduced at day 5 with control virus or virus encoding Bmi1 or PCGF5. The left column describes untransduced cells while the right describes transduced cells as identified by their positive GFP signal. Similar results were obtained for greater than 6 different experiments.
encoding PCGF5 and cultured on OP9-DL1 feeders. After five days, control and PCGF5-expressing cells both efficiently gave rise to DP cells as well as both CD4 and CD8 SP cells (Figure 4.2.13B).

**Figure 4.2.12 – DP cell generation on OP9-DL1** Flow Cytometry plots of CD45+ cells at day 14 of coculture on OP9-DL1 feeder cells. Cells were stained for CD4 and CD8 to reveal their stage of development. The fraction of cells lying within each quadrant is indicated. Cells were transduced at day 5 with control virus or virus encoding Bmi1 or PCGF5. The top row describes transduced cells while the bottom describes untransduced cells within the same culture. Similar results were obtained for greater than 6 different experiments.

In the absence of the exogenous DL1 ligand, OP9 cells robustly promoted the development of B lineage cells, while a natural mutation in their M-CSF gene does not support efficient monocyte differentiation. Since the only difference between the conditions experienced by cells grown on OP9-DL1 cells and the control OP9-GFP cells is the presence or absence of Notch ligand, this system was employed to
determine whether expression of PCGF5, which is one of the genes turned on by Notch, is able to rescue the absence of Notch stimulation on OP9-GFP. Cells were cultured on OP9-GFP or DL1 or 6 days, and then transduced with either control retrovirus or retrovirus encoding Bmi1 or PCGF5. These cells were then cultured on fresh OP9 cells for another 8 days. Cells transduced with control virus grown on OP9-DL1 yielded very few DP cells, as was always the case in the experiments undertaken for this study. However, cells grown on OP9-GFP yielded no DP cells whatsoever, regardless of their expression of Bmi1 or PCGF5 (Figure 4.2.14A).

Figure 4.2.13 – PCGF5-expressing DN thymocytes rapidly become DP on OP9-DL1  A) Example flow plot showing low DP contamination in sorted DN thymocytes from adolescent mice. B) After day 5 of coculture on OP9-DL1 feeder cells were stained for CD4 and CD8. The fraction of cells lying within each quadrant is indicated. Cells were transduced at day 0 after sorting with control virus or virus encoding PCGF5. Similar results were obtained for 5 independent experiments, each from a different thymus.
Likewise, cells transduced with control virus grown on OP9-DL1 efficiently gained CD25 expression and showed evidence on transition through DN stages 1-4. Control cells grown on OP9-GFP showed no such progression. With respect to the putative ability of overexpression of PCGF5 to rescue the absence of
Notch ligand, no evidence of T lineage differentiation was observed in PCGF5-expressing cells compared to control (Figure 4.2.14B). Indeed, cells cultured on OP9-GFP predominantly gave rise to B cells, as identified by CD19 expression, in the MigR1 control condition as well as in cells transduced with Bmi1 or PCGF5, while only a minority of cells grown on OP9-DL1 expressed CD19 (Figure 4.2.14C). Neither PCGF5, nor Bmi1, were found to cause any significant changes in the differentiation of progenitor cells on OP9-GFP.

It is possible that transduction at day six may be too late to exert an effect so, for comparison, one experiment was performed wherein transduction was performed after three days of coculture. This experiment gave the same result.

Thus, in the context of the OP9 coculture system, Bmi1 and PCGF5 overexpression showed no effects on the behavior of progenitors in either T or B cell development, suggesting that neither Bmi1 nor PCGF5 are sufficient to substitute for Notch signalling in this assay, nor are these genes able to robustly alter the qualitative potential of lymphoid progenitors.

4.2.3 PCGF5 causes reduced myeloid CFC output in bone marrow progenitors

Despite no effect having been observed for overexpression of PCGF5 in steady-state hematopoiesis it remained possible that a subtle effect on progenitors might be compensated for by homeostatic processes in vivo. The colony forming assay was thus employed to test PCGF5’s effect on progenitor colony output as it is well-suited to assessing progenitor potential and is quantitative. Cells harvested from 5-FU-treated mice the pre-stimulated in liquid culture with IL3, IL6, Flt3L, 100ng/ml SCF to induce them to cycle in preparation for infection. Cells were transduced after two days of culture with control, Bmi1, or PCGF5 virus, put back in liquid culture for two days and then sorted to deplete untransduced cells. Sort purities were routinely above 95% (Figure 4.2.15A).
Cells were then assayed for their performance in a myeloid Colony Forming Cell (CFC) assay in semi-solid media supplemented with IL-3, IL-6 and SCF. Colonies were counted after 10 days of growth.

Overexpression of Bmi1 resulted in an average decrease in colony number of 17% (p value = 0.025) while overexpression of PCGF5 resulted in an average decrease of 52% (p value < 0.0001) relative to MigR1 control over 3 independent transductions of independently harvested bone marrow representing 5 total plates for each condition.

**Figure 4.2.15 – Overexpression of Bmi1 and PCGF5 decrease myeloid colony forming cell frequency**  Myeloid CFC assay of 5-FU bone marrow infected with retrovirus and then sorted to high purity. A) Percentage of GFP positive cells in sorted population 2 days post transfection with the indicated virus. B) Average colony number after 10 days of culture as a percentage of MigR1 control over 3 independent transductions of independently harvested bone marrow representing 5 total plates for each condition. C) Colony counts per plate for 10,000 input cells for each experiment.
cells infected with the control virus (Figure 4.2.15B). Actual colony numbers for each experiment are shown in Figure 4.2.15C.

**Figure 4.2.16 – Bmi1 and PCGF5 overexpression result in reduced proportion of CD11b bright cells**  
A) FACS plots showing GFP for cells transduced with the indicated virus recovered from semi-solid methylcellulose media after 10 days of culture and untransduced P12 cells as a negative control for GFP expression. The percent of cells expressing GFP is indicated within the gate.  
B) Histogram showing fluorescence intensity for CD11b for cells transduced with the indicated retrovirus harvested after 10 days of culture in CFC. GFP and CD11b was measured post-culture in 2 or 3 CFC experiments and similar results were observed.
The cultures were then harvested and their GFP expression was assessed by flow cytometry to confirm that the colonies derived from transduced cells. For control and Bmi1-expressing cells, greater than 95% of cells were GFP positive while in the PCGF5 cultures, it was slightly lower (Figure 4.2.16A). Further,
while the large majority of cells in the control cultures expressed high levels of the monocyte/granulocyte marker CD11b, cultures of both Bmi1- and PCGF5-transduced cells had large fractions of cells that expressed CD11b at a level nearly two orders of magnitude lower (Figure 4.2.16B).

This data indicated an effect of PCGF5 overexpression on the performance of extant progenitors in bone marrow but does not address whether PCGF5 has an effect on the progenitor pool itself. To determine whether PCGF5 altered the progenitor frequency in mouse bone marrow bone marrow of transplant recipient mice was analyzed for markers that identify the GMP population. No significant difference in GMP frequency was observed between recipients transplanted with cells infected with control virus compared with Bmi1- or PCGF5-expressing virus (Figure 4.2.17A).

Thus, it seems PCGF5 overexpression does not alter the number of GMPs present in the bone marrow of mice but reduces their ability to produce CD11b+ colonies in an in vitro CFC assay.

4.3 Discussion

4.3.1 Comment on Experimental Design

The experiments described in this chapter sought to characterize the function of PCGF5 by overexpressing it in hematopoietic cells and observing the outcome. While the experiments were designed with the aim of relating the effects of PCGF5 overexpression to Notch function, it was also of interest to contrast PCGF5’s function with that of the prototypic PCGF family member, Bmi1.

To that end, Bmi1-overexpressing cells were assayed in parallel with those overexpressing PCGF5 in nearly all the assays described here. In was hoped that this would offer a point of comparison for any phenotypes identified for PCGF5.

Some assays were performed in cells receiving a strong Notch signal, such as the OP9-DL1 coculture experiments, which would be expected to have high levels of endogenous PCGF5 mRNA in addition to
the retrovirally overexpressed PCGF5. In other assays, such as the OP9-GFP coculture and the CFC assay, cells would be expected to experience little or no Notch stimulation and, therefore, might experience a greater fold increase, but lower absolute level, of PCGF5 expression.

The aim of these experiments was to examine the effects of enforced PCGF5 expression in the presence of absence of a Notch signal and to compare the activity of PCGF5 with that of Bmi1.

4.3.2 PCGF5 overexpression in bone marrow does not alter engraftment or multilineage reconstitution in transplant recipients

Twelve mice transplanted with PCGF5-transduced bone marrow were analyzed for this study and compared with four mice which received Bmi1-transduced marrow and three which received control marrow transduced with a virus expressing only GFP. Cumulative analysis of all recipients revealed no significant differences in mature hematopoietic subsets between the control, PCGF5 and Bmi1 recipients.

Cells expressing either PCGF5 or Bmi1 were able to efficiently give rise to both myeloid and lymphoid lineages until at least 16 weeks post-transplant suggesting that neither Bmi1 nor PCGF5 impair long-term repopulating activity of the transplanted bone marrow. Recipients were injected with 50,000 cells derived from 5-FU-treated bone marrow. Stem cell frequency in 5-FU treated marrow is estimated to be approximately 1 in 1000 [217] and as the infection rate was lower than expected, each recipient mouse would have received approximately five transduced stem cells. This appears to have been sufficient to confer multi-lineage reconstitution by transduced cells up to sixteen weeks, but it’s possibly stochastic effects caused by the transplant of so few cells contributed to the difficulty is discerning a phenotype.

Since neutrophils are very short-lived with a half-life in circulation of much less than 24 hours [218, 219], their presence in the periphery serves as a proxy measurement of ongoing progenitor activity.

Therefore, the observation at 16 weeks post-transplant of similar fractions of Gr1-positive granulocytes
in the periphery of recipients of both Bmi1- and PCGF5-expressing marrow to control suggests enduring hematopoiesis in these mice. The stable maintenance of granulopoiesis by Bmi1- and PCGF5-expressing marrow alongside untransduced marrow suggests that there is no competitive advantage or disadvantage due to the enforced transgene expression.

4.3.3 Recipients of PCGF5 overexpressing bone marrow have an unaltered distribution of hematopoietic lineages

One recurring role of the Notch pathway in biology is in the determination of fate choice in bi- or multi-potential progenitors [220-223]. As a transcription factor, Notch exerts its control over fate choice by the genes whose expression Notch regulates [163, 224]. Therefore, it was of interest to determine whether PCGF5, as a target of Notch signaling, conferred a bias towards one fate over another. Enforced expression of Notch favours lymphoid fate over myeloid, T lymphoid over B [189], and finally CD8 T cell development at the expense of CD4 [225]. Therefore, if PCGF5 was responsible for all or part of this phenotype of Notch activity, it would be expected to similarly skew development in bone marrow transplant recipients; namely, to increase T lymphoid cells, in particular CD8 SP cells, with a commensurate decrease in B lymphoid and myeloid subsets. However, the panel of surface markers employed, which demarcate granulocytic cells, B lymphoid and T lymphoid cells of both CD4 and CD8 subsets, did not reveal any evidence of alteration of fate choice was observed for PCGF5 or Bmi1.

Since both CD4 and CD8 SP cells were observed in the periphery but there was no evidence of immature DP T cells in the bone marrow or spleen signifying extrathymic T cell development, it can also be concluded that neither Bmi1 nor PCGF5 blocked thymopoiesis. Thus, hematopoiesis seems to progress as normal despite the enforced expression of these genes.
4.3.4 PCGF5 overexpressing progenitors show no difference to controls in an *in vitro* model of lymphoid development

While the bone marrow transplant model provided a means of determining effects on the steady state distribution of hematopoietic subsets, an alternate system was required to circumvent several technical and practical limitations associated with bone marrow transplant which hindered the further study of developing cells. Transplant experiments took a long time to complete, they were costly, and could only be assayed at their endpoint – no interim analysis was possible. Another challenge for analysis is the complexity of the murine hematopoietic system, which requires cells to transit through various developmental stages, to home to different tissues and to make numerous fate choices. With respect to T cells, it would be difficult to determine whether a thymic phenotype was due to altered thymocyte biology, or whether the defect arose during an earlier progenitor step. Mouse thymi also undergo involution as they age, so the study of mouse thymi would require that transplants be done with young recipients and that these be sacrificed soon after transplant which would have compromised the detection of phenotypes that arose later.

To help circumvent these problems, the co-culture system developed in the Zuniga-Pflucker laboratory was adopted [214]. In this system, progenitors from fetal liver, or other sources, are cultured on a feeder layer derived from the osteopetrotic (op/op) strain of mice. These mice, of the C57BL/6J background, carry a spontaneous mutation in the gene encoding Macrophage-Colony Stimulating Factor and thus do not produce any functional M-CSF [226]. Therefore, these feeders do not support the development of monocytes which would otherwise overwhelm such a culture. To drive T cell development at the expense of the B cells which would arise under these conditions normally, these feeder cells have been engineered to express Dl1, a ligand of Notch [227]. Provided with the environmental stimuli of the OP9 feeders coupled with the Notch simulation provided by the DL1 transgene, uncommitted progenitor cells are able to reach the CD4, CD8 double positive (DP) stage *in*
vitro, and also undergo a several thousand-fold expansion. This system has further been demonstrated to allow for CD8 SP cells to arise in the culture which show evidence of being functionally mature [214]. Thus, in this system, it is possible to drive cells from the uncommitted progenitor stage to the DP stage in vitro, via the DN2 stage where cells are committed to the T lineage and the DN3 stage where β-selection occurs. Cells can be withdrawn for analysis at interim time points and the experiment can be concluded as little as two weeks from initiation of the co-culture.

No phenotype was observed for PCGF5 or Bmi1 overexpression in this system when cells where transduced after 5 to 6 days of co-culture. In case cells had already passed some threshold after which they were insensitive to increased PCGF5, an experiment was performed wherein cells were infected after 3 days of culture. However, despite the earlier and more prolonged enforcement of PCGF5 expression, no effect was observed. It was also noted there was no evidence of the CD25 dim population described for the Eμ-Bmi1 transgene in either the Bmi1 or PCGF5-transduced populations [203].

Since the fetal liver progenitors co-cultured on OP9-DL1 yielded very few DP cells in the experiments described here, it was difficult to conclude whether PCGF5 overexpression exerted any influence on the DN to DP transition. Experiments using sorted DN cells from thymi of 5 week-old transduced with PCGF5 gave rise to DP cells extremely efficiently and showed no difference in the PCGF5-transduced cells relative to control. The DN to DP transition was so rapid and efficient though, that it is possible that many cells had already committed to the transition and undergone a proliferative burst by the time PCGF5 protein was expressed. This would mask an effect on a subset of cells, particularly if they were less proliferative and would, as a consequence, make up a small portion of the culture.

Also worth noting in this system is the presence of CD4 SP cells in the culture. CD4 T cells require the expression of MHC Class II by the thymic stroma during their development so that their newly
rearranged T cell receptor may be positively selected for its ability to interact with Class II. OP9 cells express MHC Class I, and thus have the potential to support CD8 SP cell development, but they do not express MHC Class II. This suggests that the CD4 SP cells that arose under these conditions had already undergone β-selection in vivo before being withdrawn from the mouse.

Of the two approaches described here, namely the co-culture on OP9-DL1 feeders of fetal liver progenitors or DN thymocytes, the former did not efficiently give rise to DP cells and the latter was obfuscated by the presence of cells that had already undergone β-selection. Thus, it cannot be concluded whether PCGF5 overexpression impacts the T cell development process at this critical stage.

Anecdotally, different labs seem to experience varying levels of success at generating DP cells via the OP9-DL1 system so it may yet be possible to answer this question via the OP9-DL1 coculture system. Alternatively, OP9 cells engineered to express an alternate Notch ligand, Dl4 could be employed as these have been shown to be more potent in driving T cell development than DL1 [228].

Though no effect of PCGF5 was revealed by its overexpression in cells with active Notch signaling, such as those grown on feeders expressing Dl1, it was of interest to determine whether PCGF5 overexpression simulated a Notch signal in progenitors grown on feeders that do not express exogenous Notch ligands. Under these conditions control cells acquired CD19 expression consistent with their adopting a B lymphoid fate. It was hypothesized that if PCGF5 was a sufficient component of the Notch directive toward T lymphoid fate, there may be evidence of T cell development in PCGF5-expressing cells on OP9-GFP feeders. However, the were no detectable DP cells in PCGF5-expressing cells cultured on OP9-GFP, while control cells on OP9-DL1 yielded small numbers of cells expressing CD4 and CD8. More significantly, while cells grown on OP9-DL1 expressed CD25 consistent with the DN2/3 stages on T cell development, no cells cultured on OP9-GFP expressed CD25 and this was not altered by transduction with PCGF5.
While the OP9 experiments cannot formally exclude an effect of PCGF5 overexpression on T cell development, particularly one that is exerted prior to thymic seeding or during β-selection, the expansion and differentiation of T cell progenitors during the early stages up to DN3 seem to be unaffected by PCGF5 overexpression. Likewise, in the absence of Notch stimulation, PCGF5 is not sufficient to replicate the effects on Notch signaling in lymphoid progenitors.

4.3.5 Enforced expression of PCGF5 and Bmi1 reduce clonogenic activity of bone marrow progenitors in a colony forming assay

While no significant alterations were revealed in cells overexpressing Bmi1 or PCGF5 in the context of bone marrow transplant or OP9 coculture, a significant decrease in the colony forming potential of Bmi1- and PCGF5-overexpressing cells was observed. These experiments, performed by infecting bone marrow from 5-FU treated mice with retroviruses containing the Bmi1 or PCGF5 gene, gave a consistent, significant decrease in PCGF5-expressing cells relative to control (Figure 4.2.15) while Bmi1 expressing cells showed a more modest effect. Since the starting progenitor number in each condition was identical, the decreased colony number in these experiments suggest that overexpression of PCGF5 causes some cells that would have otherwise given rise to colonies to either fail to proliferate or to die.

The fraction of CD11b bright cells was much lower in both Bmi1 and PCGF5-expressing cultures; however, it cannot be determined from the data available whether this is due to a failure of these cells to differentiate or whether the reduced CD11b population resulted from a death or proliferative failure of the upstream progenitors of the CD11b bright population. It was also noted that, while Bmi1-expressing and GFP control cells maintained nearly the same percentage of GFP expressing cells at the end of the culture as was present at initiation, the PCGF5 cultures showed a modest decrease in GFP frequency in the two experiments where this parameter was measured. This might suggest that the few
remaining GFP negative cells that remained after sorting had a competitive advantage and were able, after the period of culture, to make up a higher fraction of the total cells.

The observation of reduced colony output by Bmi1-transduced cells contrasts with published data showing the opposite effect of Bmi1 overexpression [229]. However, in the published study, sorted CD34-KSL cells were cultured for 10 days in stem cell factor and thrombopoietin, which promotes expansion of immature cells but inhibits differentiation, before initiation of the CFC culture. These cells, therefore, had the opportunity to proliferate under the influence of elevated Bmi1 for 10 days before their clonogenicity was measured while the experiments described here measured the clonogenic potential of total (5-FU-treated) bone marrow newly transduced with Bmi1. Thus, a Bmi1-mediated expansion of clonogenic progenitors would likely not be revealed in this assay.

However, whatever effect Bmi1 and PCGF5 overexpression have on myeloid growth and development is clearly moderate not absolute, since colony forming potential is not completely ablated by PCGF5, nor were CD11b bright cells completely absent from these cultures. The subtle nature of the phenotypes observed in the colony forming assay is consistent with the absence of a robust phenotype in vivo.

While the observation that PCGF5 reduces colony output of myeloid progenitors in CFC is intriguing it, unfortunately, did little to illuminate out understanding of the role of Notch-driven PCGF5 in hematopoiesis. Since Notch is not thought to play an important role in myeloid progenitors, this phenotype was not investigated further.

4.3.6 Enforced expression of PCGF5 does not replicate constitutive Notch signalling

One aim for the experiments described in this chapter was to determine whether expression of PCGF5, which is stimulated by active Notch signaling, as sufficient to drive Notch-dependent processes in the
absence of a Notch signal. Such processes include the choice of T cell fate over B lymphoid or myeloid fates, and the adoption of T lineage specificity of uncommitted progenitors.

Specifically, bone marrow transplant recipients were examined for changes in the distribution of hematopoietic lineages consistent with a bias towards T cell fate at the expense of other lineages. Hematopoietic progenitors cultured on control OP9-GFP cells, which lack the Dll1 Notch ligand, were also assayed for evidence of Notch-driven T cell development in. In neither of these cases was there any evidence that PCGF5-expressing cells behaved like cells receiving a Notch signal. These data lead to the conclusion that PCGF5 is not sufficient to replace Notch signaling as a driver of T cell development.

Exacerbating the interpretation of the absence of a phenotype was the inability to confirm that the PCGF5-infected cells were expressing elevated levels of PCGF5 protein. Since no effective antibody was available it was impossible to confirm expression of the protein. However, the decrease in colony output in the CFC assay (Figure 4.2.15) suggests that the construct was having an effect which the control vector did not have. Further, it was observed that this retroviral construct conferred an equivalent increase in ubiquitylation of histone H2A, as will be discussed in Chapter 5, as an epitope-tagged version of PCGF5 which was employed subsequently and whose expression was confirmed by western blot. These observations provide evidence that the Mig PCGF5 retrovirus does drive expression of PCGF5 protein though this cannot be formally proven.

4.3.7 Physiologic effects of altered gene dose of the Polycomb Group and tissue-specific phenotypes

As mentioned above, it was not possible to ascertain the degree of overexpression achieved in these experiments, nor was it possible to measure the level of endogenous PCGF5 protein expressed in the cells examined. If PCGF5 was already expressed at a high level, as suggested for thymic progenitors, in
particular, by the microarray data discussed in Chapter 1, then a modest degree of overexpression may have had little effect.

Indeed, there is evidence that many genes in the Polycomb Group are associated with dose-dependent phenotypes. The phenotype in RING1A transgenic mice was observed to be dose-dependent [38] and both Bmi1 [209] and RING1B have [230] have had phenotypes associated with their haploinsufficiency. Thus, it is possible the level of overexpression achieved was not sufficient to induce a detectable phenotype.

While overexpression gain-of-function phenotypes have been described for PRC1 complex members Bmi1 [34] and RING1A [38] many of the phenotypes associated with their overexpression manifest during development in non-hematopoietic tissues. Thus, it is possible the strategies employed in these studies were not capable of revealing PCGF5 gain-of-function effects.

4.3.8 Open question: does PCGF5 contribute to leukemogenicity of Notch

PCGF5 was of interest for study due to its potential to act as an oncogene in a manner similar to its paralogue Bmi1; however, the experiments described here were not designed to identify an oncogenic activity due to the relatively short duration over which the recipient mice were monitored.

Mice engineered to have a disrupted Ink4a gene, which in actuality was a Ink4a/ARF double null, were reported to have a rate of tumour development of 69% at 29 weeks of age with lymphoma being a common tumour type [231]. Subsequently, ARF null mice that retained functional Ink4a were described to become moribund due to tumour formation in 29 of 38 (72%) mice, with nearly a third of those developing lymphoma[232]. Finally, the reciprocal mutant lacking Ink4a but retaining ARF had a rate of tumour formation of 26% at one year with 4 of 10 mice displaying a splenic lymphoma [233]. Similarly,
mice with a Bmi1 transgene driven by the immunoglobulin heavy chain enhancer (Eμ-Bmi1) were reported to develop lymphoma within 42 weeks with a median survival of about 25 weeks [203].

As the PCGF5 transplant mice described herein were only monitored for 8-24 weeks post-transplant, such a phenotype, should one have existed, may not have had time to manifest. A PCGF5 transgenic, or larger cohorts of transplant recipients left for longer periods, would be required to rule out a phenotype akin to the Bmi1 transgenic or Ink4a/ARF null models.

Another approach for uncovering an oncogenic phenotype for PCGF5 would be to over-express PCGF5 in conjunction with another oncogene. For instance, Notch-driven c-myc plays in development of T-ALL [159, 160] and Bmi1 collaborates with myc to cause lymphoma. Oncogenic levels of c-myc could be provided by transducing bone marrow cells from Eμ-myc mice with PCGF5 and transplanting into recipients. These cells would thus express two Notch-driven genes: the bona fide oncogene c-myc and PCGF5. However, as mice of the Eμ-myc transgenic background commonly develop B lineage neoplasms over T [234, 235] a T-cell specific myc transgenic mouse would be better suited for the study of PCGF5 in T cell leukemogenesis. This would be advantageous as it is in this context that Notch is observed to regulate PCGF5.
Chapter 5 PCGF5 is a member of the Polycomb Group and a paralogue of Bmi1

Contribution

I designed, executed, and analyzed the data for all experiments in this chapter. The mouse T-ALL cells which I transduced for the experiment in Figure 5.2.11 were in vitro expanded and frozen in Dr. Andrew Weng’s lab by Dr. Hind Medyouf.

5.1 Introduction

The Notch pathway has direct impacts on gene expression via Notch’s role as a transcriptional activator. However, several Notch target genes themselves act to alter transcription both as activators, such as myc [159, 160] and Gata3 [236], or repressors such as the Hes family [237].

The identification, described in Chapter 3, of PCGF5 as a target of the Notch signalling pathway raised the intriguing possibility that Notch may modify PcG activity and presented another means by which Notch might alter gene expression. This link, however, relies on the assumption that PCGF5 acts as a bona fide Polycomb Group gene. At the outset of this study there was no published empirical data, save PCGF5’s homology to Bmi1 and the other PCGF family members, to support PCGF5’s inclusion in the PcG. However, other putative PCGF family members, MBLR and NsPC1 had joined Bmi1 and Mel-18 as acknowledged members of the PcG.

Recently, a handful of proteomic studies have identified peptides corresponding to PCGF5 in tandem-affinity mass spectrometry experiments. First, PCGF5 was detected in immunoprecipitates of RING1B from a murine erythroleukemia cell line [53] while it was not detected in immunoprecipitates of its paralogues Bmi1 or Mel-18 [49, 238] in HeLa and 293 cells respectively.

A more extensive study provided further evidence for PCGF5’s role in a PRC1-like complex that is quite distinct from the Bmi1- and Mel-18-containing complexes previously described [52]. This same study
also performed ChIP-seq to identify the regions of the genome where the various PCGF-containing complexes bind. PCGF5 was found to bind to an overlapping, but distinct set of genomic loci suggesting that, both via its complex composition and by its distribution on chromatin, that PCGF5 plays a role that differs from that of the PCGF family members, Bmi1 and Mel-18, that have shaped our understanding of the PRC1 complex to date.

However, many questions remain unanswered. Firstly, to date, no one has directly addressed PCGF5’s binding of PRC1 complex members, instead relying on passively detecting it via mass spectrometry. Secondly, it is not known whether PCGF5 is able to confer ubiquityl transferase activity to the PRC1 complex as Bmi1 does. Lastly, the consequences of PCGF5 expression on gene transcription are unknown. Thus, the following hypotheses were investigated:

1. PCGF5 associates with the Polycomb Group protein RING1B
2. PCGF5 is competent, as a member of PcG complexes, to effect modification of histones
3. PCGF5 is capable of modulating gene expression

Here data will be presented demonstrating PCGF5’s association with RING1B and that PCGF5 competes with Bmi1 for this interaction. It was also found that PCGF5 overexpression results in increased ubiquitylation of histone H2A to similar degree as Bmi1. Finally, overexpression of PCGF5 was found to correlate with changes in the expression of a subset of Hox genes but not of the tumour suppressors Ink4 and Arf. Collectively, these data suggest a role for PCGF5 that is mechanistically similar to that for Bmi1 but that the spectrum of genes regulated by PCGF5 and Bmi1 differ.
5.2 Results

5.2.1 PCGF5 contains a RING domain with homology to Bmi1

The identification of PCGF5 as a transcriptional target of the Notch signalling pathway in T-ALL cells was of interest to us due to its homology to the proto-oncogene Bmi1. This homology is largely limited to the N-terminal RING-type zinc finger domain which is the hallmark of the PCGF family (Figure 5.2.1).

The RING-type zinc finger is defined by a sequence comprising cysteine and histidine residues at distinct intervals which form a consensus motif, known as the cross-brace, responsible for binding zinc ions. The stable RING-type structure is often associated with E3 ubiquitin ligase activity and can mediate protein-protein or protein-DNA interactions.
The PCGF5 RING finger consensus sequence spans residues 17 to 56 in both human and mouse proteins. In human, PCGF5 and Bmi1 are 59% identical and 73% similar over this interval. Critically, the cysteine and histidine residues, and the length of the intervals between them, are exactly conserved (Figure 5.2.2A). The same is true for the rest of the PCGF family. This consensus motif is contained within the RING domain, here defined as amino acids 1-109, whose crystal structure has been solved in complex.

**Figure 5.2.1 – Domain structure schematic of the human PCGF family** RING domains corresponding to amino acids 1-109 of Bmi1 are indicated in blue. Green regions do not conform to any known domain structure. The region shaded grey of PCGF6 is not homologous to any region of the other PCGF family members.
with the RING domain of RING1B and, further, this complex possess ubiquityl transferase activity \textit{in vitro} \cite{64,65}. Closer analysis of the conservation between Bmi1 and PCGF5 reveals important similarities, and interesting differences, when the crystal structure is considered. The association of Bmi1 and RING1B is mediated in part by salt bridges between Glu11, Asp72, and Lys81 of Bmi1 \cite{65}. These residues are conserved in the RING domain of PCGF5 (Glu11 is substituted for similarly acidic Asp in PCGF5, the other two are identically conserved (\textbf{Figure 5.2.2B}). This strongly suggests that these genes were subject to evolutionary constraint at these positions by a conserved interaction with RING1B. However, the basic

\textbf{Figure 5.2.2} – \textit{Important features of Bmi1’s RING domain are conserved in PCGF5}  
\textbf{A)}

Comparison of the RING consensus sequence with the RING motifs of Bmi1 and PCGF5. Critical C and H residues are highlighted in green.  

\textbf{B)} Conservation by PCGF5 of some, but not all, residues determined empirically to be important for Bmi1 function. Residues shaded in green are critical for associated of Bmi1 and RING1B and are conserved in PCGF5. Residues shaded in red mediate association of Bmi1 with DNA and are not conserved in PCGF5.
region created by Lys62 and Arg64 of Bmi1 observed to interact with DNA in the crystal structure, and to significantly decrease ubiquitin transferase activity when both basic residues are mutated to alanines, is not conserved in PCGF5. Those positions are instead occupied by glutamic acid and asparagine, respectively.

5.2.2 Sequence divergence outside the RING domain of PCGF family members

As previously discussed the RING domain underpins the structural and functional definition of the PCGF family. Accordingly, this domain is highly conserved between all members of the PCGF family, such that each pairwise comparison is between 41% and 55% identical. The one exception is Bmi1 and Mel-18 which are 89% identical over this region (Figure 5.2.3A).

This contrasts with overall homology which is below 20% for PCGF6:Bmi1 and PCGF6:Mel-18 up to 61% for Bmi1:Mel-18. PCGF5 is most similar to PCGF3 with 39% of positions being identical. Further, if the alternatively spliced exon 9, which seems not to be present in PCCF3, is excluded from this comparison, these two proteins are 42% identical (Figure 5.2.3B).

By sequence, Mel-18 and Bmi1 seem for form a subfamily, and PCGF3 and PCGF5 are also strongly similar to each other. PCGF1 and PCGF6 are less strongly homologous to the rest of the PCGF family and are not strikingly similar to each other (Figure 5.2.4). As mentioned previously, Bmi1 and Mel-18 seem to have similar functions in some tissues, but opposing functions in others. Not enough is known yet about the rest of the PCGF family to group them functionally.
Table 5.2.3 – Homology between PCGF family members  A) Pairwise comparisons of the RING domains for all members of the PCGF family. Table shows percent identity between each pair. RING domain is defined as aa1-109 for Bmi1, Mel-18, PCGF3, PCGF5 and aa30-138 for PCGF1, and aa117-225 for PCGF6. B) Overall homology between all pairs of PCGF proteins. Table shows percent identity between each pair. * indicates exon 9 splice variant of PCGF5 which is more similar to PCGF3 than the product of the full length transcript.

<table>
<thead>
<tr>
<th></th>
<th>PCGF1</th>
<th>PCGF2</th>
<th>PCGF3</th>
<th>Bmi1</th>
<th>PCGF5</th>
</tr>
</thead>
<tbody>
<tr>
<td>PCGF2</td>
<td>55</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PCGF3</td>
<td>53</td>
<td>48</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bmi1</td>
<td>56</td>
<td>89</td>
<td>47</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PCGF5</td>
<td>41</td>
<td>42</td>
<td>47</td>
<td>44</td>
<td></td>
</tr>
<tr>
<td>PCGF6</td>
<td>55</td>
<td>47</td>
<td>51</td>
<td>50</td>
<td>48</td>
</tr>
</tbody>
</table>

* dEx9 is 42
5.2.3 PCGF5 shares some, but not all, putative post-translational modification sites with Bmi1

Phosphorylated residues have been detected by mass spectrometry for PCGF5. There are three tyrosines and three threonines, each of which is conserved between mouse and human, which have been detected to be phosphorylated in one species or the other [239]. These are Y15, Y19, Y24, T29, T30, T37 which are all located either within, or immediately adjacent to the RING structure. No phosphorylation sites have been observed C-terminal of the RING domain. While Y15 and Y19 are His and Val residues in Bmi1, the remaining sites are conserved exactly between PCGF5 and Bmi1, except that T37 is instead a serine. These modifications were identified by mass spectrometry and no function has yet been ascribed to the phosphorylation of PCGF proteins within the RING domain.

Bmi1 has many more known post-translational modifications, likely due to its more extensive history of study. Bmi1 has been described to undergo ubiquitylation and one study implicated the F-box protein βTrCP in this process. The authors also demonstrate that βTrCP required a conserved DSGS degron
motif in Bmi1 to exert its effect [240]. No such motif is present in PCGF5. However, Lys81 of Bmi1, which was found by mass spectrometry to be ubiquitylated, is conserved in PCGF5. In fact, the region around this site from amino acid 74 to 86 is 69% identical with only four conservative substitutions in this region, suggesting substantial selective pressure to conserve this motif (Figure 5.2.5). There are several phosphorylation sites described to the C-terminus of the RING finger of Bmi1 including three serines at positions 251, 253, 255 which have consequences on Bmi1’s activity towards histone H2A [241-243]. However, none of these is conserved in PCGF5.

The studies described here had focused on PCGF5 in comparison with Bmi1 due to the established roles Bmi1 has in hematopoiesis and cancer; however, within the PCGF family, Bmi1 is relatively dissimilar to PCGF5. PCGF5’s closest paralogue is PCGF3, to which it is 39% identical and 60% similar over their entire lengths and show significant similarity in their C-termini. While several regions show very high conservation suggesting an evolutionary constraint to mutation, analyzing the sequence using the Motif Scan utility found only one motif conserved between PCGF5 and PCGF3. This is a CK2 phosphorylation site at residues 188-191 of PCGF5.

5.2.4 PCGF5 associates with RING1B in competition with Bmi1

While the amino acid sequence homology to Bmi1 and other Polycomb Group Ring Finger proteins is striking, it remained to be determined whether the conserved RING domain of PCGF5 permitted
interaction with the Bmi1 binding partners RING1A and RING1B. Since no antibody was available for PCGF5, the PCGF5 coding sequence was cloned in frame with an N-terminal hemagglutinin (HA) epitope tag to facilitate detection, hereafter called HA-PCGF5.

HA-PCGF5 was co-transfected into 293T cells with a plasmid encoding FLAG-tagged RING1B (FLAG-RING1B) and lysates were subjected to immunoprecipitation with an antibody against the FLAG epitope tag. Immunoblots against the HA tag revealed expression of tagged PCGF5 protein in the presence and absence of exogenous FLAG-RING1B. The same HA immunoblot of the FLAG immunoprecipitates revealed that PCGF5 was present in the FLAG-RING1B-transfected cells but not the empty vector transfected cells, demonstrating the PCGF5 specifically co-precipitated with RING1B (Figure 5.2.6A). The same strategy was employed to perform the reciprocal experiment. An anti-HA antibody was used to precipitate HA-PCGF5 and FLAG immunoblot revealed co-precipitation of FLAG-RING1B (Figure 5.2.6B) lending further support to the hypothesis that PCGF5 and RING1B interact.

The observation that PCGF5, like Bmi1, associated with RING1B prompted the question whether this interaction occurred in competition with Bmi1 or if the two co-occupied the same PRC1-like complex. Immunoprecipitates for HA-PCGF5 were probed with an antibody against endogenous Bmi1, and while robust overexpression of Bmi1 was achieved, it was not detected to co-precipitate with HA-PCGF5 (Figure 5.2.7A). Since PCGF5 and Bmi1 did not seem to occupy the same complex, the potential that these two proteins were in competition for occupancy of the complex was investigated.

Immunoprecipitates were performed for FLAG-RING1B and probed for endogenous Bmi1 in the presence and absence of overexpressed HA-PCGF5. The quantity of endogenous Bmi1 co-precipitated with FLAG-RING1B was less when HA-PCGF5 was overexpressed (Figure 5.2.7B). Over three independent experiments, an average of 33% less co-precipitation was observed and this decrease was significant (p < 0.001). This supports the hypothesis that Bmi1 and PCGF5 are, in fact, in competition for
Figure 5.2.6 – PCGF5 interacts with RING1B  A) PCGF5 is present in immunoprecipitates of RING1B. Whole Cell Lysate (WCL) and anti-FLAG immunoprecipitates (IP FLAG) from 293T cells transfected with the indicated plasmids were probed for either FLAG (top) or HA (bottom) antibodies. B) Reciprocal immunoprecipitation showing RING1B is present in immunoprecipitates of HA-tagged PCGF5. Whole cell lysate (upper panel) probed with antibodies against HA-PCGF5 and immunoprecipitates of HA-PCGF5 probed with FLAG-RING1B (bottom panel) from 293T cells transfected with the indicated plasmids. Bands from the immunoglobulin heavy (IgH) and light (IgL) chains carried over from the immunoprecipitation are indicated.
Figure 5.2.7 – Bmi1 and PCGF5 compete for association with RING1B  

A) Bmi1 is not detectable in immunoprecipitates of HA-tagged PCGF5. Whole cell lysates (left panel) and immunoprecipitates of HA-PCGF5 (right panel) from 293T cells overexpressing HA-PCGF5 were probed with antibodies against native Bmi1 or the HA tag on PCGF5. 

B) Bmi1 co-precipitated with RING1B decreases when PCGF5 is overexpressed. Immunoprecipitates of FLAG-RING1B (FLAG IP) and whole cell lysate (WCL) were probed with antibodies against endogenous Bmi1 (top panel) or the HA tag of PCGF5 (bottom panel). WCL lanes were loaded with 1/10 total protein mass used for immunoprecipitation. Similar results were obtained in 3 independent experiments.
5.2.5 PCGF5 overexpression induces increased total H2AK119 ubiquitylation

Since the PcG proteins exert their effects by depositing specific modifications on histones, the impact of enforced PCGF5 expression on the levels of H2A ubiquitylated at K119 (H2AK119ub) was investigated. This modification is the hallmark of PRC1 activity and is recapitulated in vitro by the binary complex of the RING domains of Bmi1 and RING1B and loss of either of these proteins leads to a global decrease in ub-H2A levels.

To address this question, mouse embryonic fibroblasts (MEFs) were infected with retroviruses containing HA-PCGF5 and HA-Bmi1. The cells were left for three to seven days to allow transgene expression and for its activity to take effect. Histones were then extracted from the cells, resolved by SDS-PAGE then probed with an antibody specific for K119-ubiquitylated histone H2A. This signal was normalized against total histone H3 protein levels to determine whether total H2AK119ub was changing with addition of either HA-PCGF5 or HA-Bmi1. In MEFs, HA-Bmi1 overexpression was found to increase H2AK119ub 1.9-fold and HA-PCGF5 was found to induce a similar increase of 2.1-fold over control (Figure 5.2.8A and C). Thus, PCGF5 overexpression causes increased H2AK119ub levels and does so to a similar magnitude to Bmi1.

The increase observed upon overexpression of Bmi1 or PCGF5 represented the cumulative effect of the exogenously expressed genes in addition to whatever amount of H2AK119ub existed in the MEFs at steady state due to the endogenous PcG machinery. MEFs derived from Bmi1-null mice [31] were employed in the hopes that the absence of endogenous Bmi1 would decrease the background levels of H2AK119ub and increase the signal to noise ratio of the experiment. In these cells, overexpression of Bmi1 and PCGF5 each resulted in increased H2AK119ub levels. However, the exaggerated effect hoped for did not materialize. Bmi1 and PCGF5 overexpression each resulted in an increase in H2AK119ub of 1.4-fold (Figure 5.2.8A and C).
Figure 5.2.8 – Bmi1 and PCGF5 overexpression correlate with increased H2AK119ub

A) WT (top left panel) and Bmi1-/- (top right panel) mouse embryonic fibroblasts were infected with control virus, or virus contain Bmi1 or PCGF5 as indicated. Histones were extracted with acid, then resolved by SDS-PAGE and immunoblotted with antibodies specific for histone H2A ubiquitylated on K119 (H2AK119ub) and histone H3 (H3). B) Expression of Bmi1 and PCGF5 were confirmed in Bmi1-/- cells by immunoblotting for their HA epitope tags. C) H2AK119ub signal was normalized for each sample to total H3 and then the fold change for Bmi1- or PCGF5-transduced cells over empty virus was calculated. The mean fold change for all experiments is shown for WT (n=2) and Bmi1 null (n=4) and significant changes from control are indicated by ** for p < 0.01 and * for p < 0.05. Error bars indicate standard deviation.

Expression of the transgenes was confirmed by western blotting of the RIPA lysates with anti-HA
antibodies which recognized only the exogenous forms of the proteins. Consistent with previous experiments, and previous constructs, Bmi1 was expressed to a much higher degree than PCGF5 (Figure 5.2.8B) despite being expressed from the same vector which was observed to induce a similar fold increase in Bmi1 and PCGF5 mRNA. This suggests either that translation of PCGF5 transcripts is less efficient than for Bmi1, or that PCGF5 is more rapidly degraded than Bmi1.

5.2.6 Regulation of gene expression by PCGF5

Given that PCGF5 seemed to mediate ubiquitylation of histone H2A, and that this histone mark is known to repress gene transcription, it was of interest to determine whether PCGF5 overexpression caused changes in gene expression. Bmi1 has been shown to be a potent regulator of Hox gene expression and the Gao et al. ChIP-seq dataset revealed Bmi1 SERs within the HoxA, HoxB and HoxC clusters [52]. PCGF5 had one SER between HoxA9 and HoxA10, but no SERs were detected in the HoxB and C clusters (Appendix Figure A6). However, the ChIP-seq experiment was performed in human HEK 293T cells while the analyses of gene expression was performed in the same system as the histone modification assays, namely, in mouse embryonic fibroblasts. Nonetheless, Hox gene expression was measured under the influence of overexpressed Bmi1 or PCGF5.

MEFs were infected with control retrovirus, or retrovirus containing Bmi1 or PCGF5. After between 5 and 10 days of culture, mRNA was harvested from cells and expression of a subset of Hox genes was measured by qPCR from cDNA prepared by a reverse-transcriptase reaction.

PCGF5 expression caused some Hox genes to be repressed to a similar degree to Bmi1 (HoxA9) while other genes were not significantly affected by either Bmi1 or PCGF5 (HoxA10 and HoxB7). In other cases, Bmi1 expression caused diminished Hox gene expression while PCGF5 had did not have a significant effect (HoxB4 and HoxC8) as shown in Figure 5.2.9A. There were no significant differences between the Bmi1-expressing and PCGF5-expressing samples. While widely generalizable conclusions
are difficult to draw due to the established tissue specificity of the effect of the PcG on Hox gene expression, PCGF5 seems to have overlapping, but non-identical consequences on Hox genes in this context.

The level of overexpression was also assessed by this method and found to be 5.2-fold for Bmi1 and 4.3-fold for PCGF5 at the level of mRNA (Figure 5.2.9B). The presence of protein was also assessed by western blot with antibodies against the HA epitope tag on Bmi1 and PCGF5 (Figure 5.2.10). As in Figure 5.2.8B, PCGF5 was expressed at a much lower level than Bmi1.

Since the Ink4a locus is also a well-established target of Bmi1 repression its expression was measured under conditions of enforced PCGF5 expression. In the Gao et al. dataset, there was one Bmi1 SER upstream of this locus, while none were detected for PCGF5. Again, the caveat of species and tissue difference hinders this comparison.

To measure the relative effects of Bmi1 and PCGF5 on Ink4a expression a primary mouse leukemia that had been expanded in vitro for two weeks was identified which expressed low levels of Ink4a and ARF and that induced higher expression upon Notch inhibition. This observation was consistent with the hypothesis that Notch-driven PCGF5 was repressing Ink4a expression, which would then be upregulated upon Notch inhibition.

These leukemia cells were, therefore, infected with retroviruses containing Bmi1 and PCGF5 then enriched to greater than 90% purity. Cells were treated with GSI to block Notch then RNA was prepared for RT-PCR. In these cells, Bmi1 overexpression resulted in moderately decreased Ink4a expression though this change was not quite significant (p=0.28). PCGF5 caused no change in Ink4a expression (Figure 5.2.11). Thus, PCGF5 did not seem to be responsible for the reduction in Ink4a expression in the presence of a strong Notch signal.
Figure 5.2.9 – PCGF5 and Bmi1 modulate expression of a subset of Hox genes in MEFs

MEFs were infected to high efficiency with empty vector control retroviruses or retroviruses encoding Bmi1 or PCGF5. qPCR was performed on the transduced cells to assess expression of a subset of Hox genes (A) and of Bmi1 and PCGF5 to confirm overexpression at the mRNA level (B). qPCR performed in triplicate normalized to β-actin. Shown relative to the sample with the lowest expression in each set. Error bars indicate standard deviation (* indicates p < 0.05; ** indicates p < 0.01).
However, in wild type B6 MEFs, PCGF5 reduced Ink4a expression to a similar degree as Bmi1 though neither caused a significant change (p=0.78 and p=0.13, respectively). Furthermore, in two other strains of MEFs, neither Bmi1 nor PCGF5 significantly altered Ink4a expression. The caveat here, however, is that these strains of MEFs carried a targeted deletion of exon 1β of ARF. Since Ink4a and ARF share exons 2 and 3, it is possible that deletion of the first exon for ARF causes a disruption in the regulation of Ink4a.

It is difficult, therefore, to draw firm conclusions about which genes, specifically, are regulated by PCGF5. Overexpression of PCGF5, however, does seem to impact expression of some of the same genes targeted by Bmi1. However, the paradigm of tissue specific effects for the PcG prevails in this analysis.

Figure 5.2.10 – Transduced cells express HA-tagged Bmi1 and PCGF5 protein
Immunoblot from lysates of MEFs transduced to high efficiency with empty vector, Bmi1, or PCGF5 retrovirus resolved by SDS-PAGE and probed for the HA epitope tag carried by both Bmi1 and PCGF5.
5.3 Discussion

As discussed in Chapter 3, PCGF5 was identified as a transcriptional target of the Notch signalling pathway. At the time, its regulation by Notch and its primary sequence were all that were known about this gene. Since the hypothesis regarding its potential function in hematopoiesis could not be supported, further studies were focused on confirming its identity as a Polycomb Group gene and characterizing the biochemical activity of the PCGF5 protein.
5.3.1 Protein structure and interactions

It was first noted that the N-terminal RING finger domain of PCGF5 was highly homologous to those of other PCGF family members, such that the critical histidine and cysteine residues identified in Bmi1 were all present in PCGF5 and, despite there being significant leeway in the consensus for RING fingers, the spacing between these residues was also identical between Bmi1 and PCGF5.

It is via the RING domain that Bmi1 associates with the ubiquitin ligase RING1B to mediate ubiquitylation of histone H2A. Indeed, truncations of Bmi1 and RING1B containing only their RING domains are competent for mediating the transfer of ubiquitin to H2A in vitro [64]. Also conserved is the region around Lys81, the paralogous region of which has been observed to be ubiquitylated on Bmi1. Ubiquitylation of PCGF5 has not yet been reported, but it would be of considerable interest to address this issue directly. Ubiquitylation might be of particular interest for PCGF5 since, in the course of these studies, achieving high levels of overexpression proved difficult. It could be that high levels of PCGF5 protein trigger an active degradation mechanism preventing the expression of equivalent levels of PCGF5 to those observed for Bmi1 expressed via the same construct and promoter. Direct assessment of whether ubiquitylated PCGF5 can be detected, as well as studies of mutants lacking the conserved lysine could shed some light on this hypothesis.

While the aforementioned features are highly conserved between Bmi1 and PCGF5, there are differences as well. The basic Lys62 and Arg64 residues of Bmi1 observed to interact with DNA are not conserved in PCGF5 [65]. It was demonstrated that a Bmi1 allele carrying mutations at these positions was less efficient at causing ubiquitylation of histone H2A and suggested these residues are important for orienting the Bmi1-RING1B complex correctly. This suggests either that PCGF5 relies on another member of the complex for DNA binding or that it binds at a different location in its primary structure.
Given the differences described by Gao et al. in the nature of the extended complex for Bmi1 and PCGF5, it could be that PCGF5 relies on a different mechanism for its interaction with DNA.

In contrast with the N-terminus, there is little conservation towards the C-terminus between PCGF5 and Bmi1 except for a few discrete regions. What implications could a divergent C-terminus have?

Interaction of Bmi1 with the mouse Ph homologue, Rae28, was found to be dependent on a region outside of the RING domain [244]. Since PCGF5 does not appear to interact with Rae28 [52] it is likely that the non-conserved C-terminus plays a role in the different constituents of Bmi1- and PCGF5-containing complexes described by Gao et al. The similarity between the C-termini of PCGF5 and PCGF3, however, suggest that there is still strong selection for conservation of PCGF5 C-termini. Given that AUTS2 and FBRS have been described to interact with PCGF5 and PCGF3 but no other PCGF family member, it may be that these proteins bind near the C-terminus of PCGF5.

Further study of the interaction between PCGF5 and AUTS2 and FBRS would be particularly interesting because AUTS2 and FBRS have not previously been linked to Polycomb function. Elucidation of their role would enrich our understanding of the diversity of PRC1-like complex composition and function. Reciprocally, a greater understanding of PCGF5 and PCGF3 function may cast light on the genetic associations described for AUTS2, such as autism [245] and addiction [246], and for FBRS, such as fibrosis and wound healing [247].

A search for recognition motifs conserved between PCGF5 and PCGF3 revealed a predicted CK2 phosphorylation site. While this remains to be confirmed experimentally, it is interesting to note that CK2 was among the proteins detected by Gao et al. that bound specifically to PCGF5- and PCGF3-containing complexes. It would be of interest to determine whether this site is in fact phosphorylated, whether CK2 is involved, and whether phosphorylation modulates the activity or stability of PRCs containing PCGF5 or PCGF3.
While the C-terminus of PCGF5 bore little similarity to that of Bmi1, the high degree of sequence conservation found at the N-terminus supported the hypothesis that PCGF5 would form a complex with RING1B that is capable of ubiquitylating histones. This hypothesis was, in fact, borne out by co-immunoprecipitation studies. Overexpression of PCGF5 was found to reduce Bmi1 co-precipitation with RING1B. This is consistent with what Gao et al. described in their mass spectrometry data [52]; however, this study is the first to corroborate the mass spectrometry data by reciprocal co-immunoprecipitation followed by western blot.

This experiment relied on densitometry of chemiluminescence-exposed film after immunoblotting, an approach that presents challenge in drawing quantitative results. A repeat of these experiments using a more quantitative assay such as fluorometry of immunoblots would lend further support to our conclusions.

Additional supporting experiments, such as the reciprocal abrogation of PCGF5:RING1B interaction by Bmi1 overexpression posed technical challenges in this study but would be valuable. Evaluation the RING1B:Bmi1 association upon expression of a PCGF5 mutant lacking binding affinity for RING1B would also be an informative control.

PCGF5 had previously been detected by mass spectrometry in immunoprecipitates of RING1A as well [53]. In my hands, PCGF5 and RING1A were observed to interact by co-immunoprecipitation and western blot in one experiment. However, this was not replicated due to technical problems that arose with the immunoblot suspected to be caused by degradation of the anti-RING1A antibody reagent.

**5.3.2 Evidence for PCGF5-mediated ubiquitylation of H2AK119**

It was also demonstrated in MEFs that increased expression of either PCGF5 or Bmi1 induces an increase in the level of H2AK119ub.
Bmi1-null cells failed to deliver a greater fold change in H2AK119ub which could be due to a commensurate decrease in RING1B levels in the absence of Bmi1 resulting in reduced capacity to modify histones. Indeed, the association of RING1B with Bmi1 was been shown to enhance its stability [248]. Alternatively, the expected reduction in H2AK119ub levels expected in the absence of Bmi1 may have been compensated for by the upregulation of another PCGF family member. Consistent with this idea, the converse result, a slight diminution of PCGF5 message upon overexpression of Bmi1 in MEFs was observed (Figure 5.2.9B).

Bmi1 and PCGF5 seem equally potent in this regard, although PCGF5 achieved a similar magnitude of H2AK119ub increase despite level of its transgene expression routinely being 4 to 5-fold less than that achieved for Bmi1. This observation supports the notion that PCGF5, despite its differences in terms of binding partners and genome targeting, possesses a similar biochemical activity to Bmi1. However, it is possible in this experiment that enforced expression of PCGF5 induced other changes in the cell, such as modulation of gene expression or activity or abundance of other proteins, which had the downstream consequence of changing H2AK119ub.

Changes in H2AK119ub levels were ascertained by densitometry of film exposed by the chemiluminescence normalized to total histone H3 in immunoblotting experiments. The histone H3 signal was, as expected, much greater than the H2AK119ub signal which represents only the fraction of H2A bearing the modification being probed. As such, it was difficult to ensure both the H2AK119ub and H3 bands in the immunoblot were in the linear range of the film. Experiments were conducted running a single set of samples and then cutting the membrane between the expected sizes of H3 and H2A and then probing each separately. Alternatively, some experiments were conducted by running two sets of samples with 1/3 as much sample being loaded in the lanes to be probed for H3 as those to be probed for H2A. The results by both methods showed similar relative levels of H2AK119ub between control and
PCGF5- or Bmi1-transduced samples. However, as with the competition experiments described in section 5.2.4, an improved method of making quantitative measurements of H2AK119ub would by immunoblot followed by probing with fluorescent labelled antibodies and fluorometry.

Interestingly, while PCGF5 seemed to perform similarly to Bmi1 in this cell-based assay, differences were found in a ChIP-seq experiment recently published [52]. There was substantially less H2AK119ub at PCGF5 enriched regions than Bmi1 enriched regions (see Appendix Figure A2). This difference seems largely attributable to the absence of RING1B at the majority of PCGF5-bound sites, while Bmi1 and RING1B were frequently observed to co-localize on chromatin. When RING1B was observed to occupy PCGF5-bound sites, H2AK119ub was very high (Appendix Figure A4). Therefore, the lower average H2AK119ub observed at PCGF5 SERs is due to the abundance of sites bound by PCGF5 where no RING1B is observed.

The absence of RING1B at 90% of PCGF5-bound sites contrasted with the co-immunoprecipitation data described in this chapter and the mass spectrometry data in Gao et al. [52] showing that PCGF5 and RING1B do interact. There could be several reasons for this:

1. PCGF5 may preferentially interact with RING1A over RING1B
2. The nature of the PCGF5:RING1B complex may be such that the epitope-tag used to precipitate RING1B has sterically obstructed
3. PCGF5 may have a RING1B-independent activity at some or most of its target loci

With respect to the first possibility, while the canonical PRC1 complex is thought to contain both RING1A and RING1B, these studies rely on precipitation and detection of complexes that are quite heterogeneous [51-53, 79, 249]. It is not clear that all PRC1-like complexes contain both RING1A and RING1B. The possibility of such biases in affinity has not been extensively studies for any PCGF family members. Indeed, in the Gao et al. paper, their mass spectrometry data is presented as a heat map
representing the abundance of the co-immunoprecipitated species [52]. Their study suggests that Mel-18 preferentially binds RING1B, while PCGF3, PCGF5 and Bmi1 pulled-down more RING1A. This, at least, suggests there is a precedent for varying binding affinities.

The second possibility could potentially be experimentally verified by using a different approach to precipitation – either by using antibodies against native RING1B or against a protein with the tag appended to the C-terminus instead of the N-terminus. It would also be interesting to try to measure the stoichiometry between PCGF5 and RING1B in precipitated complexes by silver staining of a gel, or similar technique. If the PCGF5 co-precipitated equimolar amounts of RING1B, this would support the notion that RING1B is present at PCGF5-bound sites on chromatin, but somehow is not being detected.

Lastly, PCGF5 may split its duties between a RING1B-dependent role and a RING1B independent role. Indeed, if only the PCGF5 sites that are co-occupied by RING1B are considered very high levels of H2AK119ub are observed, higher even that Bmi1:RING1B co-bound sites. Thus, the 90% of sites enriched by PCGF5 have very low H2AK119ub and may lack the RING1A/B histone ubiquityl-transferase activity. It is unknown what function PCGF5 may have at these sites.

5.3.3 Implications of PCGF competition for RING1B interaction

Section 5.2.4 describes experiments providing evidence that PCGF5 and Bmi1 compete for RING1B. Thus, change in expression in one may alter the activity of the other presenting a means by which signalling pathways, such as Notch, may fine tune Polycomb function. This is an interesting possibility given that the identity of the PCGF member determines the nature of the extended complex.

For instance, Bmi1- and Mel-18-containing complexes contain ‘canonical’ PRC1 members such as the CBX and PHC families. While the core PCGF-RING1 association is preserved for the remainder of the PCGF proteins, they each form complexes devoid of CBX and PHC proteins [52]. Instead, PCGF5, as well
as the related PCGF3, associate with AUTS2 and FBRS, proteins not previously associated with Polycomb function. Thus, by titrating Bmi1 from RING1B, an increase in PCGF5 expression may have substantial qualitative, rather than simply quantitative, effects on PRC1 function.

In addition to changes in the composition of the PRC1 complex, substitution of different PCGF family members likely changes the spectrum of loci targeted for repression. Genes targeted by one PCGF family member are infrequently targeted by another, suggesting the member has its own array of targets. Indeed, a gene ontology analysis revealed that different PCGF family members seemed to preferentially modulate genes associated with distinct biological processes.

The hypothesis that Bmi1 and PCGF5 compete for RING1B under physiological conditions relies on RING1B being limiting; however, it is unknown whether such a state is common. Indeed, the observation that addition of Bmi1 or PCGF5 is able to increase total H2AK119ub suggests a pool of free RING1B, at least in the context of the MEFs studied. Alternatively, Bmi1 and PCGF5 might titrate RING1B away from yet another PCGF family member that is less active than either.

Since the different PCGF family members were observed by Gao et al. to target different regions of the genome [52] substituting one PCGF for another in PRC1 complexes could have profound consequences on gene expression and may be an important mediator of development.

**5.3.4 PCGF impacts on gene expression**

The studies performed on the effect of PCGF5 on H2AK119ub measured changes in the global level of this chromatin mark. It is unclear what the consequence of increased total H2AK119ub may be on cellular physiology. Since the PRC1 complexes act to modify chromatin in a manner the represses nearby gene expression, it would be much more informative to examine the change in the landscape of H2AK119ub across the genome while modulating PCGF5 to determine what genes might be affected.
Coupling such an analysis with gene expression profiling would give a detailed picture of the consequences of PCGF5 expression. Further, comparison of this data against a parallel set modulating Notch instead of PCGF5 would allow identification of genes whose expression is regulated by Notch via PCGF5. This would be of particular interest in T-ALL cells since Notch plays such a critical role in T-ALL biology and it was in this context that PCGF5 was observed to be regulated by Notch.

The consequences of PCGF5 overexpression on a subset of genes, chosen because they had been previously identified as being regulated by Bmi1, were reported. These genes were HoxA9, HoxA10, HoxB4, HoxB7, HoxC8 and Ink4a. Cross referencing with Gao et al.’s ChIP-seq data [52] confirmed that Bmi1 did in fact enrich regions near each of these genes, while SERs for PCGF5 were only observed in the HoxA cluster (see Appendix Figure A6).

These gene regulation results were broadly concordant with the ChIP-seq data. Overexpression of Bmi1 diminished expression of HoxA9, HoxB4 and HoxC8, each of which was enriched by Bmi1 ChIP-seq. However, among the transcripts assayed, only HoxA9 was diminished by PCGF5 overexpression and indeed only the HoxA cluster contained SERs for PCGF5.

The consequences of Bmi1 and PCGF5 expression on Ink4a expression varied and were in some cases cell type dependent. This is consistent with the role of PcG genes as maintenance factors in gene expression rather than directly up- or down-regulating specific genes in the manner of traditional transcription factors. It should be noted that measurement of Ink4a transcripts often required 30 or more cycles of PCR, which likely contributed to the difficulty in achieving the consistent measurements required to discern a significant effect.

It would be of interest to observe changes in gene expression during developmental transitions under gain- and loss-of function of PCGF5. This strategy has been useful in the past for Bmi1, and would reveal genes which require PCGF5 for the maintenance of their expression profile.
The data described herein identifies *PCGF5* as a true member of the Polycomb Group by virtue of its interaction with members of the PRC1 complex and its association with an ubiquityl transferase activity towards histone H2A. Future studies will be needed to determine the role PCGF5 has in physiology and gene regulation.
Chapter 6  Conclusions and perspectives

This thesis describes the identification of the Polycomb Group gene PCGF5 as a transcriptional target of the Notch pathway in T lineage cells. It further provides evidence that PCGF5 mediates ubiquitylation of histone H2A and is able to impact gene expression.

While PCGF5 has significant sequence homology to Bmi1, this homology is limited to the N-terminal RING domain suggesting the C-terminal portion may confer functionality not shared with Bmi1. PCGF5, and PCGF3, however, share sequence homology over their entire amino acid sequences and seem to form a distinct subfamily within the extended PCGF family. While PCGF5 associates with RING1B like Bmi1, the composition of the extended complexes differs between the two. Further, the genomic loci which they bind, and presumably repress transcriptionally, also differ.

Here, I will discuss the implications of Notch regulation of Polycomb activity and what role Notch-driven PCGF5 may play in development. Also, I will speculate as to how PCGF5 fits into the greater mammalian Polycomb Group. I will also point to several unanswered questions regarding PCGF5 and suggest experiments by which they may be addressed.

6.1 Polycomb complexes: shuffling the deck

The mammalian Polycomb Group is made up of members of approximately ten different gene families. Some are represented in the genome by a single gene while others have been the subject of gene duplication resulting in multiple paralogous genes.

It is likely that the some members of both PRC1 and PRC2 are required to be expressed more or less constitutively, lest all PcG activity be lost. However, other members which have been subject to gene duplication resulting in several paralogous copies may have some functional redundancy. Over the course of hundreds of millions of years, a time course supported by the presence of multiple paralogues
of both the CBX and PCGF families in birds and frogs as well as mammals, this redundancy would allow for the division of labour between several genes and for the evolution of specialized activities. Further, these genes could be modulated individually to fine tune PcG activity. An example of how different PcG family members may be regulated by different means is the serine-rich motif in Bmi1 whose phosphorylation by Akt regulates modulates ubiquityl transferase activity [243]. This motif is not conserved in the rest of the PCGF family presenting a mechanism of regulation specific to Bmi1.

This redundancy is of particular importance for the CBX and PCGF families which are the largest, containing five and six known members, respectively. It is worth noting the PRC1 complex is particularly heterogeneous, whereas in the PRC2 complex only one member, EZH1/2, is represented by multiple genes.

The argument for at least partial redundancy among PcG paralogues is supported by genetic evidence. Mice lacking Eed or Suz12, which have no paralogues, are not viable [250, 251] nor are mice lacking EZH2 [252], whose single paralogue EZH1 is not expressed ubiquitously [253, 254]. RING1B-deficient mouse embryos also suffer a developmental arrest at gastrulation [255].

One the other hand, mice lacking Bmi1 [31], Mel-18 [35], of the six strong PCGF family, or M33 [36, 256] and CBX8 [257], of the five member Pc/CBX family, show milder phenotypes, each being viable at least until birth.

Taken together, the knockout data suggests that while PcG activity is required for viability, the genes present as multiple paralogues have overlapping, but not wholly redundant roles, and operate collaboratively or in parallel to mediate proper development. Modulation of individual family members could, thus, subtly alter the activity and specificity of PRC1 to achieve the different transcriptional profiles required for different tissues and developmental stages. This hypothesis was first suggested by Satijn and Otte in their 1999 review [258].
<table>
<thead>
<tr>
<th>Gene</th>
<th>Viability</th>
<th>Tissues</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>PRC1</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>M33</strong></td>
<td>Viable</td>
<td>Skeletal transformation, hematopoietic defects</td>
<td>[36]</td>
</tr>
<tr>
<td><strong>CBX8</strong></td>
<td>Viable</td>
<td>n/a</td>
<td>[257]</td>
</tr>
<tr>
<td><strong>Rae-28</strong></td>
<td>Viable and fertile</td>
<td>Skeletal transformation</td>
<td>[38]</td>
</tr>
<tr>
<td><strong>Ring1a</strong></td>
<td>Viable and fertile</td>
<td>Skeletal transformation</td>
<td>[38]</td>
</tr>
<tr>
<td><strong>Ring1B</strong></td>
<td>EL; 9.5dpc</td>
<td>Gastrulation defect</td>
<td>[255]</td>
</tr>
<tr>
<td><strong>Bmi1</strong></td>
<td>Viable, fertile</td>
<td>Skeletal transformation, hematopoietic defects, neurological abnormalities</td>
<td>[31]</td>
</tr>
<tr>
<td><strong>Mel-18</strong></td>
<td>Viable at birth, death around 4wks</td>
<td>Skeletal transformation, hematopoietic defects</td>
<td>[260]</td>
</tr>
<tr>
<td><strong>PRC2</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Eed</strong></td>
<td>EL; 8.5dpc</td>
<td>Impaired mesoderm production, gastrulation defects</td>
<td>[261]</td>
</tr>
<tr>
<td><strong>Suz12</strong></td>
<td>EL; 7.5-8.5dpc</td>
<td>Gastrulation defects</td>
<td>[251]</td>
</tr>
<tr>
<td><strong>EZH2</strong></td>
<td>EL; 7.5dpc</td>
<td>Gastrulation defects</td>
<td>[252]</td>
</tr>
</tbody>
</table>

**Table 6.1.1 Phenotype of loss-of-function for PcG genes** (EL: embryonic lethal; PNL: perinatal lethal) No data is available for CBX5, 6, 7 or PCGF1, 3, 6.

The data referred to in Chapter 5 by Gao et al. suggested that the identity of the PCGF family member associated with RING1B determined the composition of the PRC1 complex [52]. They also found that PCGF5-containing complexes did not contain members of the CBX family. This was in conflict with another study which detected PCGF5 in immunoprecipitates of CBX8; however, PCGF5 was not
precipitated with CBX2, 4, 6 or 7 although, in another study, PCGF5 was able to interact with CBX6, 7, and 8 \textit{in vitro} [51].

Taken as a whole, however, the study by Vandamme et al. does not refute Satijn and Otte’s hypothesis as they performed immunoprecipitation for each of five CBX family members and found significant differences in the proteins which co-precipitated. For instance, RING1B associated with each CBX member and Bmi1 was associated with each but CBX6. By contrast, PCGF1, 3, and 5 associated with only CBX8 and the histone demethylase JHD2a associated with only CBX2. Thus, while these two proteomic studies are not completely concordant, they at least agree that PRC1 complexes are heterogeneous and that substitution of PCGF or CBX family members can dramatically alter the spectrum of proteins recruited to the complex. The diagram summarizing the interactions observed by Gao et al. is shown in Figure 6.1.1. There is evidence that such substitution can dramatically alter the repertoire of genes targeted for repression [52].

While the hypothesis that different tissues may favour the expression of different complexes is appealing, evidence in its support is lacking. Assessing gene expression by microarray may shed some light on this issue; however, in a dynamic system in which RING1B may be limiting there may be free complex members that are inactive due to a lack of binding partners. Ideally, quantitation of the PCGF- and CBX-family members in immunoprecipitates of RING1B, by either mass spectrometry or western blot, could give a ratio of each PRC1 complex member to the core RING1B subunit. Completing this analysis for a variety of cell types might reveal a signature of the prevailing PRC1-like complexes expressed in each tissue, should such a signature exist.

If this model is correct, then the precise regulation of the expression of these PcG components must be critical for maintaining the fidelity of developmental processes. However, surprisingly little is known about how members of the PCGF and Pc families are regulated.
6.2 Transcriptional regulation of the Polycomb Group

Upon review of the phenotypes described in the literature for loss of various PcG genes as well as those observed for overexpression, which are often dose-dependent, it seems clear that PcG protein expression must be tightly regulated to enable proper physiologic functioning and development.
Both E2F1 [262] and MYCN [263] have been described to enhance expression of Bmi1; however, these seem like unlikely candidates to mediate fine tuning of Polycomb complexes between tissues as neither is known to respond proximally to extracellular developmental signals.

Pathways such as Notch, Wnt, TGF-beta are important regulators of development and fate choice and are able to respond to signals derived from outside the cell. Thus, implication of any of these pathways in the regulation of PcG components might offer insight into how tissue specific expression is achieved. Indeed, it has been suggested that Bmi1 may be regulated by Wnt signalling in colon cells [264] and by the Notch ligand DL4 in naïve CD4 T cells; however, this activity is highly specific as it is not preserved in effector T cells, nor does the Notch ligand Jagged-1 confer expression of Bmi1 [265]. It was also observed in the course of this research that inhibition of Notch decreased Bmi1 transcripts.

This research implicates the Notch pathway in the regulation of PCGF5 in T cell leukemia cells. It is likely this link is conserved in some normal tissues as well. Overexpression of PCGF5 in thymocytes was found not to significantly perturb their development but the Notch pathway is highly active in these cells which may drive high levels of endogenous PCGF5. Indeed, as shown in Figures 3.2.16 to 3.2.19, PCGF5 message was observed by microarray experiments to be high in thymocytes relative to other tissues. In these cells, a significant fraction of the PRC1-like complexes in thymocytes can be presumed to contain PCGF5, along with Bmi1 and Mel-18 which are known to serve critical roles in thymopoiesis [201, 266, 267].

Withdrawal of Notch signalling, either by transiting through the DN stage where Notch signalling and PCGF5 expression is highest, or by experimental manipulation, would thus cause a decrease in PCGF5 levels. If no other PCGF family member is present to substitute, this might lead to a global decrease in H2AK119ub levels and derepression of genes. Alternatively, another PCGF member may replace PCGF5
in PRC1 complexes and change, not the total amount, but the distribution of H2AK119ub marks, leading to the upregulation of PCGF5 target genes and repression of others.

This could be one mechanism by which cell fate is fixed during differentiation. Such mechanisms are of particular interest in the study of hematopoiesis because, unlike solid tissues, hematopoietic cells are mobile and must maintain their identity despite changes in their environment. For instance, the Notch signals that thymocytes depend on for their development are delivered by ligands which are fixed to cells. Therefore, once a T cell leaves the thymus, it must “remember” the instructions it received from this cells despite having no, or infrequent, stimulation of the Notch pathway once in the periphery.

6.3 PCGF5 and maintenance of repression

One more reason to take interest in the observation that Notch regulates a Polycomb gene is the role the PcG has in maintaining transcriptional repression, rather than establishing it [26, 268]. In *Drosophila* homeotic gene expression during segmentation is maintained even after the transcription factors which initiated their pattern of expression have themselves ceased being expressed. This maintenance function is dependent on the PcG.

Notch signalling increases expression of many genes, but also, via transcriptional repressors, causes the downregulation of genes as well. Many repressors are transactivated by Notch including those of the Hes and Hey families, the helix-loop helix protein Id3 [269], the zinc-finger-containing repressor Slug [224].

Thus, when a Notch signal is newly received by a cell, a complex array of gene expression changes ensue with the induction of direct and indirect Notch targets as well as the downregulation of genes targeted by Notch-driven repressors such as repression of *PTEN* by Hes in T-ALL [270] and of VE-Cadherin by Slug in embryonic cardiac development [224]. While some of these changes may be transient, others may be
permanent, particularly in a developing cell whose potential for giving rise to different cell types becomes increasingly limited during its differentiation.

So how can Notch, in a developing thymocytes for instance, block the capacity for cells to become B or NK cells, and instead lock in an expression programme required for a T lymphocyte. While Hes and other repressors transactivated by Notch may help establish that T lymphocyte transcriptional programme, it could be that Polycomb proteins, including PCGF5, are required to maintain the programme.

Indeed, loss of Notch in thymic progenitors leads to developmental arrest and reduced proliferation [271, 272]; however, cells to not change course and instead adopt a B lymphoid or myeloid fate. This lineage commitment suggests that, even before the completion of β-selection, ablation of Notch signalling leaves a residual T programme that cannot be overcome. It would be interesting to see if contemporaneous loss of Notch and PCGF5 in early thymic progenitors led to derestriction of lineage potential and the generation of non-T cells in the thymus.

6.4 Mechanism of repression
The paradigm that prevailed until recently, that PRC2-mediated histone methylation occurred first, and served as a signal to recruit PRC1, has fallen into some doubt. Studies have shown that the distribution of PRC1 complexes is similar in cells lacking PRC2 activity to wild-type cells, suggesting they are targeted independently of PRC2.

I will also make the argument here that the complexity of the PRC1-like complexes, versus the relative simplicity of PRC2, mean that it is unlikely that PRC2 serves as the determinant of PcG targets. The diversity of PRC1 complexes make them much better candidates, along with other co-factors, to provide the variable distribution of targets required for the transcriptional programmes of different cell types.
Thus, rather than operating by a sequential mechanism, it seems likely that PRC1 and PRC2 act as two parallel systems which engage in cross-talk with the other.

In this way, PRC2 and a heterogeneous array of PRC1-like complexes work in concert to maintain repression, such that loss of either class [78, 273] can lead to derepression. Further, the various PRC1-like complexes serve to recruit co-repressors such as Ncor2 and Bcor, additional histone modifying enzymes such as JHD2A, Fbxl10 and SETD2, kinases such as CK2 and FER and DNA methyltransferase DNMT as well as histone deacetylases, ubiquitin ligases and proteases [51-53]. In this way, the Polycomb proteins serve as a hub in the extended repression machinery recruited to silenced loci.

6.5 In vivo future directions

PCGF5 loss of function

None of the data presented in this thesis addresses whether a function for PCGF5 in hematopoiesis might be revealed in a loss-of-function experiment. Knock-down experiments were attempted using lentivirally expressed shRNA sequences targeting PCGF5 and but no phenotype was observed in these experiments. Since no antibody was available, it was not possible to determine whether the knock-down was successful, and since no phenotype was observed, less reliable means of measuring knock-down, such as qRT-PCR quantitation of mRNA, were not pursued.

However, future work examining the effects of loss-of-function of PCGF5, either via knock-down or conditional knock-out mouse, would be extremely enlightening. Indeed, most of the largest advances in the field have been achieved in the context of loss-of-function. Of course, the early work describing mutants in Drosophila was largely dependent on loss-of-function mutants which resulted in inappropriate derepression on genes during embryogenesis. Besides the initial identification of Bmi1 as a cooperating oncogene in the Eμ-myc mouse model in the context of a gain-of-function [27], loss-of-function experiments have provided significant insight into Polycomb biology in mammals as well.
The generation of a conditional PCGF5 knock-out mouse would be of particular value. Such a model could be used to examine the effect of PCGF5 deficiency in embryonic development to determine whether PCGF5 loss results in similar skeletal dysplasia to those found in mice deficient for Bmi1, Mel-18, and RING1A. Crossing these mice to animals expressing Cre-transgenes in specific hematopoietic lineages has the potential to uncover processes in blood development where PCGF5 is required. For instance, crossing mice with a LoxP site-flanked PCGF5 gene with Lck-Cre transgenic mice would reveal whether PCGF5 is required for T lymphopoiesis and T-ALL.

**PCGF5 in stem cells and self-renewal**

Bmi1 has well-established roles in the maintenance of hematopoietic stem cells and self-renewal. Bmi1-null fetal liver cells compete poorly with wild type cells in competitive repopulation assays and have reduced output in CFC assays [229229]. Conversely, Bmi1 overexpression increased colony output cells and increased chimerism in transplant recipient mice. However, these experiments focused specifically on CD34-KSL cells cultured for 10 days so cannot be fairly compared to the data presented here. It does, however, beg the question whether PCGF5 would have a similar effect under the same conditions, or whether it might result in reduced performance as was described by Iwama et al. for the Polycomb homologue M33/CBX2.

### 6.6 Conclusion

The characterization of PCGF5 in the work described here, as well as that by other groups, adds new understanding to the complexity of PRC1 complexes. The next challenge will be to map the tissue distribution of the various paralogues and to determine which signalling pathways and transcriptional regulators determine their tissue distribution. For PCGF5 expression in T cells, Notch seems critical; however, whether this is applicable to other tissues, and indeed, how the rest of the PcG are regulated remains to be determined.
Further, it will be of great interest to determine the gene expression profile imparted by the PcG and how alteration of the activity or expression of various PcG members may impact human disease.
References


Appendix

Contribution

This data was provided to me by Dr. Danny Reinberg’s lab as lists of filtered SERs. I generated the figures from pre-analyzed data using Cistrome Galaxy [152] as described in the Chapter 2 Materials and Methods.

A1 Introduction

This section describes my review and analysis of data published by Gao et al. [52] of ChIP-seq experiments performed for NSPc1, mel-18, Bmi1, PCGF5, MBLR, RING1B, CBX2 and the histone marks H3K27me3 and H2AK119ub. Significantly enriched regions (SERs) were shared by the authors. Their methods are described in the manuscript.

Shortly after our observation that enforced expression of PCGF5 caused an increase in global H2K119ub levels in mouse embryonic fibroblasts ChIP-seq data was published and made publically available for the PCGF family, including PCGF5. This allowed us to determine whether PCGF5 co-localized on chromatin with its binding partner, RING1B, and the histone mark which it is assumed to help catalyze, namely H2AK119ub.

A2 PCGF5 binds chromatin at locations distinct from Bmi1

Review of Gao et al.’s filtered ChIP-seq data [52] for the PCGF family, RING1B as well as several histone marks revealed that, in their analysis, Bmi1 was enriched many more regions than PCGF5; specifically 2374 SERs were found for Bmi1 while only 668 regions were significantly enriched by PCGF5. This compared with 1364 for PCGF1/NSPc1, 921 for PCGF2/Mel-18, and 1387 for PCGF6/MBLR. No data was available for PCGF3. So while PCGF5 had the fewest SERs, it was no further from the mean of 1342 than Bmi1, albeit on the low end instead of the high end (Figure A1.C). Further, despite their high similarity in terms of sequence and complex composition, Bmi1 and Mel-18 occupied near opposite ends of the spectrum with Bmi1 being associated with the largest number of SERs and Mel-18 having fewer than average.

Unfortunately, data was not available to show relative expression of Bmi1 and PCGF5 so it cannot be determined whether this difference was simply a result of varying expression levels. Interestingly, only
32 SERs were shared between Bmi1 and PCGF5 (1.3% of Bmi1 were also enriched by PCGF5; 4.8% for PCGF5 SERs were also enriched by Bmi1) suggesting the PCGF5 is recruited to distinct loci than its paralogue (Figure A1.A).

Since the PCGF-family is thought to function as part of the PRC1 complex in association with RING1B, the proportion of PCGF5 sites co-occupied by RING1B was determined. In fact, only 66 (9.9%) of PCGF5 SERs overlapped RING1B SERs, while 55% of Bmi1 SERs were also enriched by RING1B (Figure A1.B).

**Figure A1 PCGF5 has a different target repertoire than Bmi1**  
A) Venn diagram showing significantly enriched regions (SERs) for PCGF5 in red and for Bmi1 in blue. Overlapping regions are SERs that share at least 1 bp between them. The number of unshared SERs for each, and the number of overlapping SERs are indicated beneath the figure. B) Histogram showing the percentage of SERs for Bmi1 and PCGF5 which are shared with RING1B. C) The distribution of SER abundance for the PCGF family. SER numbers for Bmi1 and PCGF5 are indicated.
Since the PRC1 complex and the PCGF family proteins it contains are thought to repress gene expression, it was of interest to determine whether PCGF5 and Bmi1, despite not binding the same genomic regions, at least targeted a common set of genes. Lists were generated of all genes whose TSS was within 300 kb of an SER for both Bmi1 and PCGF5 which numbered 2844 and 790, respectively. The majority of genes targeted by both Bmi1 and PCGF5 (79.2% and 64.4%, respectively) were also enriched by RING1B. Further, it was observed that 258 genes had each Bmi1 and PCGF5 SER within 300 kb of their TSS which represented 32.7% of PCGF5-linked genes.

Figure A2 PCGF5 SERs correlate with increased H2AK119ub signal
Composite H2AK119ub signal normalized to input control for all Bmi1 SERs (red line) and PCGF5 SERs (blue line) is plotted over +/- 5kb from the centre of the SER. A taller peak indicates higher average levels of H2AK119ub.
A3 PCGF5 SERs are associated with increased H2AK119ub when RING1B is also present

Next, the correlation between PCGF5 enrichment and H2AK119ub levels was determined. **Figure A2** shows the average enrichment for H2AK119ub is much higher for Bmi1 SERs than for PCGF5 SERs; however, there PCGF5 SERs still showed modestly elevated H2AK119ub over the input control. Since a minority of PCGF5 SERs were co-bound by Bmi1, it was of interest to determine how much of the H2AK119ub signal in this composite was due to co-bound sites and might thus be attributed to Bmi1. To the end, the same analysis was performed for sites bound by PCGF5, but not Bmi1. The sites bound by PCGF5 but not Bmi1 had a slightly diminished H2AK119ub profile, but there remained some signal despite the absence of Bmi1 (**Figure A3**). Therefore, while Bmi1 co-occupancy contributed a portion of the PCGF5-associated H2AK119ub signal, it could not account for all, or even the majority of H2AK119ub at PCGF5-bound loci.

It had been previously observed that many PCGF5 sites were not co-enriched by RING1B while 55% of Bmi1 SERs shared occupancy with RING1B. We, therefore, wondered whether the lower average H2AK119ub levels at PCGF5-bound sites could be explained by the lower RING1B co-occupancy. To determine if this underpinned the difference in H2AK119ub signal between Bmi1 and PCGF5 the H2AK119ub levels were compared over the Bmi1 and PCGF5 sites that were co-enriched by RING1B. In this analysis it was found that the H2AK119ub signal was, in fact, greater over PCGF5-RING1B sites than at Bmi1-RING1B sites (**Figure A4**). Thus, PCGF5 seems to possess equivalent activity to Bmi1 for the ubiquitylation of histones when associated with RING1B. However, PCGF5 binds DNA independent of RING1B much more commonly than Bmi1. Whether this difference is an artifact of the overexpression system or is due to a difference in the ability of PCGF5 and Bmi1 to recruit RING1B is unknown at this time.

While determining the genes that fell near the SERs it was noted that while SERs for Bmi1 and RING1B fell within 300 kb of a promoter TSS over 98% of the time, PCGF5 SERs were beyond 300 kb from the nearest TSS 16.8% of the time, an over 8-fold increase in “orphan” SERs not closely associated with any gene. To shed further light on this trend, the frequency of SERs for Bmi1, PCGF5 and RING1B were plotted against their distance to the nearest TSS.
While the maximum density of PCGF5 SERs was, like for Bmi1 and RING1B, located within 50 kb of a TSS, a greater fraction of SERs for PCGF5 were found distant from SERs (Figure A5.A). Indeed, while 58.7% and 55.4% of SERs for Bmi1 and RING1B, respectively, were centred within 5 kb of a TSS, only 14.3% of PCGF5 SERs were so close (Figure A5.B). Whether this difference represents a divergent mechanism of recruitment than that employed by Bmi1, or whether the distribution observed here is an artifact of overexpression is not known. It was noted, however, that PCGF6 had a similar number of orphan SERs with 13.2% of its SERs beyond 300 kb of the nearest TSS. Unfortunately, no data was available for PCGF3, the closest homologue of PCGF5.
To complement the qPCR experiments performed in MEFs overexpression Bmi1 or PCGF5 I plotted the enrichment tracks and SERs for Bmi1 and PCGF5 over the loci HoxA, HoxB, HoxC and Ink4a loci (Figure A6). Note, however, the tissue and species difference between the mouse fibroblasts from which the qPCR was performed and the embryonic kidney cells (HEK 298T) which were used as source material for the qPCR.

**Figure A4 PCGF5 sites co-occupied by RING1B have high H2AK119ub** Comparison of composite H2AK119ub signal normalized to input control over RING1B SERs that overlap by at least 1 bp either PCGF5 SERs (blue line) or Bmi1 SERs (red line).
Figure A5 PCGF5 SERs locations are less biased towards TSSs than Bmi1 and RING1B

SERs A) Plot showing distance from the nearest TSS of SERs for each Bmi1, PCGF5 and RING1B. Distances were binned into 50kb intervals and counts for each interval were normalized as a percent of the most common interval. B) Percentage of SERs for Bmi1, PCGF5 and RING1B located within 5kb of the nearest TSS.
Figure A6 Distribution of Bmi1 and PCGF5 SERs on selected Bmi1 target loci in human. ChIP-seq enrichment profiles for Bmi1, PCGF5 and H2AK119ub are shown for the HoxA/B/C clusters as well as the region around the CDKN2A/B locus. SERs for Bmi1 (B) and PCGF5 (P) are indicated. Y-axes are scaled to top of highest peak for each track and represent normalized tag fold enrichment.