Transcriptional Regulation of Clr-b (*Clec2d*) in Health and Disease

by

Christina Kirkham

A thesis submitted in conformity with the requirements for the degree of Doctor of Philosophy
Department of Immunology
University of Toronto

© Copyright by Christina Kirkham 2017
Transcriptional Regulation of Clr-b (Clec2d) in Health and Disease

Christina Kirkham
Doctor of Philosophy

Department of Immunology
University of Toronto
© 2017

Abstract

Innate immunity provides the first line of defense against infectious disease and cancer. Natural killer (NK) cells are innate lymphocytes that recognize and eliminate pathological cells. Target cell recognition is facilitated by germline-encoded NK cell receptors, including the NKR-P1B inhibitory receptor. The NKR-P1B-ligand, Clr-b, is broadly expressed by normal healthy cells, yet rapidly downregulated during pathological alterations, resulting in NK cell disinhibition via NKR-P1B. This thesis investigates transcriptional regulation of Clr-b during health and disease.

A whole-genome lentiviral shRNA screen was used to identify novel transcription factors involved in the normal expression of the Clr-b gene (Clec2d). MeCP2 was identified and validated as a transcription factor that represses Clr-b expression. TBP and CBP were also identified as ubiquitous transcription factors required for normal basal Clr-b surface levels. Accordingly, a core Clec2d promoter with an inverted TATA motif was identified ~20bp upstream of the transcriptional start site (TSS) as sufficient to drive basal Clec2d expression. During NIH3T3 cell infection using MCMV-GFP, infected cells (GFP⁺) lost Clr-b expression, while uninfected ‘bystander’ cells (GFP⁻) upregulated Clr-b levels. This was due to reciprocal regulation of the Clec2d promoter and nascent transcripts in virus-infected versus bystander cells exposed to paracrine type-I interferon (IFN). Infected cells downregulated Clec2d nascent transcripts, mainly due to absence of RNAPII recruitment to the Clec2d promoter. Moreover, a
cell-autonomous role for the MCMV ie3 gene product was also discovered, whereby ectopic
expression of IE3 directly repressed Clec2d activity. In uninfected cells, paracrine or exogenous
type-I IFN upregulated Clec2d promoter activity via canonical IFNAR-mediated signaling. This
was dependent upon recruitment of STAT2 to a consensus cluster of IRF motifs (IRF3/7/9,
termed the IRFC) located ~200bp upstream of the Clec2d TSS. Clec2d induction by type-I IFN
was IRF9- and STAT1-dependent, abrogated in IRF9−/− or STAT1−/− MEF or upon mutation of
the IRFC. Thus, in addition to Clr-b being an MHC-I-independent ‘missing-self’ marker on
virus-infected cells, Clec2d is an interferon-stimulated gene on uninfected bystander cells,
collectively facilitating greater dynamic range of NKR-P1B-dependent self-nonself
discrimination during virus infection.
This thesis is dedicated to my family for their unconditional love and support. Especially to Mom, Dad, and Steven who have been huge pillars of encouragement and love from the beginning and who raised me to never quit, and who made me a strong person. Going home to spend time with you always makes me so happy – I always laugh the most when I’m with you. To my grandparents Kathleen and Paul Cherrie, and Lola Kirkham for always believing in me and thinking my work is meaningful. Finally, to my own little family – Mat and Oliver – thank you for being the best parts of my day and helping me to keep going, especially when I didn’t think I could. I live to spend time with you guys and I consider myself so lucky to get to spend the rest of my life with the both you.
Acknowledgments

Jim, thank you for welcoming me to your lab and for providing an environment that challenged me to become self-sufficient. I gained other valuable skills from my time in your lab (managing a lab and technician-type work) that has given me an immense appreciation for those roles and how much time it takes. I know supervising me has not always been easy but I am grateful that you never gave up on me. I hope I have made you proud.

Thank you to the members of the JRC Lab, past and present. DSJA, thank you for always encouraging and supporting me. I will never be able to express how much our talks meant to me, especially during harder times. AM, for reminding me not to take things too seriously all the time. JHF, for being my first friend when I came to Toronto, training me, and always being there. PC, for teaching me how to play the game, even though I suck at playing it. JM, for the laughs and being someone I could rely on. OAA, for listening, being a goof, and for sound scientific discussion. MT, for providing the lab with more estrogen, the inapporopriate discussion, and having my back. Finally, to the undergraduates that have passed through our lab, especially KLC – thanks for being the best summer/project student I could have hoped for.

To my committee members, Michele, Rod, and JC, thank you for your time, expertise, and support. I have learned a lot from all three of you and I am a better scientist for it.

Thank you to my collaborators, Dr. Rod Bremner and Tao Yu, for assistance with the ChIP experiments. I learned a lot from working with you both on all the considerations and troubleshooting that go into one of these experiments, and I don’t think I would have gotten it up and running on my own. I thank Drs. Jason Moffat and Troy Ketela for their guidance and suggestions regarding the shRNA work. Thank you Dr. A. Makrigiannis and Dr. S. Vidal for providing us with the MCMV viruses, Dr. E. Fish and Beata Majchrzak for the IFN-α4 supernatant and Drs. J.C Züñiga-Pflücker, A. Martin, A. Nagy, and T. Schrader for plasmids. Thank you to Gisele Knowles, Courtney McIntosh, Dr. Geneve Awong, and Vincent Cheng for assistance with cell sorting and for their friendship.

To IGSA, thanks for planning all the wonderful departmental and interdepartmental events. They were a great way to meet people and get out of the lab to have fun for very little
cost. Thank you to the Department of Immunology faculty and staff for creating a nice atmosphere for learning by providing us with seminars to attend and give.

I am grateful to the friends I have made during my PhD. AC, thanks for teaching me Westerns and for keeping me company in the lab. I have enjoyed our coffee and crossword time more than you know. Thanks for being there. MM, thanks for being my second friend in Toronto. I miss not having you in the building anymore. AA and CS for the wonderful coffee breaks and moral support. Thanks KK for your scientific advice, computer expertise, and company at the Blue Jays games! I miss not seeing you at SRI anymore. DC, it has been a pleasure to get to know you and I’m glad I still see you outside of academic life. Thank you to JH, TL, and CG for not being scientists. I appreciate having friends that I don’t have to talk about science or grad school with. JH, I am amazed that you still find ways fit me into your insane schedule. I love visiting you and your family; it always reminds me of what’s truly important in life. It was hard to leave you in Ottawa and move to Toronto. I miss our weekly TV night dates. But, I’m so relieved that we’re still friends despite the distance. TL and CG, we have not been friends that long but I already can’t imagine life without you guys. I’m looking forward to more adventures!

Last, but definitely not least, thank you to my family. Thank you for supporting my decision to move away and go back to school. It could not have been easy for you – it’s still difficult for me. Thank you for always being just a text message or phone call away. Mat, thank you for coming with me to Toronto. I know that it wasn’t your dream, that we’ve been here longer than either of us intended, and that it has made our journey into adulthood rather delayed, so thank you for letting me pursue my dreams and for being there when I fall flat on my face and when I soar high. Thanks for taking care of me and picking up my slack when I’ve been too tired, sick, depressed, or anxious to do anything for myself. I’ll pay you back when we’re old 😊. Oliver, you will never know that I wrote anything about you in my thesis, but that doesn’t mean that I should not include you here, as you have been an important part of this journey too. You are so playful and silly that you can make me laugh even when I’m crying. Your purrs and cuddles after a bad day in the lab remind me of what’s important and help me snap out of my despair. Our apartment is a quieter and lonelier place when you aren’t here.

Thank you all. So much.
Table of Contents

Dedications .................................................................................................................. iv

Acknowledgments ...................................................................................................... v

Table of Contents ....................................................................................................... vii

List of Tables ............................................................................................................... x

List of Figures ............................................................................................................. xi

List of Abbreviations .................................................................................................. xiii

List of Publications .................................................................................................... xvii

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

1 Chapter 1: Introduction ........................................................................................... 2

1.1 Innate Immunity ..................................................................................................... 2

1.2 Natural Killer Cells ............................................................................................... 2

1.2.1 NK Cell Development ........................................................................................ 3

1.2.2 NK Cell Tolerance and Education .................................................................... 6

1.2.3 NK Cell Memory ................................................................................................. 8

1.2.4 Self-Nonself Discrimination by NK Cells .......................................................... 9

1.2.5 MHC-I-independent recognition: The NKR-P1 and Clr Families ....................... 14

1.3 Cytomegaloviruses ............................................................................................... 37

1.3.1 Overview ............................................................................................................ 37

1.3.2 Virion Structure and Replication ...................................................................... 38

1.3.3 Host Resistance and the Role of NK Cells ......................................................... 40

1.3.4 Interferon Production ...................................................................................... 42

1.3.5 IFN Signaling .................................................................................................... 47

1.4 Transcription and Regulation of Gene Expression ................................................ 50

1.4.1 Overview of Transcription .............................................................................. 50

1.4.2 RNA Polymerase II Core Promoter Elements ............................................... 55

1.4.3 Transcriptional Regulation of Gene Expression .............................................. 57

1.5 Rationale and Approach ....................................................................................... 58

Chapter 2 ................................................................................................................... 60
4.2.1 Identification of 2 transcriptional start sites and an alternative exon in the Clec2d gene...125
4.2.2 Promoter activity of the Clec2d 5’ flanking region ..........................................................128
4.2.3 MCMV infection reciprocally modulates Clr-b levels on infected and bystander cells.........132
4.2.4 Bystander Clr-b induction during MCMV infection is type-I IFN-dependent ......................132
4.2.5 Reciprocal modulation of Clec2d promoter by MCMV infection and type-I IFN ...............135
4.2.6 The Clec2d promoter is regulated by a functional cluster of IRF binding sites (IRFC) ........137
4.2.7 Requirement for IFNAR1, IRF9, STAT1 and STAT2 in Clec2d induction by type-I IFN ......143
4.2.8 IFN-dependent Clr-b induction augments NKR-P1B-ligand function .............................153
4.2.9 The MCMV ie3 gene product cell-autonomously represses Clec2d expression ...............158
4.3 Discussion ..................................................................................................................................160

Chapter 5 ........................................................................................................................................164

5 Chapter 5: Discussion and Future Directions ...........................................................................165
5.1 The major Clec2d TSS and alternative transcripts .................................................................165
5.2 The Clec2d promoter and transcriptional regulation .............................................................166
5.3 Importance of Clec2d transcription in regulating Clr-b expression .....................................170
5.4 Future Avenues of Research .................................................................................................172
5.5 Conclusions .............................................................................................................................174

References .........................................................................................................................................176
List of Tables

Table 2.1 Primers utilized throughout the thesis for cloning and qPCR analyses

Table 2.2 Version 4.0 barcoded primer sequences for multiplexing used to generate Illumina sequencing-compatible DNA and the primers used for the sequencing reactions

Table 3.1 Analysis of Clec2d-targeting shRNAs

Table 3.2 shRNA selection criteria for the secondary iterative screen

Table 3.3 List of shRNA selected for further validation

Table 3.4 Analysis of the effect of control shRNA on Clr-b expression based upon fold change from the second screen
List of Figures

Figure 1.1  NK cell target recognition
Figure 1.2  Repertoire of NK cell receptors
Figure 1.3  The NK gene complex (NKC)
Figure 1.4  NKR-P1:Clr gene composition and organization in mice, rats, and humans
Figure 1.5  Structure of cytomegaloviruses
Figure 1.6  Type-I IFN production and signaling pathways
Figure 1.7  Transcription initiation and core promoter elements in metazoans
Figure 3.1  Determination of the multiplicity of infection (MOI) of the 80K pool for the whole genome knockdown in NIH3T3 cells
Figure 3.2  Schematic representation of the workflow for the NIH3T3 cell lentiviral screen
Figure 3.3  Analysis of NIH3T3 cells transduced with the 80K pool library
Figure 3.4  Determination of the multiplicity of infection (MOI) of the 1K pool for the secondary iterative knockdown screen in NIH3T3 cells
Figure 3.5  Analysis of NIH3T3 cells infected with the 1K pool library
Figure 3.6  Correlations between the 80K pool screen and the iterative 1K pool screen
Figure 3.7  Schematic representation of the workflow of the validation of single shRNA
Figure 3.8  Analysis of control shRNA using lentiviral shRNA viruses
Figure 3.9  Analysis of the effect of individual shRNA on Clr-b expression by flow cytometry
Figure 3.10  Knockdown of MeCP2 increases Clr-b expression
Figure 3.11  Knockdown of IRF7 does not change Clr-b expression
Figure 3.12  Knockdown of TBP or Crebbp decreases Clr-b expression
Figure 3.13  Knockdown of Med1, Atf7, or Satb2 decreases Clr-b expression at the protein level
Figure 4.1  Mapping of the mouse Clec2d transcriptional initiation site by 5’ rapid amplification of cDNA ends (5’-RACE)

Figure 4.2  Functional analysis of the Clec2d 5’ upstream region

Figure 4.3  Reciprocal modulation of Clr-b expression on MCMV-infected and IFN-stimulated bystander cells

Figure 4.4  MCMV infection and type-I IFN oppositely regulate Clec2d nascent transcript levels

Figure 4.5  MCMV infection and type-I IFN reciprocally regulate Clec2d promoter activity

Figure 4.6  Herpes simplex virus (HSV)-1 infected MEF cells exhibit reduced RNA polymerase II (RNAPII) occupancy at the Clec2d promoter

Figure 4.7  An overlapping IRF3/7/9 consensus cluster (IRFC) is required for type-I IFN-dependent promoter activity

Figure 4.8  Type-I IFN-dependent induction of Clr-b is IRF9-dependent

Figure 4.9  Expression of Clr-b on MEF cells during viral infection

Figure 4.10  Type-I IFN-mediated Clr-b induction is dependent on STAT1

Figure 4.11  STAT2 is required for Clr-b induction following type-I IFN treatment

Figure 4.12  Time course of Clec2d nascent transcript and protein levels during IFN treatment

Figure 4.13  BWZ.CD3ζ/NKR-P1B (BWZ.P1B) reporter cell analysis of Clr-b function

Figure 4.14  BWZ.P1B reporter cell analysis of Clr-b ligand function on type-I IFN-treated stimulator cells

Figure 4.15  Immediate early gene-3 (ie3) cell-autonomously promotes Clr-b downregulation by repressing Clec2d promoter activity in uninfected cells

Figure 4.16  Model of Clr-b transcriptional regulation during MCMV infection
## List of Abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ab</td>
<td>Antibody</td>
</tr>
<tr>
<td>aCGH</td>
<td>Microarray-based comparative genomic hybridization</td>
</tr>
<tr>
<td>ADCC</td>
<td>Antibody-dependent cellular cytotoxicity</td>
</tr>
<tr>
<td>APC</td>
<td>Antigen presenting cell</td>
</tr>
<tr>
<td>ATCC</td>
<td>American type culture collection</td>
</tr>
<tr>
<td>β2m</td>
<td>Beta-2-microglobulin</td>
</tr>
<tr>
<td>BAC</td>
<td>Bacterial artificial chromosome</td>
</tr>
<tr>
<td>BCR</td>
<td>B cell receptor</td>
</tr>
<tr>
<td>BM</td>
<td>Bone marrow</td>
</tr>
<tr>
<td>bp</td>
<td>base-pair</td>
</tr>
<tr>
<td>BRE</td>
<td>TFIIB recognition element</td>
</tr>
<tr>
<td>BSA</td>
<td>Bovine serum albumin</td>
</tr>
<tr>
<td>CARD</td>
<td>Caspase-activation and recruitment domain</td>
</tr>
<tr>
<td>CBP</td>
<td>CREB-binding protein</td>
</tr>
<tr>
<td>CD</td>
<td>Cluster of differentiation</td>
</tr>
<tr>
<td>cDNA</td>
<td>Complementary DNA</td>
</tr>
<tr>
<td>ChIP</td>
<td>Chromatin immunoprecipitation</td>
</tr>
<tr>
<td>ChIP-Seq</td>
<td>Chromatin immunoprecipitation sequencing</td>
</tr>
<tr>
<td>Clec</td>
<td>C-type lectin-like</td>
</tr>
<tr>
<td>CLP</td>
<td>Common lymphoid progenitor</td>
</tr>
<tr>
<td>Clr</td>
<td>C-type lectin-related</td>
</tr>
<tr>
<td>CMV</td>
<td>Cytomegalovirus</td>
</tr>
<tr>
<td>CREB</td>
<td>Cyclic AMP responsive element-binding protein</td>
</tr>
<tr>
<td>CPRG</td>
<td>Chlorophenol-red-β-D-galactopyranoside</td>
</tr>
<tr>
<td>CPSF</td>
<td>Cleavage and polyadenylation specificity factor</td>
</tr>
<tr>
<td>CRISPR</td>
<td>Clustered regularly interspersed short palindromic repeats</td>
</tr>
<tr>
<td>CTD</td>
<td>Carboxy-terminal domain</td>
</tr>
<tr>
<td>CyTOF</td>
<td>Cytometry by time of flight</td>
</tr>
<tr>
<td>DAP12</td>
<td>DNAX activation protein 12</td>
</tr>
<tr>
<td>DBD</td>
<td>DNA-binding domain</td>
</tr>
<tr>
<td>DC</td>
<td>Dendritic cell</td>
</tr>
<tr>
<td>DMEM-HG</td>
<td>Dulbecco’s modified eagle medium-high glucose</td>
</tr>
<tr>
<td>DMSO</td>
<td>Dimethylsulfoxide</td>
</tr>
<tr>
<td>DN</td>
<td>Double negative</td>
</tr>
<tr>
<td>DNA</td>
<td>Deoxyribonucleic acid</td>
</tr>
<tr>
<td>DPE</td>
<td>Downstream promoter element</td>
</tr>
<tr>
<td>DRB</td>
<td>5,6-Dichloro-1-β-D-ribofuranosylbenzimidazole</td>
</tr>
<tr>
<td>dsDNA</td>
<td>Double-stranded DNA</td>
</tr>
<tr>
<td>DSIF</td>
<td>DRB sensitivity inducing factor</td>
</tr>
<tr>
<td>EBV</td>
<td>Epstein Barr virus</td>
</tr>
<tr>
<td>EGFR</td>
<td>Epidermal growth factor receptor</td>
</tr>
<tr>
<td>EST</td>
<td>Expressed sequence tag</td>
</tr>
<tr>
<td>FACS</td>
<td>Fluorescence-activated cell sorting</td>
</tr>
<tr>
<td>FBS</td>
<td>Fetal bovine serum</td>
</tr>
<tr>
<td>Fc</td>
<td>Fragment crystallization</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Definition</td>
</tr>
<tr>
<td>--------------</td>
<td>------------</td>
</tr>
<tr>
<td>FcR</td>
<td>Receptor for fragment crystallizable region of antibodies</td>
</tr>
<tr>
<td>Flt3</td>
<td>Fms-like tyrosine kinase-3</td>
</tr>
<tr>
<td>GFP</td>
<td>Green fluorescent protein</td>
</tr>
<tr>
<td>GM-CSF</td>
<td>Granulocyte macrophage colony-stimulating factor</td>
</tr>
<tr>
<td>GRO-Seq</td>
<td>Global run-on sequencing</td>
</tr>
<tr>
<td>HCMV</td>
<td>Human cytomegalovirus</td>
</tr>
<tr>
<td>HBSS</td>
<td>Hank’s balanced salt solution</td>
</tr>
<tr>
<td>HEPES</td>
<td>4-(2-Hydroxyethyl)piperazine-1-ethanesulfonic acid</td>
</tr>
<tr>
<td>HIV</td>
<td>Human immunodeficiency virus</td>
</tr>
<tr>
<td>h.p.i.</td>
<td>hours post infection</td>
</tr>
<tr>
<td>HSV-1</td>
<td>Herpes simplex virus-1</td>
</tr>
<tr>
<td>IAD</td>
<td>IRF activation domain</td>
</tr>
<tr>
<td>IE</td>
<td>Immediate early</td>
</tr>
<tr>
<td>IEL</td>
<td>Intraepithelial lymphocyte</td>
</tr>
<tr>
<td>IFN</td>
<td>Interferon</td>
</tr>
<tr>
<td>IFNAR</td>
<td>Interferon-α/β receptor</td>
</tr>
<tr>
<td>IFNGR</td>
<td>Interferon-γ receptor</td>
</tr>
<tr>
<td>IFNLR</td>
<td>Interferon-λ receptor</td>
</tr>
<tr>
<td>Ig</td>
<td>Immunoglobulin</td>
</tr>
<tr>
<td>IKK</td>
<td>IkappaB kinase</td>
</tr>
<tr>
<td>IL</td>
<td>Interleukin</td>
</tr>
<tr>
<td>ILC</td>
<td>Innate Lymphoid Cell</td>
</tr>
<tr>
<td>Inr</td>
<td>Initiator element</td>
</tr>
<tr>
<td>IRAK</td>
<td>IL-1 receptor-associated kinase 1</td>
</tr>
<tr>
<td>IRF</td>
<td>Interferon regulatory factor</td>
</tr>
<tr>
<td>ISGF</td>
<td>Interferon-stimulated gene factor</td>
</tr>
<tr>
<td>ISRE</td>
<td>Interferon-stimulated response element</td>
</tr>
<tr>
<td>JAK1</td>
<td>Janus activated kinase-1</td>
</tr>
<tr>
<td>Kb</td>
<td>Kilobase</td>
</tr>
<tr>
<td>KIR</td>
<td>Killer-cell immunoglobulin-like receptor</td>
</tr>
<tr>
<td>Klr</td>
<td>Killer-cell lectin-like receptor</td>
</tr>
<tr>
<td>LAK</td>
<td>Lymphokine activate killer</td>
</tr>
<tr>
<td>LCR</td>
<td>Locus control region</td>
</tr>
<tr>
<td>Lin−</td>
<td>Lineage negative (Mac-1−CD4−CD8−B220−Ter-119−)</td>
</tr>
<tr>
<td>LLT1</td>
<td>Lectin-like transcript 1</td>
</tr>
<tr>
<td>LMP</td>
<td>Low molecular weight protein</td>
</tr>
<tr>
<td>LPS</td>
<td>Lipopolysaccharide</td>
</tr>
<tr>
<td>LRC</td>
<td>Leukocyte receptor cluster</td>
</tr>
<tr>
<td>LTi</td>
<td>Lymphoid tissue-inducer</td>
</tr>
<tr>
<td>mAb</td>
<td>Monoclonal antibody</td>
</tr>
<tr>
<td>MAVS</td>
<td>Mitochondrial antiviral signaling</td>
</tr>
<tr>
<td>Mda5</td>
<td>Melanoma differentiation-association protein-5</td>
</tr>
<tr>
<td>MEF</td>
<td>Mouse embryonic fibroblast</td>
</tr>
<tr>
<td>MEMα</td>
<td>Minimum essential medium α</td>
</tr>
<tr>
<td>MCMV</td>
<td>Murine cytomegalovirus</td>
</tr>
<tr>
<td>MHC</td>
<td>Major histocompatibility complex</td>
</tr>
<tr>
<td>MICA/B</td>
<td>MHC class I chain-related gene A/B</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Full Name</td>
</tr>
<tr>
<td>--------------</td>
<td>-------------------------------------------------------------</td>
</tr>
<tr>
<td>MIEP</td>
<td>Major immediate early promoter</td>
</tr>
<tr>
<td>MFI</td>
<td>Median fluorescence intensity</td>
</tr>
<tr>
<td>MOI</td>
<td>Multiplicity of infection</td>
</tr>
<tr>
<td>mRNA</td>
<td>Messenger RNA</td>
</tr>
<tr>
<td>MyD88</td>
<td>Myeloid differentiation primary response 88</td>
</tr>
<tr>
<td>NCR</td>
<td>Natural cytotoxicity receptor</td>
</tr>
<tr>
<td>NELF</td>
<td>Negative elongation factor</td>
</tr>
<tr>
<td>NFAT</td>
<td>Nuclear factor of activated T cells</td>
</tr>
<tr>
<td>NF-κB</td>
<td>Nuclear factor-κB</td>
</tr>
<tr>
<td>NK</td>
<td>Natural killer</td>
</tr>
<tr>
<td>NKC</td>
<td>Natural killer gene complex</td>
</tr>
<tr>
<td>NKG2</td>
<td>Natural killer group 2</td>
</tr>
<tr>
<td>NKP</td>
<td>NK precursor</td>
</tr>
<tr>
<td>NKp46</td>
<td>Natural killer cell p46-related protein</td>
</tr>
<tr>
<td>NKR</td>
<td>Natural killer cell receptor</td>
</tr>
<tr>
<td>NKR-P1</td>
<td>Natural killer cell receptor protein-1</td>
</tr>
<tr>
<td>NLR</td>
<td>Nucleotide-binding oligomerization domain (NOD)-like receptor</td>
</tr>
<tr>
<td>OAS</td>
<td>2′-5′-oligoadenylate synthase</td>
</tr>
<tr>
<td>OCIL</td>
<td>Osteoclast inhibitory lectin</td>
</tr>
<tr>
<td>OD</td>
<td>Optical density</td>
</tr>
<tr>
<td>ORF</td>
<td>Open reading frame</td>
</tr>
<tr>
<td>PAMP</td>
<td>Pathogen-associated molecular pattern</td>
</tr>
<tr>
<td>PBS</td>
<td>Phosphate buffered saline</td>
</tr>
<tr>
<td>PCR</td>
<td>Polymerase chain reaction</td>
</tr>
<tr>
<td>pDC</td>
<td>Plasmacytoid dendritic cell</td>
</tr>
<tr>
<td>PFU</td>
<td>Plaque-forming unit</td>
</tr>
<tr>
<td>PGE₂</td>
<td>Prostaglandin E2</td>
</tr>
<tr>
<td>PHA</td>
<td>Phytohaemagglutinin</td>
</tr>
<tr>
<td>PIC</td>
<td>Preinitiation complex</td>
</tr>
<tr>
<td>PKR</td>
<td>RNA-dependent protein kinase</td>
</tr>
<tr>
<td>PMA</td>
<td>Phorbol 12-myristate 13-acetate</td>
</tr>
<tr>
<td>PRR</td>
<td>Pattern recognition receptor</td>
</tr>
<tr>
<td>P-TEFb</td>
<td>Positive transcription elongation factor-b</td>
</tr>
<tr>
<td>qRT-PCR</td>
<td>Quantitative real-time polymerase chain reaction</td>
</tr>
<tr>
<td>RA</td>
<td>Rheumatoid arthritis</td>
</tr>
<tr>
<td>RACE</td>
<td>Rapid amplification of cDNA ends</td>
</tr>
<tr>
<td>RBP</td>
<td>RNA polymerase B</td>
</tr>
<tr>
<td>RCN</td>
<td>Relative cell number</td>
</tr>
<tr>
<td>RCTL</td>
<td>Rat cytomegalovirus C-type lectin-like</td>
</tr>
<tr>
<td>REF</td>
<td>Rat embryonic fibroblast</td>
</tr>
<tr>
<td>RIG-I</td>
<td>Retinoic acid-inducible gene</td>
</tr>
<tr>
<td>RIP</td>
<td>Receptor interacting protein</td>
</tr>
<tr>
<td>RNA</td>
<td>Ribonucleic acid</td>
</tr>
<tr>
<td>RNAPII</td>
<td>RNA polymerase II</td>
</tr>
<tr>
<td>RNase L</td>
<td>Ribonuclease L</td>
</tr>
<tr>
<td>rRNA</td>
<td>Ribosomal RNA</td>
</tr>
<tr>
<td>SA-APC</td>
<td>Streptavidin-allophycocyanin</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Full Form</td>
</tr>
<tr>
<td>--------------</td>
<td>-----------</td>
</tr>
<tr>
<td>SA-PE</td>
<td>Streptavidin-R-phycoerythrin</td>
</tr>
<tr>
<td>Sca-1</td>
<td>Stem cell antigen 1</td>
</tr>
<tr>
<td>SCF</td>
<td>Stem cell factor</td>
</tr>
<tr>
<td>SD</td>
<td>Standard deviation</td>
</tr>
<tr>
<td>SDS-PAGE</td>
<td>Sodium dodecyl sulfate polyacrylamide gel electrophoresis</td>
</tr>
<tr>
<td>SEM</td>
<td>Standard error of the mean</td>
</tr>
<tr>
<td>SH2</td>
<td>Src-homology 2</td>
</tr>
<tr>
<td>SHP</td>
<td>SH2-containing protein tyrosine phosphate</td>
</tr>
<tr>
<td>shRNA</td>
<td>Short hairpin ribonucleic acid</td>
</tr>
<tr>
<td>ssRNA</td>
<td>Single-stranded RNA</td>
</tr>
<tr>
<td>STAT</td>
<td>Signal transducer and activator of transcription</td>
</tr>
<tr>
<td>STING</td>
<td>Stimulator of IFN genes</td>
</tr>
<tr>
<td>TAF</td>
<td>TBP-associated factor</td>
</tr>
<tr>
<td>TAP</td>
<td>Transporter associated with antigen processing</td>
</tr>
<tr>
<td>TBK</td>
<td>Tank-binding kinase</td>
</tr>
<tr>
<td>TBP</td>
<td>TATA-binding protein</td>
</tr>
<tr>
<td>TCR</td>
<td>T cell receptor</td>
</tr>
<tr>
<td>TF</td>
<td>Transcription factor</td>
</tr>
<tr>
<td>TFIIA</td>
<td>Transcription factor for RNA polymerase II-A</td>
</tr>
<tr>
<td>Th1</td>
<td>Type-I helper</td>
</tr>
<tr>
<td>TLR</td>
<td>Toll-like receptor</td>
</tr>
<tr>
<td>TNF</td>
<td>Tumour necrosis factor</td>
</tr>
<tr>
<td>TRAF</td>
<td>TNF receptor-associated factor</td>
</tr>
<tr>
<td>TRAIL</td>
<td>TNF-related apoptosis-inducing ligand</td>
</tr>
<tr>
<td>TRF</td>
<td>TBP-related factor</td>
</tr>
<tr>
<td>TRIF</td>
<td>Toll-IL-1 receptor domain-containing adaptor inducing IFN-β</td>
</tr>
<tr>
<td>tRNA</td>
<td>Transfer RNA</td>
</tr>
<tr>
<td>TSS</td>
<td>Transcriptional start site</td>
</tr>
<tr>
<td>Tyk2</td>
<td>Tyrosine kinase 2</td>
</tr>
<tr>
<td>UTR</td>
<td>Untranslated region</td>
</tr>
<tr>
<td>UV</td>
<td>Ultraviolet</td>
</tr>
<tr>
<td>VCAM</td>
<td>Vascular cell adhesion molecule</td>
</tr>
<tr>
<td>WT</td>
<td>Wild-type</td>
</tr>
</tbody>
</table>
List of Publications


**Acknowledgements**


Chapter 1
Introduction

Christina L. Kirkham$^{1,2}$ and James. R. Carlyle$^{1,2}$

$^1$Sunnybrook Research Institute, Platform of Biological Sciences, Toronto, ON

$^2$University of Toronto, Department of Immunology, Toronto, ON

The sections on the NKR-P1 and C1r families, Figure 1.3, and Figure 1.4 were adapted from Frontiers in Immunology, 02 June 2014 http://dx.doi.org/10.3389/fimmu.2014.00214, and are reproduced and updated here with permission.
1 Chapter 1: Introduction

1.1 Innate Immunity

The immune system is composed of cells and effector molecules that serve to protect the body from invading pathogens. An immune response is first triggered by cells of the innate branch of the immune system, which include monocytes, macrophages, dendritic cells (DC), neutrophils, eosinophils, basophils, and natural killer (NK) cells [1]. These cells express a variety of germline-encoded pattern recognition receptors (PRR) that detect moieties broadly expressed by pathogens or microbes, called pathogen or microbe-associated molecular patterns (PAMP; MAMP) [1]. PRR stimulation results in several effector functions depending on the cell type, including secretion of cytokines, phagocytosis, and cytotoxicity [1]. The invading pathogen can be effectively controlled by innate immunity, whereas activation of the adaptive immune system by these mechanisms can facilitate sterilizing immunity or promote stable chronic infection.

1.2 Natural Killer Cells

Natural killer cells were first documented in studies that demonstrated rapid “spontaneous” cytolytic activity of splenic lymphocytes from healthy animals against tumour cells. The cells responsible for this so-called background killing were originally described as “null” cells, owing to their lack of surface immunoglobulin (Ig) or T cell-associated markers such as CD3 [2, 3]. In the 1970’s, these cells were renamed natural killer (NK) cells due to their ability to spontaneously kill both syngeneic and allogeneic Moloney leukemia cells in the mouse [2].

It is now known that NK cells are capable of recognizing other non-tumour target cells, including antibody-coated cells (via antibody-dependent cellular cytotoxicity; ADCC), transplanted cells (mainly allogeneic hematopoietic cells), infected cells (including viruses and cytosolic pathogens), transformed and cancerous cells, and “stressed” cells. The cytotoxic function of NK cells is mediated via synaptic exocytosis of pre-formed granules containing perforin (a membrane pore-forming protein), granzymes (serine proteases), and granulysin (antibacterial toxin), or through the activation of intracellular caspases by cell surface receptors.
of the TNF receptor family (e.g., Fas:FasL; TRAIL-R:TRAIL) [4]. In addition to cytotoxicity, activated NK cells are also early and potent secretors of cytokines, most notably interferon-gamma (IFN-γ), tumour necrosis factor-alpha (TNF-α), and granulocyte-macrophage colony stimulating factor (GM-CSF) [5].

With the recent discovery of innate lymphoid cells (ILC), and re-classification of NK cells as a subset of ILC, knowledge of NK cell development and function has required some reinterpretation, depending on how NK cells were defined in individual publications. NK cells belong to a subset of group-1 ILC (or ILC1), based upon their ability to secrete type-I helper (Th1) cytokines such as IFN-γ, and by their expression of NKP46 (NCR1), NK1.1 (NKR-P1C<sup>B6</sup>), and transcription factors such as T-bet (<i>TBX21</i>)[6]. NK cells are unique among ILC1 cells, in that they display cytotoxic function and express the transcription factor, Eomes (<i>Eomesodermin</i>) [7, 8], separating them from helper ILC1 cells. Thus, previous work that examined NKP46<sup>+</sup>NK1.1<sup>+</sup> “NK cells” may not have strictly studied NK cells per se, but rather included some ILC1. There are also group-2 ILC (ILC2) and group-3 ILC (NCR± ILC3; LTi cells), but discussion of their development and function are beyond the scope of this thesis; suffice it to say that helper ILC subsets are largely dependent upon IL-7, while cytotoxic conventional NK cells are dependent upon IL-15.

1.2.1 NK Cell Development

Mouse NK cells are known to develop primarily in two locations, the bone marrow (BM) and the thymus; although the majority of conventional NK cell development occurs in the BM. NK cells have also been reported to develop in the liver, uterus, mucosal tissues, and lymph nodes (in humans), but these NK cell subsets are functionally and/or phenotypically different from conventional BM-derived splenic NK cells. These differences highlight the fact that NK cells are an extremely phenotypically and functionally diverse population, capable of carrying out different functions, depending on their organ of development and distribution in the periphery.
In the BM, NK cells (along with all ILC subsets and B cells) develop from common lymphoid progenitor (CLP) cells, which display a Lin−, cKit^low, CD127^+, Sca1^low, Flt3^+, and α4β7− phenotype. CLP express Nfil3 (E4BP4), a transcription factor (TF) important in NK cell development, as mice lacking Nfil3 possess a specific deficiency in developing NK cells [9, 10]. Notch signaling through the Jagged1 and Jagged2 ligands may preferentially drive NK cell development from the CLP [11], although other groups have shown that Notch signaling does not play a significant role in NK cell development [12].

From CLP, NK lineage progenitor cells become lineage-specified NK cell precursor (NKP) cells, as defined by expression of CD122 (IL-2Rβ), CD27, CD244 (2B4), and CD127 (IL-7Rα) [13]. No single factor has been identified that acts as a switch to generate NKP from hematopoietic precursors, but several factors are important in maintaining NKP once they are generated, including stem cell factor (SCF, ligand for c-kit), fms-like tyrosine kinase-3 ligand (Flt3L), IL-7, and IL-15 [14]. CD122 is essential for NK cells as this subunit, together with the IL-2R common-γ chain (CD132), forms the functional IL-15 receptor, which is required for homeostasis of mature peripheral NK cells and drives the expression of Nfil3 [15, 16]. Not all NKP develop into mature NK cells, as they maintain some ability to generate T cell subsets in vitro and in vivo [17]. As NKP gain expression of NK1.1 (Nkrp1C^B6), T-bet, and high levels of CD94/NKG2 heterodimers, they undergo NK lineage commitment and transition to the immature NK (iNK) cell stage [14]. T-bet is important for the expression of S1P5, a receptor used by mature NK cells to egress from the lymph node and bone marrow, as well as for stabilization of an iNK cell state [18]. As they mature, iNK cells lose expression of IL-7Rα, and begin to express NKp46, DX5 (CD49b), Eomes, and inhibitory Ly49 receptors, which are required for the full acquisition of effector function, variably called NK education, licensing, or disarming [19]. In general, for the purpose of educating NK cells, inhibitory isoforms of paired NK cell receptors are expressed before activating isoforms. However, subsets of NK cells that lack inhibitory receptors for self MHC-I still develop, but are rendered hyporesponsive to prevent autoimmunity. Eomes expression is required for the transition from the immature to mature NK cell state, for the acquisition of Ly49 receptors, and for the expression of CD122 [20]. Without T-bet and Eomes expression, no NK cells develop in any organs. As cells transition from an
immature to a mature NK cell, they upregulate CD11b and CD43 expression. Id2 activity is important for this transition, as loss of Id2 expression halts NK cell development [21]. As NK cells gain expression of CD11b, DX5, CD43, and Ly49 receptors, they become functionally competent (i.e., can produce IFN-γ and exhibit cytotoxic activity), and exit the BM [22]. Peripheral NK cells are classified by their expression of CD27 – whereby the more mature CD27$^\text{low}$ subset is more cytotoxic and produces more cytokines than the less mature CD27$^\text{high}$ subset [23] – as well as the sequential upregulation of CD11b – whereby the least mature subset is CD27$^+$CD11b$^-$, progressing through the CD27$^+$CD11b$^-$ and CD27$^+$CD11b$^+$ subsets, followed by the most mature CD27$^+CD11b^+$ subset [24].

In the thymus, the thymic NK cell (tNK) population as a whole is distinct from the BM-derived splenic conventional NK cell (cNK) cohort described above. While conventional NK cells are Ly49$^\text{high}$ and CD11b$^\text{high}$, thymic NK cells are Ly49$^\text{low}$, CD11b$^\text{low}$, and CD16$^-$, similar to iNK cells [25]. However, unlike splenic iNK cells, thymic NK cells are also GATA-3$^+$CD127$^+$CD69$^\text{high}$B220$^\text{low}$ [25, 26]. This has led some groups to suggest that thymic NK cells develop as a result of defaulted NK lineage commitment away from a T cell-specified fate, as some thymic NK cells possess germ-line rearranged T cell receptor-γ (TCRγ) genes, in turn suggesting they arose from a DN2 thymocyte subset [27]. Other groups suggest that thymic NK cells develop independently of T cell precursors, since most thymic NK cells do not express Rag2 and do not rearrange their TCR genes [12]. It is also possible that, given their expression of functional IL-7Rα (CD127/CD132) receptors, tNK cells may represent a unique ILC lineage subset, retaining the ability to give rise to both ILC1-like and ILC3-like cells, at least during fetal thymopoiesis [28]. Regardless of whether tNK cells develop from a T cell-specified precursor or not, they represent a phenotypically and functionally distinct subset compared to splenic cNK cells. In terms of function, CD127$^+$ tNK cells travel to the lymph nodes (and perhaps mucosal tissues) and are less cytotoxic than splenic cNK cells, but they also produce substantially more IFN-γ, TNF-α, and GM-CSF in response to cytokines [26]. Again, this reflects more of an ILC-like functional phenotype than cNK cells.
1.2.2 NK Cell Tolerance and Education

Regardless of their origin, developing NK cells acquire self-tolerance through a process called education. Education of NK cells is an incredibly important event, as uneducated NK cells have the potential to be self-reactive, yet instead are rendered hyporesponsive, a state similar to T and B cell anergy. There are two classical models for NK cell tolerance: The licensing (arming) model and the disarming model. The disarming model states that NK cells lacking self-MHC-I-specific inhibitory receptors become hyporesponsive, either due to chronic activation via generalized stimulatory receptors, or due to a block in the final maturation steps, locking the NK cell in an immature state with drastically reduced effector functions [29]. The licensing or arming model states that MHC-I-specific inhibitory receptors are required for functional maturation of NK cells, delivering a positive maturation signal during development; thus, lacking these receptors results in immature hyporesponsive NK cells [29]. While both models sound deceptively similar, there are some key differences. For instance, in the arming model, only immature NK cells undergo licensing, while the disarming model applies to both immature and mature NK cells. Hyporesponsive yet tolerant NK cells have indeed been found in mice deficient in MHC-I-expression (e.g., $\beta_2m^{-/-}$ mice) and in normal B6 mice [30, 31]. These hyporesponsive NK cells appear to be mature based upon cell surface markers, suggesting that a block in development, a key aspect of the licensing model, is not supported [31]. It is possible that NK cells without MHC-I-specific inhibitory receptors can also retain tolerance to self by downregulating stimulatory receptors [29]. This was examined in some detail, and although many hyporesponsive NK cells express normal levels of the activating receptors tested and intracellular perforin, it is still possible that other stimulatory receptors that have not been examined in this context are upregulated [31]. It is also unclear under the licensing model how a qualitatively distinct stimulatory signal can be elicited from inhibitory Ly49 receptors during NK cell development and maturation. Nonetheless, NK cell responsiveness is also known to be proportional to the number of self-MHC-I inhibitory receptors they express, whereby NK cells with higher numbers of Ly49 receptors are more responsive to stimuli than those NK cells with fewer Ly49 receptors [32, 33].
Another more recent model for NK cell education, called the confining model states that confined NK receptor compartmentalization prompts efficient NK:target cell contacts and sequential NK cell activation [34]. In support of this model, NKG2D and NKp46 form microclusters at the immune synapse, which results in a lowered threshold for activation in educated NK cells, but not in MHC-I-deficient NK cells [35].

A final model for NK cell tolerance, called the tuning or rheostat model, suggests that the strength of the MHC-I:Ly49 interaction translates into quantitative responsiveness in NK cells, as opposed to a binary on-off switch [36]. Predictions from this model are that NK cells with several self-specific receptors would be more responsive in an environment where those self-ligands are expressed, and that NK cells expressing a particular MHC-I-specific receptor would be more responsive in mice with higher expression of MHC-I ligands compared to low expression [37]. Indeed, transfer of hyporesponsive splenic NK cells from MHC-I deficient donor into WT recipient results in increased responsiveness to MHC-I [38]. Here, NK cell maturation is sequentially induced upon exposure to self-ligands, and can be quantitatively fine-tuned, akin to an adjustable or variable resistor switch, such as a thermostat.

Since not all NK cells express an inhibitory receptor for MHC-I [31], MHC-I-independent education must also be important. For example, our laboratory and others have demonstrated that the inhibitory interaction between Clr-b (Clec2d) on healthy cells and NKR-P1B (Klrb1b) on NK cells (see below) plays a role in NK cell education. NKR-P1B⁺ NK cells from WT mice are unable to kill cells expressing Clr-b, but they readily kill Clr-b-deficient targets [39, 40]. Lack of Clr-b ligand expression during development (e.g., in Clr-b⁻/⁻ mice) results in a hyporesponsive NK phenotype (i.e., less IFN-γ secretion and reduced ability to reject Clr-b-deficient targets), while lack of NKR-P1B receptor expression (e.g., in Nkrp1b⁻/⁻ mice) results in an NK cell inability to discriminate between Clr-b-sufficient and Clr-b-deficient targets, as well as an augmented reliance upon MHC-I-specific Ly49-dependent education [41, 42].

Notably, since NK cells express both MHC-I and Clr-b, it is possible for NK cell education to occur in trans, in cis, or both. Some groups have tried to address this question using
animal models of Ly49A variants that can only be recognized in trans or in cis, depending on the variant; here, the cis interaction of Ly49 receptors with MHC-I on the NK cells reduces the threshold for NK cell activation, and may provide a qualitatively distinct signal during NK cell development that may promote NK cell maturation [34, 43-45].

The precise mechanism(s) regarding how NK cells become educated is still unclear. The original dogma was that NK cells are educated during development in the BM (discussed above in section 1.2.1). However, when uneducated yet mature NK cells from MHC-I-deficient mice are transferred into WT mice, their function is restored, suggesting that mature NK cells can also be educated [31, 46]. It seems that although education is important for the acquisition of effector functions, uneducated or unlicensed NK cells do mount an effective response against target cells under certain conditions, such as when target cells express high levels of MHC-I to evade detection via missing-self recognition with Ly49 receptors [47] or during viral infection [48].

1.2.3 NK Cell Memory

One of the distinguishing features between innate and adaptive immune cells is the capacity for memory, or the ability of cells to specifically and more quickly respond to a particular infectious agent that they have encountered previously. Traditionally, innate immune cells were not thought to exhibit memory, while adaptive immune cells displayed the classical hallmarks of memory or recall responses. This idea was challenged a number of years ago with the concept that NK cells possess memory capacity, even though they are only capable of expressing a limited number of germline-encoded receptors that detect self and non-self antigens.

The notion that NK cells possess the capacity for memory began in the early 2000’s with studies showing that mouse NK cells possess an activating Ly49H receptor specific for the murine cytomegalovirus (MCMV) m157 gene product, an MHC-I-like protein [49, 50]. Although neither initial study examined memory, they provided the first evidence of a prerequisite for memory – antigen specificity. The concept of NK memory was first formally demonstrated in the hapten-induced contact hypersensitivity response, where agents that chemically modify proteins are painted onto an animal’s dorsal skin followed several days later
by a rechallenge at a previously unexposed site. Sensitization-dependent infiltration of inflammatory cells in challenged \( \text{Rag}^2/^{--} \) mice (lacking B and T cells) was identical in magnitude to that in WT mice, suggesting a T- and B cell-independent hapten-specific memory, which was lost upon NK-cell depletion using the anti-asialo-GM1 or NK1.1 mAbs [51]. In the Ly49H:m157 MCMV model, memory NK cells can be detected even at 70 days post-infection and remain functional (i.e., produce IFN-\( \gamma \) and degranulate \textit{ex vivo}) [52]. The memory NK cells expressed elevated levels of Ly49H (\( Kira8 \), KLRG1, CD43, and Ly6C, and less CD27, compared to naïve Ly49H\({ }^+\) NK cells. When rechallenged with MCMV, the memory Ly49H\({ }^+\) NK cells had an amplified response that was more protective than that of naïve NK cells. The clonal expansion and generation of memory NK cells is dependent on IL-12 and STAT4 [53-56]. A similar memory response by NK cells has been seen in humans with CMV infection and Hantavirus infection [57, 58].

Non-pathogenic cytokine-induced or cytokine-primed NK cell memory also exists [53, 54, 56]. For example, NK cells pretreated with IL-12, IL-18, and IL-15 that were adoptively transferred into Rag-deficient hosts persisted for weeks post transfer, produced more IFN-\( \gamma \) following cytokine restimulation than untreated NK cells, and were more protective in a transplantable tumour model [59]. The transcription factor Zbtb32 was required for NK cell memory. Zbtb32 is induced downstream of IL-12, IL-18, and type-I IFN where it controls the proliferation of activated NK cells by antagonizing Blimp-1 (\( Prdm1 \)) [60].

1.2.4 Self-Nonself Discrimination by NK Cells

NK cells express both stimulatory and inhibitory receptors on their cell surface. Whether an NK cell becomes activated to kill a target cell depends on the balanced or integrated signals received from both receptor types. For example, healthy cells broadly express a number of ‘self’ inhibitory ligands (including MHC-I molecules), which are frequently lost during pathological alterations such transformation and infection; this mode of target cell sensing is termed “missing-self” recognition [5, 61]. On the other hand, most stimulatory ligands are expressed minimally on healthy cells, but are strongly induced during cellular pathologies; this is termed “induced-self”
recognition [29]. Typically, both missing-self and induced-self recognition events operate simultaneously to shift the balance from NK cell tolerance to activation [29, 62](Figure 1.1).

**Figure 1.1. NK cell target recognition.** NK recognition of target cells is mediated through integrated signals from both inhibitory and stimulatory receptors. A healthy cell is protected from NK cytotoxicity as it expresses inhibitory ligands such as MHC Class I. Following transformation and infection, a cell can lose the expression of the inhibitory ligand thus creating a net activation signal. This is termed missing-self recognition. A target cell can also upregulate the expression of stimulatory ligands like NKG2DL, which also creates a net activation signal and cytotoxicity. This is termed induced cell recognition.
Many NK cell receptors are encoded by genes linked to the NK gene complex (NKC), located on chromosome 6 in mice, chromosome 4 in rats, and chromosome 12 in humans [63, 64]. The NKC is also conserved among several other species, including dogs [63, 65], cattle (where it is split between two chromosomes) [63], and chickens (where it is linked to the MHC region) [66]. However, a number of NK cell receptor gene families are linked to the leukocyte receptor cluster (LRC) located on mouse chromosome 7, rat chromosome 1, and human chromosome 19 [67-70]. There are also numerous other NK cell receptors located in other regions of the genome, including the SAP/SLAM family of receptors found on mouse and human chromosome 1 [71, 72], the natural cytotoxicity receptors (NCRs) [73, 74], and others. Within these regions, most NK cell receptors can be broadly classified into two structurally divergent categories: immunoglobulin-like receptors (e.g., KIR, NCR; mostly encoded within the LRC) and C-type lectin-like receptors (e.g., NKR-P1, Ly49, mostly encoded within the NKC) (Fig 1.2). Receptors for classical MHC-Ia ligands include the Ly49 family in rodents and the KIR family in humans. NK cells also possess several receptors for non-classical MHC-I and MHC-I-like molecules as well, which include the CD94/NKG2 family in rodents and humans (non-classical MHC-Ib); NKG2D (MHC-I-related), and the NKR-P1 family (non-MHC). NK cells also express other important receptors, including 2B4 (ligand, CD48), and NKp46 (ligand, influenza haemagglutinin).
Figure 1.2. Repertoire of NK cell receptors. NK cells recognize targets through the expression of a diverse range of cell surface receptors. The receptors are stimulatory or inhibitory in nature and the balance of signals generated from both types of receptors determines NK cell activation. NK cell receptors are broadly classified into two structural groups: the C-type lectin family and the immunoglobulin super family. There are receptors specific for classical MHC-I (Ly49 in mouse; KIR in humans), non-classical MHC-I (CD94/NKG2C,E in mouse and human), and distantly related MHC-I (NKG2D in mouse and human). The NKR1 family members are C-type lectin-like homodimeric receptors that recognize homodimeric C-type lectin-related ligands (Clr). The NKR-P1B:Clr-b inhibitory interaction is the best characterized NKR1:Clr interaction.
How do NK cells integrate the signals from so many receptors to elicit appropriate responses? One mechanism is the requirement for synergy or co-stimulation between activating receptor classes, which may act as a safe guard against unrestrained NK cell activation. For example, stimulation of resting NK cells with an activating NKp46 mAb may not promote degranulation in isolation, but NK cell activation does occur when this stimulus is combined with NKG2D signals [75]. Notably, NKp46 signals via the CD3ζ signaling adaptor (which possesses intrinsic immunoreceptor tyrosine-based activation motifs, or ITAM), while NKG2D can signal via the DAP12 (ITAM-bearing) and DAP10 adaptors, the latter of which signals via a PI3K recruitment motif (YxxM; similar to CD28) [76]. All activating NK cell receptors do not require synergy, however. CD16 stimulation, which signals via the FcRγ adaptor (ITAM-bearing) and recognizes the Fc portions of Ab, is enough to induce degranulation in isolation, a process known as ADCC. In this case, specificity is determined by the Ab-producing B cell, which may explain why CD16-dependent NK cell activation is not subjected to the requirement for synergy, since the specificity is provided by cross-talk between the adaptive and innate immune systems [75]. Polarization of lytic granules and degranulation may also require synergy between activating receptors. For instance, in mouse NK cells, LFA-1 and NKG2D stimulation are required for granule polarization, but cannot induce degranulation [75].

All activating receptors for NK cells depend on the activity of Src-family kinases (Lck, Fyn, etc) for their phosphorylation [77]. The activity of these kinases is highly redundant, such that no single Src-family kinase is absolutely required. Exocytosis of lytic granules is dependent on PLC-γ phosphorylation, while Vav proteins are important for the reorganization of the actin cytoskeleton, a key step in polarizing the lytic granules towards a target cell [77]. For inhibitory signals, the immunoreceptor tyrosine-based inhibitory motif (ITIM) is phosphorylated by Src-family kinases, which in turn recruits the SHP-1 and/or SHP-2 phosphatases. Inhibitory signaling greatly diminishes the phosphorylation of receptors and signaling adaptors during activation, due to inhibition of a proximal signaling step involving dephosphorylation of Vav1 by SHP-1 [77].

The distribution of NK cell receptors is stochastic, making individual NK cells heterogeneous and functionally diverse. A recent study examining 28 receptors on human NK
cells by cytometry coupled to mass spectrometry (CyTOF) revealed that there are up to 30,000 diverse NK cell subpopulations per person and more than 100,000 NK cell subpopulations in an analyzed group of individuals [78].

1.2.5 MHC-I-independent recognition: The NKR-P1 and Clr Families

Among the earliest group of NK cell receptors discovered is the NKR-P1 family (encoded by the *Klrb1* genes; centromeric to *Cd69*) (reviewed in [79] and [80]). This family of receptors is somewhat unique within the NKC, because their ligands are other C-type lectin-related proteins (Clr; encoded by the *Clec2* genes), and the *Clr/Clec2* loci are genetically interspersed amongst the *Nkrp1/Klrb1* receptor genes themselves [79, 81-83](Fig1.3). Although a number of functional interactions have been demonstrated to date between different NKR-P1 receptors and Clr ligands in mice [84], rats [85], and humans [86], many of the receptor-ligand interactions remain biologically uncharacterized, and most have unknown function in vivo.

**Figure 1.3. The NK gene complex (NKC).** The NKR1 receptors and their Clr ligands are found interspersed between *A2m* and *Cd69* within the NKC. This co-localization helps to ensure co-inheritance of both receptor and ligand. The mouse Nkrp1 receptor genes (official gene nomenclature, *Klrb1*) are highlights at the top, with the Clr ligand genes (official gene nomenclature, *Clec2*) denoted below. The triangles depict the gene orientation and their colour signifies the known or suspected function of the gene products in regulating NK cell activity, as follows: co/stimulatory (green); co/inhibitory (red); bifunctional (green/red); pseudogene (*ps*, blue); or unknown (purple). Other select NK receptor genes located telomeric to *Cd69* are also depicted with common and official (*Klr*) gene nomenclatures. Not to scale.
1.2.5.1 Historical Perspective

**Discovery of the NKR-P1 (Klrb1/CD161/Ly55/Ly59/Clec5b) receptors**

The first receptor identified to be selectively expressed by NK cells was the mouse NK-1 alloantigen, which was discovered by Glimcher *et al.* in 1977 and found to be differentially expressed across mouse strains [87]. The development of a specific monoclonal antibody (PK136 mAb) facilitated its designation as the NK1.1 antigen, permitting the detection and purification of NK cells in select inbred mouse strains [88-90]. Subsequently, the NK1.1 antigen was shown to possess activating function [89, 91], providing direct evidence that NK cells express receptors that may be capable of recognizing cognate ligands on target cells [92]. However, the identity of the mouse NK1.1 antigen would remain unknown for several years [93].

In 1989, Chambers *et al.* generated a mAb (3.2.3) against a cell surface antigen present at high density on rat NK cells and purified rat lymphokine activated killer (LAK) cells [94]. The 3.2.3 antibody was shown to induce redirected NK cell cytotoxicity against FeR\(^+\) targets, as well as exocytosis of NK cell cytolytic granules, classifying it as an activating receptor. They called the antigen NKR-P1A [95]. Since ligation of mouse NK1.1 and rat NKR-P1A both induced NK cell-mediated cytotoxicity, the hypothesis arose that they could represent homologous structures [89].

Consequently, Giorda *et al.* screened a B6-strain mouse LAK cDNA library using the rat NKR-P1A cDNA and identified 3 mouse NKR-P1 homologues, which they called NKR-P1A (clone 2), NKR-P1B (clone 34), and NKR-P1C (clone 40) [96]. The cloned sequences corresponding to the mouse NKR-P1 cDNA were found to exist in different sizes, suggestive of alternative splicing. Overall, these cDNA shared between 61-74% identity at the amino acid level with rat NKR-P1A, with high similarity existing in the extracellular lectin-like region, including several conserved C residues and N-linked glycosylation sites. The discovery and designation of the NKC in mice in 1991 showed that the NKR-P1 receptor loci were distinct from the Ly49 receptor loci, despite their common expression and structural similarity [97].
Importantly, however, their genetic linkage demonstrated that a specific location on mouse chromosome 6 was dedicated to NK cell function [64]. With the physical mapping of the NKR-P1 genes and the locus encoding the NK1.1 antigen to the same region of mouse chromosome 6, along with their similar expression, structure, and function, it became increasingly likely that the NK1.1 antigen belonged to the NKR-P1 family, and this was formally demonstrated by Ryan et al. in 1992, via expression cloning of the mouse Nkrp1c<sup>B6</sup> cDNA using PK136 mAb [98].

However, it later became clear that the strain-dependent expression of the NK1.1 antigen was not only due to allelic expression of the Nkrp1c<sup>B6</sup> gene product. In 1999, two groups demonstrated that Nkrp1b gene products from the Swiss-NIH and SJL strains also reacted with the NK1.1 mAb, PK136 [99, 100]. Furthermore, the NKR-P1B<sup>Sw/SJL</sup> receptors inhibited NK cell function rather than activating NK cells like NKR-P1C. In these studies the cloned NKR-P1B<sup>B6</sup> (NKR-P1D) gene products did not react with NK1.1 mAb, nor did the NKR-P1C<sup>Sw/SJL</sup> gene products, making it unclear whether they were alleles of existing genetic loci or new genes. In any case, these results demonstrated that polymorphisms exist at both the mouse Nkrp1b and Nkrp1c loci (see below).

In 2001, a BAC contig of the Nkrp1 gene cluster in the B6 mouse strain allowed for the identification of several new genomic sequences, including Nkrp1d (Nkrp1b<sup>B6</sup>), Nkrp1e, and Nkrp1f [101]. The Nkrp1d sequence is 90% similar to that of Nkrp1b, and likely represents an allele of the Nkrp1b locus, rather than a new gene, since the coding sequence resembles that of the cloned NKR-P1B<sup>B6</sup> cDNA [100, 102]. Nkrp1e contains an early stop codon in exon-3, and splicing of intron-5 is predicted to create a frame-shift in the open reading frame (ORF), suggesting it may represent a pseudogene, at least in the B6-strain [101]. The Nkrp1f gene appears to be intact and is predicted to code for a functional protein. Work in our lab has identified the latest mouse family member, Nkrp1g, which is conserved in the genomes of B6, BALB/c, and 129-strain mice [84, 102](Fig1.3).

Subsequent investigation into the nature of strain-dependent NK1.1 reactivity showed that it was due, at least in part, to a single amino acid substitution in the NKR-P1B gene products (and presumably the NKR-P1C gene products, although this remains to be shown) [102]. The
more recent development of NKR-P1B\textsuperscript{B6} (NKR-P1D) mAb [39, 82] has shown that, unlike NKR-P1C, which is expressed uniformly by all NK cells in B6 mice, NKR-P1B is variegated and only expressed at high levels on a subset of mature NK cells [99, 100, 103]. Previous Southern blot analyses and more recent microarray-based comparative genomic hybridization (aCGH) analyses, paired with phenotypic NK subset comparisons, have shown this to be a generalized phenomenon across many strains [97, 104].

A recent comprehensive study by Hao \textit{et al.} allowed an in-depth analysis of the NKC regions from many species, including the rat \textit{Nkrp1} genes, although no functional data were determined from this computational study [63]. The rat NKC has at least 4 \textit{Nkrp1} genes, predicted to represent orthologs of the \textit{Nkrp1a/c}, \textit{Nkrp1b/d}, \textit{Nkrp1f}, and \textit{Nkrp1g} gene sets (Fig 1.4). Like the \textit{Nkrp1b}\textsuperscript{B6}, the rat \textit{Nkrp1c} gene likely represents a divergent \textit{Nkrp1b} allele present in the PVG rat strain (\textit{Nkrp1b}\textsuperscript{PVG}; see below) [105]. While the activating function of the rat NKR-P1A receptor was known since the late 1980’s, the inhibitory function of the rat NKR-P1B isoform was not shown until much later. Here, the Miller lab transfected human YTS cells with a rat cDNA library and sorted cells using the 10/78 mAb (similar NKR-P1A reactivity to 3.2.3 mAb) [106]. Sequence analysis of positive clones fell into two groups corresponding to either rat NKR-P1A or NKR-P1B. Using $^{51}$Cr-release assays, they showed that the rat NKR-P1B clones inhibited NK cytotoxicity. It is now clear that both the NKR-P1A/B gene products from some rat strains react with the 10/78 and 3.2.3 mAb, demonstrating that polymorphisms also exist in selected rat \textit{Nkrp1} loci [107].
Figure 1.4. NKR1:Clr composition and organization in mice, rats, and humans. (Top) The mouse Nkrp1/Klrb1 receptor genes and Clr/Clec2 ligand genes on mouse chromosome 6 (B6-strain reference sequence, genome build: Dec. 2011, GRCm38/mm10) are depicted, along with their orientation (triangle direction), known or suspected function (triangle colour), and intron-exon structure (vertical exon lines); triangle representations are as in Figure 1.3. (Middle) The mouse Nkrp1/Klrb1 receptor genes and Clr/Clec2 ligand genes on rat chromosome 4 (BN-strain reference sequence, genome build: Nov. 2004, Baylor 3.4/rn4) are depicted as above. (Bottom) The human KLRB/F receptor genes and the CLEC ligand genes on human chromosome 12 (genome build: Feb. 2009, GRCh37/hg19) are depicted as above. Not to scale.
The closest human homologue, NKR-P1A (CD161/KLRB1), was identified in 1994 through the development of the DX1 mAb against human LAK cells [108]. Human NKR-P1A shares 45% identity at the amino acid level with rat NKR-P1A and 46-47% homology with the mouse NKR-P1 proteins, but significantly lower homology with other C-type lectin-like NK cell receptors. Ligation of human NKR-P1A with DX1 mAb did not induce cytotoxicity against FcR-bearing P815 cells or have an effect on lysis of K562 target cells. However, when added to NK clones with the spontaneous ability to lyse P815 cells, cytotoxicity was inhibited, while F(ab’)2 of anti-NKR-P1A mAb did not inhibit, suggesting FcR cross-linking was important. Paradoxically, NK cell clones that do not spontaneously kill P815 targets could have a small proportion induced to kill P815 in the presence of anti-NKR-P1A mAb. In 2005, two groups showed that the NKR-P1A receptor was indeed inhibitory on human NK cells, demonstrating it to be a functional homolog of the rat and mouse NKR-P1B receptors [109, 110]. Also similar to the rodent NKR-P1B, only a subset of human NK cells express NKR-P1A [108]. However, the existence of another putative inhibitory receptor in rodents, NKR-P1G, calls into question the orthology with NKR-P1B. Whereas several different NKR-P1 transcripts could be detected by Northern blot in mouse NK cells [111], only one band was detected in human NK cells [108]. Nonetheless, an apparent lack of stimulatory NKR-P1 homologues in the human genome was recently re-investigated by the Steinle group, who suggested that the human KLRF1/2 gene products, located telomeric to CD69, may actually represent divergent activating NKR-P1 homologs [86](Fig1.4).

**Discovery of the Clr (Clec2d/Ocil/LLT1) ligands**

While investigating bone morphogenesis, Zhou et al. discovered a gene product they designated osteoclast inhibitory lectin (Ocil), after its ability to inhibit osteoclast formation when expressed on osteoblast cells [112]. They went on to show that Ocil was a prototypical member of a group of similar gene products in mice (Ocil, Ocilrp1, Ocilrp2) [113]. The extracellular domains of all three gene products are similar to each other at the amino acid (>70%) and nucleotide (~90%) levels. Around the same time, Plougastel et al. identified a novel set of C-type lectin-related (Clr) genes by sequencing B6-strain BAC of the NKC region centromeric to Cd69
The Clr gene products were shown to possess 40% amino acid similarity to the lectin-like domain of CD69, and 70-80% identity at the nucleotide level to each other. They named the genes \( \text{Clr-a,b,c,d,e,f,g} \), where \( \text{Clr-b} \) was identical to \( \text{Ocil} \), \( \text{Clr-g} \) was identical to a splice variant of \( \text{Ocilrp2} \), and \( \text{Clr-d} \) (and perhaps \( \text{Clr-c} \)) were similar to splice variants of \( \text{Ocilrp1} \) [113, 114]. Since then, new family members, namely \( \text{Clr-h, Clr-i, and Clr-j} \), were discovered by hybridizing partial \( \text{Clr} \) PCR products to mouse BALB/c and 129-strain BACs [84, 102].

Subsequent to the identification of the \( \text{Ocil/Clr} \) gene family, several screens aimed at identifying NKR-P1 ligands demonstrated interactions between gene products of the two families (reviewed in [79]). Briefly, interactions between the NKR-P1B/D:Clr-b as well as NKR-P1F:Clr-g receptor-ligand pairs were shown using cellular assays (involving BWZ.36 reporter cells bearing CD3ζ/NKR-P1 fusion receptors), NKR-P1 tetramers, and blocking mAbs for NKR-P1B\(^{B6}\) (a.k.a., NKR-P1D; 2D12 mAb) or Clr-b (4A6 mAb) [81, 82]. These results were significant in that they showed an interaction between the NKR-P1 receptors, which were previously thought to recognize carbohydrates [116, 117], and protein ligands. More recently, additional interactions have been shown in the mouse system, including NKR-P1B/D:Clr-b, NKR-P1F:Clr-c,d,g, and NKR-P1G:Clr-d,f,g [84].

The first foreign ligand identified in the rat system was a viral immunoevasin, RCTL, a spliced ORF with C-type lectin-like sequence homology identified in 2001 [118], derived from the rat cytomegalovirus-English isolate (RCMV-E; Mhv8 [119]). \( \text{Rctl} \) is expressed as an early gene upon infection of rat embryonic fibroblast (REF) cells [107, 118]. The RCTL gene product appears to functionally replace the endogenous Clr-b-like ligand (rat Clr11; \( \text{Clec2d1I} \)), which is rapidly lost during RCMV-E infection; notably, both the host and viral ligands are recognized by specific alleles of the NKR-P1B receptor in certain rat strains [107]. Prior to this work, no information was available regarding the rat Ocil/Clr gene family, until a breakthrough publication in 2006 by Hao \textit{et al.} described the complete repertoire of NKC-associated \( \text{Clec} \)-like
genes in several species [63]. This in-depth analysis outlined at least 11 rat Clec2d-like genes or pseudogenes, making the rat system the most complex family identified to date. Due to close sequence homology and an inability to determine strict orthology with the mouse family members, the rat Clr genes were simply designated Clec2d1-11 (centromeric to telomeric), with the closest relative of mouse Clr-b (mouse Clec2d8) predicted to be rat Clec2d11, based upon sequence, structure, and genomic position [63]. Indeed, rat Clec2d11 was shown to be recognized by two distinct host NKR-P1B alleles [107]. In 2009, the promoter region of a rat Ocil homologue was first described [120]; however, the gene corresponding to rat Ocil is actually another Clr family member, rat Clec2d5 [63].

The first in-depth analysis of rat NKR-P1 ligands, which adopted a rat Clr1-11 nomenclature based upon the corresponding Clec2d1-11 assignments, used an NFAT-driven GFP-reporter cell assay (BWN3G, similar to the BWZ.36 LacZ-reporter cell assay), to show that rat NKR-P1A and NKR-P1B (clade 1 receptors) recognize Clr11, while NKR-P1F and NKR-P1G (clade 2 receptors) recognize an overlapping set of ligands, namely Clr2,3,4,6,7 and Clr2,6,7, respectively [121]. Follow up work by this group showed that the mouse and rat NKR-P1F/G receptors are highly xenoreactive, cross-recognizing corresponding ligands from the other species and allowing some degree of assignment of functional orthology [85].

In the human system, a C-type lectin-related sequence was discovered in the NKC in 1999 by Boles et al., by searching the expressed sequence tag (EST) database with a consensus sequence of known human C-type lectin-like receptors in the NKC [122]. They screened an NK cDNA library and identified an 850bp cDNA, designated lectin-like transcript-1 (LLT1), and later generated a mouse anti-LLT1 mAb (L9.7) by immunizing AKR/J2 mice with LLT1 fusion proteins [123]. LLT1 is similar in structure and function to the mouse Clr gene products, and indeed has been given the synonymous gene designation, CLEC2D, as the mouse Clr-b locus (Clec2d). Actually, Hao et al. identified two homologues, CLEC2D1p (a predicted pseudogene) and CLEC2D2 (which encodes LLT1) [63]. However, LLT1 also shares significant homology with other rodent Clr gene products, as well as other human gene products, such as KACL (CLEC2A), AICL (CLEC2B), and CD69 (CLEC2C) [79, 86, 122, 124, 125]. As with mouse
Ocil/Clr-b, human OCIL was also discovered separately from LLT1 in 2004 [126]. The human OCIL cDNA encodes an identical protein to LLT1, and is 42% similar to mouse Ocil/Clr-b at the protein level.

The above features of LLT1 made it an attractive candidate for the interaction with the human NKR-P1A receptor [123]. Indeed, in 2005, simultaneous publications from two groups formally demonstrated an interaction between the human NKR-P1A (KLRB1) receptor and the LLT1 (CLEC2D) ligand using LLT1 multimers and liposomes, reciprocal NKR-P1A/LLT1 reporter cell assays, as well as functional cytotoxicity and IFN-\(\gamma\) production assays [109, 110]. Interestingly, while LLT1 inhibits cytotoxicity and IFN-\(\gamma\) production via NKR-P1A on NK cells, it appears to co-stimulate IFN-\(\gamma\) production by NKT cells [109, 127]. Reciprocal studies have suggested that LLT1 itself may signal on NK and NKT cells to augment IFN-\(\gamma\) production via a Src family-ERK signaling pathway [123, 128]. Expression of LLT1 is induced on human B cells and DCs by TCR/BCR agonists, CD40 ligation, mitogens, and TLR agonists [110, 127]. More recent studies have identified splice variants of the human CLEC2D gene, only one of which (LLT1) appears to functionally bind the human NKR-P1A receptor [127] (Oscar A. Aguilar, unpublished results in our laboratory). As mentioned above, the Steinle laboratory has also reported significant CLEC2 group homology between KACL (CLEC2A; linked to its receptor, KLRF2), AICL (CLEC2B; linked to its receptor, KLRF1), CD69 (CLEC2C; unknown receptor), and LLT1 (CLEC2D; linked to its receptor, KLRB1) [86].

1.2.5.2 Expression

Mouse

Transcripts for all mouse Nkrp1 genes have been detected in the spleen, thymus, lymph nodes, and other hematopoietic tissues, basically, wherever NK cells are found [115]. Additionally, the lung and intestine have showed expression of almost all Nkrp1 transcripts, with the exception of Nkrp1a [115]. Nkrp1b/d is unique in showing expression within the tongue and bladder [115]. Within these tissues, Nkrp1a, Nkrp1b/d, Nkrp1c, and Nkrp1f transcripts are present in NK cells, while Nkrp1c is also found in NKT cells [115]. Subsets of sorted cells
classified as tissue DCs and macrophages have also been shown to express Nkrp1b/d [115]. Nkrp1b/d transcripts are also highly expressed in ILCs, including ILC3-like and LTi-like cells, where the gene may be induced by expression of the RORγt (Rorc) transcription factor [28]. They are also enriched in a CD8αα subset of intraepithelial lymphocyte (IEL); notably, these latter IEL also express greatly enriched levels of transcripts for Nkrp1a, upregulated almost 30-fold compared to CD8αβ IEL [129]. Nkrp1f transcripts also are present in BM DC/monocyte precursors and lymph node endothelial cells [115], and expression has also been reported in DC/APC [130, 131]. While Nkrp1c is found in almost all NK and NKT cells, expression of Nkrp1b/d is only found at high levels in a subset of NK cells (50-70%), but few NKT cells (~10%) [39, 41, 42]. Nkrp1g transcripts are most highly expressed in the spleen and small intestine, but are also present in the colon, BM, thymus, and lymph nodes [132]. Within these tissues, Nkrp1g transcripts are found predominately in the IEL of the small intestine [132].

Consistent with transcript data, most NKR-P1 receptors are expressed on mouse NK cells; however, few NKT cells express surface NKR-P1 receptors, with the exception of NKR-P1C, which is expressed at intermediate levels by NKT cells [39]. Mouse NKRP1B/D is only detected at high levels on a subset of NK cells (50-70%, depending on the strain), whereas the remainder expresses only low levels [39, 82]. Strain-dependent expression is detectable on NK cells from a number of strains (CE, B6, NZB, C58, Ma/My, ST, SJL, FVB, and Swiss mice, but not BALB/c, AKR, CBA, C3H, DBA, or 129 nice); however, expression of the particular NK1.1 antigen involved (NKR-P1B or NKR-P1C) seems to be dictated by whether the majority of the cells are NK1.1+ (NKR-P1C), or whether the NK1.1 expression is variegated (NKR-P1B), as determined by gating on DX5+ or NKp46+ NK cells [39, 102, 104]. Surface expression of mouse NKR-P1G is variegated, like NKR-P1B/D, and solely found in the intestines, particularly on CD3−CD103+, γδTCR+, and αβTCR+ IEL cells [132].

Interestingly, both activating and inhibitory NKR-P1 isoforms are rapidly downregulated from the cell surface following ligation by plate-bound antibody or cognate ligands [39]. On the other hand, surface NK1.1 (NKR-P1C) expression is induced on conventional CD8+ T cells following viral infection and activation by other stimuli [133-136]. Expression of NK1.1
isoforms (NKR-P1B>NKR-P1C) is also found on immature thymocytes, thymic NK cells, fetal blood progenitors, subsets of mucosal T cells, and at low levels on NKp46⁺ gut ILC [137-144].

Expression of the mouse Clr ligand genes is variable and highly regulated. Mouse Ocil (Clrb) was reported to be expressed broadly in most tissues [114, 115, 132], but is also prevalent during bone development and upregulated in response to osteotropic factors like retinoic acid, 1,25-dihydroxyvitamin D₃, parathyroid hormone, IL-1α, IL-11 [112]. This regulation of Ocil appears to be unique, as Ocilrp1 (Clrd) and Ocilrp2 (Clrg) were not regulated by these treatments, despite Ocilrp2 being co-localized in the same tissues as Ocil [113]. At the transcript level, Clra is only expressed in the intestines and even then only at low levels [115]. Clrb is expressed in all nucleated hematopoietic cells, some non-hematopoietic cells, and within most tissues, with the exception of the brain, which appears to be devoid of Clr and Nkrp1 transcript expression [81, 112, 115, 132]. In addition, Clrb transcript expression is tightly regulated on healthy versus pathological cells. While expressed highly in normal cells, the transcripts are lost following transformation [81], viral infection [107, 145], and genotoxic and cellular stress [40]. Clrc is expressed at low levels in the tongue, spleen, thymus, ovaries, testes, and lymph node tissues; Clrd transcripts are uniquely present in the eye; and Clrf is expressed in the liver and very highly in the kidney and intestine, specifically in intestinal epithelial cells and kidney tubular epithelial cells [114, 115, 132]. Clrg is expressed in the spleen, thymus, and lymph node, particularly in intestinal epithelial cells, as well as LAK cells [114, 115, 132]; notably distinct Clrg splice isoforms exist (a.k.a., LCL-1a,b,c,d) that may have distinct expression patterns [130, 131]. Expression of the remaining family members requires further investigation.

In terms of cell surface expression and protein expression, Clr-b has been the most widely studied, as determined by using the 4A6 mAb [40, 81, 145]. It appears to be a constitutive surface marker of healthy cells, as nearly all nucleated hematopoietic cells express it, as do normal embryonic and adult skin fibroblasts [40, 81, 145]. In line with this, Clr-b is frequently downregulated on pathological target cells, including most hematopoietic tumour cell lines [81], virally infected cells (e.g., MCMV, RCMV) [107, 145], and cells undergoing genotoxic or cellular stress [40]. Thus, like MHC-I molecules, Clr-b demonstrates a “missing-self”
recognition pattern, expressed by healthy cells, but lost on target cells undergoing pathological changes, making them more susceptible to NK cell recognition. Interestingly, the downregulation of Clr-b occurs at both the steady-state transcript and protein levels, suggesting that transcriptional, as well as post-transcriptional and post-translational mechanisms, play important roles in regulating surface Clr-b expression, and therefore regulating NK cell activity [40, 107, 145]. Notably, loss of cell surface Clr-b on stressed cells is abrogated by inhibitors of the ubiquitin-proteasome and endolysosomal pathways (e.g., MG132, lactacystin, chloroquine), many of which also affect autophagy [40]. Clr-f is highly and uniformly expressed on differentiated intestinal epithelial cells lining the intestinal villi, and on colon epithelial cells, but not on IEL [132]. Expression of the other mouse Clr family members at the cell surface has been less well documented due to lack of specific reagents.

**Human**

Human NKR-P1A is expressed on the majority of (but not all) CD56\(^{bright}\)CD16\(^{-}\) and CD56\(^{dim}\)CD16\(^{+}\) NK cells, subsets of CD4\(^{+}\) and CD8\(^{+}\) TCR\(\alpha\beta\) T cells, NKT cells, as well as TCR\(\gamma\delta\) T cells [110, 146, 147]. Expression is higher on memory versus naïve T cells [108, 148]. Induction of NKR-P1A occurs in the fetal liver early in developmental ontogeny [108, 146] and it is one of the earliest markers of human NK cell development [149-152]. It is also expressed de novo on CD34\(^{-}\) and CD34\(^{+}\) immature thymocytes, and is induced on thymocytes upon culture in rIL-2 [153]. Surface NKR-P1A is upregulated on immature and mature NK cells upon exposure to IL-12 [149, 154, 155]. NKR-P1A has also been reported to be a marker of all Th17 cell subsets (where it is induced by RORC [156]), human ILC and LTi subsets [157], as well as a novel subset of FoxP3\(^{+}\) Treg-like cells found in both healthy and arthritic humans that secrete proinflammatory cytokines [158]. CD4\(^{+}\)NKR-P1A\(^{+}\) T cells are capable of migration in transwell assays and transendothelial migration *in vitro* [159]. Acquisition of NKR-P1A has also been reported to be an early event in monocyte differentiation and on DC [160].

In contrast to rodent Clr-b, and perhaps more akin to other rodent Clr, human LLT1 transcripts are only expressed at low levels in NK cells, T cells, B cells, and osteoblasts, but are undetectable in monocytes [127, 146, 161]. LLT1 induction in osteoblasts has also been
documented using IL-1-α and PGE$_2$ [126]. While freshly isolated B cells and DCs express little LLT1 mRNA, transcript levels are augmented following infection with viral pathogens (e.g., influenza, HSV-1, EBV, HIV) or TLR agonists (i.e., TLR-3,4,7,8,9), in particular CpG DNA treatment [127, 146]. LLT1 expression is also rapidly induced on the surface of bronchial epithelial cells in response to respiratory syncytial virus (RSV) infection or treatment with proinflammatory cytokines, including type-I interferon, IL-1β, and TNF-α [162]. Again unlike rodent Clr-b, LLT1 protein is not detected at the surface under naïve, resting conditions; however, mirroring transcript expression, surface LLT1 is highly upregulated upon activation with the stimuli noted above. Induction on B cells also occurs following IgM crosslinking or CD40 ligation. LLT1 expression is also inducible on T cells upon stimulation via CD3 crosslinking, PMA, PHA, and IL-2 [123, 127, 146]. On DC, the most potent inducers of LLT1 expression are CpG DNA, polyI:C, LPS, TLR-7/8 agonists, and IFN-γ [127]. LLT1 is also induced on NK cells upon activation, and following incubation with certain NK-sensitive target cells [127, 146]. It remains to be determined whether NKR-P1A:LLT1 interactions regulate NK and T cell functions in cis, in addition to recognition of target cells in trans [44, 163].

Expression of the other human CLEC2 ligands is more restricted. As the names suggest, KACL is expressed in keratinocytes, AICL is induced upon activation of hematopoietic cells, CD69 is induced early after activation of T and NK lymphocytes and other hematopoietic cells, BACL (CLEC2L) is expressed in brain tissue, and the more distantly-related DCAL (CLEC1L) is expressed in DC [86, 164].

1.2.5.3 Structure

Genomic

In the mouse B6-strain reference genome, all of the Klrb1 genes have 6 coding exons: exon-1 codes for the cytoplasmic domain, exon-2 for the transmembrane domain, exon-3 for the stalk region, and exons-4,5,6 for the extracellular lectin-like domain [84, 101]. This gene structure is conserved for the Klrb1 genes in the BALB/c and 129 strains [115] and human
*KLRB1* [108]. Notably, the human *KLRF1* and *KLRF2* genes also have a 6-exon gene structure (Fig1.4).

The *Clec2* genes in the mouse also have a similar genomic structure (Fig1.4). All the genes have 5 exons, with the following exceptions: mouse *Clrg* (*Clec2i*, 6 exons), mouse *Clrd* (*Clec2g*), 7 exons; mouse *Clre* (*Clec2-ps1*, 3 exons); mouse *Clrj* (*Clec2-ps3*, 2 exons). The human *ICLEC2D* gene that encodes LLT1 also has 5 exons, as does upstream *ICLEC2Dlp* pseudogene described by Hao, et al. [63]. The human *CLEC2A* and *CLEC2B* genes also possess a 5-exon gene structure, and the same is true for the *CLEC2C* gene that encodes CD69 [86, 164]. The recently described *CLEC2L* (BACL) gene, which is not linked to the NKC, also possesses 5 exons, but the less related *CLECL1* (DCAL) gene linked to the NKC only has 4 exons.

The promoter regions of the rodent *Clec2* genes remain poorly characterized. In mouse, *Ocil/Clr* and *Ocilrp2/Clrg* have been reported to possess an inverted TATA box upstream of their transcriptional start sites, while *Ocilrp1/Clrd* has a more traditional TATA box [113]. Both *Ocilrp1/Clrd* and *Ocilrp2/Clrg* have been reported to generate alternative splice variants [113], and an independent characterization of LCL-1 (*Ocilrp2/Clrg*) transcripts suggests that up to 4 different splice variants exist (LCL-1a,b,c,d) [130, 131]. The human *Ocil* (*CLEC2D/LLT1*) gene lacks an inverted TATA sequence, instead containing a GAATCA sequence upstream of the TSS [126]. Human *ICLEC2D* has also been reported to generate several alternative splice variants (other than LLT1), some of which may be functional (e.g., coding for proteins retained in the ER as heterodimers with LLT1, or others lacking a transmembrane domain) [161].

**Protein**

The NKR-P1 and Clr proteins belong to group-V of the 14 C-type lectin superfamily groups [63, 165]. C-type lectins are a class of glycoprotein characterized by their Ca\(^{2+}\) dependence, conserved disulphide-linked cysteine residues, and functional carbohydrate recognition domains [165]. The NKR-P1 and Clr proteins are designated as “lectin-like” because they primarily bind other proteins, and they possess somewhat atypical conservation of their cysteine residues, such that they may have lost the residues and loop structures required to
coordinate divalent calcium ions, and thus lack high affinity Ca\(^{2+}\)-dependent binding to carbohydrates. For example, the mouse Clr and human LLT1 are all missing the C5 residue in their lectin-like domain, and the C4 residue is also absent from mouse Clr-b and Clr-g [114]. The implications of these changes require further investigation. However, it should be noted that the rat NKR-P1 receptors were first reported to possess functional carbohydrate recognition domains [116, 117]. In addition, the human LLT1/OCIL and mouse Clr-b/Ocil proteins have been reported to bind high-molecular weight carbohydrates (including sulphated forms) in a Ca\(^{2+}\)-independent manner, blockable using soluble carbohydrates (including heparin and chondroitin sulphate) [126, 166].

Initial work on the structure of the mouse NKR-P1 proteins demonstrated that they likely exist as disulfide-linked homodimers. For example, NKR-P1D (NKR-P1B\(^{B6}\)) is \(~90\) kDa under non-reducing conditions, but \(~47\) kDa under reducing conditions, when assessed using 2D12 mAb [82]. The Clr proteins also likely exist as homodimers, but heterodimers between the Clrs have also been speculated to exist [82]. Crystal structures for the mouse NKR-P1 proteins are just beginning to be published and assessed. Although early co-crystal structures were first reported in 2003 for the mouse NKR-P1F:Clr-g and NKR-P1D:Clr-b proteins (the latter reported under conditions requiring mutagenesis of certain residues), these have not been published [82]. Published work suggests that mouse NKR-P1A\(^{B6}\) and NKR-P1C\(^{B6}\) have a similar structure, with slight differences explainable by their amino acid sequences (e.g., NKR-P1A\(^{B6}\) has a higher Y residue content than NKR-P1C\(^{B6}\) [167]. Mouse NKR-P1C\(^{B6}\) also has a better-defined conformation of disulfide bridges than NKR-P1A\(^{B6}\), suggesting that NKR-P1C is more rigid and stable; indeed, this is supported in the literature regarding their individual stable expression at the cell surface on the NK cells [39]. However, it should be noted that the NKR-P1C\(^{BALB}\) receptor lacks a potentially crucial conserved C4 residue (position C122S) that may have implications for its expression and function, in addition to its lack of NK1.1 reactivity due to its S191T substitution (which is also found in the functional NKR-P1B\(^{BALB}\) receptor) [102, 168].

In general, all of the NKR-P1 proteins seem to possess a fold similar to other C-type lectin-like domains, including at least two \(\alpha\)-helices and two antiparallel \(\beta\)-sheets. They have a
β-core composed of long anti-parallel β-strands that form a central pillar, flanked at one end by short β-strands and at the other end by a β-sheet [167]. This core is surrounded by α-helices (2-3, depending on the isoform). All appear to have 3 disulfide bonds, except mouse NKR-P1C\textsuperscript{BALB}, which has only two [167]. Comparison of all known NKR-P1 ectodomains revealed the highest conservation in the β-core, less conservation in the loop regions, with the longest loop suggested to play a role in ligand specificity [167]. A mouse NKR-P1A crystal structure reported recently suggests a claw-like structure may be important for ligand interaction [169].

Recent biochemical studies of the human NKR-P1A:LLT1 proteins has shown that they interact with a \(K_d\sim50\ \mu M\) [170]. Reciprocal mutagenesis of the proteins also provided a partially validated structural model of the interaction, supporting a dimeric interaction with several important contact residues. Even more recent co-crystal structures of the high-affinity (\(K_d\sim2\ \text{nM}\)) human NKp65:KACL interaction have suggested a conserved docking topology for the interaction of genetically-linked C-type lectin-like receptor-ligand pairs in the NKC [171]. Here, two NKp65 receptor monomers, with limited dimeric interface between them, contact a single dimeric KACL ligand bivalently at two distinct but symmetrical sites in a butterfly-shaped complex. A similarity in this study was noted to the compact Ly49C dimer interacting independently with two monomeric H-K\textsuperscript{b} MHC-I ligands [172]. Whether or not this applies to other NKC-encoded receptor-ligand pairs remains to be elucidated, but the remaining genetically-linked receptor-ligand interactions have all been shown to interact with much lower affinities (\(\mu M\) versus nM) [171].

Much less is known about the structure of the rodent Clr and human LLT1 proteins. LLT1 has been modeled on a CD69 structural backbone [161, 173] and models have also been proposed based upon mutagenesis studies [170], as well as the human NKp65:KACL crystal structure [171, 174]. LLT1 possess a typical C-type lectin-like fold and the membrane distal part of LLT1 forms a square face that is predominately positively charged [173]. LLT1 hexamers were identified during analysis of the X-ray structure consisting of three LLT1 dimers packed together [174]. The biological relevance, if any, of these hexamers has not been determined.
Recently, the structure and biophysical properties of mouse Clr-g were determined by expressing soluble portions of the extracellular lectin-like domain [175]. In this structure, Clr-g exists as a monomer or dimer, depending on the size of the ectodomain used. In terms of secondary structure, it has two α-helices, two anti-parallel β-sheets composed of 3 β-strands each, and two disulphide bonds, with a dimeric interface involving 10 hydrogen bonds. Mouse Clr-g has mostly positive electrostatic potential, while its cognate receptor, NKR-P1F, has negative potential, suggesting electrostatic potential may be a driving force in their contact [175]. Notably, models of the other Clr, and other C-type lectin-like proteins, are predicted to possess distinct surface electrostatic potential in this work. Mouse Clr-g also shares almost ~50% sequence identity with human KACL, and predictive models of most of the human and mouse NKR-P1:Clr receptor-ligand interactions have been outlined based upon the human NKp65:KACL interaction [171].

1.2.5.4 Function

Signaling

Signaling via NK cell receptors is hypothesized to deliver either an activating signal or to exert inhibitory effects on downstream signaling cascades, and the NKR-P1 family is no exception to this notion of paired or balanced recognition systems. Early experiments demonstrated that Syk acts as a common signaling element important for both FcR-initiated and natural cytotoxicity in NK cells [176]. However, additional pathways involving Src and PI3K kinases are also involved in triggering natural cytotoxicity [177-179].

In this light, it was shown that the mouse NKR-P1C receptor activates both redirected cytotoxicity and IFN-γ production in NK and NKT cells, and that association with the FcRγ adaptor protein was crucial for these functions [89, 180]. Furthermore, while FcRγ homodimers were shown to be required for both optimal NKR-P1C expression and signaling in both NK cells and NKT cells [180], FcRγ/CD3ζ heterodimers may be sufficient for residual NKR-P1C cell surface expression, in contrast to CD16 expression, which is dependent upon FcRγ homodimers and is negatively regulated by CD3ζ (via FcRγ/CD3ζ heterodimers) [181]. In addition, the rat
NKR-P1A receptor was shown to induce Ca\(^{2+}\) flux, phosphatidylinositol turnover, \(\text{PLA}_2\) activation, arachadonic acid release, heterotrimeric G-protein activation, redirected cytotoxicity, and granule exocytosis by rat NK cells and the RNK-16 cell line [182-184]. Subsequent analysis of a number of NKR-P1 protein sequences revealed the presence of potential PLC-\(\gamma1/2\) motifs (YxxL), SH3-binding proline-rich domains, Lck-recruitment motifs (CxCPR/H), and variably present ITIM motifs (I/L/VxYxxL/V) in the rodent but not human NKR-P1 receptors [185]. The activating rodent NKR-P1 receptors were later shown to contain a conserved positively charged amino acid residue (R) in their transmembrane domain, likely for FcR\(\gamma\) adaptor association (based upon a distinct transmembrane R amino acid position versus K for DAP12-associated receptors), while the inhibitory rodent NKR-P1 receptors contain consensus ITIM motifs [99, 185, 186].

Notably, the rodent NKR-P1C CxCPR/H motifs, similar to those found in the CD4 and CD8 T cell co-receptors, were shown to functionally recruit Lck by immunoprecipitation, yeast two-hybrid analysis, and functional mutagenesis [187, 188]. In addition, Lck was also shown to be necessary for efficient NKR-P1C-mediated redirected lysis, as cytotoxicity was diminished in BM-derived LAK cells from Lck\(^{-/-}\) mice (whereas splenic LAK cells may use another Src-family member to compensate) [189]. The ITIM motif in the mouse NKR-P1B receptor was also shown to recruit the SHP-1 phosphatase upon pervanadate stimulation (and to a lesser extent SHP-2) [99, 189, 190]. A consensus ITIM motif [190] is also present in the mouse and rat NKR-P1G receptors, which are presumably inhibitory; however, mouse NKR-P1G lacks the consensus Lck-recruitment motif [84]. The functional consequences of the loss of di-cysteine motif for Lck recruitment remain unknown, but may suggest a more direct inhibitory receptor function versus a co-inhibitory receptor function, if the CD4/8 T cell co-receptors provide any reference. The mouse and rat NKR-P1F receptors may either be stimulatory, co-stimulatory, or function as co-receptors in terms of signaling [39], as they retain a charged transmembrane residue, yet mouse NKR-P1F lacks the conserved YxxL motif, instead containing a Y residue in a different position and context [84].
Like the rat NKR-P1A receptor, the mouse NKR-P1C receptor (NK1.1 in B6 mice) has also been shown to signal IFN-γ production in part via PI3K (p110γ>δ) activity [191], suggestive of some co-stimulatory function, akin to NKG2D isoforms [192]. Given the specific expression of NKR-P1C on NKT cells (but not other mouse NKR-P1 receptors [39]), its ability to recruit Lck (akin to a T cell co-receptor), and its partial signaling via the PI3K pathway (similar to CD28, NKG2D), it could serve as a co-receptor or costimulatory receptor (signal 2) for TCR activation (signal 1) on NKT cells in their recognition of CD1d-restricted ligands. Since the mouse NKR-P1C ligand(s) remain unknown (see below), it is intriguing to speculate that endogenous glycolipids (and/or glycoproteins) could be recognized simultaneously by the NKT cell TCR and this lectin-like receptor to facilitate dual recognition (co-receptor function) and/or co-stimulation. Notably, conventional NK cells express most NKR-P1 receptor isoforms (whereas NKT cells more uniquely express NKR-P1C [39]), and CD1d over-expression on target cells has been reported to inhibit NK cell function via an as yet unidentified receptor [193, 194]. Moreover, unlike conventional CD4+ and CD8+ T cells, NKT cells are not yet known to express a co-receptor for CD1. Of note, during the preparation of this thesis, we recently identified a viral ligand for the mouse NKR-P1B and NKR-P1C (NK1.1) receptors in the MCMV genome (O.A. Aguilar et al., manuscript submitted). We are currently evaluating the structural and functional effects of this ligand on NK cells, NKT cells, and other NK1.1+ immune cell subsets.

Interestingly, the cytoplasmic tail of the human NKR-P1A receptor contains neither a strong consensus ITIM motif (I/L/VxYxxL/V) [99, 189, 190], nor an intact Lck-binding motif (CxCPR/H). Nonetheless, the human NKR-P1A cytoplasmic tail does contain a tyrosine residue in an atypical motif (AxYxxL) that may function as a weak ITIM [110, 185], and the receptor has been shown by immunoprecipitation and Western blotting to form large complexes containing the Src-family kinases, Lck, Fyn, and Lyn in Brij-58-solubilized NK cells [195]. However, association with SHP-1 has not been demonstrated [109, 110], and Lck association was not observed in CD161+ NKT cells [196]. The receptor has been shown by yeast two-hybrid and coimmunoprecipitation studies to associate with acid sphingomyelinase in primary human cell lines and NK cells [197]. Ligation using CD161 mAb results in association of CD161 with aSMase in detergent-resistant membranes, activation of aSMase activity, production of ceramide,
PI3K-dependent activation of the Akt and ERK pathways, and proliferation [197]. Ligation of human NKR-P1A on thymocytes has been reported to inhibit cytotoxicity but enhance proliferation [153], expression on T cells enhances transendothelial migration [159], and ligation on macrophages and DC fluxes intracellular calcium and leads to IL-1α and IL-12 production [160]. More recently, human NKR-P1A has also been shown to be expressed specifically by Th17 cells [156, 198], MR1-restricted MAIT cells [199], and a unique Treg-like subset [158], where its function(s) are still being evaluated [147].

**Autoimmunity and Disease Association**

Human NKR-P1A has been shown to be preferentially expressed on γδ T cells expressing the Vδ2 chain [148]. Interestingly, the proportion of γδ T cells and IL-17+ T cells expressing NKR-P1A is higher in MS patients (~70%) than in healthy donors (~20%) [148, 200]. Upregulation of human NKR-P1A on γδ T cells also occurs following exposure to IL-12, as it does for NK cells, but only for the Vδ2+ population, because Vδ1+ cells express lower levels of the IL-12Rβ2 subunit [148]. These NKR-P1A+ Vδ2+ T cells also undergo transendothelial migration in an NKR-P1A-dependent manner in vitro. Since MS patients have more of these T cells, it has been proposed that they use NKR-P1A to migrate to local lymph nodes, recirculate, and extravasate into the brain, leading to exacerbated MS symptoms [148].

NKR-P1A+ T cells can also be detected in inflammatory infiltrates in patients suffering with psoriasis [198] and Crohn’s disease [201], with 20-fold more NKR-P1A−CD4+ T cells present in patients with Crohn’s disease than healthy patients [201]. Whether this is due to upregulation of NKR-P1A on reactive T cells, or exacerbated production or expansion of pre-existing NKR-P1A+ Th17 cells, requires further investigation [147].

Interestingly, an N19K substitution in the human OCIL (CLEC2D) gene has been found to be associated with reduced bone mineral density in postmenopausal women [202]. Genome-wide association studies should reveal new interactions between this receptor-ligand system and human disease.
Finally, in rheumatoid arthritis (RA) synovial fluid, LLT1 is expressed on a small proportion of monocyte cells [203]. Patients with RA also have higher levels of soluble LLT1 in their serum compared to healthy controls, suggesting a role for LLT1:NKR-P1A interactions in arthritis [203].

**Viral Infection**

Viral evasion of host innate and adaptive immunity is a well-documented evolutionary occurrence. In 2001, a spliced C-type lectin-like ORF was identified in the rat cytomegalovirus-English (RCMV-E) genome [118]. The gene product was designated RCTL, and it was later shown to closely resemble the coding sequence of an endogenous rat Clec2d-like sequence, Clr11 (Clec2d11), a predicted ortholog of mouse Clr-b (Clec2d) [107]. Indeed, a mAb specific for the RCTL protein (R3A8 mAb) was also shown to crossreact with the host rat Clr-b (Clec2d11) protein. Expression of rctl was detected as early as 3h post-infection (characterizing it as an early gene), while reciprocal expression of host rat Clec2d11 was observed during RMCV infection, such that the host protein was lost from the cell surface, replaced by RCTL (although heterodimers of host Clec2d11 and RCTL may exist). BWZ.36 reporter cells expressing a CD3ζ-RCTL fusion receptor responded only to 293T transfectants expressing the rat NKR-P1B receptor (indeed, only the NKR-P1B\textsuperscript{WAG} allele), suggesting a direct interaction between RCTL and the inhibitory NK cell receptor. RCTL expression on infected cells moderately inhibited WAG-strain LAK cell activity, an effect that was abrogated with either R3A8 mAb blockade or infection using an RCMV ΔRCTL-mutant virus [107]. The ΔRCTL-mutant virus exhibited lower splenic and liver titres \textit{in vivo}, but only in rat strains where the RCTL protein interacted strongly with the NKR-P1B allele. Notably, weak recognition of RCTL was also shown by the activating rat NKR-P1A receptor, suggesting that the host and pathogen were evolving and counter-evolving to this innate recognition mechanism. It will be interesting to investigate the roles of similar CMV-encoded genes with lectin-like homology (e.g., r153 in RCMV-Maastricht) in NK cell recognition, as well as whether these CMV immunoevasins are capable of heterodimerizing with host proteins (e.g., rat Clec2d1-11). Notably, as mentioned above, we also recently identified an MCMV-encoded ligand that directly interacts with the
mouse NKR-P1B and NKR-P1C (NK1.1) receptors; however, this decoy ligand lacks homology to the C-type lectin-like fold, instead adopting an Ig-like fold (O.A. Aguilar et al., manuscript submitted).

Induction of LLT1 surface expression has been shown to occur upon infection with several viruses (e.g., EBV, HIV, influenza, HSV) [127, 146]. In addition, LLT1 induction by TLR agonists (e.g., TLR3,4,7,8,9) as well as other immune stimuli (such as TCR/BCR crosslinking, mitogens, CD40, etc.) suggests that LLT1 interaction with human NKR-P1A also regulates immune responses during infection, particularly viral and bacterial infections [127, 146]. In this light, several lectin-like viral ORF (including those from vaccinia and other poxviruses) have been speculated to play roles in the regulation of host immune responses, but none of these has been evaluated to date [118]. It is interesting to speculate that an HCMV-encoded gene product might mediate immune evasion of human NK cells, similar to the RCMV- and MCMV-encoded immunoevasins discussed above.

T Cell Co-Stimulation and Adaptive Immunity

Very few reports to date have examined the effects of the NKR-P1:Clr receptor-ligand system on the adaptive immune response. Ligation of an Ocilrp2 splice isoform (LCL-1; Clr-g) expressed on activated T cells by NKR-P1F on B cells and DC has been reported in mice, leading to enhanced TCR/CD28-mediated T cell proliferation and IL-2 production [130]. However, this situation would not be predicted by the somewhat restricted expression of NKR-P1F mainly in NK and T cells, with splice variants of Clr-g being more broadly expressed [115]. Thus, in light of the recently-identified ‘promiscuous’ NKR-P1F:Clr-c,d,g and NKR-P1G:Clr-d,f,g interactions [84], the reagents used in these studies should be taken with caution regarding specificity (e.g., polyclonal anti-Clr-g Ab could cross-react with other Clr, Ocilrp2-Ig fusion protein should recognize both NKR-P1F/G, NKR-P1F tetramer should recognize all of Clr-c,d,g). Nonetheless, in this study, Clr-g was reported to be induced more rapidly after CD3 crosslinking than other co-stimulatory molecules (e.g., OX40, 4-1BB) [131]. Also, the costimulatory effect of Clr-g was shown to require and synergize with CD28, while OX40 and 4-1BB can act independently of CD28 [131]. In a follow-up study, silencing of Clr-g resulted in
impaired T cell proliferation and IL-2 production, due to an inability of the responding T cells to phosphorylate Lck, reorganize the actin cytoskeleton, form TCR caps, and transduce NFκB signals [130]. Notably, the mouse Clr-b protein has been suggested to possess a TRAF2-interaction motif [186], supporting the possibility of regulation of NFκB signals by the Clr proteins on target cells and APC, but the function of this motif has not been investigated yet.

Several cancers including human glioblastoma, germinal center-derived B-cell non-Hodgkin’s lymphomas, and prostate cancer display high levels of LLT1 protein and mRNA and are thereby able to escape NK recognition via NKR-P1A, but can stimulate NKR-P1A-expressing T cells, thereby promoting their activation and proliferation in some cases [204-206].

**Ribozyme**

A recent publication examined the presence of discontinuous hammerhead ribozymes in mammalian genomes and identified several such self-cleaving RNA motifs within the 3’UTRs of mouse Clec2e (Clr-a), mouse Clec2d (Clr-b), rat Clec2d11 (Clr-b), rat Clec2d10 (Clr-a-like), and rat Clec2d9 (Clr-f-like), along with others in the 3’UTRs of predicted Clec2-like genes in other species [207]. Discontinuous hammerhead ribozymes differ from other ribozymes in that they exist as two fragments (enzyme, substrate) in cis, separated by an insertion of variable length sequence into the stem-1 loop, rather than being contiguous (reviewed in [208-210]). In the mouse Clr-b, mouse Clr-a, and rat Clr-b gene products, this insertion varies from ~250 to ~690 to ~790 nucleotides, respectively, and serves to separate the substrate sequence (upstream) from the enzyme sequence (downstream) [207]. In addition, these ribozyme sequences were shown to be catalytically active (with the mouse Clr-b ribozyme being more active than the Clr-a ribozyme), reducing luciferase reporter expression by ~80% due to autocatalytic reporter RNA degradation [207]. The net effect of these cleavage events would be predicted to separate the mRNA 5’-cap and coding sequence from the poly-A tail of the transcript, leading to enhanced RNA degradation. This suggests another possible level of regulation (post-transcriptional) of the rodent Clr-b and Clr-a gene products, either enhancing mRNA turnover to provide more efficient “missing-self” recognition, or providing a molecular switch to turn on and off expression of the gene products during immune responses.
Interestingly, other closely-related mouse *Clec2* genes contain substrate-like motifs in their 3’UTRs (in the absence of corresponding enzyme motifs), each with similar sequence identity to the substrates in Clr-b and Clr-a [207]. It would be interesting to test whether these sequences can act in *trans* as substrates for either the Clr-b or Clr-a enzyme components, akin to micro-RNA-mediated regulation.

### 1.3 Cytomegaloviruses

#### 1.3.1 Overview

Cytomegaloviruses (CMV) are species-restricted viruses of the *Betaherpesvirinae* subfamily with large, double-stranded DNA (dsDNA) genomes. Their genomes are capable of encoding over 200 open reading frames (ORF) [211]. In humans, ~60-70% of adults in developed countries and up to 100% in developing countries have CMV-specific IgG antibodies, indicating past infection [212]. CMV have broad cellular tropism, yet CMV infections are largely asymptomatic in healthy, immune-competent individuals; however, complications can result in immune-compromised individuals, such as unborn fetuses, newborns of seronegative mothers, transplant recipients, the elderly, the congenitally immunodeficient, and patients with acquired immunodeficiency syndrome (AIDS) [212]. The standard treatment for someone with CMV infection is ganciclovir or valganciclovir. Human CMV (HCMV; Human Herpesvirus-5, HHV5) is capable of inducing immunopathological diseases including retinitis, atherosclerosis, and pneumonitis.

Murine or mouse CMV (MCMV; Murid Herpesvirus-1, MuHV1) has become one of the best-studied models of viral infections of laboratory mice because of the similarities between the diseases caused by HCMV and MCMV. Due to the strict species-restricted tropism of different CMV, HCMV cannot be studied experimentally in animal models. MCMV infection is of particular interest to NK cell and T cell biologists, as NK cells are important in controlling the infection at early time points, and T cell responses have been shown to be important for recovery from MCMV and entry into latency.
1.3.2 Virion Structure and Replication

As a betaherpesvirus family member, CMV virions are spherical and composed of the four classic structural components: the core, capsid, tegument (matrix), and envelope (Fig 1.5) [213]. The core is essentially the dsDNA genome of the virus, which is packaged as a single linear molecule. CMV genomes are larger than those of other herpesviruses, for example the genome of MCMV Smith is ~230 kb [214]. The MCMV genome is divided into three groups of genes or gene families. At the left or 5’-end is the m02 family; at the right or 3’-end is the m145 family; notably, many of these genes code for viral immunoevasins [214]. The middle of the genome contains evolutionarily conserved genes, many of which are essential or important for viral replication and structure. The capsid is icosahedral in shape and composed of 4 proteins [213]. The tegument is a proteinaceous layer between the capsid and the envelope. Tegument proteins function in the disassembly of the virion during entry, assembly of the virion during egress, and to inhibit the host immune response to the infection [213]. The envelope is mostly composed of lipids from the intracellular trans-Golgi network and endocytic membranes of the host cell, and a number of virally encoded glycoproteins, the most significant of which is glycoprotein B (gB), a highly conserved protein found in all herpesviruses that is essential for host cell entry via membrane fusion [215]. Viral entry may also be facilitated by viral contact with integrins and epidermal growth factor receptors (EGFR) on host cells, where the viral envelope fuses with the host cell surface in a pH-independent manner to release the nucleocapsid into the cytoplasm [216]. In some cell types, CMV gains entrance into the cell via receptor-mediated endocytosis, which requires a low pH [217]. Viral nucleocapsids then translocate to the nucleus on cytoskeletal filaments and enter the nucleus through nuclear pores for to facilitate replication, while tegument proteins are released into the cytosol and traffic to sites where they function to modulate the host cell response [213].
Cytomegaloviruses are members of the Betaherpesvirinae subfamily of viruses. They are composed of the envelope, tegument, caspid, and genome. The envelope is a lipid bilayer, composed mostly of glycoproteins from the host obtained during viral entry. The tegument is a proteinaceous matrix that holds the icosahedral nucleocapsid. The genome within the caspid is double-stranded linear DNA, ~200Kb.

MCMV gene expression is tightly temporally regulated. The first genes expressed are the immediate early (IE, or α) genes, then the early (E, β) and finally the late (L, or γ) genes. The IE genes are *ie1* (m123), *ie2* (m128), and *ie3* (M122), where *ie1/3* are driven by the major IE promoter (MIEP) [218]. Interestingly, the MIEP is subject to epigenetic repression by the host, as a means of protection. The MIEP is a common tool for eukaryotic expression vectors since it is constitutive and strong. IE gene products and tegument proteins are important in relieving this repression. The MIEP contains binding sites for many TFs, including ATF/CREB and NFκB [213]. IE gene expression starts immediately after the viral genome enters the nucleus (~2 h.p.i),
and *de novo* protein synthesis is not needed for this process [219]. M122 (IE3) and m123 (IE1) are alternative splice products from the same transcript, while IE2 is completely separate and unique to MCMV (i.e., no human homolog) [220]. Both m123 and M122 function as transcriptional activators of early genes and M122 expression is absolutely required for viral replication [220, 221]. The human homolog for MCMV IE3 is HCMV IE2, which is capable of repressing the MIEP by binding just upstream of the transcriptional start site (TSS) to block transcription, by inducing a state of repressed chromatin [213]. HCMV IE2 modulates transcription driven by RNAPII to create a viral transcriptome via direct interaction with p53, cdk9, cdk7, cycT1, Bcd4, and HDAC-1 and -2 [213]. Host cell RNA polymerase II (RNAPII) and the transcription machinery controls the transcription of the rest of the viral genes, and the translation of viral transcripts relies entirely on host ribosomes [213]. Early phase gene expression is also dependent on the expression of the IE gene products and *de novo* protein synthesis [213]. E phase genes include those required for entry into the L phase, and most immune evasion genes [222]. The L phase of replication takes place ~16 h.p.i and requires DNA synthesis and the expression of early genes [223]. Genes from this phase of infection are mostly structural proteins and proteins involved in virion maturation [213]. DNA replication occurs via circular and/or concatemeric intermediates [224] and begins 8-16 h.p.i. [223]. By 10-12 h.p.i., host genome replication is inhibited by ~95% to promote viral replication.

### 1.3.3 Host Resistance and the Role of NK Cells

Different mouse strains have different levels of susceptibility to MCMV infection, suggesting genetic control of viral infection. For instance, possession of the H-2^k^ MHC-I haplotype confers greater resistance than the H-2^b^ or H-2^d^ haplotypes [225]. Host resistance to MCMV is mediated by the innate immune system, as viral titres are controlled by 3-5 days post infection, before the induction of adaptive T and B cell activities [226]. Several lines of evidence highlight the importance of NK cells in controlling MCMV infection. First, by 24 h.p.i., NK cells acquire an activated phenotype in resistant mouse strains compared to susceptible strains [227]. Depleting NK cells in B6-strain mice causes higher viral titres (up to 1000x), enhanced viral dissemination, and MCMV-induced hepatitis soon after infection [228, 229]. Adoptive transfer of NK cells, but not cytotoxic T cells, from naïve adult mice into lethally infected suckling mice
prevents morbidity [230]. Similar experiments using beige (bg/bg) mice, which have defects in NK-cell mediated cytotoxicity and are highly susceptible to MCMV infection, showed protection against MCMV upon transfer on NK cells from normal WT mice [231].

Certain mouse strains, including B6 mice, have evolved an activating NK cell receptor, Ly49H, specific for the viral protein m157 (MHC-I-like), further highlighting the importance of NK cells in controlling MCMV infection [49]. The m157 immunoevasin is thought to have originally evolved to inhibit NK cells via direct interactions with the inhibitory Ly49C/I receptors in certain mouse strains, prior to host adaptation of Ly49H. Importantly, Ly49H engagement with m157 has a negative impact on the virus-specific CD8⁺ and CD4⁺ T cell responses [232]. This is achieved by the Ly49H-specific elimination of infected DCs expressing m157, combined with lower overall viral titres, which together cause a reduction in T cell priming. Indeed, a lack of Ly49H leads to enhanced viral titres [233, 234], and abrogating the Ly49H:m157 interaction using the Δm157-mutant MCMV virus induces a stronger CD8⁺ T cell response with increased IFN-γ and proinflammatory cytokine production, and preservation of DC function, compared to WT MCMV [235]. This suggests that NK cells are also important regulators of the T cell response to MCMV infection. Indeed, NK cells can modulate the T cell response to MCMV: (i) directly, by releasing or consuming cytokines, or by cytotoxicity towards T cells; and (ii) indirectly, by regulating antigen presenting cell activity.

Observations using mice with distinct MHC-I haplotypes that influence MCMV titres have since been shown to be due to an epistatic interaction between activating NK cell receptors, such as Ly49P/L, host MHC-I molecules, and an MCMV immunoevasin, m04, which escorts some but not all MHC-I alleles to the cell surface [236, 237]. It is thought that this function evolved to maintain interactions with inhibitory Ly49 receptors, such as Ly49A/G2, but do so in a manner that masks or interferes with MHC-I-restricted antigen presentation and T cell responses. However, certain host strains evolutionarily adapted activating receptors that recognize the m04/MHC-I complexes directly.

Another observation that highlights the importance of NK cells during CMV infection is the presence in the MCMV genome of several additional viral immunoevasin genes that target
MHC-independent recognition by NK cells. One example was already mentioned above, the RCTL decoy ligand from RCMV-English, which functionally replaces rat Clr11 (Clec2d11) on the surface of infected cells to inhibit NK cells via the NKR-P1B receptor [107]. Other examples include several non-redundant immunoevasins that prevent the expression of NKG2D ligands at the cell surface to avoid induced-self recognition via NKG2D, including m138, m145, m152, and m155. The m138 protein downregulates expression of MULT-1 and H60 by endocytosing and driving them to lysosomal degradation [238]; m145 targets MULT-1 in the secretory pathway, after the Golgi compartment, although the exact mechanism is not yet clear [239]; m152 targets all five Rae-1 isoforms (as well as MHC-I molecules) [240]; and m155 targets H60 via proteasomal degradation [240].

1.3.4 Interferon Production

Interferons (IFN) are a family of soluble cytokines that limit the ability of viral pathogens to replicate and spread. Three physiologically distinct types of IFN exist: types I, II, and III, which are produced by different cell types, bind unique receptors, and have distinct roles in antiviral and anti-proliferative cell states. Type-I IFNs include IFN-β (produced by most cell types), 14 subtypes of IFN-α (produced by hematopoietic cells, especially plasmacytoid DC), IFN-δ, IFN-ε, IFN-τ, IFN-κ, and IFN-ω (although IFN-δ and IFN-τ have only been described for some species), and these bind to the IFN-α/β receptor (IFNAR1/2) expressed on all cell types [241]. Genes encoding type-I IFN are found on chromosome 4 in mice and chromosome 9 in humans. Type-II IFN consists of IFN-γ, which is important for immune cell activation and is produced primarily by activated NK, NKT and CD8+ T cells, among others. The gene encoding IFN-γ is found on chromosome 10 in mice and on chromosome 12 in humans and it binds to the interferon-γ receptor (IFNGR1/2) [241, 242]. Type-III IFNs include IFN-λ1 (IL-29), IFN-λ2 (IL-28A), IFN-λ3 (IL-28B), and IFN-λ4, which activate signaling pathways similar to those activated by type-I IFN, except that type-III IFN bind to interferon-λ receptor (IFNLRα/IL10Rβ; a.k.a., IL-28Rα/IL-10Rβ) [243].
Type-I IFNs are important antivirals that have three major functions. First, they induce a cell-intrinsic antimicrobial state to limit the spread of pathogens, especially viruses. Second, they activate the innate immune response to promote antigen presentation and NK cell function. Lastly, they activate the adaptive immune response, promoting the development of antigen-specific B and T cells and immunological memory [244].

MCMV infection induces IFN-α/β expression as early as 6 h.p.i. followed by a second wave at 36-48 h.p.i. [245, 246]. The earlier IFN-β induction is derived from stromal cells responding to the primary virus inoculum [246], while the second round of IFN-α induction occurs in response to the second phase of MCMV infection as the virus spreads and is primarily produced by DC [247]. Expression of type-I IFN, especially IFN-β, during infection is very well characterized and follows two general, distinct signal transduction pathways, depending on the cell type [248]. Most cells induce type-I IFN expression by the so-called classical pathway that detects viral components in the cytoplasm and activates interferon regulatory factor 3 (IRF3) and nuclear factor kappa B (NFκB) to turn on IFN-β expression, while plasmacytoid DC (pDC) utilize mainly endosomal TLR7/8/9 to recognize ssRNA and unmethylated CpG DNA motifs, and have been called the natural IFN-α/β producing cells [248].

In the classical pathway, signaling is activated by dsRNA, an intermediate of viral transcription, detected by two non-redundant intracellular RNA helicases, retinoic acid-inducible gene-I (RIG-I) and melanoma differentiation-associated protein 5 (MDA5) [249, 250]. RIG-I and MDA5 interact with the mitochondrial antiviral signaling (MAVS; a.k.a., IPS-1, VISA, Cardif) molecule through caspase-activation and recruitment domains (CARD) [250]. MAVS then interacts with stimulator of interferon genes (STING), another adaptor protein that relays the detection of cytosolic nucleic acids, to activate Tank-binding kinase-1 (TBK1) and IkappaB kinase epsilon (IKKe). These kinases phosphorylate IRF3, which homodimerizes, translocates to the nucleus, recruits the co-activators p300 and CREB-binding protein (CBP), and binds to the IFN-β promoter to activate transcription [248, 251]. RIG-I signaling also stabilizes NFκB-inducing kinase (NIK), activating the non-canonical NFκB signaling pathway, which culminates in the subunit RelA binding to the IFN-β promoter [252, 253]. Interestingly, NFκB also appears
to be important for promoting early IFN-β expression prior to IRF3 activation [252]. This wave of IFN-β production induces the expression of IRF7, which is phosphorylated by IKKe and TBK-1, causing it to form heterodimers with IRF3 and stimulate more type-I IFN production, creating a positive feedback loop [254]. Thus, type-I IFN genes have been subdivided into two groups: immediate-early genes activated in response to virus infection by a protein-synthesis-independent pathway (IFN-β and murine IFN-α4) and delayed-type genes, which include the other IFN-α subtypes, whose expression is dependent on de novo protein synthesis [255]. As such, IRF7 has been called the “master regulator” of IFN gene expression [255-257].

Some hematopoietic cells also can detect dsRNA through endosomal TLR3. TLR3 signaling occurs via Toll/IL-1 receptor domain-containing adaptor inducing IFN-β (TRIF) to directly activate TBK-1 [258]. TRIF also activates NFκB by associating with receptor-interacting protein kinase 1 (RIP1) and TNF receptor-associated factor-6 (TRAF6), causing the degradation of IκBα, the inhibitor of NFκB [259]. This pathway is important because TLR3-expressing cells would not need to be infected to produce type-I IFN, but rather can respond to nucleic acid from inactivated virus particles or dead cells taken up in the endosomal compartment [258].

The TLR pathway used by pDC senses ssRNA (TLR7/8) or unmethylated CpG DNA (TLR9) and signals through the adaptor protein myeloid differentiation primary response-88 (MyD88) [260]. MyD88 complexes with TRAF6, IL-1 receptor-associated kinase 1 (IRAK1), IRAK4, and IRF7, in turn inducing the phosphorylation of IRF7, which stimulates the expression of IFN-α genes [261] (Fig 1.6). pDC are unique in that they express IRF7 without prior IFN-β induction. TRAF6 also activates the protein kinase TGF-β-activated kinase 1 (TAK1), which activates the canonical IKKα/β/γ pathway, allowing NFκB to enter the nucleus. TAK1 also phosphorylates mitogen-activated protein kinase kinase-3 (M KK3) and MKK6, which activate ATF2-c-Jun [259].

Cells also employ a variety of intracellular sensors that detect cytosolic DNA, including viral DNA. DNA-dependent activator of IFN-regulatory factors (DAI) detects double-stranded B
DNA, inducing type-I IFN production through IRF3 and NF-κB activation [262]. RNA polymerase III is capable of acting as a DNA sensor by transcribing AT-rich dsDNA into RNA, which is then sensed by RIG-I [263]. Stimulator of interferon genes (STING) is an intracellular sensor of cytosolic dinucleotides, including bacterial cyclic dinucleotides (such as diguanylate monophosphate [264]), and endogenously synthesized 2’5’-cGAMP (which is synthesized by the endogenous enzyme, cyclic GAMP synthetase, or cGAS). Upon recognition of cytosolic DNA, cGAS generates the second messenger, cGAMP, which activates STING to directly associate with TBK1 and promote activation of IRF3. Absent in melanoma 2 (AIM2) is an inflammasome sensor that upon recognition of cytosolic DNA, recruits the adaptor molecule ASC (PYCARD) which in turn recruits and activates pro-Caspase-1. Finally, interferon-γ-inducible protein-16 (IFI16) also detects viral DNA.

Once in the nucleus, IRF3 dimers (and IRF7, once expressed) bind to the positive regulatory domain (PRD) I and III in the IFN-β promoter. PRD II is bound by NFκB, and PRD IV by ATF-2/c-jun. The binding of these factors to the promoter creates the IFN-β enhanceosome. The enhanceosome then recruits histone acetyltransferases to acetylate nucleosome II to open up the region of the promoter containing the TATA box [265]. Next, CBP and p300 complexes with RNAPII and the general transcription factors (more detail below) and are recruited to the promoter to begin transcription [265].
Figure 1.6. **Type-I IFN production and signaling pathways.** Following viral entry into a cell, endosomal TLRs and cytosolic DNA sensors detect viral DNA. Signaling through TLR-7,8,9 recruits MyD88 to activate type-I IFN gene expression through TBK1/IKKe or IKKγ/IKKa/IKKβ, which phosphorylate various substrates allowing NFkB and IRF3 to translocate to the nucleus where they bind their respective DNA binding sites of the IFNβ promoter. TLR3 recruits TRIF, which interacts with TRAF6 to initiate MAPK signaling cascades resulting the translocation of ATF2 and c-Jun to the IFNβ promoter to create the enhanceosome. Once type-I IFN is produced, IRF7 is expressed, which becomes phosphorylated by IRAK1 and IRAK4 allowing it to translocate to the nucleus as homodimers or heterodimers with IRF3. Type-I IFN is detected by the IFNAR receptor expressed at the cell surface of most cells. This triggers the autophosphorylation of Tyk2 and Jak1, which phosphorylate STAT1 and STAT2. Phosphorylated STAT1 and STAT2 complex with IRF9, which translocates to the nucleus to bind ISRE of interferon stimulated genes.
1.3.5 IFN Signaling

Upon IFN-α/β binding to the IFNAR, the IFNAR1 and IFNAR2 subunits dimerize, leading to the activation of Janus kinase-1 (JAK1) and tyrosine kinase-2 (Tyk2), which are constitutively associated to the subunits. Autophosphorylation of JAK1 is followed by phosphorylation of tyrosine residues in the cytoplasmic tail of IFNAR1 and IFNAR2, which creates a docking site for the signal-transducer and activator of transcription-1/2 (STAT1/2) proteins. JAK1 and Tyk2 phosphorylate tyrosine residues on STAT1 (residue 701) and STAT2 (residue 688), leading to their dimerization via SH2 domains [266]. STAT1/2 dimers associate with IRF9 to form the heterotrimeric complex, interferon-stimulated gene factor-3 (ISGF3), which travels to the nucleus to drive the expression of genes containing IFN-stimulated response elements (ISRE; a.k.a., PRD; IFN consensus sequence, ICS; or IFN regulatory element, IRF-E) [266, 267]) in their promoters [268]. IFN signaling also activates STAT3 and STAT5 [269]. Other STAT complexes besides ISGF-3 form during type-I interferon responses, including homodimers of STAT1, STAT2, STAT3, and STAT5 and several heterodimers [269]. These other STAT complexes bind another DNA element, called the IFN-γ-activated site (GAS) element [269]. Of the known ISG, some have only ISRE or only GAS elements, while others have both, allowing for different combinations of STAT complexes. IFN signaling is a rapid process. Within 10 min of IFN-α exposure, ISG activation is detectable [270]. It’s also a very transient process, with the transcription of ISG returning to baseline levels with 8-15 hours in most cells [271].

The specific roles of many ISG remain to be discovered but, in general, ISG function to impair molecular processes essential for virus replication, by inducing apoptosis, inhibiting transcription, recruiting lymphocytes, and enhancing antigen presentation [272]. Some of the most widely studied ISG include protein kinase-R (PKR), 2'-5'-oligoadenylate synthetases (OAS), and ribonuclease L (RNase L). PKR phosphorylates the α subunit of eukaryotic initiation factor 2 (eIF2α), which blocks both host and viral translation [273]. OAS synthesizes oligoadenylates from ATP upon interaction with dsDNA, which then activates latent RNase L [274]. RNase L then cleaves mRNA and rRNA indiscriminately, halting viral replication [275].
1.3.5.1 IRF Family of Transcription Factors

The IRF family consists of 9 members: IRF1, IRF2, IRF3, IRF4 (Pip/LSIRF/ICSAT), IRF5, IRF6, IRF7, IRF8 (ICSBP), and IRF9 (ISGF3γ/p48) [276]. All family members show homology in their DNA binding domain (DBD), which contains a repeat of five tryptophan residues spaced by 10-18 amino acids [276, 277]. Analysis of the crystal structure suggests that the DBD forms a helix-turn-helix structure that recognizes the ISRE ((A/G)NGAAANNGAAACT). The C-termini of the IRF are different, conferring each member with distinct functions [278].

IRF1 was first identified as a protein that induces the expression of type-I IFN in several cell types during virus infection, at low efficiency, by binding the promoters of IFN-α/β genes [279]. It appears to function as a tumour suppressor gene, as mice deficient in IRF1 are susceptible to transformation by the Ras oncogene [280]. IRF-1−/− mice also have a defect in the development of thymic CD8+ T cells [281], due to reduced expression of transporter associated with antigen processing-1 (TAP-1) and low molecular weight protein-2 (LMP-2), leading to reduced peptide loading of MHC-I molecules, and a subsequent defect in the maturation of MHC-I-restricted T cells [282, 283]. IRF1 also regulates the expression IL-15, an important cytokine for proper NK cell development and survival [276].

IRF2 is an oncogene whose tumourogenicity can be reversed by the action of IRF1 [284]. Although usually a transcriptional repressor, IRF2 can activate transcription of a few promoters including that of Histone-4 and the vascular cell adhesion molecule-1 (VCAM-1) in some cell types [285].

As mentioned above, IRF3 is required for the induction of type-I IFN expression. In uninfected cells, two autoinhibitory domains interact to mask the IRF-association domain (IAD) and DBD to prevent nuclear translocation and DNA binding [286]. Following infection, phosphorylation of IRF3 at the C-terminus relieves the interaction between the autoinhibitory domains, unmasking the IAD and DBD. This allows the formation of IRF3 homodimers that can translocate to the nucleus, associate with CBP/p300, and bind to DNA [286]. IRF3
phosphorylation also results in its degradation via the ubiquitin-proteasome pathway [287], to limit signaling.

IRF4 is one of two IRF with tissue-restricted expression, being expressed in lymphoid cells (T and B cells only) and lacking IFN-inducibility [288, 289]. It associates with the PU.1 transcription factor to bind DNA, hence its alternative name, Pip (PU.1 interacting protein) [288]. In particular, PU.1 and IRF4 heterodimers bind to the ISRE-like λB-site in the Ig-enhancer region [290]. In T cells, IRF4 binds to the promoter of MHC-I to induce transcription, without needing to interact with another transcription factor [289].

IRF5 and IRF6 are structurally related, but little was understood about their function(s) until recently. Although originally identified as a gene involved in the antiviral response and type-I IFN production, IRF5 has very diverse functions in a number of cellular processes and diseases. IRF5 plays a role in cell cycle and apoptosis, microbial infections, and inflammation [291]. In humans, IRF5 is associated systemic lupus erythematosus (SLE), inflammatory bowel disease (IBD), and rheumatoid arthritis (RA) [291]. Much less is known about the functions of IRF6. It is rather unique in that its primary role is developmental in nature, in particular in the tissues that give rise to the head, face, skin, and genitals [292].

IRF7 resides in the cytoplasm in a latent state. Signaling through PRR triggers its phosphorylation, allowing IRF7 to translocate into the nucleus. In the nucleus, it associates with other co-activators to form a complex that binds to promoters to activate transcription, as described above [293].

IRF8 was originally identified as the protein that binds to the ISRE motif in the promoter of the H-2L^d gene [294]. IRF8 is the second IRF family member to have tissue specific expression, being expressed only in lymphocyte and monocyte/macrophage lineage cells [276]. Its expression is induced by IFN-γ, not type-I IFN [294]. IRF8 contains an IAD that allows it to form interactions with other transcription factors. IRF8^-/- mice have several immune defects including failure to produce IL-12 and IFN-γ upon stimulation, promoting susceptibility to infection [295].
Lastly, IRF9 functions as a transcription factor only in association with STAT1 and STAT2. This trimeric complex, called ISGF3, is formed within minutes of type-I IFN treatment and activates the transcription of a large number of genes.

1.4 Transcription and Regulation of Gene Expression

1.4.1 Overview of Transcription

Transcription is the process by which DNA (usually a gene or ncRNA precursor) is converted into RNA (mRNA in the case of Pol II, rRNA for Pol I, and tRNA for Pol III). RNA polymerase II (RNAPII) is a 12-subunit (RPB1-12) enzyme responsible for transcribing DNA into mRNA. The hallmark of RNAPII is the carboxy-terminal domain (CTD) of its largest subunit, RNA polymerase B1 (RPB1). The CTD contains tandem repeats of a heptapeptide with a Y_{1}S_{2}P_{3}T_{4}S_{5}P_{6}S_{7} consensus sequence [296]. The CTD is a binding platform for many factors involved at all stages of transcription and undergoes several important phosphorylation events throughout the process of transcription. For transcription of a protein-coding gene to occur, several events must take place, including de-condensation of the locus, nucleosome remodeling, histone modifications, binding of transcriptional activators and co-activators to enhancers and promoters, and the recruitment of the basal transcription machinery to the core promoter. Successful transcription has 3 phases: initiation, elongation, and termination. Repetition of these phases over a gene determines its expression level.

Transcription initiation involves the correct identification of the transcriptional start site (TSS) and the formation of the pre-initiation complex (PIC) at this site. Examination of the different patterns of transcription initiation reveals two modes of initiation: focused and dispersed. In focused transcription initiation, a single nucleotide or narrow region of a few nucleotides controls initiation [297]. In dispersed transcription initiation, there are several weak start sites over a range of 50-100 nucleotides. In simple organisms, focused transcription dominates, while 70% of vertebrate genes have dispersed promoters, which are typically found in CpG islands [297]. In fact, the general rule of thumb for vertebrates is that focused promoters
are associated with highly regulated genes, while dispersed promoters are typically found in constitutive genes, as it would be easier to regulate transcription from a single TSS, and variations in the expression of a constitutive gene would be minimized by the use of multiple start sites [297]. Focused promoters are typically associated with a TATA box, precisely positioned ~30bp upstream of the TSS [298]. Transcription initiation from dispersed promoters is thought to occur by a different mechanism than focused transcription, as dispersed promoters usually do not contain any of the core elements, which are discussed in more detail below. Rather, dispersed promoters contain several binding sites for Specificity protein-1 (Sp1) and Nuclear transcription factor-Y (NF-Y) [299]. Adding to the complexity of transcription initiation is the existence of bi-directional promoters, where two closely spaced (less than 1 Kb) transcriptional initiation elements in a head-to-head arrangement of RNAPII-driven genes are transcribed on opposite stands of a central DNA region. Bi-directional promoters exist for 10-22% of mammalian genes and the directionality is controlled in a cell-type specific manner [300]. Bioinformatics analysis has revealed that some core promoter elements, such as upstream TFIID recognition elements (BREu), CpG islands, and Sp1 binding sites, are enriched at bi-directional vs uni-directional promoters. Another system of classifying promoters of metazoan genes is based upon the configuration of promoter signals, TSS patterns, nucleosome positions and their epigenetic marks, whereby the function of the associated gene groups promoters into 3 categories: Type-I promoters are often associated with genes that are specifically expressed in terminally differentiated peripheral tissues of an adult. These promoters are focused and often contain a TATA box or other core promoter elements, have low CpG content, less ordered nucleosomes, posses H3K4me3 downstream of the TSS when active, and no RNAPII binding when not active [301]. Conversely, type-II promoters are associated with housekeeping genes, have a dispersed promoter with multiple TSS, and have a single CpG island covering the location where transcription initiates [302]. The TSS is in a nucleosome-free region, flanked by two well-positioned nucleosomes with active histone marks in all tissues and cell types. Finally, type-III promoters are characteristic of genes whose expression is developmentally regulated. They have dispersed TSS in a nucleosome-free region, flanked by nucleosomes, and have larger CpG islands. Developmental genes have several features associated with repression by polycomb
group proteins, including H3K27me3 and H3K4me3 marks. These promoters are responsive to long-range regulation, including distal enhancers [302].

To initiate transcription correctly, RNAPII requires additional factors termed the “basal” transcription factors, which include TFIIA (transcription factor for RNA polymerase-IIA), TFIIIB, TFIID, TFIIE, TFIIF, and TFIIH (Fig.1.7) [297]. TFIID is the key TF involved in recognizing most focused core promoters. TFIID is composed of TATA-binding protein (TBP; see below) and a dozen TBP-associated factors (TAF). TFIIIB can also recognize some promoter motifs (see below) and interacts with TBP to assist in recruitment of RNAPII to the core promoter. TFIIA promotes the binding of TBP to the TATA box. TFIIE, F, and H act afterwards and mediate the unwinding of DNA and the early transcription initiation steps [297]. The assembly of these TF at the core promoter forms the PIC and occurs in a step-wise manner. In the classical model of PIC formation, TFIID (or TBP) binds to the minor groove of the promoter DNA, inducing a 90° bend in the DNA [303]. This interaction is stabilized by TFIIIB and TFIIA, which flank both sides of TFIID. Although TFIIA is not required for basal transcription, it does bind to DNA upstream of the TATA box and to the underside of the TBP saddle, thereby aiding in stable complex formation [303]. TFIIIB is required for RNAPII recruitment to the promoter. TFIIB binds to the promoter on either side of the TATA box (BRE, see below). A complex of RNAPII and TFIIF is recruited to the promoter. TFIIF prevents RNAPII from binding non-specifically with DNA, stabilizes TFIIB, and influences TSS selection [303]. Finally, TFIIE and TFIIH join the complex to form a complete, closed PIC. TFIIE and TFIIH are required for opening promoter DNA. TFIIE binds to RNAPII but not to the complex on the TATA box, and facilitates the recruitment of TFIIH, acting as a bridge between TFIIH and RNAPII [303]. TFIIH contains a DNA-dependent ATPase activity essential for the initiation of transcription [303]. In the presence of ATP, a portion of the dsDNA becomes melted, creating a transcription bubble and an open promoter complex.
After the formation of the PIC at the promoter, the next step is elongation. The transition from initiation to early elongation is controlled by multiple factors and phosphorylation events of the heptad repeats within the CTD of RNAPII. When RNAPII is recruited to a promoter, its CTD is mostly unphosphorylated. TFIIH then phosphorylates serine-5 of the CTD, causing destabilization of the interactions between RNAPII and other components of the PIC, permitting promoter escape and early elongation. 5,6-Dichloro-1-β-D-ribofuranosylbenzimidazole (DRB)-
sensitivity-inducing factor (DSIF) and negative elongation factor (NELF) associate with phosphorylated RNAPII, leading to pausing at the promoter-proximal region. Chromatin immunoprecipitation sequencing (ChIP-Seq) and global run-on sequencing (GRO-Seq) studies have demonstrated that the bulk of promoter-bound RNAPII on most genes is paused 20-50 nt downstream of the TSS [304]. Pausing of RNAPII at promoters may help to maintain genes in an activated or poised state, as NELF knockdown causes nucleosomes to reposition over the nucleosome-free region, decreasing expression of genes located there [304]. Positive transcription elongation factor-b (P-TEFb) phosphorylates serine-2 of the CTD, and DSIF and NELF, resulting in productive elongation by releasing RNAPII from the promoter [305]. Once RNAPII is released from the promoter, productive elongation along the gene ensues, although this transition is not a binary switch; for example, exons, mRNA cleavage, and polyadenylation can all slow down RNAPII [306]. RNAPII transcribes DNA very quickly, and studies have reported anywhere from 1-4 kb per minute [307]. RNAPII pauses again after transcribing through the poly(A) site and before termination [306].

Termination of transcription is the final step in transcription and includes the dephosphorylation of the CTD, dissociation of RNAPII from the 3’-end of the gene, and cleavage of the pre-mRNA. Transcription termination is coupled to 3’-end processing of the nascent transcript, and up to 80 proteins play roles in this step in human cells. The most characterized method of transcription termination in metazoans is polyadenylation-dependent. Two models of termination have been proposed: the so-called “torpedo” model and the allosteric/anti-termination model. In both models, cleavage and polyadenylation specificity factor (CPSF) is recruited to RNAPII once its recognition sequence, AAUAAA, is transcribed [308]. CPSF binding induces RNAPII pausing 18-30 nt after the polyadenylation signal [308], which is typically located 10-35 nt upstream from the mRNA cleavage site. Another factor, cleavage stimulatory factor (CstF) binds the downstream GU-rich processing signal [308]. In the torpedo model, CPSF then binds to CstF, releases RNAPII, and cleaves the nascent pre-mRNA transcript, which is immediately polyadenylated by the poly(A) polymerase [308]. It seems metazoan endonucleolytic cleavage lacks a consensus sequence, but it is preceded immediately by a CA dinucleotide. Cleavage of the pre-mRNA is carried out by the 5’-3’ exonuclease Xrn2,
which degrades the mRNA in competition with ongoing RNAPII elongation [309]. Once Xrn2 catches up to RNAPII, RNAPII is released from the DNA. In the allosteric model, cleavage is not necessary, rather transcription of the poly(A) site triggers a conformational change in the elongation complex, destabilizing the complex, resulting in termination.

1.4.2 RNA Polymerase II Core Promoter Elements

The focused core promoter is the region of DNA that directs correct initiation of transcription by RNAPII, typically −40 to +40 relative to the TSS (+1) [297]. Although many core promoter elements have been identified, none of these elements are universal, rather they are found in only a fraction of core promoters in various combinations, and many promoters lack any of these elements. Core promoters are not simply passive DNA sequences that serve to direct the placement of RNAPII, rather they receive and integrate regulatory inputs and convert them into a precise rate of transcription initiation [302].

1.4.2.1 The TATA Box

The first core promoter element identified in eukaryotic protein-coding genes was the TATA box, in 1979, by comparing the 5’-flanking sequences in several protein-coding genes from flies, mammals, and viruses [310]. These early studies of a few genes suggested that the TATAAA consensus sequences was present in all protein-coding genes and was essential for transcription initiation. As more genes were identified, sequenced, and characterized, the prevalence of TATA boxes declined, and it is now accepted that ~10-15% of core promoters contain TATA boxes with the consensus TATAWAAR, where the upstream T is usually located at −31 or 30 relative to the A_{+1} or G_{+1} [298, 311, 312]. TATA-binding protein (TBP) was identified as the protein responsible for binding to TATA boxes and initiating transcription. The DNA-binding domain of TBP folds into a saddle-like structure consisting of two domains, where the N-terminus contacts the 3’-end of the TATA box and the C-terminus contacts the 5’-end. The two concave portions of the saddle are composed of 5-stranded antiparallel β-sheets, where 8 of the β-strands contact the minor groove of DNA via hydrophobic interactions. TBP also partially unwinds dsDNA by introducing kinks at the 5’- and 3’-ends of the TATA box. Although a consensus TATA box sequences has been defined, TBP is able to bind a wide variety of A/T-rich
sequences and plays only a minor role in determining the orientation of transcription [313]. The TATA box is probably the most highly conserved core element, being found in essentially all plants, animals, and fungi, and TATA-like sequences even exist in protists. The location of the TSS from the TATA box is determined by both TFIIB and RNAPII [310].

1.4.2.2 Initiator

Other early studies examining the promoter sequences from protein-coding genes revealed that many genes utilize an A at the TSS, and a C at position –1, surrounded by pyrimidines (YYANWYY) [297], termed the initiator element (Inr). Inr is probably the most commonly occurring promoter motif [311]. Inr are functionally similar to TATA boxes, but can function independently and are capable of augmenting promoter activity when inserted into a synthetic promoter [310]. When a promoter contains both an Inr and a TATA box, these elements act synergistically when spaced 25-30bp apart, but act independently when separated more than 30 bp [314]. The Inr is recognized by TFIID (via TAF1 and TAF2), RNAPII (in the absence of other basal TF), TFII-I, and YY-1.

1.4.2.3 Downstream Promoter Element

Another promoter element is the downstream promoter element (DPE), which was identified as a motif required for TFIID binding in some TATA-less promoters [310]. This motif is conserved from flies to humans and is usually, but not always, found in TATA-less promoters [315]. DPE act with the Inr, as removal of either motif from a promoter results in loss of transcriptional activity. The core sequence of the DPE is located +28 to +32 relative to the A+1 of the Inr (not necessarily the same site as the TSS). The consensus DPE motif is $A/G^2/C/T$ [315]. The spacing of the DPE with respect to the Inr is critical, as increasing or decreasing the space by even 1 nt results in less TFIID binding and reduced transcriptional activity [315]. The binding of TFIID to DPE/Inr-containing promoters requires TAF6 and TAF9 [310].
1.4.2.4  The Motif Ten Element

The motif ten element (MTE) is conserved from flies to humans and is located immediately upstream from the DPE at +18 to +27 relative to the A\textsuperscript{+1} in the Inr [297]. Like the other motifs, MTE is recognized by TFIID.

1.4.2.5  The TFIIB Recognition Elements

The TFIIB recognition element (BRE) is the only characterized core promoter element recognized by a factor other than TFIID. Studies with eukaryotic TFIIB established the consensus $G/C/G/C/CGCC$, and TFIIB binding is mediated through a helix-turn-helix at the C-terminus of TFIIB [316]. The BRE is located immediately upstream of 10-30\% of TATA box elements (BREu) [316]. A second TFIIB recognition site (the BREd, for downstream TFIIB recognition element) can be found immediately downstream of the TATA box [317].

1.4.2.6  CpG Islands

Approximately 70\% of mammalian genes with regulated expression are also associated with a CpG island. Methylation of CpG dinucleotides results in gene silencing through several mechanisms: (i) methylated cytosines can change binding sites for transcriptional activators, preventing them from binding; (ii) mCpG can act as a recognition sequence for methyl-cytosine binding domain proteins, which recruit corepressor complexes that cause compaction of chromatin; and (iii) methylation increases the affinity of certain sequences for the histone octamer, increasing nucleosome occupancy and stability [318].

1.4.3  Transcriptional Regulation of Gene Expression

There are several levels of transcriptional control of gene expression. As hinted to above, the use of different core promoter elements and different components of the PIC play roles in this regulation. Another level of regulation is the interactions between enhancers and promoters that regulate the activation of specific genes in a precise spatio-temporal manner. Enhancers are DNA binding sites for sequence-specific transcription factors that recruit co-activators. Enhancers are different from promoters in that they can function at a considerable distance from the TSS and...
their function is orientation-independent. Although enhancers function when present at a
distance, enhancer-promoter interactions occur by looping of the DNA to physically bring these
two elements in close proximity. This looping requires the Mediator complex [304]. While
multiple enhancers can interact with multiple promoters, there is specificity between certain
enhancers and promoters, although the mechanisms that determine this specificity remain poorly
understood.

Most transcriptional regulation, however, comes from transcription factors and the DNA
sequences they recognize. In eukaryotes, the DNA sequences recognized by a TF are shorter than
in prokaryotes, thus clustering of sites is often needed to ensure specific recognition [319].
Whether a TF binds to its consensus response element or not is largely governed by different
signal transduction events, depending on the TF [320]. Other factors that determine whether a
given TF will bind to its consensus recognition sequence at the proper time and place include
cooperativity between transcription factors in the context of hetero- or homo-dimer formation or
association of a TF with cofactors; for example, the homodimerization of IRF3 and its
interaction with co-activators p300/CBP discussed above [320]. Requirement for more factors
binding to a region of DNA increases the area and strength of recruitment of the PIC over areas
with few factors.

Gene expression can also be regulated at the transcriptional level by epigenetic changes,
such as methylation of DNA and several histone modifications. Histone modifications can result
in either open or closed chromatin, depending on the modification. These modifications can act
to alter DNA-histone interactions, or regulate transcription in their vicinity by acting as scaffolds
to which other proteins can bind, or by preventing proteins from binding to chromatin. As such,
these modifications are termed activating or repressing, are usually reversible, and are added or
removed by separate enzymes.

1.5 Rationale and Approach

One classical functional hallmark of NK cells is their ability to recognize and eliminate
target cells lacking expression of “self” markers – termed “missing-self” recognition. The
inhibitory interaction between NKR-P1B on NK cells and Clr-b on healthy cells has been shown to mediate missing-self recognition of tumour and virally infected cells, which lose Clr-b expression, in turn resulting in dis-inhibition (activation) of NK cells. While the loss of MHC-I molecules on tumour and infected cells is understood as a mechanism to evade T cell immunity, how or why tumour or infected cells lose Clr-b expression is not clear. Understanding how Clr-b expression is modulated in real-time will aid in our understanding of target cell recognition by NK cells in an MHC-independent system.

Previous studies have shown that Clr-b is highly expressed on many cell types, but frequently downregulated on cells undergoing various pathologies, at both the protein and steady-state transcript levels. Decreased steady-state transcripts suggest that Clr-b may be regulated transcriptionally (at the promoter) or post-transcriptionally (by mRNA decay) or both. This thesis will focus on the role of transcription in regulating Clr-b expression using two different approaches.

In Chapter 3, we used an unbiased, genome-wide lentiviral shRNA screen to identify transcription factors involved in maintaining normal or “healthy-self” levels of Clr-b at the cell surface.

In Chapter 4, we took a hypothesis-driven approach to mechanistically examine the regulated transcriptional control of the Clr-b (Clec2d) locus during primary MCMV infection, which promotes a “missing-self” loss of Clr-b steady-state transcripts and protein. In doing so, an unexpected role for type-I interferons was discovered to operate in uninfected “bystander” cells, in the vicinity of infected cells, whereby Clr-b levels are transcriptionally and functionally-augmented, resulting in a greater dynamic range of self-nonself discrimination between infected and healthy cells.
Chapter 2

Materials & Methods
2 Materials & Methods

2.1 Mice, Cell Lines, and Tissue Culture

WT B6, Stat1−/− and wild-type WT control mice (129S6-strain) were purchased from Taconic Biosciences and handled in accordance with approved animal care protocols at Sunnybrook Research Institute.

Mouse embryonic fibroblasts (MEF) were a generous gift from Dr. Karen Mossman (McMaster University, Hamilton, ON). C1498 and NIH3T3 cells were purchased from the American Type Culture Collection (ATCC) and maintained in culture for less than 30 passages. BWZ.36 cells were obtained from Dr. Nilabh Shastri (University of California, Berkeley, USA). HEK293T, RMA, RMA-S, and EL4 cells were obtained from Dr. David Raulet (University of California, Berkeley, USA). Most cells were cultured in Dulbecco’s Modified Eagle’s Medium-High Glucose (DMEM-HG, HyClone) supplemented with 2 mM GlutaMax (Invitrogen), 100 U/mL penicillin (HyClone), 100 µg/mL streptomycin, 50 µg/mL gentamicin, 110 µg/mL sodium pyruvate (HyClone), 50 µM 2-mercaptoethanol, 10 mM HEPES (HyClone), and 10% FBS (Gibco). MEF cells were grown in Minimum Essential Medium (MEM-α) supplemented with 100 U/mL penicillin, 100 µg/mL streptomycin, 110 µg/mL sodium pyruvate, 2mM GlutaMax, and 15% FBS. MEF deficient in various IFN signaling pathway components were validated by PCR following the Jackson Laboratory protocols.

To generate IFNAR1−/− BWZ.36 reporter cells, cells were transfected with pX458 plasmid containing guide RNA (see below). After 72 h, cells were single-cell sorted by fluorescence-activated cell sorting (FACS) for GFP expression and the absence of surface IFNAR1. After expansion, IFNAR1−/− cells were confirmed for lack of IFNAR1 expression by flow cytometry.

A piggyBac tetracycline/doxycycline-inducible system [321] was modified to replace the β-geo cassette with a puromycin resistance gene (PuroR); this vector was then used to generate stable NIH3T3 transfectants inducibly overexpressing viral genes. Briefly, NIH3T3 cells were
transfected with the modified PB-TET vector containing viral ORF of interest, plus PB transposase and reverse transactivator (rtTA) vectors at a 1:1:1 ratio. Doxycycline (Dox) was added at a concentration of 1.5 µg/mL the next day, then the cells were selected in 2.5 µg/mL puromycin plus 1.5 µg/mL Dox for 5 days and allowed to recover for 2 days in 10% DMEM before being used in experiments.

To generate stable NIH3T3 cell lines expressing luciferase constructs, cells were electroporated using an Amaxa Nucleofector II (see below) with modified pGL4.22 vector that had been linearized with SalI restriction enzyme. After 24 h, stable transfectants were selected with medium containing 2.5 µg/mL puromycin for 5 days. Puromycin was removed for 24 h prior to experiments.

For most treatments, 0.75-1.5x10⁵ cells were seeded in 1-2mL supplemented media and treated the following the day for 24 hours, unless otherwise indicated. Primary splenocytes and bone marrow cells were harvested from Stat1−/− and WT mice, red blood cells were lysed using ACK lysis buffer, and cells were analyzed immediately following 6 h IFN-α4 treatment (10⁵ units). Adult ear fibroblasts (AEF) were generated from minced ear tissue, dissociated, and cultured in 10% supplemented DMEM. Primary AEF cells were treated with or without the STAT1 inhibitors, nifuroxazide or fludarabine (CedarLane Labs), at titrated doses (50µM is shown) during IFN−α4 treatment. All cells were maintained at 37°C, 5% CO₂.

2.2 Chemicals and Viruses

IFN-α4 was kindly provided by Dr. Eleanor Fish (University of Toronto, ON) and used at 10³ U/mL. Dox was purchased at BioShop Canada and dissolved in water. Puromycin was obtained from Life Technologies and dissolved in water.

MCMVSmith was provided by Dr. Andrew Makrigiannis (University of Ottawa, ON) and MCMV-GFP was provided by Dr. Sylvia Vidal (McGill University, QC); both have been described previously [322, 323]. Viruses were passaged in vitro by infecting MEF cells with low viral titres and collecting the supernatant 9-10 days post infection. Plaque assays were used to
determine viral titers, as described [324], without centrifugation during infection. For \textit{in vitro} infections, viral supernatant was introduced to cells in culture at a multiplicity of infection (MOI) of \approx 0.5 PFU/cell (unless otherwise stated) and spin-infected by centrifugation for 30 minutes at 800 xg and 37°C.

\section*{2.3 PCR, Cloning, and Plasmids}

Amplification of PCR products for the Dual Luciferase Reporter assays (DLRA) was performed using primers listed in \textbf{Table 2.1} with the AccuPrime enzyme and B6-strain BAC RP24-384I3 (BacPac Resources), following the manufacturer’s protocol (Life Technologies). The PCR products were excised from gels, purified using a gel extraction kit (Life Technologies) and cloned directly into pcDNA3.1/V5/His/TOPO (PCR products <3 Kb) or TOPOXL (PCR products \geq 3 Kb). The PCR products were confirmed by sequencing (Macrogen Inc., Seoul, South Korea; or TCAG Sequencing Facility, Sick Kids Hospital, Toronto, ON). The PCR products were subcloned into either the pGL3-Basic vector (for transient transfections) or pGL4.22 (for stable transfectants, with the hygromycin resistance cassette replaced with a puromycin resistance gene; Promega) using KpnI and XhoI restriction enzymes (NEB) and T4 ligase (Promega). The mut500bp construct was generated by GeneSOE (splicing by overlap extension) using the AccuPrime enzyme and the indicated primers (\textbf{Table 2.1}). This was achieved by performing a primary amplification of the 2 mutated fragments, followed by gel purification. A secondary amplification was performed using the 2 mutated fragments as template and primers that span the full-length 500bp fragment, which was subsequently cloned and sequenced. The mutated sequence was validated by 2 independent transcription factor search algorithms to be devoid of transcription factor binding sites [325] \url{http://diyhpl.us/~bryan/irc/protocol-online/protocol-cache/TFSEARCH.html}. The pRL-TK plasmid was used as a transfection efficiency control in the DLRA and was a gift from Tom Schrader (Health Canada, Ottawa, ON).

To overexpress IRF3/7/9, their respective coding sequences were PCR amplified from MCMV-infected NIH3T3 cDNA using Q5 enzyme (NEB Biolabs) and the primers listed in \textbf{Table 2.1}. The sequences were ligated into pcDNA3.1 (Life Technologies) and sequenced as
described above. M27 and the immediate early genes \((ie1,2,3)\) were amplified as described above from MCMV-infected NIH3T3 cells and ligated into pIRES2-GFP (Clontech) for transient transfectants and the modified piggyBac Dox-inducible system [321] for stable transfectants.

2.4 Cell Transfections

NIH3T3 and C1498 cells were electroporated using a Nucleofector II (Amaxa). For both, 1 million cells were mixed with 1 or 2 µg of DNA and 100 µl Nucleofector Reagent R (for NIH3T3) or Reagent V (for C1498) before adding the suspension to the cuvette. The mixture was electroporated with program U-030 (NIH3T3) or A-020 (C1498). The cells were plated at the desired density in 6- or 12-well plates with pre-warmed media and allowed to recover overnight.

MEF and BWZ cells were transfected using the Neon transfection system with the 10 µL kit (Invitrogen). For each transfection, \(1x10^5\) cells were resuspended in 10 mL buffer R (MEF) or buffer T (BWZ) with 1 µg DNA and transfected with the following programs: 1350 V, 30 ms, 1 pulse (MEF) or 1400 V, 20 ms, 2 pulses (BWZ).

2.5 Flow Cytometry

Antibodies employed for flow cytometry included: biotinylated anti-mouse Clr-b (4A6), described previously [81], biotinylated anti-mouse IFNAR1 (MAR1-5A3; BioLegend), and anti-mouse NK1.1 APC (PK136, eBioscience). Streptavidin-R-phycoerythrin (SA-PE) or streptavidin-allophycocyanin (SA-APC) were used as secondary reagents (Life Technologies). Cells were stained in 100 µL staining buffer (Hank’s balanced salt solution [HBSS], without phenol red or \(\text{Mg}^{2+}/\text{Ca}^{2+}\), with 0.5% bovine serum albumin [BSA], and 0.03% sodium azide) for 30 min on ice with primary antibody, washed with staining buffer, stained for 30 min on ice with secondary reagent, and washed again. Cells were resuspended in staining buffer with propidium iodide and analyzed using a BD FACSCalibur flow cytometer and FlowJo software (TreeStar). Flow plots were gated on live cells, as determined by forward scatter, side scatter, lack of propidium iodide staining, and GFP (where necessary).
2.6 Gene Expression Analysis

Total RNA was isolated from cells using an RNeasy Plus Mini Kit (Qiagen), and first-strand cDNA was generated from total RNA using SuperScript III Reverse Transcriptase (Life Technologies) plus either the Oligo-dT12-18 or random hexamer primers. For analysis of nascent transcript levels (intrinsic qPCR), the RNA was digested with Amp Grade DNase1 (Life Technologies) prior to cDNA synthesis. Quantitative real-time PCR (qRT-PCR) was performed using gene-specific primers (Table 2.1), SsoFast EvaGreen Supermix (Bio-Rad) or PerfeCTa SYBR Green FastMix (Quanta Biosciences) and a CFX96 cycler (Bio-Rad). Gene-specific primers were designed using PrimerBLAST to have a standardized melting temperature of 60°C, a PCR product size of 90-120 bp, and target specificity. For SsoFast EvaGreen, cycling conditions were 95°C for 30 s, followed by 40 cycles of 95°C for 5 s, then 60°C for 5 s. For PerfeCTa, cycling conditions were 95°C for 3 min, and 40 cycles of 95°C for 10 s, then 60°C for 30 s. The resulting PCR products were confirmed by dissociation curve analysis. For all qRT-PCR experiments, primers targeting 5 independent housekeeping genes were tested for each condition, with the least variable gene(s) being selected as reference gene(s) for those conditions. Analysis was done using CFX Manager software (Bio-Rad) and validated by manual calculations.

2.7 BWZ Reporter Assay

BWZ.36 reporter cells expressing a CD3ζ/NKR-P1B fusion receptor (BWZ.P1B) [81] were used as reporters, with the following modification: the Ifnar1 receptor gene was targeted for null mutants using CRISPR-Cas9 technology. Oligonucleotide pairs 5’-caccg GCT GGT GGC CGG GCC TT-3’, or 5’-aaac AAG GCG CCC CGG CCA CCA GC c-3’, were designed following the Zhang lab protocol (http://crispr.mit.edu) and cloned into the pSpCas9-E2A-EGFP (pX458) plasmid. For cell-based assays, NIH3T3 cells or MEF cells were used as stimulator cells and plated in 3-fold serial dilutions in flat-bottomed 96-well plates. Modified Ifnar1−/− BWZ.P1B reporter cells were added at a density of 5x10⁴ cells/well. Half of the co-cultures were treated with 10³ U/mL IFN-α4, and the other half were left untreated. For a positive control, BWZ cells were stimulated with 10 ng/mL PMA and 0.5 µM ionomycin.
Purified blocking Clr-b mAb (4A6) antibody was added at 10 µg/mL. Co-cultures were incubated overnight at 37°C. Cells were washed with PBS and developed using 100 mL 1x CPRG buffer (90 mg/L chlorophenol-red-β-D-galactopyranoside (Roche), 9 mM MgCl₂, 0.1% NP-40, in PBS) at room temperature. Optical densities were measured using a Varioskan microplate reader (Thermo Scientific) using differential OD₅₉₅⁻₆₅₅ values.

For plate-bound stimulation assays, 96-well flat-bottom high binding plates (Corning) were coated with NK1.1 mAb (PK136; 50 µg/mL) overnight then washed. BWZ reporter cells bearing the P1C⁰⁶ or P1B⁰⁶ fusion receptor were plated as above in 3-fold serial dilutions of IFN-α4 and incubated overnight. Cells were washed and developed as described above.

2.8 Luciferase Bioluminescence and Protein Assays

For transient DLRA, NIH3T3, C1498, or MEF cells were transfected with a 10:1 molar ratio of pGL3:pRL-TK (see above) and plated at the desired density in 6- or 24- well plates. The next day, the media was replaced with media containing the desired treatment or vehicle control for another 24 h. The media was removed and the cells were washed with PBS. Passive Lysis buffer (Promega) was added to the cells for 10 min to release the luciferase enzyme. The lysates (20 µl) were transferred to white 96 well plates (Nunc). The substrates for the luciferase enzymes were added and the plates were read using the following protocol on a Varioskan microplate reader (Thermo Scientific): add 100 µl firefly luciferase substrate, wait 2 s, read luminescence for 10 s, add 100 µl Renilla luciferase substrate, wait 2 s, read luminescence for 10 s. Data were analyzed by dividing the relative luminescent signal from the firefly luciferase by the signal from the Renilla luciferase to control for differences in transfection efficiency and normalized to empty vector control (pGL3).

For luciferase assays using the stable NIH3T3 transfecants, the cells were plated at the desired density and treated the next day with the desired chemical or virus for 24 h. After washing the cells with PBS, the cells were lysed as above. The lysates (20 µl) were transferred to white 96 well plates and the luciferase assay was performed as above except only with the firefly luciferase substrate. To control for cell number differences, the relative luminescent signal was
divided by the total protein for each sample and normalized to empty vector control (pGL4.22). Total protein was determined by protein assay using a BSA standard curve and a colorometric assay using Bio-Rad protein assay dye reagent (Bio-Rad).

2.9 5’ Rapid Amplification of cDNA Ends (RACE)

The transcriptional start site and alternative splice isoforms of Clec2d were determined by 5’-RACE using the SMARTer™ RACE cDNA Amplification Kit from Clontech. First RNA was isolated from numerous cell lines and ex vivo cells using RNeasy Plus Mini Kit (Qiagen). For the generation of RACE-ready cDNA, the manufacturer’s protocol was followed. The resulting cDNA were amplified using Advantage 2 PCR kit (Clontech) and the gene-specific primer and nested gene-specific primer listed in Table 2.1. The gene-specific primers were designed to target the last exon (exon-5), such that near full-length transcripts and splice variants could be detected. PCR products were electrophoresed on a DNA gel, excised and gel purified using a gel extraction kit (Qiagen). Purified PCR products were ligated into pcDNA3.1 and sequenced as described above.

2.10 Chromatin Immunoprecipitation (ChIP)

NIH3T3 cells (10x10^6) were treated with IFN-α4 (10^3 U/mL) or left untreated for 1.5 h. ChIP was performed as described previously [326]. Briefly, cells were cross-linked with 1% formaldehyde at room temperature for 10 min, washed twice with ice-cold PBS, collected in 1 mL of PBS and spun down. Cells were resuspended in 1 mL of lysis buffer (1% SDS, 10 mM EDTA, and 50 mM Tris-HCl (pH 8)) plus protease inhibitors, incubated on ice for 10 min and sonicated to an average size of 500 bp. Chromatin was pre-cleared with 25 µl of Pansorbin (Merck Millipore) at 4°C for 15 min. A 100 µl aliquot of sonicated chromatin was immunoprecipitated (IP) with 2 µg of STAT2 mAb (D9J7L; Cell Signaling Technology) or normal rabbit IgG (Santa Cruz Biotechnology) at 4°C overnight. IP samples were centrifuged at 13,200 rpm and supernatants were incubated with 10 µl of Pansorbin at room temperature for 15 min. Precipitates were washed twice sequentially for 3 min in 1x dialysis buffer (2 mM EDTA, 50 mM Tris-HCl (pH 8), and 0.2% sarkosyl), then IP wash buffer four times (1% Nonidet P-40,
100 mM Tris-HCl (pH 8), 500 mM LiCl, 1% and deoxycholic acid). Samples were extracted twice with 150 µl of elution buffer (1% SDS and 50 mM NaHCO₃), heated at 65°C overnight to reverse cross-linking, then DNA fragments were purified with a QIAEX II Gel Extraction kit (Qiagen). For ChIP analysis, qPCR amplification was performed on the CFX Connect Real-Time PCR Detection System (Bio-Rad). Amplicons were detected using SYBR Green (Invitrogen) and the primers listed in Table 2.1.

Table 2.1 PCR primers used throughout the thesis for cloning and qPCR reactions

<table>
<thead>
<tr>
<th>Gene product</th>
<th>1 Forward Primer (5’-3’)</th>
<th>1 Reverse Primer (5’-3’)</th>
</tr>
</thead>
<tbody>
<tr>
<td>100bp Clec2d promoter</td>
<td><code>GGT ACC GCG TTT CCC AAT GAG ATT TG</code></td>
<td><code>CTC GAG TGG GGA GGT TTC AGT ATT TAG C</code></td>
</tr>
<tr>
<td>200bp Clec2d promoter</td>
<td><code>GGT ACC CAG AGT GTG CTG AAA GAG A</code></td>
<td><code>CTC GAG TGG GGA GGT TTC AGT ATT TAG C</code></td>
</tr>
<tr>
<td>300bp Clec2d promoter</td>
<td><code>GGT ACC CAT TGG AAA TGT AAA TGA GGA</code></td>
<td><code>CTC GAG TGG GGA GGT TTC AGT ATT TAG C</code></td>
</tr>
<tr>
<td>400bp Clec2d promoter</td>
<td><code>GGT ACC CCC AGT ACA GGA GAA C</code></td>
<td><code>CTC GAG TGG GGA GGT TTC AGT ATT TAG C</code></td>
</tr>
<tr>
<td>500bp Clec2d promoter</td>
<td><code>GGT ACC CCC TAT AGG TGG AAC ATC ATT ATG</code></td>
<td><code>CTC GAG TGG GGA GGT TTC AGT ATT TAG C</code></td>
</tr>
<tr>
<td>mut500bp-(SOE)</td>
<td><code>ACC TAG AGA ATC GAA CAA GGA GCA</code></td>
<td><code>CCT TGT TCG ATT CTC TAG GTA GAA CAG CAA</code></td>
</tr>
<tr>
<td>5’UTR-Intron1</td>
<td><code>GGT ACC CGT ATT TGT GGC TCT CAG CTT TC</code></td>
<td><code>CTC GAG CTA AAA ATA TAA ATA TGA GAA CAC CAA TCT C</code></td>
</tr>
<tr>
<td>1 Kb Clec2d promoter</td>
<td><code>GGT ACC TGC AGG CCT GTG ATA AGA AAG</code></td>
<td><code>CTC GAG TGG GGA GGT TTC AGT ATT TAG C</code></td>
</tr>
<tr>
<td>2 Kb Clec2d promoter</td>
<td><code>GGT ACC CGT ATT TGT GGC TCT CAG CTT TC</code></td>
<td><code>CTC GAG TGG GGA GGT TTC AGT ATT TAG C</code></td>
</tr>
<tr>
<td>3 Kb Clec2d promoter</td>
<td><code>GGT ACC CGT ATT TGT GGC TCT CAG CTT TC</code></td>
<td><code>CTC GAG TGG GGA GGT TTC AGT ATT TAG C</code></td>
</tr>
<tr>
<td>4 Kb Clec2d promoter</td>
<td><code>GGT ACC CGT ATT TGT GGC TCT CAG CTT TC</code></td>
<td><code>CTC GAG TGG GGA GGT TTC AGT ATT TAG C</code></td>
</tr>
<tr>
<td>5 Kb Clec2d promoter</td>
<td><code>GGT ACC CGT ATT TGT GGC TCT CAG CTT TC</code></td>
<td><code>CTC GAG TGG GGA GGT TTC AGT ATT TAG C</code></td>
</tr>
<tr>
<td>6 Kb Clec2d promoter</td>
<td><code>GGT ACC CGT ATT TGT GGC TCT CAG CTT TC</code></td>
<td><code>CTC GAG TGG GGA GGT TTC AGT ATT TAG C</code></td>
</tr>
<tr>
<td>7 Kb Clec2d promoter</td>
<td><code>GGT ACC CGT ATT TGT GGC TCT CAG CTT TC</code></td>
<td><code>CTC GAG TGG GGA GGT TTC AGT ATT TAG C</code></td>
</tr>
<tr>
<td>8 Kb Clec2d promoter</td>
<td><code>GGT ACC CGT ATT TGT GGC TCT CAG CTT TC</code></td>
<td><code>CTC GAG TGG GGA GGT TTC AGT ATT TAG C</code></td>
</tr>
<tr>
<td>9 Kb Clec2d promoter</td>
<td><code>GGT ACC CGT ATT TGT GGC TCT CAG CTT TC</code></td>
<td><code>CTC GAG TGG GGA GGT TTC AGT ATT TAG C</code></td>
</tr>
<tr>
<td>10 Kb Clec2d promoter</td>
<td><code>GGT ACC CGT ATT TGT GGC TCT CAG CTT TC</code></td>
<td><code>CTC GAG TGG GGA GGT TTC AGT ATT TAG C</code></td>
</tr>
<tr>
<td>IRFC R 100bp</td>
<td><code>GGT ACC GAG TTT CAG TAT TCC ATG AGA AAT TTT TGT TA</code></td>
<td><code>CTC GAG TGG GGA GGT TTC AGT ATT TAG C</code></td>
</tr>
<tr>
<td>IRFC F 100bp</td>
<td><code>GGT ACC GAA AGA GAA AAG AAA CAC AAG GTT TCC CCA ATG AGA AAT TTT TGT TA</code></td>
<td><code>CTC GAG TGG GGA GGT TTC AGT ATT TAG C</code></td>
</tr>
<tr>
<td>IRFC F 200bp</td>
<td><code>GGT ACC GAA AGA GAA AAG AAA CAC AAG GTT TCC CCA ATG AGA AAT TTT TGT TA</code></td>
<td><code>CTC GAG TGG GGA GGT TTC AGT ATT TAG C</code></td>
</tr>
<tr>
<td>IRFC F mut500bp</td>
<td><code>GGT ACC GAA AGA GAA AAG AAA CAC AAG GTT TCC CCA ATG AGA AAT TTT TGT TA</code></td>
<td><code>CTC GAG TGG GGA GGT TTC AGT ATT TAG C</code></td>
</tr>
<tr>
<td>IRFC R mut500bp</td>
<td><code>GGT ACC GAA AGA GAA AAG AAA CAC AAG GTT TCC CCA ATG AGA AAT TTT TGT TA</code></td>
<td><code>CTC GAG TGG GGA GGT TTC AGT ATT TAG C</code></td>
</tr>
<tr>
<td>Clec2d-q-intron-1</td>
<td><code>TGA ACC AAT GGG AGG GGC AGA GAG</code></td>
<td><code>GCA GAC CCA CAG TGA CGG GCT</code></td>
</tr>
<tr>
<td>Clec2d-q-intron-2</td>
<td><code>CCT TGG GGA ACT GTT CAA GGG AGA</code></td>
<td><code>GCA GAC CCA CAG TGA CGG GCT</code></td>
</tr>
<tr>
<td>Primers</td>
<td>Forward</td>
<td>Reverse</td>
</tr>
<tr>
<td>---------</td>
<td>---------</td>
<td>---------</td>
</tr>
<tr>
<td>Clec2d-q-intron-3</td>
<td>GCA ATG GGC AGG GAT TGG TTT GTC T</td>
<td>GGT CAT GGT GGG AGG CAC TGT</td>
</tr>
<tr>
<td>Clec2d-q-intron-4</td>
<td>TGT GGC TAT GGG AGA TGA AAG GCA</td>
<td>TGT CTC CAT GTG CTT GGG ATG GGA</td>
</tr>
<tr>
<td>Gapdh aPCR</td>
<td>GTG CAG TGC CAG CCT CGT C</td>
<td>GCC ACA CCT CCA TGG AAG GGT</td>
</tr>
<tr>
<td>Hprt qPCR</td>
<td>TCC CAG GTG GAT TAG CGA</td>
<td>TGA AGG GGT GTG TGA TGG CAA CA</td>
</tr>
<tr>
<td>Tbp qPCR</td>
<td>AGA GCC AGC AAC TGC GTT G</td>
<td>CCA TCT CCT TCA TGA CAT CTC GAG C</td>
</tr>
<tr>
<td>Gbp2b-P ChIP</td>
<td>CTC CTC ATT CCT GCA GTG CT</td>
<td>CTG GGA AGC CCA ACT TCT GCA C</td>
</tr>
<tr>
<td>Gbp2b-Ctrl ChIP</td>
<td>CCT CCA GGC AGA ATG CAT AG</td>
<td>CTT GTG AGG AGT TCG CTC TT</td>
</tr>
<tr>
<td>Clec2d-P ChIP</td>
<td>ACC TCA GTA AGG TCA GCT TGA A</td>
<td>CAC AGT AGG CTG GAG CAT GA</td>
</tr>
<tr>
<td>Clec2d-Ctrl ChIP</td>
<td>GAG AGT TAG CTA CCC CAC ACT C</td>
<td>ACC AAG AGG GAG AAG AAA GGA TGG</td>
</tr>
<tr>
<td>Irf3-Flag</td>
<td>GAA TTC ATG GAA ACC CGG AAA CGG C</td>
<td>GGA TCC CTA CTT GTC GTC GTG CTA GTC</td>
</tr>
<tr>
<td>Irf7-Flag</td>
<td>GAA TTC ATG GCT GAA GTG ACC CAC ACT C</td>
<td>GGA TCC CTA CTT GTC GTC GTG CTA GTC</td>
</tr>
<tr>
<td>Irf9-Flag</td>
<td>GAA TTC ATG GCT GAA GTG ACC CAC ACT C</td>
<td>GGA TCC CTA CTT GTC GTC GTG CTA GTC</td>
</tr>
<tr>
<td>Clec2d qPCR</td>
<td>AGC TCC ATG GCA TGG AAG GTG TAT G</td>
<td>AGG GGA GAT GGT TCC GTG CCT TT</td>
</tr>
<tr>
<td>McCP2 qPCR</td>
<td>AGG AGA GAC TGG AGG AAA AGT</td>
<td>CTT AAA CCT CAG TGG CCT GTG TCA C</td>
</tr>
<tr>
<td>Irf7 qPCR</td>
<td>AGG TTC TGC ATG ACA GCC AC</td>
<td>GGG TCC TCT GTA AAC AGC GT</td>
</tr>
<tr>
<td>Satb2 qPCR</td>
<td>TTT GCT CCA TGG ACA GAG CC</td>
<td>TGG AGA AAA ATG TCC CAC AG</td>
</tr>
<tr>
<td>Aat7 qPCR</td>
<td>TAC CAG GGG TCT TGC TCT</td>
<td>TGC CAG GTG TGC CTC ATT TG</td>
</tr>
<tr>
<td>Med1 qPCR</td>
<td>GCC CAG GGT AAT GTT GGG C</td>
<td>GCC TCT CTT AGT CCT CGG TT</td>
</tr>
<tr>
<td>Ie1 CDS</td>
<td>cgc cgc GCG GCC TGG AGG CCG CCC</td>
<td>cgc cgc TCC GAA TCA CTT CTT GCT CTT GT</td>
</tr>
<tr>
<td>Ie2 CDS</td>
<td>cgc cgc GCG GCC TGG AGG CCG CCC</td>
<td>cgc cgc TCC GAA TCA CTT CTT GCT CTT GT</td>
</tr>
<tr>
<td>Ie3 CDS</td>
<td>cgc cgc GCG GCC TGG AGG CCG CCC</td>
<td>cgc cgc TCC GAA TCA CTT CTT GCT CTT GT</td>
</tr>
<tr>
<td>Ie2 qPCR</td>
<td>TCC ACG GGC ATC ACA ACT AC</td>
<td>GTC CCT GTG CTC CAG GAT AC</td>
</tr>
<tr>
<td>Ie3 qPCR</td>
<td>TGG AGG AGA GAC GAG GTC AG</td>
<td>TCC ACC GCT CCC AAG AAG TC</td>
</tr>
<tr>
<td>Ie1 qPCR</td>
<td>TCA GCC ATC AAC TCT GTT ACC AAC</td>
<td>ATC TGA AAG AGC GGT ATA TCA TCT TG</td>
</tr>
<tr>
<td>M27</td>
<td>CGC CGG GCG GCC GCA TGG CGG ACC GGC GCC CCT GTG CTT AA</td>
<td>CGC CGG TCC GAA TCA CAC CGG CCT CTC CAC CAC AAA CTC GQA GC</td>
</tr>
<tr>
<td>5' RACE GSP</td>
<td>N/A</td>
<td>CTT GTG ACT GAT CAG CCC GTT GGT TTT</td>
</tr>
<tr>
<td>5'RACE nested GSP</td>
<td>N/A</td>
<td>TGT GCC ATG CAG AAG GCC TGG CGG</td>
</tr>
</tbody>
</table>

1Underlined sequences represent restriction enzyme sites; lower case indicates extraneous sequence added to facilitate restriction enzyme digestion. 2Identical primers were used to clone into the pIRES2-GFP vector, but using restriction cut sites for EcoRI and BamHI in lieu of NotI and BstBI, respectively.

2.11 The RNAi Consortium Lentiviral shRNA Screen

2.11.1 Whole Genome Primary Screen

To identify proteins that play a role in the expression of Clr-b, we performed a “whole genome” knockdown in NIH3T3 cells using a lentiviral RNAi library [327]. We obtained the 80K mouse pool in collaboration with Dr. Jason Moffat (University of Toronto, Toronto, ON),
which consists of a total of 77,690 short hairpin RNA (shRNA) targeting the entire mouse genome (~5 shRNA per gene). The lentiviral vector used in this system is pLKO.1, which has a puromycin resistance gene under the control of the human phosphoglycerate kinase promoter and an shRNA under the control of the human U6 promoter [327]. The 21 bp stem of the shRNA were designed to target specific transcripts using criteria that have previously been shown to maximize knockdown, while minimizing off-target effects, as well as to ensure that most genes have shRNA that target both the coding sequences (CDS) and 3’-untranslated regions (3’UTR). The generation of the 80K lentiviral shRNA library was performed by the Moffat lab and has been described previously [327].

NIH3T3 cells were seeded onto 10 cm culture dishes and infected with the 80K pool to an MOI of 0.3 (see Relative Viral Titering, below) in triplicate. After 24 h, the medium was replaced with selection medium containing 2.5 µg/mL puromycin and selected for 5 days. Following selection, the replicates were stained with biotinylated Clr-b mAb (4A6) and SA-PE secondary reagent and sorted based on Clr-b expression into 2 populations each: the top 10% of cells expressing the most Clr-b and the bottom 10% of cells expressing the least Clr-b. The cells were washed with PBS and pelleted for DNA isolation.

### 2.11.1.1 Relative Viral Titering

In order to minimize the chance of multiple shRNA entering a given cell, we chose to infect NIH3T3 cells at a relatively low MOI of 0.3. To determine the MOI in NIH3T3 cells, cells were seeded and infected with various dilutions of the stock 80K pool virus plus polybrene (hexadimethrine bromide; 10 µg/mL) and spin-fected at 800 x g for 30 min at 37°C. After 48 h the media on half the cells was replaced with selection media containing 2.5 µg/mL puromycin and the rest received fresh 10% DMEM. Following 2 days of selection, the number of live cells was counted in both the selected and unselected groups using trypan blue dye exclusion. The MOI was determined using the following equation:

\[
\text{MOI} = \frac{\text{[-(In) non-responders]}}{}
\]

where non-responders = [1-(# live cells in +Puro/# live cells in –Puro)].
2.11.1.2 Genomic DNA Isolation

The Qiagen DNA BloodAmp Maxi kit (Qiagen) was used to extract genomic DNA following the manufacturer’s protocol. The DNA was further purified with 0.2M NaCl (final concentration) and 2 volumes of -20°C 100% ethanol to precipitate the DNA. The precipitated DNA was collected by centrifugation at 13,000 rpm for 15 min at 4°C. The pellet was washed with -20°C 70% ethanol and centrifuged again. Following complete removal of the supernatant, the pellet was air dried for 5 min and resuspended in 10 mM Tris-HCl pH 7.5 to a final concentration of 450 ng/µL.

2.11.1.3 Microarray

Further preparation of the samples and the microarray analysis were conducted by Dr. Troy Ketela in the lab of Dr. Jason Moffat, as previously described [328]. The shRNA hairpins were amplified from genomic DNA in a PCR master mix containing 160 µl of H₂O, 24 µl dNTP (2.5 µM of each), 30 µl 10x ExTaq buffer, 30 µl 10 mM LKO-6738 biotinylated 5’ primer (5’-Bio-AATGGACTATCATATGCTTACCGTAACCTTGAA), 30 µl 10 mM LKO-6863 3’ primer (TGTGGATGAATACTGCCATTTGTCTCGAGGTC), and 4.5 µl TaKaRa Ex Taq (5 U/mL). The mastermix was split into 6 tubes containing 45 µl of master mix and 5 µl of genomic DNA. Amplification was carried out under the following conditions: 95°C for 5 min, and 35 cycles of 95°C for 30 s; 50°C for 30 s; and 72°C for 1 min, followed by a final extension of 72°C for 10 min. The 6 reactions were pooled together and subjected to another round of PCR with another 45 µl of master mix and the following conditions: 95°C for 7 min, 55°C for 2 min, 72°C for 1 hr and then pooled together.

The resulting amplicons were digested into half-hairpins with XhoI and PCR purified with QIAquick PCR Purification Kit (Qiagen) and QIAvac 24 Plus Vacuum Manifold (Qiagen) according to the manufacturer’s direction. To prepare the amplicons for hybridization to the array, 35 µl of purified amplicons were added to 4.5 µl of 100 µM blocking primer (LKO_6810_block 5’-GTCTTTTCACAAAGATATATAAGCAGAAATCGAAATA-3’) and 38.5 µl of H₂O and heated to 99°C for 5 min and 45°C for 5 min. The heated DNA was added to a hybridization cocktail consisting of 6 µl fragmentation buffer, 15 µl GeneChip
Control Oligonucleotide B2 (3 nM), 15 µl GeneChip 20x Eukaryotic Hybridization Controls (*bioB, bioC, bioC, cre*), 3 µl herring sperm DNA, 3 µl BSA (50 mg/mL), 150 µl 2x Hybridization buffer, and 30 µl DMSO and heated to 99°C for 5 min and 45°C for 5 min. The array was wetted with 1x Hybridization Buffer in a Hybridization Oven at 40°C for 10 min with rotation at 60 rpm. The buffer was removed from the array and replaced with the hybridization cocktail, placed in the Hybridization Oven at 40°C and 60 rpm for 16 h. Following hybridization, the cocktail was removed and the probe array was washed with non-stringent wash buffer. GeneChip Operating Software (GCOS) was used to run the microarray wash and stain protocol (FlexGE_WS2v5 Biao30) on an Affymetrix fluidics station. The array was stained with a labeling mix consisting of 1x MES stain buffer, 2 mg/mL BSA, and 10 µg/mL SA-PE. The array was then scanned and binding of the fluorophore-labeled streptavidin to the biotinylated shRNA amplicons was detected. Signal from the gene modulation array platform (GMAP) microarray [329] was extracted using Affymetrix Power Tools v.1.12.0 and the library file GMAP-Uts520601.cdf by Jason Moffat’s lab, who have developed R scripts to aid in the normalization of signal intensities, generation of quality control plots, and summarization of the replicate probe intensities. A full set of instructions can be found at [http://chemogenomics.med.utoronto.ca/supplemental/gmap/](http://chemogenomics.med.utoronto.ca/supplemental/gmap/) [328]. GC-background correction for non-specific probe binding was performed with APT and normalization of replicate arrays was performed with the Bioconductor `affy` package in R.

### 2.11.1.4 Microarray Data Analysis

An excel file containing the normalized hybridization intensity of each shRNA expressed as a log_2 value was obtained from Dr. Jason Moffat’s lab. Preliminary analysis of this screen included calculating the mean hybridization intensity and the standard deviation for each shRNA from the Clr-b^{high} and Clr-b^{low} replicates, as well as the ratios of the Clr-b^{high} over Clr-b^{low} and vice versa in log_2. The data was also transformed from log_2 to linear values for several calculations including the mean hybridization intensity, and the standard deviation as mentioned above, and the fold change. For simple analysis of statistical significance, we employed a student’s t-test (2-tailed, type 3) to establish which shRNA were overrepresented in either the Clr-b^{high} or Clr-b^{low} groups and a Z-score (or standard score), to see how many standard
deviations an shRNA was from the mean. From these simple analyses, we created a list of top hits of shRNA overrepresented in either the Clr-b\textsuperscript{high} or Clr-b\textsuperscript{low} groups. Since the focus of this thesis is on the regulation of Clr-b at the level of transcript production, we filtered the data set to only look at those shRNA that targeted TF genes using the list of genes with nucleic acid binding transcription factor activity from Gene Ontology (GO:0001071 and children).

2.11.2 The 1K Secondary Screen

To more quickly validate candidate genes and to increase the odds of having successful validations downstream, we opted to perform a secondary, iterative screen of \(~1K\) shRNA, which we felt was much more manageable than the \(~80K\) whole genome set. Since the secondary screen was designed with this thesis in mind, we decided that half of the shRNA selected should target TF (See Table 3.2 for the criteria used to select these TF-targeting shRNA). The other half of the shRNA were chosen to target non-TF genes to found new projects for future lab members (See Table 3.2 for the criteria used to select these shRNA). Finally, as controls, we included several non-targeting shRNA including those targeting LacZ, red fluorescent protein (RFP), GFP, etc, as well as shRNA targeting Clr-b itself and Irf7 and MeCP2, which validated from the primary screen.

2.11.2.1 Creation of the 1K Lentiviral Library

We obtained glycerol stocks of bacteria harbouring each shRNA of interest in 96 well plates from Dr. Jason Moffat’s lab. We inoculated the glycerol stocks into deep 96-well growth plates containing 1.2 mL of TB media containing 100 µg/mL carbenicillin, sealed them with a gas-permeable seal, and incubated them at 37°C, 300 rpm overnight. We then combined 400 µL from each well to make 1 large culture, from which we isolated plasmids using QIAamp DNA Blood Maxi Kit (Qiagen). The remaining cultures were spun down and frozen as pellets stored at -80°C.

To generate the lentivirus, HEK293T cells were seeded onto 10cm dishes in 10% DMEM without pen/strep and transfected the next day with pLKO.1 1K pool, packaging plasmid and envelop plasmid (10:10:1). After 18 h, the medium was removed and replaced with 15% DMEM for 24 h. The medium containing the virus was collected and replaced with 15% dMEM for
another 24 h. The medium containing the virus was collected again and spun down at 1250 rpm for 5 min to remove HEK293T cells. The virus was titred and used to infected NIH3T3 cells as described above. Remaining virus was stored at -80°C.

The NIH3T3 infected with the 1K library were selected, maintained, sorted, and genomic DNA isolated, as described above for the primary screen.

2.11.2.2 Deep Sequencing

We opted for deep sequencing of the genomic DNA for this screen, since the screen was much less complex than the original. We also performed the PCR amplifications ourselves, in duplicate. We created a PCR mix of 10X reaction buffer, 2.5 mM dNTP, DMSO, BC oligo mix (version 4.0, see Table 2.2, 20 µM each, 6 pairs of primers for multiplexing), 5 µg 1K pool DNA, exTaq enzyme and water. The amplification was carried out using the following cycling conditions: 95°C 5 min and 32 cycles of 94°C for 30 s; 65°C for 30 s; 72°C for 20 s and a final extension of 72°C for 5 min. The amplicons were PCR purified using the QIAquick PCR Purification Kit (Qiagen). We ran 2 µl of each sample on a 2% agarose gel and performed densitometry on each product. Equal amounts of the samples were pooled together for multiplexed sequencing reactions. Each sample was loaded into a precast NOVEX 6% TBE acrylamide gel and run at 180V for 35 min. The gel was stained in 1X TBE and ethidium bromide for 10 min and rinsed in 1x TBE. The gel was quickly photographed to minimize nicking of the DNA and the 220bp band was excised. The gel slices were transferred to a 500 µl tube with a hole poked in the bottom. The gel was passed through the hole to shred it by centrifugation at 13,000 xg for 1 min and heated in 140 µl Qiagen EB for 16 h with rotation. After a 1 min centrifugation at 13,000 xg, the supernatant was removed and the DNA was cleaned and concentrated using the QIAquick PCR Purification Kit (Qiagen) according to the manufacturer’s protocol in 12 µl EB and quantified on a Nanodrop.

Dax Torti at the Donnelly Centre for Cellular and Biomolecular Research sequencing facility (Toronto, On) performed the sequencing reaction using the primers listed in Table 2.2.
Table 2.2: Version 4.0 barcoded primers for multiplexing used to generate Illumina sequencing-compatible DNA and the primers used for the sequencing reactions.

<table>
<thead>
<tr>
<th>Primer Name</th>
<th>Primer Sequence (5’-3’)</th>
</tr>
</thead>
<tbody>
<tr>
<td>V4-6BC-1</td>
<td>CAAGCAGAAGACGGCATAACGAGATCGTGAAGAAATTTTCGATTTCTTGCTTTATATATCTTGTGGA</td>
</tr>
<tr>
<td>V4-6BC-2</td>
<td>CAAGCAGAAGACGGCATAACGAGATACATCGGAAAGTATTTTCGATTTCTTGCTTTATATATCTTGTGGA</td>
</tr>
<tr>
<td>V4-6BC-3</td>
<td>CAAGCAGAAGACGGCATAACGAGATGCCATTGAAGAAATTTTCGATTTCTTGCTTTATATATCTTGTGGA</td>
</tr>
<tr>
<td>V4-6BC-4</td>
<td>CAAGCAGAAGACGGCATAACGAGATTTGTCGAAAAGTTTTTCGATTTCTTGCTTTATATATCTTGTGGA</td>
</tr>
<tr>
<td>V4-6BC-5</td>
<td>CAAGCAGAAGACGGCATAACGAGATCACTGTGAAAGATATTTTCGATTTCTTGCTTTATATATCTTGTGGA</td>
</tr>
<tr>
<td>V4-6BC-6</td>
<td>CAAGCAGAAGACGGCATAACGAGATATTGCGAAAAGATATTTTCGATTTCTTGCTTTATATATCTTGTGGA</td>
</tr>
<tr>
<td>V4-6BC-7</td>
<td>CAAGCAGAAGACGGCATAACGAGATGATCGGAAAAGTATTTTCGATTTCTTGCTTTATATATCTTGTGGA</td>
</tr>
<tr>
<td>V4-6BC-8</td>
<td>CAAGCAGAAGACGGCATAACGAGATTAAGCTGAAAGATATTTTCGATTTCTTGCTTTATATATCTTGTGGA</td>
</tr>
<tr>
<td>V4-6BC-9</td>
<td>CAAGCAGAAGACGGCATAACGAGATCTGATCAGAAATTTTCGATTTCTTGCTTTATATATCTTGTGGA</td>
</tr>
<tr>
<td>V4-6BC-10</td>
<td>CAAGCAGAAGACGGCATAACGAGATGACATGAAAGATATTTTCGATTTCTTGCTTTATATATCTTGTGGA</td>
</tr>
<tr>
<td>V4-6BC-11</td>
<td>CAAGCAGAAGACGGCATAACGAGATGACATGAAAGATATTTTCGATTTCTTGCTTTATATATCTTGTGGA</td>
</tr>
<tr>
<td>V4-6BC-12</td>
<td>CAAGCAGAAGACGGCATAACGAGATGACATGAAAGATATTTTCGATTTCTTGCTTTATATATCTTGTGGA</td>
</tr>
<tr>
<td>V4-6BC-13</td>
<td>CAAGCAGAAGACGGCATAACGAGATGACATGAAAGATATTTTCGATTTCTTGCTTTATATATCTTGTGGA</td>
</tr>
<tr>
<td>V4-6BC-14</td>
<td>CAAGCAGAAGACGGCATAACGAGATGACATGAAAGATATTTTCGATTTCTTGCTTTATATATCTTGTGGA</td>
</tr>
<tr>
<td>V4-6BC-15</td>
<td>CAAGCAGAAGACGGCATAACGAGATGACATGAAAGATATTTTCGATTTCTTGCTTTATATATCTTGTGGA</td>
</tr>
<tr>
<td>V4-6BC-16</td>
<td>CAAGCAGAAGACGGCATAACGAGATGACATGAAAGATATTTTCGATTTCTTGCTTTATATATCTTGTGGA</td>
</tr>
<tr>
<td>V4-6BC-19</td>
<td>CAAGCAGAAGACGGCATAACGAGATGACATGAAAGATATTTTCGATTTCTTGCTTTATATATCTTGTGGA</td>
</tr>
<tr>
<td>V4UN1</td>
<td>AATGATACGGCGACCACCGAGATCTCTCTCTCTCTCTTTTTAAAAATTGTTGAATATATCTGCCATTIG</td>
</tr>
<tr>
<td>V4shRNAseq</td>
<td>CTGCCATTGTCGAGGTCGAGAATTCAAAAA</td>
</tr>
<tr>
<td>V4BCseq</td>
<td>TCCACAAGATATATAAAGCCGAAATAGATACCTTTC</td>
</tr>
</tbody>
</table>

Italics are bases that comprise the Illumina adapter sequences. Bold, underlined bases are the barcode sequences.
2.11.2.3 Sequencing Analysis

Dr. Kevin Brown in Dr. Moffat’s lab deconvoluted the sequencing data using PoolQ and provided us with an excel spreadsheet consisting of the raw number of reads per shRNA in the 2 populations and the normalized number of reads/million reads.

Analysis of the sequencing data was performed essentially as for the microarray data; the average fold enrichment of each shRNA in both populations was calculated, as well as the standard deviation. We then performed Student’s t-test to determine which shRNA were significantly overrepresented in one population or the other. The fold enrichment of each shRNA from the secondary screen was compared to that obtained in the original screen. We opted to mostly validate shRNA that came out in the same population in both screens. For reference, Oscar A. Aguilar from our lab had recently performed RNA sequencing experiment on NIH3T3 cells and NIH3T3 cells infected with either mouse cytomegalovirus (MCMV) or RCMV. Using these data, and to avoid validating shRNA designed to target those genes not expressed in the NIH3T3 cells, we opted to validate only shRNA designed to target genes with ≥ 3 reads/million reads from the RNA-seq experiment. A list of which shRNA to validate was created simply based on the fold enrichment, p ≤ 0.05, and eliminating those shRNA that did not validate in the second screen, and those targeting genes not expressed in NIH3T3 cells (Table 3.3).

2.11.3 Validation of Individual shRNA

To validate individual shRNA, lentiviruses were generated for each hairpin of interest following the protocol above. Since many hairpins were to be validated, the individual viruses were not titred. Rather, we infected 50K NIH3T3 cells with 340 µl of virus with 10 µg/mL polybrene and selected for 5 days in puromycin, as per the protocol above, along with parental (uninfected, unselected NIH3T3) and control shRNA (either an shRNA targeting LacZ or the entire 80K viral library). The cells were subjected to flow cytometric analysis for Clr-b expression as described above. shRNA that resulted in altered Clr-b expression relative to controls were chosen for further validation by qPCR, both for Clr-b and the gene targeted by the shRNA, as described above. Ultimately, to show direct effect of the shRNA on the correct target, the proteins were analyzed by Western blot.
2.11.3.1 Western Blot

Cells were pelleted at 1500 RPM for 5 min and washed twice with ice cold PBS. The pellets were resuspended in 50 µl of RIPA Buffer (50 mM Tris HCl pH 8.0, 150 mM NaCl, 0.02% sodium azide, NP-40, 0.1% SDS, 0.5% sodium deoxycholate) + 1X Complete Mini protease inhibitor cocktail tablets (Roche Diagnostics) and left on ice for 1 h. Lysates were spun down at 12,000 xg for 10 min at 4°C and the supernatants were transferred to new tubes. Protein concentrations were determined as described above. Normalized protein amounts were mixed with 5X protein loading buffer (0.29M Tris HCl pH 6.8, 8.57% SDS w/v, 30% glycerol, 4.2% β-mercaptoethanol, 0.2 mg bromophenol blue; 1X final) and boiled at 100°C for 5 min. Lysates and kaleidoscope ladder (Bio-Rad) were loaded in 8% or 10% SDS polyacrylamide gels in a Mini-Protean Tetra System with Running Buffer (with 1% SDS) at 0.3 mA until a desired band migration was achieved. Blots were transferred at 80V for 90-120 min onto an activated Immobilon PVDF membrane (Millipore). Membranes were blocked with 5% milk or 5% BSA in TBS-T for 45 min at room temperature and blotted with primary antibody overnight at 4°C in either 5% BSA or 5% milk depending on manufacturers’ guidelines. Anti-MeCP2 (rabbit pAb ab2828) and anti-IRF7 (rabbit mAb EPR4718) antibodies were purchased from Santa Cruz Biotechnology. Anti-STAT2 (rabbit mAb D9J7L) and anti-TBP (rabbit mAb D5G7Y) were purchased from Cell Signaling Technology, anti-CBP (rabbit pAb poly6064) was purchased from BioLegend, and β-Actin (mouse mAb AC-15) was purchased from Sigma. The blots were washed 3 times for 5 min in TBS-T before incubation with secondary antibody for 1 h at room temperature in 5% milk in TBS-T. Anti-rabbit and anti-mouse HRP were purchased from Cell Signaling Technology. After washing, the blots were incubated with Immobilon Western HRP chemiluminescent reagent (Millipore) and visualized on a MicroChemi 4.2 imager using DNR’s GelCapture image acquisition software and the GelQuant software (DNR Bio-Imaging Systems).

2.12 Statistical Analysis

Data were analyzed using GraphPad Prism 7 using paired Student’s t-test or ANOVA. All graphs show mean values ± SEM; *p<0.05, **p<0.01, ***p<0.001, ****p<0.0001. Only significant differences are annotated.
Chapter 3

Identification of novel transcription factors involved in the expression of Clr-b using an RNAi library

Christina L. Kirkham*, Kuan Lun Chu*, Troy Ketela§, Jason H. Fine*, Dax Torti, Jason Moffat§, David S.J. Allan*, and James R. Carlyle*

*Department of Immunology, University of Toronto, Sunnybrook Research Institute, 2075 Bayview Ave., Toronto, ON, M4N 3M5, Canada.

§Department of Molecular Genetics, University of Toronto, The Donnelly Centre, 160 College St., Toronto, ON, M5S 3E1, Canada.

All experiments were performed by C. Kirkham.
The 80K lentiviral pool was provided by Dr. J. Moffat.
Dr. T. Ketela contributed advice and expertise regarding the TRC library.
The microarray and sequencing experiments were performed by Dax Torti at the CCBR.
Additional replicates were contributed by Kuan Lun Chu.

This work is unpublished
3 Chapter 3

3.1 Introduction

Hypothesis-driven research is usually considered to be the cornerstone for successful scientific projects. However, many important discoveries have been made fortuitously, including the discoveries of penicillin and Teflon. These discoveries, and others, suggest that unbiased approaches to research in addition to hypothesis-driven research are both useful. Unbiased approaches have the advantage of finding novel factors that play roles in complex biological systems that might not be expected based on a limited current understanding.

While most healthy hematopoietic and some non-hematopoietic cells (e.g., fibroblasts) express high levels of Clr-b at the cell surface, Clr-b is frequently downregulated on tumour cell lines, virus-infected, transformed, and stressed cells. Specific mechanisms controlling Clr-b expression have not been uncovered; however, to date, transcriptional, post-transcriptional, and post-translational pathways have been implicated [40, 81, 107, 145, 330]. Therefore, with limited pathways and factors to investigate, an unbiased approach was applied to help identify factors involved in the expression of Clr-b.

Recently, several screens have been developed to meet the need for unbiased approaches to research, including The RNAi Consortium’s (TRC) whole genome lentiviral libraries, for either human or mouse, which are readily available at the University of Toronto [327]. We chose to employ the mouse library, consisting of 77,690 hairpins targeting about 15,538 mouse genes [327]. The pooled, entire library is advantageous over individual shRNA, as it allows the simultaneous analysis of all mouse genes, rather than the labour-intensive analysis of single shRNA or single genes, one at a time.

3.1.1 Screen Design

The whole genome lentiviral library (referred to as the 80K pool) was provided by Dr. Jason Moffat and has been partially validated [327-329]. Since factors influencing the normal or “healthy-self” expression of Clr-b were the focus this study, surface Clr-b expression was
measured using anti-Clr-b mAb (4A6) by flow cytometry for validating positive shRNA candidates (hits) from the screen. Since NIH3T3 cells (H-2^d) express high levels of Clr-b, grow readily and easily in the lab, and are easily infected by viruses, we chose NIH3T3 cells for our study. Depletion of genes important in maintaining high surface Clr-b levels by shRNA would be expected to cause a decrease in Clr-b expression. Conversely, depletion of genes that may cause Clr-b downregulation would be expected to be enriched in the Clr-b^high cells. Genes that play no role in Clr-b expression, likely most of the genome, would be expected to be equally represented in both groups of Clr-b-expressing cells.

To ensure adequate detection of all shRNA, the screen was designed such that each shRNA should be present in the whole-genome NIH3T3 library approximately 125 times, or 125X over-represented. To minimize confounding effects of multiple shRNA in the same cell, a low multiplicity of infection (MOI) of 0.3 was selected.

While the lentiviral library was designed with multiple shRNA targeting each gene, this does not guarantee that all or any of the shRNA cause knockdown the desired genes of interest. Off-target effects are major concerns using screens of this nature, especially for gene families where individual family members have high sequence similarity; it is very possible to target multiple family members or the incorrect family member. Another potential concern is the redundancy of several factors; such that knockdown of only one factor would not be enough to see any phenotype with respect to the experimental target, Clr-b in this case. High-throughput screen approaches also have experimental noise arising from both technical and biological artifacts. Although this can be minimized to some degree with replicates, the cost associated with performing more than 2 or 3 replicates represents an issue. The concept of RNAi itself is also a potential disadvantage, as knockdown may not be sufficient to see a phenotype and knockout of gene function may be more ideal. Nonetheless, knockdown studies are particularly useful in cases where genes may be required for cell survival.
3.2 Results

3.2.1 Transduction of NIH3T3 cells with the 80K lentiviral shRNA pool

To identify novel genes involved in the normal, healthy expression of Clr-b, we used the shRNA library developed by Dr. Jason Moffatt [327]. We performed the screen in NIH3T3 cells because they express high “healthy-self” levels of Clr-b (as opposed to most transformed hematopoietic cell lines), similar to normal ex vivo primary murine embryonic fibroblast (MEF) cells, modulate their expression of Clr-b in response to a variety of stimuli, are easily infected by viruses such as MCMV, and are readily available in large scalable numbers (as opposed to primary MEF cells or many other cell types requiring cytokines).

Since the 80K library is a pool of thousands of individual shRNA, a multiplicity of infection (MOI) of 0.3 selected to ensure that only ~1 lentiviral shRNA vector would be transduced per cell, to minimize the chances of multiple shRNA transducing individual cells. We conducted a lentiviral vector titre assay of NIH3T3 cells to determine the volume of lentiviral supernatant needed to obtain an MOI of 0.3 (Fig. 3.1). This experiment indicated that per 2x10^6 NIH3T3 cells, a total of 88.6 µL of lentiviral supernatant would result in an MOI of 0.3.
Figure 3.1. Determination of the multiplicity of infection (MOI) of the 80K pool for the whole genome knockdown in NIH3T3 cells. 2x10^6 cells were transduced with the indicated volume of virus in 30 mL media in a conical tube and 10 µg/mL polybrene with gentle shaking for 20 min. Each transduction was split into 2 plates. After 24 h the media was replaced on half the plates with selection media containing 2.5 µg/mL puromycin. The number of live cells in each plate was counted after 2 days of selection and the MOI was calculated using the given equation.

\[ y = 4.5x - 0.0987 \]
\[ R^2 = 0.99107 \]
3.2.2 Whole-genome high-throughput screen of NIH3T3 cells to determine host factors involved in normal Clr-b expression

We infected NIH3T3 cells at an MOI of 0.3 in triplicate with a ~125X over-representation of the entire library (Fig. 3.2). After puromycin selection, the NIH3T3 transductants were stained with Clr-b mAb (4A6) to monitor Clr-b cell surface expression (Fig. 3.3A). The cells were sorted based upon Clr-b expression into 2 groups: the top 10% (Clr-b<sup>high</sup>) and the bottom 10% (Clr-b<sup>low</sup>) (Fig. 3.3A). Genomic DNA was extracted from both populations after sorting for PCR-based amplification of the shRNA using barcode-specific primers. The PCR products were analyzed by microarray with assistance from Dr. Jason Moffatt’s lab, as described previously [329].
Figure 3.2 Schematic representation of the workflow of the NIH3T3 lentiviral screen. NIH3T3 cells were transduced with the 80K lentiviral library at a multiplicity of infection of 0.3. Untransduced cells were removed by puromycin selection. NIH3T3 cells were stained with the anti-Clr-b antibody 4A6 and sorted into Clr-b$^{\text{high}}$ (top 10%) and Clr-b$^{\text{low}}$ (bottom 10%). Following isolation of genomic DNA and PCR amplification of incorporated shRNAs, microarray analysis was performed to characterize the abundance of each shRNA in both cell populations.
3.2.2.1 Screen analysis and results

The relative intensity of each shRNA in both populations was obtained, on a Log2 scale (Fig. 3.3B). We had hypothesized that the majority of the genome would not be involved in Clr-b regulation, thus most shRNA were expected to be equally represented in both populations. shRNA overrepresented in the Clr-b\textsuperscript{high} population were hypothesized to potentially play roles in the downregulation of Clr-b, whereas shRNA overrepresented in the Clr-b\textsuperscript{low} population were hypothesized to be involved in maintaining high normal Clr-b expression.

To simplify the analysis, the intensities were converted to a linear scale and the relative fold-change, along with the standard deviation of each shRNA, was determined. For each shRNA, a Student’s t-test was performed to determine whether each shRNA was significantly over-represented in either the Clr-b\textsuperscript{high} or Clr-b\textsuperscript{low} population (Fig. 3.3C). Since the lentiviral pool was designed with multiple shRNA targeting each gene (~5 shRNA per target gene), the dataset was further filtered to identify genes with multiple shRNA overrepresented in the same population (Fig. 3.3D). A total of 10 genes had 5 shRNA with a fold enrichment \( \geq 1.5 \) in the Clr-b\textsuperscript{high} population, while 13 genes had 5 shRNA with a fold enrichment \( \geq 1.5 \) in the Clr-b\textsuperscript{low} population. Finally, since the nature of this project was transcriptional regulation, we focused on shRNA targeting transcription factors to further reduce the number of shRNA to validate (Fig. 3.3E-G). Although most genes do have multiple targeting shRNA, a remaining caveat is that multiple shRNA do not guarantee that each shRNA functions to knockdown the intended target or that all genes have at least 1 effective targeting shRNA.
Figure 3.3 Analysis of NIH3T3 cells transduced with the 80K pool library. (A) Flow cytometric analysis of Clr-b expression on NIH3T3 cells transduced with the 80K lentiviral pool pre- and post-sort. Dashed lines represent the secondary alone staining; solid black lines represent the population before sort; red shaded lines represent the staining of the Clr-b^{low} population; green shaded lines represent the staining of the Clr-b^{high} population. (B) Representation of each shRNA in both the Clr-b^{low} and Clr-b^{high} populations. (C) Volcano analysis of all shRNA showing fold enrichment on the x-axis and −Log p-value on the y-axis. Fold enrichment is an average of the triplicate result for each shRNA (expressed as a Log_2 value) and the p-value was calculated using Student’s t-test comparing the hybridization intensity of each shRNA from both populations, expressed as a −Log_{10} value. (D) The number of shRNA for each unique target present in the Clr-b^{high} and Clr-b^{low} groups. The data is filtered based on an enrichment ≥ 1.5 in either direction and any p-value. (E-G) Identical analyses as B-D except for shRNA targeting transcription factor genes only.
The screen was designed to segregate cohorts based upon Clr-b surface protein expression; therefore we hypothesized that some Clr-b-targeting shRNA should be over-represented in the Clr-b\text{low} population, as positive controls. While all 5 shRNA designed to target Clr-b were overabundant in the Clr-b\text{low} population, only TRCN0000065883 (TRCN65883) and TRCN65885 were significant, with fold changes of 2.76 and 1.57, respectively (Table 3.1).

<table>
<thead>
<tr>
<th>TRCN ID</th>
<th>Avg representation Clr-b\text{high}</th>
<th>Avg representation Clr-b\text{low}</th>
<th>Fold Change (Clr-b\text{low}/Clr-b\text{high})</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>0000065883</td>
<td>584.9</td>
<td>1560.83</td>
<td>2.76</td>
<td>0.003</td>
</tr>
<tr>
<td>0000065884</td>
<td>3266.44</td>
<td>4367.17</td>
<td>1.33</td>
<td>0.190</td>
</tr>
<tr>
<td>0000065885</td>
<td>1805.57</td>
<td>2840.13</td>
<td>1.57</td>
<td>0.039</td>
</tr>
<tr>
<td>0000065886</td>
<td>358.84</td>
<td>530.25</td>
<td>1.48</td>
<td>0.676</td>
</tr>
<tr>
<td>0000065887</td>
<td>1224.85</td>
<td>1746.39</td>
<td>1.42</td>
<td>0.303</td>
</tr>
</tbody>
</table>

3.2.3 Selection of shRNA candidates for a secondary iterative screen

Two major issues in validating screens of this nature are: (1) the rate of false positives; and (2) the amount of time it takes to validate individual shRNA. In order to help reduce the frequency of false positives and validate hundreds of shRNA at once, we chose to perform a second, smaller screen to validate about ~1000 shRNA positive hits from the whole genome screen. About half of the shRNA selected for the iterative screen targeted transcription factors. The rest of the shRNA consisted of the available non-targeting controls (i.e. shRNA targeting LacZ or GFP), positive controls (i.e. shRNA targeting Clr-b), shRNA targeting genes of interest to other members of the lab, and any other shRNA that matched our fold change and p value criteria (Table 3.2).
Table 3.2. shRNA selection criteria for secondary screen

<table>
<thead>
<tr>
<th>Criterion</th>
<th>Number of unique shRNAs</th>
</tr>
</thead>
<tbody>
<tr>
<td>Transcription factors with fold change ≥ 1.7 in either direction, p &lt; 0.1</td>
<td>473</td>
</tr>
<tr>
<td>Fold change ≥ 1.7, p ≤ 0.05</td>
<td>113</td>
</tr>
<tr>
<td>“Clec2d optimized” fold change ≥ 2.66, p &lt; 0.003</td>
<td>69</td>
</tr>
<tr>
<td>“Fold change optimized” fold change ≥ 8.05, p &lt; 0.2</td>
<td>290</td>
</tr>
<tr>
<td>Control shRNAs</td>
<td>9</td>
</tr>
<tr>
<td>Other shRNAs of interest</td>
<td>13</td>
</tr>
<tr>
<td>GO “response to stress” fold change ≥ 4, p &lt; 0.1</td>
<td>44</td>
</tr>
</tbody>
</table>

3.2.4 Generation of the 1K lentiviral pool and NIH3T3 cell transduction

Glycerol stocks of individual shRNA were obtained from the lab of Dr. Jason Moffat, and grown in culture overnight in 96 well plates. Half of each culture was pooled together and plasmid isolated by maxiprep. The maxiprep was co-transfected with packaging and envelop vectors to make the 1K lentiviral vector pool.

As with the 80K pool, the 1K pool was titred in NIH3T3 cells to obtain an MOI of 0.3 (Fig. 3.4) and ~250X over-representation of the library. After puromycin selection, the NIH3T3 cells were stained with Clr-b mAb (4A6) to monitor Clr-b cell surface expression (Fig. 3.5A). The cells were sorted based on Clr-b expression into 2 groups: the top 10% (Clr-b\textsuperscript{high}) and the bottom 10% (Clr-b\textsuperscript{low}) (Fig. 3.5A). Genomic DNA was extracted from both populations for PCR-based amplification of the shRNA using barcode-specific primers. The PCR products were analyzed by deep-sequencing with assistance from Dr. Jason Moffatt’s lab, as described previously [329].
Figure 3.4. Determination of the multiplicity of infection (MOI) of the 1K pool for the secondary knockdown screen in NIH3T3 cells. 2x10^6 were transduced with the indicated volume of virus in 30 mL media and 10 µg/mL polybrene with gentle shaking for 20 min. Each transduction was split into 2 plates. After 24h the media was replaced on half the plates with selection media containing 2.5 µg/mL puromycin. The number of live cells in each plate was counted after 2 days of selection and the MOI was calculated using the given equation.
3.2.4.1 Screen analysis and results

The relative reads of each shRNA in both populations of the 1K pool was obtained (Fig. 3.5B). We hypothesized that the majority of the shRNA tested would play a role in Clr-b regulation if the results repeated, thus most shRNA were expected to be over-represented in one of the populations, within the limits of the false positive rate from the first screen. However, it appeared that the majority of the shRNA were more or less equally represented in both groups, suggesting that the criteria used to select shRNA for a second round of testing were not adequate, that the variability was extremely high, and/or that the statistical method of selecting candidates by triplicate according to fold-change and t-test within the experimental design for 80,000 genes in triplicate were inadequate. Nonetheless, one shRNA (TRCN77292) that targeted a transcription factor, IRF7, (Fig. 3.5B and E) was dramatically over-represented in the Clr-b\text{low} fraction. In addition, one gene, MeCP2, which is a transcription factor, had 3 targeting shRNA (TRCN39081, TRCN39079, and TRCN3983) in the Clr-b\text{high} group (Fig. 3.5D and G). In the Clr-b\text{low} group, 3 transcription factors had at least 2 shRNA targets identified: Irf7 (TRCN77292 and TRCN00091), Med14 (TRCN96283, TRCN96281), and Zfp759 (TRCN96036 and TRCN96037) (Fig. 3.5D and G).
Figure 3.5 Analysis of NIH3T3 cells transduced with the 1K pool library. (A) Flow cytometric analysis of Clr-b expression on NIH3T3 cells transduced with the 1K lentiviral pool pre- and post-sort. Dashed lines represent the secondary alone staining; solid black lines represent the population before sort; red shaded lines represent the staining of the Clr-b\textsuperscript{low} population; green shaded line represents the staining of the Clr-b\textsuperscript{high} population. (B) Representation of each shRNA in both the Clr-b\textsuperscript{low} and Clr-b\textsuperscript{high} populations. (C) Volcano analysis of all shRNAs showing fold enrichment on the x-axis and –Log p-value on the y-axis. Fold enrichment is an average of the triplicate result for each shRNA (expressed as a Log\textsubscript{2} value) and the p-value was calculated using Student’s t-test comparing the normalized number of reads for each shRNA from both populations, expressed as a –Log\textsubscript{10} value. (D) The number of shRNAs for each unique target present in the Clr-b\textsuperscript{high} and Clr-b\textsuperscript{low} groups. The data is filtered based on an enrichment ≥ 1.25 in either direction and any p-value. (E-G) Identical analyses as B-D except for shRNAs targeting transcription factor genes only.
Since the purpose of the second iterative screen was to validate as many shRNA as possible in bulk from the first screen, the results obtained from the second screen were compared to those of the first screen to look at the correlation between the 2 screens. The log$_2$ of the fold change (Clr-b$^{\text{high}}$/Clr-b$^{\text{low}}$) from this second screen was compared to that obtained from the first screen using a Spearman correlation, which does not assume Gaussian distribution of the datasets (Fig. 3.6A and B). The positive correlation between the 2 screens was relatively high, 0.47 for all shRNA and about 0.45 for transcription factor-specific shRNA, and both of these correlations were highly significant (p < 0.0001). One transcription factor-specific shRNA emerged as not correlated between the 2 screens. This shRNA, TRCN77292, was designed to target IRF7 and was the shRNA that came out most highly over-represented in the Clr-b$^{\text{low}}$ group in the second screen. The 1K pool contained a total of 1,010 shRNA. Of these, 733 shRNA validated (i.e. were over-represented in the same population in the first and second screens), while 277 did not validate (Fig. 3.6C). A total of 576 shRNA targeted TF in the second screen. Similarly, 406 validated, while 170 did not validate (Fig. 3.6D). A caveat to this analysis is that it does not take into account that multiple shRNA targeting the same gene may not have validated in the same direction (i.e. 1 shRNA targeting gene A was validated as overrepresented in the Clr-b$^{\text{low}}$ cells, while a second shRNA targeting gene A was validated as overrepresented in the Clr-b$^{\text{high}}$ cells).
Figure 3.6 Correlations between the first 80K pool screen and the second 1K pool screen. 
(A) Comparison of the log\(_2\) fold change (CLR-b\(_{\text{high}}\)/CLR-b\(_{\text{low}}\)) of each hairpin from the second screen (y-axis) with the log\(_2\) fold change obtained from the first screen. Red dotted lines demark log\(_2\) fold change of 1 and –1. Statistical analysis by two-tailed nonparametric Spearman correlation. (B) Comparison as in (A) but for shRNAs designed to target transcription factors only. (C-D) Venn diagrams depicting the number of shRNAs that validated in the same population in both screens (overlap area) and the number of shRNAs that did not validate between both screens. (C) shows all shRNAs while (D) shows only the shRNAs targeting transcription factors.
A final step to reduce the number of TF-targeting shRNA was to examine RNA-Seq data performed and analyzed by Oscar A. Aguilar in our laboratory, and filter the list of shRNA to those that target genes expressed in NIH3T3 cells. This analysis helps to prevent further validation of shRNA with off-target effects, because if an shRNA is significantly over-represented in one group over the other, yet it targets a non-expressed gene, then the shRNA may be knocking down an off-target gene (a caveat here being that lentiviral transduction may itself upregulate some immune response genes that are not expressed in parental, untransduced NIH3T3 cells). While targets that fall into this category thus remain interesting, it is simpler when facing over 1000 candidates to look at shRNA with presumed on-target effects. From the RNA-Seq data, genes with Q value > 3 were defined as expressed. The genes targeted by the top 25 shRNA over-represented in the Clr-b\textsuperscript{high} cohort and the top 25 shRNA over-represented in the Clr-b\textsuperscript{low} cohort were cross referenced with the RNA-Seq data to assess whether the genes were expressed. This analysis revealed that Hoxb6, Elf5, St18, Alx1, Gbx2, Hnf4a, Hoxb1, Hoxb6, Nr1h3, Crx, Lmo2, Myt1l, Pou4f3, Nr5a2, Hey1, and Noto were not expressed in NIH3T3 cells, and as such shRNA targeting these genes were excluded from further analysis, but are noted here for completeness. The following shRNA were deemed suitable for further validation into their role of regulating Clr-b expression (Table 3.3)

Table 3.3. List of shRNA selected for further validation

<table>
<thead>
<tr>
<th>TRCN</th>
<th>Gene Name</th>
<th>Gene Description</th>
<th>Fold Change First screen</th>
<th>Fold Change Second screen</th>
</tr>
</thead>
<tbody>
<tr>
<td>TRCN0000012678</td>
<td>Tcf7</td>
<td>Transcription factor 7, T-cell specific</td>
<td>8.59</td>
<td>12.92</td>
</tr>
<tr>
<td>TRCN0000026178</td>
<td>Esrra</td>
<td>Estrogen related receptor, alpha</td>
<td>2.56</td>
<td>7.94</td>
</tr>
<tr>
<td>TRCN0000123414</td>
<td>Phf5a</td>
<td>PHD finger protein 5A</td>
<td>8.29</td>
<td>5.36</td>
</tr>
<tr>
<td>TRCN0000012068</td>
<td>Mef2c</td>
<td>Myocyte enhancer factor 2C</td>
<td>4.82</td>
<td>4.41</td>
</tr>
<tr>
<td>Gene ID</td>
<td>Gene Symbol</td>
<td>Description</td>
<td>Fold Change Rel</td>
<td>Fold Change Abs</td>
</tr>
<tr>
<td>----------------</td>
<td>-------------</td>
<td>------------------------------------------------------------------------------</td>
<td>-----------------</td>
<td>-----------------</td>
</tr>
<tr>
<td>TRCN0000055203</td>
<td>Jun</td>
<td>Jun oncogene</td>
<td>1.99</td>
<td>3.78</td>
</tr>
<tr>
<td>TRCN0000174596</td>
<td>Lass5</td>
<td>LAG1 homolog, ceramide synthase 5</td>
<td>2.03</td>
<td>3.27</td>
</tr>
<tr>
<td>TRCN0000178405</td>
<td>Crebl2</td>
<td>cAMP responsive element binding protein-like</td>
<td>2.02</td>
<td>2.80</td>
</tr>
<tr>
<td>TRCN000012347</td>
<td>Nfkb2</td>
<td>Nuclear factor of kappa light polypeptide gene enhancer in B-cells 2, p49/p100</td>
<td>1.88</td>
<td>2.87</td>
</tr>
<tr>
<td>TRCN000042712</td>
<td>Mycl1</td>
<td>v-myc myelocytomatosis viral oncogene homolog 1, lung carcinoma derived (avian)</td>
<td>2.92</td>
<td>2.61</td>
</tr>
<tr>
<td>TRCN000001660</td>
<td>Pparg</td>
<td>Peroxisome proliferator activated receptor gamma</td>
<td>-1.87</td>
<td>2.57</td>
</tr>
<tr>
<td>TRCN0000081743</td>
<td>Foxg1</td>
<td>Forkhead box G1</td>
<td>2.00</td>
<td>2.53</td>
</tr>
<tr>
<td>TRCN0000097132</td>
<td>Nfate3</td>
<td>Nuclear factor of activated T-cells, cytoplasmic, calcineurin-dependent</td>
<td>2.75</td>
<td>2.42</td>
</tr>
<tr>
<td>TRCN000075513</td>
<td>Tfdp1</td>
<td>Transcription factor Dp 1</td>
<td>4.30</td>
<td>2.31</td>
</tr>
<tr>
<td>TRCN000039230</td>
<td>Taf5</td>
<td>TAF5 RNA polymerase II, TATA box binding protein (TBP)-associated factor</td>
<td>1.72</td>
<td>2.19</td>
</tr>
<tr>
<td>TRCN000096110</td>
<td>Zfp281</td>
<td>Zinc finger protein 281</td>
<td>33.04</td>
<td>2.19</td>
</tr>
<tr>
<td>TRCN000085846</td>
<td>Tcerg1</td>
<td>Transcription factor elongation regulator 1</td>
<td>2.53</td>
<td>2.08</td>
</tr>
<tr>
<td>TRCN000096099</td>
<td>Irx3</td>
<td>Iroquois related homeobox 3 (Drosophila)</td>
<td>11.13</td>
<td>2.0</td>
</tr>
<tr>
<td>TRCN000084564</td>
<td>Crebf</td>
<td>CREB/ATF bZIP transcription factor</td>
<td>1.81</td>
<td>1.99</td>
</tr>
<tr>
<td>TRCN000072005</td>
<td>Foxp1</td>
<td>Forkhead box P1</td>
<td>2.43</td>
<td>1.96</td>
</tr>
<tr>
<td>TRCN000039079</td>
<td>MeCP2</td>
<td>Methyl CpG binding protein 2</td>
<td>1.74</td>
<td>1.35</td>
</tr>
<tr>
<td>TRCN0000039080</td>
<td>Nfkb1</td>
<td>Nuclear factor of kappa light polypeptide gene enhancer in B-cells 1, p105</td>
<td>-11.74</td>
<td>-1.9</td>
</tr>
<tr>
<td>TRCN0000042678</td>
<td>Fos</td>
<td>FBJ osteosarcoma oncogene</td>
<td>-1.87</td>
<td>-1.93</td>
</tr>
<tr>
<td>TRCN0000070760</td>
<td>Shox2</td>
<td>Short stature homeobox 2</td>
<td>-2.02</td>
<td>-1.93</td>
</tr>
<tr>
<td>TRCN0000095218</td>
<td>Gata4</td>
<td>GATA binding protein 4</td>
<td>-2.77</td>
<td>-1.95</td>
</tr>
<tr>
<td>TRCN000125387</td>
<td>Lass4</td>
<td>LAG1 homolog, ceramide synthase 4</td>
<td>-2.50</td>
<td>-2.07</td>
</tr>
<tr>
<td>TRCN000085520</td>
<td>Rcor2</td>
<td>REST corepressor 2</td>
<td>-2.20</td>
<td>-2.14</td>
</tr>
<tr>
<td>TRCN000085334</td>
<td>Alx1</td>
<td>ALX homeobox 1</td>
<td>-1.70</td>
<td>-2.16</td>
</tr>
<tr>
<td>TRCN000099575</td>
<td>Med1</td>
<td>Mediator complex subunit 1</td>
<td>-2.58</td>
<td>-2.29</td>
</tr>
<tr>
<td>TRCN000085149</td>
<td>Satb2</td>
<td>Special AT-rich sequence binding protein 2</td>
<td>-2.76</td>
<td>-2.31</td>
</tr>
<tr>
<td>TRCN000086619</td>
<td>Zkscan5</td>
<td>Zinc finger with KRAB and SCAN domains 5</td>
<td>-2.76</td>
<td>-2.31</td>
</tr>
<tr>
<td>TRCN000098041</td>
<td>Arid3a</td>
<td>AT rich interactive domain 3A (BRIGHT-like)</td>
<td>-1.79</td>
<td>-2.36</td>
</tr>
<tr>
<td>TRCN000086246</td>
<td>Atf7</td>
<td>Activating transcription factor 7</td>
<td>-1.94</td>
<td>-3.00</td>
</tr>
<tr>
<td>TRCN00012527</td>
<td>Meis1</td>
<td>Meis homeobox 1</td>
<td>-2.16</td>
<td>-3.26</td>
</tr>
<tr>
<td>TRCN000071982</td>
<td>Tbp</td>
<td>TATA box binding protein</td>
<td>-1.85</td>
<td>-3.57</td>
</tr>
<tr>
<td>TRCN000012724</td>
<td>Crebbp</td>
<td>CREB binding protein</td>
<td>-7.59</td>
<td>-3.95</td>
</tr>
<tr>
<td>TRCN000077292</td>
<td>Irf7</td>
<td>Interferon regulatory factor 7</td>
<td>1.85</td>
<td>-128.26</td>
</tr>
</tbody>
</table>
3.2.5 Validation of the effects of individual shRNAs on Clr-b expression

With the results of many shRNA from the first screen validated at least correlatively by the second screen, the next step was to validate individual shRNA for their influence on Clr-b levels (i.e., shRNA over-represented in the Clr-b\textsubscript{low} cohort should promote Clr-b downregulation relative to a negative control shRNA or the entire 80K pool). Fig. 3.7 provides a schematic representation of how the validation of individual shRNA was carried out. First, shRNA were tested for their impact on Clr-b surface expression by flow cytometry. shRNA that significantly altered Clr-b surface expression were further validated for their impact on the desired target gene product by Western blot and/or qPCR to determine whether the target was indeed knocked down. Clr-b transcript levels were also quantitated by qPCR. We hypothesized that if a transcription factor had a direct impact on Clr-b expression, transcript levels would also be altered, whereas if Clr-b protein but not transcript levels were altered, the transcription factor would play an indirect role in Clr-b expression.
Figure 3.7 Schematic representation of the workflow of the validation of single shRNAs. NIH3T3 cells were transduced with individual shRNA lentivirus library. Untransduced cells were removed by puromycin selection. NIH3T3 cells were stained with Clr-b antibody to validate the effect of shRNA on Clr-b expression by flow cytometry. shRNA that caused a change in Clr-b expression were further validated for knockdown of the desired target by quantitative PCR and Western blot.
First, a non-targeting negative control shRNA was needed for these experiments. shRNA specific for luciferase, GFP, RFP, and LacZ were available. **Table 3.4** shows the fold change of the negative control non-targeting shRNA from the second screen. The effect of each negative control shRNA on Clr-b expression was also confirmed by flow cytometry following transduction of NIH3T3 cells with individual lentiviral vectors (**Fig. 3.8A**). We hypothesized that any of these shRNA would function as a suitable negative control, however most of these shRNA (6/9) actually increased Clr-b expression. Based on these results, LacZ (TRCN72240) was chosen as a control shRNA. The 80K pool itself also functions as a polyclonal normalization control, since most of the shRNA in the pool do not alter Clr-b expression significantly.

**Table 3.4. Analysis of the effect of negative control shRNA on Clr-b expression based on fold-change from the second iterative screen**

<table>
<thead>
<tr>
<th>TRCN ID</th>
<th>Gene</th>
<th>Fold Change</th>
</tr>
</thead>
<tbody>
<tr>
<td>TRCN0000072256</td>
<td>LUCIFERASE</td>
<td>1.58</td>
</tr>
<tr>
<td>TRCN0000072246</td>
<td>LUCIFERASE</td>
<td>-1.08</td>
</tr>
<tr>
<td>TRCN0000072181</td>
<td>GFP</td>
<td>-1.10</td>
</tr>
<tr>
<td>TRCN0000072179</td>
<td>GFP</td>
<td>-1.32</td>
</tr>
<tr>
<td>TRCN0000072186</td>
<td>GFP</td>
<td>1.12</td>
</tr>
<tr>
<td>TRCN0000072210</td>
<td>RFP</td>
<td>-2.00</td>
</tr>
<tr>
<td>TRCN0000072219</td>
<td>RFP</td>
<td>1.43</td>
</tr>
<tr>
<td>TRCN0000072235</td>
<td>LacZ</td>
<td>2.12</td>
</tr>
<tr>
<td>TRCN0000072240</td>
<td>LacZ</td>
<td>1.11</td>
</tr>
</tbody>
</table>

The shRNA targeting Clr-b were also tested individually as a positive control for Clr-b knockdown at both the cell surface and transcript levels (**Fig. 3.8B**). Based on the representation of the Clr-b-targeting shRNA in the first screen (**Table 3.1**), we hypothesized that TRCN65883 would cause Clr-b knockdown; however, TRCN65885 was the only Clr-b-targeting shRNA that promoted significant Clr-b knockdown at the cell surface and transcript levels (**Fig. 3.8B**).
TRCN65884 could not be assessed, because despite repeated attempts, cells transduced with this lentiviral vector died during puromycin selection, following the same timing as non-infected cells, suggesting either that productive packaged lentivirus could not be generated, or that the shRNA may have been lethal to the selecting NIH3T3 cells. This is interesting to note, since Clr-b is predicted to possess a TNF-receptor associated factor-2 (TRAF2)-interaction motif in its cytosolic tail that may influence cell viability versus apoptosis [40].
Figure 3.8 Analysis of control shRNA using lentiviral shRNA viruses. (A) Flow cytometric analysis of control shRNA on Clr-b expression in NIH3T3 cells. Dotted lines represent secondary reagent staining; solid black lines represent staining on parental NIH3T3 cells; shaded grey histograms represent the staining on cells transduced with shRNA and selected with puromycin. (B) Analysis of Clr-b-targeting shRNAs by flow cytometry and quantitative PCR. Dotted lines represent secondary reagent staining; solid black lines represent staining on NIH3T3 cells transduced with LacZ shRNA and selected with puromycin; shaded grey histograms represent staining on cells transduced with Clr-b shRNA and selected with puromycin. Quantitative PCR for Clec2d transcripts relative to Gapdh and normalized to LacZ. Two-tailed Student’s t-test; * = p < 0.05; n = 3
With the help of Kuan Lun (William) Chu, an undergraduate student under my supervision, 25 of 36 genes from Table 3.3 were validated by flow cytometry (Fig. 3.9). Where multiple shRNA for candidate genes were present in the second screen, all shRNA for that gene were tested, as we hypothesized that genes of this nature were likely to validate. Of the 25 genes examined, 8 genes (32%) caused a statistically significant change in Clr-b expression in the direction predicted by both screens, except TRCN12068 that targets Mef2c, which caused a significant downregulation in Clr-b expression despite being over-represented in the Clr-b^{high} cohort (Fig. 3.9). shRNA targeting these 8 genes were further validated by Western blot and/or qPCR.
Figure 3.9. Analysis of the effect of individual shRNAs on Clr-b expression by flow cytometry. NIH3T3 cells were transduced with lentivirus containing one shRNA and selected with puromycin. Plots show the average median fluorescence intensity ± SEM of at least 3 independent biological replicates. Statistical analyses were performed by two-tailed Student’s t-test. **, p<0.01; *, p<0.05. Con, control non-targeting shRNA.
3.2.5.1 MeCP2 is a negative regulator of Clr-b expression

One shRNA designed to target methyl CpG binding protein 2 (MeCP2; TRCN39081), emerged as one of the most highly significant hits in the Clr-b\textsuperscript{high} population during the first screen, and validated during the second screen. The other 4 shRNA designed to target MeCP2 were included in the second screen as well, and both TRCN39079 and TRCN39083 also validated as over-represented in the Clr-b\textsuperscript{high} population (Table 3.3). The other 2 MeCP2-targeting shRNA appeared to be equally present in both populations. Loss-of-function mutations in the MECP2 gene result in Rett syndrome in females, which is characterized by progressive dementia, motor loss, developmental stagnation, and deterioration after 1-1.5 years of age [331]. In addition, duplication of MECP2 results in MECP2-duplication syndrome (MDS) in males, which is characterized by infantile hypotonia, autistic features, gait abnormalities, and recurrent infections [332]. As the name implies, MeCP2 typically binds methylated cytosines, although it can bind DNA in a methylation-independent manner, and it represses transcription through an as yet to be identified mechanism [333].

TRCN39079 and TRCN39081 shRNA resulted in a significant increase in Clr-b protein (~2-3-fold), (Fig. 3.10A). Since MeCP2 is a known transcription factor, we hypothesized that if MeCP2 is acting directly on the Clr-b gene locus, Clec2d transcripts should increase as well. In parallel with Clr-b surface expression, TRCN39081 caused a statistically significant increase in Clr-b transcripts (~2-fold), while TRCN39079 had a trend towards increased Clec2d transcripts (Fig. 3.10B). To confirm whether any of these shRNA promoted MeCP2 knockdown, the transcript levels of Mecp2 were analyzed by qPCR, which revealed that TRCN39081 caused a 75% decrease in Mecp2 transcripts (Fig. 3.10B). The protein levels of MeCP2 were examined by Western blot to determine whether the decreased MeCP2 transcripts resulted in less protein or whether TRCN39079 may act post-transcriptionally to regulate MeCP2 expression. TRCN39079, TRCN39081, and TRCN39083 caused significant reduction in MeCP2 protein (0% MeCP2 remaining for TRCN39079 and TRCN39081 and 53% remaining for TRCN39083) (Fig. 3.10C), indicating that these shRNA are indeed on-target for MeCP2. Thus, loss of MeCP2 protein results in increased Clec2d transcripts and Clr-b cell surface expression, suggesting that, when expressed, MeCP2 would negatively regulate the expression of Clr-b. Notably, the Clec2d
gene possesses a CpG island ~10kb upstream of the promoter, which could be a direct target of MeCP2 activity.
Figure 3.10 Knockdown of MeCP2 increases Clr-b expression. (A) Depletion of MeCP2 increases cell surface expression of Clr-b. NIH3T3 cells were transduced with lentivirus containing one of the shRNAs specific for MeCP2 and selected with puromycin. Clr-b expression was assessed by flow cytometry using 4A6 anti-Clr-b antibody. Dotted lines represent secondary reagent staining; solid black lines represent staining from the control 80K pool; gray shading represents staining from the MeCP2-targeting shRNAs. (B) Quantitation of Clec2d and Mecp2 transcripts from (A) normalized to Gapdh and 80K pool. Statistical analysis was performed by two-tailed Student’s t-test; n=3. (C) Levels of MeCP2 and Actin from cells in (A) were assessed by Western blot. An unidentified band in the anti-MeCP2 blot is denoted by *. Quantitation of the Western is shown to the right, depicting the percent MeCP2 remaining indexed to the 80K pool control.
3.2.5.2 IRF7 is not involved in regulating Clr-b expression

An shRNA targeting IRF7, TRCN77292, was the biggest hit in the second screen being the most overrepresented gene in the Clr-b low population. While TRCN77292 did not validate between the first and second screens (Table 3.3), TRCN77291 did validate. We therefore examined both of these shRNA further, which revealed no effects on Clr-b protein levels by flow cytometry (Fig. 3.11A). On the other hand, both shRNA caused reduction in IRF7 protein, 30% for TRCN77291 and 88% for TRCN77292 (Fig. 3.11B), suggesting that the shRNA are on-target for IRF7. Thus, while both shRNA were specific for IRF7 and reduced IRF7 protein levels significantly, the reduced IRF7 expression had no effect on Clr-b expression, suggesting IRF7 is not important in maintaining Clr-b expression. A more detailed examination of the role, if any, for IRF7 in Clr-b regulation will be presented in Chapter 4 using IRF7−/− primary MEF cells and type-I IFN.
Figure 3.11 Knockdown of IRF7 does not change Clr-b expression. (A) NIH3T3 cells were transduced with lentivirus containing one of the shRNAs specific for IRF7 or the entire 80K pool and selected with puromycin. Clr-b expression was assessed by flow cytometry using 4A6 anti-Clr-b antibody. Dotted lines represent secondary reagent staining; solid black lines represent staining from the control 80K pool; shaded histograms represent staining from the cells transduced with MeCP2-targeting shRNA. (B) Protein levels of IRF7 and Actin from (A) were assessed by Western blot. Quantitation of the Western is shown to the right, depicting the percent IRF7 remaining remaining indexed to the 80K pool control. Representative of 2 independent experiments.
3.2.5.3  TBP and CREBBP positively affect Clr-b transcription

TBP (TATA box binding protein), the best-known subunit of RNAPII, seemed an obvious choice for validation, as the Clr-b promoter contains an inverted TATA box ~25bp upstream of the TSS [113] (See also Fig. 4.5A in Chapter 4). Infection of NIH3T3 cells with TBP shRNA (TRCN71982) resulted in as much as 64% reduction of Clr-b surface expression (Fig. 3.12A). Analysis of Tbp transcripts by qPCR revealed a 4-fold (75%) loss, suggesting the shRNA is on-target for Tbp. At the same time, Clec2d transcripts were decreased by 1.7-fold (38%) (Fig. 3.12B). Western blot analysis revealed that TBP protein levels were also reduced by ~50% following transduction with TRCN71982 shRNA (Fig. 3.12C). Since a partial loss of Tbp transcripts and protein results in a partial reduction of Clec2d transcripts and protein, this suggests that the presence of TBP is required for normal or constitutive high levels of Clec2d transcripts and Clr-b surface protein.
Figure 3.12 Knockdown of TBP, or Crebbp decreases Clr-b expression. (A) and (D) NIH3T3 cells were transduced with lentivirus containing an shRNA specific for Tbp or Crebbp and selected with puromycin. Clr-b expression was assessed by flow cytometry using 4A6 anti-Clr-b mAb. Dotted lines represent secondary reagent staining; solid black lines represent staining from the 80K pool; gray shading represents staining from the cells transduced with shRNA targeting Tbp or Crebbp. (B) and (E) Quantitation of Clec2d, Tbp, or Crebbp transcripts from A, D normalized to Gapdh and 80K pool. Statistical analyses were performed by two-tailed Student’s t-test; n=3 for each. (C) Protein level of TBP and Actin from the cells in (A) were assessed by Western blot. Quantitation of TBP protein from Western is depicted to the right showing the percent of TBP remaining indexed to the 80K pool control.
Cyclic adenosine monophosphate (cAMP) response element-binding protein (CREB)-binding protein (CREBBP or CBP), is a transcriptional co-activator of many genes with histone acetylase activity [334]. When phosphorylated, CBP binds to c-AMP-regulated enhancer sequences in DNA. CBP contains 3 cysteine/histidine-rich (CH) domains, facilitating interactions with several transcription factors, including PU.1, GATA-1/2/3, TFIID, TBP, and RNAPII [335]. An acetyltransferase domain acetylates all 4 core histone proteins and some nonhistone proteins. Mutations in CREBBP have been implicated in cancers, including childhood high hyperdiploid acute lymphoblastic leukemia (ALL) [336], acute myeloid leukemia (AML) [337], and the cancer predisposition syndrome, Rubinstein-Taybi syndrome, which is characterized by mental retardation and craniofacial malformations [338]. Transduction of TRCN12724, designed to target Crebbp, resulted in as much as 75% reduction in Clr-b expression at the cell surface (Fig. 3.12D) and caused a significant loss in both Clec2d (36%) and Crebbp transcripts (34%), (Fig. 3.12E), suggesting that the shRNA is on-target for Crebbp and that loss of CBP causes Clr-b reduction at the transcript level. Western blots to assess CBP protein levels were attempted, but failed to yield any CBP band, even in parental NIH3T3 cells. CBP expression has been detected by Western in NIH3T3 cells by others, using a different antibody clone, making the antibody used a potential reason for the lack of CBP detection [339]. CBP is a large protein, ~265 kDa, so it is also possible that altered transfer conditions would be necessary to transfer CBP from the gel to the membrane for detection. As with TBP above, knockdown of Crebbp transcripts resulted in less Clec2d and Clr-b, indicating that CBP plays a role in activating Clr-b expression at the genomic level. Functional cAMP response elements (CRE) are usually located 100-200bp upstream of a TATA box and exist as either a full palindrome, TGACGTCA, or a less active half palindrome, CGTCA [340, 341]. The Clec2d promoter and CDS were analyzed for CRE consensus sequences on either the + or – strand and a TGACCTCA sequence 182bp upstream from the TATA box. Since CBP also forms interactions with other proteins, CBP may also influence Clr-b expression by complexing with other transcription factor(s) that bind directly to the Clr-b promoter. Depletion of Crebbp resulted in loss of Clr-b expression, indicating that expression of CBP is important in maintaining high Clr-b surface levels.
3.2.5.4 Med1, ATF7, and Satb2 downregulate Clr-b protein but not transcripts

Med1, a component of the Mediator complex, is another logical choice for validation, as Mediator is involved in the transcription of almost all genes transcribed by RNAPII [342, 343], through a direct interaction between Mediator, the CTD domain of RNAPII [344], and DNA-bound transcriptional activators. This indicates that Mediator is required for both basal and regulated gene expression. Although detailed information regarding the function of all Mediator subunits remains elusive, Mediator itself is a heterogeneous oligomeric complex, where different isolations of the complex contain differing subunit compositions [345]. Subunit Med1 is a component of the middle module of the Mediator complex [346], and is present in only a subpopulation of Mediator complexes. It has been shown to be important for direct communication between nuclear receptors and components of the basal transcription factor machinery [345]. Studies using Med1−/− fibroblasts demonstrated that Med1 is required for the intracellular function of thyroid hormone receptor [347], peroxisome proliferator-activated receptor gamma (Pparγ) [348] and estrogen receptor (ER)-mediated transcription [345]. Med1 protein contains 2 nuclear receptor recognition motifs (LXXLL) [347] that are required for the physical interaction of Med1 with Pparγ [349]. Transduction of NIH3T3 cells with Med1 shRNA (TRCN99575) resulted in a ~1.8-fold reduction in Clr-b protein (Fig. 3.13A). Analysis of Med1 transcripts by qPCR demonstrated that Med1 was knocked down by ~ 2.5-fold (60%) (Fig. 3.13B). Although a trend towards a decrease in Clec2d transcripts was observed upon transduction of TRCN99575, this difference was not statistically significant, suggesting that Med1 does not regulate Clr-b expression by directly affecting relative transcript abundance (Figure 3.13B). Further validation of the specificity of TRCN99575 on Med1 is required to know whether the decrease in transcripts translates to a significant loss of protein, or whether the shRNA acts post-transcriptionally to regulate Med1 protein levels.

ATF-7 (activating transcription factor 7, also known as ATFa) is a member of the ATF/CREB family of basic region-leucine zipper proteins and the sub-family of AP-1 transcription factors. ATF-7 exists as 3 splice isoforms, ATF7-1, ATF7-2, and ATF7-3, although only ATF7-2 is expressed in mice. ATF7 homodimers are known to bind CRE elements
(described above) [350] but when ATF7 forms heterodimers with c-Jun, it can also bind to tetracycline response element (TRE) sequences. Atf7 is expressed ubiquitously [350]. ATF-7 interacts with the E1a adenovirus oncoprotein, c-Jun and c-Fos protein via its C-terminal leucine-zipper [351]. Using in vitro and in vivo studies, it has been shown that ATF-7 forms an interaction with TAF12 in the TFIID subunit of RNAPII, which activates transcription of genes bound by ATF-7 [352]. This activation of ATF7-mediated transcription by TAF12 can be blocked by TAF4 [352]. TRCN86246, targeting ATF7, was identified as a hit that caused Clr-b downregulation in both screens (Table 3.3; ratios of −1.94 and −3.00 for the first and second screens, respectively). Validation of this phenotype with the single TRCN86246 shRNA lentivirus resulted in ~33-49% loss of Clr-b surface protein by flow cytometry (Fig. 3.13C). Knockdown of Clec2d and Atf7 transcripts was assessed by qPCR, which demonstrated that TRCN86246 caused ~85% reduction in Atf7 transcripts, while Clec2d transcripts remained unchanged (Fig. 3.13D). These findings suggest that ATF7-mediated Clr-b downregulation is not caused by direct interaction of ATF7 at the Clec2d promoter, although promoter activity assays and ChIP would formally demonstrate whether this is true. Again, the shRNA could also be acting post-transcriptionally on ATF7.

Satb2 (special AT-rich binding protein-2) binds to nuclear matrix-attachment regions (MAR) to activate transcription; MAR have been implicated in regulating gene expression by modifying the organization of chromosomes and assisting enhancers to act over large distances. In humans, haploinsufficiency of SATB2 results in cleft palate [353]. Recently, the clinical term SATB2-associated syndrome has been proposed, characterized by severe intellectual disability, delayed speech, behavioural and sleep problems, osteoporosis, and more disabilities [354]. While a related gene, Satb1, is thymocyte-specific, Satb2 expression is found in developing bone (osteoblasts) during skeletogenesis. In Satb2−/− mice, for example, the deletion of Satb2 is perinatal lethal, resulting in numerous bone malformations, several missing teeth, shorter bones, and a cleft palate compared to WT mice [355]. ChIP assays have demonstrated that this is caused, in part, by SATB2 binding to the EII enhancer of the Hoxa gene to repress transcription in osteoblast cells [355]. SATB2 is also expressed in early B cell development (pre-B cells), in the brain, and kidney [356]. SATB2 has been shown to bind to specific MAR DNA sequences,
and to other transcription factors, including ATF4 and Runx2, to enhance the transcriptional activity of these transcription factors [355]. In pre-B cells, ChIP analysis demonstrated that tagged SATB2 binds the 5'-MAR sequences in the Eμ intronic enhancer of the immunoglobulin (Ig) heavy chain gene and augments transcription [356]. Satb2 contains 2 SUMOylatable lysine residues (233 and 250) that, when SUMOylated, function to antagonize SATB2-mediated transcription by downregulating its contact with MAR sequences in DNA and targeting SATB2 to the nuclear periphery [356]. Transduction of NIH3T3 cells with TRCN85149 targeting Satb2 was predicted from Table 3.3 to cause Clr-b downregulation, as it was over-represented in the Clr-b^{low} populations in both screens. Validation by flow cytometry resulted in ∼50-60% loss in Clr-b protein at the cell surface (Fig. 3.13E). To validate whether TRCN85149 is specific for Satb2, qPCR was performed, demonstrating a ∼30% reduction in Satb2 transcripts (Fig. 3.13F). Analysis of Clec2d transcripts revealed no change in steady-state transcript levels, suggesting that Satb2-mediated downregulation of Clr-b protein does not occur at the genomic level (Fig. 3.13F), and may again be post-transcriptional.
Figure 3.13 Knockdown of Med1, Atf7, or Satb2 decreases Clr-b expression at the protein level. (A,C,E) NIH3T3 cells were transduced with lentivirus containing one of the shRNAs specific for Med1, Atf7, or Satb2 and selected with puromycin. Clr-b expression was assessed by flow cytometry using 4A6 anti-Clr-b antibody. Dotted lines represent secondary reagent staining; solid black lines represent staining from the control LacZ shRNA or 80K pool; gray shading represents staining from the shRNA targeting Med1, Atf7, or Satb2. (B,D,F) Quantitation of Clec2d, Med1, Atf7, or Satb2 transcripts from A,C,E) normalized to Gapdh and either LacZ shRNA or 80K pool. Statistical analyses were performed by two-tailed Student’s t-test; n=3 for each
3.3 Discussion

Previous studies examining regulation of Clr-b expression have taken a direct approach and yielded some functional findings. These studies have consisted of treating cells with chemicals, such as DNA-damage agents or cytokines, then examining Clr-b or Ocil expression by flow cytometry and/or relative RNA abundance by RT-PCR [40, 145, 357]. While these studies have aided in the understanding of the processes by which Clr-b expression is modulated, they have not provided much insight into specific mechanism(s) by which this regulation occurs. Since we have previously shown that Clr-b has a relatively short half-life at the cell surface (3h), using Brefeldin A treatment followed by flow cytometric analysis for Clr-b expression [145], we hypothesized that Clr-b may need to be constitutively and highly transcribed in healthy cells to ensure high levels of surface protein. Conversely, during pathological alterations, Clr-b transcription would decrease or halt all together, resulting in a dramatic loss of surface protein within hours. If this were true, the high expression of Clr-b on healthy cells and subsequent decrease in Clr-b expression may rely on the presence or absence of various transcription factors and RNAPII to drive constitutive levels of Clec2d transcripts, depending on the condition of the cell. To help identify novel transcription factors that play a role in the normal or regulated expression of Clr-b, we employed a lentiviral shRNA high-throughput screen approach that surveyed the whole mouse genome.

The genome-wide shRNA screen was carried out using The RNAi Consortium 80K lentiviral shRNA library, which consists of 77,690 hairpins targeting about 15,538 mouse genes. NIH3T3 cells transduced with the lentiviral vectors were sorted based upon Clr-b levels into the top and bottom 10% and the relative abundance of each shRNA in the two cohorts was determined by microarray. Validation of individual shRNA was a labour-intensive and time-consuming process. To validate hundreds of shRNA at once, a second iterative screen of ~1000 shRNA was manually selected based on the first 80K screen, and carried out the same way, except that the genomic DNA was analyzed by deep sequencing instead of microarray. The data from both screens were analyzed using a few approaches. These included comparing the fold enrichment of each shRNA in both populations with an arbitrary cut-off and determining significance by Student’s t-test; filtering the data set based on shRNA that were over-represented
in the same population in both screens, and further filtering based on shRNA designed to target
genes known to be expressed in NIH3T3 cells using RNA-Seq data. The second screen and
filtering the data in this way helped to decrease the overwhelming size of the initial data set
(~80,000 shRNA).

While screens of this nature are simple to perform and unbiased, interpreting relevant hits
from the noise is challenging. While using arbitrary fold enrichment cut-offs is a simple way to
analyze a large data set, it does not take into account higher-level statistical approaches, or the
biological significance of the target genes. By incorporating mainly transcription factors or
proteins with nucleic acid-binding activity and genes involved in the cellular response to stress
from Gene Ontology for the second screen, we hoped to address this latter limitation. Another
issue faced was that since little is known about Clr-b (Clec2d) regulation, the only positive
control shRNA examined were those targeting Clr-b itself, which only had 1 shRNA out of 5 that
causedsignificant knockdown at the protein and transcript levels (Fig. 3.8). Perhaps the most
difficult issue to overcome, was the fact that Clr-b expression can be variable depending on cell
cycle and confluence. We tried to address this by strictly adhering to a consistent infection,
trypsinization, and plating protocol to minimize variations in Clr-b expression. Finally, the
analysis of the effects of the non-targeting negative control shRNA (GFP, RFP, etc) on Clr-b
expression was discouraging. We expected that any of these shRNA would be suitable negative
controls, but this did not prove to be the case, as most of these shRNA caused some up- or down-
regulation of Clr-b at the protein level. A possible explanation could be non-specific off-target
effects at the mRNA and protein levels. Off-target effects can be assessed using multiple unique
shRNA targeting the same gene, as each shRNA would have unique off-target effects, yet ideally
similar on-target effects. Another possibility is stressors induced by the experimental design,
including introduction of foreign lentiviral DNA into cells, which may induce a stress response,
or the stress of long-term puromycin selection, each of which may in turn result in the
downregulation of Clr-b expression. Quality of the lentiviral library and transduction conditions
may play a role as well, since poor lentiviral production or transduction efficiencies may kill the
cells during selection, or alter the confluence of the cells. Induction of the interferon response is
also possible using lentiviral vectors, but since these lentiviral vectors are non-replicating, any
effects of interferon should be transient, and negligible after 5-7 days of puromycin selection and regular changing of the culture medium, unless the shRNA hairpins themselves induce dsRNA recognition pathways.

With the experience from this screen and hindsight, other analyses in addition to the ones performed here may have been more helpful. One can calculate $q$-values and utilize them to determine false detection rates. This could be a more correct statistical approach given the size of the dataset obtained from the screens. With a $p$-value of 0.05, there is a 5% probability that each result may be a false positive. While 5% is deemed scientifically acceptable for a single test between two cohorts, with a larger number of tests approaching 80,000, one would expect ~4000 false positives by chance alone. This can be corrected by assigning an adjusted $p$-value to each test or reducing the $p$-value threshold from 5% to a value that better reflects the dataset of interest. A $q$-value is a $p$-value that has been corrected to take into account the false discovery rate, generated by incorporating the $p$-value distribution when determining significance. There are caveats to this approach as well, including: (i) the fact that statistical significance is not necessarily a valuable predictor of biological relevance; (ii) calculation of $q$-values is best done using advanced software and a strong background in statistics and screens of this nature; (iii) and finally $q$-value calculations here would treat each shRNA independently, while the lentiviral library contains multiple shRNA targeting a given gene.

Despite these difficulties and caveats, several TF genes validated, including TBP and the co-activator, CBP, as playing a role in the normal expression of Clr-b in healthy cells. This suggests that Clec2d is a TATA-driven gene, although further investigation is needed to confirm that TBP functionally binds to the Clec2d promoter (for example, by TBP ChIP). It would be possible that downregulation of Clr-b expression may be simply a consequence of reduced TBP, CBP, and RNAPII binding to the Clec2d promoter; other assays of interest include ATAC-Seq, FAIRE-Seq, or DNAse-Seq datasets, which survey cells for open chromatin regions in various cell types in response to a variety of reagents.

MeCP2 is another novel and interesting factor in regulating Clec2d expression. MeCP2 usually binds to methylated CpG dinucleotides, and there is a predicted CpG island in the 5'-
upstream region of the Clr-b gene locus, ~10Kb upstream of the TSS. In addition to its DNA-binding domain, MeCP2 also possess a transcriptional repression domain, which binds to the corepressors, Sin3A, N-CoR, and c-Ski, to recruit histone deacetylases [358, 359]. ChIP-chip experiments have revealed that MeCP2 can also occupy active promoters, where it mostly binds non-methylated sites in between genes that are >10Kb away from TSS [360]. In addition, MeCP2 is interesting because of its known disease associations, as mentioned previously. A discussion on potential avenues of future experiments regarding these transcription factors can be found in section 5.3.
Chapter 4

Interferon-dependent induction of Clr-b during MCMV infection protects bystander cells from NK cells via NKR-P1B-mediated inhibition

Christina L. Kirkham\textsuperscript{a}, Oscar A. Aguilar\textsuperscript{a}, Tao Yu\textsuperscript{b}, Miho Tanaka\textsuperscript{a}, Aruz Mesci\textsuperscript{a}, Kuan-Lun Chu\textsuperscript{a}, Jason H. Fine\textsuperscript{a}, Karen L. Mossman\textsuperscript{c}, Rod Bremner\textsuperscript{b}, David S. J. Allan\textsuperscript{a,d}, and James R. Carlyle\textsuperscript{a,e}

\textsuperscript{a}Department of Immunology, University of Toronto, Sunnybrook Research Institute, 2075 Bayview Ave., Toronto, ON, M4N 3M5, Canada.
\textsuperscript{b}Department of Laboratory Medicine & Pathobiology, University of Toronto, Lunenfeld-Tanenbaum Research Institute, 600 University Ave., Toronto, ON, M5G 1X5, Canada.
\textsuperscript{c}Department of Pathology and Molecular Medicine, McMaster University, McMaster Immunology Research Centre, 1200 Main Street West, Hamilton, ON, L8N 3Z5, Canada.
\textsuperscript{d}Current Address: National Heart, Lung, and Blood Institute, National Institutes of Health, Bethesda, MD, USA

This work has been accepted for publication at the Journal of Innate Immunity with the exception of Figures 4.1 and 4.2, which are unpublished.

Tao Yu performed the ChIP experiments
Oscar A. Aguilar and Aruz Mesci propagated, purified, and titred the viruses
Aruz Mesci performed the type-I IFN pre-treatment with MCMV-GFP infection

Oscar A. Aguilar and Miho Tanaka performed the WT and Stat1\textsuperscript{−/−} mouse experiments

Kuan Lun Chu performed the genotyping PCR assays on MEF gDNA

All other experiments and replicates were designed and performed by C.L.K.
Chapter 4:

4.1 Introduction

Natural killer (NK) cells are an important component of innate immune responses to pathological target cells, including tumour, virus-infected, antibody-coated, transplanted, and “stressed” cells. NK cell effector functions include cellular cytotoxicity, mediated by exocytosis of preformed granules (containing perforin, granzymes, granulysin), expression of apoptosis-inducing surface molecules (such as Fas-L, TRAIL), cytokine secretion (most notably IFN-γ, TNF-α, GM-CSF), and chemokine responses [5]. Recently, NK cells have been reclassified as a subset of group-1 innate lymphoid cells (ILCs) that share with ILC1 both T-bet (Tbx21) expression and IFN-γ secretion, yet also possess cytotoxic function and enhanced expression of eomesodermin (Eomes) [361, 362]. NK cells distinguish between healthy and pathological target cells through a complex integration of signaling events mediated by inhibitory and stimulatory cell surface receptors, which in turn recognize cognate ligands either downregulated or induced on target cells under surveillance [361].

In mice, self-specific NK cell receptors (NKR) include the Ly49 (Klra) family of inhibitory and activating receptors that mainly recognize classical MHC-I molecules, the bi-functional CD94/NKG2 (Klrd1/Klrc) family that recognize non-classical MHC-I molecules, the stimulatory NKG2D (Klrk1) receptor that recognizes MHC-I-related stress ligands, and the NKR-P1 (Klrb) family of inhibitory and stimulatory receptors that recognize C-type-lectin-related (Clr/Clec2) glycoproteins [80]. All of these receptors are encoded within the NK gene complex (NKC) located on mouse chromosome 6, rat chromosome 4, and human chromosome 12 [63], yet a number of loci encoding other NKR, including NKp46 (Ncr1) and 2B4 (Cd244) are located outside the NKC (reviewed in [363]).

Among this diversity, the NKR-P1 receptor family is somewhat conspicuous in being genetically linked to its cognate Clr ligand family, akin to one other MHC-independent recognition system, CD244:CD48; this arrangement may ensure co-inheritance of self-specific receptor-ligand interactions [79, 81, 82]. In the mouse, five functional NKR-P1 receptors have
been identified, including the activating receptors, NKR-P1A (*Klrb1a*; unknown ligand), NKR-P1C (NK1.1; *Klrb1c*; unknown ligand), and NKR-P1F (*Klrb1f*; recognizes Clr-c,d,g), as well as the inhibitory isoforms, NKR-P1B/D (*Klrb1b*; recognizes Clr-b), and NKR-P1G (*Klrb1g*; recognizes Clr-d,f,g) [84]. At least one other receptor pseudogene locus is annotated (NKR-P1E; *Klrb1-ps1*), while other Clr loci with unknown or pseudogene function also exist (Clr-a,e,h,i,j) [114, 115]. In humans, only a single inhibitory NKR-P1A receptor exists (CD161/*KLRB1*), genetically linked to its cognate ligand (LLT1/*CLEC2D*); however, related stimulatory receptors are also encoded within the NKC, tightly linked to their respective ligand loci, including NKp80 (*KLRF1*; ligand, AICL/*CLEC2B*) and NKp65 (*KLRF2*; ligand, KACL/*CLEC2A*) [86].

The inhibitory NKR-P1B:Clr-b interaction is currently the most well-characterized recognition pair. Like the MHC-specific Ly49 receptors, NKR-P1B:Clr-b interactions are involved in both NK cell education and target cell recognition [41, 42]. Clr-b (*Clec2d*) is a type-II transmembrane C-type lectin-like glycoprotein expressed on most hematopoietic and some non-hematopoietic cells, in a pattern similar to that of MHC-I molecules [81]. However, under pathological conditions such as oncogenesis, virus-infection, and genotoxic stress, Clr-b is rapidly downregulated at both the cell surface and steady-state transcript levels, rendering these cells more sensitive to NK cell-mediated “missing-self” recognition [40, 81, 107, 145, 330]. The regulatory mechanisms governing modulation of Clr-b expression remain to be elucidated, although previous work has suggested roles for transcriptional, post-transcriptional, as well as ubiquitin-proteosomal and endolysosomal control [40, 107, 145, 330].

To investigate the genomic control of the Clr-b/Clec2d gene at the promoter and nascent transcript levels in healthy versus virus-infected cells, we used MCMV as a model pathogen. MCMV is a β-herpesvirus with a large double-stranded DNA genome capable of accommodating numerous immunoevasin genes that subvert host immune responses. Previous studies have shown that MCMV, RCMV-E, and Vaccinia virus infections promote a rapid loss of mouse Clr-b/Clec2d and rat Clr-11/Clec2d11 on fibroblasts [107, 145, 330]. Interestingly, at early time points during MCMV infection *in vitro*, uninfected fibroblasts actually upregulate Clr-b expression, as do cells exposed to passaged viral supernatants. This reciprocal regulation may
represent a means to ensure optimal self-nonself discrimination between uninfected “bystander” cells in the vicinity of infected “missing-self” targets.

Here, we demonstrate that MCMV-mediated downregulation of Clr-b steady-state transcripts is controlled by disruption of Clec2d promoter activity, mediated at least in part by the cell-autonomous action of the MCMV ie3 gene product in trans. In contrast, Clr-b upregulation on uninfected bystander cells is driven by paracrine type-I interferon (IFNαβ), in a manner that is dependent upon IFNAR1 signaling and occupancy of the Clec2d promoter in cis by a complex containing IRF9, STAT1, and STAT2, most likely the ISGF3 heterotrimer (IRF9/STAT1/STAT2). Discerning how NKR ligands are regulated on both healthy and pathological target cells is an important facet in further understanding NK recognition and harnessing NK cell activity in disease therapy.

4.2 Results

4.2.1 Identification of 2 transcriptional start sites and an alternative exon in the Clec2d gene

In order to begin analyses on the transcriptional regulation of Clr-b, we wanted to be confident in the published transcriptional start site (TSS) utilized by RNAPII to transcribe Clec2d. A previous report had identified a single G as the TSS (called +1) in bone marrow and spleen cells by 5’-RACE [113] (Fig. 4.1). For a more complete understanding of the TSS of Clec2d and to identify any potential alternative splice isoforms, 5’-RACE was performed on numerous cell lines with varying degrees of Clr-b expression (high or low), as well as ex vivo cells, using a reverse primer located within the last exon (exon-5).

In agreement with the previous report, Clec2d from all cells tested was found to be transcribed from a main G (+1) TSS (Fig. 4.1). In addition, however, some cells, including NIH3T3, BWZ.36, C1498, Swiss-strain fibroblasts, and RMA had at least 1 transcript with a TSS within the 5’-end of intron-1.
The human CLEC2D gene (which encodes the human NKR-P1A-ligand, LLT1) is known to give rise to several additional splice variants with unknown function. To investigate the hypothesis that the same is true for mouse Clec2d, the clones obtained from 5'-RACE were aligned to the mouse genome using the UCSC Genome Browser’s Blat function and default settings. This analysis identified 2 main splice variants (Fig. 4.1). One variant lacks exon-3; this transcript was found in the tumour cell lines, C1498, EL4, BWZ.36 (in which this was the only Clec2d detected), as well as fresh ex vivo splenocytes and thymocytes. Exon-3 codes for the transmembrane domain of Clr-b, suggesting these transcripts would not be capable of being expressed at the cell surface upon translation. In fact, the published Ocil/Clr-b−/− mouse was generated by targeted removal of exon-3, resulting in a frameshift mutation in the stalk region, and Clr-b mAb staining of cells from these mice has demonstrated a complete lack of cell surface Clr-b protein [41, 357]. The second splice variant lacks the standard exon-1 and instead generates an alternative exon-1 within intron-1 (Fig. 4.1). This variant was present in C1498, NIH3T3, RMA, ex vivo splenocytes, and Swiss-strain ear fibroblasts. The largest ORF from this transcript is only 477 bp (159 residues) and would generate a protein lacking the entire cytosolic domain and 6 amino acids within the transmembrane domain.
Figure 4.1. Mapping of the mouse Clec2d transcriptional initiation site by 5' rapid amplification of cDNA ends. The nucleotide sequence for the genomic portion of the 5'-flanking region, exon 1, and part of intron 1 is shown. Underlined nucleotides correspond to exon 1 with the ATG translation start in bold italics. Lower case nucleotides correspond to a portion on intron 1. The previously identified transcriptional initiation site is in bold (+1). The symbols correspond to the unique sites identified in different cell lines and ex vivo cells. Unique clones from each cell line were aligned to the mouse genome using UCSC genome browser build mm10 to identify splice variants.
4.2.2 Promoter activity of the Clec2d 5’ flanking region

The functional promoter activity of the Clec2d upstream region was assessed by transient and stable expression of luciferase reporter constructs (Fig. 4.2). Genomic fragments of various sizes of the Clec2d upstream region were cloned into the pGL3 (transient) or pGL4.22 (stable) vectors (Fig. 4.2A). Since cis-acting regulatory elements can be found downstream of the TSS, a few constructs were created that contain genomic DNA downstream of the TSS (Fig. 4.2A). To examine promoter activity in “normal” versus “cancerous” cells, the constructs were introduced transiently into NIH3T3 and C1498 cells, respectively, and the luciferase activity in cell extracts was measured (Fig. 4.2B). The pGL3-Basic vector alone was used as a negative control, while pGL3 containing the SV40 promoter was used as a positive control for promoter activity, and an HSV-TK Renilla (NIH3T3) or CMV Renilla (C1498) vectors were used to normalize transfection efficiency. As expected, the SV40 control produced a large induction in luciferase activity in both cell lines, compared to pGL3 empty vector (Fig. 4.2B). In both cell lines, the Clec2d upstream region yielded similar luciferase activity to the SV40 promoter. In NIH3T3 cells, the various sizes of the Clec2d upstream region all yielded similar promoter activity, with the exception of the 5Kb, 5’-UTR-intron-1, and 500bp-intron-1 constructs, which drove less expression of luciferase (Fig. 4.2B). A similar trend was seen in the C1498 cell line, where the 4-5Kb, and 9Kb constructs drove lower expression, and the 5’-UTR-intron-1 and 500bp-intron-1 constructs did not drive any luciferase expression (Fig. 4.2B). To identify a core promoter capable of driving gene expression, smaller upstream regions from 100-500bp were cloned into the pGL3 vector and assessed in NIH3T3 and C1498 cell lines. Interestingly, in NIH3T3 cells, the first 100bp upstream from the TSS is sufficient to drive some luciferase expression (although the first 200bp drives enhanced expression), while in C1498 cells, the first 200bp upstream from the TSS was necessary for luciferase expression (Fig. 4.2C).

To assess the function of these promoter constructs in the context of euchromatin, NIH3T3 cells were stably transfected with the promoter fragments in pGL4.22, selected with puromycin, and assessed for luciferase activity. To normalize for differences in cell numbers, the values were normalized to the total protein content of the samples. Importantly, all constructs tested functioned within chromatin (Fig. 4.2D). Note that the 7Kb and 9Kb constructs were not
assessed further as their transfectants did not produce luciferase, despite having survived puromycin selection.

With this basic understanding how the *Clec2d* promoter fragments function in resting NIH3T3 cells, we extended these analyses to address *Clec2d* and Clr-b regulation during viral infection in more detail.
Figure 4.2 Functional analysis of the Clec2d 5’ flanking region. (A) Cloning strategy for the creation of Clec2d upstream regulatory sequences for luciferase reporter assays. (B-C) Luciferase reporter assays comparing the relative activities from the fragments in (A). Fragments were cloned into the pGL3 basic vector in front of the luciferase gene and assayed for luciferase activity upon transfection in NIH3T3 and C1498 cells. The mean ± SEM from at least 3 independent experiments is shown. Bars represent the fold induction relative to that obtained from the Renilla vector and indexed to the pGL3 empty vector. (D) Luciferase reporter assay in NIH3T3 cells stably expressing the constructs in (A) in the pGL4.22 vector. The mean ± SEM from at least 3 independent experiments is shown. Bars represent the fold induction relative to total protein and indexed to the pGL3 empty vector.
4.2.3 MCMV infection reciprocally modulates Clr-b levels on infected and bystander cells

Previous studies have shown that infection of various mouse and rat cells with a number of viruses (MCMV; RCMV-E; Vaccinia; Ectromelia) promotes a rapid loss of mouse Clr-b (Clec2d) and the rat Clr-b homolog, rClr-11 (Clec2d11), at both the steady-state transcript and cell surface levels [107, 145, 330]. To distinguish between infected and uninfected (bystander) cells at the single cell level, we infected mouse NIH3T3 fibroblasts over an early time course using a modified MCMV-GFP reporter virus, in which an enhanced GFP transgene is driven by an immediate early gene (ie1/3) promoter in the MCMV-Smith (VR-194) strain [145, 322, 323]. While we consistently observed loss of Clr-b surface expression on the infected population at later time points (GFP+; 12-24 hours post-infection; h.p.i.), at early time points, the uninfected bystander population expressed elevated Clr-b surface levels relative to mock-infected parental cells (GFP−; 3-12 h.p.i.; Fig. 4.3A). Notably, exposure of NIH3T3 cells to UV-irradiated viral supernatants still promoted early Clr-b upregulation (Fig. 4.3B), while infection using highly purified MCMV viral particles did not cause significant bystander Clr-b upregulation [145]. This suggested that a soluble mediator in viral supernatants may promote bystander Clr-b upregulation.

Time-course analysis of total steady-state mRNA levels by qRT-PCR confirmed (at the population level) that MCMV-infected NIH3T3 cells initially induced then downregulated Clr-b transcripts relative to mock-infected cells (Fig. 4.3C). Collectively, this suggests that MCMV infection may differentially modulate Clr-b levels in infected and bystander cells, whereby early induction of Clr-b may be due to exogenous factors produced upon MCMV infection acting in trans.

4.2.4 Bystander Clr-b induction during MCMV infection is type-I IFN-dependent

We hypothesized that the induced levels of Clr-b protein and Clec2d transcripts in uninfected bystander cells might be due to paracrine type-I IFN cytokine stimulation. To test this, we exposed primary murine embryonic fibroblasts (MEF) from both wild-type (WT) B6 and
Ifnar1−/− mice to IFN-α4 then examined Clr-b levels by flow cytometry. Treatment of WT MEF with IFN-α4 upregulated Clr-b (~2-fold, at low dose; ~6-fold, at high dose), while Ifnar1−/− MEF showed no change in Clr-b levels (Fig. 4.3D). Thus, Clr-b levels are upregulated on primary MEF cells in response to type-I IFN in an IFNAR1-dependent manner.

Since type-I IFNs function to promote an anti-viral state, we hypothesized that pre-conditioning cells with type-I IFN would block MCMV infection-mediated Clr-b loss. NIH3T3 cells were pre-treated overnight with a high dose (10⁴ U/mL) of IFN-α4 prior to infection. Notably, IFN-α4 pre-exposure did not prevent MCMV-GFP virus infection (Fig. 4.3E). In addition, while an IFN-mediated induction of Clr-b was observed on all cells, infected (GFP⁺) cells still displayed reduced Clr-b levels in comparison to uninfected (GFP⁻) cells. This suggests that the IFN-induced antiviral state is insufficient to prevent wild-type MCMV infection in vitro, and that infection-mediated Clr-b loss is IFN-independent. To probe this reciprocal regulation further, we examined genomic control, at the level of the Clr-b (Clec2d) promoter.
Figure 4.3 Reciprocal modulation of Clr-b expression on MCMV-infected and IFN-stimulated bystander cells. NIH3T3 cells were infected with MCMV-GFP supernatants at an MOI of 0.5 PFU/cell over an extended time course. (A) Cells were analyzed by flow cytometry for Clr-b surface expression. Numbers represent median fluorescence intensity (MFI). (B) NIH3T3 cells were infected with UV-inactivated MCMV-GFP for 24 h followed by analysis for Clr-b expression by flow cytometry. Bars represent the mean MFI ± SEM normalized to Mock. (C) Quantitation of steady-state Clec2d transcripts in (A) by qRT-PCR normalized to Tbp expression and 0 h. Significance was determined by two-tailed t-test (n=3). (D) Flow cytometric analysis of WT and Ifnar1−/− primary MEF cells with or without IFN-α4 treatment for 24 h. Shaded histograms represent Clr-b expression; dotted lines represent secondary reagent alone; vertical dotted lines show MFI of untreated cells. Numbers represent MFI. (E) NIH3T3 cells were exposed to medium alone (Untreated) or 10^4 U/mL IFN-α4 overnight, followed by MCMV-GFP infection at an MOI of 1 PFU/cell or mock infection for 24 hrs. Representative of at least 3 independent experiments.
4.2.5 Reciprocal modulation of Clec2d promoter by MCMV infection and type-I IFN

To elucidate the role of the Clec2d promoter, independent of mRNA stability, we performed intronic qRT-PCR to assess nascent (pre-mRNA) Clec2d transcript levels, since intron removal during mRNA splicing usually occurs co-transcriptionally. Clec2d is a relatively compact gene, consisting of 5 known exons separated by 4 introns; therefore, the nascent abundance of introns-1-4 (i1-i4) was quantitated during MCMV infection, which we predicted would decrease nascent transcript abundance, relative to healthy cells. Following overnight MCMV-GFP infection, nascent Clec2d transcript levels were reduced ~5-fold relative to mock-infected cells (Fig. 4.4A; MCMV-GFP vs. Mock). In contrast, following overnight IFN-α4 treatment, a ~2-3 fold increase in nascent Clec2d transcripts was observed (Fig. 4.4B; IFN-α4 vs. Untreated). Thus, MCMV infection promotes a significant decrease in nascent Clr-b transcript levels, while type-I IFN treatment causes a significant increase, suggesting that Clec2d promoter activity is reciprocally modulated in infected and bystander cells.
Figure 4.4 MCMV infection and type-I IFN oppositely regulate *Clec2d* nascent transcript levels. (A) NIH3T3 cells were infected with MCMV-GFP at an MOI of 0.5 PFU/cell or (B) treated with $10^3$ U/mL IFN-α4 for 24 h. Levels of *Clec2d* nascent pre-mRNA transcripts across introns 1-4 were quantified by qRT-PCR and normalized relative to (a) *Gapdh* or (b) *Hprt* and *Tbp*. Significance was determined by two-tailed t-test (n=3-5 experiments).
4.2.6 The *Clec2d* promoter is regulated by a functional cluster of IRF binding sites (IRFC)

Luciferase reporter assays were employed to test the hypothesis that infected and bystander cells would display differential *Clec2d* promoter activity. To this end, *Clec2d* promoter fragments of various sizes were subcloned from a BAC vector into luciferase reporter vectors (pGL3 for transient, and pGL4.22 for stable transfectants) (Fig. 4.5A). NIH3T3 transfectants were subsequently exposed to either MCMV or IFN-α4 overnight and assayed for luciferase activity. As observed at the transcript level, MCMV-infection caused a ~4-5-fold decrease in *Clec2d* promoter activity, relative to empty pGL4.22 vector, in transfectants containing the ≥300bp upstream of the *Clec2d* transcriptional start site (TSS) (Fig. 4.5B). In contrast, IFN-α4 treatment resulted in a ~3-fold increase in *Clec2d* promoter activity, relative to empty pGL3 vector or the SV40 promoter, in transfectants containing ≥200bp upstream of the *Clec2d* TSS (Fig. 4.5C). Collectively, this suggests that MCMV-mediated downregulation and IFN-mediated induction of *Clec2d* are regulated by distinct regulatory regions, and that DNA element(s) within the first ~200bp upstream of the *Clec2d* TSS are required for IFN-mediated Clr-b induction.

We next analyzed the *Clec2d* promoter region to identify putative transcription factor binding sites responsible for Clr-b regulation. Several interferon regulatory factor (IRF) binding sites were predicted within the ~200-300bp promoter responsible for IFN-mediated induction, which we predicted to be functional (Fig. 4.5A,C). A putative IRF4 binding site was identified within the first 100bp, although this region is not sufficient to augment *Clec2d* promoter activity (Fig. 4.5A,C). Importantly, the 200bp promoter fragment, which is sufficient to induce *Clec2d* in response to IFN-α4, contains a cluster of overlapping binding sites for IRF3, IRF7, and IRF9 (see Fig. 4A for IRF cluster (IRFC) nucleotide sequence); in addition, another IRF3 binding site was predicted within the 300bp fragment (Fig. 4.5A). To investigate the hypothesis that the ~200bp IRF cluster (IRFC) was necessary for type-I IFN induction, we mutated the IRF3/7/9 consensus motifs within the 500bp promoter fragment (mut500bp) to sequences devoid of transcription factor activity (as determined by MatInspector and TfSearch). In contrast to the intact 500bp promoter fragment, the mut500bp fragment failed to augment *Clec2d* promoter
activity in response to IFN-α4, indicating this region is necessary for Clr-b induction (Fig. 4.5C). However, reduced Clec2d promoter activity was still observed during MCMV infection using the mut500bp fragment (Fig. 4.5B), suggesting that virus infection more broadly affects the Clec2d promoter. Interestingly, a recent study examined whole-genome RNA polymerase II (RNAPII) occupancy by ChIP-Seq in MEF cells during HSV-1 infection [364]; analysis of the Clec2d promoter from these data revealed that RNAPII was essentially absent following HSV-1 infection (Fig. 4.6). This supports our earlier hypothesis that loss of Clr-b at the cell surface may result simply from failure to recruit RNAPII and TBP to the Clec2d promoter. Thus, loss of Clec2d promoter activity and nascent Clr-b transcripts may be a generalized response to herpesvirus infection, and perhaps other viruses [107, 330, 364].
Figure 4.5 MCMV infection and type-I IFN reciprocally regulate Clec2d promoter activity. (A) Top: Schematic representation of the Clec2d upstream regulatory region showing selected consensus motifs; ATG, translational start site; +1, transcriptional start site (TSS); TBP, TATA binding protein; IRF, Interferon regulatory factor; Sp, specificity protein. Bottom: Nucleotide sequence of the cluster of predicted IRF-binding sites (IRFC) within the first 200bp of the Clec2d promoter. The upstream GAAA motifs highlighted in bold may enhance STAT1 (and thus ISGF3) recruitment to the IRF9 motif, in turn minimizing a requirement for the IRF3/7 motifs in response to type-I IFN [365] (B-C) NIH3T3 cells were transfected with luciferase-reporter plasmids containing varying sizes of the Clec2d promoter region as outlined in A). (B) Stable transfectants were infected with MCMV at an MOI of 0.5 PFU/cell for 24 h. (C) Transient transfectants were treated with 10^3 U/mL IFN-α4 for 24 h the next day. Promoter activity was assayed by luciferase reporter assay relative to (B) total protein and empty vector (pGL4.22) or (C) Renilla and empty vector (pGL3). Significance was determined by two-tailed t-test on log-transformed values (n=3–7 experiments)
Figure 4.6. Herpes simplex virus (HSV)-1 infected MEFs have reduced RNA Polymerase II (RNAPII) occupancy at the Clec2d promoter. RNAPII ChIP-seq data from Mock (GSM1623231) and 4hr HSV-1 infected (GSM1623232) MEF cells were mapped onto the mouse genome (mm10 assembly) on UCSC genome browser. GEO Series accession number GSE66487. Experiments were performed by Abrisch and others, 2015.
Next, we examined the IRFC in more detail to test the hypothesis that the IRFC is sufficient to mediate IFN-inducible expression in a minimal promoter/enhancer assay. In transient NIH3T3 transfectants, the IRFC alone provided limited IFN-dependent induction in the forward orientation only, although this was only significant using a one-tailed $t$-test (Fig. 4.7). More importantly, the IFN-nonresponsive 100bp promoter fragment became IFN-responsive upon upstream addition of a forward-oriented IRFC (Fig. 4.7). However, providing an additional IRFC to the IFN-responsive 200bp promoter fragment did not further increase expression, nor did addition of an intact IRFC upstream of the mut500bp promoter fragment (Fig. 4.7). Taken together, these orientation/proximity-dependent results suggest that the IRFC acts as part of an IFN-inducible promoter but not an independent enhancer. It remains possible that the additional IRF3 site may contribute to this context dependence.
Figure 4.7 An overlapping IRF3/7/9 consensus cluster (IRFC) is required for type-I IFN-dependent promoter activity. NIH3T3 cells were transfected with luciferase reporter vectors then treated with 10^3 U/mL IFN-α4 for 24 h the next day. Promoter activity was assayed by dual luciferase reporter assay relative to Renilla luciferase and normalized to pGL3 empty vector. Significance was determined by one-tailed (^p) or two-tailed (^2p) t-test on log-transformed values (n=3 experiments).
4.2.7 Requirement for IFNAR1, IRF9, STAT1 and STAT2 in *Clec2d* induction by type-I IFN

To examine the mechanism of IFN-induced Clr-b induction further, we utilized MEF cells deficient in type-I IFN signaling pathway components. To this end, MEF cells from WT, *Ifnar1*\(^{-/-}\), and various IRF-deficient mice were exposed to IFN-α4 or MCMV and examined for Clr-b levels by flow cytometry. As shown previously, IFN-α4 upregulated Clr-b on WT but not *Ifnar1*\(^{-/-}\) MEF (*Fig. 4.8A*), demonstrating a requirement for IFNAR1 signaling. IFN-dependent Clr-b induction was also observed on *Irf3*\(^{-/-}\) and *Irf7*\(^{-/-}\) MEF, but not on *Irf3*\(^{-/-}\)Irf9\(^{-/-}\) or *Irf9*\(^{-/-}\) MEF, demonstrating a requirement for IRF9 function (*Fig. 4.8A,C*).
Figure 4.8 Type-I IFN-dependent induction of Clr-b is IRF9-dependent. (A) Primary MEF cells derived from WT, Ifnar1−/−, Irf3−/−, Irf9−/−, and Irf9−/− mice were treated with 10^3 U/mL IFN-α4 for 24 h and analyzed for Clr-b cell surface expression by flow cytometry. Dotted lines represent secondary reagent alone; solid lines represent untreated cells; shaded histograms represent treated cells. (B) MEF cells as in A) were transfected with luciferase reporter constructs and analyzed by dual luciferase reporter assay. Graphs show fold luciferase induction relative to Renilla and normalized to pGL3 empty vector. Significance was determined by two-tailed t-test on log-transformed values (n=3-5 experiments). (C) MEF cells were transfected with overexpression vectors containing IRF3, IRF7, and/or IRF9 for 24 h, then left untreated or treated with 10^3 U/mL IFN-α4 for an additional 24 h before analysis of Clr-b expression by flow cytometry. Histograms are gated on GFP^+ cells. Dotted lines represent secondary reagent alone; solid lines represent untreated cells; shaded histograms represent IFN-treated cells. Representative of at least 4 independent experiments.
On the other hand, upon MCMV-GFP exposure, infected (GFP⁺) MEF cells downregulated Clr-b independent of genotype, and WT bystander (GFP⁺) MEF cells routinely induced Clr-b, yet Irf3+/Irf9⁻⁻ and Irf9⁻⁻ bystander (GFP⁻) MEF cells did not upregulate Clr-b (Fig. 4.9). Results using Irf3⁻⁻ and Irf7⁻⁻ MEF were not as clear, with bystander cells inducing Clr-b in some but not all experiments (dependent upon resting Clr-b levels and viral MOI; Fig. 4.9). Taken together, this suggests that IFN-mediated Clr-b induction requires IRF9, while MCMV-mediated Clr-b downregulation is more complex and appears to be IRF1/3/7/9-independent.
Figure 4.9 Expression of Clr-b on MEF cells during viral infection. MEF cells were seeded and infected with MCMV-GFP at an MOI of 1 (MCMV-GFP 1) or 3 (MCMV-GFP 2). Clr-b expression was analyzed by flow cytometry 18 h.p.i. Cells were gated based on forward and side scatter and lack of propidium iodide staining. Numbers at the top of the gates represent the MFI.
We next used mutant MEF cells to re-analyze Clec2d promoter activity via luciferase reporter assays. As shown previously for NIH3T3 cells, the 500bp Clec2d promoter fragment augmented luciferase activity in response to IFN-α4 in WT and Irf3−/− MEF cells, while the mut500bp fragment did not (Fig. 4.8B). In contrast, neither the intact 500bp nor mut500bp promoter fragments augmented luciferase activity in response to IFN-α4 treatment in Ifnar1−/−, Irf3−/−Irf9−/− or Irf9−/− MEF transfectants (Fig. 4.8B). This suggests that both IRF9 and the IRF3/7/9 cluster, in addition to intact IFNAR1 signaling, are required for IFN-mediated Clec2d induction. Importantly, complementation studies using mutant MEF cells and overexpression of their deleted gene products demonstrated that re-introduction of IRF9 into Irf3−/−Irf9−/− and Irf9−/− MEF cells re-established IFN-dependent Clr-b induction, further confirming the importance of IRF9 (Fig. 4.8C).

IRF9 commonly activates transcription of ISG by forming a complex with STAT1 and STAT2 (termed ISGF3). To test the hypothesis that ISGF3 binds to the Clec2d promoter, we analyzed the IFN-α4 responses of primary cells from Stat1−/− and WT control (129S6-strain) mice. Notably, Clr-b induction on Stat1−/− splenocytes (SPL), bone marrow cells (BMC), and primary adult ear fibroblasts (AEF) was abrogated in response to IFN-α4 treatment, while the same tissues from strain-matched (129S6) control WT mice upregulated Clr-b in response to IFN-α4 (1.8-fold, 3.1-fold, and 2.4-fold, respectively; Fig. 4.10A). Furthermore, two independent STAT1 inhibitors, nifuroxazide (which inhibits STAT1/3/5 phosphorylation by Jak2/Tyk2) and fludarabine (which specifically depletes STAT1 protein and mRNA) [366-368], also blocked Clr-b upregulation upon IFN-α4 treatment of NIH3T3 cells (Fig. 4.10B). Taken together, these results suggest that type-I IFN-mediated Clr-b upregulation is dependent upon STAT1. Notably, it has been previously suggested that GAAA sequences localized upstream of IRF9 binding sites help to facilitate the recruitment of STAT1 and IRF9 to ISGF3-dependent interferon-stimulated response elements (ISRE) [365]; thus, the enrichment of GAAA sequences in the IRFC (Fig. 4.5A) may explain both the STAT1-dependence and IRF3/7-independence of this cluster.
Figure 4.10 Type-I IFN-mediated Clr-b induction is dependent on STAT1. (A) Splenocytes (SPL), bone marrow cells (BMC), and primary adult ear fibroblasts (AEF) were harvested from WT and Stat1−/− mice, treated with 10³ U/mL IFN-α4 for 6 h and analyzed by flow cytometry for Clr-b expression. (B) NIH3T3 cells were treated with 50μM vehicle alone (DMSO), Nifuroxazide, or Fludarabine for 22 h, at which point 10³ U/mL IFN-α4 was added for 3 h prior to flow cytometric analysis of Clr-b expression. Dotted lines, secondary reagent alone; solid line, cells not treated with IFN-α4; shaded histogram, IFN-α4-treated cells. Numbers reflect the MFI of untreated cells (left) or IFN-α4-treated cells (right); the lower number (x) represents the fold change in Clr-b MFI (IFN-α4 treated/untreated).
To demonstrate a role for STAT2 in type-I IFN-mediated *Clec2d* induction, we overexpressed the MCMV immunoevasin, M27, known for its ability to target and degrade STAT2 [369], in NIH3T3 cells. Strikingly, M27 overexpression resulted in a complete loss of STAT2 protein, in turn preventing CLR-b upregulation in response to IFN treatment; thus, STAT2 is necessary for IFN-dependent CLR-b induction (Fig. 4.11A,B). To formally show recruitment of STAT2 to the *Clec2d* promoter, we performed chromatin immunoprecipitation (ChIP) experiments on IFN-treated NIH3T3 cells. Here, IFN-α4 treatment was reduced to ~1.5h, a time-point reflecting an optimal increase in nascent *Clec2d* transcript levels prior to protein induction (Fig. 4.12A,B). Interestingly, an enrichment of *Clec2d* promoter gDNA (containing the IRFC) was observed upon ChIP of IFN-treated NIH3T3 cells using a STAT2 mAb (Fig. 4.11C), while this was not observed for downstream *Clec2d* exon-5 gDNA. In addition, similar results were observed using stable Dox-inducible NIH3T3.M27 transductants [321], whereby Dox induction of M27 expression mitigated the IFN-dependent recruitment of STAT2 to the *Clec2d* promoter containing the IRFC (Fig. 4.11D). This demonstrates that STAT2 is directly and specifically recruited to the *Clec2d* promoter IRFC element upon type-I IFN treatment. Notably, attempts to perform ChIP on mouse cells using multiple anti-IRF9 antibodies and Flag-tagged IRF9 failed, due to specificity and sensitivity problems. Nonetheless, the above results collectively demonstrate that the ISGF3 heterotrimer (IRF9/STAT2/STAT1) recruitment to the IRFC is likely responsible for CLR-b induction.
Figure 4.11 STAT2 is required for Clr-b induction following type-I IFN treatment. (A) NIH3T3 cells were transfected with a vector encoding the MCMV M27 immunoevasin or pIRES2-GFP empty vector, then treated with $10^3$ U/mL IFN-α4 for 24 h the next day. Clr-b expression was analyzed by flow cytometry, gated on GFP$^+$ cells. Dotted lines represent secondary reagent alone; solid lines represent untreated cells; shaded histograms represent IFN-treated cells. (B) Cells were transfected as in (A) then examined by Western blot for STAT2 expression the next day. (C) ChIP using an anti-STAT2 mAb was performed on NIH3T3 cells treated for 1.5 h with $10^3$ U/mL IFN-α4, using qRT-PCR primers spanning the IRFC of the Clec2d promoter or exon-5 (n=2 experiments). (D) NIH3T3 cells stably expressing M27 using a modified Dox-inducible piggyBac vector (NIH3T3.M27) were used to perform STAT2 ChIP as in C) (n=2 experiments).
Figure 4.12. Time course of Clec2d nascent transcript and protein levels during IFN treatment. NIH3T3 cells were seeded and treated with 10^3 U/mL IFN-α4 for the indicated times. (A) Cells were analyzed for Clec2d nascent transcript expression relative to Tbp and Hprt by qRT-PCR. N=4, significance determined by ANOVA with Bonferroni multiple comparisons test. (B) Representative flow plots showing Clr-b expression on cells from (A). Histograms represent the median fluorescent intensity of the staining with secondary reagent only (dashed line), untreated (solid black line), and treated (shaded) cells. Δ represents the fold change in Clr-b expression between the treated and untreated cells. Representative of 4 independent experiments.
4.2.8  IFN-dependent Clr-b induction augments NKR-P1B-ligand function

We predicted that increased Clr-b expression on the surface of target cells would translate to increased signaling on cells bearing the NKR-P1B receptor. To assess the functional consequences of Clr-b upregulation by type-I IFN, we utilized BWZ reporter cell assays. Here, BWZ.36 reporter cells expressing a CD3ζ-fusion receptor of the NKR-P1B ectodomain (BWZ.P1B cells) were used to specifically quantitate NKR-P1B-ligand (Clr-b) function on various stimulator cells in a colourimetric assay. Preliminary experiments using native BWZ.P1B cells suggested that reporter cell signaling was partially reduced by type-I IFN in a dose-dependent fashion (Fig. 4.13A,C), possibly due to altered signaling [370, 371] or threshold effects on the BWZ.36 (T-lineage) cell line. Therefore, we utilized CRISPR-Cas9 gene-editing technology to generate Ifnar1−/− BWZ.P1B and control BWZ(−) cells. IFNAR1 deficiency was confirmed for multiple BWZ clones by flow cytometry using IFNAR1 mAb (Fig. 4.13B). Notably, IFN-α4 treatment of NIH3T3 stimulator cells augmented NKR-P1B-ligand function using IFNAR1-deficient BWZ.P1B versus BWZ(−) cells (Fig. 4.14A and Fig. 4.13C). This signal was Clr-b-specific, as it was blocked using Clr-b mAb.
Figure 4.13. BWZ.36 reporter cell analysis of Clr-b expression. (A) Plate-bound PK136 (NK1.1) stimulation of BWZ.P1B and BWZ.P1C in the presence of increasing IFN. Graphs show mean ± SEM and ANOVA statistical analysis compared to no interferon treatment (not shown). (B) BWZ(−) and BWZ.P1B reporter cells were engineered with CRISPR-Cas 9 technology to be devoid of IFNAR1 expression. Histograms represent the median fluorescent intensity of the staining with secondary reagent only (dashed line), WT BWZ reporter cells (solid black line), and IFNAR1−/− BWZ CRISPR clone 7.1 (shaded) cells. (C) IFNAR1−/− BWZ.CD3ζ/NKR-P1B reporter cell analysis of ligand expression on IFN-α4-treated fibroblasts. NIH3T3 and MEF cells were used as stimulator cells upon co-culture with BWZ reporter cells in various concentrations of IFN-α4. Graphs show mean ± SEM.
To confirm a role for IRF9 in Clr-b induction, we repeated these assays using primary MEF from WT, *Irf9*<sup>−/−</sup> and *Irf3*<sup>−/−</sup>*Irf9*<sup>−/−</sup> mice as stimulator cells. Notably, treatment of WT MEF with IFN-α4 greatly augmented BWZ.P1B reporter activity, which was blocked using Clr-b mAb (Fig. 4.14B). In contrast, treatment of *Irf9*<sup>−/−</sup> or *Irf3*<sup>−/−</sup>*Irf9*<sup>−/−</sup> MEF with IFN-α4 yielded almost no BWZ.P1B responses, with minimal dose-dependent induction (Fig. 4.14C,D). Taken together, these results confirm that *Irf9*<sup>−/−</sup> MEF cells possess an intrinsic functional defect in IFN-dependent induction of the NKR-P1B-ligand, Clr-b.
Figure 4.14. BWZ.NKR-P1B reporter cell analysis of Clr-b ligand function on type-I IFN-treated stimulator cells. (A) NIH3T3 cells, (B) primary WT MEF cells, (C) primary Irf9<sup>−/−</sup> MEF cells, or (D) primary Irf3<sup>−/−</sup> Irf9<sup>−/−</sup> MEF cells were used as stimulator cells for Ifnar1<sup>−/−</sup> CD3ζ/NKR-P1B receptor-bearing BWZ.36 reporter cells, either untreated or treated overnight with 10<sup>3</sup> U/mL IFN-α4, in the absence or presence of blocking Clr-b mAb. Control Ifnar1<sup>−/−</sup> BWZ.36 cells are not shown. Experiments were analyzed using ANOVA with Bonferroni’s post-hoc analysis comparing treated to untreated. Representative of at least 4 independent experiments, graphs show mean ± SEM.
4.2.9 The MCMV ie3 gene product cell-autonomously represses *Clec2d* expression

In an attempt to understand mechanistically how MCMV infection downregulates Clr-b, we have tested a number of MCMV large genomic-deletion mutants, and cloned numerous MCMV gene products for independent overexpression. Importantly, neither m02 or m145 family immunoevasin genes nor several core genes were found to negatively regulate Clr-b levels in isolation (Oscar A. Aguilar, unpublished observations). However, overexpression of a single immediate-early gene product, *ie3*, but not *ie1* or *ie2*, consistently promoted a significant (≥50%) Clr-b downregulation when transiently transfected into NIH3T3 cells (Fig. 4.15A).

To further assess the effects of *ie3* on Clr-b levels, we generated stable Dox-inducible NIH3T3 transfectants of the alternately spliced *ie1/3* genes [321]. Following Dox induction, *ie3* was found to promote a significant (≥66%) loss of surface Clr-b, with minimal downregulation observed for *ie1* and empty vector controls (Fig. 4.15B). Since the function of immediate early genes is to activate the expression of viral early genes, we hypothesized that *ie3* may play a direct role in modulating Clr-b transcription as well. To examine the effect of *ie3* on *Clec2d* gene expression, we next quantitated *Clec2d* nascent and steady-state transcript levels by qRT-PCR. Here, Dox-induced *ie3* expression promoted a significant decrease in both *Clec2d* nascent and steady-state transcript levels (≥2-3 fold; Fig. 4.15C,D). To determine whether the *ie3* gene product directly repressed *Clec2d* promoter activity, dual-luciferase constructs of the promoter fragments were transfected into the Dox-inducible stable NIH3T3.IE1/3 transfectants. In line with Clr-b protein and transcript data, overexpression of *ie3* resulted in significantly decreased *Clec2d* promoter activity using the intact 500bp promoter construct, while *ie1* had no significant effect (Fig. 4.15E). Taken together, these findings suggest that the *ie3* gene product cell-autonomously represses the *Clec2d* promoter, in turn partially extinguishing *Clec2d* nascent transcripts and downregulating surface Clr-b protein early during MCMV infection.
Figure 4.15. Immediate early gene 3 (ie3) cell-autonomously promotes downregulation of Clr-b levels by repressing Clec2d promoter activity. (A) Vectors encoding the MCMV immediate early gene ORF (ie1,2,3) or empty pIRES2-GFP vector were transiently transfected into NIH3T3 cells, then analyzed for GFP and Clr-b expression after 24 h. Dotted lines represent secondary reagent alone; solid black lines represent Clr-b expression gated on GFP (untransfected) cells, and shaded histograms represent Clr-b expression gated on GFP + (transfected) cells. Numbers represent the fold change in Clr-b expression (GFP +/GFP – cells).

(B) NIH3T3 stable transfectants of Dox-inducible piggyBac vectors encoding the ie gene products or empty vector controls were induced for 2 days in 1.5 µg/mL Dox then analyzed by flow cytometry. Dotted lines represent secondary reagent alone; solid black lines represent Clr-b expression on untreated cells (–Dox); shaded histograms represent Clr-b expression on Dox-treated cells (+Dox). Numbers represent the fold change in Clr-b expression (+Dox/–Dox).

RNA from the cells in (A) was analyzed by qRT-PCR for Clec2d nascent pre-mRNA transcripts of introns 1-4 normalized relative to Tbp (n=3 experiments). Significance was determined by two-tailed t-test.

(D) RNA from the cells in (B) was analyzed for Clec2d steady-state levels by qRT-PCR relative to Tbp (n=3-4 experiments). (E) Cells as in (B) were transfected with luciferase-reporter vectors containing various Clec2d promoter fragments, then promoter activity was assayed by luciferase reporter assay relative to Renilla luciferase and normalized to empty vector control (pGL3). Significance was determined by two-tailed t-test on log-transformed values (n=3 experiments).
4.3 Discussion

This study provides mechanistic evidence that Clr-b (*Clec2d*) is an inducible interferon-stimulated gene (ISG), in addition to its demonstrated function as a “missing-self” marker of cell health that interacts with the NKR-P1B inhibitory receptor on NK cells and ILC subsets [82]. In keeping with this dual function, we propose that Clr-b enhances self-nonself discrimination by NK cells during viral infection, by two means: (i) MCMV-infected cells downregulate resting levels of Clr-b via a host pattern recognition mechanism, in order to render the “missing-self” cells more susceptible to clearance via NK cell disinhibition; while (ii) normal “bystander” cells in the vicinity of infected cells induce Clr-b as a marker of cell health in response to paracrine type-I IFN, thereby reinforcing NK cell inhibition via NKR-P1B. The net effect is that NK cells encounter a broader dynamic range of Clr-b-mediated inhibition/disinhibition during self-nonself discrimination, in turn rendering them better able to integrate signaling inputs and redirect cytotoxic machinery to recognize pathological versus healthy target cells in a battlefield of infection (Fig. 4.16).

This work has shown that both of the above responses are facilitated at the genomic level, where significant changes in *Clec2d* promoter activity lead to direct and immediate changes in nascent Clr-b transcript and protein levels. In this model, infected cells lose Clr-b in a manner partially attributable to *ie3*-mediated *Clec2d* repression, presumably involving loss of transcription factors and/or RNAPII occupancy at the *Clec2d* promoter (see below). Meanwhile, infected cells secreting type-I IFN may be sensed in a paracrine manner by uninfected bystander cells via the IFNAR1 receptor. Subsequently, canonical IFNAR1 signaling activates JAK1/TYK2 to phosphorylate and activate STAT1 and STAT2 to heterotrimerize and complex with IRF9 as ISGF3. An ISGF3-containing transcription factor complex is then recruited to the *Clec2d* promoter at the proximal ~200bp IRFC element (the IRF3/7/9 cluster, which acts as an ISRE), in turn augmenting Clr-b nascent transcripts and protein levels. This combined regulation mechanism is significant because it creates a greater disparity between the infected and uninfected cell populations, facilitating enhanced NK cell recognition.

On the other hand, in infected cells, pattern-recognition receptor (PRR) signaling events
following the detection of MCMV-specific pathogen-associated molecular patterns (PAMP, such as cytosolic nucleic acids), in combination with the effects of ie3 on host gene transcription (and/or unidentified immunoevasins), lead directly to the loss of Clec2d promoter occupancy, Clr-b transcripts and surface protein. While the complexity of this “missing-self” host response mechanism remains to be elucidated, the direct repressive effects of ie3 on host gene expression in general and Clec2d in particular, follow the tenets of a pattern recognition system, implicating the NKR-P1B:Clr-b interaction as a PRR axis, where the pattern itself is host transcriptional integrity.

In support of this, previous studies have shown that ie3 (M122) is critical for MCMV fitness, such that without it, infectious viral progeny can only be produced by cellular ie3 complementation [221, 372, 373]. Thus, MCMV (the pathogen) cannot easily mutate ie3 (the molecule) to avoid cellular detection of its effects on host cells (the pattern – shutting down host gene expression and turning on early viral genes) without compromising viral fitness. In turn, the Clec2d gene serves by proxy as an innate detector (the recognition axis) of the active repression of cellular host gene expression by diverse viruses. Following the shut-off of Clec2d, this signal is then relayed to NK cells via the “missing-self” loss of Clr-b, which acts as a marker of cell health and the loss of which disinhibits (activates) NK cells via NKR-P1B (the receptor). In contrast, normal bystander cells augment Clec2d and Clr-b expression in response to type-I IFN as a mechanism to relay to NK cells that they are indeed healthy yet responding to paracrine IFN as a byproduct of pathological infected cells in the nearby vicinity.

One remaining unanswered question is how infection with diverse viruses (CMV, HSV, Poxviruses) mechanistically promotes a rapid loss of Clec2d transcription in infected cells. One possibility is general viral shut-off and subversion of host mRNA transcription. Another possibility is that MCMV may encode an immunoevasin that directly modulates Clr-b levels by design. Interestingly, we have previously shown that both rat and mouse CMV encode decoy (Clr-b-surrogate) immunoevasins to elude missing-self recognition via NKR-P1B [107, 145], perhaps suggesting that the viruses cannot otherwise subvert general host mRNA loss and instead directly target host Clr-b and/or NKR-P1B function. Notably, while previous attempts
have failed to identify CMV genes that directly lead to loss of Clr-b expression, here we show that exogenous expression of *ie3* autonomously promotes Clr-b downregulation in the absence of infection. Since *ie3* is required for productive virus infection and early CMV gene expression [221, 372, 373] the virus cannot easily circumvent this requirement, making Clr-b a pattern recognition axis and the NKR-P1B receptor a self-specific PRR. IE3 is known to promote cell cycle arrest in G\(_1\)/G\(_0\) and interact in MCMV replication compartments with PML, Daxx, and ND10 complex proteins involved in viral restriction, and the HCMV homolog, IE2 functions to repress important host transcription factors (including HDAC, TBP/TFIID) [374, 375], and is thought to function as a transcription-associated factor [376]. This is interesting because the Clr-b promoter contains an inverted TATA motif and may function as a TBP-dependent promoter. Future studies will elucidate *ie3*-interacting partners in host gene transcriptional regulation. It is also possible that *ie3* is not the sole gene involved in Clr-b downregulation during infection. Rather, since it is an immediate early gene, *ie3* likely acts as an immediate repressor of Clec2d, whereas early or late gene products may then contribute to the complete loss of surface Clr-b, and/or counteract the loss of Clr-b ligand or NKR-P1B inhibition specifically. Indeed, our current work implicates an m02 family member as a decoy ligand for the NKR-P1B receptor (O.A. Aguilar *et al.*, manuscript submitted to *Cell*), and other ongoing work implicates an m145 family member in moderately antagonizing Clr-b downregulation during MCMV infection (O.A. Aguilar *et al.*, manuscript in preparation).

In conclusion, this work has demonstrated that Clr-b, a marker of cell health and a self-ligand for the NKR-P1B inhibitory receptor, is reciprocally regulated on MCMV-infected and bystander cells. While the mechanism of ‘missing-self’ Clr-b downregulation remains to be fully characterized, Clr-b induction on bystander cells occurs by a canonical paracrine type-I IFN signaling mechanism, making Clr-b a *bona fide* ISG. This dual tuning mechanism involving the inhibitory NKR-P1B:Clr-b axis facilitates NK cell self-nonself discrimination during viral infection by increasing the recognition threshold between pathological (missing-self) targets and normal (healthy-self) bystander cells.
Figure 4.16. Model of transcriptional regulation of Clr-b during MCMV infection. Cells infected with MCMV respond by manufacturing and secreting type-I IFN. The type-I IFN is sensed by neighbouring cells through its interaction with IFNAR receptor. This initiates IFN signaling in the neighbouring cell by phosphorylation of Tyk1 and Jak1, which phosphorylate STAT1 and STAT2. Phosphorylated STAT1 and STAT2 form a heterotrimer with IRF9 that translocates into the nucleus and binds to the ISRE within the first 200bp of the Clec2d promoter, augmenting Clr-b transcription and protein levels. Within the infected cell, Clec2d transcripts are downregulated via loss of RNAPII at the promoter, possibly due to IE3, resulting in less protein expression. This creates disparity between the infected and uninfected cells, where Clr-b<sup>low</sup> infected cells are easily recognized by NKR-P1B-bearing NK cells, while the Clr-b<sup>high</sup> cells are spared. Figure designed by C. Kirkham and D. Grossi and created by D. Grossi. Used with permission.
Chapter 5

Discussion and Future Directions
5 Chapter 5: Discussion and Future Directions

5.1 The major Clec2d TSS and alternative transcripts

The study of the Clec2d TSS by 5’-RACE revealed a single predominant TSS in all cells tested, regardless of primary or transformed cell line status and mostly independent of the levels of Clr-b expressed. The presence of a single major TSS indicates that Clec2d transcription initiation appears to be focused, rather than dispersed, which is indicative of genes that are highly regulated. This fits with the hypothesis that Clr-b expression levels must be tightly controlled in order to facilitate its function as a missing-self marker of cellular health.

Two novel minor splice variants of Clec2d transcripts were also identified in this study: one using an alternative exon-1 within the first intron, and one lacking exon-3. The identification of splice variants was not surprising in itself, as most genes seem to possess at least one alternatively spliced transcript, including other Clr, such as documented Clr-d, and Clr-g splice variants [113, 130, 131]. What was interesting, however, was the nature the alternative transcripts identified. Utilization of the alternative exon-1 is predicted to result in a protein lacking the cytosolic portion of Clr-b and 6 amino acids of the transmembrane domain. The ‘reverse’ signaling events that occur within a target cell when Clr-b is engaged by NKR-P1B have not been investigated meticulously, thus it is unclear at present what the consequences would be if an alternative Clr-b protein were translated from this transcript. The Clr-b cytoplasmic tail has previously been analyzed for putative post-translational modifications using the ELM database website, where consensus CK1 and CK2 phosphorylation motifs, several S/T phosphorylation sites, a consensus TRAF2 recruitment motif, and several K and C residues as possible substrates for ubiquitination were identified [40] (Jason H. Fine, Ph.D. thesis). Since the cytosolic anchor and a portion of the transmembrane domain coding region were also lacking in this transcript, we hypothesize that this protein is not expressed at the cell surface, but rather secreted, retained intracellularly, or flipped to a type-I orientation if a consensus or cryptic signal peptide might be created by the deletion. This is interesting for a number of reasons, but primarily because LLT1, the closest human Clr homolog, has been reported to generate several alternatively spliced transcripts, most of which remain uncharacterized and some which code for
proteins retained in the ER, while others lack a putative transmembrane domain; orientation flipping has not yet been reported, however. Whether or not the splice variants of LLT1 are functional in any capacity, in a positive functional or negative regulatory role, remains to be seen.

The other interesting Clr-b splice variant lacked only the majority of the transmembrane domain (encoded by exon-3), which was intriguing, as the existing Clr-b\(^{+/−}\) (Ocil\(^{+/−}\)) mouse was made by targeted deletion of exon-3, resulting in an expressed frameshift mutant allele, rather than a null allele, for unknown reasons [357]. Analysis of Clec2d transcripts in BWZ.36 cells, which express little to no Clr-b at the cell surface, suggested that this and other lymphoid tumour cell lines could only generate Clec2d transcripts lacking exon-3 (Fig.4.1). Thus, it is possible that alternative splicing to remove exon-3 is one mechanism that transformed cells intrinsically utilize to reduce Clr-b expression levels at the cell surface, perhaps by an evolutionarily conserved mechanism. Clr-b splicing modalities thus remain an interesting pursuit for future investigations, from a structure-function relationship.

5.2 The Clec2d promoter and transcriptional regulation

Functional analysis of the Clec2d promoter revealed the presence of a single promoter immediately upstream from the TSS that was active in both healthy normal and Clr-b\(^{+}\) transformed cell lines, functionally maintained when recombined into euchromatin by stable transfection and genomic integration. The 100bp core promoter element contains an inverted TATA motif that is capable of minimally driving gene expression in the forward orientation. In Chapter 3, partial knockdown of TBP by shRNA resulted in a loss of Clr-b transcripts and protein levels, suggesting that TBP may directly bind the inverted TATA motif to drive Clec2d transcription. A computational analysis of numerous promoters has revealed a correlation between the presence of a TATA box and core promoters with a single TSS, which also fits with what was observed for Clr-b. The location of the TATA box in the Clec2d promoter is 23-28 bp upstream of the TSS, which is also in line with the consensus location of other functional TATA motifs [298]. Our analysis of the Clr-b promoter by luciferase reporter assays did not explicitly reveal any consensus enhancer or silencer elements, while it did provide evidence that such
elements exist. The Clr-b promoter, therefore, appears to be rather simple, as far as inhibitory NKR ligand promoters have been characterized. The promoters of MHC-I genes, for example, contain multiple TSS in the core promoter, with a prominent TSS at +1 [377]. The core MHC-I promoter contains TATA box-like motif, as well as an Inr element and three DPE elements [377, 378]. The conserved expression of Clr-b and MHC-I at least in hematopoietic cells and fibroblasts suggests that additional commonalities remain to be characterized.

Previous work by others has analyzed one of the rat Clec2d family promoters (rClec2d5) [120], yet found no proximal TATA, Inr, or CAAT motifs; however, binding sites for several transcription factors including C/EBP, Sp1, GATA-1, and Oct-1 were identified yet not further validated. In this work, binding sites for C/EBP and Sp1 in the mouse Clr-b/Clec2d promoter (a.k.a., mClec2d8) were also identified (Fig. 4.5A). However, the gene annotated rat Ocil is actually rat Clr5/Clec2d5, which is another rat Clr family member distinct from rat Clr11/Clec2d11, the ligand for rat NKR-P1B [63]. Unfortunately, this annotation has resulted in nomenclature confusion within the Clr gene family [80]. However, the identification of similar transcription factor binding sites in the promoters of the Clr-b/Ocil and other Clr/Ocilrp genes suggests that several members of the Clr family may be regulated similarly, although this has not been formally tested.

Recently, a comparison of six commonly used constitutive promoters for protein expression in mammalian systems was undertaken across several cell lines [379]. The promoters (UBC, PGK, EF1A, CMV, CAGG, SV40 and TRE) were used to drive the expression of GFP in cells stably expressing a transgene, then quantified by flow cytometry. This revealed that the SV40 promoter was found to be a medium strength promoter [379], and thus a good positive control for normalization. The human elongation factor-1α promoter (EF1A), chicken actin promoter coupled with CMV early enhancer (CAGG), and the cytomegalovirus immediate-early (CMV) promoter were identified as strong promoters in this study (although CMV-driven expression is variable depending on the cell type); on the other hand, the human Ubiquitin C promoter (UBC) was the weakest among the group, with the phosphoglycerate kinase-1 (PGK) promoter a close second weakest in all cell types tested [379]. Clr-b promoter activity is similar
in strength to that of the SV40 early promoter, which was used as a positive control for promoter activity in the luciferase reporter assays, suggesting it is also a medium-strength promoter. Nonetheless, full-length polyadenylated Clr-b transcripts remain scarce in most cells, suggesting that constitutive mRNA degradation mechanisms strongly affect Clr-b turnover under normal circumstances, perhaps to facilitate ‘missing-self’ recognition; on this note, an intrinsic autocatalytic hammerhead ribozyme (HHR) has been identified within the Clr-b 3’-UTR that separates the Clr-b CDS from the poly-A tail [207], in turn resulting in enhanced temporal mRNA degradation.

In terms of regulated transcription of Clec2d, the evidence provided in Chapter 4 suggests that Clr-b is reciprocally regulated at the nascent transcript level during viral infection, in infected versus uninfected ‘bystander’ cells. A canonical mechanism for Clr-b upregulation (akin to that of an ISG) on bystander cells was elucidated that requires type-I IFN-mediated and IFNAR1-dependent signaling, followed by STAT1-, STAT2-, IRF9-dependent recruitment to the IRFC motif within the first ~200bp upstream of the TSS (which acts as an ISRE). Upregulation of Clr-b by type-I IFN is reminiscent of MHC class-I molecules, which are also upregulated by type-I (and type-II) IFN. MHC-I molecules serve as ligands for other prominent NKR family members, including the Ly49 receptors in mice and the KIR receptors in humans. Clr-b is structurally and functionally distinct from MHC-I molecules (i.e., Clr-b is a type-II transmembrane, C-type lectin-related protein, while MHC-I molecules are type-I transmembrane, β2m/peptide-dependent, adopt a chimeric dual α-helix/β-platform/IgC-like fold, and also recognized by T cells) [114, 380, 381]; however, they share some interesting features: (i) both exhibit a broad “healthy-self” expression pattern on normal hematopoietic and some non-hematopoietic cells (e.g., fibroblasts) [81]; (ii) both display a “missing-self” expression pattern, being downregulated or lost during pathological alterations, such as viral infection or oncogenic transformation [40, 81, 107, 145, 330, 382]; (iii) both play important roles in NK cell education and recognition [41, 42, 145, 383, 384]; and (iv) both are inducible by type-I IFN at the genomic level via IRF (i.e., MHC-I expression is induced via IRF1) [281, 385], an effect similar to the results shown here for Clr-b via IRF-9 and canonical STAT signaling.
In Chapter 3, a whole-genome high-throughput shRNA screen was employed to identify transcription factors important for the normal or regulated expression of Clr-b, followed by a second iterative screen to aid in shRNA validation. Performing the screens by themselves were not intellectually challenging, although the scale of the experiments were unprecedented in the lab; however, sifting through the large volumes of data generated and identifying positive ‘hits’ above the statistical ‘noise’ was quite challenging. Despite these challenges, MeCP2 was identified and validated as a transcription factor that negatively regulates Clr-b expression at both the transcript and protein levels. MeCP2 studies have been typically performed in other studies using brain cells, since MeCP2 is highly expressed in different parts of the brain and mutations in MeCP2 primarily cause cognitive impairments and neuro-developmental disorders [386]. As suggested in Chapter 3, with the established role for MeCP2 in the brain, it is possible that MeCP2-mediated repression contributes directly to the fact that Clr-b is not expressed in the brain. However, MeCP2 expression has also been reported in the spleen, lung, kidney, heart, and liver [387], implying that expression is not limited to neuronal cells. Some more recent studies have identified novel phenotypes of MeCP2-deficiency in the periphery, including cardiovascular [388] and lung abnormalities [389], bone [390] and skeletal muscle [391] defects, and altered cholesterol biosynthesis [392], suggesting that MeCP2-deficiency outside the nervous system could potentially contribute to Rhett Syndrome-like phenotypes. Ross et al. recently attempted to address nervous system versus peripheral expression of MeCP2 by creating a mouse model where MeCP2 could be silenced in peripheral tissues only [393]. This revealed that the key Rhett Syndrome phenotypes can be attributed to lack of MeCP2 expression in the nervous system. Peripheral absence of MeCP2 led to pronounced exercise fatigue and defective bone mineralization properties [393]. The bone defects are intriguing, as Clr-b is also known as Ocil, an osteoclast inhibitory ligand expressed on osteoblasts that impacts osteoclast development and is important in bone remodeling [112, 113]. In this context, MeCP2 could regulate Clr-b/Ocil expression to control bone remodeling, although this would require further investigation. Notably, NK cells are educated by bone marrow stromal cells within the bone marrow microenvironment, and Clr-b/Ocil does contribute to NK cell education mechanisms [41, 42, 357].
5.3 Importance of Clec2d transcription in regulating Clr-b expression

Previous work has examined Clr-b protein level regulation on NIH3T3 fibroblasts using brefeldin-A [145], showing that Clr-b has a half-life of ~3h at the cell surface. This leads us to hypothesize that the maintenance of normal ‘healthy-self’ levels of Clr-b at the cell surface requires abundant steady-state levels of Clr-b transcripts, in particular via constitutively high levels of transcription. MCMV infection results in a rapid loss of Clr-b transcripts, consistent with a loss of RNAPII promoter occupancy, as observed during HSV-1 infection [364]. Similarly, removal of type-I IFN from the cell culture medium following stimulation results in a rapid reversion to baseline Clr-b levels after ~2-4 hours, consistent with a relatively short half-life. While this does not rule out a role for mRNA stability in Clr-b regulation, it does make a strong case for a direct relationship between Clec2d promoter activity and cell surface protein levels, especially given that the Clec2d (and Clec2e) mRNA have been shown to contain self-cleaving HHR motifs in their 3'-UTR that affect their stability [207]. The internal HHR ribozyme cleaves the Clr-b mRNA following transcription, resulting in the reduction of both transcript and protein levels in luciferase reporter cell assays [207]. While this study did not investigate whether ribozyme activity could be regulated by exogenous factors, it did show that the ribozyme motif can actively cleave the Clec2d mRNA auto-catalytically in the presence of divalent cations alone. Unpublished work in our lab has investigated whether the HRR activity is modulated during genotoxic stress or viral infection, but its activity appears to be constitutive in our hands (Jason H. Fine, Aruz Mesci, and Oscar A. Aguilar, unpublished observations). Thus, if the mRNA is constantly being cleaved, thereby affecting mRNA turnover constitutively, there exists a direct connection between a requirement for basally high levels of transcription and maintenance of steady-state levels of Clr-b protein on the surface of normal, healthy cells. Thus, this represents an ideal situation for a ligand that requires healthy cellular expression and a rapid shut-off mechanism to permit ‘missing-self’ recognition by NK cells. A long mRNA or protein half-life would be inconsistent with missing-self recognition, in the absence of alternative protein-level regulation.

With respect to the regulated expression of Clr-b in normal, uninfected (bystander) cells,
this level of Clr-b regulation functionally opposes its missing-self expression pattern during viral infection, likely in order to increase the dynamic range between infected and normal cells. Most cells can make type-I IFN (especially upon virus infection), although the IFN subtypes secreted largely depend on the cell type. Fibroblasts are known producers of IFN-β, while dendritic cells (DC), especially plasmacytoid DC (pDC) are potent producers of IFN-α. Here, IFN-α4 was chosen due to low cost and immediate availability, but the same responses were induced by purified IFN-β, IFN-α4 in particular mirrors early IFN-β production, and IFN-β production leads to the autocrine induction of many other IFN-α isotypes. There are likely subtle temporal differences between infecting cells with MCMV to make endogenous type-I IFN versus treating healthy cells with purified type-I IFN. Notably, the mutant MEF cells used in Chapter 4, deficient for various IRF, may also possess significantly impacted ability to produce endogenous type-I IFN. Type-I IFN production is controlled transcriptionally and occurs upon PRR activation by PAMP, in turn resulting in the translocation of numerous transcription factors (including ATF2, c-jun, IRF, and CBP/p300) to ISG promoters. There are 9 IRF family members (IRF1-9), which all possess N-terminal DNA binding domains (DBD) that facilitate their recruitment to specific DNA elements (canonical motif: 5’-AANNGAAA-3’) [394-396]. With the exception of IRF1-2, most IRF contain an IRF association domain that allows the formation of homodimeric or heteromeric complexes with other proteins like the STAT proteins. IRF1,2,3,5,7 have been shown to be important for the production of type-I IFN. Thus, the Irf3−/− and Irf7−/− MEF cells used here have defects in their ability to make endogenous type-I IFN in response to viral infection, thereby impairing their ability to promote paracrine induction of Clr-b expression on neighboring cells. This is overcome here by directly treating cells with IFN-α4. IRF3 is constitutively expressed in the cytosol and activated via phosphorylation to form homodimers that are translocated into the nucleus during viral infection. In contrast, IRF7 itself is an ISG, and is only basally expressed in most resting cells. This effectively means that the Irf3−/−Irf9−/− MEF are functionally triple IRF-deficient, since their basal IRF7 levels are diminished. Notably, the Irf9−/− MEF used here have similarly strict Clr-b regulation and function to the Irf3−/−Irf9−/− MEF cells in response to IFN-α4 treatment, but this may not be the case for endogenous IFN production if IRF7 plays an additional role in the auto/paracrine ISG response.
5.4 Future Avenues of Research

The lentiviral shRNA screens performed in Chapter 3 sought to identify transcription factors that play a role in the normal or regulated expression of Clr-b. MeCP2 was the most novel transcription factor identified in the screens. Since MeCP2 mice are available, it would be interesting to further validate the role of MeCP2 on Clr-b expression by simply comparing Clr-b levels on cells from various tissues from MeCP2–/– mice to WT mice by flow cytometry. Since MeCP2 is a negative regulator of Clr-b expression, it would also be interesting to examine the CpG island ~10Kb upstream from the Clec2d TSS to see if it is capable of being methylated under conditions that promote stable Clr-b downregulation (such as oncogenic transformation, as opposed to transient virus infection or DNA-damage responses), or in the tissue-specific expression of Clr-b (Clr-b is not expressed in the brain). ChIP experiments would then demonstrate whether MeCP2 binds to the CpG island. Since MeCP2 can bind unmethylated DNA as well, it is also possible MeCP2 binds elsewhere to the Clec2d promoter or enhancer regions. This hypothesis could be assessed using MeCP2 knockdown or CRISPR/Cas9-mediated gene editing in cells, followed by luciferase reporter assays in comparison to WT cells to narrow down the region(s) of DNA that may mediate MeCP2 binding. These experiments rely on the assumption that MeCP2 does bind to the Clec2d promoter directly to repress transcription. It is also possible that MeCP2 acts indirectly on another transcription factor required for Clr-b expression, which in turn affects Clr-b levels.

TBP and CBP were also identified as potential ‘housekeeping’ factors involved in Clr-b expression, although we put forward a hypothesis that this aspect of Clr-b regulation may be intrinsically important for its role in detecting viral infection, by serving as a proxy marker for normal transcriptional integrity in viable cells (thus, a pattern recognition axis). ChIP assays would be needed to demonstrate functional binding of either factor to the Clec2d promoter. Since the promoter lacks any consensus motifs for CBP, this TF likely functions in concert with other factors, such as TBP or RNAPII, rather than binding to the promoter in isolation. To test this hypothesis, one could perform co-immunoprecipitation of TBP or RNAPII with CBP in ChIP experiments to demonstrate whether complexes of CBP/TBP or CBP/RNAPII are enriched at the Clec2d promoter.
While the shRNA screen seemed like an ideal unbiased approach to investigate Clr-b expression at the time, with hindsight and technological advances, there are aspects that could perhaps be more fruitful. For example, with the innovation of CRISPR/Cas9-mediated genome editing to homozygosity in primary cells, this technology provides extremely rapid functional knockouts (e.g., the *Ifnar1*+/− BWZ.36 cells produced in Chapter 4), compared to transient or partial RNAi-mediated knockdown; thus, a whole-genome CRISPR/Cas9/sgRNA-mediated knockout screen could rapidly result in a library of complete null phenotypes rather than partial or hypomorphic gene product deficiencies. Indeed, CRISPR/Cas9/sgRNA screens of this nature have already been used with success [397, 398]. Potential caveats here include the fact that off-target effects are still possible, although less likely, and in addition, complete null or loss-of-function phenotypes of some proteins can be lethal to primary cells, essentially making these targets invisible in the screens; however, recent data indicate that this may not be a significant concern, since lethal null genes frequently do not result in homozygous gene deletion in practise, rather heterozygous hypomorphic phenotypes, when applied to cell lines in vitro. However, this approach would still result in an overwhelming volume of data. Therefore, similar to the shRNA technology, it might prove useful to validate functional ‘hits’ using multiple sgRNA targeting the same genes that have a similar overall effect on Clr-b expression, rather than just looking at single sgRNA in isolation. Analyzing the data with the help of a bioinformatician would be prudent as well.

In Chapter 4, alternative Clec2d transcripts were also identified. It would be interesting to overexpress the ORF generated from these transcripts, for example tagged alternative exon-1 variants, to see if they are expressed, secreted, retained intracellularly, flipped, or expressed at all at the protein level. Since all cells, with exception of the BWZ.36 cell line, expressed full-length Clec2d transcripts regardless of endogenous Clr-b levels, the use of promoter activity assays, and intronic qPCR for Clec2d transcript levels in Clr-b<sup>lo</sup> versus Clr-b<sup>hi</sup> cells, could help address whether the low expression of Clr-b in some transformed or cancerous cells is transcriptionally controlled or regulated at another level, such as epigenetic regulation that might mirror MHC-I loss in the same cancer cells.
It would also be important to analyze the promoter regions of several Clr family members to see if they are similarly regulated during viral infection, as well as type-I and/or type-II/III IFN treatment. A current issue is that no commercially available antibodies to the other Clr exist, which would limit study of mRNA levels without protein expression validation. Examining the effect of type-I/II/III IFN and HCMV infection on human LLT1, KACL, and AICL expression would also be important to draw parallels to the human system. Currently, there are no reliable commercially available antibodies for LLT1; however, LLT1 mRNA and protein levels have been shown to be upregulated upon infection with a wide variety of viruses and PRR agonists [127, 146], including induction in lung epithelial cells during infection by respiratory syncytial virus (RSV) and IFN-β treatment, providing further justifactio for the hypothesis that LLT1 is the human ortholog of Clr-b [162].

Also important would be dissection of the mechanism by which IE3 promotes early loss of Clr-b expression during viral infection. Since its human homolog, IE2, has been shown to interact with TBP and TFIID [374, 375], it is possible that IE3 functions in a similar way. A simple way to test this hypothesis could include co-immunoprecipitation of TBP followed by Western blot for tagged-IE3 protein, since commercial IE3 mAb are not readily available. One could also attempt to block Clr-b downregulation during MCMV infection by overexpressing TBP inducibly and early during infection as well.

### 5.5 Conclusions

The mechanism(s) behind Clr-b regulation, and those regulating the human LLT1 homolog, are only beginning to be elucidated. While this body of work has provided a detailed examination of Clr-b regulation at the genomic level, there is still much to be learned not only about how Clr-b is regulated transcriptionally, but also post-transcriptionally and post-translationally as well. While the expression of well-known inhibitory NKR ligands such as MHC-I molecules have been studied in depth (likely due to their important role in activating CD8 T cells during adaptive immune responses), very little is known about how other MHC-independent NKR inhibitory ligands are regulated. The regulation of NKR ligands is important to understand, since integrated signaling from numerous stimulatory and inhibitory receptors
interacting with their cognate ligands is ultimately what determines the signaling outputs required for full NK cell activation. Thus, this thesis provides a framework for understanding how the signaling from an MHC-independent recognition system (NKR-P1B:Clr-b) is regulated to assist in distinguishing self from non-self, by increasing the dynamic range between how normal and infected cells respond to pathological stimuli, such as virus infection.
References


