Regulatory mechanisms driving the random monoallelic expression of the natural killer cell receptor genes

By

Djem U. Kissiov

A thesis submitted in partial satisfaction of the

Requirements for the degree of

Doctor of Philosophy

in

Molecular and Cell Biology

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor David H. Raulet, Chair
Professor Jasper D. Rine
Professor Donald C. Rio
Professor Michael W. Nachman

Spring 2021

Abstract

Regulatory mechanisms driving the random monoallelic expression of the natural killer cell receptor genes

by

Djem U. Kissiov

Doctor of Philosophy in Molecular and Cell Biology

University of California, Berkeley,

Professor David H. Raulet, Chair

Natural killer (NK) cells constitute the first line of defense against many foreign pathogens and cancerous cells. Unlike T and B lymphocytes, NK cells do not rearrange their receptor genes and instead generate diversity for MHC I ligands by drawing on a pool of germline-encoded receptors in a stochastic fashion. These receptors are encoded by C-type lectin domain-containing genes in a tandem array on mouse chromosome 6, and are expressed in a random, monoallelic and mitotically stable pattern (RME).

Genes are generally transcribed from both alleles, but in recent years RME has arisen as a notable exception and may describe up to ~10% of genes. While progress has been made toward the understanding of the prevalence and phenomenology of RME, the driving mechanisms are not understood. Research has been hampered by the lack of an established *in vivo* genetic model to dissect the role of specific regulatory elements in RME expression patterns. NK cell receptors provide the opportunity to generate such a model. NK cell receptors are regulated proximally, greatly simplifying the search for the relevant regulatory elements as they should occur near the gene locus itself. Furthermore, our lab has previously developed allele-specific antibodies, allowing the assessment of allelic expression on single cells rapidly and with a high degree of confidence by flow cytometry, circumventing the technical challenges, costs and time associated with experiments such as single cell RNA-seq. Finally, primary NK cells are readily cultured *in vitro* in medium containing IL-2 such that questions about mitotic stability of expression states are easily addressed.

Deletion of enhancers *in vivo* has been greatly simplified with the advent of CRISPR/Cas9-based genome editing. Germline enhancer deletion in mice can now be achieved on timescales of months rather than years, and much more reliably than by traditional gene targeting methods. Additionally, analysis of the chromatin states of silent and active alleles has been historically limited by the requirement of large numbers of cells. Recent advances in chromatin profiling technologies (ATAC-seq and CUT&RUN) allow experiments to be performed with tens of thousands of cells rather than tens of millions, allowing the profiling of

subsets of NK cells sorted with respect to allelic expression status using allele-specific antibodies.

Using the power and flexibility provided by these new approaches, this thesis addresses the following questions. First, what is the role of enhancer elements in regulating the expression frequencies of the variegated NK receptor genes? Chapter 3 addresses this question through a series of enhancer deletions and F_1 hybrid genetics *in vivo*. Furthermore, in Chapter 3 I leverage the power of allele-specific antibodies and flow cytometry to search for RME expression patterns in genes previously thought to be ubiquitously expressed by a given cell type. Strikingly, RME-like expression patterns are identified in *all* assayed receptors: NKG2D by NK cells, CD45 by T cells and B cells, CD8α by cytotoxic T cells, and Thy1 by both cytotoxic and helper T cells. This supports a model where RME is the consequence of generalized stochastic properties of gene expression and can be detected in many and perhaps all genes.

Next, this thesis addresses the chromatin features of both silent and active NK receptor gene alleles *in vivo* and deduces clues as to the mechanism of mitotic stability in RME. Chapter 4 discusses the results of chromatin analyses in sorted primary cells from F_1 hybrid mice. Additionally, this chapter addresses the role of enhancer elements in the maintenance of active RME alleles.

The data presented in this thesis results in a unified model of RME and enhancer function derived from the broadly probabilistic properties of gene expression. Enhancers display constitutive activation but only probabilistic effects on target gene expression, suggesting enhancer activation is generally decoupled from target gene activation. Deletion of individual enhancers from a set of enhancers regulating a target gene results in a reduction of the proportion of cells expressing the gene at both the *Ly49g* and *Nkg2d* loci. A particularly notable result is the transformation of the (apparently) ubiquitously expressed *Nkg2d* gene to a stable RME gene via enhancer deletion, displaying all the fundamental properties of the natural RME NK receptor genes. These deletions had no large effect on the expression level of these genes per cell. These results strongly support the binary on/off model of enhancer action. Rather than being specific to a set of genes, stochastic allele activation appears to be a general property of gene expression and is not restricted to a biologically meaningful set of genes. Surprisingly, silent NK receptor alleles lack a repressive chromatin state, and more closely resemble the chromatin of lineage non-specific genes. Mitotic stability in RME is likely a result of allelic classification as lineage appropriate or lineage non-specific by stochastic enhancer action. We propose that previously documented examples of RME are extreme manifestations of a general property, rather than a result of a dedicated mechanism. RME, therefore, does not seem to be an exception to the rules and instead describes gene expression broadly.

This model of gene expression conceptually resembles a bistable multivibrator, in which an initial signal determines one of two possible states which are then maintained in the absence of the initial signal. In the context of developmental gene regulation, this signal is presumably provided during cellular differentiation, and may be provided to inducible genes in fully differentiated cells. Importantly, gene induction (whether developmental or through some stimulatory signal) is read out by the varying strength of enhancer activity as a rheostat, raising or lowering the probability of stable allelic activation.

Table of Contents

List of Figures and Tables

Acknowledgements

The successful completion of this thesis was the result of the support and investment made in me by many individuals, not all of whom I can thank here. Science is a strange animal; it requires a degree of cold objectivity that runs counter to everything that makes us who we are. To perform this objective work dispassionately, it is—perhaps ironically—necessary to be supported by subjective things such as love and friendship, both of which I have been lucky to receive in abundance.

The first thanks go to my parents, Yusuf Kissiov and Havva Kissiova. Thanks for your never-ending support and providing a net to catch me in my toughest moments. The value I place on learning and knowledge comes directly from you. To my brother Salih—thanks for all of your enthusiasm about my science on road trips to Yosemite!

A major thanks you to Jennifer Wolff, the architect of my scientific interests. You don't know this but after taking your genetics class at Carleton I went right to the Libe and found some books on genomic imprinting on the 3rd floor. I guess there was something about how the diversity and wonder of life is encoded the genome captured my imagination and hasn't let go since. Imprinting seemingly broke those rules (how can the rules by which code of life be broken!?). I guess unsurprisingly, my Ph.D. focused on monoallelic gene expression, and I hope that one day this work might inspire another young student fascinated by life's knack for rule breaking. Thanks for forgiving my many lab errors as a new student. The burned lab bench, the Requiem mass played for the worms I killed and the attempts to use an incubator as a refrigerator. You said that one day I'd gain lab common sense. I'd like to think that's happened.

Thank you to Julie Ahringer, Ron Chen, Mike Chesney, Moritz Hermann, Przemyk Stempor, Adam Knight, Dan Saxton and the many other amazing scientists at the Gurdon Institute. Through all of your help and teaching I transformed from a doe-eyed aspiring scientist into a molecular biologist. A special thanks to Ron, without whom I would not have been aware of a distinguished immunologist named David Raulet.

To David: thanks for the greatest Ph.D. journey I could have hoped for. I came to you with the (I thought) unreasonable request that I join your lab, but work on something related to the genomic organization of NK cells rather than cancer immunology. I was not previously aware of the problem of NK receptor variegation. I remember our fateful meetings during "awkward week," and having my imagination absolutely captured by genes and alleles that ignored each other and behaved completely irresponsibly, yet somehow worked together to create the functional diversity of the NK cell population. I may have behaved just as irresponsibly as these genes by taking up such a challenging thesis project, but I simply could not say no. Your complete support throughout my Ph.D., your willingness to let me go in directions that called to me, and perhaps most importantly your willingness to entertain my wild ideas were incredible. I remember shaking your hand as we decided to take this problem on. You said, "we're going to figure this out, man." Now that we've come to the end of the journey, I hope we can look back on that moment and smile.

Thanks to my thesis committee, Jasper Rine, Don Rio and Michael Nachman. A special thanks to Japser. You were one of the main reasons I chose to come to Berkeley, fully intending to do my Ph.D. in your lab. Deciding not to do so was perhaps the hardest decision I had to make. Despite not being a member of your lab, your scientific insight and input in this work has been invaluable. I hope the science has been enlightening for you and contributed to your understanding of gene silencing!

Among my lab mates, my first thank you must go to Chris Nicolai. We rotated together and shared the experience each step of the way. From you I learned to be fearless in approaching experiments. I also learned how to put the work down and have a beer at Jupiter, or 6. To Thornton, my bay bae, I learned so much from you as someone completely new to immunology. Thanks for humoring the most basic of my questions and joining us for that 7th round at Jupiter. To Natalie Wolf, my second bay mate after Thornton's graduation. You provided a greatly needed grounding presence in both my work and the lab generally. Thanks for bringing Hotsy Totsy into my life on a regular basis. And thanks for that 8th round. To Michele Ardolino, I simply could not have become an immunologist with your help. The basic principles you taught me both about the immune system and the bench when it comes to flow and immune assays (Functionalisssimo!) were exactly what I needed when I started in David's lab. To my undergrads Ishan, Susanna, Irissa, Katrine and Alec: I couldn't have done it without you!

To my classmates, thanks for the amazing ride. Special thanks to Lydia Lutes and Eva Nichols. With the incredible highs of the scientific experience come the inevitable lows. The support system we had as classmates in immunology was essential. Eva, you are inspiring. AoT discussions with you were always a highlight. Lydia, the lab just isn't the same without your laugh on the other side of the wall. I hope we can all see each other again soon.

To the Carls, Maggles the Frog, Kevin, Ben, Adriana, Eric. I couldn't have been luckier to have you all here in the Bay during my Ph.D. We've had incredible Californian adventures, but more important were nights with Ben and Jerry's and netflix, wine and murder mysteries and basketball in Potrero. These experiences will define my memories of my Ph.D. outside the lab.

To my incredible wife, Dr. Breanna Jane Kissiova Tetreault. There simply are not words. We got there—we're both doctors! After 1500+ days of long distance, much of it transatlantic we're finally together, doctorates in hand. You know and understand me better than anyone in this world. Because of that, you knew just what to say every time I hit a necessary bump in the road on my scientific journey. As you watched me become a scientist, I saw you become an awe-inspiring woman and doctor. Each time something great or awful happened you were the first person I wanted to tell. You were there with me when all of this started: in Jenn's Genetics class, in the lab craning our necks over the worms and at wine night on a certain Monday before Thanksgiving, 10 years ago. The work in this thesis belongs to us, not only because of your essential and undying support, but because my love for the work and the science grew from our shared experience all those years ago. I love you, always.

Chapter 1
Introduction

The Immune System

Multicellular life evolved on Earth over billions of years under external selective pressures including competition between life forms for resources, driving much of evolutionary adaptation. To add to the challenges posed to early life, it needed to hold at bay direct molecular attack from other organisms seeking to steal cellular resources. Immunologists traditionally study multicellular organisms under attack by pathogenic unicellular life forms or viruses (that at the time of writing of this thesis are holding the human species hostage in a global pandemic). Recent advances in microbiology, however, have found that unicellular bacteria and archaea developed their own strategies for fending off viral invaders that constitute a form of adaptive immunity, where organisms remember unique molecular features of previous attacks to allow a more powerful and specific response. The immune system is broadly divided into two categories: the innate immune system which recognizes unchanging molecular features of foreign invaders, and the adaptive immune system, which can generate vast diversity to recognize unique and rapidly changing features of pathogens.

The vertebrate adaptive immune system is thought to be approximately 500 million years old, forming via the evolutionary co-opting of a virally-derived transposase called the recombination-activating gene (RAG) (Flajnik and Kasahara 2010). RAG mediates the DNA level rearrangement of antigen receptor genes, allowing for the generation of virtually unlimited diversity in specificity for foreign antigen. T and B lymphocytes both express rearranged antigen receptor genes and mediate cellular and humoral immunity, respectively. Adaptive immune responses are inherently slower than innate responses (discussed shortly), since the prevalence of T and B cells with antigenic specificity for a given pathogen is extremely small—orders of magnitude lower than 1% of naïve cells. These cells must first recognize their cognate antigen introduced by the invader and multiply to an abundance that allows an effective immune response. Cells of the innate immune system produce rapid responses, as a much larger proportion of the "naïve" or pre-infection population are competent to respond immediately since they recognize generalized features of infection. Broadly, the innate immune system recognizes pathogen-associated molecular patterns (PAMPs) such as components of bacterial cell walls or foreign nucleic acids introduced by either bacteria or viruses. Furthermore, invasion and subsequent cell death generate damage-associated molecular patterns (DAMPs), such as extracellular ATP and cytosolic DNA. DAMPs can be thought of as a "rearranging" of the natural order and location of host-derived molecules. Both the innate and adaptive immune systems are comprised by defined cell types that differentiate from common precursors found in adult bone marrow called the hematopoietic stem cells. Emblematic of the innate immune system, neutrophils are a rapidly responding cell type that migrates toward sites of injury and secretes a slew of factors that further recruit immune cells, and other cells called macrophages act as large cellular garbage disposals that track down foreign invaders and dead host cells alike and consume them. As previously mentioned, T and B lymphocytes are key players in the adaptive immune system and display remarkable antigenic specificity and proliferative capacity. Yet other cells are required to present foreign antigen in the priming of T cell responses, with dendritic cells leading the charge. The differentiation of these cell types is a defined—albeit complicated—model of differentiation in mammalian development which unfolds over timescales of weeks, rather than months or years in the case of the entire organism.

Also of importance is the concept that the immune system attacks not only foreign invaders, but also cancerous host cells. In the case of advanced disease the immune system has failed to control tumor growth and metastases. It is thought that the immune system plays a large role in preventing the occurrence of cancer in a process known as immune surveillance. Cancer presents a unique problem for the immune system since it is derived from the host, and there are generally limited antigens that distinguish the tumor from the host (called neoantigens) to which the adaptive immune system can mount a response. Responses to neoantigens are the basis for a new generation of cancer therapies that mobilize the immune system to combat cancer more effectively, collectively called immunotherapies. As one would predict, cancers with a lower abundance of neoantigens are refractory to immunotherapy (Alexandrov and Stratton 2014). Additionally, even cancers with sufficient neoantigens to be recognized by the adaptive immune system can lose the ability to present antigen on major histocompatibility (MHC) class I molecules (through mutations in the antigen processing machinery or MHC molecules themselves), also rendering them recalcitrant to therapies targeted at boosting cytotoxic T lymphocyte (CTL) anti-tumor activity (Garrido et al. 2016). Attacking cells that evade CTL immunity this way is the job of another lymphocyte, but one of the innate immune system, natural killer (NK) cells.

Natural Killer cells

The hero of this thesis is the NK cell. NK cells are lymphocytes, like T and B cells, and represent a key arm of the innate immune system but lack the awe-inspiring power of the T and B lymphocytes to generate mail-order antigen receptor genes to mount a response with an astounding degree of specificity. Instead, NK cells recognize their targets (mainly virally infected or transformed host cells) via germline-encoded receptors (as opposed to the rearranged antigen receptors of T and B cells). These receptors allow NK cells to recognize other cells that induce stress-associated ligands, a process known as induced-self recognition, or to lose inhibition through inhibitory receptors that recognize class I MHC, known as missing-self recognition (Fig. 1.1A). Unlike T or B cells, NK cells are capable of spontaneously killing tumor cells. Indeed, that is how NK cells were first characterized (Herberman et al. 1975). As previously mentioned, the innate immune system recognizes generalized features called PAMPs and DAMPs. Cancer cells do not provide PAMPs to the immune system. The transformed state, however, is associated with various forms of stress that can be recognized by NK cells through their germline-encoded receptors. Furthermore, NK cells can be indirectly activated by activation of other immune populations—often of myeloid origin—and secretion of immune mediators such as cytokines. While work in our laboratory group focuses heavily on the role of NK cells in anti-tumor immunity, NK cells also play a key role in the response to foreign pathogens, chiefly viruses, as well (Cerwenka and Lanier 2001). In all cases, the expression of the germline-encoded receptors is paramount to NK cell function.

Development of NK cells.

NK cells are derived from lymphoid precursor populations that also give rise to T and B cells as well as innate lymphoid cells (ILCs). NK cells are developmentally distinct from type 1 ILCs but express a remarkably similar suite of cell surface receptors that regulate their immune function (more on the NK cell receptors below). The NK cell fate is sequentially defined by

expression of the E-protein suppressing ID2 transcription factor (TF) (Zook et al. 2018). ID2 represses a naïve T cell fate, and is expressed by NK cells and other ILC1s, and later in NK development the TFs T-bet and Eomes enforce late-stage maturation and cytotoxic function (Gordon et al. 2012). Expression of both Eomes and T-bet is shared with ILC1s, while non-NK cell group 1 ILCs are delineated by the expression of the PLZF transcription factor (Harly et al. 2018).

Biological roles and natural cytotoxicity mechanisms of NK cells.

NK cells play a key role in early responses against infection as well as in the initiation of adaptive immune responses via secretion of pro-inflammatory molecules, including IFNγ, TNFα, and others, in response to activation via receptor crosslinking. Unlike T and B cells, NK cells do not require prior antigenic exposure in order to impart their cytotoxic immune function. Virally infected cells or otherwise stressed cells often upregulate stimulatory ligands for NK receptors—a process known as "induced self." Alternatively, but not mutually exclusively, loss of MHC I molecules, which are ligands for inhibitory NK cell receptors, results in "missing self" recognition and subsequent target cytolysis (Karre et al. 1986).

Mechanisms of cytotoxicity.

Canonical NK cell killing involves the secretion of cytotoxic granules containing the pore-forming protein perforin and procaspase-cleaving granzyme B, resulting in target cell apoptosis. NK cells also induce target cell apoptosis via the death receptors TRAIL and FasL. Interestingly, recent evidence suggests that granzyme B-based killing takes place early in sequential serial killing events, and a switch is made to death receptor mediated killing over the course of killing events (Prager et al. 2019). Killing via granule secretion induced kinetically fast killing, while death receptor-mediated killing was characterized by delayed apoptotic events in target cells. NK cells also exert their effects through secretion of IFNγ and TNFα, which can induce cytotoxic and cytostatic effects on target cells in addition to further stimulating broader cellular immune responses (Wang et al. 2012).

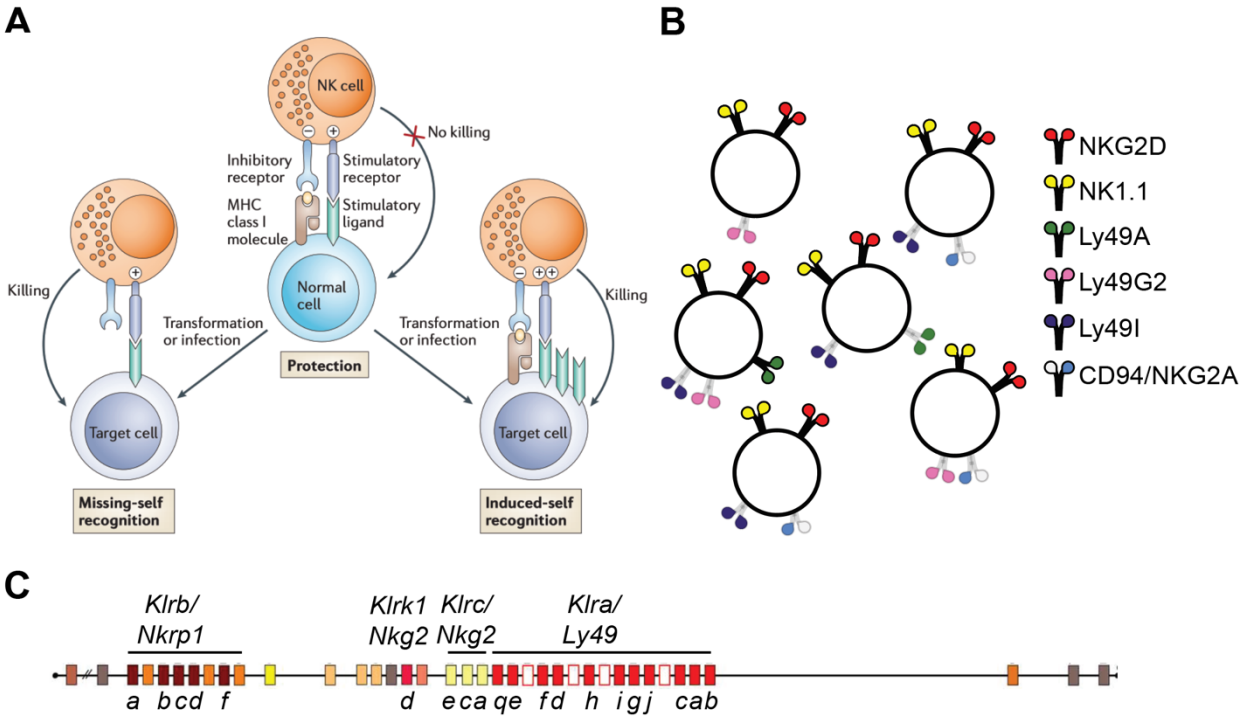


Fig. 1.1. Overview of the mouse NK cell receptors and their variegated expression pattern. (**A**) Schematic model of NK cell target recognition via induced-self and missing-self. Stimulatory ligands are induced in infected or otherwise stressed cells, tipping the balance of stimulatory and inhibitory signaling toward killing. Missing-self recognition involves the loss of a self MHC I allele and the associated inhibitory signal, resulting in a net positive signal resulting in killing. Modified from the following reference: (Raulet and Vance 2006). (**B**) Simplified schematic of the variegated expression pattern of select NK cell receptors. NKG2D and NK1.1 (red and yellow, respectively) are generally expressed by all NK cells, while the other receptors are expressed only on a subset of NK cells. (**C**) Representation of selected multi-gene families in the natural killer cell gene complex on mouse chromosome 6. Distances are not to scale. The gene family is represented by labels above the cluster, while the names of individual members are denoted along the bottom. All genes depicted are the result of evolutionary tandem duplications and belong to the C-type lectin superfamily. Modified from the following reference: (Kelley, Walter, and Trowsdale 2005)

Mechanisms of NK cells target recognition—activating and inhibitory receptors.

NK cells recognize their targets via a system of activating and inhibitory receptors that recognize membrane-bound ligands on potential target cells. Unlike T and B lymphocytes, activating receptors are germline-encoded and recognize ligands that are upregulated by stressed cells (Gowen et al. 2015; Vivier et al. 2011). Inhibitory receptors largely recognize MHC I proteins. The overall logic is simple; ligands that are upregulated in pathological conditions stimulate NK cell cytotoxic activity, while ligands expressed by normal cells that are often downregulated in pathological conditions inhibit NK cell killing, thus protecting normal cells while rendering abnormal cells susceptible. Inhibition and activation are to some extent integrated, or balanced against each other, such that MHC I engagement by inhibitory receptors may only partially inhibit cells receiving strong activating signals. Cellular adhesion molecules expressed by NK cells such as LFA-1, which recognizes ICAM-1 on target cells, play a crucial role by mediating stable interaction of NK cells with target cells, and this axis can be regulated by NK cells via the secretion of TNFα which upregulates ICAM-1 on target cells (Wang et al. 2012).

Activating receptors expressed by NK cells are divided between several broad families; C-type lectins (NKG2D, NK1.1), natural cytotoxicity receptors (NCRs; NKp46 in mouse and human; NKp44 and NKp30 in human) and the Fc receptors CD16 and CD32. While largely associated with inhibition, certain members of the Ly49 family in mice (Ly49D and Ly49H) and the Killer Ig-Like Receptors (KIRs) in humans are activating receptors. Activating receptors usually signal through associated signaling adapter proteins: NKG2D with DAP10/12; activating KIRs with DAP12; NCRs with CD3ζ or FcRg; CD16 with CD3ζ or FcRγ; Ly49D and Ly49H with DAP12) (Barrow, Martin, and Colonna 2019). Most of the adapters contain immunoreceptor tyrosine-based activation motifs (ITAMs) in their cytoplasmic domain, whereas DAP10 contains a distinct motif that recruits PI-3-kinase.

NKG2D is an activating receptor of key importance to this thesis and is expressed by all NK cells that recognizes numerous ligands, all of which are distant relatives of MHC I proteins. The human NKG2D ligands include ULBP1-ULBP6, MICA and MICB, while in mice the ligands include the RAE-1 family members (RAE-1α--RAE-1ε), MULT1 (a ULBP family member), and three H60 isoforms (Wensveen, Jelencic, and Polic 2018). NKG2D ligands are induced by stress (Bauer et al. 1999) of various forms, including genotoxic stress (Gasser et al. 2005), protein/ER stress (Gowen et al. 2015), heat shock (Nice, Coscoy, and Raulet 2009), hyperproliferation (Jung et al. 2012), and activated p53 (Textor et al. 2011).

Inhibitory receptors expressed by NK cells largely recognize MHC I molecules and signal through immunoreceptor tyrosine-based inhibitory motifs (ITIMs) in their cytoplasmic domains. There are 13 KIR genes encoded head to tail in a cluster, 7 of which are inhibitory (KIR2DL1-5; KIR3DL1-3) (Pende et al. 2019). Different inhibitory KIR molecules discriminate different MHC I allelic products. In a striking example of convergent evolution, mouse NK cells express the evolutionarily unrelated Ly49 receptors of the C-type lectin family, which discriminate mouse MHC allelic variants. Both of these receptor families mediate “missing self” recognition, or recognition by NK cells of cells that lose expression of self MHC I molecules.

NK cells are therefore especially promising candidates for therapy against MHC I-deficient tumors, which are resistant to recognition by CD8+ T cells.

The NKG2 and CD94 C-type lectin receptors are expressed by both human and mouse NK cells and form heterodimers that recognize non classical MHC I molecules (HLA-E in human and Qa-1 in mouse). HLA-E and Qa-1 present peptides from classical MHC I proteins, and these complexes are bound by CD94/NKG2 receptors. NKG2C and NKG2E form activating receptors with CD94, but the CD94/NKG2A receptor is an inhibitory isoform. Roughly half of NK cells express inhibitory CD94/NKG2A, which, like inhibitory KIR or Ly49 receptors, mediates missing-self recognition.

NK cell tuning and education.

An important consequence of signaling through both activating and inhibitory receptors on NK cells is the tuning of the functional responsiveness of NK cells. An abundance of activating signal (e.g., from ligands of the NKG2D receptor expressed either on tumor cells or normal endothelial cells) results in excess stimulatory signaling, and consequent down-tuning of the basal ability of NK cells to respond to acute stimulus from a potential target cell (Thompson et al. 2018; Thompson et al. 2017; Deng et al. 2015). Conversely, inhibitory signaling prevents overstimulation of NK cells. In fact, the functional responsiveness of NK cells is directly related to the number of self MHC I recognizing inhibitory molecules expressed (Fernandez et al. 2005; Joncker et al. 2009). This model of the tuning of NK cell responsiveness, wherein chronic activating and inhibitory signaling strength varies quantitatively, raising or lowering the threshold of acute net stimulatory signaling required to induce degranulation was named the rheostat model. The mechanisms by which this is achieved are not well understood and are an area of active investigation in our lab group.

Variegated Expression of the NK cell receptor genes

As this thesis focuses on the variegated expression of the mouse NK receptor genes, this section will focus on the murine *Ly49* and *Nkg2* gene families, although some attention will also be paid to the human *KIR* genes. The major histocompatibility genes are among the most polymorphic in mammalian genomes (Radwan et al. 2020). As NK cells mediate missing-self recognition they must be sensitive to the loss of the various MHC I molecules that represent normal "self" expression on nucleated cells. To deal with this diversity of alleles in the natural population, subpopulations of NK cells with distinct but overlapping specificities are required to be sensitive to loss of any given allele of MHC I that defines immunological "self."

The number of *Ly49* genes is highly variable between inbred mouse lines, with 7 functional inhibitory genes in BALB/c, 9 in C57BL/6 (B6) and 10 in 129 mice, displaying the rapid speed of evolution of the gene cluster (Rahim et al. 2014). The inhibitory *Ly49* genes in the B6 haplotype are: *Klra1* (Ly49A), *Klra2* (Ly49B), *Klra3* (Ly49C), *Klra5* (Ly49E), *Klra6* (Ly49F), *Klra7* (Ly49G), *Klra9* (Ly49I), *Klra10* (Ly49J), *Klra17* (Ly49Q) (Rahim et al. 2014). The encoded Ly49 receptors have overlapping but distinct specificity for MHC I molecules as shown by direct binding assays (Hanke et al. 1999), e.g., Ly49A was found to have a strong interaction with H-2D^{d} , while the closely related Ly49G2 had a similar specificity with a

slightly weaker interaction. Ly49C and the closely related Ly49I, meanwhile, display promiscuous binding to MHC molecules from diverse haplotypes (Schenkel, Kingry, and Slayden 2013).

The practical consequences of this are that only certain receptors have specificity for self-MHC I depending on the genetic background. In addition to Ly49C and Ly49I, which recognize MHC I molecules in the B6 MHC I haplotype (H-2^b), NKG2A/CD94 recognizes the non-classical Qa-1 MHC in the B6 background. In order to render subsets of NK cells sensitive to loss of particular MHC I molecules, NK cells express only a subset of their inhibitory receptors (Fig. 1.1B), e.g., a given cell might express Ly49A and Ly49C, but not Ly49I, Ly49G2 or NKG2A. This expression pattern is variegated (i.e., not all otherwise ontogenetically identical cells express a given receptor), monoallelic (i.e., most, but not all, cells expressing a given receptor express only the maternal or paternal copy) and stochastic with respect to the expression of other receptors (Tanamachi et al. 2001; Tanamachi et al. 2004; Held, Roland, and Raulet 1995). Each receptor is expressed at a characteristic frequency. Ly49A is expressed by ~17% of NK cells in the B6 background, while Ly49G2 is expressed by ~50% (Tanamachi et al. 2004). Expression of one gene or allele has little bearing on the likelihood of expression of any of the other genes (including close neighbors). Co-expression of two receptors occurs at the frequency calculated by multiplying their individual frequencies, showing that the regulation of each receptor gene is largely independent of the others. This mathematical description of independent Ly49 frequencies is known as the "product rule" (Raulet et al. 1997). Even though NKG2A belongs to another subfamily of NK cell receptors (albeit distantly related), it is also regulated according to these rules of independence with respect to the Ly49 receptors (Vance et al. 2002). Once selected, the complement of chosen alleles is largely mitotically stable. How this is achieved at the molecular level is poorly understood, and the chief aim of my thesis work is to unravel this mystery.

Discovery of Ly49 variegation and monoallelic expression.

Ly49a was the first *Ly49* gene discovered as an activation marker expressed by T cells (Yokoyama et al. 1989). Only a subset of T cells, however, express Ly49 receptors and the Ly49s are predominantly expressed by NK cells. A *Ly49a* cDNA probe identified many bands by Southern Blot, implying the existence of multiple family members. Positional cloning demonstrated tight linkage of the *Nk1.1* NK cell antigen and *Ly49a* (Yokoyama et al. 1990). Approximately 20% of NK cells isolated from the spleen were found to stain with a Ly49A specific antibody (YE1/48) in the same study. As more C-type lectin family members expressed by NK cells were discovered in linkage with these genes on chromosome 6, the gene cluster became known as the Natural Killer cell gene Complex (NKC) (Fig. 1.1C) (Yokoyama and Seaman 1993).

Our group made the surprising finding that the Ly49A receptor was monoallelically expressed (Held, Roland, and Raulet 1995). At the time, the concept of monoallelic gene expression (discussed in the next section) had not entered the collective scientific consciousness. Allelic exclusion, however, as it applied to the T and B antigen receptors and the odorant receptors (Chess et al. 1994) was in vogue. By using a Ly49A^{B6} allele-specific antibody clone (A1) and a pan-Ly49A reactive clone (JR9), our lab found that the *Ly49a* gene was

monoallelically expressed, invoking a direct comparison with the allelic exclusion of the antigen and odorant receptors (Held, Roland, and Raulet 1995). Through the analysis of transcripts from NK cell clones it became clear that biallelic expression, while rare, was observed and so the concept of allelic exclusion of Ly49s was replaced with predominantly monoallelic expression (Held and Raulet 1997). With the development of more Ly49 allele-specific antibodies and testing in (B6 x BALB/c)F_1 hybrids, the frequencies of cells expressing both alleles *in vivo* were further elucidated (Tanamachi et al. 2001), and were generally found to approximate independent or stochastic regulation with respect to each other according to the product rule. Importantly, cells sorted according to allelic expression status maintained that configuration during *in vitro* expansion, showing that expression states are mitotically maintained. Through similar approaches, the related NKG2A receptor (also encoded in the NKC) was found to be monoallelically expressed, and again displayed independent and stochastic regulation with respect to the other variegated and monoallelically expressed NKC genes (Vance et al. 2002). It was further proposed that the reason monoallelic expression of the NK receptors for MHC I evolved was not to prevent expression of both alleles, as is the case in allelic exclusion, but rather is the product of a stochastic expression mechanism that results in an overall variegated expression pattern within the gene family, resulting in the diverse NK cell repertoire (Vance et al. 2002).

The human *KIR* genes are likewise expressed in a variegated and monoallelic fashion (Moretta et al. 1990; Valiante et al. 1997), and recognize overlapping subsets of MHC I (HLA in humans). Surprisingly, they are not related to the murine *Ly49* genes, and independently evolved similar functionality and a similar monoallelic and variegated expression pattern is. Presumably the molecular regulation of NK receptor variegation is highly complex. Why then, would the system not have evolved once predating the divergence of the mouse and human lineages? An intriguing possibility is that while variegation seems to break the "rules" of gene expression, the process is stochastically enacted by a generalized feature of gene expression not yet appreciated. If this were the case, it would be much less surprising to envision the evolution of variegation *de novo* in multiple closely related evolutionary lineages. Why the task of providing NK cells with sensitivity for the loss of the diverse array of MHC I molecules falls completely within the purview of the *KIRs* in humans and the unrelated *Ly49s* in mice is not completely understood. Intriguingly, some NKC genes of the *Nkrp1* family (to which NK1.1 belongs) are variegated, suggesting variegation frequently arises in evolution (Kirkham and Carlyle 2014).

The variegated expression pattern of the Ly49 *genes is proximally regulated.*

A key question that arose early in the investigation of *Ly49* regulation was whether the genes were regulated individually by their own proximal elements, or whether the gene family was regulated as a whole. One possibility was that the genes compete *in cis* for a regulatory element such as a locus control region (LCR), as was made famous by the β-globin regulatory elements (Deng et al. 2012). Importantly, the variable exon segments in the protocadherin gene cluster of the nervous system (another example of stochastic, monoallelic expression) are regulated in such a fashion (Esumi et al. 2005; Canzio et al. 2019), where the 5' gene segments stochastically compete for interaction with an enhancer downstream of the cluster. To address this question, a former graduate student in our lab, Dawn Tanamachi, generated an ~30kb *Ly49a* transgene harboring the *Ly49a* gene body, ~3kb of downstream sequence and ~10kb of upstream

sequence (Tanamachi et al. 2004). Eight transgenic lines were generated via microinjection of the linear transgene into single cell embryos. Amazingly, the general result was that the transgene was expressed at a frequency approximating that of natural Ly49A expression (~17%). This was true of most of the transgenic lines, a key consideration given the known powerful effects of genomic position on transgene expression. This result strongly suggested that the variegated expression pattern of *Ly49a* was regulated by elements proximal to the gene locus, and localization within the NKC itself is not required for gene variegation.

Another key finding in this work was the identification of a DNase I hypersensitive site found ~5kb upstream of the *Ly49a* locus. It was named hypersensitive site 1 (HSS-1) and was the first of several sites identified by DNase I-Southern Blotting that fall in the sequence defined by the 30kb transgene. Subsequently identified elements contained in the upstream region of the transgene were designated as HSS-2, 3 and 4, but HSS-1 displayed the highest degree of DNase I sensitivity by far. HSS-1 turned out to be a crucial element and was required for expression of the transgene, since transgenic mice generated using a construct lacking the site were not expressed (Tanamachi et al. 2004). Whether this element played a direct role in variegation or was simply a required enhancer element upon which a variegating epigenetic program is imposed remained an open question at the time.

A final point of interest from this seminal study was that B cells, which do not express Ly49 receptors, expressed the transgene in almost all transgenic lines. Even more curiously the transgene was expressed by all B cells and was not variegated. This immediately suggested the existence of a putative variegating factor or factors that are expressed in NK cells but not in B cells. Furthermore, the 30kb transgene presumably lacked some B cell-specific repressor site that is relevant in the endogenous locus and prevents B cells from expressing *Ly49* genes. The existence of this hypothetical variegating protein factor was part of the impetus for an *in vivo* ENU forward genetic screen focused on NK receptor phenotypes, which will be discussed in the penultimate chapter of this thesis. An alternative explanation, which is more in line with the results presented in this thesis, is that B cells more powerfully induce expression of *Ly49a*. This induction could take the form of a much higher expression level of key TFs in the development of B lymphocytes relative to NK cells.

This study from our group (Tanamachi et al. 2004) remains as perhaps the most important work carried out on the regulation of *Ly49* expression to date. A model describing the HSS-1 element as developmental switches was published shortly afterwards and is discussed in the next section. Importantly, other members of the *Ly49* family harbor sequences highly paralogous to HSS-1 ~5kb upstream. Despite being non-coding, the sequence identity among the different HSS-1 elements is similar to the sequence identity in the coding sequence between genes, and for some gene pairs (e.g., the closely related *Ly49a* and *Ly49g*) is upwards of 90% (Tanamachi et al. 2004).

Prevailing bidirectional promoter switch model and alternative enhancer identity model.

Work that was published shortly after Tanamachi et al. study (Saleh et al. 2004) focused on the HSS-1 (called Pro1 in that study) elements. This study used a pGL3 reporter modified to harbor 2 distinct fluorescent reporter genes flanking the inserted Pro1 elements. Transcription in

one direction would result in expression of one reporter, and transcription in the other direction would drive expression of the opposing reporter. The construct was transfected into cell lines modeling immature and mature NK cells (although, due to the lack of a widely used mature NK cell line, EL4 cells which are of T lymphocyte origin were used). Fluorescence was generally observed to be restricted to one or the other reporter in the immature LNK cell line in clones derived from single cells sorted to express only one color. The concept behind this approach is that a single copy transfectant should only express one of the reporters—a multicopy transfectant could express both colors from different genes, forming a somewhat circular argument. The conclusion that was drawn is that the Pro1 elements transcribe either in the forward or reverse direction (but not both) in immature cells, and that forward transcription establishes the locus for expression in later mature cells, presumably by removal of a transcriptional repressor (Saleh et al. 2004). The propensity to transcribe towards the downstream associated *Ly49* gene would then determine the expression frequency of the gene in the NK cell population. It had been previously observed that transcripts initiating in the Pro1 elements transcribe through the locus only in immature NK cells not yet expressing Ly49 receptors, and that mature cells used different, downstream promoters to express transcript that results in protein (Saleh et al. 2002). Divergent TATA box sequences in the Pro1 elements were shown to play a role in the likelihood of transcription in either direction, and it was proposed that a complex of relevant transcription factors (including Runx3 and NF-κB) could assemble at competing binding sites such that transcription proceeds in either the forward or reverse direction. Because of these results Pro1 (HSS-1) was classified as a promoter that plays a role only early in development, acting as a stochastic "switch" element.

A more recent study proposed that Pro1 instead functions as an enhancer element in mature NK cells (Gays, Taha, and Brooks 2015). Furthermore, its enhancer activity in EL4 cells was not dependent on the TATA elements. The authors note that for a handful of reasons, the bidirectional model required updating. Included among these is that bidirectional transcription initiating at enhancers has, since the publication of the original model, been observed at many if not most active enhancer elements. Furthermore, Pro1 displays weak promoter activity in reporter assays. Therefore, the transcriptional behavior of Pro1 in reporters no longer strongly supports promoter identity of the element and is instead consistent with an enhancer identity. Transcripts emanating from Pro1 elements elongating in both directions were observed in mature primary splenic NK cells, which demonstrated enhancer RNA (eRNA) production and strongly argued for activity of the Pro1 elements in mature cells. Perplexingly, results using similar reporter constructs greatly varied depending on the exact cloned insert (Gays, Taha, and Brooks 2015). Using similar constructs, a counterargument was made that the Pro1 elements do *not* display enhancer activity in mature cells because very few transcripts emanating from Pro1 were detected in mature NK cells by RNA-seq (McCullen et al. 2016). It is important to note that RNA-seq is not generally sufficiently sensitive to detect eRNAs, and therefore cannot provide evidence for a lack of eRNA production. Importantly, whether HSS-1/Pro1 elements are important for the continued expression of *Ly49* genes in mature cells could not be addressed in these studies and would necessitate the ability to abrogate endogenous HSS-1 activity in cells expressing the associated receptor. A major caveat to all of the work using reporters in the LNK and EL4 cell lines is that these cell lines are poor models for *in vivo* NK cell development, and the acquisition and expression pattern of the Ly49 receptors. In order to clarify issues surrounding the molecular identity of the HSS-1/Pro1 elements (hereafter referred to strictly as

HSS-1) and develop a molecular model of the role of HSS-1 in variegated NK receptor expression, it will be necessary to carry out experimental manipulation of regulatory elements *in vivo* and observe the resultant effects on the frequency of expression of individual receptors.

A similar but distinct model based on bidirectional promoter switches was proposed by the Anderson group to regulate the human *KIR* genes. Once again, this model invokes two promoters, one proximal and one distal (Davies et al. 2007). The proximal promoter abuts the first exon. In this case, the distal promoter is required for activation of the proximal promoter (as is the case for the *Ly49* model), but it is the proximal promoter that serves as the bidirectional switch. In a curious parallel with *Xist*/*Txis* dynamics in X-inactivation (discussed below), antisense transcription from the proximal/downstream promoter results in silencing. In the event of probabilistic sense transcription, the gene is activated and maintained as "on." Antisense transcription was proposed to result in the production of a 28 nt Piwi-interacting RNA (piRNA), which was necessary for *KIR* gene silencing (Cichocki et al. 2010). Of note, in recent years it has become appreciated that both enhancers and promoters initiate transcription in both directions, and this appears to be a property of all nucleosome-depleted (hypersensitive) sites (Young et al. 2017). Therefore, what initially appeared to be a unique property of the "switch" elements at the NK receptor genes actually seems to be a pervasive feature of gene regulatory elements.

Epigenetic regulation of variegated NK cell receptor expression.

The epigenetic regulation of the NK cell receptor genes is poorly understood. How are the silent and active states both maintained through mitosis? Mitotic maintenance of gene expression states generally is not well understood, but investigators often turn to post-translational modifications (PTMs) histone proteins and DNA methylation for clues. An important caveat to this approach is that the sufficiency of such modifications—especially of histone N-terminal tails—to mediate mitotic stability of gene expression states is not well established (Saxton and Rine 2019). What is known regarding such modifications at NK receptor loci will be briefly summarized here.

Acetylation of histone-H3 at lysine 9 and H4 was observed by ChIP-qPCR at the *Ly49a* promoter of lymphoid cell lines expressing Ly49A, and the same was observed for the *Ly49g* and *Nkg2a* promoters in cell lines expressing the respective receptors (Rouhi et al. 2007; Rouhi et al. 2006; Rogers et al. 2006). Furthermore, methylation of CpG sites at the promoters of silent alleles was observed in both cell lines and sorted primary NK cells. Importantly, the *Ly49* promoters are CpG poor. Methylation of the promoter regions of these genes was found in both immature developing cells (Rogers et al. 2006) and in non-hematopoietic cells (Rouhi et al. 2006), suggesting that silent variegated alleles are not specifically targeted for methylation, and instead CpG methylation is a feature of the basal "off" or inactive state. Treatment with an inhibitor of DNA methyltransferase (DNMT) activity, 5-azacytidine, was not sufficient to derepress silent *Ly49* or *Nkg2a* alleles, but it was observed that co-treatment with trichostatin A (TSA)—a widely used inhibitor of multiple classes of histone deacetylase enzymes—did result in derepression of silent alleles in EL4 cells. Unpublished data from our group using primary NK cells also indicated that inhibition of DNMT activity is not sufficient for silent allele derepression, and TSA appears to be toxic to primary NK cells in doses well below those used in inhibition assays. *KIR* genes likewise display CpG methylation at silent promoters, but these

promoters harbor a higher density of CpG sites. Treatment of a KIR3DL1-negative human NK cell line, NK92.26 resulted in robust derepression of KIR3DL1 (Chan et al. 2005). Furthermore, active histone modifications (H4K8ac) were associated with promoters at both active and silent alleles, in apparent contrast with the *Ly49* genes (Santourlidis et al. 2008). Importantly, the landscape of repressive histone modifications at the *Ly49* genes has not been explored.

As was the case with the studies discussed in the previous section that attempted to functionally dissect the NK receptor gene regulatory elements and their role in NK cells, the key work done with respect to the chromatin state of silent and active alleles was largely performed in cell lines due to technical cell number limitations with chromatin immunoprecipitation, and the difficulty of obtaining sufficient numbers of primary cells. For a more complete picture of the histone PTM landscape of NK receptor genes, a detailed chromatin analysis in primary NK cells is needed.

Random Monoallelic Gene Expression

When the variegated and monoallelic expression patterns of the NK receptor genes were first uncovered in the 1990s, expression of autosomal genes was widely assumed to occur from both alleles in a deterministic fashion and only a handful of unrearranged autosomal genes had been observed to display monoallelic expression (Vance et al. 2002). Since then, numerous monoallelically expressed loci have been identified. Especially important is the discovery that a significant proportion of expressed genes in a given cell type, on the order of low single digit percentages, are thought to be expressed in a random, monoallelic and mitotically stable fashion (RME), in much the same manner as the NK receptor genes. Analyses of monoallelic expression rely on allelic polymorphisms in heterozygous animals—usually F_1 hybrid mice—that can be detected via transcript or protein analysis. The key modes of monoallelic expression will be discussed with an eye toward the monoallelic expression pattern of the NK receptors.

Classes and mechanisms of monoallelic gene expression.

The classic example of monoallelic gene expression is the random silencing of one X chromosome in each cell during preimplantation development in XX animals (Gendrel et al. 2016). The full complement of genes across the X is silenced, with the exception of a few escape genes. The mechanisms of silencing and mitotic maintenance of silencing are relatively well studied. Transcription of the long non-coding *Xist* transcript results in silencing of the X chromosome in *cis*. Antisense transcription (the product of which is cutely named *Tsix*) prevents expression of *Xist*. The transcript coats the entire chromosome, and repressive histone modification associated with polycomb (H3K27me3) and heterochromatin (H3K9me3) are enriched chromosome wide. Furthermore, the two chromosomes are replicated asynchronously with respect to each other, which has been implicated in the maintenance of monoallelic expression in various contexts (Gendrel et al. 2016; Mostoslavsky et al. 2001). X inactivation is a form of random monoallelic gene expression but is distinct from the monoallelic expression of NK receptor genes in that there is no stochasticity of the two alleles of a gene with respect to each other. The randomness is on the level of the chromosome rather than the gene locus itself. Genomic imprinting, where expression of a gene occurs only from one predetermined parental copy is more similar to NK receptors in that the effect is on the level of the single gene locus but

is fundamentally different in that there is no randomness or stochasticity, as the silenced allele is fixed with respect to parental origin.

Allelic exclusion, a concept made famous by the antigen receptors, is superficially similar to the random monoallelic expression of the NK receptors, but fundamentally differs in that expression of both alleles by a given cell is nearly always prevented. The T cell receptor and Immunoglobulins in B cells are allelically excluded such that one rearranged copy is expressed (Khamlichi and Feil 2018). A feedback mechanism ensures that the cell is defined by a single antigenic specificity. A single olfactory receptor (OR) allele, of which there are over 1,000 genes encoded in both the human and mouse genomes, is selected for expression in the olfactory epithelium sensory neurons (OSNs) at the expense of the remaining thousands of alleles (Monahan and Lomvardas 2015; Chess et al. 1994). Again, a feedback mechanism is thought to ensure that expression of only a single allele is maintained. Many of these cases display some combination of DNA methylation and repressive histone modifications that are thought to be involved in the tight regulation of silent alleles. At the olfactory receptor loci, the repressive H3K9me3 and H3K20me3 are found in abundance in developing sensory neurons that do not yet express ORs (Magklara et al. 2011). Lysine specific demethylase 1 (LSD1) is thought to remove the aforementioned marks at the chosen allele as a key step in allele selection. This tight regulation of silent alleles is associated with biological systems in which biallelic expression would destroy specificity. This is in contrast to the NK receptors, where biallelic expression does not abrogate cellular specificity for MHC I and is observed in approximately the expected proportion of cells according to the product rule.

An interesting case of monoallelic expression is that of the immune cytokine genes. The *Il-2, Il-4, Il-5, Il-10* and *Il-13* genes have all been observed to be monoallelically expressed by different T cell subsets (Bix and Locksley 1998; Riviere, Sunshine, and Littman 1998; Gendrel et al. 2016; Kelly and Locksley 2000). Cytokine stimulation is induced by stimulation through the T cell receptor (TCR). While expression of these genes itself is inherently unstable since they are inducible, an impressive degree of mitotic stability was observed in stimulated T cell clones cultured in conditions that allowed for maintained IL-4 expression (Bix and Locksley 1998). Importantly, the proportion of cells that induce both *Il-4* alleles increased with a stronger TCR signal, demonstrating that allelic expression was stochastically regulated and a stronger induction would result in a higher probability of biallelic activation (Riviere, Sunshine, and Littman 1998). Why cells would express these cytokines monoallelically is not immediately clear, but it was hypothesized that this could be a mechanism to diversify the functional responses of T cell subsets (Bix and Locksley 1998; Kelly and Locksley 2000; Riviere, Sunshine, and Littman 1998).

The NK receptors are generally not inducible via stimulation (with a couple of notable exceptions) but could share many principles with these inducible cytokine genes. If the signal to induce *Ly49* gene expression during NK cell development is relatively weak, this may result in stochastic activation and monoallelic expression. Of the classes of monoallelic expression described above, the cytokine genes most closely resemble the monoallelic expression pattern of the NK receptor genes in that allelic activation appears to be stochastic, biallelic expression is often observed, and there is at least some degree of mitotic stability.

Widespread monoallelic gene expression.

The previously described cases are by no means exhaustive and many more examples of monoallelic gene expression exist. While it is tempting to associate monoallelic expression with genes belonging to certain classes or families that perform given biological functions, it is perhaps more useful to consider the prevalence of monoallelic expression generally. Monoallelism gives the distinct impression of a highly regulated process that must evolve specifically to confer beneficial characteristics to biologically meaningful gene groupings. However, it has more recently been appreciated that RME is widespread and applies to many genes.

Studies of RME have broadly relied on clonal cell lines derived from F_1 hybrids to address several technical limitations to studying RME *in vivo*, precluding population level analysis of the frequency of monoallelic expression. First, it is difficult to reliably assess allelic expression in single cells or clones *in vivo*. Second, it is a major challenge to observe clonal mitotic stability *in vivo*, although a recent study did so by establishing primary cell clones and performing single cell RNA-seq within several clones (Reinius et al. 2016). Third, clonal lines are required in order to obtain cells of known allelic expression status in sufficient numbers to perform chromatin analysis or any experimental manipulation. A major limitation posed by this approach is that the number of clones studied must be manageable, greatly limiting resolution. This is true of both studies in cell lines (Gendrel et al. 2014; Xu et al. 2017) and primary cell clones (Reinius et al. 2016). In order to increase resolution to a level allowing population analysis of RME as is possible with flow cytometry and allele-specific antibodies raised against the NK receptors, many hundreds or thousands of clonal lines would be required. A further major limitation of working with clonal cell lines is the difficulty of extrapolating relevant biology to *in vivo* systems. The NK receptor genes provide a unique opportunity to address each of these limitations, as discussed shortly.

It is important to distinguish between the apparently ubiquitous dynamic form of RME, which is the product of the inherently stochastic nature of transcription. Gene alleles are transcribed in transcriptional bursts which are asynchronous, such that at any given point in time the abundance of transcript derived from one allele might be vastly greater than from the other allele (Deng et al. 2014; Reinius and Sandberg 2015; Reinius et al. 2016). This is highly distinct from clonally stable RME, since both alleles are fully expressed despite dynamic readouts of transcript abundance at a given point in time. Unless otherwise specified, RME will refer to clonally stable RME, where in some cells a single allele is expressed while the other allele is not expressed.

Widespread random monoallelic expression (RME) was first appreciated in a seminal study investigating clonal human lymphoblast cell lines (Gimelbrant et al. 2007). Using a SNP-sensitive Human Mapping array modified to hybridize with RNA, the authors found that nearly 10% of a set of 4,000 assayed genes displayed RME that was clonally stable. This was a surprisingly high number, and more recent studies utilizing RNA-seq in clonal cell populations have placed the estimate closer to 0.5%-3% of genes (Gendrel et al. 2014; Eckersley-Maslin et al. 2014; Reinius et al. 2016). There appears to be a general negative correlation between gene expression level and prevalence of RME, but highly expressed genes may also exhibit RME

(Gendrel et al. 2014; Reinius et al. 2016). Furthermore, it has been difficult to associate RME genes with biological function; there may be a slight bias for cell surface receptors, but gene ontology (GO) analysis shows that RME genes are broadly distributed across GO terms (Gendrel et al. 2016). Interestingly, genes displaying RME are highly cell type specific such that the set of genes displaying RME in one cell type has little bearing on the RME genes in another (Eckersley-Maslin et al. 2014; Xu et al. 2017; Vigneau et al. 2018), further supporting the stochastic nature of the RME phenomenon.

Importantly, there has been recent controversy over the prevalence of RME, or even its relevance as a mode of gene regulation *in vivo* (Reinius et al. 2016; Reinius and Sandberg 2018; Vigneau et al. 2018). Several lines of evidence, however, suggest that RME is a *bona fide* natural phenomenon, chief among which is the reproducibility of identified RME genes across studies of similar cell types (Gendrel et al. 2014; Eckersley-Maslin et al. 2014; Rv et al. 2021). As previously stated, studies of clonal cell populations limit resolution and the ability to identify RME genes that might be monoallelically expression only occasionally. The data presented in this thesis strongly argue not only for the wide prevalence of RME *in vivo* but raise the striking possibility that clonally stable RME is a ubiquitous feature of gene expression.

Chromatin features of RME genes.

Analysis of the chromatin at RME genes has failed to identify a unifying chromatin feature such as that at the OR genes. This is almost certainly because RME genes are a diverse set of genes that represent distinct mechanisms of regulation. Importantly, correlative analyses revealed that most known monoallelic genes in a collection of cell lines could be identified by a bivalent chromatin signature, comprised of co-enrichment of the polycomb repressive mark H3K27me3 and the active transcriptional elongation associated H3K36me3 mark (Nag et al. 2013). Presumably, the active allele would be enriched in H3K36me3 while the silent allele would be associated with H3K27me3. It was proposed that this combination of marks could be used to identify as of yet undiscovered monoallelic genes, circumventing the need for clonal analysis. Importantly this study did not differentiate between general monoallelic expression and random monoallelic expression (Eckersley-Maslin and Spector 2014). Furthermore, the signal for both of these modifications at monoallelically expressed genes was relatively low. This signature might simply be associated with poorly expressed genes, among which RME is more common (Reinius and Sandberg 2018).

Experiments targeted at identifying chromatin modifications that may be causal for maintenance of silenced RME alleles have relied on pharmacological inhibition of DNMT activity (5-azacytidine) or histone methyltransferase activity (various inhibitors targeted a multiple individual histone methyl marks). The results have generally been disappointing (Gendrel et al. 2014; Eckersley-Maslin and Spector 2014; Eckersley-Maslin et al. 2014), and silent RME alleles are not usually derepressed in these assays. Furthermore, only the minority of known RME genes display enrichment of any given repressive feature, and even in those cases, inhibition of the enzymes responsible usually does not result in derepression. The data cumulatively suggest that DNA and histone methylation are not sufficient to explain the bistable phenotype of the RME phenomenon broadly, although in individual cases do play some role.

Constitutive enhancer accessibility at RME loci and probabilistic enhancer action.

Perhaps the most striking and consistent chromatin feature of RME genes is that RME gene-proximal accessible sites are constitutively accessible irrespective of gene expression status (Levin-Klein et al. 2017; Xu et al. 2017). Promoters, on the other hand, are accessible only at expressed alleles. From these data it was extrapolated that promoters are the point of stochastic regulation of RME, while enhancers are constitutively active and permissive for expression. This is in opposition to a model where stochastic regulation is at the point of activation of a critical enhancer element. That enhancer accessibility is decoupled from RME gene expression is puzzling. Curiously, the HSS-1 element at the *Ly49a* gene was accessible to DNase I in both Ly49A+ and Ly49A- cells (Tanamachi et al. 2004), suggesting the NK receptor genes follow the general RME accessibility pattern.

These results were surprising, since activation of critical enhancers that drive expression of a target gene is generally thought of as a 1:1 relationship. However, there exists an underappreciated literature about the probabilistic nature of enhancer action (Weintraub 1988; Fiering, Whitelaw, and Martin 2000; Walters et al. 1995; Walters et al. 1996; Blackwood and Kadonaga 1998; De Gobbi et al. 2017). Two opposing models of enhancer action purport that 1) enhancers act in a stochastic and binary fashion to raise the *probability* of target gene expression, known as the probability (on or off) model, and 2) they act to increase the rate of transcription, known as the progressive response model (Blackwood and Kadonaga 1998). The progressive response model gained traction early on, since enhancers increased the amount of gene product in luciferase reporter assays (Fiering, Whitelaw, and Martin 2000). However, this could be achieved through both mechanisms in bulk analyses that do not divulge single cell information.

The evidence supporting the binary or probability model chiefly comes from assays where enhancers are present, or not, and single cell information can be obtained. The first experiments were done in CV-1 and HeLa cells transiently transfected with an SV40 T-antigen reporter with or without the SV40 enhancer (Weintraub 1988). A small proportion of cells transfected with a reporter lacking the enhancer expressed normal levels of SV40 T-antigen. Similar experiments were done in K562 and HeLa cells with a transiently transfected β-galactosidase reporter under the control of the γ-globin promoter and inserted enhancers (or not). These assays showed that enhancers raised the proportion of cells that were reporter positive, but expression per cell remained constant (Walters et al. 1995). These results were followed by examples of similar effects from targeted deletion of endogenous enhancers, first in B cell hybridomas and variegated expression of IgM after deletion of the intronic *Igh* enhancer (Ronai, Berru, and Shulman 1999), then in the *Cd8a* locus *in* vivo (Garefalaki et al. 2002; Ellmeier et al. 2002), and followed by more recent examples in the globin genes in a humanized mouse model (De Gobbi et al. 2017) and at the *Bcl11b* locus in developing T cells (Ng et al. 2018). In each case, the major effect of enhancer deletion was to reduce the probability of target gene expression. Of note, deletion of a P element enhancer in OSNs results in a reduced expression probability of nearby OR genes *in cis*, suggesting direct relevance of probabilistic enhancer action to monoallelic expression (Khan, Vaes, and Mombaerts 2011). Importantly and surprisingly, where it was tested enhancer deletion was sufficient to result in variegated expression that displayed a significant degree of mitotic stability (Ronai, Berru, and Shulman 1999, 2002, 2004; Garefalaki et al. 2002). Furthermore, where it could be measured, the

variegation observed was regulated *in cis* and displayed allelic stochasticity (Garefalaki et al. 2002; Ellmeier et al. 2002; Ng et al. 2018). These phenotypes are strikingly similar to RME.

In these cases, it is presumed that the deleted enhancer is one of a set of enhancers functioning in a coordinate fashion to increase the probability of gene expression, such that deletion of one does not preclude gene expression completely. Strong evidence of this comes from two studies of the *Cd8a* locus (encoding the α-subunit of the CD8 co-receptor on cytotoxic T cells) (Garefalaki et al. 2002; Ellmeier et al. 2002). In each study, two distinct enhancers out of a group of at least four were deleted using classical gene targeting in mice, resulting in an RME-like phenotype of CD8α in developing thymocytes. This strongly suggested that these enhancers function in a partially redundant fashion to "ensure" expression of the co-receptor in the appropriate cells, rather than regulate the gene via a qualitatively different mechanism for each enhancer. Multiple enhancers regulating a single gene is a common feature of gene expression and is thought to increase the robustness of the system especially in sub-optimal conditions (Perry, Boettiger, and Levine 2011; Hobert 2010). As a final important point on the function of enhancers, it has recently become appreciated that enhancers regulate the *frequency*, rather than size of transcriptional bursts (Fukaya, Lim, and Levine 2016; Bartman et al. 2016; Larsson et al. 2019), indicating that enhancers play a key role in regulating the dynamic form of RME discussed earlier. How enhancers regulate clonally stable RME remains an open question, but based on previous data from enhancer deletions and enhancer accessibility dynamics at RME loci, the underlying hypothesis of this thesis is that the probabilistic properties of enhancer action are a crucial mechanistic driver of clonally stable RME.

Of great importance, the probabilistic enhancer properties deduced from early experiments (Weintraub 1988; Walters et al. 1995; Walters et al. 1996) were predicted to be a potential mechanism for the generation of diversity in otherwise homogenous cell populations (Fiering, Whitelaw, and Martin 2000). This would involve the stochastic expression of alleles—constituting a prediction of the RME phenomenon before it was observed. At the time it was noted that to generalize the early results, it would be necessary to delete enhancers *in situ* from genes that are expressed in a variegated fashion and observe the effects on expression probability. This requires the ability to robustly measure allelic expression status at the single cell level. The NK receptor genes provide the ideal *in vivo* model for these experiments, and they form the bulk of the key results presented in this thesis.

Chapter 2
Methodology

Animals and animal procedures

All mice were maintained at the University of California, Berkeley. *Nkg2d*$^{-/-}$ mice are available at the Jackson Laboratory (JAX Stock No. 022733). C57BL/6J (hereafter referred to as B6) mice were purchased from the Jackson Laboratory and bred at UC Berkeley. 129-*Ncr1*^{tm1Oman}/J (Ncr^{gfp}), BALB/cJ, CBA/J and AKR/J and B6.SJL-*Ptprc*^{a} *Pepc*^{b}/BoyJ mice were purchased from the Jackson Laboratory. All F_1 hybrid mice were generated through crosses carried out at UC Berkeley.

For the generation of CRISPR edited mice, Cas9 RNP was delivered to single-cell embryos either through microinjection or CRISPR-EZ electroporation, both of which are described in reference (Modzelewski et al. 2018). *Ly49a*$_{Hss1\Delta}$ mice were generated by microinjection, while *Nkg2a*$_{5'E\Delta}$, *Nkg2d*$_{5'E\Delta}$ and *Ly49g*$_{5'E\Delta}$ mice were generated by CRISPR-EZ electroporation. Whether through microinjection or electroporation, enhancer deletion mice we generated using paired sgRNAs flanking the enhancer. sgRNAs were selected using the GPP web portal from the Broad Institute. Guides with highest predicted editing efficiencies were prioritized, while also minimizing for predicted off-target cutting in protein-coding genes. sgRNAs were generated using the HiScribe T7 Quick High Yield RNA Synthesis Kit (New England Biolabs). Founder mice (F_0) harboring deletion alleles were backcrossed to C57BL/6J (B6) mice to generate heterozygous F_1 mice, and were then intercrossed to generate WT, heterozygous and homozygous littermates for experiments. All sgRNAs used for the generation of enhancer deletion mice are listed in Table 2.1. Primers used to PCR identify edited founders and genotype subsequent filial generations are listed in Table 2.2 and Table 2.3. (B6 x BALB) F_1 were used for sequencing experiments with the exception of ATAC-seq in NK cells sorted according to Ly49G2 allelic expression, in which case (Ncr^{gfp} x BALB) F_1 hybrids were used. All animals were used between 8-32 weeks of age, and all experiments were approved by the UC Berkeley Animal Care and Use Committee (ACUC).

Target enhancer	sgRNA sequence 5' → 3'	
	Upstream sgRNA	**Downstream sgRNA**
***Ly49a*$_{Hss1}$**	CTTAGTGCTTGAGCCCATGA	CAGCATAATACAGGAGGTAA
***Nkg2a*$_{5'E}$**	CAGGATAATTATTATGATTG	GAGGCACCGTTCAGATGCAG
***Nkg2d*$_{5'E}$**	ATAGCCAACATTATACTAGA CTTTCAACTATTATTTAACA	AGCACAAGGTGAGTCCTAGG TTCTACAACCATTATGTGGG
***Ly49g*$_{Hss5}$**	AGAACAGTCATTTCTTTAAA	AGAGGGATTACTCTGGGGAA

Table 2.1. sgRNAs used to generate germline enhancer deletion mice. Guide RNAs (sgRNAs) used to generate germline enhancer deletion mice via electroporation or microinjection are displayed. A flanking guide pair was used to delete the indicated enhancer, except for in the case of *Nkg2d*$_{5'E}$, where two sets of flanking guides were used (all four sgRNAs were simultaneously delivered to embryos).

Target enhancer	Deletion genotyping primer sequence 5'→3'	
	Upstream	**Downstream**
Ly49a$_{Hss1}$	AAGGCACATACCACATTGTCAC	GAGCAGTACCTTCCTCTAAGTTC
Nkg2a$_{5'E}$	1-ATGAGTGTGCAGTGGTGTCTTC 2-TGTGCCAGCCATAAGAGTTTG	1-TCATCCAAAGAGCCACAGCA 2-TCCAGATGATGGCTAACTCTCCAT
Nkg2d$_{5'E}$	1-AAGGAAACCAGAACCCTGATG 2-AGAGTGATCTCAGTGATGCAAGGA	1/2-AGCTGAGGACAAGCTGCACA
Ly49g$_{Hss5}$	1-AGCTTCTGCCTATACTCCTGATTG 2-AAACTTGGGAAAGAGATTCGGAC	1/2-AAGGACAACTTTGATGGGTTATGG

Table 2.2. Germline deletion genotyping primers. Primers used to genotype mice carrying a deletion allele are shown. More than one primer is shown if PCR was performed as a nested reaction; "1" indicates use in the first amplification and "2" indicates use in the subsequent amplification.

Target enhancer	Homozygosity genotyping primers 5' → 3'	
	Upstream	**Downstream**
Ly49a$_{Hss1}$	GTCCAAGGGTGTGACTGGAAG	GAGCAGTACCTTCCTCTAAGTTC
Nkg2a$_{5'E}$	AGAGACACTTTGTACCTTCCAC	TGTGTCATTGAAGGTTGAACAG
Nkg2d$_{5'E}$	CATCCATTCAA CATAGTTCTGG	TAACTGATGGTTTTCTTGGCCA
Ly49g$_{Hss5}$	TTTTCGCTCATGTCTACCCAG	GGGAGAAGTAGCAGCAGTGT

Table 2.3. Deletion homozygosity genotyping primers. Primers used to genotype mice lacking a wildtype, non-deleted allele are shown. These primers allow delineation of WT, heterozygous and homozygous enhancer deletion animals with respect to the indicated enhancer element.

Flow Cytometry

Single cell splenocyte suspensions were generated by passing spleens through a 40 μm filter. Red blood cells were lysed with ACK buffer. Fresh splenocytes, or where indicated cells cultured with 1000 U/ml recombinant human IL-2 (National Cancer Institute) were stained for flow cytometry in PBS containing 2.5% FCS (FACS Buffer). Before staining with antibodies, FcγRII/III receptors were blocked for 15 minutes at 4C using 2.4G2 hybridoma supernatant. Cells were washed with FACS buffer and then stained with antibodies directly conjugated to fluorochromes or biotin at 4°C for 15 to 30 minutes. In order to differentiate between alleles of a receptor in (B6 x BALB/c)F_1 hybrid NK cells, the B6-specfic clone was used first in order to block epitopes in competition with the clone recognizing both alleles. For example, to discriminate Ly49G2 alleles, cells were stained for at least 15 minutes with 3/25 which recognizes Ly49G2^{B6}, and then 4D11 was added. For discriminating alleles of NKG2A, cells were stained first with the NKG2A^{B6}-specific 16a11, followed by 20d5, which binds to both

alleles. Ly49A^{B6} (A1) was added before the non-discriminating JR9 clone, but in this case, cells expressing only the B6 allele did not resolve from the population of cells expressing both alleles. When necessary, cells were washed and then stained with secondary antibody or fluorochrome-conjugated streptavidin. Near-IR viability dye (Invitrogen L34975) or DAPI (Biolegend 422801) were used to discriminate live cells. Flow cytometry was carried out using an LSR Fortessa or X20 from BD Biosciences, and data were analyzed using FlowJo software. In all cases, NK cells were defined as CD3$^-$NKp46$^+$ splenocytes. For sorting on a BD FACSAria II sorter, the samples were prepared nearly identically as they were for flow cytometric analysis with the exception that the medium used was sterile RPMI 1640 (ThermoFisher) with 5% FCS.

Antibodies used in flow cytometry

From Biolegend: anti-CD3ε (145-2C11), anti-CD4 (GK1.5), anti-CD19 (6D5), anti-F4/80 (BM8), anti-Ter119 (TER-119), anti-NKp46 (29A1.4), anti-NKG2A^{B6} (16a11), anti-Ly49A^{B6} (A1), anti-NKG2D (CX5), anti-CD8β (YTS156.7.7), anti-CD45.1 (A40), anti-CD45.2 (104), anti-CD90.2 (53-2.1), goat-anti-mouse IgG (Poly4053). From eBioscience/ThermoFisher: anti-NKG2A (20d5), anti-Ly49I (YLI-90), anti-Ly49G2 (4D11), anti-CD90.1 (HIS51), anti-rat IgG F(ab')2 (polyclonal, lot 17-4822-820). From BioXCell: anti-CD8.1 (116-13.1), anti-CD8.2 (2.43). Purified in-house: anti-Ly49A (JR9), anti-Ly49G2^{B6} (3/25), anti-NKG2D (MI6).

Ex vivo NK cell cultures for analysis of the stability of monoallelic expression of NKG2D

Splenocytes were prepared as a single cell suspension by passage through a 40 μm filter and red blood cells were lysed with ACK buffer. Cells were washed three times in RPMI 1640 media with 5% FCS, and were then cultured in 40 mL of the same media with 1000 U/mL recombinant IL-2. NKG2D+/- NK cells were sorted from *WT* or *Nkg2d*$^{5'E\Delta/5'E\Delta}$ mice on day 2 or 3. Cells were cultured *in vitro* in IL-2 containing media for a further 8-10 days, during which cells expanded ~10-100 fold based on hemocytometer counts. Cells were analyzed for NKG2D expression by flow cytometry. In all cases media contained 5% FCS (Omega Scientific), 0.2 mg/mL glutamine (Sigma), 100 U/mL penicillin (ThermoFisher), 100 μg/mL streptomycin (Thermo Fisher Scientific), 10 μg/mL gentamycin sulfate (Fisher Scientific), 50 μM β-mercaptoethanol (EMD Biosciences), and 20 mM HEPES (ThermoFisher). Cells were incubated at 37°C at 5% CO_2.

Ex vivo assay for the stability of monoallelic expression in T cells

Cells from the spleens and a collection of lymph nodes (brachial, axial, inguinal, mesenteric) from F_1 hybrid mice and parental inbred line controls were combined and passed through a 40 μm filter, and red blood cells were lysed with ACK buffer. Cells were prepared for sorting as described above, staining with the relevant allele-specific antibodies. For CD45 monoallelic expression, Thy1+ cells were further gated according to CD45 allelic expression. For CD8α monoallelic expression, CD3+CD8b+ cells were further gated on CD8α allelic expression. Cells expressing either the paternal or maternal allele (or both) of the receptor studied were sorted and expanded for 1 week in RPMI 1640 (ThermoFisher) containing 200 U/mL recombinant IL-2, Dynabeads mouse T-activator CD3/CD28 (ThermoFisher) beads at a 1:1 cells to beads ratio, 10% FCS, and supplemented as RPMI 1640 above. After 1 week of

expansion, cells were harvested, counted by hemocytometer and prepared for a second sort. After sorting for expression of the relevant receptor allele again, cells were once again expanded in a restimulation, this time with a cells to beads ratio of 10:1. After the second expansion, cells were again counted, stained and prepped for final analysis of monoallelic receptor expression by flow cytometry.

F_1 hybrid genetics and calculations of expected changes in receptor-expressing NK cell populations

F_1 hybrid genetics were carried out by breeding WT or CRISPR/Cas9-edited males on the B6 background to females from the following backgrounds: BALBc/J, CBA/J, AKR/J. Edited alleles were crossed only to BALBc/J, while CBA/J and AKR/J were used in the F_1 hybrid analysis of monoallelic expression of CD8α and Thy1, respectively.

We estimated the expected frequencies of NK cells in (*Nkg2a*$^{B6\text{-}5'E\Delta/BALB/c+}$) F_1 mice by assuming independence of allelic expression. That assumption leads to the following predictions:

The percentage of cells expressing neither allele in the mutant will equal the sum of the percentages of the two NK cell populations in WT (B6 x BALB/c)F_1 hybrids that lack NKG2A$^{BALB/c}$.

The percentage of cells expressing NKG2A$^{BALB/c}$ only in the mutant will equal the sum of the percentages of the NK cell populations in WT (B6 x BALB/c)F_1 hybrids that express NKG2A$^{BALB/c}$.

The percentages of cells expressing NKG2A^{B6} only or both NKG2A^{B6} and NKG2A$^{BALB/c}$ will be 0, since NKG2A^{B6} is not expressed.

The expected changes in populations with respect to Ly49G2 alleles in *Ly49g*$^{B6\text{-}Hss5\Delta/BALB/c+}$ mice were calculated with the same assumption of independent regulation of alleles.

We started by calculating the overall percentage of expression of Ly49G2^{B6} in the mutant, which averaged 47.7% of that in WT F1 mice.

The predicted percentage of cells expressing only Ly49A^{B6} in the mutant F1 was then 47.7% of the percentage of cells expressing only Ly49A^{B6} in WT mice.

And the predicted percentage of cells expressing both alleles in the mutant F1 was 47.7% of the percentage of cells expressing both alleles in WT mice.

The predicted percentage of cells expressing neither allele in the mutant F1 was calculated as the percentage of cells expressing neither allele in WT mice + 52.3% (100%-47.7%) of the precentage of NK cells that express only Ly49G2^{B6} in WT mice.

Finally, the predicted percentage of NK cells expressing only Ly49G2$^{BALB/c}$ in the mutant was calculated as the percentage expressing only Ly49G2$^{BALB/c}$ in WT mice plus 52.3% of the NK cells expressing both alleles in WT mice.

ATAC-seq

ATAC-seq was performed as previously described in reference (Buenrostro et al. 2013). Briefly, 50,000 sorted NK cells were washed in cold PBS and resuspended in lysis buffer (10 mM Tris-HCl, pH 7.4; 10 mM NaCl; 3 mM $MgCl_2$; 0.1% (v/v) Igepal CA-630). The crude nuclear prep was then centrifuged and resuspended in 1x TD buffer containing the Tn5 transposase (Illumina FC-121-1030). The transposition reaction was incubated at 37C for 30 minutes and immediately purified using the Qiagen MinElute kit. Libraries were PCR amplified using the Nextera complementary primers listed in reference (Buenrostro et al. 2013) and were sequenced using an Illumina HiSeq 4000.

CUT&RUN

CUT&RUN was performed essentially as previously described (Skene, Henikoff, and Henikoff 2018). Briefly, 50,000-500,000 NK cells were washed and immobilized on Con A beads (Bangs Laboratories) and permeabilized with wash buffer containing 0.05% w/v Digitonin (Sigma-Aldrich). Cells were incubated rotating for 2 hours at 4°C with antibody at a concentration of 10-20 µg/mL. Permeabilized cells were washed and incubated rotating at room temperature for 10 minutes with pA-MNase (kindly provided by the Henikoff lab) at a concentration of 700 ng/mL. After washing, cells were incubated at 0°C and MNase digestion was initiated by addition of $CaCl_2$ to 1.3 mM. After 30 minutes, the reaction was stopped by the addition of EDTA and EGTA. Chromatin fragments were released by incubation at 37°C for 10 minutes, purified by overnight proteinase K digestion at a concentration of 120 µg/mL with 0.1% wt/vol SDS at 55°C. DNA was finally purified by phenol/chloroform extraction followed by PEG-8000 precipitation (final concentration of 15% wt/vol) using Sera-mag SpeedBeads (Fisher) (https://ethanomics.files.wordpress.com/2012/08/serapure_v2-2.pdf).

Libraries were prepared using the New England Biolabs Ultra II DNA library prep kit for Illumina as described online (https://www.protocols.io/view/library-prep-for-cut-amp-run-with-nebnext-ultra-ii-bagaibse?version_warning=no) with the following specifications and modifications. The entire preparation of purified CUT&RUN fragments from a reaction were used to create libraries. For histone modifications, end repair and dA-tailing were carried out at 65°C. NEB hairpin adapters (From NEBNext Multiplex Oligos for Illumina) were diluted 25-fold in TBS buffer and ligated at 20°C for 15 minutes, and hairpins were cleaved by the addition of USER enzyme. Size selection was performed with AmpureXP beads (Agencourt), adding 0.4X volumes to remove large fragments. The supernatant was recovered, and a further 0.6X volumes of AmpureXP beads were added along with 0.6X volumes of PEG-8000 (20% wt/vol PEG-8000, 2.5 M NaCl) for quantitative recovery of smaller fragments. Adapter-ligated libraries were amplified for 15 cycles using NEBNext Ultra II Q5 Master Mix using the universal primer and an indexing primer provided with the NEBNext oligos. Amplified libraries were further purified with the

addition of 1.0X volumes of AmpureXP beads to remove adapter dimer and eluted in 25 μL H_2O. Libraries were quantified by Qubit (ThermoFisher) and Bioanalyzer (Agilent) before sequencing on an Illumina HiSeq 4000 or MiniSeq as paired-ends to a depth of 10-32 million.

The following antibodies were used for CUT&RUN: Abcam: anti-H3K4me1(ab8895), anti-H3K4me2 (ab7766), anti-H3K4me3 (ab8580), anti-H3K27ac (ab4729), anti-H3K9me3 (ab8898). Cell Signaling: anti-H3K27me3 (C36B11), anti-H2AUb1 (D27C4). Biolegend: Mouse IgG2a k

Datasets, processing and visualization

All mined data were downloaded from NCBI Gene Expression Omnibus (GEO) or the European Bioinformatics Institute (EBI) in FASTQ format. NK cell ATAC-seq histone modification (H3K4me1, H3K4me2, H3K4me3, H3K27ac) were mined from reference (Lara-Astiaso et al. 2014) under GEO accession numbers GSE59992 and GSE60103. NK cell Runx3 ChIP-seq data was mined from reference (Levanon et al. 2014) (GSE52625). T-bet ChIP-seq data were mined from reference (Jeevan-Raj et al. 2017) (GSE77695). p300 ChIP-seq raw data was mined from reference (Sciume et al. 2020) (GSE145299). p300 ChIP-seq peaks were called in reference (Sciume et al. 2020) and downloaded in .csv format

All data—mined or generated by our group—were aligned and processed using an in-house pipeline. Raw data were aligned to the mm10 reference genome build with bowtie2 using the parameter "--sensitive". All reads aligned to the mitochondrial chromosome were removed. Aligned reads were then sorted, indexed, and filtered for a mapping quality of ≥10 with samtools. PCR duplicates were removed with picard (Broad Institute). Reads covering blacklisted regions (ENCODE mm10 database), were removed with bedtools. Data were normalized by signal per million reads (SPMR) using macs2 and converted to bigWig format using the bedGraphToBigWig program from UCSC Genome Browser. Data were visualized using the Integrative Genomics Viewer (Thorvaldsdottir, Robinson, and Mesirov 2013).

Ranking of accessible sites in NK cells according to H3K4me1:me3 ratio

Reads from replicate ChIP-seq datasets (for both H3K4me1 and H3K4me3) from reference __ were merged to ensure robust signals, and the resultant files were processed and normalized as above. NK cell ATAC-seq peaks were called in the Ly49G2$^{B6+BALB+}$ NK cell ATAC-seq dataset using macs2 narrowpeaks. Before ranking, ATAC-seq peaks were filtered such that only peaks that fell within the top 95% of both H3K4me1 and H3K4me3 signal computed over a 2 kb window from the peak midpoint computed using pandas and numpy in Python 3.7.4, resulting in 51,650 usable peaks. H3K4me1:me3 raw ratio and log2 ratio bigwigs were generated with the bamCompare utility from deepTools (v2.5.4). The log2 ratio track was visualized on IGV, and the raw ratio was used to rank ATAC-seq peaks. Heatmaps were generated with the computeMatrix and plotHeatmap utilities from deepTools (v2.5.4). Heatmaps were sorted by the mean H3K4me1:me3 ratio signal over a 2 kb window centered at the midpoint of the 51,650 ATAC-seq peaks. *Hss1* and *5'E* enhancer regions and corresponding promoters at NKC genes were individually predefined and the position of each was then marked on the heatmap.

Definition of NK cell promoters and enhancers and ranking of regulatory elements according to H3K4me1:me3 ratio

Annotated mouse promoters (defined as the TSS at a single nucleotide) in the mm10 genome assembly were downloaded as a BED file from the EDPNew (http://epd.vital-it.ch) database. To identify likely active promoters in NK cells, broad regions of H3K27ac were called based on ChIP-seq data mined from reference (Lara-Astiaso et al. 2014) using the "macs2 callpeak --broad" command. Mouse EDPNew promoters falling within broad H3K27ac domains were identified using the "bedtools intersect -wa" command, resulting in a set of 9901 active promoters in mouse NK cells.

Enhancers in naïve mouse NK cells were defined as the intersection of ATAC-seq and p300 peaks not found at the promoters as defined above. p300 ChIP-seq peaks in resting NK cells were previously defined and downloaded from reference (Sciume et al. 2020). ATAC-seq peaks that were enriched in p300 binding were identified using the "bedtools intersect -wa" command. To define enhancers that do not overlap annotated promoters, EDPNew promoters were subtracted from p300-enriched ATAC-seq peaks using the "bedtools subtract" command resulting in 10,246 NK cell enhancers.

SNPsplit chromosome of origin reads analysis

Delineation of allele-informative reads was performed similarly as in reference (Xu et al. 2017). SNPs between the C57BL/6 (B6) and BALB/cJ (BALB) mouse strains were sourced from the Wellcome Sanger Institute Mouse Genomes Project dbSNP (v142). In order to perform unbiased alignment of reads originating from both the B6 and BALB genomes, SNPs marked by the database were replaced by 'N' in the mm10 reference genome that we use for alignment using SNPsplit (Babraham Institute) (Krueger and Andrews 2016). ATAC-seq datasets generated in (B6 x BALB/c) F_1 hybrid NK cells were then aligned to the N-masked genome using bowtie2 and further processed and normalized as above. Reads that overlapped the annotated sites and contained were marked as allelically informative reads after alignment and quality control using SNPsplit. Allele-informative reads were then processed and normalized as described above. ~4% of ATAC-seq reads across the dataset were allele-informative.

Enhancer deletion in primary NK cells via Cas9-RNP nucleofection

Ex vivo editing of primary mouse NK cells was carried out according to a modified version of the protocol used to modify primary human T cells described in reference (Roth et al. 2018). Cas9 was purchased from the UC Berkeley Macro Lab core (40 uM Cas9 in 20 mM HEPES-KOH, pH 7.5, 150 mM KCl, 10% glycerol, 1 mM DTT), and sgRNAs were transcribed *in vitro* according to the Corn lab online protocol (https://www.protocols.io/view/in-vitro-transcription-of-guide-rnas-and-5-triphos-bqjbmuin). NK cells were prepared by sorting day 5 IL-2 cultured NK cells from (B6 x BALB)F_1 hybrids. CD3-NKp46+ Cells were sorted to be positive for either $NKG2A^{B6}$ using the 16a11 clone or $Ly49G2^{B6}$ using the 3/25 clone, and cells were further cultured overnight in RPMI 1640 media containing 5% FCS and 1000 U/mL IL-2 (National Cancer Institute). On day 6, 1 million sorted NK cells were prepared for nucleofection using the Lonza 4D-Nucleofector per condition. Cas9 and sgRNAs were complexed at a molar

ratio of 1:2 (2.5 µL of 40 µM Cas9 was added to 2.5 µL of sgRNA suspended at 80 µM (6.5 µg) in nuclease-free H_2O). If two flanking guides were used, 1.25 µL of each were used maintaining the Cas9 to sgRNA molar ratio. Cas9-RNP was complexed for 15 minutes at 37°C and transferred to a single well of a 96-well strip nucleofection cuvette from Lonza for use with the Nucleofector 4D. 1 million sorted day 6 IL-2 cultured NK cells were resuspended in 18 µL of supplemented Lonza P3 buffer from the P3 Primary Cell kit, and added to the Cas9-RNP complex. Cells were nucleofected using the CM137 nucleofection protocol and 80 µL pre-warmed RPMI 1640 with 5% FCS was immediately added. After a 15-minute recovery period at 37°C, cells were returned to culture in 1 mL of RPMI 1640 with 5% FCS and 1000 U/mL IL-2. After 5-7 days in culture maintaining a density of approximately 1 million cells/mL, receptor expression was assayed by flow cytometry. In order to validate enhancer flanking guides (Table 2.4) an identical protocol was followed with either day 5 IL-2 cultured splenocytes, or day 5 IL-2 cultured NK cells isolated using the MojoSort NK isolation kit from Biolegend, but instead of analysis by flow cytometry, gDNA was prepared and used as a template for PCR to detect the expected deletion. Primers are shown in Table 2.5.

sgRNA target	**sgRNA sequence 5' → 3'**	
	Upstream	**Downstream**
Nkg2a$_{5'E}$	CAGGATAATTATTATGATTG	GAGGCACCGTTCAGATGCAG
Non-targeting pair 1 (nt1)	CAGTCAGGACAGGAACAAGC	CCTCACAGAAACATACATAA
Non-targeting pair 2 (nt2)	AGCTGATTCAAGACCAGCCA	CCTTTATGTATGTTTCTGTG
Cd45 (Ptprc)	GAGCCTACCAATAGTGCTG	N/A
Ly49g$_{Hss1}$	GGGCTCAAGCACTCAGAGCA	ACCAAAGTACAGCATAATAT

Table 2.4. *Ex vivo* deletion guides. Flanking sgRNAs used to delete the *Nkg2a$_{5'E}$* and *Ly49g$_{Hss1}$* are shown non-targeting (nt) controls guide pairs, and a single control guides targeting *Cd45* are also displayed.

sgRNA target	**Deletion detection primers 5' → 3'**	
	Upstream	**Downstream**
Nkg2a$_{5'E}$	1-ATGAGTGTGCAGTGGTGTCTTC 2-TGTGCCAGCCATAAGAGTTTG	1-TCATCCAAAGAGCCACAGCA 2-TCCAGATGATGGCTAACTCTCCAT
Ly49g$_{Hss1}$	1-CTGTGCATGTTGTCAATACAGTG 2-GTAGTTAGTGTCTGTTGGTTAG	1/2-CAGAATGGGCTTCTTTGTTGGTT

Table 2.5. *Ex vivo* deletion genotyping primers. Deletions were detected via nested PCR reactions. "1" indicates a primer used in the first reaction, while "2" indicates a primer used in the second reaction.

Chapter 3
Stable random monoallelic expression is pervasive and is controlled by the probabilistic properties of transcriptional enhancers

Portions of this chapter were adapted and/or reprinted with permission from "Stable random monoallelic expression is pervasive and is controlled by the probabilistic properties of transcriptional enhancers. Djem U. Kissiov[1], Sean Chen[1], Ishan D. Paranjpe[1], Alec Ethell[1], Katrine N. Madsen[1], David H. Raulet[1*"] to be submitted for publication in 2021.

Abstract

Mitotically stable random monoallelic gene expression (RME) is thought to occur in a small percentage of autosomal genes and has been viewed as the autosomal analog of X-inactivation. Here we establish an *in vivo* genetic model in the variegated natural killer (NK) cell receptor genes to study the role of transcriptional enhancers in RME. We find that the constitutively accessible enhancers of RME genes directly control the probability of target allele expression. We further find that genes previously thought to be ubiquitously expressed in defined hematopoietic lineages, *Cd45, Thy1* and *Cd8a*, are regulated in an RME fashion, suggesting stable RME is more widespread than previously appreciated. We present a probabilistic model of gene expression where gene allele expression probability is controlled by constitutively active enhancers, and we propose that previously documented RME is the extreme manifestation of this property.

Introduction

Gene alleles are generally co-regulated such that both alleles of autosomal genes are transcribed. In recent years random monoallelic expression (RME) has emerged as an important exception that may apply to ~0.5-10% of expressed genes in a given tissue (Deng et al. 2014; Gendrel et al. 2014; Gendrel et al. 2016; Gimelbrant et al. 2007; Reinius and Sandberg 2015; Xu et al. 2017; Reinius et al. 2016; Eckersley-Maslin and Spector 2014; Eckersley-Maslin et al. 2014). In this mode of expression, either allele may be expressed, or not, independently of the opposing allele, and expression is mitotically stable. RME is distinct from other classes of monoallelic expression (e.g., X-inactivation, genomic imprinting, and the allelic exclusion of antigen receptor genes and odorant receptors), in that biallelic expression occurs at an appreciable frequency, the effect is locus-specific, and that expression is largely stochastic rather than being imposed by strict feedback regulatory mechanisms (Gendrel et al. 2016).

The molecular determinants of RME are poorly understood. Recent progress has been made by analysis of clonal cell lines derived from F1 hybrids, an approach that has circumvented technical limitations in studying RME in single cells from primary tissue (Gendrel et al. 2014; Xu et al. 2017; Eckersley-Maslin et al. 2014). Surprisingly, two recent studies in F_1 clones suggested that enhancers associated with RME genes are constitutively accessible irrespective of gene or allelic expression status, whereas promoters are accessible only at active alleles (Levin-Klein et al. 2017; Xu et al. 2017). Therefore, promoter accessibility, rather than enhancer opening and activation, might be the "gatekeeper" of RME (Xu et al. 2017). Because they were constitutively open, enhancers were instead proposed to be permissive for expression at RME alleles.

The possibility that enhancers play more than a permissive role in RME is suggested by findings, more than twenty years ago, that transcriptional enhancers act primarily to influence the probability of mitotically stable expression, rather than to determine expression level per cell

(Fiering, Whitelaw, and Martin 2000; Walters et al. 1995). This was supported by studies of both ectopically inserted reporter constructs and endogenous loci where enhancer sequences were deleted, resulting in variegated and largely mitotically stable expression that was at normal levels per cell, notably at the *Igh* locus in hybridomas and at the *Cd8a* locus in thymocytes (Sleckman et al. 1997; Xu et al. 1996; Walters et al. 1995; Walters et al. 1996; Ronai, Berru, and Shulman 1999, 2002, 2004; Garefalaki et al. 2002; Ellmeier et al. 2002). It was further shown at the *Cd8a* locus that the enhancer deletion-associated variegation effect in thymocytes is imparted entirely *in cis* and is stable over several mitotic divisions (Garefalaki et al. 2002). RME-like phenotypes resulting from enhancer deletions in various systems broadly support the notion that gene alleles are stochastically activated, and enhancers influence the probability of expression *in cis* to ensure the faithful lineage-appropriate expression of genes.

Analysis of the role of enhancers in naturally variegated/RME genes has been hampered by the lack of a powerful *in vivo* genetic model of RME. The *Ly49* receptor family genes encoded on mouse chromosome 6 are a frequently cited example of RME (Gendrel et al. 2016; Chess 2012; Reinius and Sandberg 2015; Eckersley-Maslin and Spector 2014). They are expressed in a variegated (Raulet, Vance, and McMahon 2001; Yokoyama et al. 1990) monoallelic (Held, Roland, and Raulet 1995), stochastic and largely mitotically stable fashion (Raulet, Vance, and McMahon 2001), resulting in subpopulations of NK cells that express random combinations of the receptors and consequently exhibit distinct reactivities for cells expressing diverse MHC I ligands. Regulation of each gene is independent, and expression of one *Ly49* gene has only a small effect on the likelihood that the others are expressed (Tanamachi et al. 2001). While many examples of RME are not associated with a known biological function, the cellular subsets generated by MHC I receptor variegation form the basis of the "missing self" mode of NK cell target recognition (Raulet, Vance, and McMahon 2001). As with RME broadly, the molecular underpinnings of NK receptor variegation remain poorly understood. Importantly, competition for *cis*-regulatory elements (seen in the olfactory receptor and protocadherin loci) is not required for variegation, as a *Ly49a* genomic transgene ectopically integrated in different genomic sites was usually expressed with a frequency similar to the frequency of expression of the native *Ly49a* gene (~17% of NK cells) (Tanamachi et al. 2004).

Our central hypothesis is that enhancers, rather than simply being permissive for RME, directly control the probability of expression of *Ly49* genes—and RME alleles generally—in a stochastic and binary fashion. In this model RME reflects limiting enhancer action, and enhancer activation (which is constitutive and deterministic) and stable target gene expression (which is probabilistically enacted by enhancers) are decoupled. Enhancer action by one or more enhancers directly controls the likelihood of stable gene expression but does not ensure it. This model explains the previous observations that a) enhancers of RME genes are constitutively accessible while the promoters are accessible only on expressed alleles (Levin-Klein et al. 2017; Xu et al. 2017), and b) that enhancer deletion results in stable RME-like phenotypes in various systems (Garefalaki et al. 2002; Ronai, Berru, and Shulman 1999, 2002). In the case of enhancer deletion-associated variegation of genes that are normally expressed from both alleles, the residual yet constitutive enhancer activity is presumably sufficient to enact target gene activation at only a fraction of alleles, and we propose that this mechanism is shared by natural RME genes.

We previously identified a DNaseI hypersensitive element, *Hss1*, ~5kb upstream of the *Ly49a* gene that is conserved in other *Ly49* genes (Tanamachi et al. 2004). Based on the requirement for *Hss1* for *Ly49a* gene expression in a transgene construct, *in vitro* enhancer assays, knockout data herein and in Chapter 4, this element functions as an enhancer (Tanamachi et al. 2004; Gays, Taha, and Brooks 2015). We also previously developed mAbs that discriminate allelic differences in several Ly49 receptors as well as in the related NKG2A inhibitory receptor, providing key tools for investigating RME of genes in this family (Vance et al. 2002; Tanamachi et al. 2001). We have exploited this system to directly test the effects of enhancers on allelic expression probabilities at the single cell level and have built on those results to provide a more general model of the role of enhancers in RME as well as in other developmentally regulated genes.

Results

Constitutively accessible enhancers upstream of the RME Ly49a and Nkg2a genes are required for expression by NK cells

The natural killer cell complex (NKC) genes are clustered in an ~1 Mb stretch of chromosome 6. The *Ly49* family members are expressed in a mitotically stable RME fashion and each harbors an accessible chromatin site (*Hss1*) ~5 kb upstream of the TSS (Fig. 3.1, A and B; Fig. 3.2A). Importantly, $Ly49a_{Hss1}$ and the other *Hss1* elements are enriched in H3K4me1 relative to H3K4me3, indicative of enhancer-like chromatin (Fig. 3.1B and Fig. 3.2A). Previous results generated with a genomic transgene suggested that $Ly49a_{Hss1}$ is essential for expression (Tanamachi et al. 2004). We tested this prediction by deleting $Ly49a_{Hss1}$ in the germline via CRISPR/Cas9 with flanking sgRNAs (Fig. 3.1 and Fig. 3.3A). $Ly49a^{Hss1\Delta/Hss1\Delta}$ mice completely lacked Ly49A expression (Fig. 3.1, B to D; Fig. 3.3B), but expression of other Ly49 receptors was unaffected (Fig. 3.3B). Furthermore, $Ly49a^{+/Hss1\Delta}$ mice displayed an intermediate phenotype (~9% Ly49A+ NK cells) compared to wildtype littermates (~15% Ly49A+) (Fig. 3.1, C and D).

Approximately 2 kb upstream of the related *Nkg2a* inhibitory receptor gene, which is also an RME gene (Vance et al. 2002), we noticed an enhancer-like element, which, like *Hss1* elements. is bound by Runx3 and T- bet (Fig. 3.1B, Fig. 3.2A). We named this element $Nkg2a_{5'E}$. We deleted $Nkg2a_{5'E}$ in the germline via embryo electroporation with Cas9 RNP (CRISPR-EZ) (Modzelewski et al. 2018) (Fig. 3.1B and Fig. 3.3C). As in $Ly49a^{Hss1\Delta/Hss1\Delta}$ mice, NK cells in $Nkg2a^{5'E\Delta/5'E\Delta}$ mice did not express NKG2A, and heterozygous mice displayed a reduced frequency of NKG2A+ NK cells (Fig. 3.1, E and F; Fig 3.3).

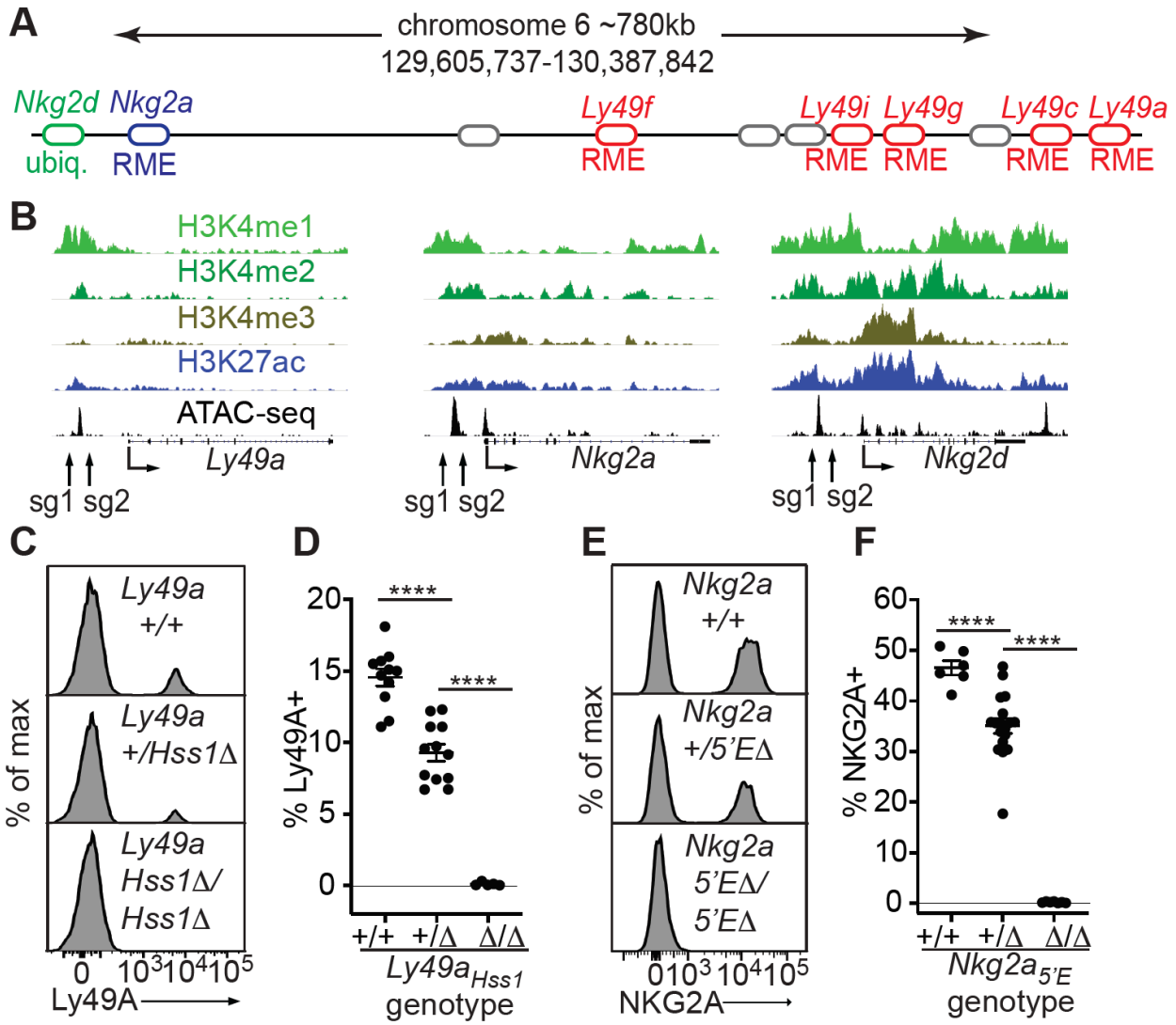


Fig. 3.1. The constitutively accessible *Ly49a*$_{Hss1}$ and *Nkg2a*$_{5'E}$ enhancers are required for gene expression. (**A**) Schematic depicting approximate location of key *Ly49* (red), *Nkg2a* (blue) and *Nkg2d* (green) genes in the natural killer cell gene complex on mouse chromosome 6. Grey ovals depict locations of other *Ly49* gene not discussed in the present study. (**B**) Normalized ChIP-seq and ATAC-seq results mined from raw data from reference (Lara-Astiaso et al. 2014), showing enhancer and promoter associated histone modifications at *Ly49a*, *Nkg2a* and *Nkg2d*. The locations of sgRNAs used to delete enhancers in this study are shown. (**C-D**) Ly49A staining by wildtype, heterozygous, and homozygous *Ly49a*$_{Hss1}$ deletion littermates. In **D**, data are combined from two independent experiments (****P <0.0001 using Welch's t-test). Error bars denote SEM, (n=5-12) (**E-F**) Data as in **C-D** for *Nkg2a*$_{5'E}$. Data in **E** are combined from two independent experiments with the *Nkg2a*$_{5'E}$(B3Δ) allele (Fig. 3.3) and are representative of three independent experiments conducted with both the *Nkg2a*$_{5'E}$(B3Δ) and *Nkg2a*$_{5'E}$(B3Δ) alleles. (****P <0.0001 using Welch's t-test). Error bars denote SEM, (n=6-18).

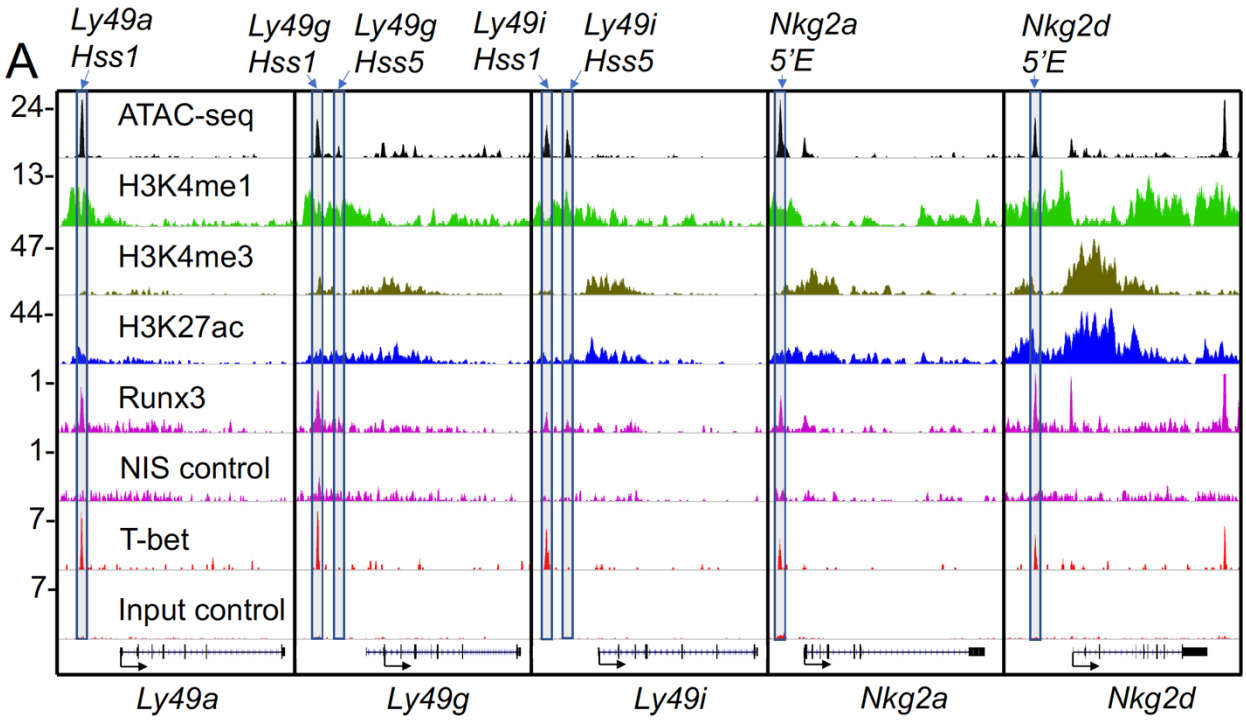


Fig. 3.2. Enhancer landscape at *Ly49a, Ly49g, Ly49i, Nkg2a* and *Nkg2d* loci in NK cells. (**A**) The ATAC-seq data was generated in house, while the H3K4me1, H3K4me3 and H3K27ac ChIP-seq in NK cells was mined from raw data in reference (Lara-Astiaso et al. 2014). The Runx3 ChIP-seq was mined from raw data in reference (Levanon et al. 2014). The T-bet ChIP-seq was mined from raw data in reference (Shih et al. 2016).

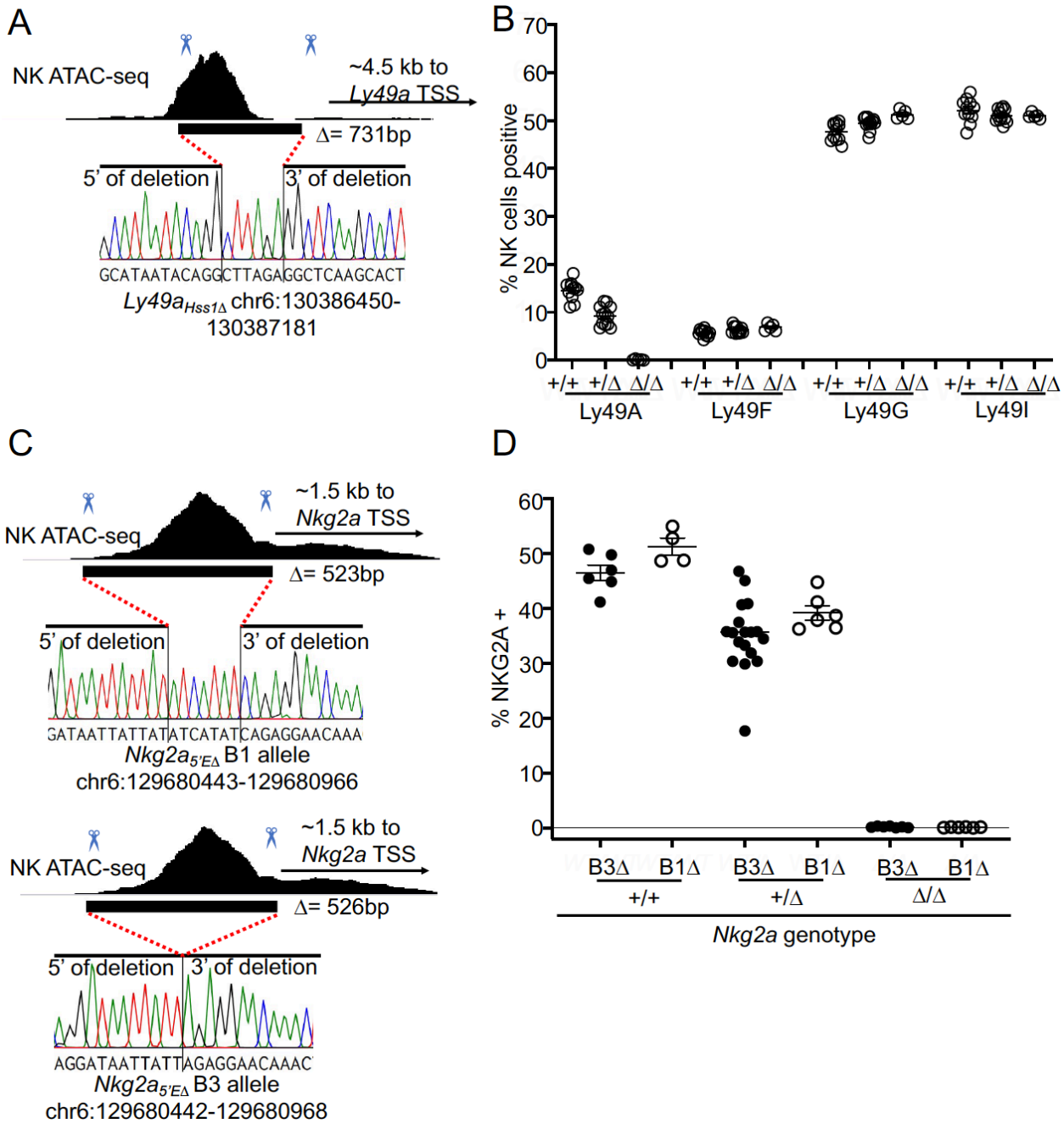


Fig. 3.3. *Ly49a*$_{Hss1\Delta}$ and *Nkg2a*$_{5'E\Delta}$ alleles employed in the study. (**A**) Genomic position and sequence of the *Ly49a*$_{Hss1\Delta}$ allele analyzed in this study. The black bar shows the location of the CRISPR/Cas9 generated in/del based on Sanger sequencing of a PCR amplicon spanning the region. (**B**) Percentages of cells expressing indicated Ly49 receptors in *Ly49a*$^{Hss1\Delta/Hss1\Delta}$ mice, heterozygous and wildtype littermates, from flow cytometry analyses. Data are combined from two independent experiments, n=5-12. (**C**) *Nkg2a*$_{5'E\Delta}$ alleles generated and analyzed in this study, as in "A." (**D**) Percentages of cells expressing NKG2A in mice with the genotypes shown. Data are combined from two independent experiments with the *Nkg2a*$_{5'E}$-B3Δ allele, and one experiment with the *Nkg2a*$_{5'E}$-B1Δ allele.

$Nkg2a_{5'E}$ and $Ly49a_{Hss1}$ enhancers act entirely in cis

We show in Chapter 4 that $Nkg2a_{5'E}$ (and $Ly49g_{Hss1}$) is constitutively accessible upstream of both active and silent alleles, consistent with the behavior of elements proximal to RME genes genome-wide (Levin-Klein et al. 2017; Xu et al. 2017). Additionally, we previously observed that $Ly49a_{Hss1}$ is hypersensitive to DNase I in Ly49A- cells as well as Ly49A+ cells (Tanamachi et al. 2004). Whether the action of these constitutively accessible enhancers at RME genes is entirely *in cis* (as is characteristic of classical enhancers of ubiquitously expressed genes) or is coordinated in *trans* via an RME-specific mechanism is not known. We addressed this problem with allele-discriminating antibodies to test the role of $Nkg2a_{5'E}$ and $Ly49a_{Hss1}$ in regulating allelic expression *in vivo*. We hypothesized that the reduced expression frequencies in enhancer deletion heterozygotes (Fig. 3.1, C to F) were entirely due to complete loss of expression of the allele harboring the enhancer deletion, i.e. each copy of the constitutively accessible enhancer acts independently and *in cis*.

In order to assess allelic *Nkg2a* regulation by the constitutively accessible $Nkg2a_{5'E}$, we generated F_1 hybrid mice in which one allele lacked $Nkg2a_{5'E}$ in order to assess the activity of $Nkg2a_{5'E}$ on the other allele. We crossed heterozygous deletion mice to BALB/c, generating $Nkg2a^{B6\text{-}5'E\Delta/BALB+}$ and wildtype littermate controls ($Nkg2a^{B6+/BALB+}$). Using an $NKG2A^{B6}$ reactive mAb (16a11) (Vance et al. 2002) in conjunction with an $NKG2A^{B6+BALB}$ mAb (20d5) we could discriminate cells expressing either, both or neither of the NKG2A alleles (Fig. 2A). $Nkg2a^{B6\text{-}5'E\Delta/BALB+}$ NK cells did not express $Nkg2a^{B6}$ (Fig. 3.4, B and C). In parallel, the percentages of NK cells expressing only the BALB/c allele, or neither allele, increased commensurate with expectations calculated under the assumption that expression is not influenced by the status of $Nkg2a_{5'E}$ on the other allele (Fig. 3.4C). We conclude that the constitutively accessible $Nkg2a_{5'E}$ functions *in cis* and independently of the activity of the other copy.

We peformed the same genetic experiment with the $Ly49a^{Hss1\Delta}$ allele, and generated $Ly49a^{B6\text{-}Hss1\Delta/BALB+}$ mice and wildtype F_1 littermate controls. We observed complete loss of cells expressing $Ly49A^{B6}$, while cells expressing $Ly49A^{BALB}$ were unaffected (Fig. 3.4, E and F), mirroring the results we obtained for the *Nkg2a* locus.

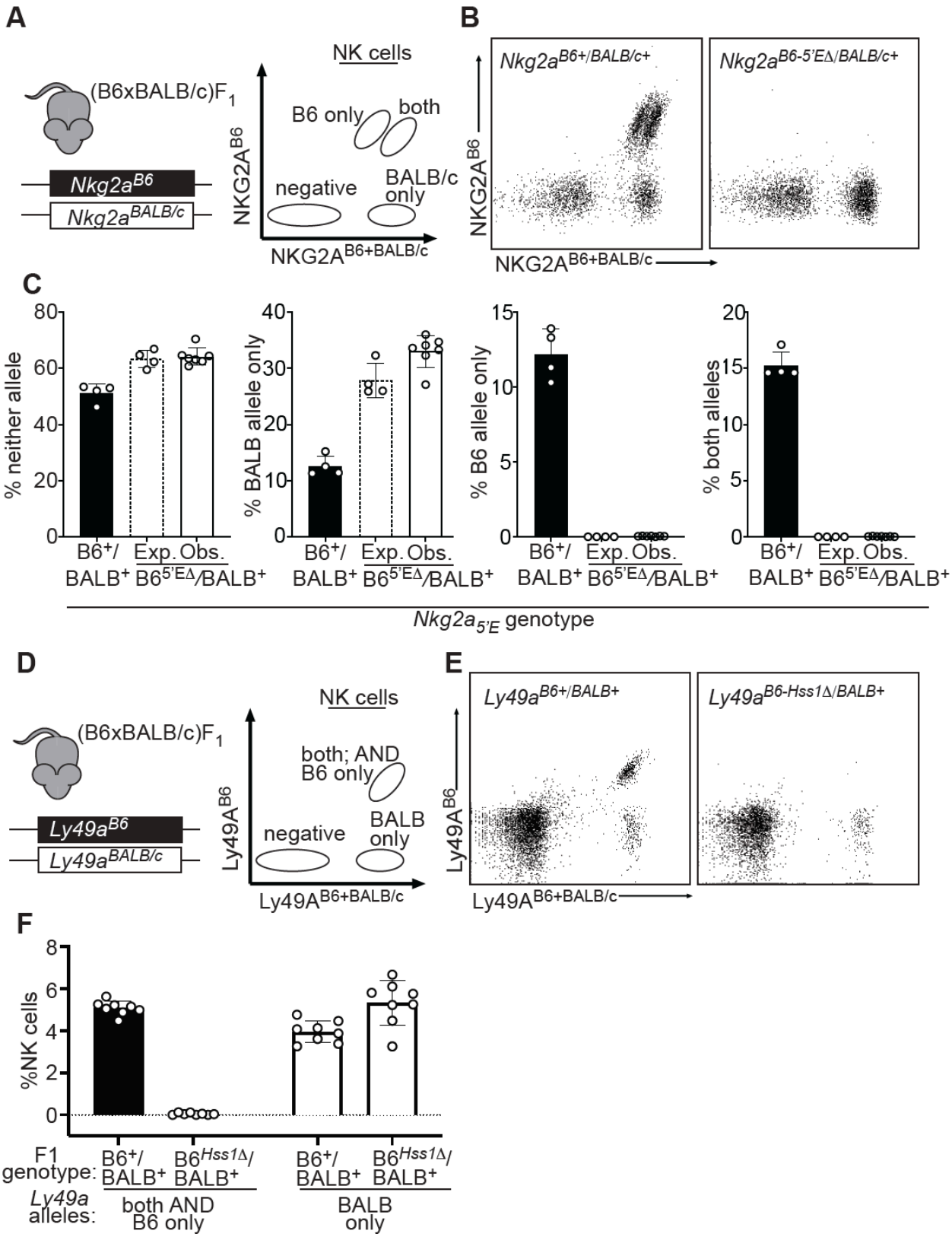


Fig. 3.4. The constitutively accessible *Ly49a*$_{Hss1}$ and *Nkg2a*$_{5'E}$ RME gene enhancers act entirely *in cis*. (**A**) Schematic of (B6 x BALB/c)F_1 hybrid NK cell staining pattern using 16a11 (NKG2A^{B6} reactive) and 20d5 (NKG2A$^{B6+BALB}$ reactive) antibodies. (**B**) Representative dot plots

displaying staining of (B6 x BALB/c)F1 hybrid splenic NK cells using 16a11 and 20d5 (*n*=4-7). (**C**) Expected (dotted bar) and observed (solid white bar) percentages of populations in *Nkg2a*$^{B6\text{-}5'E\Delta/BALB\text{-}5'E+}$ mice, compared to wildtype littermate (B6 x BALB/c)F1 hybrid mice (black bar). Expected frequencies are calculated assuming stochastic *cis* regulation of alleles, and assumes that cells expressing only the B6 allele in WT hybrids merge with cells expressing neither allele in the mutant hybrids, and that cells expressing both alleles in the WT hybrids merge with those expressing only the BALB allele in the mutant hybrids (detailed in methods). Data are representative of two independent experiments. (**D**) Schematic of (B6 x BALB/c)F_1 NK cell staining pattern using A1 (Ly49A^{B6} reactive) and JR9 (Ly49A$^{B6+BALB}$ reactive) antibodies. (**E**) Representative dot plots displaying (B6 x BALB/c)F1 hybrid NK cells using A1 and JR9. (**F**) Percentages of NK cells expressing the indicated Ly49A alleles (*n*=8-9). Error bars in all panels denote SEM.

Deletion of Nkg2d$_{5'E}$ is sufficient to recapitulate stable RME in Nkg2d.

In theory, variegation might be actively imposed *in cis* by a putative specialized element dedicated to establishing variegation, i.e. a "variegator" or "switch," which might be the promoter, the enhancer or an entirely independent element, as has been previously proposed (Saleh et al. 2004). Instead, we hypothesize that RME of NK receptor genes, among others, is an extreme manifestation of fundamental properties of gene activation by enhancers at many loci. Surprisingly, many loci are competent to display stable variegation, as revealed by experiments in which variegation accompanies the deletion of enhancer elements (Walters et al. 1995; Ronai, Berru, and Shulman 1999; Garefalaki et al. 2002; Ellmeier et al. 2002; De Gobbi et al. 2017; Ng et al. 2018), implying this expression pattern is rooted in general rather than specialized properties. As one test of this hypothesis in the case of NK receptor genes, we predicted that we could generate an RME NK receptor gene *de novo* by deleting an enhancer that contributes to (but is not strictly required for) the expression of a receptor gene that is normally expressed by all NK cells.

We chose to analyze the *Klrk1/Nkg2d* gene encoding the NKG2D immunostimulatory receptor, since it is expressed by all NK cells (Wensveen, Jelencic, and Polic 2018), is distantly related to NKG2A and Ly49 genes, and is flanked by sequences rich in enhancer-like chromatin, suggesting possible regulation by multiple enhancers (Fig. 3.1B). We targeted an ATAC-accessible site ~5 kb upstream of the *Nkg2d* gene (*Nkg2d$_{5'E}$*), which displays classical chromatin features of enhancers and binds Runx3 and T-bet, as do *Hss1* and *Nkg2a$_{5'E}$* elements (Fig. 3.1B and Fig. 3.2) via CRISPR-EZ using flanking guides. We isolated two founders harboring independent *Nkg2d$^{5'E\Delta}$* deletion alleles (Fig. 3.1B and Fig. 3.6A). Remarkably, expression of NKG2D was variegated in *Nkg2d$^{5'E\Delta/5'E\Delta}$* animals, with only ~65% of NK cells expressing NKG2D (Fig. 3.5, A and B). Expression level per cell, depicted as mean fluorescence intensity (MFI) of staining, was only modestly affected in *Nkg2d$^{5'E\Delta/5'E\Delta}$* animals, and to an extent consistent with largely monoallelic vs biallelic expression (Fig. 3.5C and Fig. 3.6C), suggesting the primary role of *Nkg2d$_{5'E}$* is in regulating the probability rather than the degree of *Nkg2d* expression.

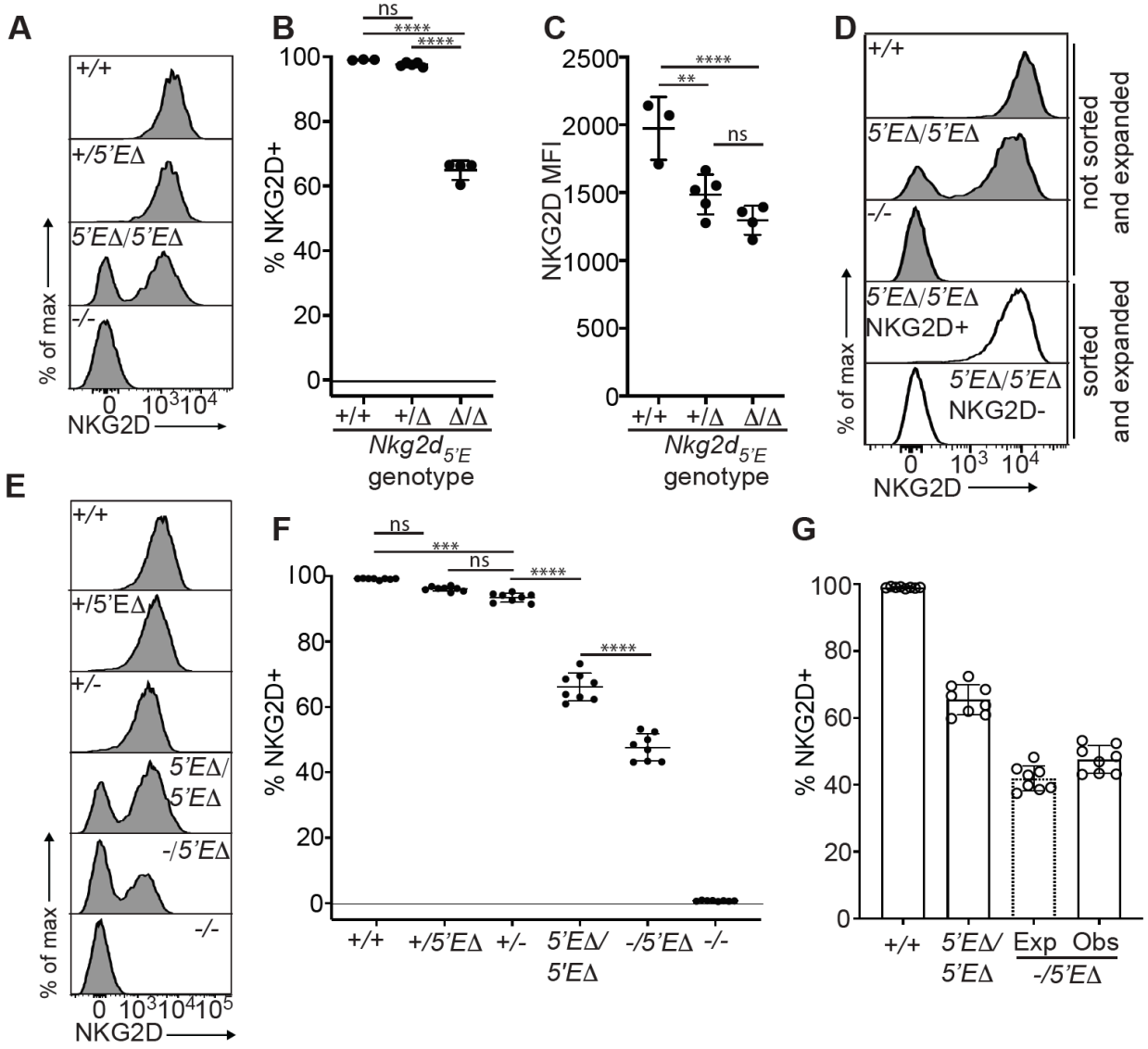


Fig. 3.5. *Nkg2d*$_{5'E}$ deletion results in RME that is mitotically stable. (**A-C**) NKG2D staining of splenocytes from wildtype, heterozygous and homozygous *Nkg2d*$_{5'E}$ deletion littermates, and an *Nkg2d*$^{-/-}$ mouse for comparison. Results with the "*Nkg2d*$_{5'E}$-B1Δ" allele are shown (n=3-5) and are representative of at least four independent experiments with two independently isolated deletion alleles (Fig. 3.6). (**D**) Splenocytes from a *Nkg2d*$^{5'E\Delta/5'E\Delta}$ mouse were activated by culturing in medium containing IL-2 for 2-3 days before sorting NKG2D+ and NKG2D- NK cells. Sorted cells were expanded in fresh IL-2 containing medium for 8-10 days before staining for NKG2D expression. Unsorted control NK cells in grey, expanded NKG2D+ and NKG2D- cultures in white. (**E**) Representative staining of splenic NK cells from mice of six genotypes from the indicated genotypic series. "+" refers to the wildtype allele, "-" refers to the gene knockout allele, "Δ" refers to *5'E* deletion. (**F**) Combined staining data from from two independent experiments. (**G**) Expected and observed percentages of NKG2D+ NK cells in *Nkg2d*$^{5'E\Delta/-}$ mice. Expected expression is calculated based on observed NKG2D+ cells in *Nkg2d*$^{5'E\Delta/5'E\Delta}$ mice, assuming stochastic allelic expression. Statistics were computed by Ordinary one-way ANOVAs ((*P <0.05; **P <0.01; ***P <0.001; ****P <0.0001). All error bars represent SEM.

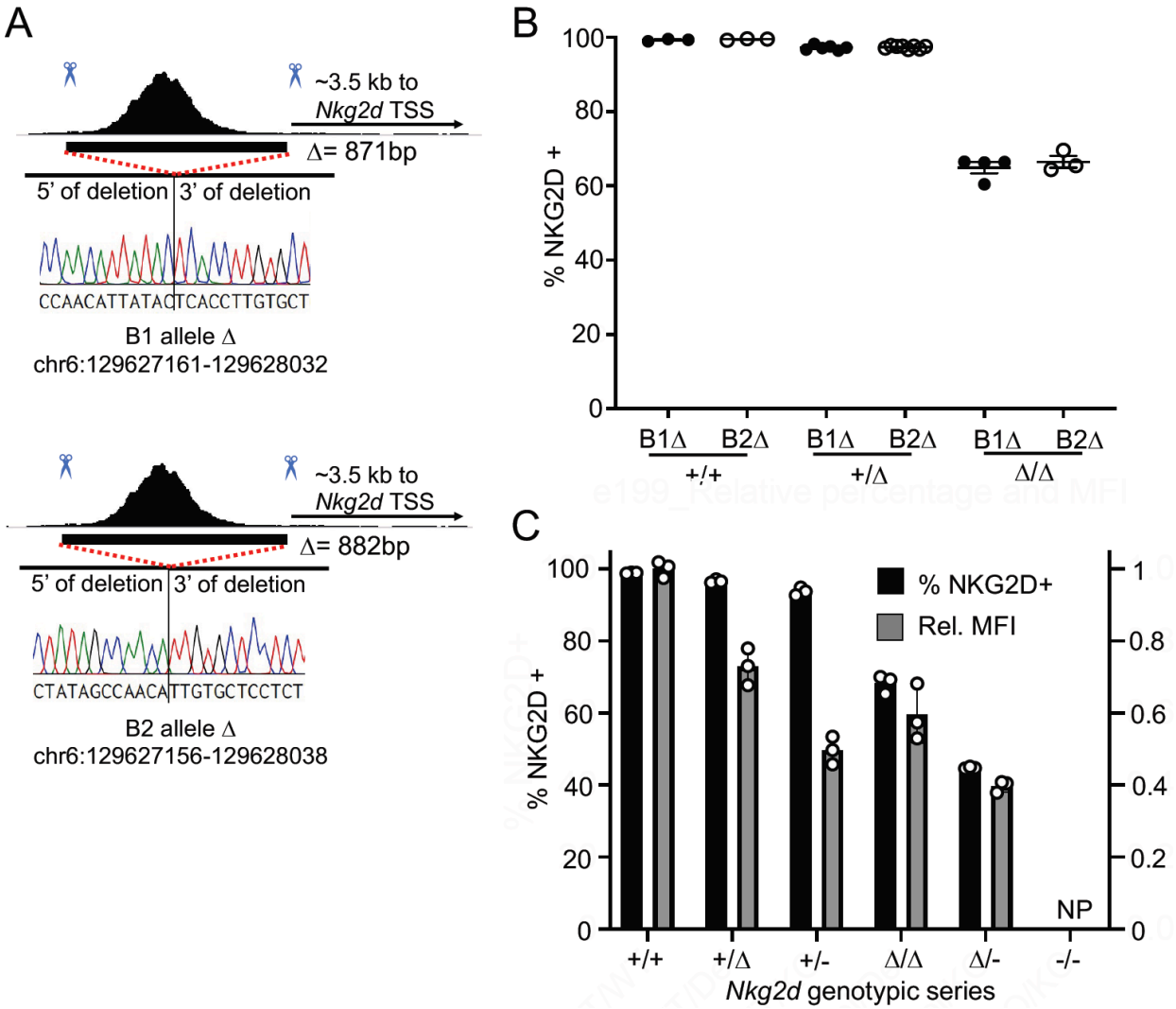


Fig. 3.6. *Nkg2d*$_{5'E\Delta}$ alleles employed in this study. (**A**) Genomic position and sequences of *Nkg2d*$_{5'E\Delta}$ alleles (see Fig. 3.3 legend for details). (**B**) Percentages of NKG2D+ cells in mice with the indicated *Nkg2d* genotypes, depicting one experiment each with the B1Δ and B2Δ alleles. (**C**) Comparison of percentages of % NKG2D+ NK cells (left y-axis) and mean staining intensity of NKG2D staining (right y-axis) from mice with the indicated *Nkg2d* genotypes. Data are representative of two independent experiments.

We next asked whether NKG2D variegation was mitotically stable—a key feature of both variegated NK receptor expression and RME broadly. NK cells from the knockouts were stimulated for 2-3 days in IL-2 before sorting NKG2D+ and NKG2D- populations, which were returned separately to culture in IL-2 for an additional 8-10 days, where they underwent an ~10-100 fold expansion. The NKG2D+ and NKG2D- phenotypes were highly stable despite the extensive proliferation (<2%-8% presence of the opposing phenotype) (Fig. 3.5D).

In order to determine whether *Nkg2d* alleles in *Nkg2d*$^{5'E\Delta/5'E\Delta}$ animals are regulated independently resulting in RME, we generated heterozygous mice with the *Nkg2d*$^{5'E\Delta}$ allele on one chromosome and an *Nkg2d* knockout allele on the other (*-/5'E*Δ). The frequency of NKG2D+ cells was reduced in these mice compared to *5'E*Δ/*5'E*Δ mice (Fig. 3.5,E and F), and nearly matched the expected frequency calculated under the assumption of independent regulation of alleles, strongly arguing that expression follows an RME pattern (Fig. 3.5G). The NKG2D fluorescence intensity of NKG2D+ cells in *Nkg2d*$^{5'E\Delta/5'E\Delta}$ animals appears slightly higher than in *Nkg2d*$^{+/-}$ animals, consistent with a proportion of cells expressing both *Nkg2d* alleles, a feature characteristic of natural RME (Gendrel et al. 2016; Eckersley-Maslin and Spector 2014) (Fig. 3.5E and Fig. 3.6C).

Nkg2d expression in Nkg2d$^{5'ED/5'ED}$ *mice mimics the expression and accessibility features of the variegated NK receptor genes.*

We asked whether the RME of *Nkg2d* observed in *Nkg2d*$^{5'E\Delta/5'E\Delta}$ mice fully recapitulates the stochastic expression pattern of the naturally variegated NK receptor genes. We observed that expression of NKG2D in *Nkg2d*$^{5'E\Delta/5'E\Delta}$ mice was approximately randomly distributed with respect to other variegated NK receptors, including NKG2A, Ly49G2 or Ly49I (Fig. 3.7A), suggesting that variegation of NKG2D in these mice is stochastic and does not reflect loss of expression in a specific NK cell subpopulation. The stochasticity of expression was further assessed by calculating a predicted frequency of "double positive" (e.g NKG2D+NKG2A+) cells, by multiplying the frequencies of cells expressing one or the other receptor (the "product rule" (Raulet et al. 1997)), and comparing to the observed frequencies. The observed frequencies nearly matched the calculated expected frequencies for NKG2A, Ly49G2 and Ly49I (Fig. 3.7B). These results established that the stochastic, mosaic and mitotically stable expression of variegated NK receptor genes can be fully explained by weakened enhancer complexes, with no need to invoke more elaborately functioning specialized "variegation" elements.

We next asked whether the molecular architecture of the RME *Nkg2d* locus in *Nkg2d*$^{5'E\Delta/5'E\Delta}$ mice resembles that of the natural RME receptors in NK cells and RME genes broadly (Xu et al. 2017) (*cite other paper in issue*). We sorted NKG2D+ and NKG2D- cells from *Nkg2d*$^{5'E\Delta/5'E\Delta}$ mice (Fig. 3.7C), and performed ATAC-seq. Cells of the *Nkg2d*$^{5'E\Delta/5'E\Delta}$ genotype lack the *5'E* element, and therefore no reads map to *5'E* (Fig. 3.7D). Cells expressing NKG2D displayed robust accessibility at the *Nkg2d* promoter region, while cells negative for NKG2D had markedly reduced promoter accessibility, mirroring the promoter dynamics of other RME NK receptor genes. A proximal site 3' of the gene, however, had unaltered accessibility in NKG2D- cells, reminiscient of proximal elements at RME loci genome-wide (Xu et al. 2017)

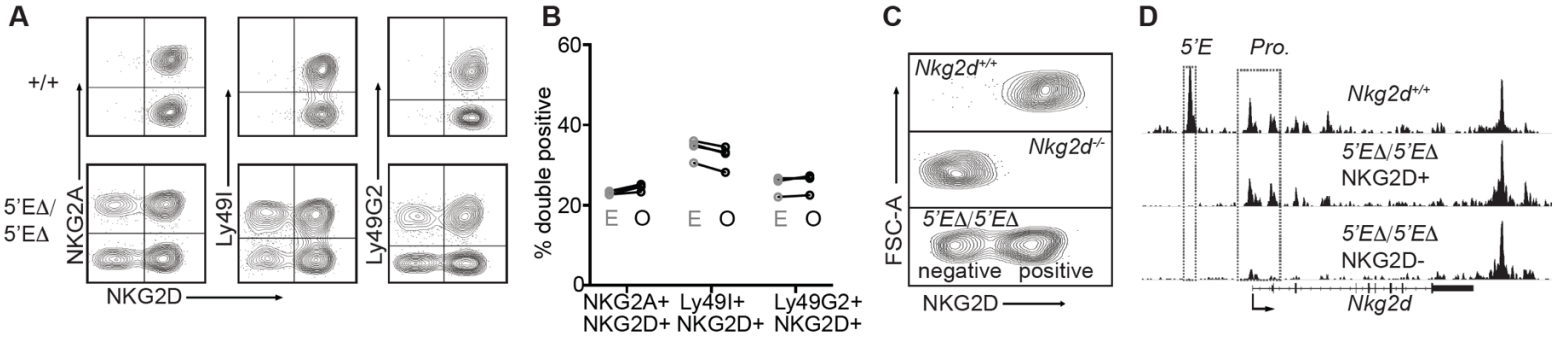


Figure 3.7. Expression pattern and chromatin accessibility features of the *Nkg2d*$_{5'E\Delta}$ allele recapitulate that of RME NK receptor genes. (**A**) Stochastic co-expression of NKG2D and NKG2A, Ly49I or Ly49G by NKp46+ NK cells in *Nkg2d*$^{5'E\Delta/5'E\Delta}$ mice. WT mice are shown for comparison. (**B**) Expected and observed percentages of cells coexpressing the indicated receptors in *Nkg2d*$^{5'E\Delta/5'E\Delta}$ mice. Expected ("E") percentageswere calculated by mutiplying individual frequencies of positive cells. Each expected percentage is paired with observed ("O") data from the same *Nkg2d*$^{5'E\Delta/5'E\Delta}$ animal (*n*=4). Data are representative of two independent experiments. (**C**) NKG2D staining of gated NK cells from *Nkg2d*$^{5'E\Delta/5'E\Delta}$ mice before sorting, with wildtype and knockout staining controls. (**D**) Normalized ATAC-seq tracks generated from NKG2D+ and NKG2D- cells sorted from the *Nkg2d*$^{5'E\Delta/5'E\Delta}$ mouse shown in (C), with control WT cells for comparison.

The striking similarity between RME of *Nkg2d* and that seen at other NK receptors and broadly in other tissues (Gendrel et al. 2016) suggests that the mechanism driving naturally occuring RME is a cognate of enhancer deletion-associated variegation. The results of our experiments with the normally ubiquitously-expressed *Nkg2d* gene locus powerfully argue that stable RME can be recapitulated in full by weakened or limiting enhancer action.

Expression likelihood of the RME Ly49g locus is controlled by cis-acting enhancers

We wondered whether the principles underlying enhancer deletion-associated variegation directly control the probability of natural RME gene expression. We hypothesized that if limiting enhancer action is responsible for RME and acts quantitatively, the frequency of expression of the RME gene will be reduced by deleting a "secondary" enhancer that is not absolutely required for gene expression. We chose to target the *Ly49g* locus since it is expressed relatively frequently by NK cells (~50%) and contains both an *Hss1* element and another accessible site with enhancer marks ~2kb downstream of *Hss1*, which we named $Ly49g_{Hss5}$ (Fig. 3.8A and Fig. 3.2). We show in Chapter 4 that like *Hss1* elements, $Ly49g_{Hss5}$ is accessible in both Ly49G2+ and Ly49G2- cells, suggesting it is a constitutively active enhancer of the RME *Ly49g* gene. Notably, the cognate site at the less frequently expressed *Ly49a* locus (~17% of NK cells) is not accessible to an appreciable degree, leading us to hypothesize that A) *Hss5* might not be strictly required for expression and B) the $Ly49g_{Hss5}$ element may contribute to the relatively high expression frequency of Ly49G2 (Fig. 3.8A). We isolated two independent $Ly49g^{Hss5\Delta}$ founder mice generated by CRISPR-EA (Fig. 3.9A).

Ly49G2 expression frequency was reduced in $Ly49g^{Hss5\Delta/Hss5\Delta}$ mice, from ~50% of NK cells in wildtype to ~35%, while heterozygous mice displayed an intermediate phenotype (Fig. 3.8, B and C). These results suggest that the primary role of the $Ly49g_{Hss5}$ enhancer is to raise the probability of gene expression (Fig. 3.8, B and D; Fig. 3.9). To test whether $Ly49g_{Hss5}$ acts entirely *in cis*, we crossed $Ly49g^{Hss5\Delta}$ to BALB/c mice, generating $Ly49g^{B6\text{-}Hss5\Delta/BALB+}$ mice and wildtype F_1 controls. We employed a Ly49G2-allele-specific staining protocol with NK cells from F1 mice that distinguishes B6/BALB single positive, double positive and double-negative cells (Fig. 3.8E) (Tanamachi et al. 2001). The populations expressing the $Ly49G2^{B6}$ alleles were reduced in $Ly49g^{B6\text{-}Hss5\Delta/BALB+}$ mice, in the proportion expected under probabilistic action of *Hss5 in cis* (Fig. 3.8, E and F). Furthermore, the populations expressing neither allele or only $Ly49G2^{BALB}$ increased in the expected proportion. We conclude that the $Ly49g_{Hss5}$ enhancer contributes to the relatively high expression frequency of *Ly49g*, and RME of *Ly49g* is the sum total of independent regulatory events on each allele. These data show that the RME frequency of *Ly49g* is directly controlled by enhancers, and suggest that the enhancer deletion-associated variegation phenomenon is directly relevant to the regulation of a gene that is naturally expressed in an RME fashion.

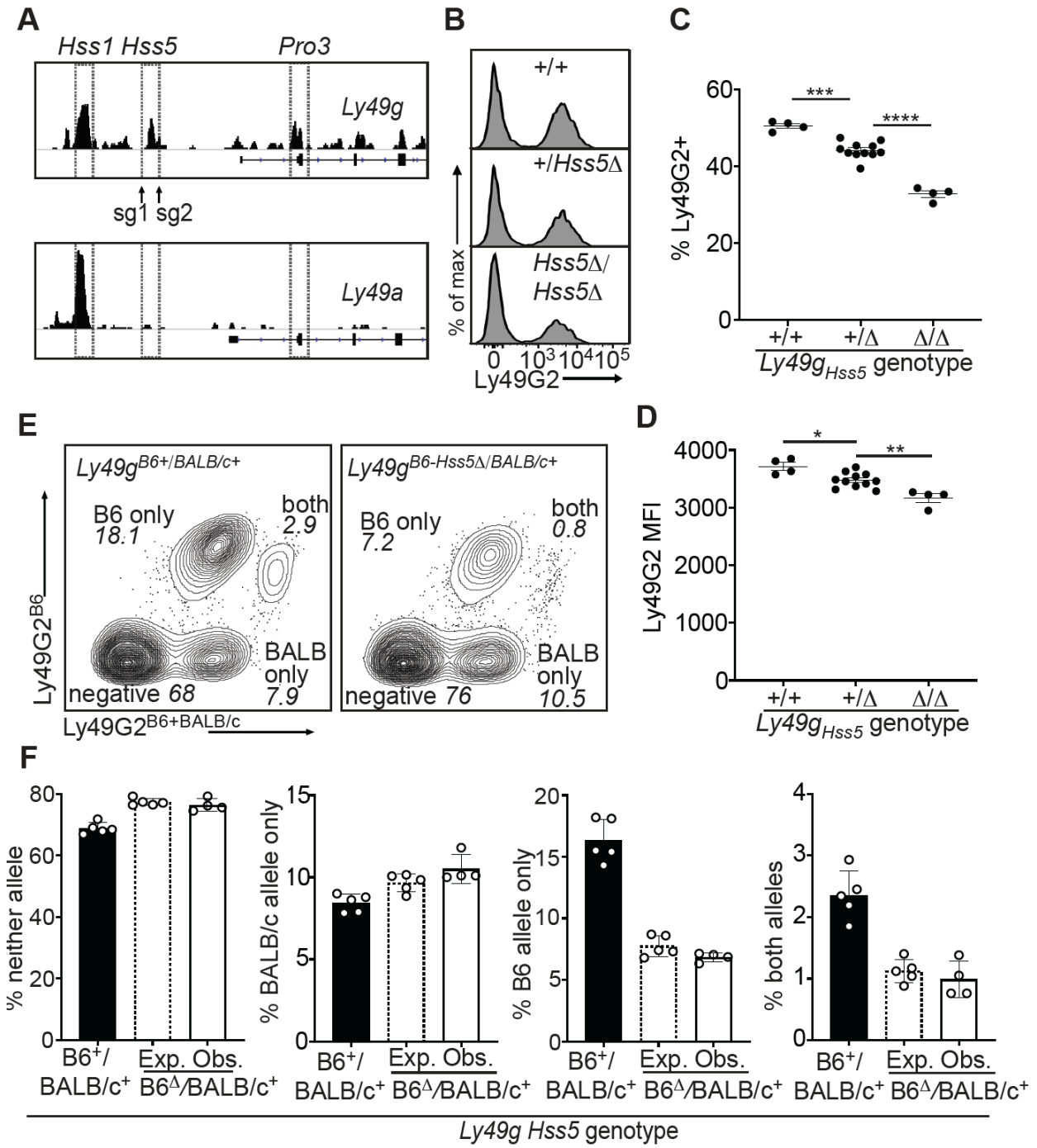


Fig. 3.8. A minor enhancer acts in *cis* to contribute to Ly49G2 expression frequency. (**A**) Normalized ATAC-seq tracks of the *Ly49a* and *Ly49g* loci in bulk NK cells, mined from reference (Lara-Astiaso et al. 2014). *Hss1 and Hss5* enhancers and the *Pro3* promoter are highlighted. The location of sgRNA sequences used to generate $Ly49g_{Hss5\Delta}$ alleles are shown. (**B**) Ly49G2 staining of NK cells in representative $Ly49g^{Hss5\Delta/Hss5\Delta}$ mice, compared to wildtype and heterozygous littermate controls. (**C-D**) Quantified Ly49G2 staining showing percentages and mean fluorescence intensities. Data are from mice with the "$Ly49g_{Hss5\Delta}$-B2Δ" allele, and are representative of two experiments conducted with independently generated alleles (Fig. 3.9) (*n*=4-11). (**P* <0.05; ***P* <0.01; ****P* <0.001; *****P* <0.0001 computed using an ordinary one-way

ANOVA). (**E**) Representative flow cytometry plots using 3/25 (Ly49G2^{B6} reactive) and 4D11 (Ly49G2$^{B6+BALB/c}$ reactive) antibodies to stain NK cells in a *Ly49g*$^{B6\text{-}Hss5\Delta/BALB/c\text{-}Hss5+}$ animal (right) and a wildtype littermate control (left). Population percentages are indicated in italics under the population name. (**F**) Expected and observed percentages of the populations depicted in "E" in wildtype (B6 x BALB/c)F_1 hybrid mice (black bar) and in F1 *Ly49g*$^{B6\text{-}Hss5\Delta/BALB/c\text{-}Hss5+}$ littermates (dotted line is expected, solid line is observed). Expected frequencies were calculated assuming stochastic *cis* regulation of alleles, and assumed that the population of cells expressing only the B6 allele in WT mice merge with cells expressing neither allele in mutant mice, and that cells expressing both alleles that would express the B6 allele in WT hybrids merge with those expressing only the BALB/c allele in mutant hybrid mice (calculations explained in depth in methods). Note that the genetic background of the mice significantly influences *Ly49g* expression even in WT mice, presumably reflecting trans-acting events (e.g. each *Ly49g*$^{B6+}$ allele is expressed on ~31% of NK cells in B6 mice, but only ~19% in F_1 hybrid mice). Therefore expected data are calculated using Ly49G2^{B6} expression frequencies in *Ly49g*$^{B6\text{-}5'E+/BALB/c+}$ mice. Data are representative of two independent experiments. All error bars represent SEM.

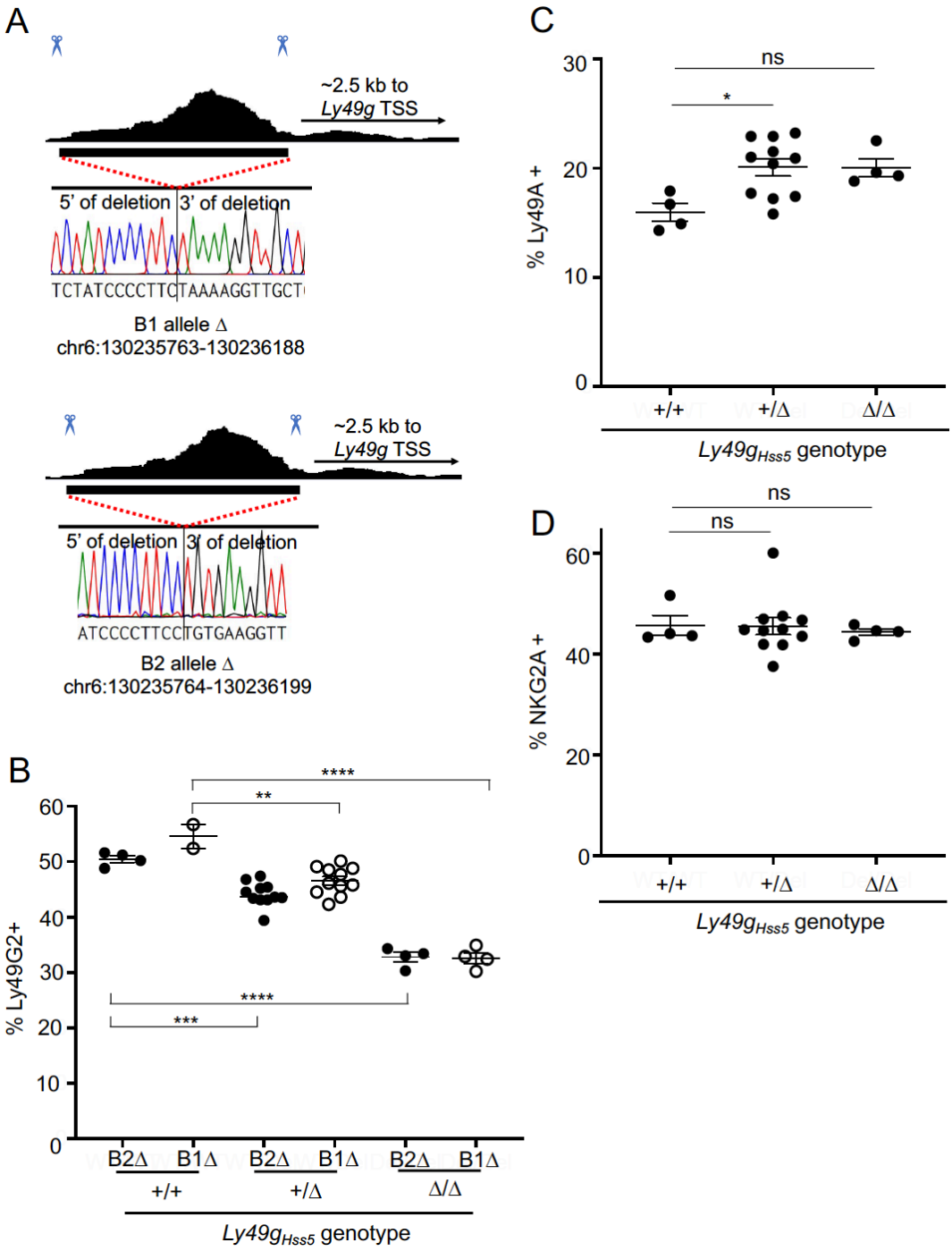


Fig. 3.9. *Ly49g*$_{Hss5D}$ alleles employed in this study. (**A**) Genomic position and sequences of *Ly49g*$_{Hss5Δ}$ alleles (see Fig. 3.3 legend for details). (**B**) Percentages of cells expressing Ly49G2 in mice with the indicated genotypes, comparing one experiment each with the two alleles (*Ly49g*$_{Hss5}$-B1D and *Ly49g*$_{Hss5}$-B2Δ). **P* <0.05; ***P* <0.01; ****P* <0.001; *****P* <0.0001 computed by an ordinary one-way ANOVA. (**C**) Percentages of Ly49A+ NK cells from one representative experiment of 2 performed with the *Ly49g*$_{Hss5}$-B2Δ allele. **P* <0.05 computed by an Ordinary one-way ANOVA (**D**) Percentages of NKG2A+ splenic NK cells from one representative experiment of 2 performed, with the *Ly49g*$_{Hss5}$-B2Δ allele. In all panels ns=not significant

Mitotically stable RME is likely far more common than previously appreciated

Our findings suggesting that RME is a natural consequence of enhancer action raised the possibility that mitotically stable RME is more frequent than previously appreciated. RME genes have previously been identified by bulk RNA analysis of allelic gene expression in heterozygous clonal cell lines, generated either from F_1 hybrid mice or human cells (Gimelbrant et al. 2007; Gendrel et al. 2014; Eckersley-Maslin et al. 2014), or by single cell RNA-seq in primary cell clones (Reinius et al. 2016). We reasoned that those methods will fail to detect RME of genes in which monoallelic expression occurs to only a minor extent, since the number of single cell clones (or single cells) studied is generally limiting, and stochastic effects and technical noise complicate interpretation of allelic expression in single-cell RNA-seq data (Gendrel et al. 2014; Eckersley-Maslin et al. 2014; Kim et al. 2015; Gregg 2017). We therefore tested whether we could identify previously overlooked RME of genes encoding cell surface receptors using flow cytometry to analyze thousands or millions of single cells *in vivo*.

The first receptor we analyzed was NKG2D, based on the fortuitous observation that ~2% of NK cells in *Nkg2d*$^{+/5'E\Delta}$ mice lacked expression of NKG2D altogether, despite the presence of a WT allele, whereas the percentage was close to 0% in WT mice (Fig. 3.10, A and B). Further analysis revealed that *Nkg2d*$^{+/-}$ mice exhibited an even higher frequency of NKG2D- cells (2.5%), ruling out the possibility that these cells arose due to *trans* effects of the *Nkg2d*$^{5'E\Delta}$ allele (Fig. 3.10, A and B). A 2.5% allelic failure rate of each wildtype allele translates to an overall frequency of only 0.063% cells lacking expression of both WT alleles in WT mice, explaining why *Nkg2d* has not been described as variegated or RME previously. We conclude that the WT *Nkg2d* gene is expressed in an RME fashion by NK cells.

We sought to extend this finding to other genes that are thought to be ubiquitously expressed for which we can analyze expression of both alleles. We focused on the CD45 receptor encoded by the *Ptprc* gene, commonly acknowledged as a marker of all immune cells (known as the common leukocyte antigen). Two allelic forms of CD45 (*Ptprc*^{a} encoding CD45.1 and *Ptprc*^{b} encoding CD45.2) are easily discriminated using monoclonal antibodies in congenic mice. We stained *B6-Ptprc*$^{a/a}$, *B6-Ptprc*$^{a/b}$ and *B6-Ptprc*$^{b/b}$ cells from spleen and lymph nodes with antibodies specific for each allele-product. We analyzed T cells (Thy1+) and B cells (CD19+), as these are highly abundant in secondary lympoid tissues and are expected to express both *Ptprc* alleles ubiquitously. Remarkably, *B6-Ptprc*$^{a/b}$ T and B cells included clearly defined, albeit very rare (~0.01%) subpopulations of cells expressing only one allele or the other (Fig. 3.10C and Fig. 3.11). As expected, we could not detect populations of T or B cells lacking CD45.1 or CD45.2 expression in homozygous mice (Fig. 3.11, A and B). Sorted CD45.1 and CD45.2 single positive T cells from *B6-Ptprc*$^{a/b}$ mice retained monoallelic expression over several hundred-fold expansion after stimulation with CD3/CD28 beads *in vitro*, showing RME of *Ptprc* is mitotically stable, further demonstrating *bona fide* stable RME of *Ptprc* (Fig. 3.10D).

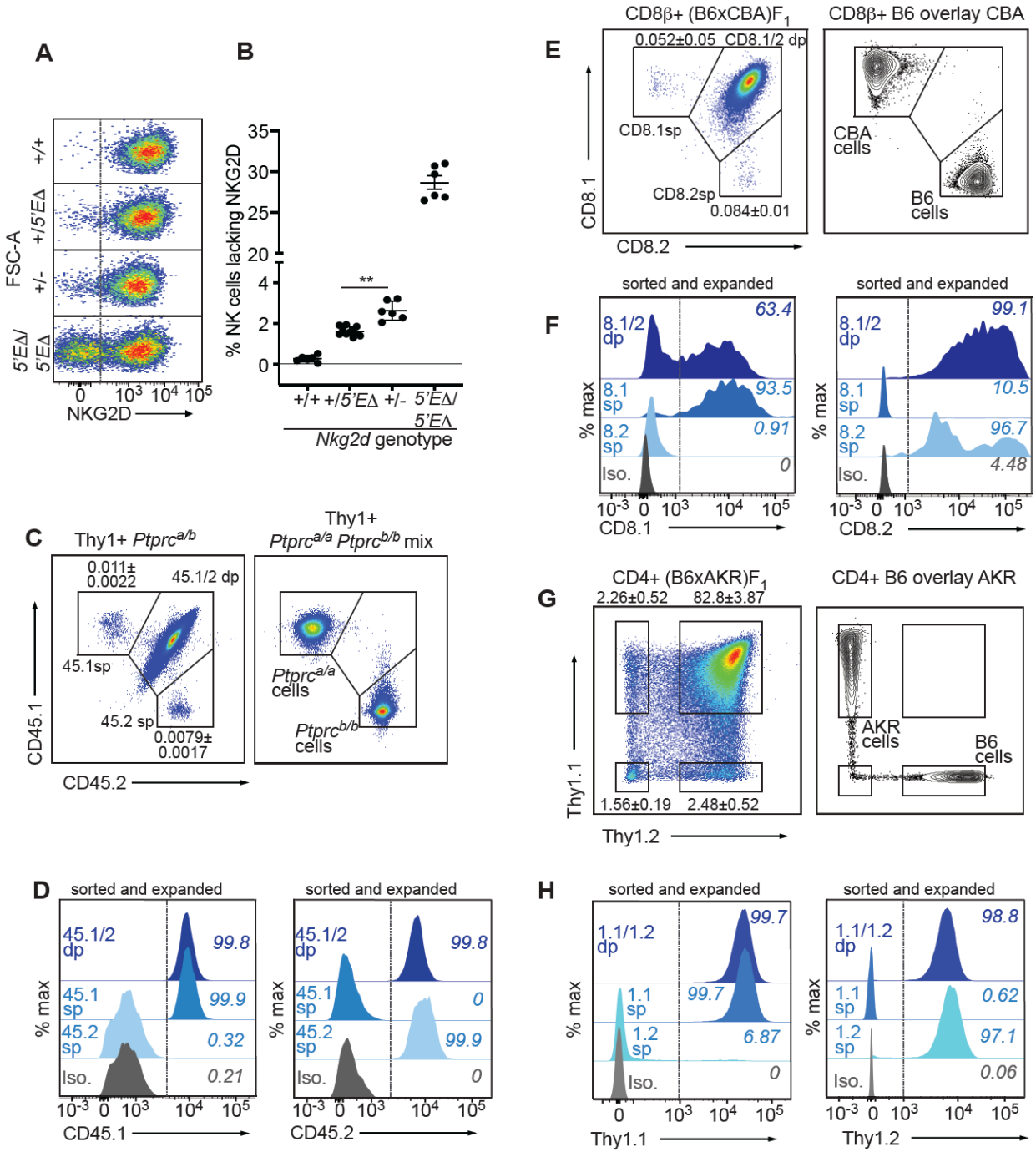


Fig. 3.10. *Nkg2d, Ptprc* and *Cd8a* are all RME genes. (**A**) Representative flow cytometry dot plots of selected *Nkg2d* genotypes. (**B**) Quantification of NKG2D- cells from mice of the genotypes depicted in panel A, combining data from two independent experiments. *Nkg2d*$^{+/-}$ and *Nkg2d*$^{+/5'E\Delta}$ results were compared using a student's *t*-test, *P*=0.002. (**C**) Monoallelic expression of Thy1. Flow cytometry scatter plot of Thy1+ cells from spleens and assorted lymphnodes pooled from 2 *Ptprc*$^{a/b}$ mice (left). The mean percentages and SEM of monoallelic cells from 3 independent experiments are depicted within the representative plot. For comparison, the right panel shows staining of a mixture of cells from a *Ptprc*$^{a/a}$ mouse and a *Ptprc*$^{b/b}$ mouse. (**D**) CD45 single positive and double positive T cell populations were sorted from the *Ptprc*$^{a/b}$ mice using

the gates shown in panel C, stimulated with anti-CD3/CD28 beads and expanded for 1 week *in vitro*, resorted and restimulated for a further week resulting in a ~700-2000 fold total expansion expansion over the course of the first and second stimulations. Histograms show CD45.1 and CD45.2 staining for the depicted sorted populations after expansion. Isotype staining control is in grey. (**E-F**) Monoallelic expression of CD8α in (B6 x CBA)F1 mice. Data are presented as in (C) and (D). E, left panel, depicts a scatter plot of splenic CD8b+ cells from a single F_1 animal, whereas the right panel shows a mixture of B6 and CBA splenocytes, similarly gated. Panel F shows CD8β+ cells from F1 mice sorted and expanded twice using the same protocol as in (D). Cells maintaining expression of CD8β were gated on, and CD8α mAb reactivity by flow cytometry is shown. Note that only the selected allele is maintained and *de novo* expression of the silent allele is not observed. (**G-H**) Data are displayed as in (C-F) but with respect to Thy1 allelic expression in (B6 x AKR)F_1 hybrid mice. All data are representative of 2-3 independent experiments. All error bars and error depicted in scatter plots are calculated as SEM.

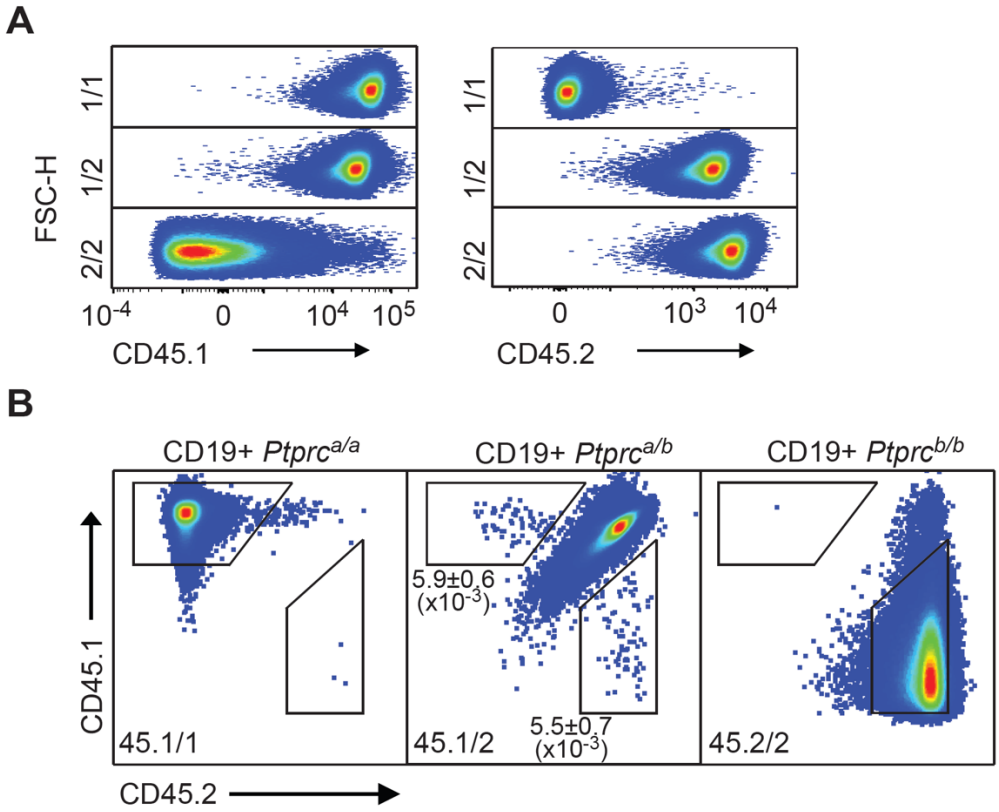


Fig. 3.11. Monoallelic expression of CD45 in B cells. (**A**) Scatterplot staining patterns of gated CD19+ splenic B cells from mice with the indicated genotypes with CD45.1 and CD45.2 mAbs vs FSC-H. The genotypes are as follows: 1/1: *Ptprc*$^{a/a}$, 1/2: *Ptprc*$^{a/b}$, and 2/2: *Ptprc*$^{b/b}$. One mouse from each of the depicted genotypes is displayed; data are representative of 3 independent experiments. (**B**) Staining of B cells on a two-dimensional scatter plot showing CD45.1 and CD45.2. A single representative mouse is displayed for each depicted genotype. Gates outlining cells with monoallelic CD45 expression are shown. Mean percentages and SEMs from 3 experiments are shown in the panels.

We further extended the analysis by analyzing RME of *Cd8a* and *Thy1*, since both are thought to be ubiquitously expressed on T cell populations and mAbs that distinguish allelic variations are available. We assayed expression of two allelic forms of the CD8α co-receptor (CD8.1 and CD8.2) on CD8β+ T cells from CBA mice (CD8.1/CD8.1), B6 mice (CD8.2/CD8.2) and (B6 x CBA)F_1 mice (CD8.1/CD8.2). *Cd8a* displayed a similar RME pattern, with an approximate allelic failure rate of 0.1% (Fig. 3.10E; Fig. 3.12). Importantly, this pattern was mitotically stable in the same *in vitro* stability assay; however, CD8α expression overall is not stable in CD8+ T cells stimulated as in our assay (Fig. 3.10F). Monoallelic loss of CD8α expression has previously been observed under certain activation conditions previously (Harland et al. 2014). However, cells that maintained expression of CD8β (which is dependent on expression of CD8α), maintained expression of the originally selected allele of CD8α (Fig. 3.10F), indicating stability of monoallelic expression when any expression is maintained. Enhancer deletion in the *Cd8a* locus has previously been shown to result in high allelic failure rates and RME (Garefalaki et al. 2002; Ellmeier et al. 2002), but our new results demonstrate RME of the *WT Cd8a* gene with a much lower allelic failure rate. Therefore, like *Nkg2d*, natural *Cd8a* alleles display low allelic failure rates that are exacerbated by enhancer deletion, further demonstrating the pervasiveness of RME and the role enhancers play in minimizing allelic failure.

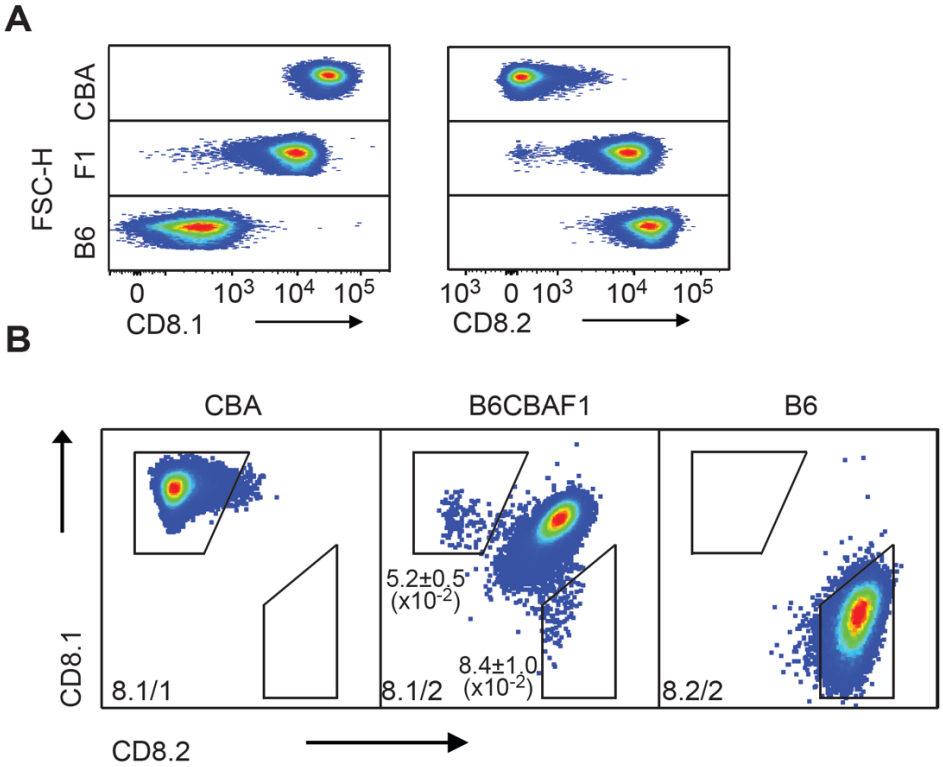


Fig. 3.12. CD8α monoallelic expression by T cells in B6, CBA and (B6 x CBA)F_1 mice. (**A**) Scatterplot staining patterns of gated CD8β+ splenic T cells from mice with the indicated genotypes with CD8.1 or CD8.2 mAbs vs FSC-H. Data are representative of 4 independent experiments. (**B**) Two-dimensional scatter plots showing CD8.1 vs CD8.2 staining of gated CD8b+ cells. A single representative mouse is displayed for each genotype. Gates outlining cells with monoallelic CD8α expression are shown. Mean percentages and SEMs from 4 experiments from (B6 x CBA)F_1 mice are shown in the panels.

Surprisingly, *Thy1*, which is thought to be expressed by all T cells, displayed an appreciable RME pattern with much higher allelic failure rates than *Ptprc* or *Cd8a*. (B6 x AKR)F_1 hybrids are heterozygous at *Thy1*, and allele-specific mAbs distinguish the Thy1.1 (encoded by $Thy1^a$ in AKR) and Thy1.2 (encoded by $Thy1^b$ in B6) alleles. ~5% of CD4+ T cells in (B6 x AKR)F_1 mice display expression of only one allele or the other (Fig. 3.10G). As with *Cd45* and *Cd8a*, *Thy1* monoallelic expression stability displayed an impresive degree of mitotic stability after two rounds of invitro stimulation and expansion (Fig. 3.10H). Importantly, silent *Thy1* alleles were to be induced in up to half of CD4+ T cells after the first stimulation (data not shown). Therefore, *Thy1* displays the features of a stable RME gene that is competent to be induced by stimulation, similarly to the monoallelic induction of the cytokine genes (Bix and Locksley 1998; Riviere, Sunshine, and Littman 1998). Intriguingly CD8+ T cells display robust expression of Thy1.2 (only ~0.1% of cells expressed Thy1.1 but not Thy1.2), but ~13% of cells expressed only Thy1.2 but not Thy1.1 (Fig. 3.13). These data suggest that both *cis* and *trans* differences regulate *Thy1* allelic failure rates, and in particular *Thy1.1* has a large failure rate in CD8+ T cells.

In conclusion, RME was detectable for all four genes we examined that were previously thought to be ubiquitously expressed, and we propose that RME is the result of naturally-occurring allelic failure rates that are directly determined by the enhancer complex (Fig. 3.14). These findings support the notion that RME is characteristic of many genes, consistent with the notion that it is a natural extension of enhancer-promoter interactions, rather than a specialized form of gene expression. Apparently, RME often occurs at such a low rate that it is both beneath ready detection and presumably irrelevant for normal gene function. In other cases, we propose, evolution has exploited this phenomenon to generate variegated gene expression patterns that limit gene expression to functional subsets of cells.

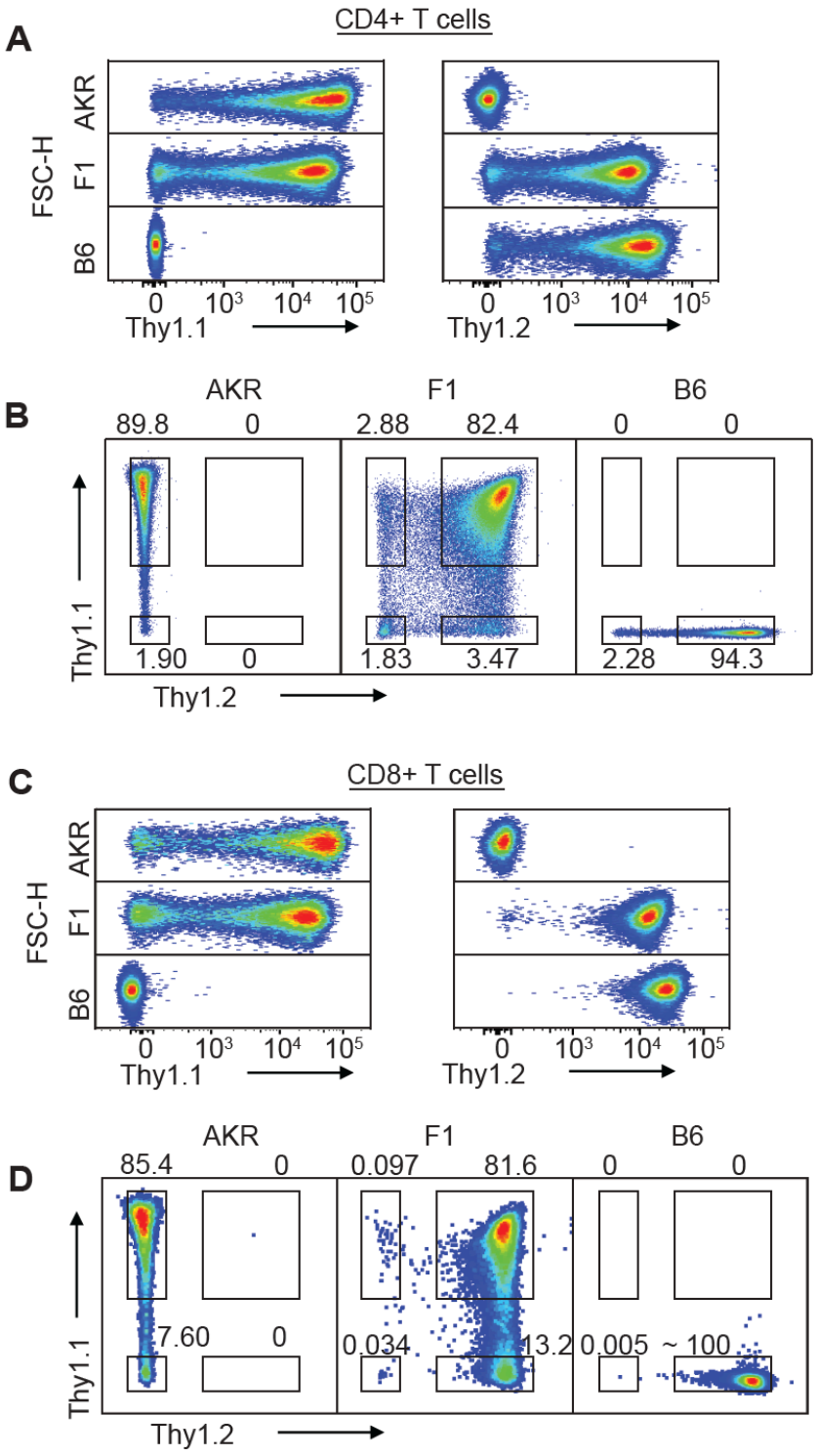


Fig. 3.13. Thy1 monoallelic expression in CD4+ and CD8+ T cells. (**A**) Scatterplot staining patterns of gated CD4+ splenic T cells from B6, AKR and (B6xAKR)F_1 mice. Thy1.1 (left) or Thy1.2 (right) staining is depicted against FSC-H. Data are representative of 3 independent

experiments. (**B**) Scatterplots of the data depicted in (A) but showing the Thy1.1 vs Thy1.2 parameters. (**C**) Scatterplot staining of CD8+ splenic T cells from B6, AKR and (B 6x AKR)F_1 mice as in (A). (**D**) Scatterplots of the data depicted in (C) gated on CD8+ T cells showing Thy1.1 vs Thy1.2. In all cases, a single representative mouse is displayed.

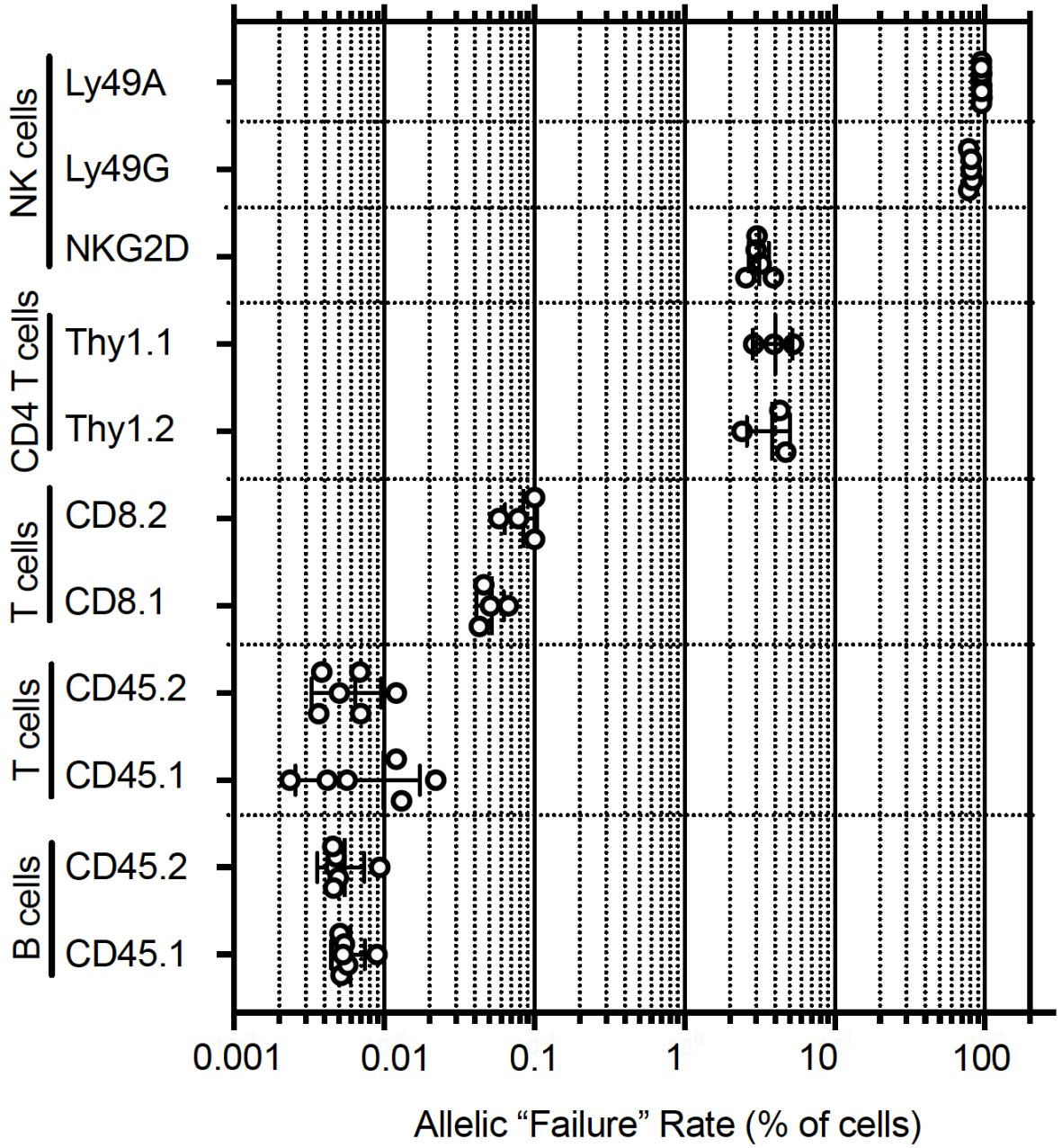


Fig. 3.14. Allelic failure rates of selected alleles from this study. Quantification of failure rates of selected alleles in this study. Failure rate is defined as the percentage of cells that fail to express a particular allele as measured in a genetic background that allows detection of such cells by flow cytometry. *Ly49* allelic failure rates were based on analysis in a (B6 x BALB/c)F_1 hybrid background; *Nkg2d* alleles in mice where the opposing chromosome harbors the *Nkg2d* knockout allele; *Cd8a* alleles in (B6 x CBA)F_1 hybrids; *Cd45* alleles in $Ptprc^{a/b}$ F1 congenic mice on the B6 genetic background. Data are compiled from 4-6 mice per group from multiple experiments. The horizontal axis depicting failure rates as a percentage of cells is on a $\log_{10}$ scale. In each case, error bars represent the SEM.

Discussion

The study of both RME and the enhancer deletion-associated variegation phenomenon have been hampered by the lack of an *in vivo* model such as the one employed in this study which allows definitive detection and sorting of allelic expression states in large numbers of single primary cells (Walters et al. 1995; Gendrel et al. 2016). We found that the required and constitutively accessible enhancers of the RME *Ly49a* and *Nkg2a* genes act stochastically and independently of the other copy, arguing that mitotically stable RME is a product of independent and stochastic events on each allele rather than of a specialized epigenetic mechanism that imposes an RME pattern of expression. That mitotically stable RME is regulated by enhancers *in cis* is highly reminiscent of *Igh* variegation in B cell hybridomas. Deletion of the intronic *Igh* LCR led to variegation, and both the positive and negative states were maintained *in cis* within the same nucleus in fused hybridomas (Ronai, Berru, and Shulman 2002). Similarly, heterozygous deletion of *Cd8a* gene enhancers also resulted in stable RME-like expression in thymocytes *in cis* (Garefalaki et al. 2002; Ellmeier et al. 2002).

The recurring findings of stable RME resulting from enhancer deletion strongly argue that the epigenetic stability of both RME and enhancer deletion-associated variegation is a manifestation of the probabilistic nature of stable gene activation, as opposed to an active gene repression mechanism dedicated to RME. Enhancers appear to raise the probability of stable allelic activation, but if activation is not achieved then the silent state is stably maintained in the absence of the inductive signal received during differentiation.

In strong support of the probabilistic model of enhancer action and stable gene activation, deletion of an enhancer upstream of the otherwise ubiquitously expressed *Nkg2d* gene imparted an RME expression pattern that was indistinguishable from that of the naturally variegated NK receptor genes in terms of stochasticity, stability through mitosis, and the accessibility of the promoter only in cells with active alleles. The similarity of these patterns with RME genes in F_1 neural progenitor clones implies a common regulatory mechanism rooted in probabilistic regulation by constitutively active enhancers.

The mechanism of maintenance of active RME alleles—which is likely shared by active lineage-appropriate genes generally—remains an open question and may involve mitotic persistence of an open and active state of promoters, known as bookmarking (Teves et al. 2016). The previous findings that promoters of both RME genes and ubiquitously expressed genes are accessible during prometaphase is consistent with the possibilty that stable expression of both types of genes is maintained by a bookmarking mechanism (Xu et al. 2017). Consistent with this model of maintenance of the active state is our finding that the only apparent chromatin differences between silent and expressed *Ly49g* alleles is that the active alleles have accessible promoters that are enriched for active H3K4me3/ H3K27ac histone modifications, whereas the inactive alleles have neither (see Chapter 4)

An additional significant finding of this study emphasizes the quantitative impact of enhancer strength on allelic expression frequencies of RME genes. The deletion of $Ly49g_{Hss5,}$ a relatively minor enhancer element based on accessibility resulted in a reduced frequency of expression of a natural RME gene, directly tying the enhancer deletion-associated variegation phenomenon to RME. This result powerully argues that enhancers are not only permissive at RME

genes, but are also instructive regarding the likelihood of expression. We hypothesize that the broad range of frequencies with which different Ly49 genes are naturally expressed (~5%-60%) reflects differences in enhancer strength, at least in part.

Previous studies that generated RME via enhancer deletion *in vivo* at the *Cd8a* locus (Garefalaki et al. 2002; Ellmeier et al. 2002) did not generalize these findings to a broad mechanism of enhancer action and RME, likely because 1) RME was not yet appreciated as a pervasive phenomenon, and 2) there were no well known systems of RME genes, such as the NK receptor genes in NK cells, for comparison in the cells studied. Our concordant results in the *Ly49a, Ly49g, Nkg2a,* and *Nkg2d* loci argue that stable RME is the result of a generalized phenomenon driven by probabilistic action of constitutively active enhancers. Our results link enhancer deletion-associated variegation with naturally-occuring RME and place the previous results in the context of a pervasive biological phenomenon. Notably, deletion of a P element enhancer of the OR genes resulted in a reduced probability of OR gene expression *in cis* (Khan, Vaes, and Mombaerts 2011). That result, along with our finding that $Ly49g_{Hss5}$ bolster the argument that probabilistic enhancer action provides the intial stochastic event across different familes of RME genes in spite of differences in molecular regulation (see Introduction).

The findings as a whole uncover broad and fundamental properties of gene regulation. They provide powerful *in vivo* support for a probabilistic, binary "on/off" model of enhancer action, where enhancers are constitutively activated, but their primary effect on target gene expression is to raise the probability of expression rather than raise the expression level per cell (Fiering, Whitelaw, and Martin 2000; Walters et al. 1995). Recent findings that enhancers are probabilistic regulators of transcription burst frequency rather than burst size are consistent with this model (Larsson et al. 2019; Bartman et al. 2016). How enhancer control of the probability of stable gene expression interfaces with the control of transcription burst frequency should be an area of intense and exciting investigation in the future.

Previous studies have estimated the prevalence of RME genes at ~0.5-3% of genes (Eckersley-Maslin et al. 2014; Gendrel et al. 2014; Reinius et al. 2016), or up to 30% of all tissue specific genes (Nag et al. 2013). Consistent with the generality of the mechanisms underlying RME, our data suggest that RME is possibly even more pervasive than these earlier estimates. We detected low but appreciable allelic failure rates of wildtype *Nkg2d* alleles in NK cells, *Ptprc* alleles in T and B cells, *Cd8a* alleles in cytotoxic T cells and *Thy1* alleles in both cytotoxic and helper T cells. Whereas RME has been associated in some previous studies with genes that are expressed at low levels per cell (Gendrel et al. 2014; Reinius et al. 2016) the RME NK receptor genes and *Ptprc* are relatively highly expressed in the cells studied. That RME is characteristic of both highly and poorly expressed genes and is regulated by enhancer strength is in accord with the aforementioned findings that enhancers regulate the frequency of transcriptional bursts, while promoters determine the number of mRNA molecules produced per burst (Larsson et al. 2019).

We propose that genes lie along a spectrum of allelic failure rates that are largely controlled by enhancer strength, with documented stable RME genes on the highest end of that spectrum. These considerations may help resolve recent controversy concerning the prevalence of RME and concerns over its significance as a specialized mechanism (Vigneau et al. 2018; Reinius and Sandberg 2018), as they suggest that RME is characteristic of many or all genes, but extremely

variable in the extent of allelic failure. We predict that the use of higher-resolution genome-wide approaches that provide high-confidence information about both positive and negative allelic expression states in millions of polyclonal cells will reveal that RME applies to many and perhaps all genes. Biallelic expression of genes, rather than the "default" state of expression of autosomal genes, must be achieved by independent successful gene activation events on each allele.

In addition to various mammalian examples, deletion of an enhancer within a set (i.e., a shadow enhancer) was previously observed to result in a reduced frequency of gap gene expression in the *Drosophila* embryo (Perry, Boettiger, and Levine 2011), indicating the conservation of probabilistic enhancer action across evolution. Enhancer redunduancy (regulation by multiple/shadow enhancers) has been suggested to introduce robustness in gene expression, especially under adverse conditions (Perry, Boettiger, and Levine 2011; Hobert 2010). Our results in the *Ngk2d* and *Ly49g* loci suggest enhancer redundnacy is also an important aspect of robustness in gene expression under optimal conditions.

The pervasive and conserved nature of probabilistic gene expression provides an obvious template for evolution to generate diversity in an otherwise ontogenetically identical population of cells. Stable RME could be readily generated by mutations that weaken relevant strong enhancers in precursor genes or by initially providing the genes with weak enhancers. In the case of the NK receptor genes, weak enhancer activity would be selected for as they confer evolutionary advantage by endowing the cells with discriminatory powers in detecting loss of MHC expression on other cells. We suspect that similar mechanisms are responsible for the evolutionarily convergent variegated phenotype of human *KIR* genes. Even more speculatively, by regulating expression of fate-determining mediators, such a mechanism may underlie cell fate decisions in some instances of cellular development, as suggested by recent results showing that a distal enhancer controls the rate of the probabilistic *Bcl11b* expression and subsequent commitment of thymocytes to a T-cell fate (Ng et al. 2018).

A previously published model of *Ly49* gene variegation proposed that *Hss1* elements act as bidirectional promoters that serve as developmental switches only in immature cells, transcribing either towards the gene to establish stable gene expression in descendant cells, or away, to stably extinguish expression (Saleh et al. 2004). Other studies, consistent with ours, emphasized that *Hss1* elements display properties of enhancers in mature cells, including bidirectional transcript production reminicient of eRNAs (Gays, Taha, and Brooks 2015). The data herein show that *Hss1* is rich in enhancer marks, and that a nearby minor enhancer, $Ly49g_{Hss5}$, plays a clear role in regulating the frequency of Ly49G2+ cells, precluding models where *Hss1* acts alone as a switch element to determine expression frequency. Furthermore, we showed that deletion of an enhancer element, rather that introduction of a variegating switch element, resulted in a variegated state indistinguishable from natural variegation in a ubiquitously expressed gene, *Nkg2d*. Finally, Chapter 4 shows that deletion of $Ly49g_{Hss1}$ in mature Ly49G2+ cells resulted in the loss of Ly49G2 expression, inconsistent with a solely developmental role of *Hss1*).

It has been shown that the dosage of transcription factors such as TCF-1, Runx3 and c-Myb directly affect the probability of the expression of different Ly49 receptors (Held et al. 1999; Ohno et al. 2008; Bezman et al. 2011). Our model is not at odds with these results, as *trans*-acting factor availability, in combination with the underlying DNA sequence, likely function to set the

"strength" of *cis*-acting enhancers, i.e. how powerfully they are able to increase the probability of target gene expression. Futhermore, pioneer transcription factors and other transcription factors might act directly at the promoter within a developmentally critical time window to confer competence for the promoter to be probabilistically activated by enhancers. The molecular details of how *trans* factors and *cis* elements act in concert to determine the probability of stable gene expression remains to be uncovered.

Chapter 4
The chromatin features of the monoallelically expressed NK receptor genes suggest bistability is an intrinsic property of probabilistic enhancer action

Portions of this chapter were adapted and/or reprinted with permission from "The chromatin features of the monoallelically expressed NK receptor genes suggest bistability is an intrinsic property of probabilistic enhancer action." Djem U. Kissiov[1], Alec Ethell[1], Natalie K. Wolf[1], Chenyu Zhang[1], Susanna M. Dang[1], Bryan Chim[2], Stefan A. Muljo[2], David H. Raulet[1*]

Abstract

Mitotically stable random monoallelic gene expression (RME) has emerged as a notable exception to the notion that autosomal genes are generally biallelically expressed. In recent years various chromatin modifications have been described at active and silent RME alleles, but no unifying or causative chromatin features have thus far been identified. Here, we analyze the chromatin features of the RME natural killer (NK) cell receptor gene alleles in sorted primary mouse NK cells. Strikingly, critical proximal enhancers are constitutively accessible and activated irrespective of gene expression status, while promoters are active only at expressed alleles. We found no evidence of repressive histone modifications associated with polycomb and heterochromatic repression at silent NK receptor alleles. Instead, silent alleles fell within an inactive chromatin state, similar to genes expressed in non-NK lineages, suggesting chromatin-based repression is not required for the stability of silent RME alleles. Sustained enhancer activity was required to maintain expression of active *Ly49g* and *Nkg2a* alleles in primary, mature NK cells. Together, these data suggest a model where stochastically activated RME alleles require maintenance of active expression states. Despite constitutive proximal enhancer activity, silent alleles remain inactive in a manner similar to that of many lineage non-specific genes, which also lack repressive modifications.

Main Text

Mitotically stable random monoallelic expression (RME) of autosomal genes is an important exception to the standard model of gene expression where both gene copies are expressed. It reportedly occurs in ~0.5-10% of genes (Gimelbrant et al. 2007; Reinius et al. 2016; Gendrel et al. 2016). RME has been likened to monoallelic expression of X-linked genes but is notably different in that most RME genes occur in scattered singletons throughout autosomes and randomness is on the level of the individual gene locus rather than chromosome-wide (Gendrel et al. 2016). RME differs from allelic exclusion of antigen receptor or odorant receptor genes in that coexpression of both alleles, and failure to express either allele, occurs with a higher frequency in RME, suggesting independence of allelic expression. The chromatin features that mediate the mitotic stability of RME are poorly understood (Eckersley-Maslin and Spector 2014). While a bivalent chromatin state (minimally defined by co-localization of the active elongation-associated H3K36me3 and the inactivating H3K27me3 modifications) has been associated with monoallelically expressed genes (Nag et al. 2013), there are no consistent or signature repressive histone modifications at RME genes (Gendrel et al. 2014; Eckersley-Maslin and Spector 2014; Eckersley-Maslin et al. 2014; Xu et al. 2017), and importantly the overwhelming majority of RME genes are insensitive to pharmacological perturbation of chromatin factors (Gendrel et al. 2014; Eckersley-Maslin et al. 2014; Eckersley-Maslin and Spector 2014). DNA methylation is not consistently found at the promoters of silent RME alleles, and inhibition of DNA methyltransferase activity has not resulted in wholesale loss of

monoallelic expression across many genes but does appear to play a role in some cases (Eckersley-Maslin and Spector 2014; Eckersley-Maslin et al. 2014; Gendrel et al. 2014).

The ability to analyze chromatin features of RME alleles *in vivo* has been a challenge due to the difficulty of isolating cell populations that are pure with respect to allelic expression status at a given RME locus (Gendrel et al. 2016; Xu et al. 2017). Therefore, most previous work has been carried out in clonal cell lines derived from F_1 hybrids, where allelic expression status is known and mitotically stable within a clone (Gendrel et al. 2014; Xu et al. 2017; Eckersley-Maslin et al. 2014). Here, we circumvent this limitation to analyze the chromatin features of RME alleles in FACS purified homogenous primary cell populations by taking advantage of allele-specific antibodies that we previously generated against the RME murine natural killer (NK) cell receptors, allowing us to identify and purify primary cells of known RME expression status with high precision.

The $Ly49_{Hss1}$ and $Nkg2_{5'E}$ elements are enhancers

RME of the NK receptor genes creates the variegated expression pattern characteristic of the genes, which underlies the requisite diversity within the NK cell population for discriminating target cells that downregulate specific MHC I alleles in transformation or infection (Raulet, Vance, and McMahon 2001). In mice, the receptors are members of the C-type lectin-related family of proteins and include the Ly49 subfamily, and the NKG2A receptor, both of which are RME genes. These genes are tightly clustered in the NK cell gene complex (NKC) on mouse chromosome 6 (Fig. 4.1A).

The *Ly49* genes are regulated proximally and are not thought to compete *in cis* for interaction with a regulatory element, in contrast to the monoallelically expressed protocadherin gene cluster (Tanamachi et al. 2004; Esumi et al. 2005). Each *Ly49* gene harbors a nucleosome-depleted region ~5kb upstream of the TSS, denoted *Hss1* (Tanamachi et al. 2004) (Fig. 4.1, A and B). We noticed that the *Nkg2a* gene, as well as a linked ubiquitously expressed NK receptor gene, *Nkg2d*, harbor similar proximal upstream elements ($Nkg2a_{5'E}$ and $Nkg2d_{5'E}$) (Fig. 4.1, A and B). These upstream elements for *Ly49*, *Nkg2a* and *Nkg2d* genes are all similarly bound by Runx3 and T-bet in NK cells (Fig. 4.2), suggesting possible common regulation of these elements across the NKC genes. We show in Chapter 3 that $Ly49a_{Hss1}$ and $Nkg2a_{5'E}$ are required for the expression of their respective genes, while $Nkg2d_{5'E}$ is required for ubiquitous expression of NKG2D by all NK cells.

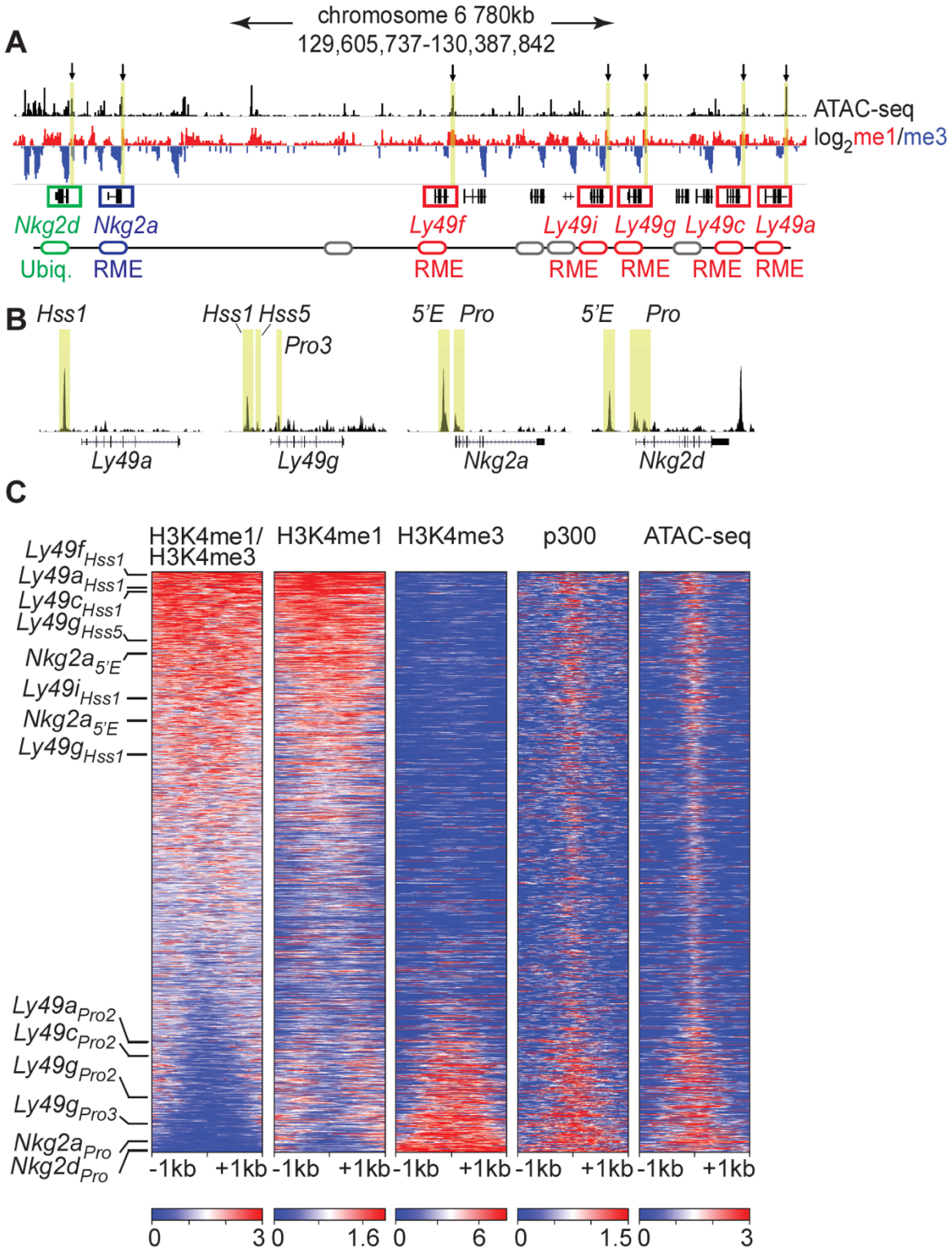


Fig. 4.1. The *Ly49*$_{Hss1}$ and *Nkg2*$_{5'E}$ elements display classical chromatin features of enhancers (**A**) Representation of the 800kb stretch of the NKC encoding the *Ly49* and *Nkg2*

NK receptor genes. ATAC-seq and H3K4me1:me3 log_2 ratio ChIP-seq data in primary NK cells are displayed; red denotes positive me1:me3 ratios, which are characteristic of enhancer chromatin, while blue indicates negative values, which are characteristic of promoter chromatin. Approximate locations of RME and ubiquitously expressed NK receptor genes are indicated in the schematic at the bottom, and *Ly49* loci not discussed in this study are depicted as grey ovals. Vertical yellow bars and arrows denote the positions of the *Hss1* and *5'E* enhancers at the indicated genes. NK cell ChIP-seq and ATAC-seq data are mined from ref (Lara-Astiaso et al. 2014) (**B**) IGV screenshots of ATAC-seq data from NK cells expressing both Ly49G2 alleles (described in Figure 4.4) at the *Ly49a*, *Ly49g*, *Nkg2a* and *Nkg2d* NK receptor gene loci. Positions of relevant proximal regulatory elements are indicated as a reference. Vertical yellow bars denote the positions of the indicated regulatory elements. (**C**) Heatmap depiction of ChIP-seq and ATAC-seq data in primary mouse splenic NK cells. All heatmaps depict 51,650 ATAC-seq peaks called using MACS2 and ranked according to H3K4me1:me3 ratio of average ChIP-seq signal calculated over a 2kb window centered on the ATAC-seq peak midpoint. H3K4me1:me3 ratio, H3K4me1, H3K4me3, p300 and ATAC-seq at these genomic sites are displayed. Data range in signal per million reads (SPMR) for each heatmap is depicted at the bottom. The locations of selected NK receptor gene *Hss1*, *5'E* and promoter elements within the me1:me3 ranking are shown. H3K4 methylation data are mined from ref (Lara-Astiaso et al. 2014) while p300 is mined from ref (Sciume et al. 2020).

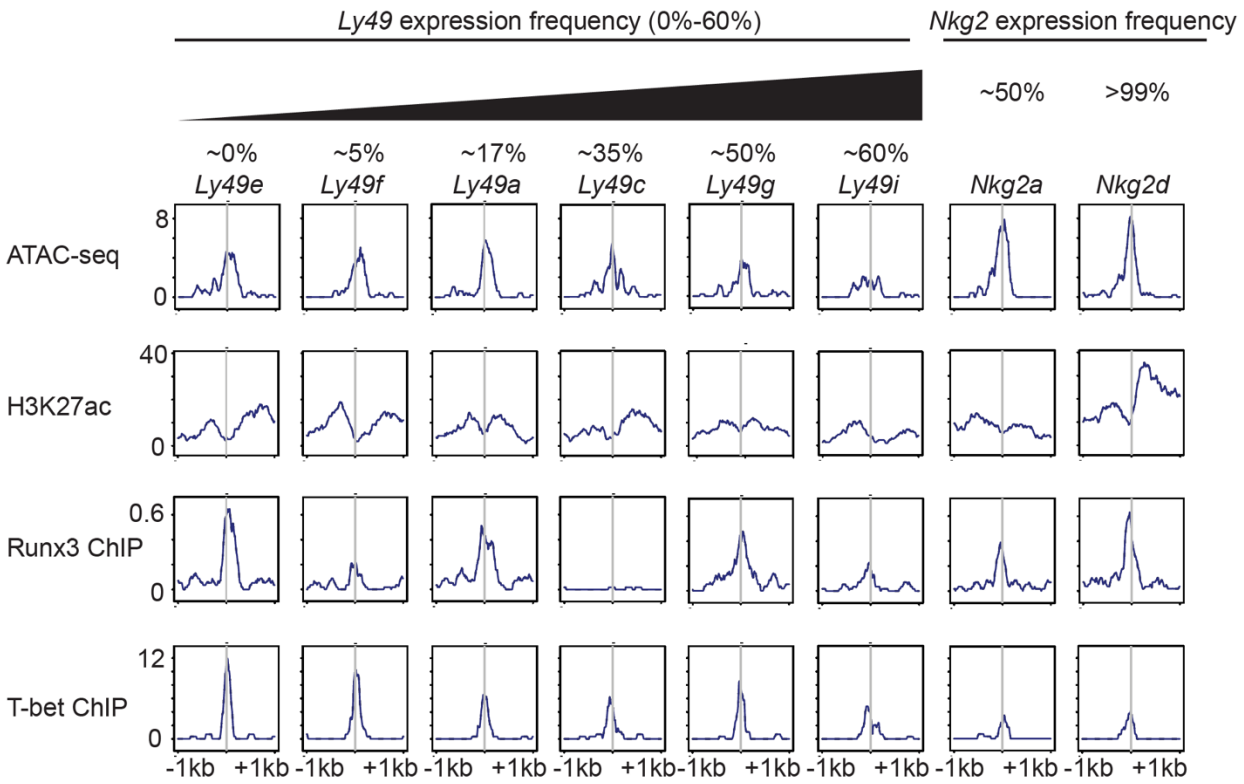


Fig. 4.2. Chromatin features and TF binding profile of the *Ly49*$_{Hss1}$ and *Nkg2*$_{5'E}$ enhancers. *Hss1* elements of selected *Ly49* genes are depicted left to right according to expression frequency. *Nkg2*$_{5'E}$ elements are depicted on the right. ATAC-seq and ChIP-seq data profiles are shown over a 2kb window centered at the midpoint of the called ATAC-seq peak. ATAC-seq and H3K27ac are mined from ref (Lara-Astiaso et al. 2014), while Runx3 data are from ref (Levanon et al. 2014) and T-bet data are from ref (Shih et al. 2016).

The *Ly49*$_{Hss1}$ elements were hypothesized in one set of studies to serve as upstream bidirectional promoters active only in immature, developing NK cells, which act to randomly transcribe either towards or away from the associated coding sequence, resulting in stable gene expression or silencing, respectively, in the descendant mature NK cells (Saleh et al. 2004; McCullen et al. 2016). Recent evidence based on enhancer reporters in cell lines instead suggested that the *Ly49*$_{Hss1}$ elements act as transcriptional enhancers (Gays, Taha, and Brooks 2015).

In order to address recent controversy over the molecular nature of the *Ly49*$_{Hss1}$ elements *in vivo,* we analyzed published ChIP-seq data generated in primary splenic NK cells, using the H3K4me1:me3 ratio as an indicator of regulatory element identity (Calo and Wysocka 2013). The *Hss1* and *5'E* elements are all enriched in H3K4me1 relative to H3K4me3 (Fig. 1A), suggesting enhancer identity. We further characterized *Hss1* and *5'E* elements of the NK receptor genes, along with their promoters, in a genome-wide comparison. *Hss1* and *5'E* elements all ranked in the top 32% of peaks with respect to the H3K4me1:me3 ratio, while NK receptor gene promoters ranked in the bottom 21% (Fig. 4.1C).

We verified this analysis by independently defining enhancers and promoters in mature NK cells using p300 binding and previously annotated murine promoters, respectively. NK cell promoters were defined as previously annotated EDPNew promoters enriched in H3K27ac in NK cells, while enhancers were defined as ATAC-seq peaks bound by the p300 histone acetyltransferase that do not overlap with the promoter list. This independent classification of NK cell promoters and enhancers validated our approach of stratifying accessible sites by H3K4me1:me3 ratio, as enhancers were grouped at the highest me1:me3 values, while promoters were clustered at the lowest (Fig. 4.3). Importantly, all *Hss1* and *5'E* elements were classified as enhancers based on the p300-bound enhancer dataset (Fig. 4.3). These findings in mature primary NK cells corroborate the conclusions that *Ly49*$_{Hss1}$ and *Nkg2*$_{5'E}$ represent enhancer elements.

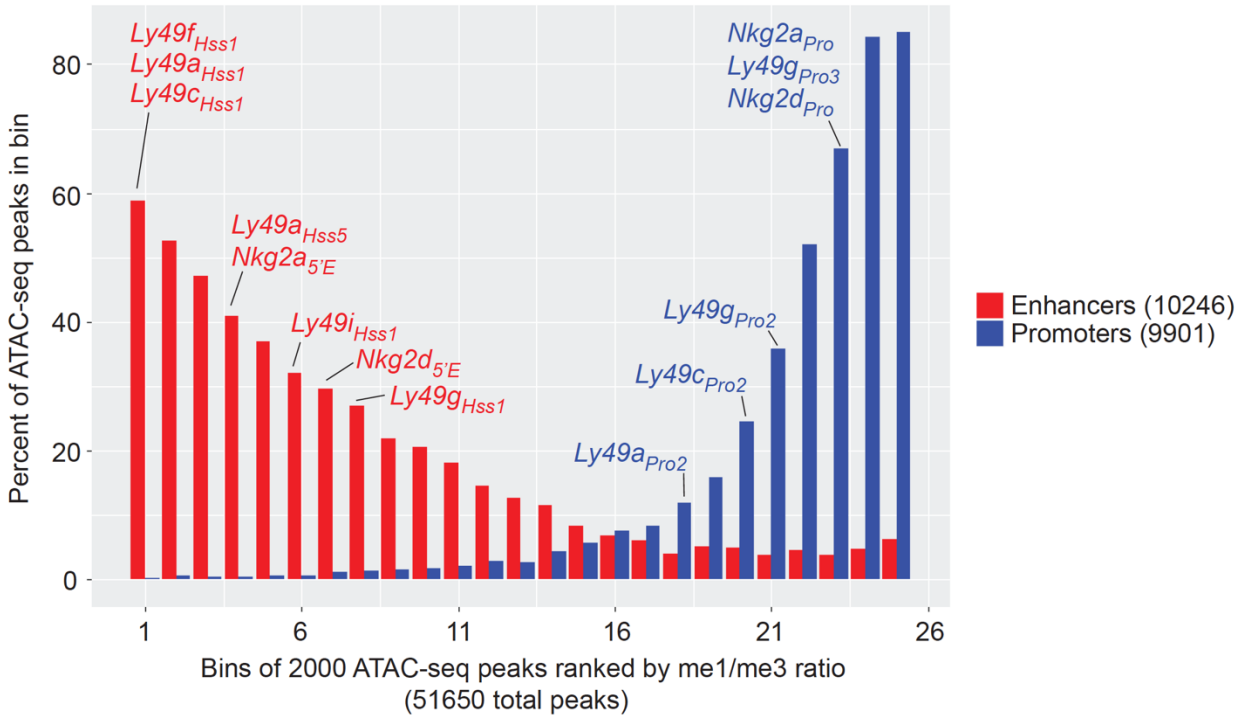


Fig. 4.3. Systematic definition of NK promoters and enhancers based on stratified H3K4me1:me3 ratios. Datasets of H3K4me1 and H3K4me3 modifications in NK cells expressing both alleles of Ly49G2 were used to examine these modifications in 51,560 MACS2-called ATAC-seq peaks (also identified in cells expressing both Ly49G2 alleles, Fig. 4.4). The data were first filtered for peaks found in the top 95% of both me1 and me3 signal in NK cells, and then ranked by me1:me3 ratio over a 2kb window as in Fig. 4.1C. The filtered ATAC-seq peaks were then binned in sets of 2,066 peaks according to me1:me3 ratio, with the highest me1:me3 ratio as bin 1. Separately, a total of 9,901 NK cell promoters were defined by mouse mmEPDnew (v003) as promoters that overlap with broad H3K27ac peaks called from mined from ref (Lara-Astiaso et al. 2014), and a total of 10,246 enhancers were defined as ATAC-seq peaks that are enriched in p300 ChIP-seq signal, mined from ref (Sciume et al. 2020). Within each bin, the percentage of peaks that overlap with enhancers (red) or promoters (blue) defined in this manner are depicted. Bins that contain ATAC-seq peaks corresponding to key selected NK receptor gene promoters or enhancers are indicated.

$Ly49g_{Hss1}$ and $Nkg2a_{5'E}$ are constitutively accessible enhancers

If the activation of the *Hss1* and *5'E* enhancers is the stochastic event that controls RME, we would expect that the accessibility and TF occupancy of these elements would correlate with gene expression frequency in bulk NK cell analysis. Instead, we noticed that accessibility, TF occupancy, and H3K27ac do not correlate with receptor expression frequency in bulk NK cell data (Fig. 4.2). This raised the intriguing possibility that these enhancers are equally active and occupied upstream of both silent and active alleles. For other RME genes studied in clonal cell lines from F1 hybrid mice, two groups recently reported the surprising finding that enhancers are accessible on both expressed and silent alleles, while promoters are accessible only at expressed alleles (Xu et al. 2017; Levin-Klein et al. 2017). It has not previously been possible, however, to verify this finding in *ex vivo* populations comprising mixtures of cells expressing different alleles of RME genes.

To address the accessibility of expressed and silent NK receptor genes, we took advantage of a monoclonal antibody that we previously generated that specifically recognizes the $NKG2A^{B6}$ allele while ignoring $NKG2A^{BALB/c}$ (16a11), and another antibody that binds to both alleles (Vance et al. 2002). Using these antibodies, we sorted NK cells from (B6 x BALB/c)F_1 hybrid mice expressing all four configurations of alleles: both, only B6, only BALB, or neither, and performed ATAC-seq (Fig. 4.4, A and B). SNP-split (Krueger and Andrews 2016) analysis of reads demonstrated that $Nkg2a_{5'E}$ was accessible on both active and inactive alleles in all four populations, whereas the *Nkg2a* promoter was accessible only at active alleles (Fig. 4.4B).

Similarly, we used an antibody that recognizes $Ly49G2^{B6}$ but not $Ly49G2^{BALB}$, combined with a pan-Ly49G2 antibody, to sort and analyze cells expressing either, both or neither Ly49G2 allele (Fig. 4.4, C-D). The $Ly49g_{Hss1}$ enhancer was accessible on both active and inactive alleles in all four populations, whereas the *Pro3* promoter region was accessible only on the active allele (Fig. 4.4D). Note that $Ly49g_{Hss5}$ (located between *Hss1* and *Pro3* in Fig. 4.4D) is also constitutively accessible, supporting its identity as a constitutively active enhancer that contributes to the probability of target *Ly49g* gene expression (see Chapter 3).

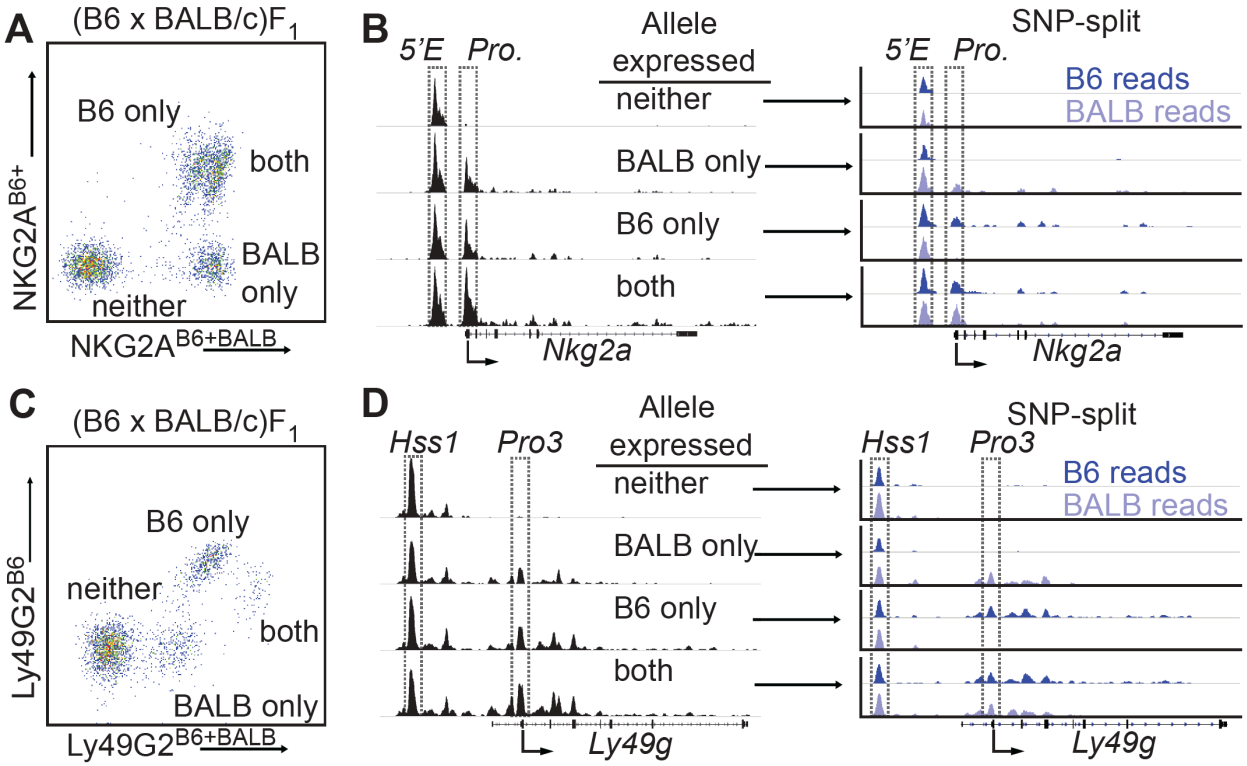


Fig. 4.4. *Nkg2a*$_{5'E}$ and *Ly49g*$_{Hss1}$ are constitutively accessible, while promoters are accessible only at expressed alleles. (**A**) FACS plot depicting splenic NK cells from a (B6 x BALB/c)F_1 hybrid mouse stained with antibodies specific for NKG2A^{B6} (16a11) and pan-NKG2A (20d5), allowing separation of NK cells expressing both, either, or neither NKG2A allele. (**B**) (left) SPMR normalized ATAC-seq data generated from the 4 cell populations depicted in (A) aligned to the mm10 reference genome. (right) Reads from the same 4 ATAC-seq datasets were instead aligned to an "N-masked" genome in which SNPs between the B6 (mm10) and BALB/cJ genomes were replaced by an "N" allowing any base as a match in the alignment process. Allele informative reads were then binned according to chromosome of origin, and displayed as signal mapping to the B6 or BALB/c chromosome. (**C and D**) Data are as in (A and B), but with respect to the Ly49G2 receptor. To separate the populations, F1 cells were stained with antibodies specific for Ly49G2^{B6} (3/25) and pan-Ly49G2 (4D11).

Silent alleles fall within an inactive rather than repressed chromatin state

These data demonstrate that enhancers within both the *Ly49* and *Nkg2* NK receptor gene families behave similarly to those of other RME genes analyzed in F_1 hybrid clones, exhibiting an accessible configuration whether or not the gene is expressed. Importantly, this further validates the NK receptor genes as a model for RME.

In Chapter 3, we show that the NK receptor gene enhancers directly regulate the probability of allelic expression, in line with a previously described model of enhancer action wherein enhancers regulate the binary "on or off" probability (rather than the analog per-cell amount) of gene expression (Fiering, Whitelaw, and Martin 2000; Walters et al. 1995; Blackwood and Kadonaga 1998). The decoupling of enhancer and promoter accessibility seen at NK receptor genes and other RME loci, while initially surprising, is highly consistent with a binary model of enhancer action, where enhancer activation is ubiquitous in a given cell type and in turn regulates the probability of stochastic promoter activation.

We next investigated repressive chromatin at NK receptor genes. For this analysis we focused on H3K27me, H3K9me3 and H2AUb1, as combinations of these marks have previously been found at inactive alleles of the monoallelically expressed odorant receptor genes, the protocadherin genes and subsets of singleton RME genes in F_1 clones (Gendrel et al. 2014; Eckersley-Maslin et al. 2014; Xu et al. 2017; Eckersley-Maslin and Spector 2014; Jiang et al. 2017; Magklara et al. 2011). We performed CUT&RUN with antibodies against these modifications in primary IL-2 expanded NK cells, sorted to be Ly49G2-negative in order to enrich cells with silent *Ly49g* alleles. The repressive histone modifications were no more enriched on silent *Ly49g* alleles than they were on expressed lineage appropriate genes in NK cells, such as *Ncr1* and *Gzmb* (Fig. 4.5A). In contrast, other genes such as *Pdcd1* (encodes PD-1) and *Spi1* (encodes the master macrophage transcription factor PU.1) displayed all 3 marks. Importantly, many other genes associated with non-NK cell lineages (e.g., *Cd19* and *Mstn*, expressed in B cells and myocytes, respectively) were not enriched for these repressive modifications, similar to silent *Ly49g* alleles (Fig. 4.5B and fig. S3). Therefore, with respect to repressive chromatin marks, silent *Ly49g* alleles resembled several lineage non-specific genes, rather than repressed genes.

We further addressed whether active and silent *Ly49g* alleles might differ with respect to active histone modifications, as opposed to repressive ones, at their promoters and gene bodies. We sorted IL-2-expanded NK cells expressing neither (N) or both (B) Ly49G2 alleles from F_1 mice and performed CUT&RUN against the active regulatory element-associated H3K4me1/2/3 and H3K27ac modifications. The *Ly49g* promoter and gene body displayed striking enrichment of H3K4me2/3 and H3K27ac in "B" cells that expressed both *Ly49g* alleles, and as predicted lacked these modifications in "N" cells where both alleles were silent (Fig. 3B). Notably, the $Ly49g_{Hss1}$ enhancer displayed equal enrichment of H3K27ac in cells expressing both or neither allele (Fig. 3B), suggesting constitutive enhancer activation as this mark delineates active from poised enhancers (Calo and Wysocka 2013).

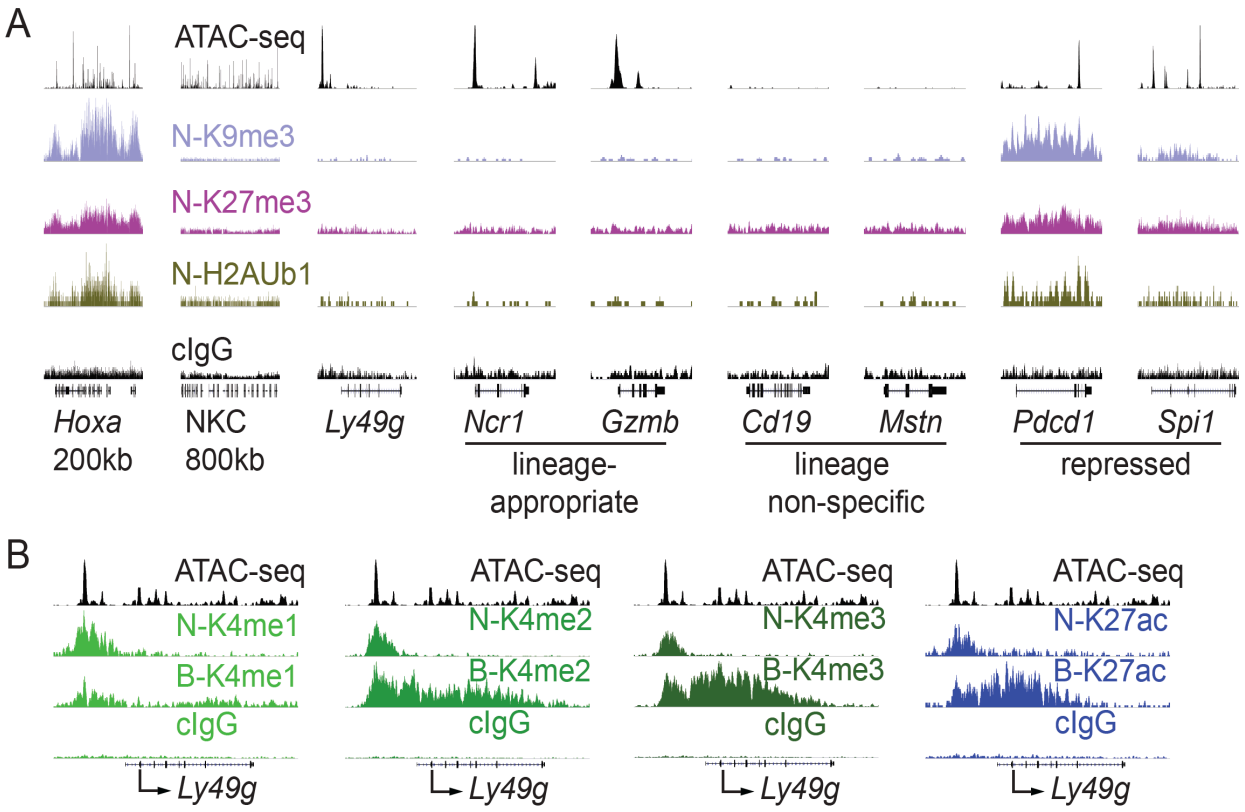


Fig. 4.5. Silent NK receptor gene alleles resemble inactive genes expressed in non-NK lineages, rather than repressed genes. (**A**) Repressive histone modification CUT&RUN data generated in primary IL-2 expanded NK cells sorted to express neither allele of Ly49G2 ("N" cells). IGV screenshots depicting the indicated histone modification or control analyses with protein A-binding mouse IgG2a (cIgG). (**B**) CUT&RUN data depicting each of 4 indicated histone modifications of the *Ly49* gene in IL-2 expanded NK cells sorted to express neither "N" or both "B" alleles of *Ly49g*. The ATAC-seq pattern is shown for reference above each analysis. Arrows depict the locations of dominant transcription start sites. CUT&RUN data are representative of a single experiment with sorted paired samples.

As silent *Ly49g* alleles appeared more similar to lineage non-specific genes in our analysis of repressive chromatin marks, we extended our analysis of chromatin states using ChromHMM, which integrates multiple datasets to classify the genome into subdomains based on their chromatin signatures (Ernst and Kellis 2012). We constructed our states model using H2AUb1 and H3K9me3 (both repressive), and H3K4me3 and H3K27ac (both active). Using data from cells expressing neither or both Ly49G2 alleles, we constructed a minimal 3 state model corresponding to transcriptionally active chromatin, inactive chromatin and repressed chromatin (Fig. 4.6, A and B). Lineage-appropriate genes expressed in NK cells (e.g., *Ncr1, Nk1.1, Ifng*) displayed high levels of H3K27ac and H3K4me3 at their promoters in both cell populations, as expected, and largely fell into the "active" chromatin state 1 (Fig. 4.6C). Notably, NK lineage non-specific genes commonly regarded as markers of other hematopoietic lineages (e.g., *Cd3e, Cd19, Ly6g, Siglech*) displayed neither active nor repressive marks, falling into an "inactive" chromatin state 2 (Fig. 4.6D). Finally, other genes—often encoding transcription factors that regulate non-NK cell fates, such as *Bcl11b*, *Batf3* and *Pax5*—were highly enriched in repressive H2AUb1 and H3K9me3 at their promoters or across the entire gene, falling into the "repressed" state 3 (Fig. 4.6E). This state analysis revealed an overarching logic wherein immune effector molecules associated with non-NK lineages are not actively repressed but are inactive and stably silent, whereas other genes that determine non-NK cell fates are repressed.

In cells expressing both copies of *Ly49g*, the entire locus spanning the enhancer and gene body falls within the active state 1 (Fig. 4.6F). In cells expressing neither copy, the enhancer still falls in the active state 1, and the promoter and gene body fall within the inactive state 2 rather than the repressed state 3. Indeed, it was striking that the NKC as a whole was notably lacking in repressive state 3 chromatin (Fig. 4.5A; Fig 4.6F).

The lack of a repressive chromatin state at silent *Ly49g* alleles and the other NK receptor genes raises the possibility that repressive chromatin may not be required for stable RME, and this may explain why no generalized repressive chromatin signatures have yet emerged at silent RME alleles. Rather, stability appears to be a property of the types of genes studied here and elsewhere, including some genes subjected to enhancer deletions (see Chapter 3) (Garefalaki et al. 2002; Ronai, Berru, and Shulman 1999).

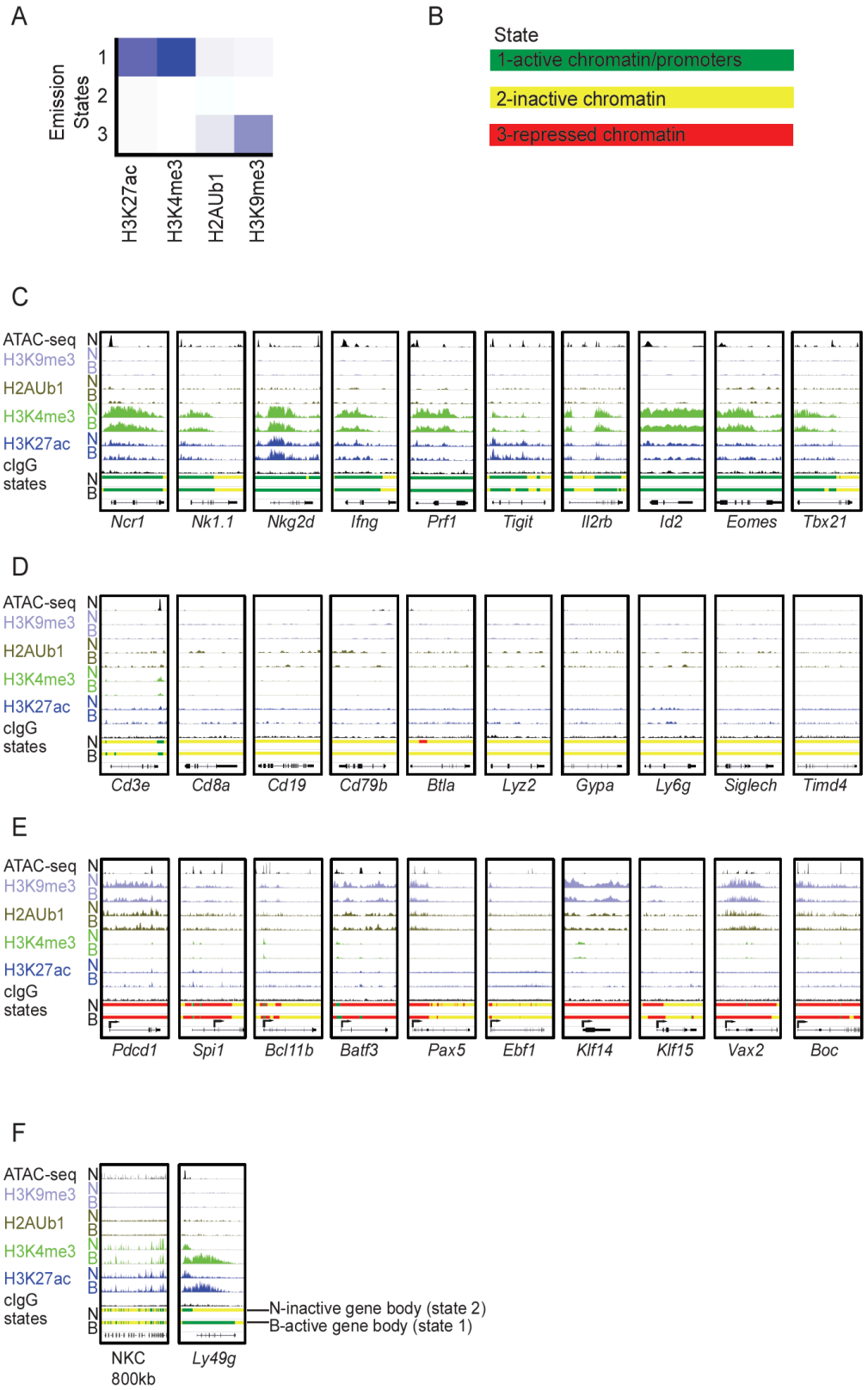


Fig. 4.6. Chromatin state analysis of NK cells expressing neither (N) allele or both (B) alleles of Ly49G2. (**A**) Emission chromatin states determined by ChromHMM (Ernst and Kellis 2012) in a 3 state model, based on active (H3K27ac and H3K4me3) and repressive (H2AUb1

and H3K9me3) chromatin modifications in the two NK cell populations, from CUT&RUN analyses. (**B**) State 1 (green), defined by active modifications, is denoted "active chromatin". State 2 (yellow) lacks both active and repressive marks and is denoted "inactive chromatin". State 3 (red) is defined by repressive modifications and is denoted "repressed chromatin". (**C**) IGV screenshots depicting the modifications and, at the bottom of each panel, the color-coded chromatin states of selected genes characteristically expressed by NK cells, including lineage-specific receptors, effector molecules, and transcription factors. For each modification and state, results with cells expressing neither ("N") Ly49G2 allele, or both ("B") are shown. (**D**) Data as in (C), except depicting selected genes encoding cell surface receptors emblematic of non-NK cell hematopoietic lineages. (**E**) Data as in (C-D), except depicting select genes expressed in non-NK cells lineages that exhibit state 3 or "repressed" chromatin either across the entire gene locus or proximal to the promoter. (**F**) Data as in (C-F), depicting the entire 800 kb segment of the NKC containing the *Ly49* and *Nkg2* loci (left) and zoomed in on the *Ly49g* locus (right).

The difference in chromatin states suggested the possibility that in the case of the genes studied, the active state of gene expression may require maintenance through cell division, as opposed to a requirement for maintaining the inactive state of silent genes. Maintenance of the active state could reflect various mechanisms, including positional memory of histones through DNA replication in S-phase (Schlissel and Rine 2019) so as to maintain nucleosome-depleted regions at promoters, and bookmarking mechanisms that maintain promoter accessibility through M-phase, as has been seen at active promoters of RME and ubiquitously expressed genes alike (Xu et al. 2017).

$Nkg2a_{5'E}$ and $Ly49g_{Hss1}$ are required for maintenance of active alleles

To address whether enhancer action is required to maintain gene expression, we developed a CRISPR/Cas9-based assay to delete the constitutively active enhancers of RME genes in primary *ex vivo* cultured NK cells, based on a modified protocol recently developed to edit primary human T-cells (Roth et al. 2018) (Fig. 4.7A). We sorted receptor positive IL-2 expanded NK cells, excised the enhancer by nucleofection with Cas9 complexed with flanking sgRNAs (Fig. 4.7, B and D; Fig. 4.8, A and B), and assayed receptor expression 6 days later. To increase the probability of successfully targeting a homogeneously expressed allele, we used NK cells from F_1 hybrid mice and sorted $NKG2A^{B6}$+ or $Ly49G2^{B6}$+ cells, allowing us to ignore the BALB alleles. Editing efficiency of NK cells was less efficient than that of T cells, approximately 30% or less, as indicating by successful disruption of the *Ptprc* locus encoding CD45 (Fig. 4.8C). Interestingly, targeting $Nkg2a_{5'E}$ increased the percentage of $NKG2A^{B6}$-negative cells to ~20-40% compared to <10% in non-nucleofected (no zap) or non-targeting sgRNA (*Cd45*) controls (Fig. 4.7C), in line with our theoretical maximum editing efficiency. Of note, $NKG2A^{B6}$ displayed largely stable expression over 6 days post-sort in control samples, with ~10% loss of expression. Similarly, targeting $Ly49g_{Hss1}$ resulted in marked loss of $Ly49G2^{B6}$ expression, with minimal (<5%) loss of expression in non-targeting and non-nucleofected controls (Fig. 4.7E; Fig. 4.8D). These data show that the NK receptor gene enhancers play critical roles in the maintenance of active alleles in mature NK cells, and strengthen the argument against the proposal that the 5' flanking elements in Ly49 genes are bidirectional promoters essential only in immature NK cells to determine the on/off state of the genes (Saleh et al. 2002; Saleh et al. 2004).

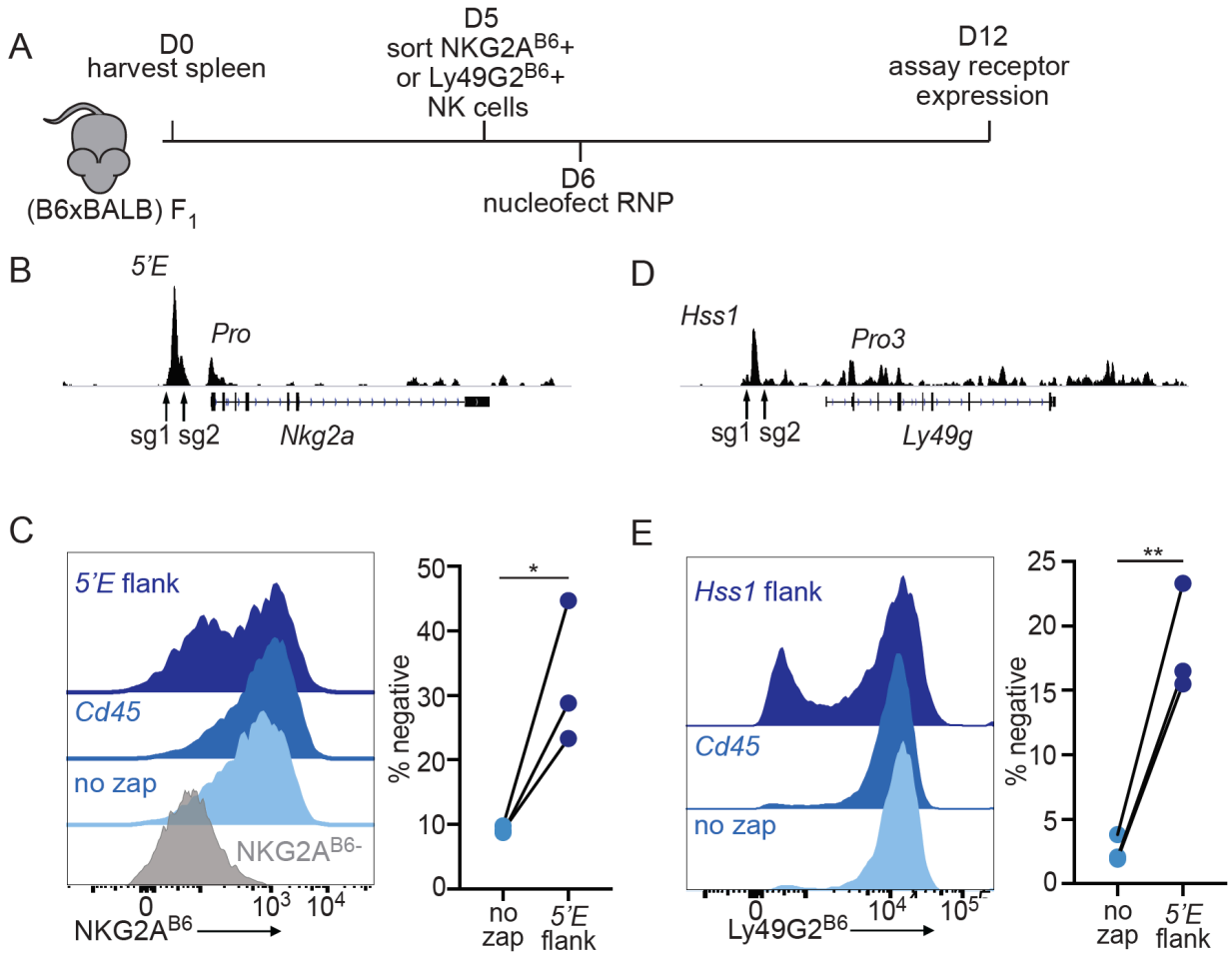


Fig. 4.7. The *Nkg2a*$_{5'E}$ and *Ly49g*$_{Hss1}$ enhancers are required to maintain gene expression in cultured NK cells. (**A**) Schematic of experimental design. Briefly, (B6 x BALB)F_1 splenocytes were cultured with IL-2 for 5 days, at which time NK cells positive for NKG2A^{B6} or Ly49G2^{B6} were sorted. After 1 day in culture in medium containing IL-2 to recover, the cells were nucleofected with Cas9-RNP complexed with the indicated sgRNA, or were not treated ("no zap"). and cultured for a further 6 days in IL-2 containing medium. Receptor expression was examined by flow cytometry on day 12. (**B**) Location of flanking guide RNAs used relative to accessibility peaks determined by ATAC-seq at the *Nkg2a* and *Ly49g* loci. (**C**) Flow cytometric analysis of NKG2A^{B6} expression by NK cells on day 12 (6 days after nucleofection with sgRNAs with the depicted targets). Control cells for comparison were sorted to be NKG2A^{B6}-negative on day 6 (grey). Data from 3 independent experiments are quantified on the right. (**D**) Flow cytometric analysis of Ly49G2^{B6} expression, as in (C). Data from 3 independent experiments are quantified on the right. Statistics carried out using a ratio paired t-test. *$P < 0.05$, **$P < 0.01$.

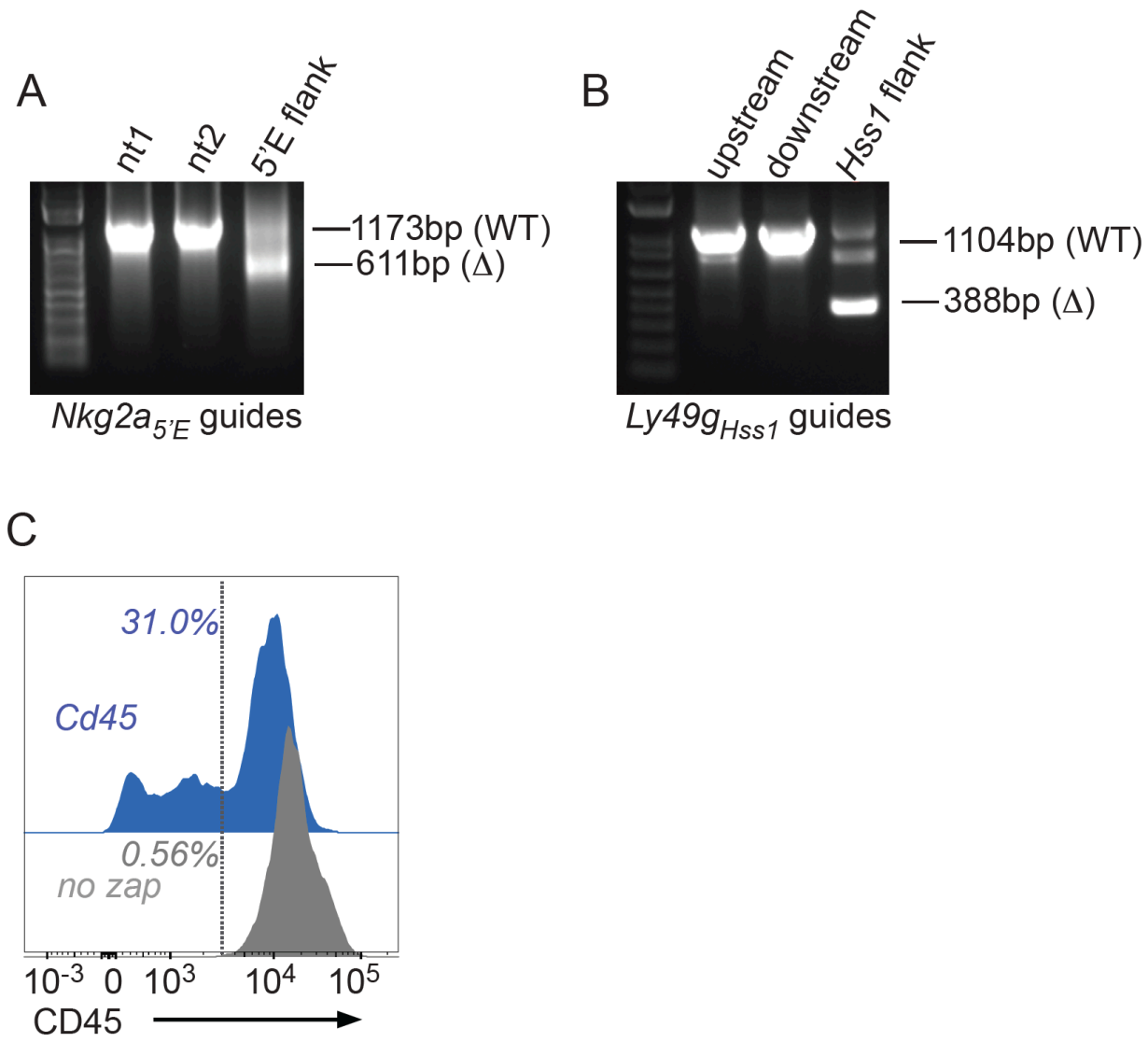


Fig. 4.8. Cas9-RNP editing of primary mouse NK cells. (**A**) Deletion test of sgRNAs used to delete *Nkg2a*$_{5'E}$. Agarose gel electrophoresis of PCR amplified fragments from the region surrounding *Nkg2a*$_{5'E}$ in cells nucleofected with either non-targeting (nt) guides, or guides flanking *NKG2a*$_{5'E}$ (5'E flank) in IL-2 cultured mouse splenocytes. Amplicon sizes of the WT and D bands are depicted. NK cells enriched using the MojoSort Mouse NK cell isolation kit from Biolegend were cultured for 6 days in IL-2 containing media before samples of 1 x 10^6 cells were nucleofected on day 6. On day 9, gDNA was prepared and used as a template for PCR. (**B**) Deletion test, as in (A), of the sgRNA pair used to delete *Ly49g*$_{Hss1}$ in comparison to results with only the upstream or downstream gRNAs. (**C**) CD45 staining of samples from the experiment depicted in Fig. 4C. Cells sorted to be NKG2A^{B6+} or NKG2A^{B6-} were nucleofected (or not) with Cas9 and the indicated sgRNAs. (**D**) CD45 staining of samples from the experiment depicted in Fig. 4D. (**E**) CD45 staining of IL-2 cultured NK cells isolated from splenocytes using the Mojosort kit on day 2 of culture, nucleofected with a single against *Cd45* (or not) on day 5 of culture and stained for CD45 expression on day 10 of culture.

Discussion

Taken together, our results show that silent and active alleles of the *Nkg2a* and *Ly49g* RME genes differ with respect to accessibility and active histone modifications at their promoters and gene bodies, and lack a repressive state altogether. Their respective enhancers are constitutively accessible and activated. These findings are consistent with a binary model of enhancer action where enhancers are activated in a deterministic fashion, while their effect on stable target gene activation is probabilistic, thus explaining the decoupling of enhancer and promoter accessibility seen at RME loci. In combination with the findings in Chapter 3, these results further suggest that RME is the consequence of limiting enhancer action, and may evolve in some cases, such as the NK receptor genes, to generate diversity in otherwise ontogenetically identical cells. In other instances, RME may arise fortuitously simply because of the absence of purifying selection for highly penetrant gene expression in a lineage.

Our data provide evidence that silent RME alleles may not require a repressed chromatin state for mitotically maintained silencing. Instead, they represent an inactive chromatin state similar to lineage non-specific genes, which also lack traditional repressive chromatin modifications, and are generally not subject to subsequent activation after the initial failure to be activated.

Notably, silent alleles of the NK receptor genes have been reported to have CpG-methylated promoters (Rogers et al. 2006; Rouhi et al. 2007; Rouhi et al. 2006; Rouhi et al. 2009). We believe DNA methylation plays at most a minor role since inhibitors of DNA methylation did not appreciably activate expression of silent alleles in our studies or those of others (Rogers et al. 2006; Rouhi et al. 2007; Rouhi et al. 2006; Rouhi et al. 2009). Furthermore, NK receptor genes are CpG poor (Rouhi et al. 2006).

Our findings also indicate that the active alleles of the NK receptor genes require enhancer-driven maintenance of the active state. The molecular and biochemical mechanisms that maintain expression through cell division are unknown but are likely shared with active genes broadly, and may employ the maintenance of nucleosome-free promoters through S and M phase. Indeed, Chapter 3 demonstrates RME in genes previously thought to be ubiquitously expressed, albeit at much lower frequencies than the RME NK receptor genes. These data support the notion that the distinction between RME and ubiquitously expressed genes is artificial, or quantitative rather than qualitative, and thus the mechanisms driving mitotic stability of expression states are likely shared by RME genes and other genes.

Chapter 5
Generation and testing of a *Ly49a*$_{Hss1\text{-}GintoA}$ knockin allele

Introduction

The variegated expression pattern of the NK cell receptors is regulated proximally (Tanamachi et al. 2004). We previously identified a DNase I hypersensitive site, HSS-1, located 5kb upstream of the *Ly49a* promoter that was required for the expression of a genomic *Ly49a* transgene. Each *Ly49g* gene harbors a similar upstream element. It was subsequently proposed that this element acts as a bidirectional switch in developing NK cells (Saleh et al. 2004; Saleh et al. 2002). In this model, transcription at HSS-1 can proceed in either the sense or antisense direction, but not both, due to the overlapping nature of the divergent promoters, and the probability with which transcription elongates in the sense direction toward the downstream gene is what determines whether the locus will be expressed in the fully differentiated NK cell.

Whether HSS-1 determines the expression frequency of the Ly49 receptors has not been established *in vivo*. To address this, we set out to create an allele where the HSS-1 element from the *Ly49g* gene (expressed on ~50% of NK cells) is knocked in to the *Ly49a* locus (expressed on ~17% of NK cells). The difference in the expression frequency between these two receptors provided a large dynamic range within which we hoped to observe changes in the expression frequency of the Ly49A receptor. If HSS-1 determines the expression frequency of the downstream proximal *Ly49* gene, we predicted that Ly49A would be expressed at a similar frequency to Ly49G2. Alternatively, if HSS-1 does not determine the expression frequency of the downstream gene, we expect to see a Ly49A expression frequency closer to the WT 17%.

Our group has tried to generate this knockout allele over the course of the last ~15 years through traditional gene targeting approaches in mouse ESCs. At least partially due to the highly homologous nature of the *Ly49* gene cluster, it was difficult to isolate correctly targeted ESC clones, and these did not contribute to the germline of resultant founder pups. With the advent of CRISPR/Cas9 genome editing, new approaches allowing direct targeting in the single cell embryo have become available. We decided to use a recently published approach that uses a long single strand DNA (lssDNA) as the repair template, and two guides flanking the targeted enhancer, called *Easi*-CRISPR (Quadros et al. 2017). Luckily, we obtained a single founder animal with the correctly targeted allele, which we designate $Ly49a_{Hss1\text{-}GintoA}$.

Approach and Results

To generate the $Ly49a_{Hss1\text{-}GintoA}$ allele, we followed the approach recently described (Quadros et al. 2017). The targeting is outlined in Fig. 5.1. The guide RNAs used are the same as those used to generate the $Ly49a_{Hss\Delta}$ allele in Chapter 3 (Table 2.1). Importantly, the sgRNAs chosen are not predicted to cut at the *Ly49g* locus, due to mismatches in the 3' region of the guide RNA and/or a disrupted protospacer adjacent motif. The single strand donor construct was designed by aligning the sequences including and surrounding the HSS-1 elements from the *Ly49a* and *Ly49g* genes. The DNA sequences were derived from the mm10 annotation of the B6 genome (UCSC Genome Browser). LoxP sites were inserted in the construct at the exact sites that were cognate to the double strand break points in the *Ly49a* locus. This served the purpose of eliminating any off-targeting cutting of the construct after repair and integration, and allows future experiments where the inserted enhancer can be excised via Cre-Lox recombination.

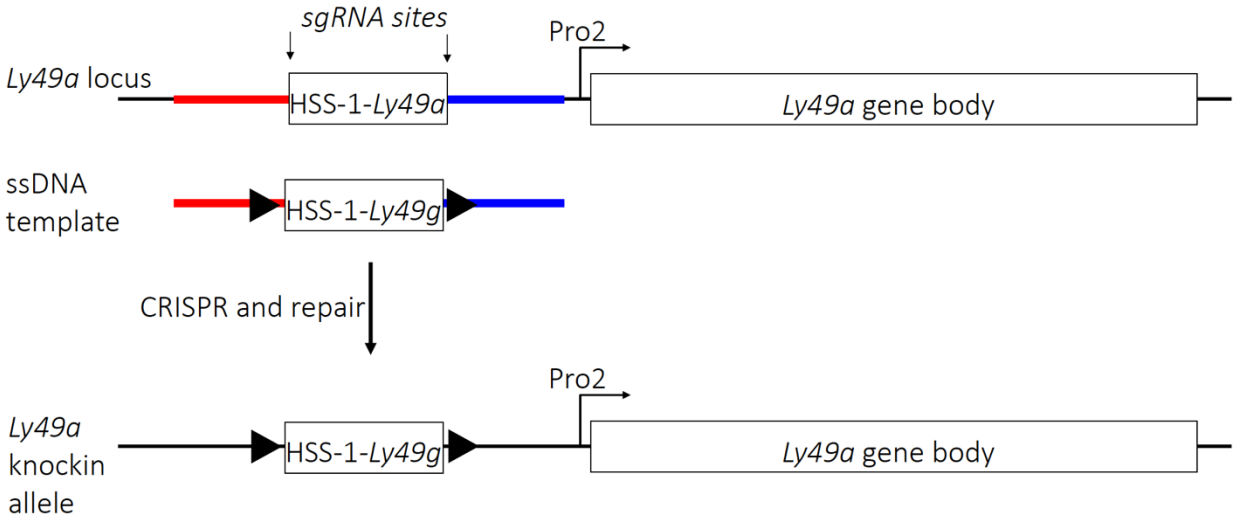
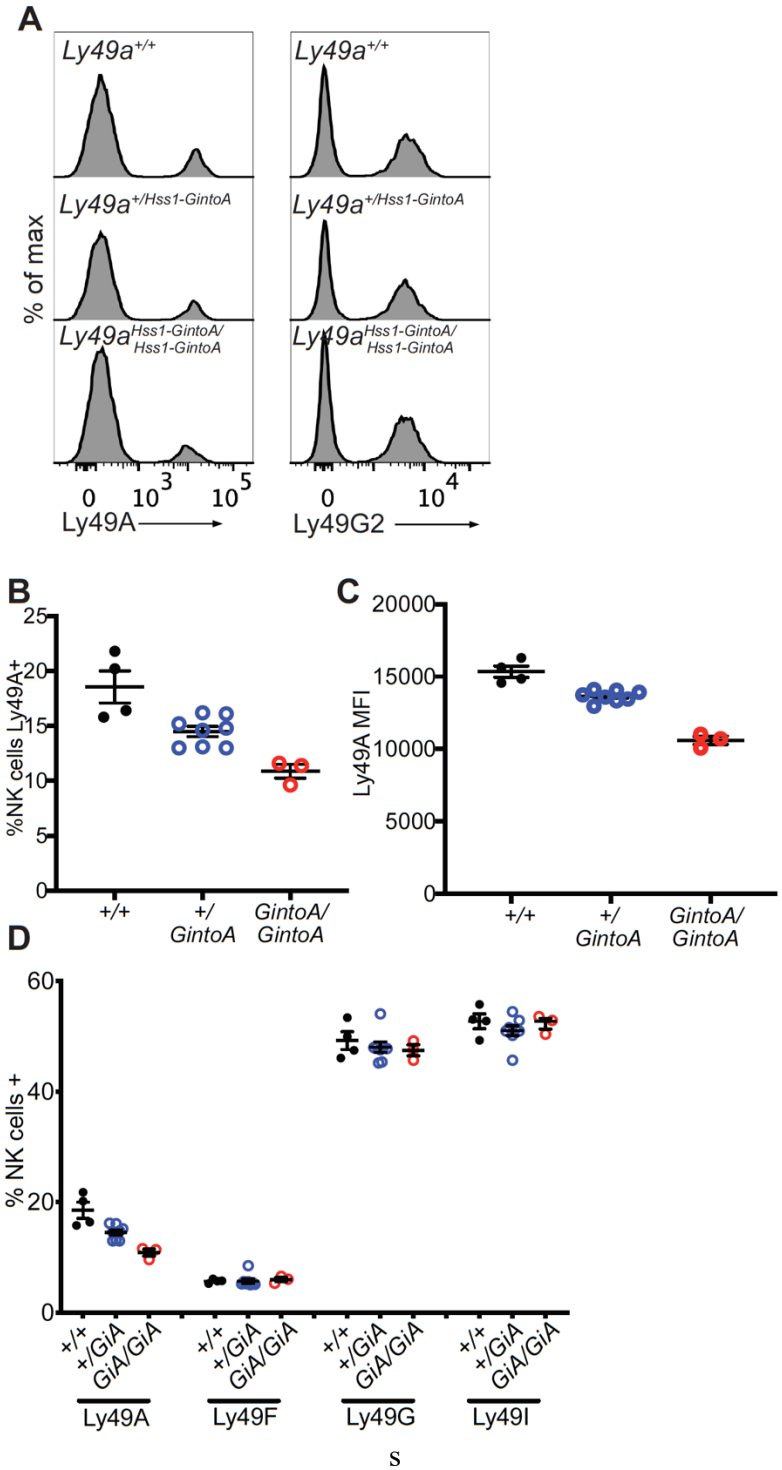


Fig. 5.1. Schematic of the strategy to generate the *Ly49a*$_{Hss1\text{-}GintoA}$ allele. The endogenous *Ly49a* locus is depicted (top). sgRNA cut sites are depicted with arrows. Homologous sequences flanking the *Ly49a*$_{Hss1}$ enhancer and the knockin enhancer are denoted in red and blue, and represent ~100 bp of sequence (not to scale). LoxP sites are indicated as black arrows in the repair template (middle). The synthetic knock in allele is shown on the bottom. DNA sequence between the LoxP sites is derived from the *Ly49g* locus, while sequence outside of the LoxP sites is natural *Ly49a* sequence.

The lssDNA repair construct was custom ordered from Genewiz as a 992bp double stranded fragment gene product. The GUIDE-IT strandase kit from Clonetech allows generation of lssDNA using a 5' phosphate-dependent single strand digestion. One primer was phosphorylated on the 5' end, while the other primer was not phosphorylated. The resultant ssDNA template was checked for quality by gel electrophoresis, bioanalyzer and Sanger sequencing. The guide RNAs were purchased from IDT as a separate cripsr and tracr RNAs (the Alt-R system), Cas9 was purchased from the UC Berkeley Macro lab and embryo microinjections were performed as previously described (Modzelewski et al. 2018). The injection mix contained 100 ng/ul Cas9, 50 ng/ul of both crispr and tracr RNAs, and 40 ng/ul of the lssDNA repair template, and 90 embryos were injected. Only 3 live pups were born, and one female harbored the *Ly49a*$_{Hss1\text{-}GintoA}$ allele. This allele was confirmed by PCR amplification and Sanger sequencing of the entire allele.

The *Ly49a*$_{Hss1\text{-}GintoA}$ allele was backcrossed to B6, and heterozygous F1 animals were intercrossed to generate WT, heterozygous and homozygous animals with respect to *Ly49a*$_{Hss1\text{-}GintoA.}$ Splenocytes from 8 week old animals were stained for flow cytometry to measure the frequency of Ly49A and Ly49G2 expression on NK cells (Fig. 5.2). Curiously, expression of Ly49A did not increase in *Ly49a*$^{Hss1\text{-}GintoA/\ Hss1\text{-}GintoA}$ animals (Fig. 5.2, A and B), an in fact appeared to decrease slightly. Expression per cell measured by the mean fluorescence intensity of staining (MFI) suggested a slight decrease in expression per cell (Fig. 5.2C). The expression frequency of Ly49G2 was unaltered, as were Ly49F and Ly49I (Fig. 5.2D).



s

Fig. 5.2. Phenotype of mice harboring the *Ly49a*$_{Hss1\text{-}GintoA}$ allele. (**A**) Representative Ly49A and Ly49G2 flow cytometry staining of splenic NK cells in WT, heterozygous and homozygous mice with respect to the *Ly49a*$_{Hss1\text{-}GintoA}$ allele. (**B**) Quantification of data depicted in (A), n=3-7 per group. (**C**) Expression per cell of Ly49A depicted as mean fluorescence activity of staining (MFI) for all 3 genotypes. (**D**) Expression frequencies of Ly49A as in (B), and the other indicated Ly49 receptors.

Discussion

The results of our successfully generated *Ly49a*$_{Hss1\text{-}GintoA}$ allele suggest that HSS-1 is not completely responsible for the determination of Ly49 receptor expression frequency by NK cells. A caveat to this interpretation is that in the generation of the synthetic enhancer allele, disruptions of the endogenous sequence might have resulted in a weakened enhancer element. However, this result is consistent with the observation in Chapter 3 that *Ly49g*$_{Hss5}$ plays a major role in determining the expression frequency of Ly49G2.

This result suggests that the differences in enhancer strength between the various *Ly49* loci that drive the different expression frequencies are not due to differences in HSS-1 strength. We note in Chapter 4 that the various HSS-1 elements are similarly accessible and occupied by transcription factors, providing further support for this view. Rather, the relevant differences in enhancer strength are likely found at non-HSS-1 enhancers that function in a partially redundant manner to raise *Ly49* expression frequencies. It is also possible that the promoter regions contribute to the expression frequencies of the *Ly49* genes. While our results in enhancer deletion mice strongly suggest that enhancers play the dominant role in determining expression frequency, we cannot completely discount the effects of promoter elements. Determining the contributions of promoter sequences and non-HSS-1 enhancers such as HSS-5 will require the generation of further knock in mice.

Chapter 6
A forward ENU mutagenesis screen to identify variegating factors

Abstract

The natural killer (NK) cell receptors are expressed in a variegated, monoallelic and mitotically stable fashion. This expression pattern is required to generate diversity within the NK cell population for different alleles of MHC I, which serves as an inhibitory ligand for NK cells. The molecular mechanisms by which this expression pattern is enacted are poorly understood. Here, we undertook a forward ENU mutagenesis screen in mice to identify factors involved in the generation of the variegated Ly49 expression pattern. We isolated a mutant that displayed an ~3-fold reduction in the expression frequency of the Ly49A receptor on NK cells. Mutant mice harbored a point mutation in the DNA-binding R2-R3 domain of the transcription factor, c-Myb. The mutant allele, *c-myb*^{D100G}, appeared to result in homozygous lethality supporting a severe loss of function, and the mutant phenotype segregated as dominant. We identify the causative lesion in the transcription factor c-Myb, which directly controls the expression frequency of the Ly49A inhibitory receptor.

Introduction

Natural Killer (NK) cells constitute a critical arm of the innate immune system. NK cells are lymphocytes, like T and B cells, but rather than recognize foreign antigen they recognize stress signals encoded by the host (Raulet and Vance 2006; Raulet, Vance, and McMahon 2001; Shifrin, Raulet, and Ardolino 2014). Additionally, NK cells scan host cells for loss of MHC I expression, a feature of viral infection and tumorigenic transformation (Raulet and Vance 2006). This detection is mediated by a loss of inhibitory signaling through receptors that recognize class I MHC. Unlike T and B cells, NK cells do not rearrange their antigen receptor genes to generate diversity for their target cells. Instead, NK cells draw on a pool of germline-encoded receptors that are expressed in a variegated, monoallelic, overlapping and mitotically stable fashion to generate an NK cell population comprised of a diverse repertoire of specificities for MHC I (Raulet, Vance, and McMahon 2001). In mice the major family of MHC I inhibitory receptors is the Ly49 family. These receptors are encoded in the natural killer cell gene complex (NKC) on mouse chromosome 6 in a tandem array. The mechanisms regulating expression are poorly understood, but of great importance to NK cell biology and the field of random monoallelic gene expression (RME).

Previously, our group found that the *Ly49* genes are regulated proximally using a 30kb genomic *Ly49a* transgene, comprising the gene body and 10kb of upstream sequence (Tanamachi et al. 2004). Transgenic mice expressed the transgene at frequencies approximating the endogenous *Ly49a* gene, at ~17% of NK cells. Furthermore, a DNase I hypersensitive site, HSS-1, was identified ~5kb upstream of the *Ly49a* promoter. This element was necessary for expression of the transgene. Curiously, B cells, which do not normally express Ly49 receptors, expressed the transgene uniformly. This result suggested that, A) the transgene lacked a B cell specific repressor element, and B) NK cells may express a factor (or factors) that actively impose a variegated expression pattern, and this variegation machinery is lacking in B cells.

We performed an *N*-Ethyl-*N*-Nitrosourea (ENU) forward mutagenesis screen in mice of the C57BL/6J (B6) genetic background to identify factors responsible for the variegated expression pattern of the Ly49 receptors in NK cells. ENU screens generally generate single

nucleotide variants (SNVs), biasing the nature of hits to protein-coding and generally hypomorphic mutations (Cordes 2005). Here, we identify an ENU-generated mutant in the transcription factor c-Myb that results in a reduced frequency of Ly49A expression in NK cells.

Results

ENU screens generally result in recessive loss-of-function mutations in protein coding genes (Cordes 2005). We employed a screening strategy to search for such mutants in the G3 generation generated under the following screening scheme: ENU treated males (G0) were mated with B6 females to generate heterozygous G1 mutants. G1 males were again outcrossed to B6 females to generate G2 animals. G2 females were backcrossed to the G1 male to generate homozygous mutants in the G3 generation (Fig. 6.1A). G3 animals were again outcrossed to B6, set up in a G4 x G3 backcross to propagate mutant alleles, and were then sacked and screened for NK cell receptor phenotypes by clow cytometry with a panel of antibodies against the variegated NK receptor genes.

After screening ~450 animals, a G3 mutant with the identifier 3921 was found to display a low frequency of Ly49A expression. B6 mice express Ly49A on ~17% of natural killer cells while 3921 displayed a frequency of Ly49A 2 standard deviations (SDs) lower than control B6 mice (Fig. 6.1C). Progeny of 3921 were propagated in a series of crosses and backcrosses to determine the segregation behavior of the mutant phenotype (defined as a frequency of Ly49A+ cells at least 2 SDs below the mean of B6 controls). It was initially assumed that the phenotype would be recessive. The segregation of the phenotype was not, however, that of a simple Mendelian recessive trait. The progeny of mutant mice with WT B6 mice frequently displayed the mutant phenotype, suggesting a dominant behavior. Additionally, the phenotype was not true-breeding, as the progeny of two mutant mice were occasionally WT (Fig. 6.1C). These observations were consistent with a dominant phenotype. To identify the molecular lesion responsible for the mutant Ly49A low phenotype, we selected 4 animals from the pedigree for whole exome sequencing with 60x coverage of the coding genome. We selected two mutant siblings (42-1 and 42-3), a mutant cousin (45-1), along with a WT mouse (35-3) (Fig. 6.1C). WT mouse 35-3 is not displayed but is an “uncle” of mutant mouse 45-1.

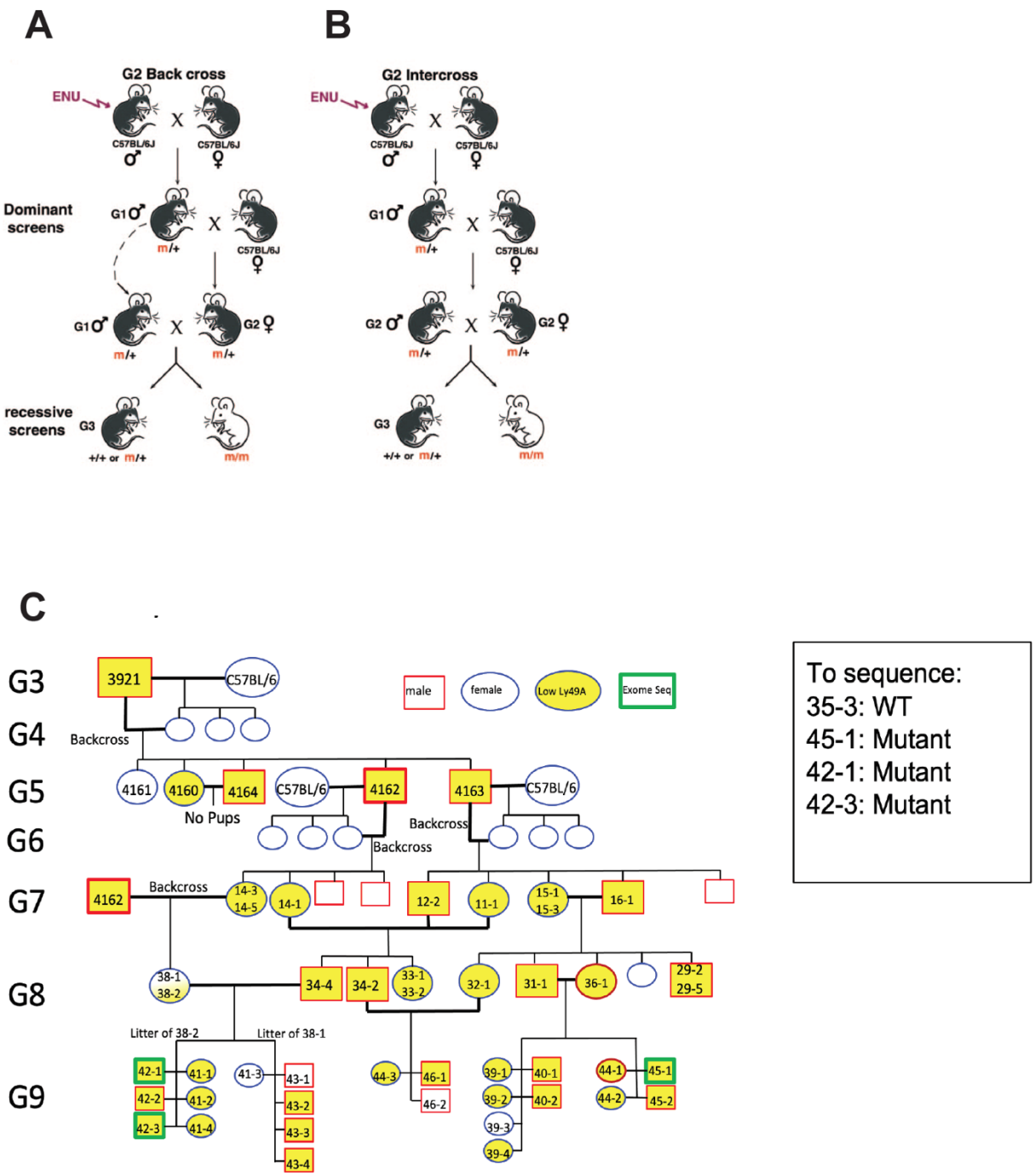


Fig. 6.1. ENU screen scheme and pedigree. (**A-B**) Schematic of the ENU screen breeding strategy. The screen was performed as depicted in (A), where G2 females were backcrossed to G1 males. G3 animals were screened for recessive phenotypes. The figure was adapted from (Cordes 2005) (**C**) Abridged pedigree of the progeny of the G3 mutant mouse. Male animals are depicted by squares while female animals are shown as ovals. Yellow animals displayed the low Ly49A phenotype, defined as 2 standard deviations below the mean value of a concurrently stained group of B6 mice. Animals that were exome sequenced are indicated in green boxes, and are listed in a box on the right. The "WT" control animal is not depicted and is a WT male littermate of the mutant female 36-1, the mother of 45-1.

gDNA from the selected animals was captured with a library of exon probes, sequenced on a HiSeq 2500, and mutations with respect to the reference B6 mm10 genome were called using the GATK software package. This resulted in a variant calls output format and a paired text file annotating the predicted effects of identified variants. The number of protein coding changes per animal ranged from 13,891-24,020, while the number of missense mutations was between 911-957 per animal. We wrote scripts in Python to parse predicted protein-coding mutations in each of the 3 mutants and the WT control animal. First, under the assumption of a recessive phenotype, we searched for mutation that was homozygous in all 3 mutant animals and was either heterozygous or absent entirely in the control WT animal. This yielded no such mutations. Initially, this was surprising given the design of the screen: we reasoned that the variegating factor would suffer a hypomorphic loss of function mutation, resulting in a phenotype that segregates as recessive.

As discussed above, the segregation of the phenotype in the ENU pedigree is more consistent with a dominant phenotype. We wrote a new Python script targeted at the identification of any protein coding mutation present in the 3 mutant animals and lacking in the WT control animal. This identified many genes with multiple mutations that were not specific to the screen since these genes harbored dozens of mutations (data not shown). We further narrowed the script to identify only missense mutations present in the 3 mutants and lacking in the WT control. This resulted in 37 hits (Fig. 6.2A). Many of these genes (e.g., *Ugt1a1*, *Muc4* and *Vmn2r114*) harbored many mutations, as depicted in (Fig. 6.2, D and E). These genes overlapped with the hits identified by the script searching for all protein coding changes. These genes were ignored as artefactual. The most interesting hit was in the gene encoding the transcription factor (TF) c-Myb (Fig. 6.2, A-C). This gene harbored a single identified point mutation which was present in a heterozygous configuration in the 3 mutant mice and was not present in the WT control animal (Fig. 6.2, B and C). The mutation changes a single amino acid in the R2-R3 DNA binding domain; D100G (Fig. 6.2F). The c-Myb transcription factor plays a key role in the development of many hematopoietic cell types, and importantly homozygous knockout mice do not survive past embryonic day 15 due to severe anemia (Mucenski et al. 1991). Interestingly, previous analysis on our lab identified a c-Myb binding motif upstream of the *Ly49a* gene, suggesting direct regulation (Tanamachi et al. 2004). We reasoned that the *c-myb*^{D100G} allele generated in our screen may be a severe hypomorph or perhaps even an amorphic allele explaining why all 3 mutant mice we sequenced were heterozygous since the homozygous configuration would be embryonic lethal.

A

Chromosome	Position	Gene
chr1	85100560	*A530032D15Rik*
chr1	88212536	*Ugt1a1*
chr1	88234599	*Heatr7b1*
chr1	172490322	*Igsf9*
chr4	147179483	*Gm13139*
chr4	147755594	*Gm13157*
chr4	156339025	*VMN2R-PS160*
chr5	26105203	*5031410I06Rik*
chr6	77605186	*Ctnna2*
chr9	3001048	*Gm10722*
chr9	3026368	*Gm10716*
chr9	103355216	*Cdv3*
chr9	103365299	*Cdv3*
chr9	103365303	*Cdv3*
chr10	21152899	*Myb*
chr14	19415942	*AC242409.1*
chr15	10147462	*Krt86*
chr15	10147463	*Krt86*
chr16	32753802	*Muc4*
chr16	32753819	*Muc4*
chr16	32753845	*Muc4*
chr16	32753990	*Muc4*
chr16	32755843	*Muc4*
chr17	23291755	*Vmn2r114*
chr17	23291782	*Vmn2r114*
chr17	23291787	*Vmn2r114*
chr17	23291821	*Vmn2r114*
chr17	23291829	*Vmn2r114*
chr17	23291838	*Vmn2r114*
chr17	23291839	*Vmn2r114*
chr17	23291841	*Vmn2r114*
chr17	23388841	*Vmn2r116*
chr17	23401589	*Vmn2r116*
chr17	23401601	*Vmn2r116*
chr17	23478268	*Vmn2r117*

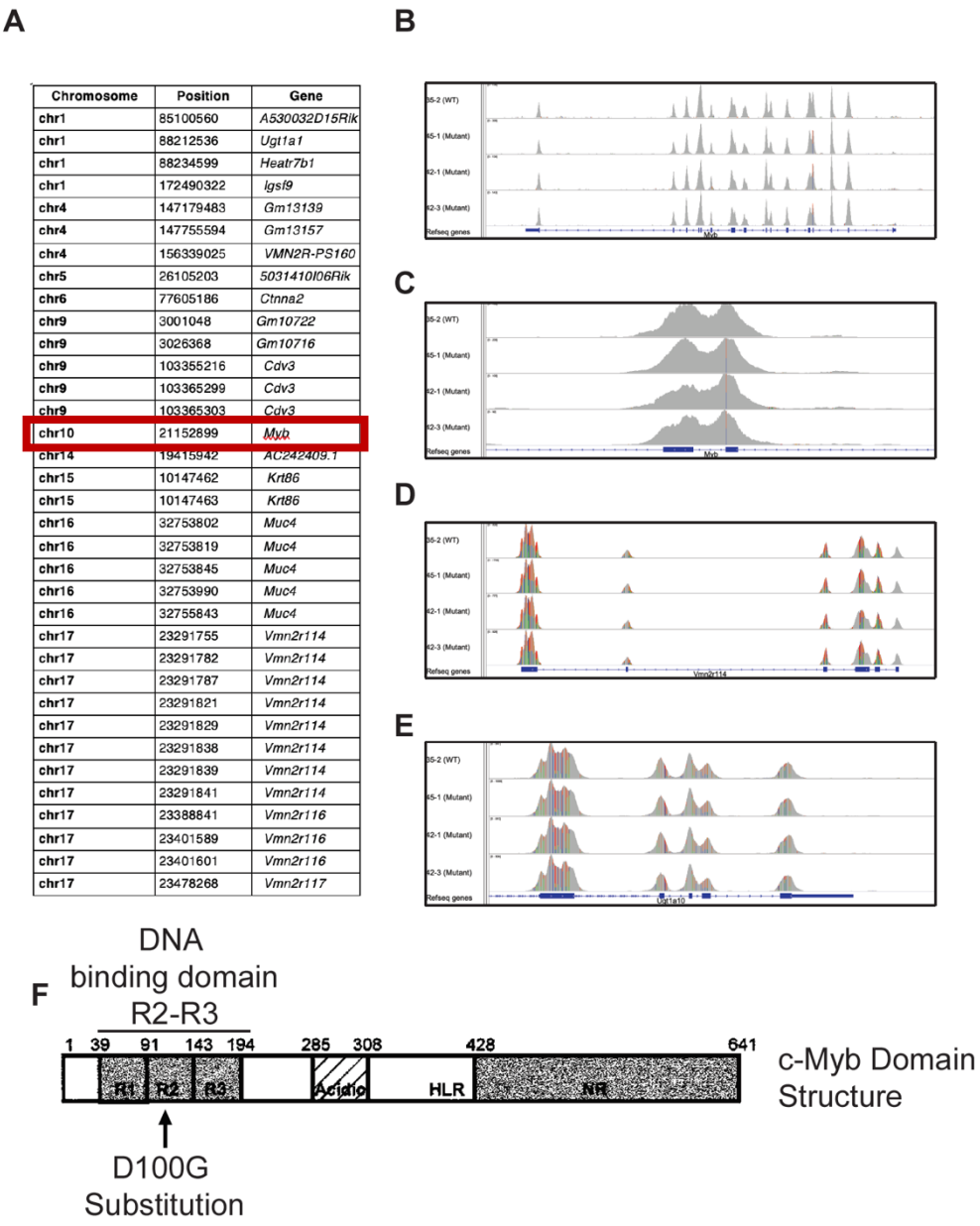


Fig. 6.2. Identified molecular lesions through exome sequencing. (**A**) Table of genes with identified variations resulting in a missense mutation as identified by the GATK software (Broad Institute). (**B**) IGV screenshot depicting aligned exome-sequence data from each sequenced mouse (BAM format) at the *c-myb* locus. The location of the mutation is denoted by the vertical red and blue lines, which represent an approximate 50:50 distribution of the reference "T" nucleotide and the mutant "C" nucleotide. No *c-myb* variant was identified in 35-2, so the signal is a continuous grey. **(C**) Expanded view of the *c-myb* locus as in (B). (**D-E**) Examples of genes identified to have many mutations that arose as hits in the screen. Genes that displayed a high apparent mutation rate were ignored as non-specific as they had many mutations in the WT animal. (**F**) c-Myb domain structure and location of the predicted c-Myb^{D100G} mutation in the DNA-binding R2-R3 domain. Modified from(Ladendorff, Wu, and Lipsick 2001)

To screen a wider panel of animals from our mutant pedigree, we designed a PCR primer pair flanking the identified mutation. To validate the primers, we used them to amplify the region of the *c-myb* gene harboring the mutation from gDNA isolated from the 4 mice that were exome sequenced. Sanger sequencing of the PCR product confirmed the results of the exome sequencing (Fig. 6.3A) and validated that we could identify heterozygous mutations through this approach.

To verify the association of the identified mutation, we analyzed further progeny of the mutant sibling males 42-1 and 42-3. These males were bred with their female progeny of both wildtype and mutant phenotype to generate an assortment of progeny with respect to the Ly49A phenotype. Splenocytes from these animals were stained for flow cytometry and gDNA was isolated from their tail for PCR with our primer pair. In addition to staining for Ly49A, we also analyzed Ly49F, Ly49C, Ly49G and Ly49I (Fig. 6.3B). Animals harboring the *c-myb*^{D100G} allele displayed a more than 2-fold reduction in the expression frequency of Ly49A. Additionally, we observed statistically significant reductions in Ly49F and Ly49G. As was seen during the early screening process, the mutant animals also displayed a slightly increased frequency of Ly49I expression. Importantly, none of the animals harbored the *c-myb*^{D100G} allele in a homozygous configuration, supporting the hypothesis that the *c-myb*^{D100G} allele is null. These results strongly supported the *c-myb*^{D100G} allele as the causative lesion in our screen. This result was initially exciting, but we promptly found that the association between the c-Myb transcription factor and Ly49A expression frequency had previously been identified (Bezman et al. 2011). In that study, heterozygous *c-myb* knockout animals displayed a strikingly similar phenotype with respect to Ly49A expression on NK cells. While disappointing, this further suggested the *c-myb*^{D100G} allele was amorphic and causative.

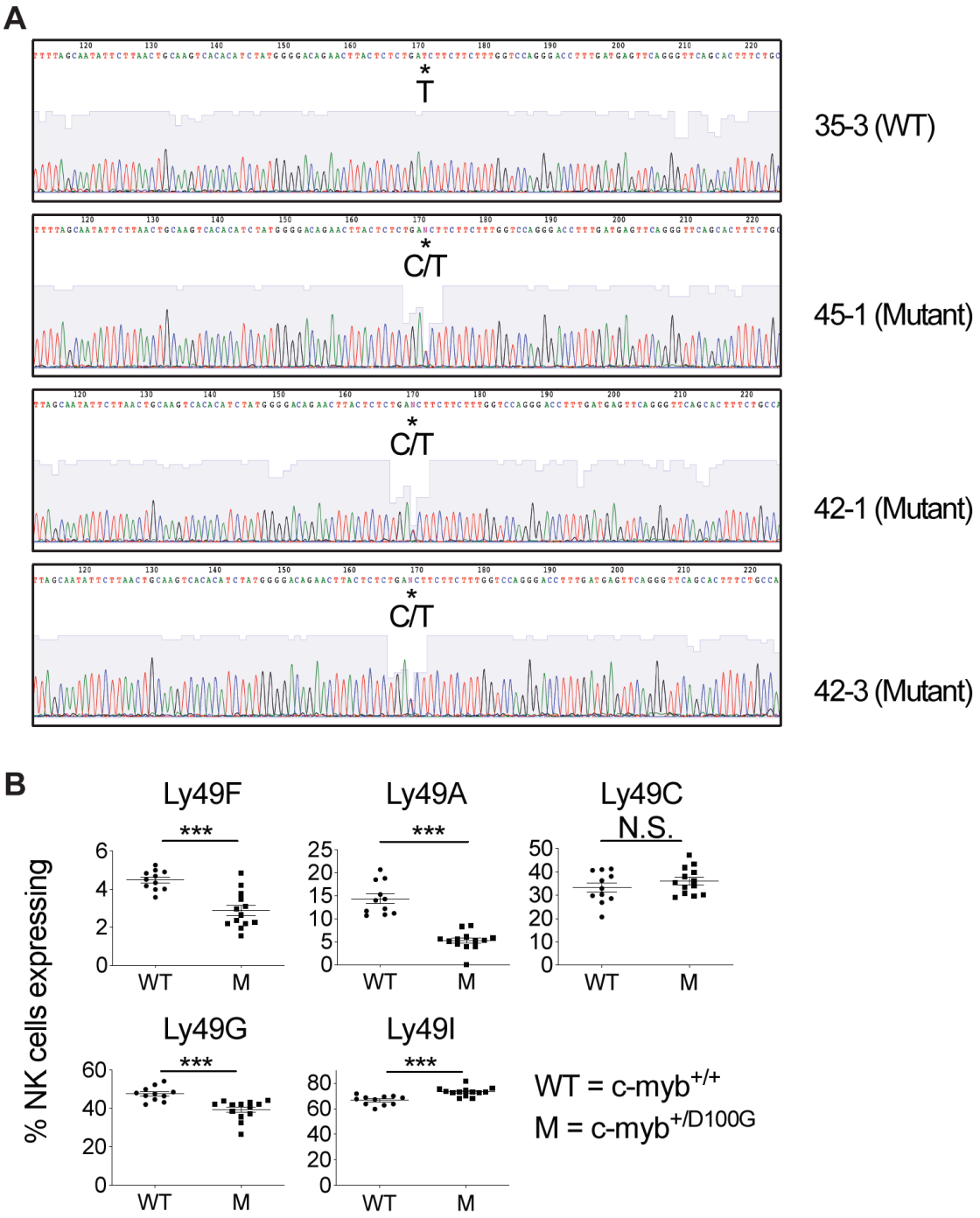


Fig. 6.3. Validation of the *c-myb*^{D100G} variant in mutant mice. (**A**) Sanger sequence traces of PCR products amplified from each of the 4 sequenced mice. Primers flank the location of the "A" to "G" transition mutation in the coding sequence, changing Aspartate 100 to Glycine (D100G). The read corresponds to the minus strand, so the reference allele is a "T" and the mutant allele is a "C." Shaded blue indicates confidence level of base calls; note the dip in the confidence at the location in the trace where the read gives signal from both "C" and "T" nucleotides. (**B**) Quantified flow cytometry data generated from the progeny of mutant mice 42-1 and 42-3. Blood was stained with antibodies against the indicated receptors on NK cells (all data is gated on NK cells). Statistical analysis was done by student's *t*-tests. ****P<0.001*.

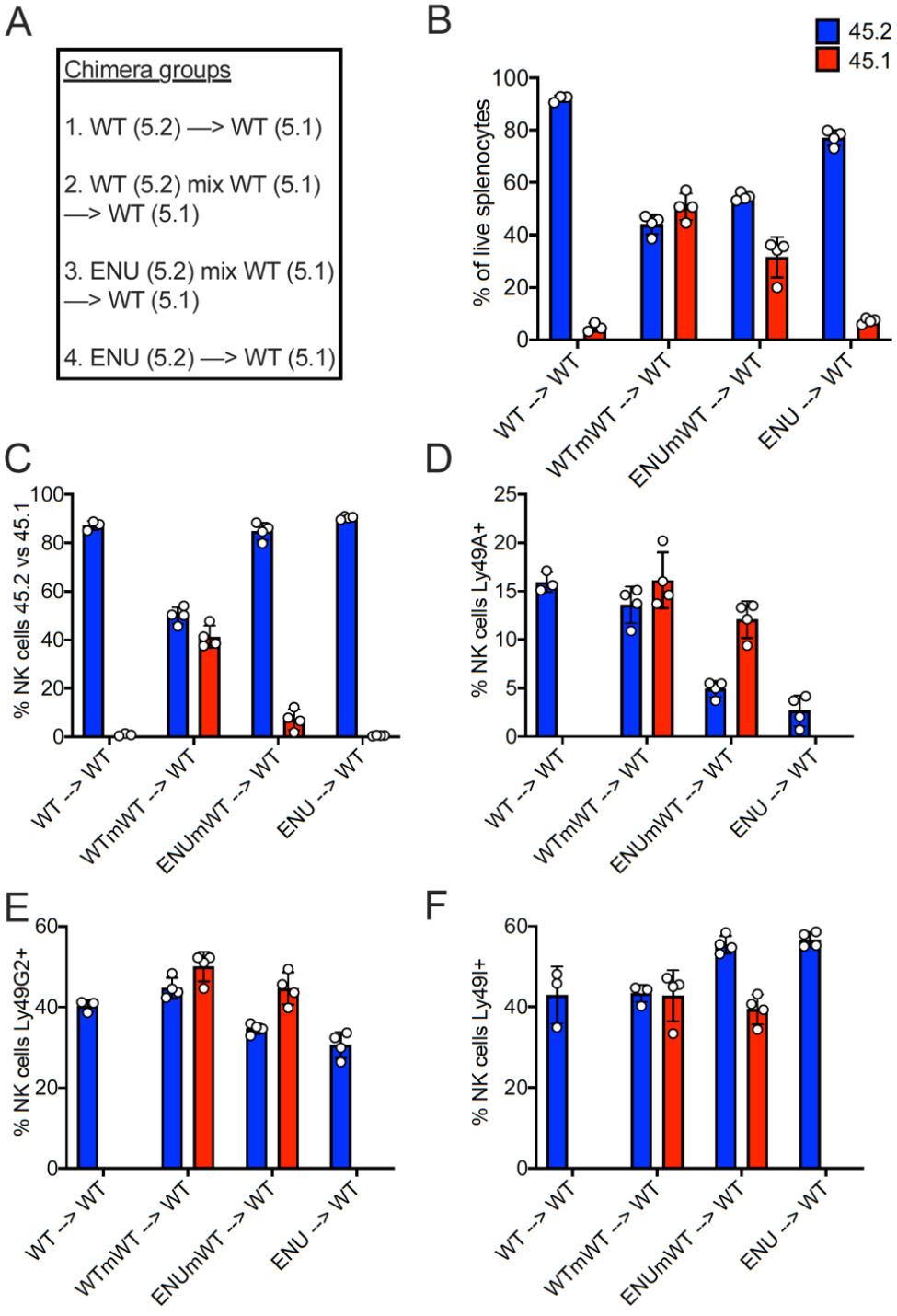


Fig. 6.4. The mutant phenotype is cell intrinsic. (**A**) Experimental and control groups in the chimera experiment. (**B**) Chimerism is measured by percent of live splenocytes that are either CD45.1+ or CD45.2+. The two mixed chimera groups (middle) are indicated as WTmWT (WT mix WT) or ENUmWT (*c-myb*^{D100G} ENU mutant chimera mixed with WT). (**C**) Chimerism in the NK cell population in chimeric mice as in (B). (**D-F**) Expression of the indicated Ly49 receptors delineated by donor origin.

We asked whether the Ly49A phenotype resulting from the *c-myb*^{D100G} allele is cell-intrinsic. To address this, we generated a series of bone marrow chimeras. We lethally irradiated B6.SJL mice (CD45.1) to serve as hosts to transferred bone marrow from either (1) WT B6 (CD45.2), (2) a mix of WT B6 (CD45.2) and WT B6.SJL (CD45.1), (3) a mix of the *c-myb*^{D100G} (CD45.2) and WT B6.SJL (CD45.1), or (4) the *c-myb*^{D100G} allele (CD45.2) alone (Fig. 6.4A). Eight weeks after reconstitution, we stained splenocytes for flow cytometry to assess whether the low Ly49A mutant phenotype is determined by the presence of the *c-myb*^{D100G} allele in a cell intrinsic fashion. Chimerism was approximately equal in the mixed bone marrow chimeras (groups 2 and 3), with a slight overrepresentation of the hematopoietic compartment derived from the *c-myb*^{D100G} donor relative to wild type in group (3) (Fig. 6.4B). NK cells were overwhelming derived from the *c-myb*^{D100G} donor in this group (Fig. 6.4C). This was consistent with the relative chimerism observed in mixed *c-myb* knockout and wild type chimeras generated previously (Bezman et al. 2011). This effect suggests a competitive advantage in the hematopoietic compartment when one *c-myb* allele is mutated despite the complete dependence of hematopoiesis on at least one *c-myb* allele. The observed effects on Ly49A, Ly49G2 and Ly49I expression were completely cell intrinsic (Fig. 6.4, D-F).

Since the *c-myb*^{D100G} mutation affects the DNA binding domain, we asked what the structural implications of the loss of the positively charged aspartate 100 might be. We found a crystal structure of the DNA-binding R2-R3 domain of c-Myb complexed with DNA on Protein Data Bank (PDB) and visualized the electrostatic interactions formed (Fig. 6.5). D100 does not interact directly with DNA, but may be involved in forming a stabilizing salt bridge important to the overall integrity of the DNA binding domain.

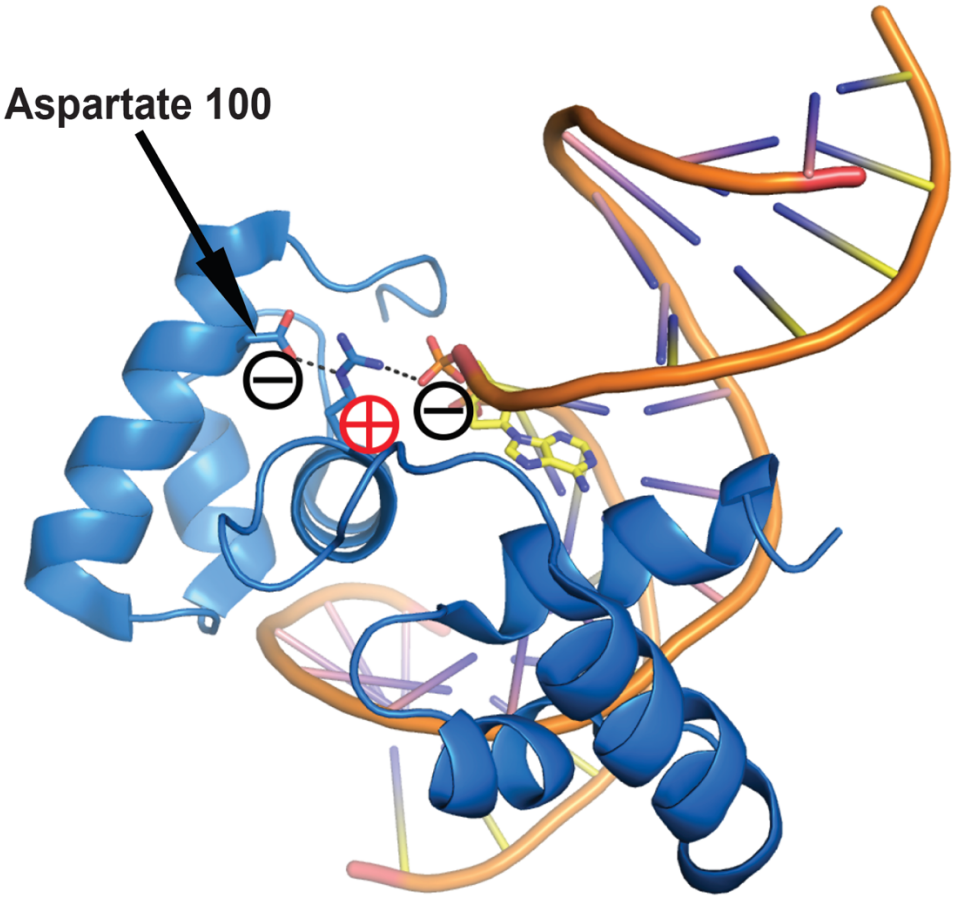


Fig. 6.5. Position of Aspartate 100 in the crystal structure of the c-Myb R2-R3 DNA-binding domain. Cristal structure of the c-Myb R2-R3 DNA binding domain was sourced from the Protein Data Bank (PDB). Structure ID is 1GV2. The position of the Aspartate 100 residue is displayed. A putative salt bridge formed with an Asparagine on the adjacent alpha helix and the DNA backbone is indicated.

Discussion

We isolated an ENU mutant mouse line that displayed a lowered frequency of Ly49A expression. We provide evidence that the *c-myb*^{D100G} allele is the causative lesion in this mouse. The association between c-Myb and Ly49A expression frequency on NK cells was previously determined by using a heterozygous knockout mouse and provides further support for our hypothesis (Bezman et al. 2011).

Interestingly, the mutant phenotype specifically affected the frequency of Ly49A expression, with slight effects on the other assayed receptors. It is tempting to speculate that this is the result of a direct regulatory interaction, although further experimentation will be necessary to determine this. At the time of writing, no widely used anti-c-Myb antibodies that are suitable for ChIP-seq or CUT&RUN experiments are available. The generation of such an antibody will greatly assist in answering whether c-Myb directly regulates the *Ly49* genes. The c-Myb transcription factor is not highly expressed in NK cells (data not shown) but is expressed in lymphoid and earlier hematopoietic progenitors. This suggests that the effect that c-Myb exerts on the *Ly49a* locus likely takes place during the differentiation of NK cells, supporting a model where the decision to activate an RME NK receptor gene may be temporally decoupled from the regulation of the gene in adult NK cells.

The fact that the *c-myb*^{D100G} mutation can only exist in a heterozygous conformation provides further clues about the nature of the regulatory events involved in the generation of variegated expression. Importantly, the cells that express the Ly49A receptor in mutant mice express a comparable amount per cell. All of our evidence points toward the amorphic nature of the *c-myb*^{D100G} allele, suggesting the dominant phenotype of our mutant is due to the haploinsufficiency of the heterozygous conformation to drive *Ly49a* expression in the normal 17% of NK cells. Analog TF dosage, then, is likely to be read out as digital (the decision to express or not). Other transcription factors have been suggested to regulate the frequency of Ly49 receptor expression in a similar dose-dependent fashion, such as TCF-1 (Ioannidis et al. 2003) and Runx3 (Ohno et al. 2008). Importantly, our results—and those with other TFs—are most consistent with a mechanism where variegation is resultant from limiting activation from *cis* and *trans* regulators. As this activation signal varies quantitatively, it alters the probability with which activation is achieved which is a binary output. At least in the case of c-Myb, the *trans* regulatory input is important only in the activation, but not the maintenance of expression in mature cells.

Chapter 7
Concluding Remarks

Concluding Remarks

This thesis began with a simple question: how is the variegated and monoallelic expression program of the NK cell receptors enacted? This pattern is critical to NK cell function and was originally defined in large part by our lab group (Held, Roland, and Raulet 1995; Held and Raulet 1997). Since then, models to explain the expression pattern have been put forward based in work in cell lines (Saleh et al. 2004), and *in vivo* work was limited to the generation of transgenic mice (Tanamachi et al. 2004). Important headway was made, but a mechanistic model of variegated NK cell receptor expression was lacking. Since these studies were carried out, the variegated expression pattern displayed by the NK cell receptors was realized to be an example of a much broader phenomenon (RME) (Gimelbrant et al. 2007; Gendrel et al. 2014; Gendrel et al. 2016; Xu et al. 2017; Eckersley-Maslin and Spector 2014; Eckersley-Maslin et al. 2014). Work on the phenomenon of RME has been similarly limited for technical reasons, and much of the informative work on RME has been done using clonal cell lines from both humans and F_1 hybrid mice (Eckersley-Maslin et al. 2014; Gendrel et al. 2014; Reinius et al. 2016; Xu et al. 2017).

This thesis establishes the NK receptor genes as a powerful *in vivo* genetic model to study RME. CRISPR/Cas9-based genome editing has allowed precise targeting of regulatory elements, and allele-specific antibodies allow rapid and cheap detection of allelic expression states in primary cells by flow cytometry, circumventing the difficulty of similar experiments in primary cells where allelic expression states can only be analyzed at the RNA level. Furthermore, using the relatively new approaches of ATAC-seq and CUT&RUN with primary NK cells sorted according to allelic expression status with allele-specific antibodies allowed us to generate a detailed description of the chromatin at silent and active alleles.

We find that the constitutively accessible and activated enhancers of the *Ly49a* and *Nkg2a* genes control expression of the downstream allele *in cis*, precluding a model where the apparently stochastic expression of alleles is regulated by an RME-specific epigenetic program the coordinates allelic expression. The $Ly49g_{Hss5}$ and $Nkg2d_{5'E}$ enhancer elements were found to directly contribute to the probabilistic expression of their target genes. The deletion of $Nkg2d_{5'E}$ resulted in the *de novo* generation of a variegated and monoallelically expressed NK receptor gene, strongly supporting the notion that stable RME is a consequence of limiting enhancer action rather than a dedicated program. This result links the field of RME to earlier work showing that enhancer deletion results in variegated expression in various systems. (Ronai, Berru, and Shulman 1999; Fiering, Whitelaw, and Martin 2000; Garefalaki et al. 2002; Ellmeier et al. 2002). Enhancers appear to function in an intrinsically probabilistic manner, largely controlling the probability of target gene expression rather than expression per cell. As a broad phenomenon, RME seems to be the result of general properties of gene expression. This is consistent with the recent observations (along with our own in Chapter 4) that enhancer elements near RME loci are constitutively accessible. Activation of enhancers is likely a deterministic phenomenon; within a given *trans* nuclear environment both copies of an enhancer are always activated. When enhancer action at a locus is limiting and the probability of expression is much less than 100%, RME is the observed result.

Perhaps the most consequential result in this thesis is the finding that genes previously thought to be ubiquitously expressed by given cells (*Ptprc/Cd45*, *Cd8a*, *Thy1* and *Klrk1/Nkg2d*) are expressed in a mitotically stable RME fashion. These experiments were borne out of the hypothesis that probabilistic enhancer action and consequent RME are rooted in a generalized mechanism of gene expression and therefore should apply at all gene loci. Strikingly, 4/4 genes tested displayed an RME expression pattern. Our ability to deduce this pattern relied on the existence of allele-specific antibodies and F_1 hybrid genetics, which allowed us to analyze millions of primary cells with high precision. Previous studies of RME were limited to the use of manageable numbers of cell clones, greatly limiting their resolution and ability to detect RME expression in genes with very low allelic failure rates. We propose that RME is a ubiquitous feature of gene expression and does not require a dedicated epigenetic mechanism. Rather, mitotic stability of expression states is likely mediated by the same epigenetic programs that confer the stability of active and inactive expression states broadly. Enhancer action simply toggles the probability that a gene will fall into either the active or inactive state (as described in Chapter 4), and both states are maintained.

The answer to the original question posed in this thesis is simultaneously mundane and shocking. At the outset of investigation into RME, the pattern appeared so striking and analogous to X-inactivation that it was reasonably assumed that there must be very specific regulatory mechanisms that generate RME. Instead, the findings in this thesis suggest that RME is a consequence of limiting enhancer action. While this appears to be the driving mechanism behind RME, it also gives the distinct impression of a *lack* of a mechanism. Importantly, the phenomenon of RME was predicted as a consequence of probabilistic enhancer action over 20 years ago (Fiering, Whitelaw, and Martin 2000). RME was not yet observed when this model was put forward. The probabilistic nature of enhancer action was, at that point in time, based on transfected reporter assays in cell lines, and subsequent work extended these findings to endogenous loci in B cell hybridomas (Ronai, Berru, and Shulman 1999) and *in vivo* (Garefalaki et al. 2002; Ellmeier et al. 2002; De Gobbi et al. 2017; Khan, Vaes, and Mombaerts 2011). The work in this thesis places this collective work in the context of the natural phenomenon it predicts, RME, and finally extends that phenomenon to gene expression broadly.

Further questions remain to be addressed. The epigenetic mechanisms driving stability of active and inactive gene expression states are of broad interest. RME of the NK receptor genes, and other genes expressed by immune cells in a manner easily measured by flow cytometry, provide a unique opportunity to study these stable states. One hypothesis is that active promoters of both RME and ubiquitously expressed genes (an artificial distinction) are maintained as nucleosome-depleted through mitosis, a model known as mitotic bookmarking (Xu et al. 2017). The RME genes studied and identified in this thesis could provide a powerful model for the study of the bookmarking model in primary cells and in an allele-specific manner.

Even more speculatively, the ubiquitous nature of RME must have consequences for natural biology. Expression of lineage-determining factors could be controlled, in some instances, by RME. Indeed, it was recently found that expression of the T cell determining gene *Bcl11b*, is regulated in a manner highly reminiscent of RME, and that allelic activation of this factor (and consequent commitment of thymocytes to the T cell fate) is kinetically delayed in mice where a distal *Bcl11b* enhancer is deleted (Ng et al. 2018). While it is our view that

pervasive RME is generally not deleterious, it is conceivable that mutation in an enhancer element can lead to a reduction in the proportion of cells producing a particular factor by a given tissue, such as a hormone. Or, some cells may express only a mutant allele, abrogating their cellular function or preventing their development and survival. This may constitute a novel mechanism of genetic haploinsufficiency. Finally, in rare instances, probabilistic enhancer driven RME allows the generation of diversity within otherwise identical cells. One such example is the generation of the functional NK cell receptor repertoire.

References

Alexandrov, L. B., and M. R. Stratton. 2014. 'Mutational signatures: the patterns of somatic mutations hidden in cancer genomes', *Curr Opin Genet Dev*, 24: 52-60.

Barrow, A. D., C. J. Martin, and M. Colonna. 2019. 'The Natural Cytotoxicity Receptors in Health and Disease', *Front Immunol*, 10: 909.

Bartman, C. R., S. C. Hsu, C. C. Hsiung, A. Raj, and G. A. Blobel. 2016. 'Enhancer Regulation of Transcriptional Bursting Parameters Revealed by Forced Chromatin Looping', *Mol Cell*, 62: 237-47.

Bauer, S., V. Groh, J. Wu, A. Steinle, J. H. Phillips, L. L. Lanier, and T. Spies. 1999. 'Activation of NK cells and T cells by NKG2D, a receptor for stress-inducible MICA', *Science*, 285: 727-9.

Bezman, N. A., T. Chakraborty, T. Bender, and L. L. Lanier. 2011. 'miR-150 regulates the development of NK and iNKT cells', *J Exp Med*, 208: 2717-31.

Bix, M., and R. M. Locksley. 1998. 'Independent and epigenetic regulation of the interleukin-4 alleles in CD4+ T cells', *Science*, 281: 1352-4.

Blackwood, E. M., and J. T. Kadonaga. 1998. 'Going the distance: a current view of enhancer action', *Science*, 281: 60-3.

Buenrostro, J. D., P. G. Giresi, L. C. Zaba, H. Y. Chang, and W. J. Greenleaf. 2013. 'Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position', *Nat Methods*, 10: 1213-8.

Calo, E., and J. Wysocka. 2013. 'Modification of enhancer chromatin: what, how, and why?', *Mol Cell*, 49: 825-37.

Canzio, D., C. L. Nwakeze, A. Horta, S. M. Rajkumar, E. L. Coffey, E. E. Duffy, R. Duffie, K. Monahan, S. O'Keeffe, M. D. Simon, S. Lomvardas, and T. Maniatis. 2019. 'Antisense lncRNA Transcription Mediates DNA Demethylation to Drive Stochastic Protocadherin alpha Promoter Choice', *Cell*, 177: 639-53 e15.

Cerwenka, A., and L. L. Lanier. 2001. 'Natural killer cells, viruses and cancer', *Nat Rev Immunol*, 1: 41-9.

Chan, H. W., J. S. Miller, M. B. Moore, and C. T. Lutz. 2005. 'Epigenetic control of highly homologous killer Ig-like receptor gene alleles', *J Immunol*, 175: 5966-74.

Chess, A. 2012. 'Mechanisms and consequences of widespread random monoallelic expression', *Nat Rev Genet*, 13: 421-8.

Chess, A., I. Simon, H. Cedar, and R. Axel. 1994. 'Allelic inactivation regulates olfactory receptor gene expression', *Cell*, 78: 823-34.

Cichocki, F., T. Lenvik, N. Sharma, G. Yun, S. K. Anderson, and J. S. Miller. 2010. 'Cutting edge: KIR antisense transcripts are processed into a 28-base PIWI-like RNA in human NK cells', *J Immunol*, 185: 2009-12.

Cordes, S. P. 2005. 'N-ethyl-N-nitrosourea mutagenesis: boarding the mouse mutant express', *Microbiol Mol Biol Rev*, 69: 426-39.

Davies, G. E., S. M. Locke, P. W. Wright, H. Li, R. J. Hanson, J. S. Miller, and S. K. Anderson. 2007. 'Identification of bidirectional promoters in the human KIR genes', *Genes Immun*, 8: 245-53.

De Gobbi, M., A. J. Brazel, J. A. Sharpe, J. A. Sloane-Stanley, A. J. Smith, W. G. Wood, and D. Vernimmen. 2017. 'Enhancer deletion generates cellular phenotypic diversity due to bimodal gene expression', *Blood Cells Mol Dis*, 64: 10-12.

Deng, Q., D. Ramskold, B. Reinius, and R. Sandberg. 2014. 'Single-cell RNA-seq reveals dynamic, random monoallelic gene expression in mammalian cells', *Science*, 343: 193-6.
Deng, W., B. G. Gowen, L. Zhang, L. Wang, S. Lau, A. Iannello, J. Xu, T. L. Rovis, N. Xiong, and D. H. Raulet. 2015. 'Antitumor immunity. A shed NKG2D ligand that promotes natural killer cell activation and tumor rejection', *Science*, 348: 136-9.
Deng, W., J. Lee, H. Wang, J. Miller, A. Reik, P. D. Gregory, A. Dean, and G. A. Blobel. 2012. 'Controlling long-range genomic interactions at a native locus by targeted tethering of a looping factor', *Cell*, 149: 1233-44.
Eckersley-Maslin, M. A., and D. L. Spector. 2014. 'Random monoallelic expression: regulating gene expression one allele at a time', *Trends Genet*, 30: 237-44.
Eckersley-Maslin, M. A., D. Thybert, J. H. Bergmann, J. C. Marioni, P. Flicek, and D. L. Spector. 2014. 'Random monoallelic gene expression increases upon embryonic stem cell differentiation', *Dev Cell*, 28: 351-65.
Ellmeier, W., M. J. Sunshine, R. Maschek, and D. R. Littman. 2002. 'Combined deletion of CD8 locus cis-regulatory elements affects initiation but not maintenance of CD8 expression', *Immunity*, 16: 623-34.
Ernst, J., and M. Kellis. 2012. 'ChromHMM: automating chromatin-state discovery and characterization', *Nat Methods*, 9: 215-6.
Esumi, S., N. Kakazu, Y. Taguchi, T. Hirayama, A. Sasaki, T. Hirabayashi, T. Koide, T. Kitsukawa, S. Hamada, and T. Yagi. 2005. 'Monoallelic yet combinatorial expression of variable exons of the protocadherin-alpha gene cluster in single neurons', *Nat Genet*, 37: 171-6.
Fernandez, N. C., E. Treiner, R. E. Vance, A. M. Jamieson, S. Lemieux, and D. H. Raulet. 2005. 'A subset of natural killer cells achieves self-tolerance without expressing inhibitory receptors specific for self-MHC molecules', *Blood*, 105: 4416-23.
Fiering, S., E. Whitelaw, and D. I. Martin. 2000. 'To be or not to be active: the stochastic nature of enhancer action', *Bioessays*, 22: 381-7.
Flajnik, M. F., and M. Kasahara. 2010. 'Origin and evolution of the adaptive immune system: genetic events and selective pressures', *Nat Rev Genet*, 11: 47-59.
Fukaya, T., B. Lim, and M. Levine. 2016. 'Enhancer Control of Transcriptional Bursting', *Cell*, 166: 358-68.
Garefalaki, A., M. Coles, S. Hirschberg, G. Mavria, T. Norton, A. Hostert, and D. Kioussis. 2002. 'Variegated expression of CD8 alpha resulting from in situ deletion of regulatory sequences', *Immunity*, 16: 635-47.
Garrido, F., N. Aptsiauri, E. M. Doorduijn, A. M. Garcia Lora, and T. van Hall. 2016. 'The urgent need to recover MHC class I in cancers for effective immunotherapy', *Curr Opin Immunol*, 39: 44-51.
Gasser, S., S. Orsulic, E. J. Brown, and D. H. Raulet. 2005. 'The DNA damage pathway regulates innate immune system ligands of the NKG2D receptor', *Nature*, 436: 1186-90.
Gays, F., S. Taha, and C. G. Brooks. 2015. 'The distal upstream promoter in Ly49 genes, Pro1, is active in mature NK cells and T cells, does not require TATA boxes, and displays enhancer activity', *J Immunol*, 194: 6068-81.
Gendrel, A. V., M. Attia, C. J. Chen, P. Diabangouaya, N. Servant, E. Barillot, and E. Heard. 2014. 'Developmental dynamics and disease potential of random monoallelic gene expression', *Dev Cell*, 28: 366-80.

Gendrel, A. V., L. Marion-Poll, K. Katoh, and E. Heard. 2016. 'Random monoallelic expression of genes on autosomes: Parallels with X-chromosome inactivation', *Semin Cell Dev Biol*, 56: 100-10.

Gimelbrant, A., J. N. Hutchinson, B. R. Thompson, and A. Chess. 2007. 'Widespread monoallelic expression on human autosomes', *Science*, 318: 1136-40.

Gordon, S. M., J. Chaix, L. J. Rupp, J. Wu, S. Madera, J. C. Sun, T. Lindsten, and S. L. Reiner. 2012. 'The transcription factors T-bet and Eomes control key checkpoints of natural killer cell maturation', *Immunity*, 36: 55-67.

Gowen, B. G., B. Chim, C. D. Marceau, T. T. Greene, P. Burr, J. R. Gonzalez, C. R. Hesser, P. A. Dietzen, T. Russell, A. Iannello, L. Coscoy, C. L. Sentman, J. E. Carette, S. A. Muljo, and D. H. Raulet. 2015. 'A forward genetic screen reveals novel independent regulators of ULBP1, an activating ligand for natural killer cells', *Elife*, 4.

Gregg, C. 2017. 'The emerging landscape of in vitro and in vivo epigenetic allelic effects', *F1000Res*, 6: 2108.

Hanke, T., H. Takizawa, C. W. McMahon, D. H. Busch, E. G. Pamer, J. D. Miller, J. D. Altman, Y. Liu, D. Cado, F. A. Lemonnier, P. J. Bjorkman, and D. H. Raulet. 1999. 'Direct assessment of MHC class I binding by seven Ly49 inhibitory NK cell receptors', *Immunity*, 11: 67-77.

Harland, K. L., E. B. Day, S. H. Apte, B. E. Russ, P. C. Doherty, S. J. Turner, and A. Kelso. 2014. 'Epigenetic plasticity of Cd8a locus during CD8(+) T-cell development and effector differentiation and reprogramming', *Nat Commun*, 5: 3547.

Harly, C., M. Cam, J. Kaye, and A. Bhandoola. 2018. 'Development and differentiation of early innate lymphoid progenitors', *J Exp Med*, 215: 249-62.

Held, W., B. Kunz, B. Lowin-Kropf, M. van de Wetering, and H. Clevers. 1999. 'Clonal acquisition of the Ly49A NK cell receptor is dependent on the trans-acting factor TCF-1', *Immunity*, 11: 433-42.

Held, W., and D. H. Raulet. 1997. 'Expression of the Ly49A gene in murine natural killer cell clones is predominantly but not exclusively mono-allelic', *Eur J Immunol*, 27: 2876-84.

Held, W., J. Roland, and D. H. Raulet. 1995. 'Allelic exclusion of Ly49-family genes encoding class I MHC-specific receptors on NK cells', *Nature*, 376: 355-8.

Herberman, R. B., M. E. Nunn, H. T. Holden, and D. H. Lavrin. 1975. 'Natural cytotoxic reactivity of mouse lymphoid cells against syngeneic and allogeneic tumors. II. Characterization of effector cells', *Int J Cancer*, 16: 230-9.

Hobert, O. 2010. 'Gene regulation: enhancers stepping out of the shadow', *Curr Biol*, 20: R697-9.

Ioannidis, V., B. Kunz, D. M. Tanamachi, L. Scarpellino, and W. Held. 2003. 'Initiation and limitation of Ly-49A NK cell receptor acquisition by T cell factor-1', *J Immunol*, 171: 769-75.

Jeevan-Raj, B., J. Gehrig, M. Charmoy, V. Chennupati, C. Grandclement, P. Angelino, M. Delorenzi, and W. Held. 2017. 'The Transcription Factor Tcf1 Contributes to Normal NK Cell Development and Function by Limiting the Expression of Granzymes', *Cell Rep*, 20: 613-26.

Jiang, Y., Y. E. Loh, P. Rajarajan, T. Hirayama, W. Liao, B. S. Kassim, B. Javidfar, B. J. Hartley, L. Kleofas, R. B. Park, B. Labonte, S. M. Ho, S. Chandrasekaran, C. Do, B. R. Ramirez, C. J. Peter, W. Jt C, B. M. Safaie, H. Morishita, P. Roussos, E. J. Nestler, A. Schaefer, B. Tycko,

K. J. Brennand, T. Yagi, L. Shen, and S. Akbarian. 2017. 'The methyltransferase SETDB1 regulates a large neuron-specific topological chromatin domain', *Nat Genet*, 49: 1239-50.

Joncker, N. T., N. C. Fernandez, E. Treiner, E. Vivier, and D. H. Raulet. 2009. 'NK cell responsiveness is tuned commensurate with the number of inhibitory receptors for self-MHC class I: the rheostat model', *J Immunol*, 182: 4572-80.

Jung, H., B. Hsiung, K. Pestal, E. Procyk, and D. H. Raulet. 2012. 'RAE-1 ligands for the NKG2D receptor are regulated by E2F transcription factors, which control cell cycle entry', *J Exp Med*, 209: 2409-22.

Karre, K., H. G. Ljunggren, G. Piontek, and R. Kiessling. 1986. 'Selective rejection of H-2-deficient lymphoma variants suggests alternative immune defence strategy', *Nature*, 319: 675-8.

Kelley, J., L. Walter, and J. Trowsdale. 2005. 'Comparative genomics of major histocompatibility complexes', *Immunogenetics*, 56: 683-95.

Kelly, B. L., and R. M. Locksley. 2000. 'Coordinate regulation of the IL-4, IL-13, and IL-5 cytokine cluster in Th2 clones revealed by allelic expression patterns', *J Immunol*, 165: 2982-6.

Khamlichi, A. A., and R. Feil. 2018. 'Parallels between Mammalian Mechanisms of Monoallelic Gene Expression', *Trends Genet*, 34: 954-71.

Khan, M., E. Vaes, and P. Mombaerts. 2011. 'Regulation of the probability of mouse odorant receptor gene choice', *Cell*, 147: 907-21.

Kim, J. K., A. A. Kolodziejczyk, T. Ilicic, S. A. Teichmann, and J. C. Marioni. 2015. 'Characterizing noise structure in single-cell RNA-seq distinguishes genuine from technical stochastic allelic expression', *Nat Commun*, 6: 8687.

Kirkham, C. L., and J. R. Carlyle. 2014. 'Complexity and Diversity of the NKR-P1:Clr (Klrb1:Clec2) Recognition Systems', *Front Immunol*, 5: 214.

Krueger, F., and S. R. Andrews. 2016. 'SNPsplit: Allele-specific splitting of alignments between genomes with known SNP genotypes', *F1000Res*, 5: 1479.

Ladendorff, N. E., S. Wu, and J. S. Lipsick. 2001. 'BS69, an adenovirus E1A-associated protein, inhibits the transcriptional activity of c-Myb', *Oncogene*, 20: 125-32.

Lara-Astiaso, D., A. Weiner, E. Lorenzo-Vivas, I. Zaretsky, D. A. Jaitin, E. David, H. Keren-Shaul, A. Mildner, D. Winter, S. Jung, N. Friedman, and I. Amit. 2014. 'Immunogenetics. Chromatin state dynamics during blood formation', *Science*, 345: 943-9.

Larsson, A. J. M., P. Johnsson, M. Hagemann-Jensen, L. Hartmanis, O. R. Faridani, B. Reinius, A. Segerstolpe, C. M. Rivera, B. Ren, and R. Sandberg. 2019. 'Genomic encoding of transcriptional burst kinetics', *Nature*, 565: 251-54.

Levanon, D., V. Negreanu, J. Lotem, K. R. Bone, O. Brenner, D. Leshkowitz, and Y. Groner. 2014. 'Transcription factor Runx3 regulates interleukin-15-dependent natural killer cell activation', *Mol Cell Biol*, 34: 1158-69.

Levin-Klein, R., S. Fraenkel, M. Lichtenstein, L. S. Matheson, O. Bartok, Y. Nevo, S. Kadener, A. E. Corcoran, H. Cedar, and Y. Bergman. 2017. 'Clonally stable Vkappa allelic choice instructs Igkappa repertoire', *Nat Commun*, 8: 15575.

Magklara, A., A. Yen, B. M. Colquitt, E. J. Clowney, W. Allen, E. Markenscoff-Papadimitriou, Z. A. Evans, P. Kheradpour, G. Mountoufaris, C. Carey, G. Barnea, M. Kellis, and S. Lomvardas. 2011. 'An epigenetic signature for monoallelic olfactory receptor expression', *Cell*, 145: 555-70.

McCullen, M. V., H. Li, M. Cam, S. K. Sen, D. W. McVicar, and S. K. Anderson. 2016. 'Analysis of Ly49 gene transcripts in mature NK cells supports a role for the Pro1 element in gene activation, not gene expression', *Genes Immun*, 17: 349-57.

Modzelewski, A. J., S. Chen, B. J. Willis, K. C. K. Lloyd, J. A. Wood, and L. He. 2018. 'Efficient mouse genome engineering by CRISPR-EZ technology', *Nat Protoc*, 13: 1253-74.

Monahan, K., and S. Lomvardas. 2015. 'Monoallelic expression of olfactory receptors', *Annu Rev Cell Dev Biol*, 31: 721-40.

Moretta, A., C. Bottino, D. Pende, G. Tripodi, G. Tambussi, O. Viale, A. Orengo, M. Barbaresi, A. Merli, E. Ciccone, and et al. 1990. 'Identification of four subsets of human CD3-CD16+ natural killer (NK) cells by the expression of clonally distributed functional surface molecules: correlation between subset assignment of NK clones and ability to mediate specific alloantigen recognition', *J Exp Med*, 172: 1589-98.

Mostoslavsky, R., N. Singh, T. Tenzen, M. Goldmit, C. Gabay, S. Elizur, P. Qi, B. E. Reubinoff, A. Chess, H. Cedar, and Y. Bergman. 2001. 'Asynchronous replication and allelic exclusion in the immune system', *Nature*, 414: 221-5.

Mucenski, M. L., K. McLain, A. B. Kier, S. H. Swerdlow, C. M. Schreiner, T. A. Miller, D. W. Pietryga, W. J. Scott, Jr., and S. S. Potter. 1991. 'A functional c-myb gene is required for normal murine fetal hepatic hematopoiesis', *Cell*, 65: 677-89.

Nag, A., V. Savova, H. L. Fung, A. Miron, G. C. Yuan, K. Zhang, and A. A. Gimelbrant. 2013. 'Chromatin signature of widespread monoallelic expression', *Elife*, 2: e01256.

Ng, K. K., M. A. Yui, A. Mehta, S. Siu, B. Irwin, S. Pease, S. Hirose, M. B. Elowitz, E. V. Rothenberg, and H. Y. Kueh. 2018. 'A stochastic epigenetic switch controls the dynamics of T-cell lineage commitment', *Elife*, 7.

Nice, T. J., L. Coscoy, and D. H. Raulet. 2009. 'Posttranslational regulation of the NKG2D ligand Mult1 in response to cell stress', *J Exp Med*, 206: 287-98.

Ohno, S., T. Sato, K. Kohu, K. Takeda, K. Okumura, M. Satake, and S. Habu. 2008. 'Runx proteins are involved in regulation of CD122, Ly49 family and IFN-gamma expression during NK cell differentiation', *Int Immunol*, 20: 71-9.

Pende, D., M. Falco, M. Vitale, C. Cantoni, C. Vitale, E. Munari, A. Bertaina, F. Moretta, G. Del Zotto, G. Pietra, M. C. Mingari, F. Locatelli, and L. Moretta. 2019. 'Killer Ig-Like Receptors (KIRs): Their Role in NK Cell Modulation and Developments Leading to Their Clinical Exploitation', *Front Immunol*, 10: 1179.

Perry, M. W., A. N. Boettiger, and M. Levine. 2011. 'Multiple enhancers ensure precision of gap gene-expression patterns in the Drosophila embryo', *Proc Natl Acad Sci U S A*, 108: 13570-5.

Prager, I., C. Liesche, H. van Ooijen, D. Urlaub, Q. Verron, N. Sandstrom, F. Fasbender, M. Claus, R. Eils, J. Beaudouin, B. Onfelt, and C. Watzl. 2019. 'NK cells switch from granzyme B to death receptor-mediated cytotoxicity during serial killing', *The Journal of experimental medicine*, 216: 2113-27.

Quadros, R. M., H. Miura, D. W. Harms, H. Akatsuka, T. Sato, T. Aida, R. Redder, G. P. Richardson, Y. Inagaki, D. Sakai, S. M. Buckley, P. Seshacharyulu, S. K. Batra, M. A. Behlke, S. A. Zeiner, A. M. Jacobi, Y. Izu, W. B. Thoreson, L. D. Urness, S. L. Mansour, M. Ohtsuka, and C. B. Gurumurthy. 2017. 'Easi-CRISPR: a robust method for one-step

generation of mice carrying conditional and insertion alleles using long ssDNA donors and CRISPR ribonucleoproteins', *Genome Biol*, 18: 92.
Radwan, J., W. Babik, J. Kaufman, T. L. Lenz, and J. Winternitz. 2020. 'Advances in the Evolutionary Understanding of MHC Polymorphism', *Trends Genet*, 36: 298-311.
Rahim, M. M., M. M. Tu, A. B. Mahmoud, A. Wight, E. Abou-Samra, P. D. Lima, and A. P. Makrigiannis. 2014. 'Ly49 receptors: innate and adaptive immune paradigms', *Front Immunol*, 5: 145.
Raulet, D. H., W. Held, I. Correa, J. R. Dorfman, M. F. Wu, and L. Corral. 1997. 'Specificity, tolerance and developmental regulation of natural killer cells defined by expression of class I-specific Ly49 receptors', *Immunol Rev*, 155: 41-52.
Raulet, D. H., and R. E. Vance. 2006. 'Self-tolerance of natural killer cells', *Nat Rev Immunol*, 6: 520-31.
Raulet, D. H., R. E. Vance, and C. W. McMahon. 2001. 'Regulation of the natural killer cell receptor repertoire', *Annu Rev Immunol*, 19: 291-330.
Reinius, B., J. E. Mold, D. Ramskold, Q. Deng, P. Johnsson, J. Michaelsson, J. Frisen, and R. Sandberg. 2016. 'Analysis of allelic expression patterns in clonal somatic cells by single-cell RNA-seq', *Nat Genet*, 48: 1430-35.
Reinius, B., and R. Sandberg. 2015. 'Random monoallelic expression of autosomal genes: stochastic transcription and allele-level regulation', *Nat Rev Genet*, 16: 653-64.
———. 2018. 'Reply to 'High prevalence of clonal monoallelic expression'', *Nat Genet*, 50: 1199-200.
Riviere, I., M. J. Sunshine, and D. R. Littman. 1998. 'Regulation of IL-4 expression by activation of individual alleles', *Immunity*, 9: 217-28.
Rogers, S. L., A. Rouhi, F. Takei, and D. L. Mager. 2006. 'A role for DNA hypomethylation and histone acetylation in maintaining allele-specific expression of mouse NKG2A in developing and mature NK cells', *J Immunol*, 177: 414-21.
Ronai, D., M. Berru, and M. J. Shulman. 1999. 'Variegated expression of the endogenous immunoglobulin heavy-chain gene in the absence of the intronic locus control region', *Mol Cell Biol*, 19: 7031-40.
———. 2002. 'Positive and negative transcriptional states of a variegating immunoglobulin heavy chain (IgH) locus are maintained by a cis-acting epigenetic mechanism', *J Immunol*, 169: 6919-27.
———. 2004. 'The epigenetic stability of the locus control region-deficient IgH locus in mouse hybridoma cells is a clonally varying, heritable feature', *Genetics*, 167: 411-21.
Roth, T. L., C. Puig-Saus, R. Yu, E. Shifrut, J. Carnevale, P. J. Li, J. Hiatt, J. Saco, P. Krystofinski, H. Li, V. Tobin, D. N. Nguyen, M. R. Lee, A. L. Putnam, A. L. Ferris, J. W. Chen, J. N. Schickel, L. Pellerin, D. Carmody, G. Alkorta-Aranburu, D. Del Gaudio, H. Matsumoto, M. Morell, Y. Mao, M. Cho, R. M. Quadros, C. B. Gurumurthy, B. Smith, M. Haugwitz, S. H. Hughes, J. S. Weissman, K. Schumann, J. H. Esensten, A. P. May, A. Ashworth, G. M. Kupfer, S. A. W. Greeley, R. Bacchetta, E. Meffre, M. G. Roncarolo, N. Romberg, K. C. Herold, A. Ribas, M. D. Leonetti, and A. Marson. 2018. 'Reprogramming human T cell function and specificity with non-viral genome targeting', *Nature*, 559: 405-09.
Rouhi, A., C. G. Brooks, F. Takei, and D. L. Mager. 2007. 'Plasticity of Ly49g expression is due to epigenetics', *Mol Immunol*, 44: 821-6.

Rouhi, A., L. Gagnier, F. Takei, and D. L. Mager. 2006. 'Evidence for epigenetic maintenance of Ly49a monoallelic gene expression', *J Immunol*, 176: 2991-9.
Rouhi, A., C. B. Lai, T. P. Cheng, F. Takei, W. M. Yokoyama, and D. L. Mager. 2009. 'Evidence for high bi-allelic expression of activating Ly49 receptors', *Nucleic Acids Res*, 37: 5331-42.
Rv, P., A. Sundaresh, M. Karunyaa, A. Arun, and S. Gayen. 2021. 'Autosomal Clonal Monoallelic Expression: Natural or Artifactual?', *Trends Genet*, 37: 206-11.
Saleh, A., G. E. Davies, V. Pascal, P. W. Wright, D. L. Hodge, E. H. Cho, S. J. Lockett, M. Abshari, and S. K. Anderson. 2004. 'Identification of probabilistic transcriptional switches in the Ly49 gene cluster: a eukaryotic mechanism for selective gene activation', *Immunity*, 21: 55-66.
Saleh, A., A. P. Makrigiannis, D. L. Hodge, and S. K. Anderson. 2002. 'Identification of a novel Ly49 promoter that is active in bone marrow and fetal thymus', *J Immunol*, 168: 5163-9.
Santourlidis, S., N. Graffmann, J. Christ, and M. Uhrberg. 2008. 'Lineage-specific transition of histone signatures in the killer cell Ig-like receptor locus from hematopoietic progenitor to NK cells', *J Immunol*, 180: 418-25.
Saxton, D. S., and J. Rine. 2019. 'Epigenetic memory independent of symmetric histone inheritance', *Elife*, 8.
Schenkel, A. R., L. C. Kingry, and R. A. Slayden. 2013. 'The ly49 gene family. A brief guide to the nomenclature, genetics, and role in intracellular infection', *Front Immunol*, 4: 90.
Schlissel, G., and J. Rine. 2019. 'The nucleosome core particle remembers its position through DNA replication and RNA transcription', *Proc Natl Acad Sci U S A*, 116: 20605-11.
Sciume, G., Y. Mikami, D. Jankovic, H. Nagashima, A. V. Villarino, T. Morrison, C. Yao, S. Signorella, H. W. Sun, S. R. Brooks, D. Fang, V. Sartorelli, S. Nakayamada, K. Hirahara, B. Zitti, F. P. Davis, Y. Kanno, J. J. O'Shea, and H. Y. Shih. 2020. 'Rapid Enhancer Remodeling and Transcription Factor Repurposing Enable High Magnitude Gene Induction upon Acute Activation of NK Cells', *Immunity*, 53: 745-58 e4.
Shifrin, N., D. H. Raulet, and M. Ardolino. 2014. 'NK cell self tolerance, responsiveness and missing self recognition', *Semin Immunol*, 26: 138-44.
Shih, H. Y., G. Sciume, Y. Mikami, L. Guo, H. W. Sun, S. R. Brooks, J. F. Urban, Jr., F. P. Davis, Y. Kanno, and J. J. O'Shea. 2016. 'Developmental Acquisition of Regulomes Underlies Innate Lymphoid Cell Functionality', *Cell*, 165: 1120-33.
Skene, P. J., J. G. Henikoff, and S. Henikoff. 2018. 'Targeted in situ genome-wide profiling with high efficiency for low cell numbers', *Nat Protoc*, 13: 1006-19.
Sleckman, B. P., C. G. Bardon, R. Ferrini, L. Davidson, and F. W. Alt. 1997. 'Function of the TCR alpha enhancer in alphabeta and gammadelta T cells', *Immunity*, 7: 505-15.
Tanamachi, D. M., T. Hanke, H. Takizawa, A. M. Jamieson, and D. R. Raulet. 2001. 'Expression of natural killer receptor alleles at different Ly49 loci occurs independently and is regulated by major histocompatibility complex class I molecules', *J Exp Med*, 193: 307-15.
Tanamachi, D. M., D. C. Moniot, D. Cado, S. D. Liu, J. K. Hsia, and D. H. Raulet. 2004. 'Genomic Ly49A transgenes: basis of variegated Ly49A gene expression and identification of a critical regulatory element', *J Immunol*, 172: 1074-82.
Teves, S. S., L. An, A. S. Hansen, L. Xie, X. Darzacq, and R. Tjian. 2016. 'A dynamic mode of mitotic bookmarking by transcription factors', *Elife*, 5.

Textor, S., N. Fiegler, A. Arnold, A. Porgador, T. G. Hofmann, and A. Cerwenka. 2011. 'Human NK cells are alerted to induction of p53 in cancer cells by upregulation of the NKG2D ligands ULBP1 and ULBP2', *Cancer Res*, 71: 5998-6009.

Thompson, T. W., B. T. Jackson, P. J. Li, J. Wang, A. B. Kim, K. T. H. Huang, L. Zhang, and D. H. Raulet. 2018. 'Tumor-derived CSF-1 induces the NKG2D ligand RAE-1delta on tumor-infiltrating macrophages', *Elife*, 7.

Thompson, T. W., A. B. Kim, P. J. Li, J. Wang, B. T. Jackson, K. T. H. Huang, L. Zhang, and D. H. Raulet. 2017. 'Endothelial cells express NKG2D ligands and desensitize antitumor NK responses', *Elife*, 6.

Thorvaldsdottir, H., J. T. Robinson, and J. P. Mesirov. 2013. 'Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration', *Brief Bioinform*, 14: 178-92.

Valiante, N. M., M. Uhrberg, H. G. Shilling, K. Lienert-Weidenbach, K. L. Arnett, A. D'Andrea, J. H. Phillips, L. L. Lanier, and P. Parham. 1997. 'Functionally and structurally distinct NK cell receptor repertoires in the peripheral blood of two human donors', *Immunity*, 7: 739-51.

Vance, R. E., A. M. Jamieson, D. Cado, and D. H. Raulet. 2002. 'Implications of CD94 deficiency and monoallelic NKG2A expression for natural killer cell development and repertoire formation', *Proc Natl Acad Sci U S A*, 99: 868-73.

Vigneau, S., S. Vinogradova, V. Savova, and A. Gimelbrant. 2018. 'High prevalence of clonal monoallelic expression', *Nat Genet*, 50: 1198-99.

Vivier, E., D. H. Raulet, A. Moretta, M. A. Caligiuri, L. Zitvogel, L. L. Lanier, W. M. Yokoyama, and S. Ugolini. 2011. 'Innate or adaptive immunity? The example of natural killer cells', *Science*, 331: 44-9.

Walters, M. C., S. Fiering, J. Eidemiller, W. Magis, M. Groudine, and D. I. Martin. 1995. 'Enhancers increase the probability but not the level of gene expression', *Proc Natl Acad Sci U S A*, 92: 7125-9.

Walters, M. C., W. Magis, S. Fiering, J. Eidemiller, D. Scalzo, M. Groudine, and D. I. Martin. 1996. 'Transcriptional enhancers act in cis to suppress position-effect variegation', *Genes Dev*, 10: 185-95.

Wang, R., J. J. Jaw, N. C. Stutzman, Z. Zou, and P. D. Sun. 2012. 'Natural killer cell-produced IFN-gamma and TNF-alpha induce target cell cytolysis through up-regulation of ICAM-1', *J Leukoc Biol*, 91: 299-309.

Weintraub, H. 1988. 'Formation of stable transcription complexes as assayed by analysis of individual templates', *Proc Natl Acad Sci U S A*, 85: 5819-23.

Wensveen, F. M., V. Jelencic, and B. Polic. 2018. 'NKG2D: A Master Regulator of Immune Cell Responsiveness', *Front Immunol*, 9: 441.

Xu, J., A. C. Carter, A. V. Gendrel, M. Attia, J. Loftus, W. J. Greenleaf, R. Tibshirani, E. Heard, and H. Y. Chang. 2017. 'Landscape of monoallelic DNA accessibility in mouse embryonic stem cells and neural progenitor cells', *Nat Genet*, 49: 377-86.

Xu, Y., L. Davidson, F. W. Alt, and D. Baltimore. 1996. 'Deletion of the Ig kappa light chain intronic enhancer/matrix attachment region impairs but does not abolish V kappa J kappa rearrangement', *Immunity*, 4: 377-85.

Yokoyama, W. M., L. B. Jacobs, O. Kanagawa, E. M. Shevach, and D. I. Cohen. 1989. 'A murine T lymphocyte antigen belongs to a supergene family of type II integral membrane proteins', *J Immunol*, 143: 1379-86.

Yokoyama, W. M., P. J. Kehn, D. I. Cohen, and E. M. Shevach. 1990. 'Chromosomal location of the Ly-49 (A1, YE1/48) multigene family. Genetic association with the NK 1.1 antigen', *J Immunol*, 145: 2353-8.

Yokoyama, W. M., and W. E. Seaman. 1993. 'The Ly-49 and NKR-P1 gene families encoding lectin-like receptors on natural killer cells: the NK gene complex', *Annu Rev Immunol*, 11: 613-35.

Young, R. S., Y. Kumar, W. A. Bickmore, and M. S. Taylor. 2017. 'Bidirectional transcription initiation marks accessible chromatin and is not specific to enhancers', *Genome Biol*, 18: 242.

Zook, E. C., Z. Y. Li, Y. Xu, R. F. de Pooter, M. Verykokakis, A. Beaulieu, A. Lasorella, M. Maienschein-Cline, J. C. Sun, M. Sigvardsson, and B. L. Kee. 2018. 'Transcription factor ID2 prevents E proteins from enforcing a naive T lymphocyte gene program during NK cell development', *Sci Immunol*, 3.