Endogenous Retroviruses and the Immune System

George Robert Young

Submitted to University College London in part fulfilment of the requirements for the degree of Doctor of Philosophy

2012

Division of Infection and Immunity

University College London
Gower Street
London
WC1E 6BT

Division of Immunoregulation

National Institute for Medical Research
The Ridgeway
London
NW7 1AA

I, George Robert Young, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

Abstract

Initial sequencing of the human and mouse genomes revealed that substantial fractions were composed of retroelements (REs) and endogenous retroviruses (ERVs), the latter being relics of ancestral retroviral infection. Further study revealed ERVs constitute up to 10% of many mammalian genomes. Despite this abundance, comparatively little is known about their interactions, beneficial or detrimental, with the host.

This thesis details two distinct sets of interactions with the immune system. Firstly, the presentation of ERV-derived peptides to developing lymphocytes was shown to exert a control on the immune response to infection with Friend Virus (FV). A self peptide encoded by an ERV negatively selected a significant fraction of polyclonal FV-specific $CD4^+$ T cells and resulted in an impaired immune response. However, $CD4^+$ T cell-mediated antiviral activity was fully preserved and repertoire analysis revealed a deletional bias according to peptide affinity, resulting in an effective enrichment of high-affinity $CD4^+$ T cells. Thus, ERVs exerted a significant influence on the immune response, a mechanism that may partially contribute to the heterogeneity seen in human immune responses to retroviral infections.

Secondly, a requirement for specific antibodies was shown in the control of ERVs. In a range of mice displaying distinct deficiencies in antibody production, products from the intestinal microbiota potentially induce ERV expression. Subsequent recombinational correction of a defective murine leukaemia virus (MLV) results in the emergence of infectious virus. In the long term, this leads to retrovirus-induced lymphomas and morbidity. ERVs, therefore, provide a potential link between the intestinal microbiota and a range of pathologies, including cancer.

Finally, a new computational tool, *REquest*, was developed for use in the above studies. *REquest* allows the mining of retroelement (RE) and ERV expression data from the majority of commercially available human and murine microarray platforms and allows rapid hypothesis testing with publicly available data.

Acknowledgements

I would like to express my thanks to my supervisor, George Kassiotis, without whom none of this work would have been possible. Thank you for having the faith to mentor an Oxford graduate with little but theoretical knowledge, for the discussion and guidance, and for encouraging the tangential research that our progress generated.

Many others within the institute have given useful critique and analysis, specifically Jonathan Stoye and my thesis committee, Kate Bishop and Pavel Tolar. My thanks also to Anne O'Garra and the members of the Division of Immunoregulation.

The National Institute for Medical Research provides an unparalleled support infrastructure, and I am indebted to members of the flow cytometry, electron microscopy, microarray, and large scale laboratory facilities, as well as the animal technicians and staff within and biological services.

Underpinning both projects described here, Ula Eksmond has provided unfailing effort and assistance, all whilst maintaining the smooth-running of the laboratory. I'm extremely appreciative of her contributions to my research.

Over many years and in numerous ways, my parents have put my education and potential above virtually all else. I am deeply grateful for their continued support and am extremely fortunate to be where I am today. Rachel has listened to, and sympathised with, the highs and the lows of experimental research and has provided the necessary motivation at the opportune moments. Now that this is complete, we can hopefully spend some more time together, and with our fantastic little William! He may be too young to understand this work, but perhaps I can read sections to him at bedtime.

Table of Contents

List of Figures

List of Tables

Abbreviations

EIA	Equine infectious anaemia
AB	Air buffered
AIDS	Acquired immune deficiency syndrome
ALV	Avian leukosis virus
APC	Antigen presenting cell
	Allophycocyanin fluorochrome
bAb	Binding antibody
BCR	B cell receptor
BM	Bone marrow
CFSE	Carboxyfluorescein succinimidyl ester
CLP	Common lymphoid progenitor
CMP	Common myeloid progenitor
CMV	Cytomegalovirus
csv	Comma separated value
CVID	Chronic variable immunodeficiency
DC	Dendritic cell
DNA	Deoxyribonucleic acid
EBV	Epstein-Barr Virus
eMLV	Ecotropic MLV
ERV	Endogenous retrovirus
ETn	Early transposon
F-MLV	Friend murine leukaemia virus
FACS	Fluorescence activated cell sorting
FCS	Foetal calf serum
FDR	False discovery rate
FITC	Fluorescein isothiocyanate fluorochrome
FSC	Forward scatter
FV/FV-B	Friend virus
FVA	Anaemia-inducing Friend virus
FVP	Polycythemia-inducing Friend virus
GALT	Gut-associated lymphoid tissue
GF	Germ free
GFP	Green fluorescent protein
GM-CSF	Granulocyte-macrophage colony stimulating factor
HAT	Hypoxanthine-Aminopterin-Thymidine
HIV	Human immunodeficiency virus
HSV	Herpres simplex virus
HT	Hypoxanthine-Thymidin
HTLV	Human T-lymphotropic virus
IAP	Intracisternal A particle

IAV	Influenza A virus
IFN	Interferon
Ig	Immunoglobulin
IL	Interleukin
IMDM	Iscove's modified Dulbecco's medium
iu	Infectious units
IVC	Individually ventilated cage
JSRV	Jaagsiekte sheep retrovirus
LDV	Lactate dehydrogenase-elevating virus
LINE	Long interspersed nuclear element
LPS	Lipopolysaccharide
LTR	Long terminal repeat
MAIDS	Murine AIDS
MHC	Major histocompatibility molecule
MLV	Murine leukaemia virus
MMTV	Mouse mammary tumour virus
mpMLV	Modified polytropic MLV
mRNA	Messenger RNA
mTEC	Medullary thymic epithelial cell
nAb	Neutralising antibody
NKT	Natural killer T cell
ORF	Open reading frame
PAMP	Pathogen-associated molecular patterns
PBS	Phosphate-buffered saline
PCR	Polymerase chain reaction
PE	Phycoerythrin
PMA	Phorbol-12-myristate-13-acetate
pMLV	Polytropic MLV
poly(I:C)	Polyinosine-polycytidylic acid
qRT-PCR	Quantitative real time PCR
RARV	*Rag1*$^{-/-}$ mouse-associated retrovirus
RBC	Red blood cell
RE	Retroelement
RNA	Ribonucleic acid
RSV	Rous sarcoma virus
SCID	Severe combined immunodeficiency
SEM	Standard error of the mean
SFFU	Spleen focus-forming units
SFFV	Spleen focus forming virus
SINE	Short interspersed nuclear element

SPF	Specific pathogen free
TCR	T cell receptor
TLR	Toll-like receptor
TNF	Tumour necrosis factor
tRNA	Transfer RNA
UV	Ultra-violet
XLA	X-linked agammaglobulinemia
xMLV	Xenotropic MLV

Chapter 1

Introduction

1.1 Retroviruses and retroelements

1.1.1 A brief history of retrovirology

Although unappreciated at the time, retroviruses as the causative agents of disease have been studied since the late 1800s, with equine infectious anaemia (EIA) being the first example (Vallée & Carré, 1904). Early work in domestic fowl similarly identified diseases experimentally transmissible through cell-free filtrates (Ellermann & Bang, 1908; Rous, 1910, 1911), and avian leukosis virus (ALV) and the closely-related Rous-sarcoma virus (RSV) were rapidly identified as the aetiologic agents. Somewhat later, cell-free extracts were also shown to cause leukaemias in mice (L. Gross, 1951, 1957; Friend, 1957; Rauscher, 1962).

RSV virions were shown to contain RNA rather than DNA (L. V. Crawford & E. M. Crawford, 1961; Temin, 1963), so called 'RNA tumour viruses', but cell transformation could also be shown to be stable through mitosis – a property not previously associated with RNA viruses. To explain this discrepancy, Howard Temin proposed the 'provirus' hypothesis, where a DNA copy of the viral genome is produced that can integrate into the host chromosomes (Temin, 1964). Whilst a logical explanation of the observations, the hypothesis conflicted with a central tenet of molecular biology at the time – that DNA could be transcribed to RNA, but not *vice versa* – and hence received little immediate attention. This concept only gained acceptance through the independent and parallel discovery of reverse transcriptase, the missing RNA-directed DNA-polymerase, by Temin and Baltimore (Temin & Mizutani, 1970; Baltimore, 1970).

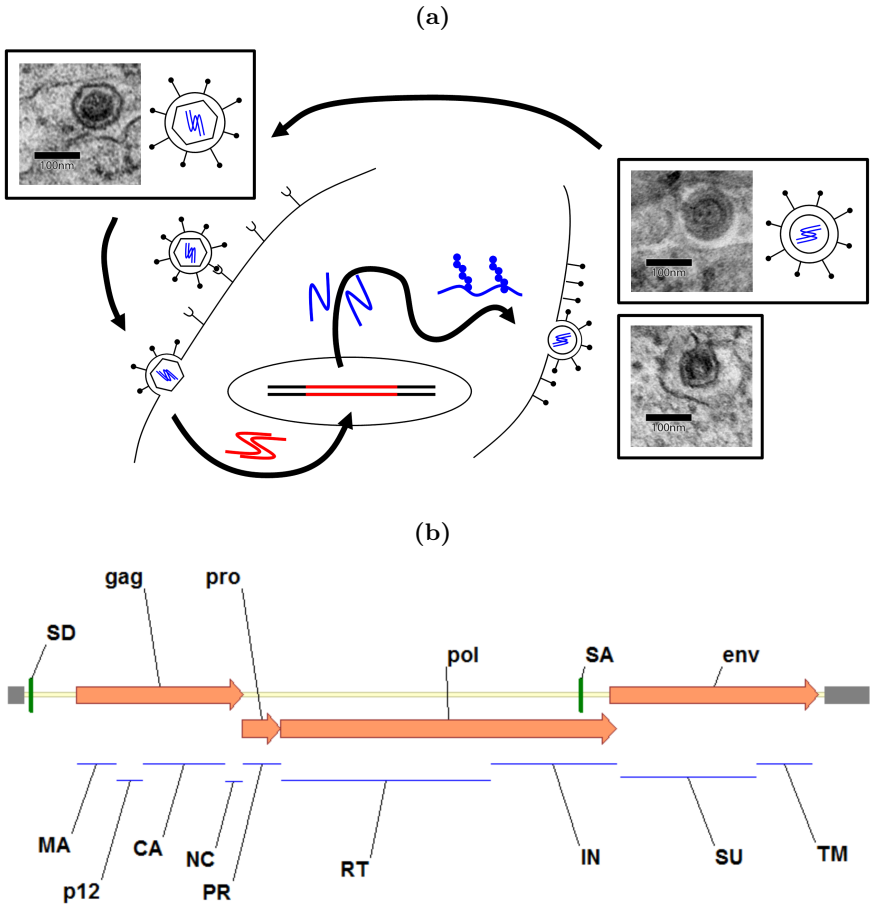
The requirement for stable chromosomal integration of proviruses raised the possibility of their vertical inheritance through germ-line infection. Support for this principle came through spontaneous oncogenesis in uninfected animals (Mühlbock, 1955) and the potential for radiological and chemical induction of infectious virions from untransformed cells (Lieberman & H. S. Kaplan, 1959; W. P. Rowe, Hartley, Lander, Pugh & Teich, 1971; Weiss, Friis, Katz & Vogt, 1971). Inherited in a dominant Mendelian manner, 'endogenous retroviruses' (ERVs) have since been recognised as essentially universal genomic features and have been found in virtually all well-studied eukaryotic genomes, including that of humans (M. A. Martin, Bryan, Rasheedt & Khan, 1981; Repaske, Steele, O'Neill, Rabson & Martin, 1985).

1.1.2 The *Retroviridae*

Whilst there is increasing evidence that a variety of viruses have the potential for stable chromosomal integration (Horie, Honda et al., 2010; Horie & Tomonaga, 2011; Tanaka-Taya et al., 2004; Belyi, Levine & Skalka, 2010), the requirement for a proviral

stage remains the defining feature of retroviral replication (Fig. 1.1a). In this way, retroviruses rely entirely on the transcriptional machinery of the host cell.

Figure 1.1: Retroviral replication and structure



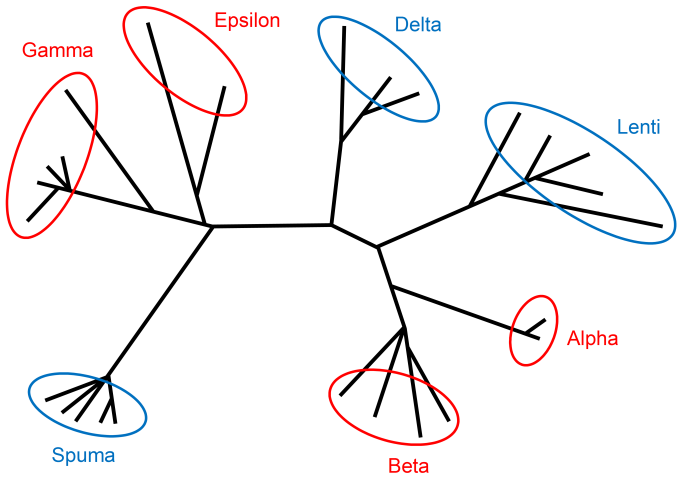
(a) A simplified retroviral 'life cycle' showing receptor binding and entry, reverse transcription and integration, translation of viral mRNA, protein product formation and assembly at the cell membrane. RNA species are shown in blue, DNA species are shown in red. Virions bud from the cell and proteolytic cleavage of proteins leads to condensation of the viral core to the 'mature', electron-dense state. (b) The genome structure of a murine leukaemia virus. Shown are the four coding genes along with the splice donator (SD) and acceptor (SA) for the *env* mRNA product. Corresponding protein products are shown – matrix (MA), p12, capsid (CA), nucleocapsid (NC), protease (PR), reverse transcriptase (RT), integrase (IN), surface (SU), and transmembrane (TM).

Production of proviruses requires the presence of both a viral reverse transcriptase and an integrase, and has a major bearing on the genome structure. Retroviral particles contain dimers of the single-stranded positive-sense linear RNA genome, which is typically 7–11 kb in size. All retroviral genomes contain four main coding sequences (*gag*, *pro*, *pol* and *env*), the order of which is vital to the correct level of gene expression, subsequent cleavage and product formation, and is hence completely conserved amongst known retroviruses (Jern & Coffin, 2008) (Fig. 1.1b).

'Simple' retroviruses, such as murine leukaemia viruses (MLVs) (Fig. 1.1b), first sequenced completely in 1981 (Shinnick, Lerner & Sutcliffe, 1981), code only these four genes, wherease 'complex' retroviruses encode for one or more additional accessory genes. These are generally inserted before or after *env* and may translate into transcriptional regulators (e.g. *tax* in HTLV-1), infectivity modulators (e.g. *vif* in HIV-1), or superantigens (e.g. *sag* in MMTV), to name but a few functions. Whilst there is some evidence of alternate splicing in simple retroviruses (Déjardin et al., 2000; Rasmussen et al., 2010), translation of complex retroviral genomes necessitates multiple splicing events and can initiate from transcription start sites giving both positive- and negative-sense mRNA products.

Traditional classification of retroviral particles relied on the study of morphology in negatively-stained electron micrographs (Bernhard, 1958). Virions were classified into four groups: A-type (non-enveloped intracellular particles with an uncondensed, 'immature', electron-lucent core), B-type (extracellular, enveloped particles with condensed, electron-dense acentric cores and prominent envelope spikes), C-type (extracellular, enveloped particles with condensed, electron-dense central cores and unclear envelope spikes (Fig. 1.1a)) and D-type (slightly larger extracellular particles with condensed, electron-dense, bar-shaped cores and unclear envelope spikes). This system has, however, largely been replaced with placement according to phylogenetic relationships (Fig. 1.2).

Figure 1.2: Retroviral phylogenetic tree



Phylogenetic tree of exogenous retroviruses, with colouring denoting simple (red) or complex (blue) genomes. Modified from Weiss, 2006.

Retroviral infection and hence host tropism relies on the specific recognition of cell surface receptors, and much work has identified the varying genes involved in viral entry in many species. Viral evolution and host escape through receptor modification are active areas of research (Soll, Neil & Bieniasz, 2010; Ribet et al., 2008; Tipper,

Bencsics & Coffin, 2005) and the varying tropisms of MLVs, dictated by the *env* gene sequence, have been particularly well studied (Stoye & Coffin, 1987) (Table 1.1).

Table 1.1: MLV cell-surface receptors

Tropism	Receptor	Infectivity
Ecotropic	*mCAT-1* (*Slc7a1*)	Certain species of *Murinae*, including mouse and rat cells (Wang et al., 1991; Albritton et al., 1989; Oie et al., 1978).
Xenotropic	*Xpr1*	Wide host range, lacking infectivity in common strains of laboratory mice (Tailor et al., 1999; Yang et al., 1999).
Polytropic	*Xpr1*	Wide host range, including infectivity in mice (Tailor et al., 1999; Yang et al., 1999).
Modified polytropic	*Xpr1*	Wide host range, including infectivity in mice (Tailor et al., 1999; Yang et al., 1999).
Amphotropic	*Pit2* (*Slc20a2*)	Wide host range, including infectivity in mice (Miller et al., 1994; van Zeijl et al., 1994).

On successful binding of a receptor and entry of the cell, the viral genome is reverse transcribed and, for the majority of retroviruses, enters the nucleus upon initiation of mitosis (Roe, Reynolds, Yu & Brown, 1993) (Fig. 1.1a). Insertion site preferences are determined by differences within the integrase (Harper, Sudol & Katzman, 2003; Zhu, Dai, Fuerst & Voytas, 2003), but are also likely influenced by complex networks of interactions with other retroviral and cellular proteins (Schaller et al., 2011), as well as by secondary structures of target DNA (Snášel, Rosenberg, Pačes & Pichová, 2009). Accordingly, integration sites have been shown to vary considerably between retroviruses (De Palma et al., 2005), and highly preferred target sites have been suggested in a variety of systems (Shih, Stoye & Coffin, 1988). MLVs, for example, have been shown to preferentially integrate in promoter-proximal regions and CpG islands (Lewinski et al., 2006; Ambrosi, Cattoglio & Di Serio, 2008).

1.1.3 Retroelements and endogenous retroviruses

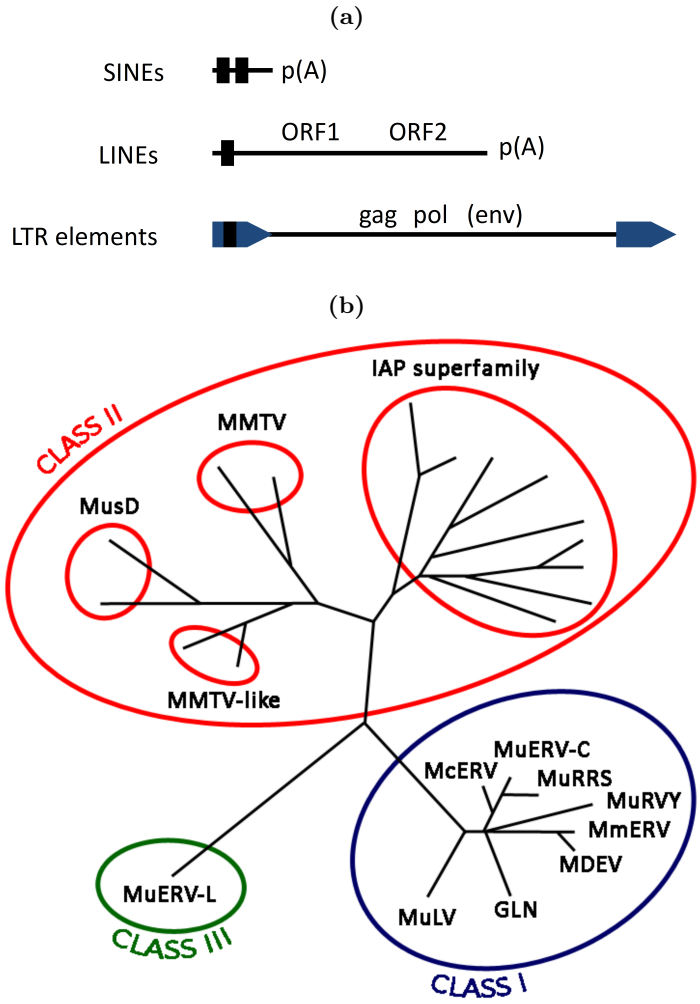
The C-value paradox describes the discrepancy between genome size and apparent organismal complexity for a wide array of eukaryotes. Whilst ploidy has a major bearing on genome size, much of remaining variability can be assigned to the relative proportions of repetitive genetic elements. One of the most surprising results from the sequencing of the human genome (Dunham, Shimizu, Roe & Chissoe, 1999; International Human Genome Sequencing Consortium, 2001) was the overall proportion of transposable elements – ∼45% of the genome sequence, or ∼3x10^6 elements in total (International Human Genome Sequencing Consortium, 2001). The later release of the mouse genome (Mouse Genome Sequencing Consortium, 2002), as well as assemblies for increasing numbers of other species, has revealed largely comparable figures, with ∼38% of the mouse genome corresponding to transposable elements (Mouse Genome Sequencing Consortium, 2002).

Around 97% of transposable elements in the mouse genome are retroelements (REs) (Mouse Genome Sequencing Consortium, 2002), utilising an RNA intermediate that is reverse transcribed prior to re-insertion. The remaining fraction is composed of DNA transposons, first identified by Barbara McClintock in the 1940s (McClintock, 1950). REs can be broadly divided into two groups according to the presence or absence of long terminal repeats (LTRs) (Fig. 1.3a). RE composition varies between species, with humans having a lower fraction of LTR-bounded elements than mice and a substantially different composition within this category (Mouse Genome Sequencing Consortium, 2002).

Non-LTR REs comprise long and short interspersed nuclear elements (LINEs and SINEs) (Fig. 1.3a). LINEs are the most abundant REs within both the human and mouse genomes (Mouse Genome Sequencing Consortium, 2002). Full-length LINE elements are autonomous, although the reverse transcriptase encoded frequently fails to successfully transcribe the full length of the element, resulting in the majority of insertions being 5'-truncated. Hence, although full length elements are 6–8 kb, the median length of insertions is only around 1 kb (International Human Genome Sequencing Consortium, 2001). SINEs (Fig. 1.3a) are non-autonomous elements, encoding no proteins and relying on LINE-encoded proteins for their replication. These elements are thought to be derived from small cellular tRNAs or the 7SL signal recognition particle component (Kramerov & Vassetzky, 2011).

LTR-bounded elements, including ERVs, comprise nearly 8 and 10% of the human and mouse genomes, respectively (International Human Genome Sequencing Consortium, 2001; Mouse Genome Sequencing Consortium, 2002). The flanking LTRs contain promoters and regulatory regions necessary for transcription of the coding regions *gag* and *pol*. Exogenous retroviruses are thought likely to have arisen through the addition of

Figure 1.3: Retroelements and endogenous retroviruses



(a) Structure of the three main families of retroelements. Shown are promoter regions (black), LTRs (blue), coding regions and polyadenylation (p(A)). *Env*, required only for exogenous virion production, is present only in some cases and is hence included in parentheses. (b) Phylogenetic tree of murine ERV reverse transcriptase domains. Modified from McCarthy and McDonald, 2004; Stocking and Kozak, 2008.

an *env* gene (Temin, 1980; Xiong & Eickbush, 1990) to this backbone, an event that appears to have occurred independently several times to create separate viral lineages (Malik, Henikoff & Eickbush, 2000). As with non-LTR elements, the majority of LTR-bounded elements are not full length sequences, most being truncated to a solo-LTR through homologous recombination (International Human Genome Sequencing Consortium, 2001).

Murine ERVs can be classified phylogenetically into three distinct classes (Fig. 1.3b). Endogenous MLVs, falling into Class I, constitute relatively small fractions of mouse genomes and have been extensively characterised. Endogenous ecotropic MLVs (eMLVs) are generally present in single copies in strains with low incidence of leukaemia,

with *Emv1*, *Emv2*, and *Emv3* being present in the most commonly-used inbred strains (Jenkins, Copeland, Taylor & Lee, 1982) (Table 1.2). Endogenous xenotropic, polytropic and modified polytropic MLVs (xMLVs, pMLVs and mpMLVs) are present in a further ∼50 copies, with considerable polymorphism between strains (Stoye & Coffin, 1988; Frankel, Stoye, Taylor & Coffin, 1990).

Table 1.2: Commonly-used inbred mouse lines possessing single endogenous eMLVs

	Emv1	*Emv2*	*Emv3*	*Emv4*	*Emv5*
A	✓				
BALB/c	✓				
CBA	✓				
C3H	✓				
SM	✓				
C57BL/6		✓			
C57BL/10		✓			
C57BR		✓			
DBA			✓		
LG				✓	
LP					✓

Compiled from Jenkins et al., 1982.

Analysis of wild mouse species and outbred strains has suggested that endogenous MLVs entered the mouse germline comparatively recently (Kozak & O'Neill, 1987). As such, both endogenous eMLV and xMLV classes contain infectious members. There is evidence across multiple species of continued acquisition of germ-line proviral insertions (Tarlinton, Meers & Young, 2006; Elleder et al., 2012) and evidence from humans suggests re-infection is a prominent factor in ERV accumulation (Belshaw, Pereira et al., 2004), although loss of *env* and subsequent intracellular replication has equally resulted in massive expansion within certain lineages (Magiorkinis, Gifford, Katzourakis, De Ranter & Belshaw, 2012).

1.1.4 Retroelements in the genome – mutagenesis and disease

Although most REs are inactivated relics that have become fixed within a species, often pre-dating a speciation event (Goodchild, Wilkinson & Mager, 1993; Tristem, 2000), there is unappreciated polymorphism for younger elements within many species. Studies of the human genome have suggested that retrotransposition plays a significant

role in individual genomic structural variation (Konkel, Wang, Liang & Batzer, 2007; Kidd, Cooper et al., 2008; Bennett et al., 2008; Kidd, Graves et al., 2010; C. R. L. Huang et al., 2010; Beck et al., 2010; Karimi et al., 2011; Reichmann et al., 2012) and can result in somatic mosaicism (Baillie et al., 2011). Whilst the majority of these variations are amongst SINEs and LINEs, there is also some polymorphism for ERVs (Kidd, Graves et al., 2010; Belshaw, Dawson et al., 2005).

Given the mutagenic potential of transposition, there are significant epigenetic and transcriptional controls on REs (Bourc'his & Bestor, 2004; Lavie, Kitova, Maldener, Meese & Mayer, 2005; Changolkar, Singh & Pehrson, 2008). Analysis of the human and mouse transcriptomes has revealed around 10 times the number of transcripts than recognised genes (Cheng et al., 2005; The FANTOM Consortium & RIKEN Genome Exploration Research Group and Genome Science Group, 2005), however, and a further analysis of 1% of the human genome suggested around 80% of bases were transcribed (The ENCODE Project Consortium, 2007). Thus, these control mechanisms are unlikely to be absolute even after embryogenesis, where necessary epigenetic modifications are suggested to allow a transient period of activity (Kano et al., 2009; H. M. Rowe et al., 2010; Matsui et al., 2010; Macfarlan, Gifford, Agarwal et al., 2011; Macfarlan, Gifford, Driscoll et al., 2012; Guallar et al., 2012). Some control may also be imposed post-transcriptionally, therefore, and a suite of cellular factors are suggested to have roles in the control of REs (Thomas & Schneider, 2011; Arjan-Odedra, Swanson, Sherer, Wolinsky & Malim, 2012), including ABOBEC family members (MacDuff, Demorest & Harris, 2009; Wissing, Montano, Garcia-Perez, Moran & Greene, 2011) and *TREX1* (Crow et al., 2006; Stetson, Ko, Heidmann & Medzhitov, 2008).

Where these mechanisms are evaded and novel integrations are seen, long-term evolutionary pressure can act as a further protection mechanism by conferring an advantage on their subsequent mutation or deletion. ERV and LINE insertions both within and in the vicinity of genes are comparatively under-represented, a phenomenon increasingly apparent with increasing age of the insertion (Medstrand, van de Lagemaat & Mager, 2002; Nellåker, Keane et al., 2012). Probability of permanent integration is also associated with neighbouring gene function, with highly-conserved housekeeping genes having few proximate insertions (van de Lagemaat, Landry, Mager & Medstrand, 2003; Nellåker, Keane et al., 2012).

The potential for novel integrations to cause disease or obvious phenotypic change has long been studied. In mice alone, three separate integrations of REs are known to alter fur characteristics. Integration of an intra-cisternal A particle (IAP) has been seen to cause the agouti phenotype (also causing susceptibility to obesity, diabetes and tumourigenesis (Duhl, Vrieling, Miller, Wolff & Barsh, 1994)) and separate MLV integrations to cause the hairless (Stoye, Fenner, Greenoak, Moran & Coffin, 1988)

and dilute phenotypes (Jenkins, Copeland, Taylor & Lee, 1981; Hutchison, Copeland & Jenkins, 1984). Interestingly, proviral insertions have also recently been seen to cause the hooded and Irish hooded coat phenotypes in rats (Kuramoto et al., 2012). Further work in mice has also associated REs with cleft-palate (Juriloff et al., 2005) disorders and lupus-nephritis (Yoshiki, Mellors, Strand & August, 1974), amongst other conditions.

The involvement of REs in human disease is far more controversial; although novel integrations can be seen in certain cases (Iskow et al., 2010; E. Lee et al., 2012), varied expression in others (Herbst, Sauter & Mueller-Lantzsch, 1996; Wentzensen et al., 2007; Frank et al., 2008; R. Contreras-Galindo, M. H. Kaplan, Leissner et al., 2008; Lamprecht et al., 2010), and the establishment of immune responses against their products (Sauter et al., 1996; Wang-Johanning et al., 2008), determining causal links is fraught with difficulty. In many cases, expression of REs may simply result from global transcriptional changes resulting from cellular transformation, with immune responses to RE-derived proteins merely representing their status as novel immune targets.

1.1.5 Retroelements in the genome – co-option and function

Whilst housekeeping genes may be relatively protected from nearby insertion of REs, 'rapidly evolving' genes, such as those involved in stress response and defence, have an increased likelihood of proximate insertion events (van de Lagemaat et al., 2003; Nellåker, Keane et al., 2012). The occasional presence of REs within exonic regions further suggests a potential role for transposable element integration in gene evolution and diversification (Nekrutenko & W.-H. Li, 2001). Indeed, given the density of repetitive elements within mammalian genomes and the potential variability in the impacts of their integration, it is perhaps unsurprising that many integrations might have positive impacts for the host.

Bioinformatic screens of human gene promoter regions suggest that ∼25% contain sequences derived from transposable elements (Jordan, Rogozin, Glazko & Koonin, 2003; van de Lagemaat et al., 2003; Conley, Piriyapongsa & Jordan, 2008), including ERVs (Buzdin, Kovalskaya-Alexandrova, Gogvadze & Sverdlov, 2006). In certain cases, insertions have obvious and definable impacts. In the human genome, for example, the integration of an ERV upstream of a pancreatic amylase allows its expression and release into saliva (Samuelson, Phillips & Swanberg, 1996), whilst another likely contributes the NF-κB-responsive expression of interferon (IFN)λ1 (Thomson et al., 2009).

Through co-option by the host, REs can have an even more direct impact. An abundance of retroviral envelope protein expression in the human placenta suggested a potential functional role (Venables, Brookes, Griffiths, Weiss & Boyd, 1995), a hypothesis

later borne-out in the discovery of syncytins (Mi et al., 2000; Blaise, de Parseval, Bénit & Heidmann, 2003). Interestingly, the fusogenic and immunomodulatory role of syncytins in placentation has been independently replicated in rodents (Dupressoir, Marceau et al., 2005), lagomorphs (Heidmann, Vernochet, Dupressoir & Heidmann, 2009), carnivores (Cornelis et al., 2012), and bovids (Dunlap et al., 2006; Baba et al., 2011), and knock-out and knock-down studies have shown their crucial significance (Dunlap et al., 2006; Dupressoir, Vernochet et al., 2009).

ERV-derived proteins can also be used by the host in the prevention of retroviral infection through mechanisms such as receptor blockade, with the mouse *Fv4* gene, a co-opted retroviral envelope (Ikeda, Laigret, Martin & Repaske, 1985; Ikeda & Sugimura, 1989; Limjoco, Dickie, Ikeda & Silver, 1993), and endogenous Jaagsiekte sheep retroviruses (JSRVs) (Spencer, Mura, Gray, Griebel & Palmarini, 2003) both functioning in this manner. Similarly, *Fv1*, operating after viral entry but before proviral integration, is derived from a retroviral *gag* sequence (Best, Le Tissier, Towers & Stoye, 1996) and shows evidence of strong positive selection through mouse evolution (Yan, Buckler-White, Wollenberg & Kozak, 2009).

Larger structural changes, such as genomic translocations, inversions, duplications, or deletions, can also be mediated by REs. An analysis of 35 human ERVs, for example, suggested that 6 of those studied had undergone inter-element recombination (Hughes & Coffin, 2001). Events such as these can impact surrounding genomic areas and have allowed the recent expansion of the glycophorin gene family during primate evolution (Rearden, Magnet, Kudo & Fukuda, 1993). Where elements occur in unusually high density, for example LINE elements in both the human and mouse allosomes (International Human Genome Sequencing Consortium, 2001; Mouse Genome Sequencing Consortium, 2002) (reaching densities of 89% of a 0.5 Mb region of human chromosome X (International Human Genome Sequencing Consortium, 2001)), these events can be far larger. Analysis of the human allosomes suggests a LINE-mediated $\sim$4 Mb translocation from the X to the Y chromosome, followed by its subsequent inversion, again mediated by LINE-LINE recombination (Schwartz et al., 1998).

1.2 The immune system

1.2.1 A high-level perspective

After the physical and chemical barriers to infection provided by environmental surfaces, an immune system constitutes the principal means of pathogen control, providing discrimination between 'self' and 'non-self'. Pathogens vary greatly, not only in their modes of pathogenesis, but also in structure, size, and complexity, meaning a dynamic and coordinated system is required for their control. Ultimately, a successful immune response will result in the control or clearance of a pathogen, whilst limiting off-target impacts and damage to the host.

Immune systems differ considerably between organisms and whilst some elements are shared, differing systems of varying natures and complexities have evolved in response to the ubiquitous presence of pathogens. Antimicrobial peptides, such as defensins, represent the oldest shared system of pathogen control and comparisons of genome sequences from plants and humans shows their conservation throughout evolutionary history (Ganz, 2003). Innate pathogen-associated molecular pattern (PAMP) sensors that control the expression of these molecules are intrinsically linked to this history, and receptors such as Toll, in *Drosophila melanogaster*, are again highly conserved (as Toll-like receptors – TLRs – in other organisms) (Hoffmann, Kafatos, Janeway & Ezekowitz, 1999; Barreiro et al., 2009). Some lineages have diversified these receptors to provide highly specific innate immune responses, and 222 TLR genes are seen in *Strongylocentrotus purpuratus* (the purple sea urchin) (Rast, Smith, Loza-Coll, Hibino & Litman, 2006) in comparison to only 10 in humans. The mammalian adaptive immune system is thought to descend from the acquisition of enzymes to allow controlled somatic recombination (Agrawal, Eastman & Schatz, 1998), first seen in cartilaginous fish, which allowed the virtually instantaneous ability to produce variable antigen receptors (see Section 1.2.2).

Broadly, the mammalian immune system can be divided into the innate (non-specific) and adaptive (pathogen-specific) responses, each comprising both cellular and humoral components. Innate immune components are for the most part constitutively present and hence provide an immediate and maximal response to infection. In contrast, the adaptive immune system requires time to develop an effective response, but is pathogen-specific and includes facility for immunological memory, allowing a more rapid response in the case of re-infection.

All immune cells ultimately derive from a pluripotent haematopoietic stem cell population in the bone marrow. This gives rise to both the common myeloid progenitor (CMP – for innate immune cells) and the common lymphoid progenitor (CLP – for adaptive immune cells). The CLP in turn gives rise to both B and T lymphocytes, although T

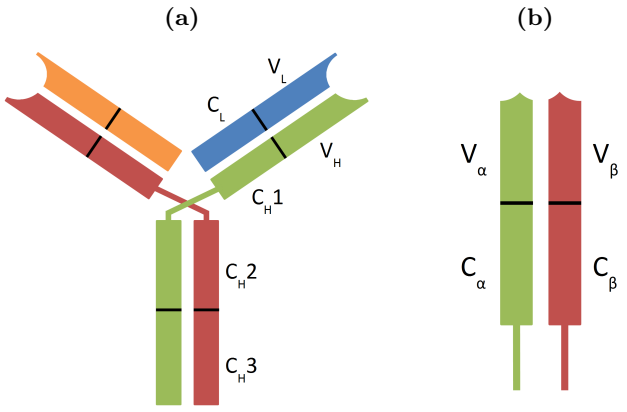
cell progenitors migrate to the thymus for their continued development into either the $\alpha\beta$ and $\gamma\delta$ T cell lineages.

1.2.2 Variable antigen receptors – form and function

Effective functioning of the adaptive immune response relies on the ability of responding cells to target a highly diverse array of antigens with a high level of specificity. Whilst some pathogens may have single immunodominant antigen epitopes, against which the majority of either a T or B cell response is directed, others will present many targets simultaneously (Doolan et al., 2003). To further complicate the requirements, the selection pressures imposed by an immune response can provoke a pathogen to vary its antigen repertoire. This may be through methods such as escape mutation, but complex pathogens may also invoke expression changes and cyclic or phased use of redundant proteins. Where a particular pathogen is endemic to an area, this variability can occur both at the individual and population level (A. Jalloh, van Thien et al., 2006; A. Jalloh, M. Jalloh & Matsuoka, 2009).

Both T and B cells possess variable antigen receptors – T cell and B cell receptors (TCRs and BCRs, Fig. 1.4), respectively – produced through a complex system of somatic gene re-arrangement first realised for the BCR (Hozumi & Tonegawa, 1976; Weigert, Gatmaitan, Loh, Schilling & Hood, 1978).

Figure 1.4: The T and B cell receptors



(a) Stylised structure of IgG. Colouring shows heavy ($_{\mathrm{H}}$ – red and green) and light ($_{\mathrm{L}}$ – orange and blue) chains and constant (C) and variable (V) regions are marked. The light chain, C_H1 and V_H comprise the Fab region and the C_H2 and C_H3 regions comprise the Fc region. (b) Stylised structure of the TCR. Colouring shows the α (green) and β (red) chains, with the variable (V) and constant (C) regions marked. Transmembrane region and cytoplasmic tail are shown underneath.

The BCR (Fig. 1.4a), or immunoglobulins (Ig), can either be cell-associated or secreted as antibodies. Immunoglobulins are constructed in a uniform manner from heavy and light chains to form a variable region, involved in antigen binding, and a constant

region, which distinguishes the five classes of immunoglobulin (IgM, IgD, IgG, IgA, and IgE). The two 'arms' of the immunoglobulin molecule are termed Fab (fragment antigen binding) and the 'stem' as Fc (fragment crystallisable).

The light chain variable region is formed through the recombination of two gene regions (V – variable, and J – joining), whereas the variable region of the heavy chain is formed from three (V, D – diversity, and J). The potential diversity created by this system comes from the number of genes within each region (∼70 V and ∼10 J genes for the light chain, and ∼40 V, ∼25 D, ∼5 J genes for the heavy chain within the human genome) and the potential for random selection within this pool, but also from variability in the recombination process causing small nucleotide insertions or deletions.

The $\alpha\beta$TCR (Fig. 1.4b), analagous to the Fab region of the BCR, is comprised of two chains, each possessing a constant and variable region. Similarly again, the variable region of the TCR α locus consists of a V and J region, whereas the β locus also contains a D gene region. Gene diversity is larger for the TCR loci, however, with ∼70 V and ∼60 J genes for the α chain and ∼50 V, 2 D and ∼15 J genes for the β chain.

The system of somatic recombination used to generate these receptors provides a far larger potential receptor pool than the total number of circulating T or B cells. The estimated size of the TCR repertoire in mice, for example, is around 10^{13} (Nikolich-Zugich, Slifka & Messaoudi, 2004) and whilst many re-arrangements may result in non-functional TCRs, this figure remains far larger than the total number of circulating T cells (1–2 x 10^8 in an adult mouse). Even amongst genetically identical mouse lines, therefore, this sampling effect means that between two individuals, only around 20% of TCRs are shared (Bousso et al., 1998; Maryanski et al., 2001).

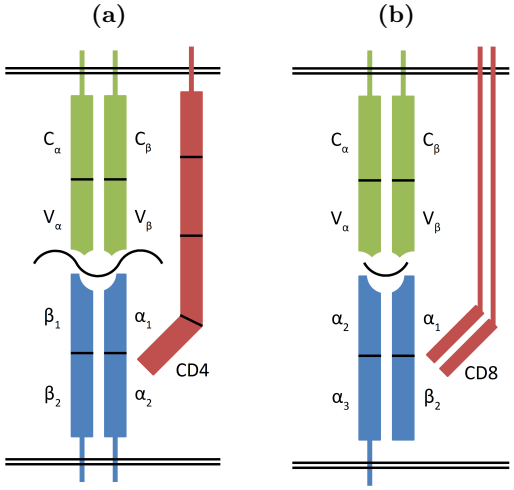
1.2.3 The T cell

T lymphocyte development in the thymus results in the formation of two lineages, divided according to the structure of their TCR – $\alpha\beta$ and $\gamma\delta$. The former further differentiate into two groups categorised according to their co-receptor expression – $CD4^+$ ('helper') and $CD8^+$ ('cytotoxic') T cells. On activation, $CD4^+$ T cells are primarily immunomodulatory, whereas $CD8^+$ T cells are involved in the direct killing of infected cells.

T cell activation depends on two complementary signalling routes through an interaction with an antigen presenting cell (APC). Peptides are presented in complex with major histocompatibility complex (MHC) molecules and are bound by a TCR, along with the appropriate CD4 or CD8 co-receptor (Fig. 1.5). Secondly, a positive signal must come through the binding of the T cell CD28 receptor to the co-stimulatory molecules CD80 or CD86 on the surface of the APC. Further interactions, such as between

CD40 and CD40L, or CTLA-4 and CD80 and CD86, may either contribute to or inhibit activation.

Figure 1.5: Antigen presentation to T cells



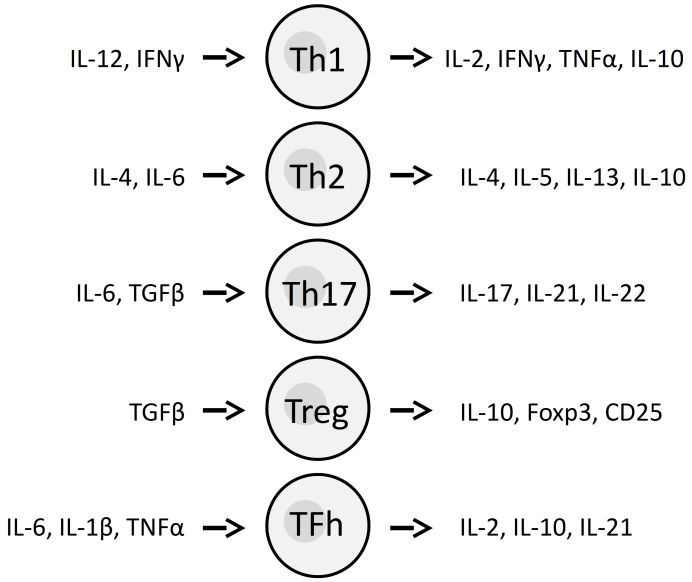
The interactions of (a) MHC class II with the TCR and co-receptor of $CD4^+$ T cells and (b) MHC class I with the TCR and co-receptor of $CD8^+$ T cells. The peptide bound between the TCR (green) and either MHC molecule (blue) is shown in black. CD4 and CD8 co-receptors are shown in red.

When a strong positive interaction occurs, the binding T cell is caused to clonally divide to give large numbers of cells with identical specificities. This allows the enrichment of pathogen-specific T cells within the circulating population and efficient control of the infection. The specific path of differentiation taken by a T cell depends on the cytokine milieu present at the point of its activation, however, and determines its cytokine production and cellular interactions (J. Kim, Woods, Becker-Dunn & Bottomly, 1985; Mosmann, Cherwinski, Bond, Giedlin & Coffman, 1986; Cherwinski, Schumacher, Brown & Mosmann, 1987; Amsen et al., 2004) (Fig. 1.6).

$CD4^+$ T cells activated in the presence of interleukin (IL)-12 differentiate into Th1 cells, producing IL-2 and IFNγ. Conversely, activation in the context of IL-4 causes differentiation into Th2 cells, producing IL-4, IL-5, and IL-13 (Fig. 1.6). Although some plasticity remains, either route of polarisation, and subsequent cytokine production, inhibits the differentiation of cells into the opposing subset (Fernandez-Botran, Sanders, Mosmann & Vitetta, 1988). Further $CD4^+$ T cell subsets have since been identified, each defined by the cytokines required for polarisation, cell-surface protein expression, and cytokine production and function.

$CD8^+$ T cells prompt the apoptosis of infected cells through the targeted release of vesicles containing perforin and granzymes. Perforin polymerises to create cylindrical

Figure 1.6: CD4$^+$ T cell subsets



Necessary polarising factors (left) and the effector profile (right) of the commonly recognised CD4$^+$ T cell subsets.

transmembrane pores in the target cell that allow the influx of water and salts. Granzyme proteases are also suggested to enter target cells through these pores, activating a cascade that results in the initiation of programmed cell death and cleavage of cellular DNA. Whilst cytotoxic killing is primarily achieved by CD8$^+$ T cells, there is also some evidence of CD4$^+$ T cells having direct antiviral effects. MHC class II-restricted killing was first seen *in vitro* (Jacobson et al., 1984; Lukacher, Morrison, Braciale, Malissen & Braciale, 1985), but has subsequently also been noted *ex vivo* (Appay et al., 2002) and *in vivo* (Jellison, Kim & Welsh, 2005).

A successful immune response to an infection results in the clearance of the pathogen and hence clearance of the epitopes against which T cell responses were mounted. At this point, the majority of cells are induced to apoptose through 'neglect', but a small fraction remain as memory cells. These remain for long periods and provide the basis for immunological memory, rapidly dividing to give a population of effector cells on re-infection.

1.2.4 The B cell

Where T cells act by recognising peptide epitopes with membrane-bound receptors, B cells act by secreting their BCR to circulate in the blood and lymph. Antibodies, and humoral immunity, work in a variety of ways to clear pathogens and circulating products or toxins through engagement of the innate immune system.

The majority of B cell activation occurs through interaction with CD4$^+$ T cells. Naïve

B cells express both IgM and IgD, and cell-surface immunoglobulins can bind and internalise antigens for processing and presentation on MHC class II molecules. This process acts as a first positive signal to the cell, but full activation requires a further stimulus delivered through CD40 when a $CD4^+$ T cell binds the presented peptide-MHC complex. The interaction of CD40 with CD40L, as well as the binding of cytokines, such as IL-4 from Th2 cells, promotes the proliferation of the B cell and affinity maturation.

Affinity maturation describes the incremental gains in antibody specificity brought about by multiple rounds of somatic hypermutation (targeted point mutation of the immunoglobulin V regions) and cell division. Upon each round, cells with increased affinity for antigen are selected, whilst mutations resulting in detrimental changes to the BCR cause cell death and, subsequently, phagocytosis of the cell. Signalling through CD40 also induces class switch recombination, where the previously re-arranged V region is paired with a different C region to give different classes of immunoglobulin. Similarly to T cell differentiation, the specific nature of the recombination largely depends on the local cytokine milieu. In the mouse, IL-4 (from Th2 cells) primarily induces switching to IgG1 and IgE, for example, whereas IFNγ (from Th1 cells) promotes recombination to IgG2a and IgG3.

Activated B cells differentiate from their naïve state into plasmablasts, beginning to secrete antibodies, but continue to divide until forming terminally differentiated plasma cells. Alternatively, cells will progress to a memory phenotype to contribute to immunological memory functions. Plasma cells are highly specialised to secrete antibodies of a designated class, each of which contribute to humoral immunity in a different manner.

Humoral immunity can be broadly divided into three functions: neutralisation, opsonisation, and complement activation. Neutralisation of pathogens or their products involves their binding by antibodies to block their access to targeted cellular receptors and is primarily through IgG and IgA antibodies. Neutralised pathogens are subsequently phagocytosed and degraded. In a similar manner, opsonisation involves the coating of a pathogen such that it can be recognised by innate immune cells for phagocytosis. This is predominantly through the action of IgG1 antibodies. Antibody coating with either IgM or IgG3 in this manner can also recruit proteins of the innate complement system to directly kill or promote the phagocytosis of bacteria.

The ultimate function of antibodies is, therefore, in mediating ingestion by phagocytic cells, a process that relies on the invariant nature of the constant regions of the immunoglobulin heavy chain. The varying heavy chains present on class switched antibodies are also selectively bound by membrane-bound Fc receptors to allow their specific internalisation and transport across epithelial surfaces. This allows the secretion of IgA across the gastro-intestinal epithelium, for example, an important primary defence against infection by food and water-borne pathogens.

1.2.5 Importance and immunodeficiency

Many immunodeficiencies have been characterised in humans and subsequently studied in animal models. Deficiencies can be seen in both the innate and adaptive immune systems and can have varying impacts according to the severity, number, and relative redundancy of systems impacted. Highlighting the co-operative nature of the immune response, many deficiencies have convergent effects and are characterised by similar susceptibilities and diseases.

The earliest immunodeficiencies studied in humans related to antibody production, with X-linked agammaglobulinemia (XLA) being the first described in 1952 (Bruton, 1952). This disease has subsequently been linked to the Btk tyrosine kinase, involved in B cell development. Mutation of this protein largely arrests development of B cell precursors, giving greatly reduced titres of all antibody classes. XLA patients (along with those suffering similar deficiencies, such as chronic variable immunodeficiency – CVID) are characterised by their poor defence against extracellular bacteria, including *staphylococci* and *streptococci*, but can be treated effectively with regular antibody infusions. Deficiencies in the complement system have analagous impacts, including increased susceptibility to extracellular bacterial infections.

More severe immunodeficiencies generally result from defects involving T cell differentiation, thus impacting both arms of the adaptive immune system through the inability to produce T cell-dependent antibody responses. These conditions are grouped as severe combined immunodeficiency (SCID) disorders and are most commonly linked to mutations impacting cytokine signalling and to mutations in the enzymes required for the somatic recombination of TCR and BCR genes. Both RAG1 and RAG2 proteins are required for V(D)J recombination in developing lymphocytes, for example, and mutations result in complete blocks in T and B cell development, or limited and abnormal development of T cells alone, such as in Omenn syndrome. Similar defects can result from mutations impacting thymic function, MHC function, and antigen presentation.

The negative effects of primary immunodeficiencies are generally seen by adulthood in humans, but a range of secondary (acquired) conditions can develop throughout life. One of the most striking examples of this is immunocompromisation through human immunodeficiency virus (HIV) infection and subsequent development of acquired immune deficiency syndrome (AIDS). Infection leads to the gradual decline in $CD4^+$ T cell numbers and their functional impairment, eventually resulting in inadequate immune responses to opportunistic infections of ubiquitous pathogens. Ultimately, over 25% of AIDS-related deaths are due to opportunistic infections (Bonnet et al., 2005).

1.3 Aims of this work

Whilst endogenous retroviruses and retroelements have long been studied by virologists, comparatively little work has centred around their potential interactions with the immune system. As such, these concepts represent under-studied and highly intriguing areas of research.

The relevance of this area is only highlighted by issues surrounding the use of retroviral vectors in human gene therapy (Donahue et al., 1992; Purcell et al., 1996; Howe et al., 2008; Zhang et al., 2008), and in the use of non-human tissues and cells in engraftment or in therapy of conditions such as liver and renal failure (Wilson et al., 1998; Cunningham et al., 2004; Di Nicuolo et al., 2010). Further, the rogue expression of these elements in transformed or infected cells, discussed in Section 1.1.4, suggests the possibility of their use as secondary immune targets (Takahashi et al., 2008). In the absence of better understanding of their ranges of expression, including in situations of infection and disease (Evans, Alamgir et al., 2009), as well as potential unknown functionality, however, the induction of a response against such ubiquitous self-antigens is a potentially dangerous possibility (Ludewig et al., 2000).

The work described in this thesis examines the interaction between ERVs and the immune system, investigating these links from both angles: the impact of ERVs on the immune system, and the impact of the immune system on ERVs. The mouse model has numerous benefits in the field of immunology and the extensive use of congenic, transgenic, and knock-out mice has greatly facilitated this work. A large amount of further work has centred on the parallel development of a software, *REquest*, that allows the analysis of ERV and RE expression using commercial microarrays.

1.3.1 A word on structure

This section has served to provide a global background to the areas covered within the thesis, but given the varied directions of the research, both chapters contain a further introduction and aims, as well as independent results and conclusions. Thus, each chapter can largely be considered separately, and indeed have been published as independent studies. *REquest* is presented as a longer section of the methods chapter.

Publications resulting from this work have been bound at the rear of the thesis.

Chapter 2

Methods

2.1 Mice

Mouse strains used within this thesis are detailed in Table 2.1. Origins or breeding colony identifiers are listed alongside full and shorter strain names, the latter being used to identify strains throughout the thesis. Sources listed are detailed in Table 2.2. All congenic strains used are backcrossed to the recipient strain for a minimum of 10 generations. All mice obtained from external sources were introduced into animal houses via implantation of embryos into pseudopregnant females to facilitate the removal of adventitious agents.

Mice bred within NIMR are pathogen tested according to guidelines from the Federation of European Laboratory Animal Science Associations (FELASA, www.felasa.eu) (Nicklas et al., 2002) to confirm absence of various viral, bacterial, mycoplasmal, and fungal pathogens, as well as endo- and ecto-parasites.

Unless otherwise stated, all mice used for this work were aged between 4 and 8 weeks.

Table 2.1: Mouse strains

Strain	Alias	Source	Ref
C57BL/6J	B6	NIMR-B1 JAX UMICH CIML RCHCI NCI	
129S8/SvEvNimrJ	129S8	NIMR-B1	
A/J	A	NIMR-B1	
B6.A-*Emv2*$^{-/-}$	*Emv2*$^{-/-}$	NIMR-B2	Young, Ploquin et al., 2012
B6 mice (*Emv2*$^{+/+}$) congenic for the null allele at the *Emv2* integration site (*Emv2*$^{-/-}$) from A strain mice. See Section 2.1.1 for more information.			
B6 EF4.1	B6 EF4.1	NIMR-B2	Antunes et al., 2008
B6 mice transgenic for a TCRβ chain from an env$_{122\text{-}141}$-specific CD4^{+} T cell clone. Mice have a polyclonal T cell population with increased frequencies of clones specific to env$_{122\text{-}141}$L.			
B6-*Emv2*$^{-/-}$ EF4.1	*Emv2*$^{-/-}$ EF4.1	NIMR-B2	Young, Ploquin et al., 2012

... Table 2.1 continued ...

Strain	Alias	Source	Ref
Emv2$^{-/-}$ mice crossed to carry the env$_{122\text{-}141}$-specific TCRβ chain from B6 EF4.1.			
129 EF4.1	129 EF4.1	NIMR-B2	
129 mice crossed to carry the env$_{122\text{-}141}$-specific TCRβ chain from B6 EF4.1.			
B6.A-*Fv2*^{s}	*Fv2*^{s}	NIMR-B2	Marques et al., 2008
B6 mice (*Fv2*^{r}) congenic for the dominant *Fv2*^{s} allele from A strain mice, conferring susceptibility to Friend virus.			
B6.C3H-*Fv1*^{n}	*Fv1*^{n}	NIMR-B2	Pike et al., 2009
B6 mice (*Fv1*^{b}) congenic for the *Fv1*^{n} allele from C3H mice, conferring permissiveness to N-tropic viruses.			
B6.SJL-*Ptprc*^{a}*Pep3*^{b}	B6 CD45.1	NIMR-B1	
Mice have an allelic form of CD45 (CD45.1 / Ly5.1), allowing determination of donor and host-derived (CD45.2) cells.			
(B6 CD45.1 x B6)F$_1$	CD45.1/2	NIMR-B1	
Progeny of a B6 x B6 CD45.1 mating, such that mice are CD45.1/2 syngeneic. This allows the determination of donor cells from both CD45.1 or CD45.2 mice within a single mouse.			
B6.129S7-*Rag1*^{tm1Mom}	*Rag1*$^{-/-}$	NIMR-B1 JAX UMICH RCHCI	Mombaerts et al., 1992
Targeted disruption of *recombination activating gene 1* (*Rag1*), preventing V(D)J recombination and blocking development of mature T and B lymphocytes.			
B6-*Rag1*$^{-/-}$*Emv2*$^{-/-}$	*Rag1*$^{-/-}$*Emv2*$^{-/-}$	NIMR-B2	
Emv2$^{-/-}$ crossed to *Rag1*$^{-/-}$ and hence lacking mature T and B lymphocytes.			
B6.129P2-*Tcra*^{tm1Mjo}	*Tcra*$^{-/-}$	NIMR-B2	Philpott et al., 1992
Targeted disruption of the *T cell receptor alpha chain* (*Tcra*), preventing formation of the $\alpha\beta$ T cell lineage.			
B6.129S7-*Tcrd*^{tm1Mom}	*Tcrd*$^{-/-}$	NIMR-B2	Itohara et al., 1993

... Table 2.1 continued ...

Strain	Alias	Source	Ref
Targeted disruption of the *T cell receptor delta chain* (*Tcrd*), preventing formation of the $\gamma\delta$ T cell lineage.			
B6.129S2-*H2*$^{dlA^{b1}\text{-}E^{a}}$	*H2-A,E*$^{-/-}$	NIMR-B2	Cosgrove et al., 1991
Targeted disruption of *histocompatibility 2* (*H2*) loci A^b to E^a, preventing peptide presentation on MHC class II and $CD4^+$ T cell development.			
B6.129S7-*Igha*^{tm1Grh}	*Igha*$^{-/-}$	RCHCI	Harriman et al., 1999
Targeted disruption of the *immunoglobulin heavy constant alpha chain* region, blocking production of IgA.			
B6.129P2-*Pigr*^{tm1Rast}	*Pigr*$^{-/-}$	RCHCI	Uren et al., 2003
Targeted disruption of the *polymeric immunoglobulin receptor*, blocking transcytosis of dimeric IgA and pentameric IgM across mucosal surfaces.			
B6.CBA-*Aicda*^{tm1Hon}	*Aicda*$^{-/-}$	NCI	Muramatsu et al., 2000
Targeted disruption of *activation-induced cytidine deaminase*, preventing production of class-switched antibodies.			
B6.129S7-*Igh-J*^{tm1Dhu}	*Igh-j*$^{-/-}$	RCHCI	Chen et al., 1993
Targeted disruption of the *immunoglobulin heavy chain joining* region, blocking development of mature B lymphocytes.			
B6.129S2-*Ighm*^{tm1Cgn}	*Ighm*$^{-/-}$	NIMR-B2 JAX NCI	Kitamura et al., 1991
Targeted disruption of *immunoglobulin heavy constant mu* (*Igh-6*), blocking development of mature B lymphocytes.			
B6-*Ighm*$^{-/-}$ MD4	*Ighm*$^{-/-}$ MD4	NIMR-B2	Goodnow et al., 1988
Ighm$^{-/-}$ mice transgenic for a Hen-egg lysozyme (HEL)-specific BCR.			
B6.129P2-*Myd88*^{tm1Aki}	*Myd88*$^{-/-}$	NIMR-B2	Adachi et al., 1998
B6.129P2-*Myd88*$^{tm1.1Defr}$	*Myd88*$^{-/-}$	JAX	Hou et al., 2008
Targeted disruption of *Myeloid differentiation primary response gene 88* (*Myd88*), impairing TLR signalling, amongst other pathways.			

... Table 2.1 continued ...

Strain	Alias	Source	Ref
B6.129S1-*Tlr7*^{tm1Flv}	*Tlr7*$^{-/-}$	JAX CIML	Lund et al., 2004
B6.129P2-*Tlr7*^{tm1Aki}	*Tlr7*$^{-/-}$	LRI	Hemmi, Kaisho et al., 2002
Targeted disruption of *Toll like receptor 7* (*Tlr7*), preventing the recognition of single-stranded RNA.			
B6.129P2-*Tlr9*^{tm1Aki}	*Tlr9*$^{-/-}$	CIML	Hemmi, Takeuchi et al., 2000
Targeted disruption of *Toll like receptor 9* (*Tlr9*), preventing the recognition of unmethylated CpG DNA.			

Table 2.2: Mouse sources

NIMR-B1	SPF facility handing the large-scale breeding of frequently used mouse strains. Mice are maintained with autoclaved diet and neutral pH, doubly filtered, and UV-irradiated water.
NIMR-B2	SPF IVC breeding facility handing congenic, transgenic and mutant mouse strains. Mice are bred with autoclaved diet and acidified (pH 2.5) water.
UMICH	Germ-free mouse strains provided by the University of Michigan Unit for Laboratory Animal Medicine (ulam.med.umich.edu) (Ann Arbor, US). Strains are re-derived via sterile caesarian section of pregnant females and maintained in germ-free conditions with autoclaved diet and autoclaved distilled water.
JAX	The Jackson Laboratory (jaxmice.jax.org) (Bar Harbor, US) SPF breeding facility provides mouse strains bred in IVCs with sterilised diet and acidified (pH 2.8–3.2) water.
CIML	SPF breeding colony located at the University Marseille-Luminy Centre for Immunology (ciml.univ-mrs.fr) (Marseille, France). Mouse strains are maintained with autoclaved neutral pH water.

... Table 2.2 continued ...

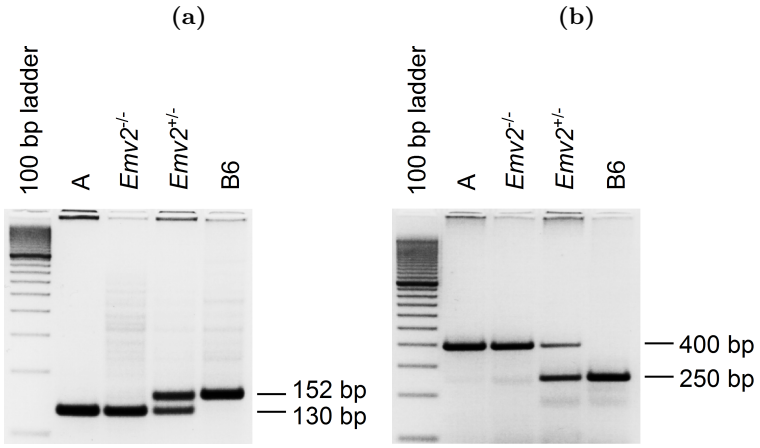
NCI	SPF breeding facilities provided by the US National Institutes of Health National Cancer Institute (cancer.gov) (Frederick / Bethesda, US). Mouse strains are offered autoclaved neutral pH water.
RCHCI	SPF breeding facilities provided by the Swiss Federal Institute of Technology Zürich (ethz.ch) (Zürich, Switzerland). Mouse lines are maintained on IVCs with autoclaved neutral pH water.
LRI	SPF breeding facilities at the London Research Institute (LRI), Cancer Research UK (london-research-institute.org.uk). Strains are maintained on autoclaved water. Mice from this source were transferred into the NIMR quarantine facility and bred before testing.

All *Rag1*$^{-/-}$ mice detailed were established by independent re-derivation of a common stock obtained from JAX. Between units, these stocks have been maintained by homozygous breeding for between 12 and 45 filial generations (F_{12} to F_{45}). The *Ighm*$^{-/-}$ JAX stock had been maintained to F_{15}. *Ighm*$^{-/-}$ mice at NIMR have recently been re-derived twice. The first colony was maintained for 16 months (F_4) on neutral pH water before being offered acidified water for F_{12}, whilst the second was maintained on a diet of neutral pH water for 7 months (F_1–F_2) prior to sampling. *Myd88*$^{-/-}$ mice at NIMR and JAX have been established for around F_{16} and F_5, respectively. *Tlr7*$^{-/-}$ mice have been maintained by homozygous breeding for F_{11} at LRI, F_5 at JAX, and F_9 at CIML.

2.1.1 Generation of *Emv2*$^{-/-}$ congenics

Emv2$^{-/-}$ B6 background mice were created by the serial backcrossing of the *Emv2* integration site from the A strain, null for this proviral integration on chromosome 8. Congenics were backcrossed for 12 nuclear generations. Lack of *Emv2* was validated by PCR using both the nearby *D8Mit49* microsatellite marker (Fig. 2.1a), polymorphic between in A (*Emv2*$^{-/-}$) and B6 (*Emv2*$^{+/+}$), and using primers flanking and within the proviral integration (Fig. 2.1b).

Figure 2.1: *Emv2* typing by PCR



Typing PCRs for (a) *D8Mit49* and (b) the *Emv2* integration. PCR products are shown on a 2% agar gel run in TBE buffer. Primers are listed in Table 2.3.

2.1.2 Ethics statement

All animal experiments were approved by an independent ethical committee within NIMR, and conducted according to local guidelines and UK Home Office regulations under the Animals Scientific Procedures Act 1986 (ASPA). Where mice have been obtained from external sources, all experiments, breeding, and procedures were conducted according to local guidelines and regulations, and subsequent to approval from respective ethical committees.

Animals were culled by an approved Schedule I method according to Home Office policy.

2.2 Primers

All primers listed were obtained from Eurofins MWG Operon (Ebersberg, Germany).

Table 2.3: Primers

Primer	Sequence (5'–3')	Length (nt)
Genotyping primers		
D8Mit49_F	TCTGTGCATGGCTGTGTATG	20
D8Mit49_R	TGGTGTGCTGCTGATGCT	18
D8Mit49 microsatellite marker ~640 kb downstream the *Emv2* integration on B6 chromosome 8. Product is 130 or 152 bp in A or B6 mice, respectively.		
Emv2_3'_F	ACCCACTAAGTAACCCAGGCTGCCTCAGCT	30

... Table 2.3 continued ...

Primer	Sequence (5'–3')	Length (nt)
Emv2_5'_R	GACCAGAATAGAAAGACGTTCAAGTGAGCT	30
Emv2_LTR_R	CGCGGGTACAGAAGCGAGAAGCGAGCTGAT	30
Flanking and internal primers for the exact integration site of *Emv2* on B6 chromosome 8. Emv2_3'_F and Emv2_LTR_R give a 250 bp product when *Emv2* is present. Emv2_3'_F and Emv2_5'_R give a 400 bp product when *Emv2* is absent (Young, Ploquin et al., 2012).		
Housekeeping and normalisation primers		
HPRT_F	TTGTATACCTAATCATTATGCCGAG	25
HPRT_R	CATCTCGAGCAAGTCTTTCA	20
Hypoxanthine-guanine phosphoribosyltransferase (*HPRT*) is used as a housekeeping gene against which RNA expression levels can be normalised. 92 bp product.		
Ifnar1_F	AAGATGTGCTGTTCCCTTCCTCTGCTCTGA	30
Ifnar1_R	ATTATTAAAAGAAAAGACGAGGCGAAGTGG	30
Interferon alpha beta receptor 1 (*Ifnar1*) is used as a gene against which DNA levels can be standardised. 150 bp product.		
Expression primers (spliced)		
eMLV_spenv_F	CCAGGGACCACCGACCCACCGT	22
eMLV_spenv_R	TAGTCGGTCCCGGTAGGCCTCG	22
Spliced ecotropic MLV *env*. 116 bp product (Young, Ploquin et al., 2012).		
MMTV_F	AGAGCGGAACGGACTCACCA	20
MMTV_R	TCAGTGAAAGGTCGGATGAA	20
Spliced MMTV *env*. 105 bp product.		
Expression and DNA quantitation primers (non-spliced)		
eMLV_env_F	AGGCTGTTCCAGAGATTGTG	20
eMLV_env_R	TTCTGGACCACCACACGAC	19
Ecotropic MLV *env*. 1 bp mismatch to *Emv2* corrected from original reverse primer sequence. 169 bp product (Yoshinobu et al., 2009).		
eMLV_pol_F	CACTTTGAGGGATCAGGAGCC	21

... Table 2.3 continued ...

Primer	Sequence (5'–3')	Length (nt)
eMLV_pol_R	CTTCTAGGTTTAGGGTCAACACCTGT	26
Ecotropic MLV *pol*. 76 bp product (Lötscher et al., 2007).		
xMLV_F	TCTATGGTACCTGGGGCTC	19
xMLV_R	GGCAGAGGTATGGTTGGAGTAG	22
Xenotropic MLV *env*. 201 bp product (Yoshinobu et al., 2009).		
pMLV_F	CCGCCAGGTCCTCAATATAG	20
pMLV_R	AGAAGGTGGGGCAGTCT	17
Polytropic MLV *env*. 163 bp product (Yoshinobu et al., 2009).		
mpMLV_R	CGTCCCAGGTTGTATAGAGG	20
Modified polytropic MLV *env*. mpMLV_R used with pMLV_F to give 151 bp product (Yoshinobu et al., 2009).		
MusD_F	GTGCTAACCCAACGCTGGTTC	21
MusD_R	CTCTGGCCTGAAACAACTCCTG	22
175 bp product (Karimi et al., 2011).		
IAP_F	AAGCAGCAATCACCCACTTTGG	22
IAP_R1	CAATCATTAGATGTGGCTGCCAAG	24
IAP_R2	CAATCATTAGATGCGGCTGCCAAG	24
IAP_R1 and IAP_R2 to be mixed 1:1 for use. 97 bp product (Karimi et al., 2011).		
GLN_F	CGTAAGGACCCTAGTGGCTG	20
GLN_R	GCACTCACTCTTCTTCACTCTG	22
155bp product (Karimi et al., 2011).		
MuERV-L_F	ATCTCCTGGCACCTGGTATG	20
MuERV-L_R	AGAAGAAGGCATTTGCCAGA	20
50 bp product (Macfarlan, Gifford, Agarwal et al., 2011).		
Emv2_AC_F	CCTGGGTTTGCGGAAATGGCAC	22
Emv2_GG_F	CCTGGGTTTGCGGAAATGGCGG	22
Emv2_mut_R	TTTGGCGTAGCCCTGCTTCTCG	22

... Table 2.3 continued ...

Primer	Sequence (5'–3')	Length (nt)
Primers to distinguish between transcripts corrected at G-3576-C and deriving from RARVs (Emv2_GG_F and Emv2_mut_R) or uncorrected transcripts (Emv2_AC_F and Emv2_mut_R) deriving from *Emv2* itself. 192 bp product for both reactions.		
eMLV / RARV DNA amplification primers		
RARV_pol_F	ATCGGGCCTCGGCCAAGAAAG	21
RARV_pol_R	CCGGGAGAGGGAGTAAGGTGGC	22
Primers used to amplify the region surrounding the *Emv2* inactivating point mutation. 708 bp product.		
RARV_5-pol_F	GCGCCAGTCCTCCGATAGACT	21
RARV_5-pol_R	CCGGGAGAGGGAGTAAGGTGGC	22
Primers used to amplify the first half of RARV genomes for subsequent sequencing. 4074 bp product.		
RARV_pol-3_F	ATCGGGCCTCGGCCAAGAAAG	21
RARV_pol-3_R	TGCAACAGCAAAAGGCTTTATTGG	24
Primers used to amplify the second half of RARV genomes for subsequent sequencing. 4909 bp product.		

2.3 RNA expression assays

Tissue samples were disturbed through vibration with 3 mm tungsten carbide beads in a TissueLyser LT machine (Qiagen, Hilden, Germany). Total RNA was isolated with TRI-reagent (Sigma-Aldrich, St Louis, US) according to the manufacturer's instructions, precipitated with isopropanol (Fisher Scientific, Loughborough, UK) and washed in 75% ethanol (Fisher Scientific). DNase digestion and cleanup was performed with the RNeasy Mini Kit (Qiagen) and RNase-free water (Qiagen) was used throughout the protocol.

cDNA was produced with the high capacity reverse transcription kit (Applied Biosystems, Carlsbad, US) with an added RNase-inhibitor (Promega Biosciences, Madison, US). cDNA was cleaned with the QIAquick PCR purification kit (Qiagen) and eluted with nuclease-free water (Qiagen).

To prevent 'contamination' with other endogenous MLV sequences in the relative

quantitation of corrected and uncorrected eMLV transcripts, cDNAs were instead reverse transcribed using a single primer specific to the eMLV *env*. Samples were split in half and reverse transcribed using both the high capacity reverse transcription kit (Applied Biosystems) and the Omniscript RT kit (Qiagen) using the primer 5'-TTCTGGACCACCACACGAC-3'.

Purified cDNA was used as template for the amplification of target gene transcripts with SYBR Green PCR Master Mix (Applied Biosystems). PCR protocols were run on an ABI Prism SDS 7000 or 7900HT (Applied Biosystems) cycler. Target gene expression values were calculated against the housekeeping gene *Hprt* using the ΔCT method:

$$\text{Value} = 2^{(\text{Hprt } C_T - \text{target } C_T)} \text{ x } 10^4$$

This was chosen such that for a representative C_T value for *Hprt*, a C_T of 40 (the length of a standard PCR program) determined for a target gene would have a value of 1 arbitrary unit. Related to this, a theoretical detection limit of 2 arbitrary units is also displayed as dashed horizontal line on graphs of qRT-PCR data. For eMLV expression in particular, values above 10^3 were considered to be elevated above background levels of variation.

For transcriptional profiling with microarrays, purified RNA samples were firstly checked for quality using the Agilent bioanalyzer (Agilent, Santa Clara, US). Synthesis of cDNA, probe labelling, and hybridisation were performed using the Mouse Gene 1.0 ST oligonucleotide microarray kit (Affymetrix, Santa Clara, CA, USA). GeneChips were scanned with an Affymetrix GeneChip Scanner and analysed using GeneSpring GX (Agilent).

2.4 DNA extraction and assays

DNA was extracted from small tissue samples, cell suspensions and cell lines by Proteinase K digestion in a buffer of 100 mM TRIS, 5mM EDTA, 200 mM NaCl, and 0.2% SDS. DNA was precipitated with isopropanol and washed with ethanol before re-suspension. Larger tissue samples were similarly digested, but DNA was extracted with phenol prior to precipitation with isopropanol and ethanol washing.

Determination of eMLV copy number was performed by qRT-PCR on DNA isolated from tissues of interest. Copy number was calculated with the ΔΔCT method using *Ifnar1* as a reference, using DNA from the spleen of a B6 mouse as a control (assigned a value of 1 copy/haploid genome, N). The same DNA samples were used for array comparative genome hybridisation analysis, performed by Atlas Biolabs (Berlin, Germany) using SurePrint G3 Mouse CGH 4x180K microarray kits (Agilent).

2.5 Sequencing and sequence analysis

PCR products were purified with the QIAquick PCR purification kit (Qiagen) and sequenced at Source BioScience (Cambridge, UK). Sequence analyses, comparisons and alignments were performed with Vector NTI v11.5 (Invitrogen, Carlsbad, US).

RARV contigs were aligned against B6 MLVs (Jern, Stoye & Coffin, 2007) using MAFFT (Katoh, Misawa, Kuma & Miyata, 2002) within UGENE (Okonechnikov, Golosova & Fursov, 2012). Distance plots were calculated with RDP (Recombination Detection Program) v4.16, using a 100 bp scanning window and a 10 bp shift (D. P. Martin et al., 2010). RARV phylogenetic analyses were performed with PHYLIP (Phylogeny inference package) v3.2 within UGENE.

For assessment of microbial diversity, colonic content was collected from culled mice either within NIMR or at RCHCI (frozen samples were shipped to NIMR for further processing). DNA was isolated using the QIAamp DNA Stool Mini Kit (Qiagen). DNAVision (Gosselies, Belgium) performed high-throughput sequencing of amplicons of bacterial DNA encoding the 16S ribosomal RNA with a Roche FLX Genome Sequencer and conducted read assignation. Validity of assignation was confirmed by comparison to assignation performed on the Ribosomal Database Project website (rdp.cme.msu.edu) (Cole et al., 2009).

2.6 Media and culture conditions

Room temperature cell incubation, antibody staining, FACS sorting, and tissue collection was performed with air-buffered Iscove's Modified Dulbecco's Medium (AB IMDM) containing 25 mM HEPES buffer and L-glutamine, and supplemented with 0.21% NaCl, 60 μg/ml penicillin, and 100 μg/ml streptomycin (Invitrogen).

Erythrocyte lysis was performed with Ammonium-Chloride-Potassium (ACK) buffer (0.15 M NH_4Cl, 1 mM $KHCO_3$, 0.1 mM EDTA, pH 7.2-7.4).

Cells were cultured with Iscove's Modified Dulbecco's Medium (Sigma-Aldrich) supplemented with 5% heat inactivated foetal calf serum (FCS) (BioSera, Ringmer, UK), 2 mM L-glutamine, 100 μg/ml penicillin, 100 μg/ml streptomycin, and 10^{-5} M mercaptoethanol. Cell lines and primary cells for longer study were incubated in an atmosphere of 95% humidity, 5% CO_2, and 37°C.

Unmodified Dulbecco's phosphate-buffered saline (DPBS with no added Ca or Mg) (Gibco/Invitrogen) was used for dilution of viral stocks prior to inoculation.

2.7 Isolation of primary cells

Single cell suspensions were prepared by mechanical disruption of tissues through a 70 μm cell strainer (Falcon/BD, Franklin Lakes, US). Splenocyte suspensions were subjected to erythrocyte lysis by brief ACK treatment, before being washed and re-suspended in AB IMDM.

Peritoneal macrophages were collected by injecting 5 ml AB IMDB at 4°C into the peritoneal cavity of culled mice. The peritoneum is massaged to disturb loosely adherent cells and the fluid aspirated into the same syringe.

Bone marrow cells were flushed from the femurs and tibiae of culled mice with cell culture medium. For the production of bone marrow-derived dendritic cells (BM-DCs), bone marrow cell suspensions were supplemented with 10% granulocyte macrophage colony-stimulating factor (GM-CSF) obtained from supernatants of X63 cells transfected with mouse *Csf2*. 5x10^6 cells were plated in 10 cm cell culture plates and incubated for a period of 7 days. At this point, dendritic cells could typically be obtained at a purity of 40–70%.

Enrichment of primary cells was performed using EasySep immunomagnetic positive selection (StemCell Technologies, Vancouver, Canada) according to the manufacturer's instructions. Briefly, single cell suspensions were produced at a concentration of 1x10^8 cells/ml and stained with a relevant phycoerythrin (PE)-conjugated antibody for 30 minutes at 4°C. Cells were washed and a PE selection cocktail was added at 25 μl/ml. Cells were incubated at room temperature for 15 minutes and magnetic beads added at 25 μl/ml. Cells were incubated at room temperature for a further 10 minutes and the tube containing the cells placed on an EasySep magnet (StemCell Technologies). Cells are left for 7 minutes to ensure those bound with magnetic beads are attracted to the sides of the tube. For positive selection of target cells, the supernatant, containing cells not bound to magnetic beads, was poured off and captured cells re-suspended. This selection was repeated twice more, balancing purity against incremental losses of target cells. For negative selection of target cells, the supernatant was collected and replaced on the magnet twice more, with cells captured by the magnet being discarded. Typical purity, as assessed by FACS, of purified populations was greater than 92%.

For adoptive transfer of immunomagnetically-enriched primary cells, 1x10^6 $CD4^+$ T cells were injected into the tail vein of recipient mice in 100 μl IMDM.

2.8 Fluorescence Activated Cell Sorting

Cell suspensions obtained from disrupted tissues or peritoneal exudates were stained with a range of directly fluorochrome-conjugated antibodies against cell-surface markers. Antibodies were obtained from eBiosciences (San Diego, US), CALTAG/Invitrogen, BD Biosciences, or BioLegend (San Diego, US) and are listed in Table 2.4.

Table 2.4: Directly fluorochrome-conjugated antibodies

Specificity	Clone	Fluorochrome
CD4	GK1.5 / RM4-5	APC / APC a750 / FITC / PE / PE TxR
CD8a	53-6.7 / 5H10	APC / FITC / PE / PE TxR
CD11b	M1/70	APC / PE
CD19	MB19-1 / 6D5	APC / FITC / PE / PE TxR
CD25	PC61.5	APC / FITC / PE
CD43	1B11	PE
CD44	IM7	APC / APC a750 / FITC / PE
CD45.1	A20	APC / APC eF780 / FITC / PE
CD45.2	104	APC / APC Cy7 / PE
CD69	H1.2F3	FITC
B220	RA3-6B2	APC / PE / PE TxR
F4/80	BM8	APC / PE
Gr1	RB6-8C5	APC / FITC / PE
Ly6-C	HK1.4	FITC
MHCII	M5/114.15.2	FITC / PE
Ter119	TER-119	APC / PE
TCRβ	H57-597	APC / FITC
Vα2	B20.1	APC / FITC / PE
Vα3.2	RR3-16	FITC
IgG1	A85-1	FITC
IgG2a/c	R19-15	FITC
IgG2b	R12-3	FITC
IgM	R6-60.2	PE

MLV SU staining was determined with use of the 83A25 rat IgG2a antibody (Evans, Morrison, Malik, Portis & Britt, 1990). This primary reagent was stained with a biot-

inylated anti-rat IgG2a secondary antibody (clone RG7/1.30, BD Biosciences), which was in turn treated with a streptavidin conjugated to PE or PE TxR (BioLegend or CALTAG/Invitrogen) for analysis by FACS.

4- and 8-colour cytometry was conducted on FACSCalibur (BD Biosciences), CyAn (Dako, Fort Collins, US), or FACSVerse (BD Biosciences) machines. Data analysis was performed with FlowJo v8.7 (Tree Star, Ashland, US) or Summit v4.3 (Dako) analysis softwares.

Cell suspensions for FACS sorting were maintained on ice or at 4°C and stained with relevant fluorochrome-conjugated antibodies for surface markers. Sorting was performed using MoFlo cell sorters (Dako) by the NIMR Cell Sorting facility, typically yielding cell purities of higher than 98%.

Where necessary to prevent non-specific antibody binding, FcBlock (antibody clone 2.4G2, produced within NIMR) was added to stains. Being a rat IgG2a antibody, this was not used in combination with 83A25 for the determination of MLV SU product levels.

2.9 Isolation and analysis of infectious MLVs

Plasma samples were obtained from mice and centrifuged for 45 mins at an 256 g at 20°C with *Mus dunni* cells transduced with the XG7 replication-defective retroviral vector. This encodes a green fluorescent protein (GFP) under the control of a human cytomegalovirus (CMV) promoter and contains a neomycin-resistance gene under the control of the LTR (Bock, Bishop, Towers & Stoye, 2000). Maintenance of GFP expression was ensured by constant selection with 1 mg/ml G418 antibiotic.

Cells were cultured for 10-14 days and culture supernatant transferred to *Mus dunni* cells. The presence of infectious MLVs was thus assessed by the quantity of pseudotyped XG7 vector in this supernatant, the presence of which causes *Mus dunni* cells to be GFP^+ by FACS after a period of 3 days.

N- or B-tropism of isolated viruses was determined by addition of serially diluted culture supernatants from *Mus dunni* XG7-transduced cells to either B- or N-3T3 cells. Infection was quantified according to the GFP positivity of these cells after 3 days. N- and B-tropic F-MLV were used to provide estimates of infectivity in cell lines of the opposing tropism. Results were expressed as a ratio of the percentage of GFP^+ B-3T3 cells to that of N-3T3 cells.

2.10 Histology

Tissue samples were collected into formalin immediately after culling of mice. Sectioning for histological analysis, staining with haematoxylin and eosin, and examination was performed by Dr Mark Stidworthy (IZVG Pathology, Leeds, UK).

2.11 Electron microscopy

Electron microscopy was performed by the dedicated service provided within NIMR. Samples were immersion fixed in 2% glutaraldehyde/2% paraformaldehyde and post-fixed in 1% osmium tetroxide using 0.1 M sodium cacodylate buffer (pH 7.2). Aqueous uranyl acetate was followed by dehydration through a graded ethanol series and propylene oxide. Samples were then embedded in Epon and 50 nm sections mounted on pioloform coated grids and stained with ethanolic uranyl acetate followed by Reynold's lead citrate. Samples were viewed with a JEOL 100EX transmission electron microscope (JEOL, Tokyo, Japan) equipped with an ORIUS 1000 CCD camera (Gatan, Pleasanton, US).

2.12 Friend Virus

Friend virus (FV) is a complex of a replication incompetent spleen focus-forming virus (SFFV) and Friend murine leukaemia virus (F-MLV), a replication competent helper virus. Stocks used within NIMR contain the polycythemia-inducing SFFV variant ($SFFV_P$). The FV-B complex is propagated for 12 days *in vivo* in a permissive strain (BALB/c) and prepared for use as a 10% w/v spleen homogenate. All viral stocks were free of Sendai virus, Murine hepatitis virus, Parvoviruses 1 and 2, Reovirus 3, Theiler's murine encephalomyelitis virus, Murine rotavirus, Ectromelia virus, Murine cytomegalovirus, K virus, Polyomavirus, Hantaan virus, Murine norovirus, Lymphocytic choriomeningitis virus, Murine adenoviruses FL and K87, and Lactate dehydrogenase elevating virus.

Prepared spleen homogenate is briefly centrifuged to pellet debris, and supernatant diluted 1:5 with DPBS for injection. Experimental mice receive an inoculum of ∼1,000 spleen focus-forming units (SFFU) injected in 100 μl via the tail vein. Where adoptive transfer was combined with FV infection, two separate injections were given at the same time, or else within a 24 hour period.

FV infection was determined by FACS using surface staining for the glycosylated viral *gag* matrix (MA) product (glyco-Gag) using the monoclonal mouse IgG2a 34 antibody (Chesebro, Britt et al., 1983), followed by a secondary anti-mouse IgG2a conjugated

to a fluorochrome. Alternatively, anaemia was assessed by analysis of complete blood counts determined with a VetScan MHII haematology analyser (Abaxis, Union City, US). Blood was collected into heparinised capillary tubes from a small incision of the tail. Red blood cell (RBC) counts of uninfected controls were $\sim$9.95x10^6/mm^3 of whole blood. FV-induced splenomegaly was expressed as spleen index, a ratio of spleen weight (mg) to body weight (g).

2.13 FV-neutralising and cell-binding antibody titre assays

FV-neutralising serum antibody titres were determined *in vitro* using *Mus dunni* cells transduced with the replication-defective retroviral XG7 vector (see Section 2.9). These cells were infected with B-tropic F-MLV and the culture supernatant, containing $\sim$1500 infectious units (iu)/ml of psuedotyped XG7 vector, was harvested. Serial dilutions of sera from infected and control mice were mixed with this supernatant and incubated for 30 mins at 37°C in IMDM +5% FCS. This was finally added to *Mus dunni* cells and incubated for 3 days, after which the percentage of GFP^+ cells was determined by FACS. The dilution of serum resulting in 75% neutralisation was taken as the neutralising antibody titre.

The F-MLV-infected cell-binding serum antibody titre was determined by FACS. Serial dilutions of serum were added to *Mus dunni* cells infected with F-MLV and antibodies binding cells were secondarily bound with fluorescently labelled anti-mouse IgG1 (clone A85-1), anti-mouse IgG2a/c (clone R19-15), anti-mouse IgG2b (clone R12-3), or anti-mouse IgM (clone R6-60.2). Starting from an initial dilution of 1:50, sera were 2-fold serially diluted and the median fluorescence intensity (MFI) of staining was recorded. Data were fitted to a sigmoidal curve, and the last positive dilution on the curve to give staining over twice the level of background was taken as the binding antibody titre. This measure was used, as opposed to determination of titre based on half the maximal response, as no samples showed signs of reaching a maximum plateau.

2.14 Hybridoma production

Cell suspensions were prepared from spleens and lymph nodes of B6 EF4.1 or *Emv2*$^{-/-}$ EF4.1 mice and stimulated in vitro with 10^{-7} M or 10^{-5} M $env_{122\text{-}141}$L peptide and 4 ng/ml recombinant human IL-2 for 4 days. $CD4^+$ T cells were immunomagnetically enriched and plated in 96-well plates with TCR$\alpha\beta$-negative BW5147 thymoma cells to produce hybridoma cell lines. Fusion and hybridoma generation over the following 5 days was assessed by microscopy, with individual clones being selected and transferred

to tissue culture flasks. Hybridomas were selected in culture medium supplemented with Hypoxanthine-Aminopterin-Thymidine (HAT) and weaned into normal culture medium with the use of Hypoxanthine-Thymidine (HT).

2.15 Stimulation assays

F-MLV $env_{124\text{-}138}$L peptide, as well as the Y and YS env mutants, were produced within NIMR. Peptides used for determination of hybridoma TCR specificity, mutating the three TCR contact points through the range of natural amino acids, were produced by GenScript (Piscataway, US).

Single cell suspensions were prepared from spleens, thymi, and lymph nodes, as appropriate. $5x10^5$ cells/well of a 96-well plate were stimulated with peptides at varying concentrations. Responding cells were determined as having upregulated CD69 after 18 hours (before cell division or death) by FACS. Co-staining for CD40L was performed in control experiments to confirm the accurate identification of cells responding to peptide stimulation.

For assessment of T cell activity over a period of 3 days, cells were labelled with CFSE prior to stimulation. CFSE dilution, occurring at cell division, was determined by FACS.

$1x10^5$ hybridoma cells were stimulated with peptides presented by $5x10^4$ B6 BM-DCs in 96-well plates and responsiveness was assessed through determination of IL-2 production. After 18 hours, culture supernatants were added to wells containing the IL-2 dependent CTLL-2 cell line and cell viability was determined after a further 18 hours with the use of AlamarBlue (Invitrogen). Fluorescence at 590 nm (excitation at 530 nm) was determined with a Safire2 plate reader (Tecan Group, Männedorf, Switzerland).

2.16 *Tra* gene usage determination

Proportions of Vα2$^+$ and non-Vα2 CD4$^+$ T cells were determined with anti-Vα2 (B20.1) and anti-Vα3.2 (RR3-16) monoclonal antibodies. Only a proportion of Vα3$^+$ CD4$^+$ T cells stain with RR3-16, however, hence accurate determination of *Trav* and *Traj* region usage was determined by qRT-PCR amplification of expressed genes and subsequent sequencing using primers from the literature (Casanova, Romero, Widmann, Kourilsky & Maryanski, 1991). Identification of gene usage from sequence data and confirmation that rearrangements were capable of producing functional TCRs were performed on the International Immunogenetics Information System website (imgt.org).

2.17 Computational and statistical methods

Local installs of NCBI BLAST+ 2.2.27 were used to create BLAST databases and perform BLASTn. Python 3.2.2 was used to produce scripts to create, handle, and query multi-process BLASTn.

Statistical analysis and graphing was performed within SigmaPlot 12 (Systat Software, Erkrath, Germany). Where normally-distributed data satisfied variance criteria, comparisons were performed with unpaired Student's *t*-tests. Where data were not normally-distributed or failed variance criteria, non-parametric two-tailed Mann-Whitney Rank Sum or Wilcoxen Signed Rank test were used instead. Assessment of bacterial diversity was performed by ANOVA with Bonferroni multiple comparison correction. Tumour incidence rates were compared by log-rank survival analysis of Kaplan-Meier curves.

Microarray data was separately analysed within GeneSpring GX (Agilent) using Benjamini-Hochberg false discovery rate (FDR) correction for multiple comparisons.

2.18 *REquest*: determining retroelement expression with microarrays

RNA microarrays, similarly to Northern blotting, rely on nucleotide complementarity to hybridise specific sequences from within a sample of interest. A sample to be interrogated is labelled with a fluorescent marker before being allowed to hybridise to fixed complementary cDNA probes. Captured nucleotide fragments are retained in place upon washing, allowing their abundance to be quantified by relative fluorescence. The first publication of this technique monitored the expression of 45 *Arabidopsis thaliana* cDNAs (Schena, Shalon, Davis & Brown, 1995), using an automated gridding robot to print cDNA probes onto a glass slide. This technique, the 'spotted microarray', was developed rapidly to accommodate over 1,000 individual probes (DeRisi et al., 1996), but separate development of *in situ* methods of probe synthesis (Fodor et al., 1991) allowed for massive parallelisation and the rapid production of microarrays with over 16,000 and 65,000 separate probes (Lockhart et al., 1996).

The necessity of RNA intermediates in the replication of REs and ERVs allows assessment of their transcriptional activity by both PCR-based approaches and RNA microarray. Significant labour has resulted in the production of spotted microarrays specifically for the assessment of retroviral activity (Seifarth, Spiess et al., 2003). The requirement for custom equipment, technologies, and analysis techniques prevents their application in the majority of exploratory settings, however, as use requires either pre-

existing or *a priori* knowledge. In the absence of specific targeting, the high levels of sequence specificity required in the design of microarray probes (He, Wu, Li, Fields & Zhou, 2005) also prevents the determination of expression for more diverse or distantly related members of various RE families. Further, the expense and expertise required makes these tools available to only a small subset of researchers.

In comparison, commercial microarray platforms are frequently used in exploratory settings and across varying research areas, ranging from developmental biology to immunology. In addition, increasing numbers of species can be investigated in this manner. The possible existence of probes within commercial microarrays that report RE or ERV expression is, therefore, a tempting prospect, and would allow researchers to monitor or identify expression patterns in parallel with those of cellular genes.

Successful identification and annotation of RE and ERV-specific microarray probes relies on the accurate and unambiguous identification of repetitive sequences within the genome of interest. This can either be done *de novo*, with predictive algorithms and self-alignment techniques, or by comparison against a compiled database of elements. Database-reliant programs such as *RepeatMasker* (Smit, Hubley & Green, 1996) and *CENSOR* (Kohany, Gentles, Hankus & Jurka, 2006) are capable of identifying far larger total areas of repetitive sequences and can be used to 'mask' genomes to facilitate faster indexing and unambiguous searching.

Comparison of the genome co-ordinates of microarray probes between masked and unmasked sequences can thus be used to identify probes corresponding to REs, and significant numbers of probes have previously been identified in this way (Barbosa-Morais et al., 2010). However, the majority of work has centred on the exclusion of these probes as confounding factors, rather than on their specific analysis. An extension of this theory has allowed the identification and analysis of RE-specific probes within three popular microarray platforms (Reichmann et al., 2012). ∼3% of probes within the Illumina Mouse WG-6 v2.0 platform potentially reported RE expression, and, respectively, ∼2% and ∼5% of probes within the Affymetrix Murine Genome U74Av2 and Mouse Expression 430 2.0 microarrays (Reichmann et al., 2012).

This methodology, whilst successfully identifying differentially regulated REs from within microarray datasets, lacks fine annotation of individual REs, and clusters elements at a relatively high level. For example, an element found to correspond to *RLTR-int*, *MuLV-int* and *RLTR4_Mm* sequences (Reichmann et al., 2012) can be recognised as the single-copy ecotropic MLV *Emv2*. Probes reporting *Emv2* expression were correctly identified as *Mela*-specific (*Melanoma antigen*, corresponding to the *Emv2 env* mRNA, NCBI GeneID 17276) in the manufacturer's annotations and hence re-annotation in this instance resulted in the loss of valuable information.

The published methodology has thus been extended to further microarray platforms

(Table 2.5), but also improved to allow the annotation, where possible, of probes corresponding to low copy number RE classes. The use of commercially available microarrays in this manner not only facilitates the discovery of novel associations and future translational research, but also allows the enhanced re-analysis of previously obtained data and informative data-mining from thousands of publicly available microarray experiments.

Table 2.5: Mouse microarray platforms

Microarray	Total probes	Probe length (nt)
Affymetrix (affymetrix.com)		
Murine Genome U74Av2	~ 200,000	25
Mouse Genome 430A 2.0	~ 250,000	25
Mouse Genome 430 2.0	~ 500,000	25
Mouse Gene 1.0 ST	~ 900,000	25
Mouse Exon 1.0 ST	~ 4,780,000	25
Agilent (genomics.agilent.com)		
Mouse GE 44K v2	~ 45,000	60
SurePrint G3 Mouse GE 60K	~ 62,000	60
SurePrint G3 Mouse Exon 180K	~ 181,000	60
SurePrint G3 Mouse Exon 400K	~ 420,000	60
Illumina (illumina.com)		
Mouse Ref-8 v2.0	~ 25,000	50
Mouse WG-6 v2.0	~ 46,000	50
Operon (operon.com)		
Mouse AROS v4	~ 36,000	50–70
Exonic Evidence Based Oligonucleotide (microarray.com)		
MEEBO	~ 39,000	70

The latest version of the mouse RefSeq (Pruitt, Tatusova & Maglott, 2007) genome (GRCm38, Mm10) was downloaded, along with the latest RepBase Update database (Jurka et al., 2005). *RepeatMasker* was used to identify repetitive elements within the genome sequence, with 46.4% in total being masked. Probes from one previously studied microarray platform, Illumina Mouse WG-6 v2.0, were downloaded as a test set and

a Python script was produced to localise probes to the genome through BLASTn and record the total number of sites sharing ≥95% homology with the probe. A further Python script was produced to determine if individual probe co-ordinates were contained within sequence regions masked by *RepeatMasker*, as described previously (Reichmann et al., 2012), and lists of probes identified were compared to those published.

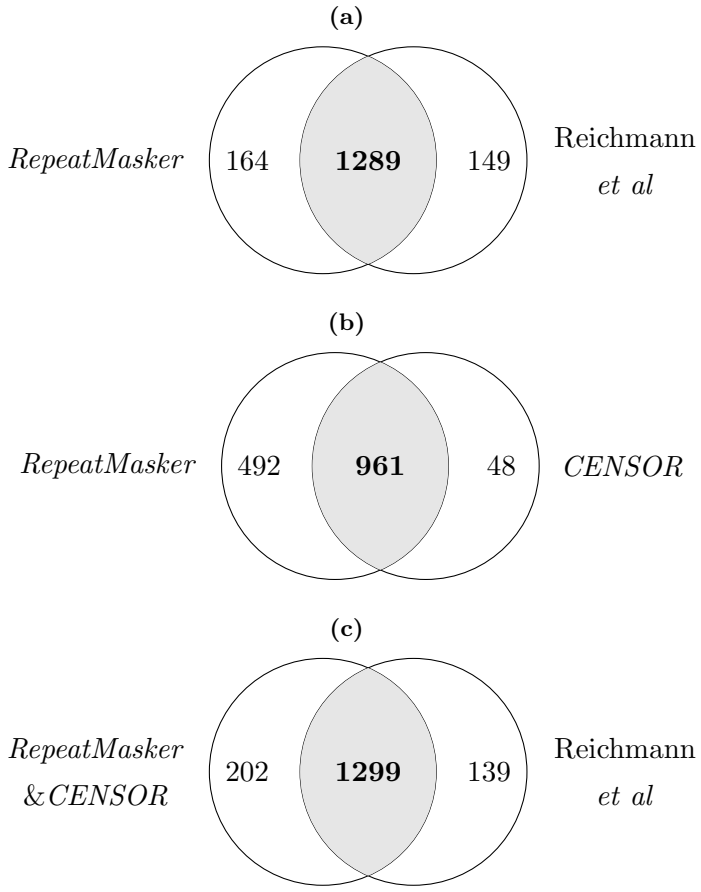
This workflow identified marginally more elements in total (1453 compared to 1438 – Fig. 2.2a), with broadly similar overall frequencies (data not shown). Overall, ∼90% of probes identified in this workflow were identical to those described previously (Fig. 2.2a). Much of the difference is likely explained by the combined use of a more recent RepBase Update library, the latest genome assembly and potentially varying program settings. To investigate the latter possibility, the analysis was repeated using *CENSOR* to produce the required masking data. *CENSOR* masked a lower overall percentage of the genome (39.1%), which resulted in identification of fewer total probes, although overall correspondence with those identified using *RepeatMasker* was high (Fig. 2.2b). Combination of probes identified with the use of either program improved overall identification and reduced discrepancy to those identified previously (Reichmann et al., 2012) (Fig. 2.2c).

∼81% of probes identified within this microarray were likely to be specific to individual elements within the genome, with those remaining sharing ≥95% homology with a median of 12 sites. The methodology outlined, using masking data from both *RepeatMasker* and *CENSOR*, was applied to a number of popular microarray platforms (Table 2.5) and numbers of identified LTR-bounded elements, SINEs, and LINEs are shown in Table 2.6.

Further, microarray probes were additionally screened for specific reactivity to a library of MLVs and MMTVs endogenous to the B6 genome and compiled from the literature (Jern, Stoye & Coffin, 2007) and genome searches. Individual elements within B6 mice were identified in this manner, but these elements are highly polymorphic (Stoye & Coffin, 1988; Frankel et al., 1990) and thus mappings are non-transferable between strains. Further work could, however, apply this method to the genomes of other sequenced mouse strains (Yalcin et al., 2011).

This workflow has also applied to a number of human commercial microarrays and together with probes identified for murine microarrays, a Python tool, *REquest*, was developed to allow the mining of RE and ERV expression from commercial microarray data. After normalisation routines and determination of significantly up- or down-regulated probes, *REquest* can be applied to comma separated value (csv) lists of probes to identify those reporting the expression of REs and ERVs, allowing the user to specifically investigate their expression patterns in relation to various treatments or conditions.

Figure 2.2: RE-specific probe identification



(a) Comparison of probes identified previously (Reichmann et al., 2012) with those identified using *RepeatMasker*. (b) Probes identified with masking data from *RepeatMasker* or *CENSOR*, and (c), probes identified using both *RepeatMasker* and *CENSOR* in comparison to those identified previously (Reichmann et al., 2012).

Table 2.6: RE- and ERV-specific probes within commercial murine microarrays

Microarray	**LTR**	**SINE**	**LINE**
Affymetrix (affymetrix.com)			
Murine Genome U74Av2	1625	507	1424
Mouse Genome 430A 2.0	2341	841	2590
Mouse Genome 430 2.0	8580	4431	11803
Mouse Gene 1.0 ST	4294	2229	4528
Mouse Exon 1.0 ST	65787	37480	48308
Agilent (genomics.agilent.com)			
Mouse GE 44K v2	639	233	323
SurePrint G3 Mouse GE 60K	709	256	336

... Table 2.6 continued ...

Microarray	**LTR**	**SINE**	**LINE**
SurePrint G3 Mouse Exon 180K	216	69	83
SurePrint G3 Mouse Exon 400K	1114	303	378
Illumina (illumina.com)			
Mouse Ref-8 v2.0	132	65	147
Mouse WG-6 v2.0	453	305	556
Operon (operon.com)			
Mouse AROS v4	425	209	199
Exonic Evidence Based Oligonucleotide (microarray.com)			
MEEBO	225	103	59

Chapter 3

Endogenous retroviruses and T cell development

3.1 Introduction

3.1.1 Friend virus

Charlotte Friend first described a retroviral agent responsible for transmission of leukaemia between mice in 1957 (Friend, 1957; de Harven & Friend, 1960), showing that the acute erythroblastosis seen on inoculation of newborn mice with Ehrlich murine carcinoma cells could subsequently cause disease in adult mice through serial transfer of cell-free splenic filtrates.

Friend originally described a virus that caused the development of anaemia, anaemia-inducing Friend virus (FVA); however, after some years of research a new disease characteristic was observed in some viral isolates. Rather than the development of anaemia, infection resulted in polycythemia, and the viral isolate was termed polycythemia-inducing Friend virus (FVP) or, subsequently, spleen focus-forming virus (SFFV).

Both types of virus were seen to target erythroid progenitors to induce an immediate erythroblastosis and massive splenomegaly in susceptible strains of mice. Subsequent research has revealed that Friend virus (FV) exists as a pair of co-transmitted viruses – the replication competent helper virus, now known as Friend MLV (F-MLV), and a replication incompetent SFFV, which may either be anaemia-inducing ($SFFV_A$) or polycythemia-inducing ($SFFV_P$).

SFFV *env* encodes a truncated glycoprotein of 55 rather than 70 kDa, gp55, resulting from the likely recombination of its original ecotropic *env* (a truncated form being retained at the carboxy-terminus of the resulting protein) with an endogenous pMLV sequence (forming the amino-terminus) (Clark & Mak, 1984). gp55, given its unusual character, is retained in the cell membrane of infected cells and has been shown to interact with the erythropoietin receptor (EpoR) (J.-P. Li, D'Andrea, Lodish & Baltimore, 1990). This interaction results in heightened sensitivity to erythropoietin (Epo), in $SFFV_A$, or complete Epo independence in the case of $SFFV_P$. Further, gp55 recruits a short form of the Stk receptor tyrosine kinase, encoded by $Fv2^s$ (Persons et al., 1999), constitutive activation of which allows the survival and proliferation of infected erythroid precursors and the associated erythroblastosis. $Fv2^{r/r}$ mice have a deletion of the promoter for short-form Stk, and are hence resistant to $SFFV_{A/P}$-induced erythroblastosis and splenomegaly.

Although SFFV causes rapid expansion of its target cell population, the resulting cells are not tumourigenic. Thus, erythroleukaemia is induced through a 'two-hit' model, where a second integration is required to cause transformation. SFFV has been shown to cause this through insertional transactivation of *Spi-1*, with a proviral integration being found here in around 95% of tumours studied (Moreau-Gachelin, Tavitian &

Tambourin, 1988). Additionally, subsequent mutation or disruption of the *p53* tumour suppressor gene can contribute to tumour growth (Mowat, Cheng, Kimura, Bernstein & Benchimol, 1985).

3.1.2 CD4$^+$ cells in Friend virus infection

Early work with FV centred around viral immunology and, as early as 1959, Friend had demonstrated effective immunisation against viral challenge through use of a formalised viral preparation (Friend, 1959). Much further work has concentrated on determining the exact roles of the adaptive immune system in the control of infection. Even in mice that mount an effective immune response, however, FV is never completely cleared, and a low-level of chronic infection is established for the life of the host (Chesebro, Bloom, Wehrly & Nishio, 1979). Thus, FV mirrors HIV infection and, as such, presents a valuable mouse model for its study.

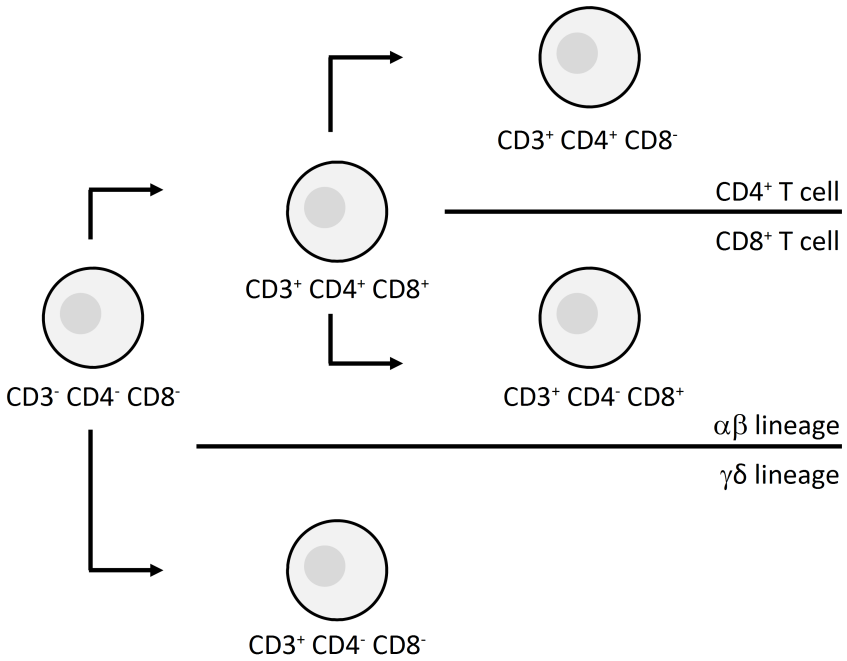
CD4$^+$ T cells act to control both CD8$^+$ T cell and B cell responses, but have also been shown to have direct anti-viral effects (Hasenkrug, Brooks & Dittmer, 1998; Iwashiro, K. Peterson, Messer, Stromnes & Hasenkrug, 2001; Pike et al., 2009; Nair et al., 2010). Their importance is borne-out by the relevance of MHC alleles in determining the outcome of infection – resistance in adult mice is strongly associated with the dominant *H2-A*^{b} allele (Miyazawa, Nishio & Chesebro, 1988; Hasenkrug & Chesebro, 1997). Mice carrying *H2-A*^{b}, such as the B6 strain, mount strong CD4$^+$ T cell responses, class switch FV-specific antibodies and respond strongly to vaccination regimens. In comparison, mice carrying *H2-A*$^{k/k}$, a low recovery allele, lack a strong CD4$^+$ T cell response and fail to class switch FV-specific antibodies, resulting in lack of responsiveness to vaccination and poor control of infection (Hasenkrug & Chesebro, 1997).

The immunodominant *H2-A*^{b}-restricted F-MLV antigen has been determined as the env$_{122\text{-}141}$ region (DEPLTSLTPRCNTAWNRLKL, Fig. 3.2) (Iwashiro, Kondo et al., 1993; Shimizu et al., 1994) and previous work has resulted in the generation of a transgenic mouse carrying a TCRβ chain from a CD4$^+$ T cell clone generated against this peptide (Antunes et al., 2008). B6 EF4.1 mice have a fixed TCRβ chain that pairs with endogenous TCRα chains to form a polyclonal CD4$^+$ T cell repertoire with elevated numbers of env$_{122\text{-}141}$-specific CD4$^+$ T cells ($\sim$4% of total). Individual CD4$^+$ T cells carrying the endogenous Vα2 chain were identified as having above average specificity and affinity for their cognate antigen (Antunes et al., 2008) and expand to be the dominant subset at the peak of FV infection (Ploquin, Eksmond & Kassiotis, 2011). Even in the absence of CD8$^+$ T and B cells, adoptive transfer of purified B6 EF4.1 CD4$^+$ T cells alone has been shown to have a significant protective effect *in vivo*, reinforcing the vital importance of CD4$^+$ T cell-mediated protection in the FV model (Pike et al., 2009).

3.1.3 T cell development and the prevention of autoimmunity

Prior to release into the periphery, developing T cells undergo both positive and negative selection. Somatic recombination of variable, diverse, and joining gene regions (V(D)J recombination) creates separate α and β chains that pair to form the heterodimeric TCR (see 1.2.2). This process results in a wide array of α and β chains, the majority of which are unable to bind peptide-MHC complexes when paired, or even produce a functional TCR. Development of thymocytes progresses through a number of discernible stages designed to remove these cells and to prevent the development of potentially auto-reactive clones (Fig. 3.1).

Figure 3.1: Thymocyte development



A schematic of T cell development in the thymus, showing $\alpha\beta$ or $\gamma\delta$ lineage choice and division of the $\alpha\beta$ lineage to $CD4^+$ and $CD8^+$ T cells.

Within the $\alpha\beta$ lineage, 'double positive' ($CD3^+CD4^+CD8^+$) thymocytes are committed to programmed cell death unless sufficient signalling through TCR binding is achieved, along with successful binding of either a CD4 or CD8 co-receptor. 13-17-mers and 8-9-mers of self-peptide are presented on MHC class II and MHC class I, respectively (Rudensky, Preston-Hurlburt, Hong, Barlow & Janeway, 1991) (Fig. 1.5). This process of positive selection ensures only TCR heterodimers capable of recognising peptides bound to the host's MHC alleles will continue development. Surviving thymocytes cease expressing one of either the CD4 or CD8 co-receptors, becoming 'single positive'.

Whilst there is also control of the peripheral T cell response through $CD4^+$ regulatory T cells (Treg cells – Fig. 1.6) (Zheng & Rudensky, 2007; Wing & Sakaguchi, 2010), removal of self-reactive thymocytes is highly important. Thus, thymocytes having strong

reactivity to presented self-peptides are prompted to apoptose prior to the release of mature naïve T cells and antigen specific responses are encouraged (Huseby, Crawford, White, Kappler & Marrack, 2003).

Both positive and negative selection of thymocytes rely on their exposure to self peptides, therefore. Many proteins have highly tissue- or temporally-specific expression patterns, however, and comparatively few are truly ubiquitous. To prevent potential autoimmune responses to these proteins, the AIRE (autoimmune regulator) protein functions as a powerful transcriptional transactivator to promote the expression of 200–1,200 tissue-specific genes, protein formation and epitope presentation to developing thymocytes (Anderson et al., 2002; Derbinski et al., 2005; Taniguchi & Anderson, 2010). This allows the negative selection of highly reactive TCRs, with mutations impacting AIRE functionality causing diseases associated with autoimmune pathology (Conteduca et al., 2010).

3.1.4 Endogenous retroviruses in T cell selection

Strong $CD4^+$ and $CD8^+$ T cell responses are known to be directed against retroviral antigens in a variety of instances, including in the CT26 colorectal carcinoma and the B16 melanoma (White, Roeder & Green, 1994; Coppola, Lam, Strawbridge & Green, 1995; A. Y. C. Huang et al., 1996; Rosato et al., 2003) and immunisation against specific viral epitopes has been shown to encourage tumour rejection in certain cases (Kershaw et al., 2001; Specht et al., 1997; Slansky et al., 2000).

When considered with regard to the immune system, ERVs can be seen in two lights: firstly as virally derived sequences (in some cases retaining replication competency and infectious potential) and secondly as 'self'. In the latter case, with around six times the number of proviruses than cellular genes in the mouse genome (see 1.1.3), if even a small number maintain partial open reading frames (ORFs), a large number of peptides may potentially be produced that could be presented to developing thymocytes.

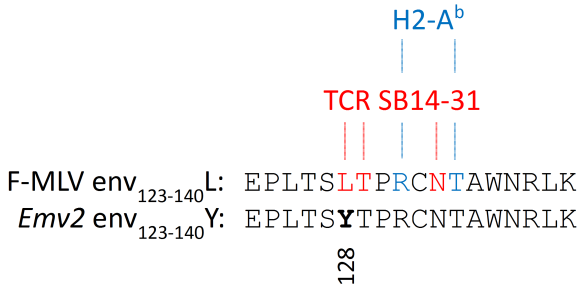
Thus, ERVs have the potential to strongly impact T cell development through either positive or negative selection of thymocytes and the imposition of biases in TCRα or β chain usage. Bias in TCR V region selections are seen frequently in immunity and variations can often be seen to correlate with differences in protective capacity (Turner, Doherty, McCluskey & Rossjohn, 2006; Bridgeman, Sewell, Miles, Price & Cole, 2011).

3.1.5 *Emv2*

A screen of the B6 proteome for peptides sharing homology with the F-MLV $env_{122\text{-}141}$ region reveals the single endogenous eMLV *Emv2*. This provirus shares 80% homology

with F-MLV at the nucleotide level and is highly similar to F-MLV within the critical $env_{122\text{-}141}$ region (Fig. 3.2).

Figure 3.2: Sequence and contact residues of the immunodominant H2-A^b-restricted F-MLV epitope



Amino acid sequence of the immunodominant *H2-A*^{b}-restricted F-MLV epitope, top, showing the central $env_{123\text{-}140}$ region. Marked are *H2-A*^{b} contact residues (blue) and contact points for the donor TCRβ chain from the SB14-31 hybridoma (red). The corresponding epitope from *Emv2* is shown underneath, with the single amino acid substitution in bold and the amino acid position indicated.

Emv2 is a complete eMLV provirus located on chromosome 8 of B6, B6-derived and closely related strains (Jenkins et al., 1982; McCubrey & Risser, 1982; Kozak & W. P. Rowe, 1982). Whilst not annotated in the reference genome assembly, the position of *Emv2* can be determined by BLASTn search with probes for the ecotropic *env* (Young, Kassiotis & Stoye, 2012). Although retaining all open reading frames, *Emv2* is inactivated by a single point mutation impacting the viral reverse transcriptase (G-3576-C within *pol*) (M. Li, Huang, Zhu & Gorelik, 1999). Further, *Emv2* encodes an N-tropic capsid protein that would be restricted by the action of Fv1^b in B6 mice.

3.1.6 Aims

Previous work has shown that the high-avidity of Vα2 CD4$^+$ T cells in the B6 EF4.1 system (seen both at a polyclonal (Antunes et al., 2008) and clonal level (Young, Ploquin et al., 2012)) is due to interactions within the core epitope region, rather than stabilising or alternate interactions with flanking sequences, as suggested for some other TCR-peptide-MHC interactions (Carson, Vignali, Woodland & Vignali, 1997). As such, reactivity is retained to variously N-truncated epitopes (Young, Ploquin et al., 2012).

The mechanism causing the observed expansion and dominance of Vα2 CD4$^+$ T cells upon FV infection is unknown, however. This bias may be introduced through a proliferative advantage or a system of favoured survival, but may also result from thymic selection. The potential role of ERVs in determining responsiveness to retroviral in-

fection is a largely unstudied area, yet due to their similarity, these elements have a large potential to impact the course of thymocyte development when expressed as self antigen. In support of this theory, peripheral B6 EF4.1 $CD4^+$ T cells have a naïve phenotype (Antunes et al., 2008), suggesting a potential role for ERV-mediated thymic deletion of cells bearing otherwise auto-reactive TCR heterodimers.

This project thus aimed to outline the contribution of the endogenous eMLV *Emv2*, if any, to FV immunity using the B6 mouse model and the pre-established B6 EF4.1 transgenic system. Ultimately, this was to investigate the mechanism of TCRα chain selection and usage, and to determine the origin of the high-avidity $CD4^+$ T cell response seen in B6 mice.

3.2 Results

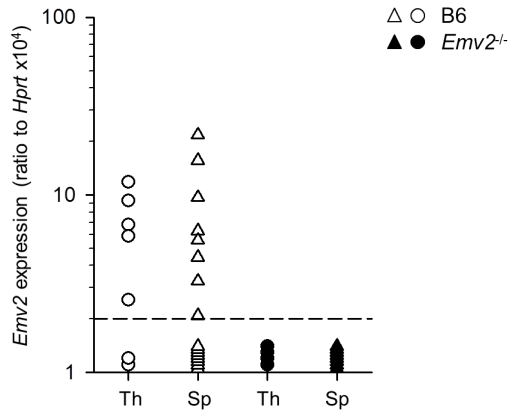
3.2.1 Characterisation of *Emv2* expression and generation of *Emv2* null congenics

Prevention of peripheral auto-reactivity through thymic selection necessitates the expression of tissue-specific self peptides in epithelial cells of the thymic medulla (mTECs). qRT-PCR can be used to assess the expression of defined transcripts, but has the potential to overlook instances of transient expression that may occur either naturally or through changes in expression mediated by infection or environmental factors. Thus, the observed absence of a particular transcript does not exclude the possible involvement of the gene in thymic selection events.

Emv2, being the only eMLV of B6 mice, is well placed for study with congenic systems, and B6 background *Emv2*$^{-/-}$ congenic mice were generated by crossing to the A strain (see Section 2.1.1). These mice were also subsequently crossed to the B6 EF4.1 line to generate *Emv2*$^{-/-}$ EF4.1.

Thymic expression of *Emv2* was determined by qRT-PCR using primers designed to distinguish between mRNA and contaminating genomic DNA (see Section 2.2) and using *Emv2*$^{-/-}$ mice to assess background amplification levels (Fig. 3.3). Expression was low, where detected, and was seen at a similar frequency in secondary lymphoid tissue from B6 mice, as well as in the liver and lung (data not shown). No expression was seen in the same tissues from *Emv2*$^{-/-}$ mice (Fig. 3.3). This was indicative of low level uniform expression in B6 mice, including expression in the thymus.

Figure 3.3: *Emv2* expression in B6 and *Emv2*$^{-/-}$ mice

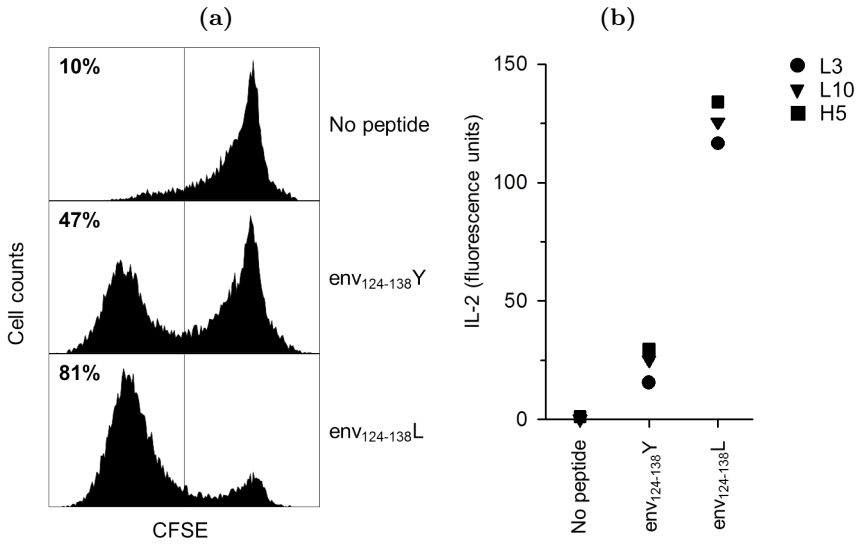


Comparison of *Emv2* expression in the thymi (Th) and spleens (Sp) of B6 and *Emv2*$^{-/-}$ mice ($p<0.001$ for both comparisons). Each symbol represents an individual mouse. Dashed line indicates detection limit.

3.2.2 Deletion of $env_{122\text{-}141}$L-specific $CD4^+$ T cells by *Emv2*

Even where mRNA is transcribed, however, a range of controls surround protein production and many of these may specifically target mRNAs resulting from RE and ERV transcription. Hence separate work by colleagues investigated the potential role for *Emv2* in thymic selection of $env_{122\text{-}141}$-specific $CD4^+$ T cells by assessment of reactivity to the core $env_{124\text{-}138}$L or Y epitopes (Fig. 3.2). Polyclonal populations of B6 EF4.1 $CD4^+$ T cells (Fig. 3.4a) and individual hybridoma cell lines derived from these cells (Fig. 3.4b) were activated when treated with $env_{124\text{-}138}$Y, although to a lesser extent than when presented with their cognate antigen, $env_{124\text{-}138}$L. Assessment of individual hybridoma clones indicated that whilst the difference in activation potential was around 5-fold (Fig. 3.4b), individual TCR heterodimers could respond to both peptide epitopes.

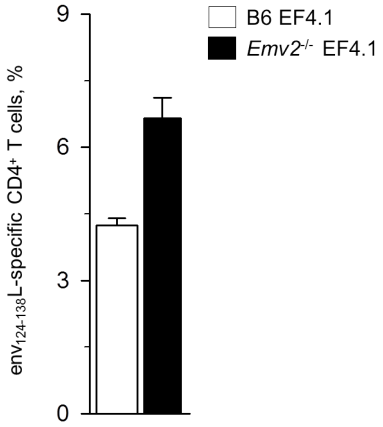
Figure 3.4: Clonal and polyclonal responses to $env_{124\text{-}138}$



In vitro determination of responsiveness to $env_{124\text{-}138}$L or Y. (a) FACS analysis of CFSE dilution in primary B6 EF4.1 $CD4^+$ T cells incubated for three days in the presence of absence of 10^{-5} M peptide. Numbering denotes percentages of cells determined as $CFSE^-$ and are averages of 4 mice per condition. (b) Fluorescence quantitation of AlamarBlue, representing the IL-2 production of 3 hybridoma lines generated in response to stimulation with $5x10^{-6}$ M $env_{124\text{-}138}$L peptide. $p<0.001$ between $env_{124\text{-}138}$L and Y stimulation, and $p=0.002$ between $env_{124\text{-}138}$Y and unstimulated control.

The retention of reactivity to both $env_{124\text{-}138}$L and Y, despite the observed thymic expression of *Emv2*, suggested that any *Emv2*-mediated deletion was not causing complete tolerance to either epitope, and that any impacts were likely to be more complex. Initial assessment of the polyclonal response to $env_{124\text{-}138}$L between B6 EF4.1 and *Emv2*$^{-/-}$ EF4.1 mice revealed around 35% of potentially $env_{124\text{-}138}$L-reactive $CD4^+$ T cells were deleted in the presence of *Emv2* (Fig. 3.5).

Figure 3.5: *Emv2*-mediated deletion of $env_{124\text{-}138}$L-reactive $CD4^+$ T cells



Frequency of $env_{124\text{-}138}$L-specific (determined as $CD69^+$) primary $CD4^+$ T cells in B6 EF4.1 and *Emv2*$^{-/-}$ EF4.1 mice. Plotted are means ±SEM of 9 mice per group from 3 experiments. p=0.003 between groups.

Thus, despite differences in affinity to $env_{124\text{-}138}$Y, and its limited expression, *Emv2* was causing the deletion of a significant proportion of developing transgenic $CD4^+$ T cells in the B6 EF4.1 strain.

3.2.3 Antiviral activity is retained despite *Emv2*-mediated selection

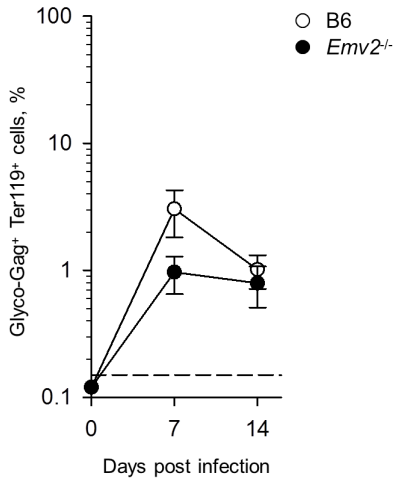
Given the large difference in the proportions of $CD4^+$ T cells responding to $env_{124\text{-}138}$L stimulation between B6 EF4.1 and *Emv2*$^{-/-}$ EF4.1 mice (Fig. 3.5), a variation in the potential response to F-MLV in a FV infection might also be expected.

To firstly investigate the size of this impact outside of the EF4.1 transgenic system, B6 and *Emv2*$^{-/-}$ mice were infected. Whilst both strains effectively controlled infection to similar levels by day 14 post-inoculation, *Emv2*-deficient mice displayed significantly lower numbers of infected erythroid precursor cells ($Ter119^+$glyco-gag^+) in the spleen at the peak of infection (Fig. 3.6).

This reinforced the importance of *Emv2* in thymic selection, but also highlighted the immunological relevance in a non-transgenic setting. Whilst $CD4^+$ T cells are known to have direct effector functions in FV infection, the difference in FV control could potentially also have been explained through impingement of 'helper' functionality. To investigate these links, separate work characterised the $CD8^+$ T cell responses at the peak of infection and assessed binding antibody titres. Levels of $CD8^+$ T cell activation were not found to significantly differ at day 7 post infection (Fig. 3.7a) and titres of IgG (Fig. 3.7b) and IgM (Fig. 3.7a) were equally low at this early time point.

Potential FV-neutralising and binding antibody titre differences were further investig-

Figure 3.6: Control of FV in *Emv2*-sufficient and -deficient mice



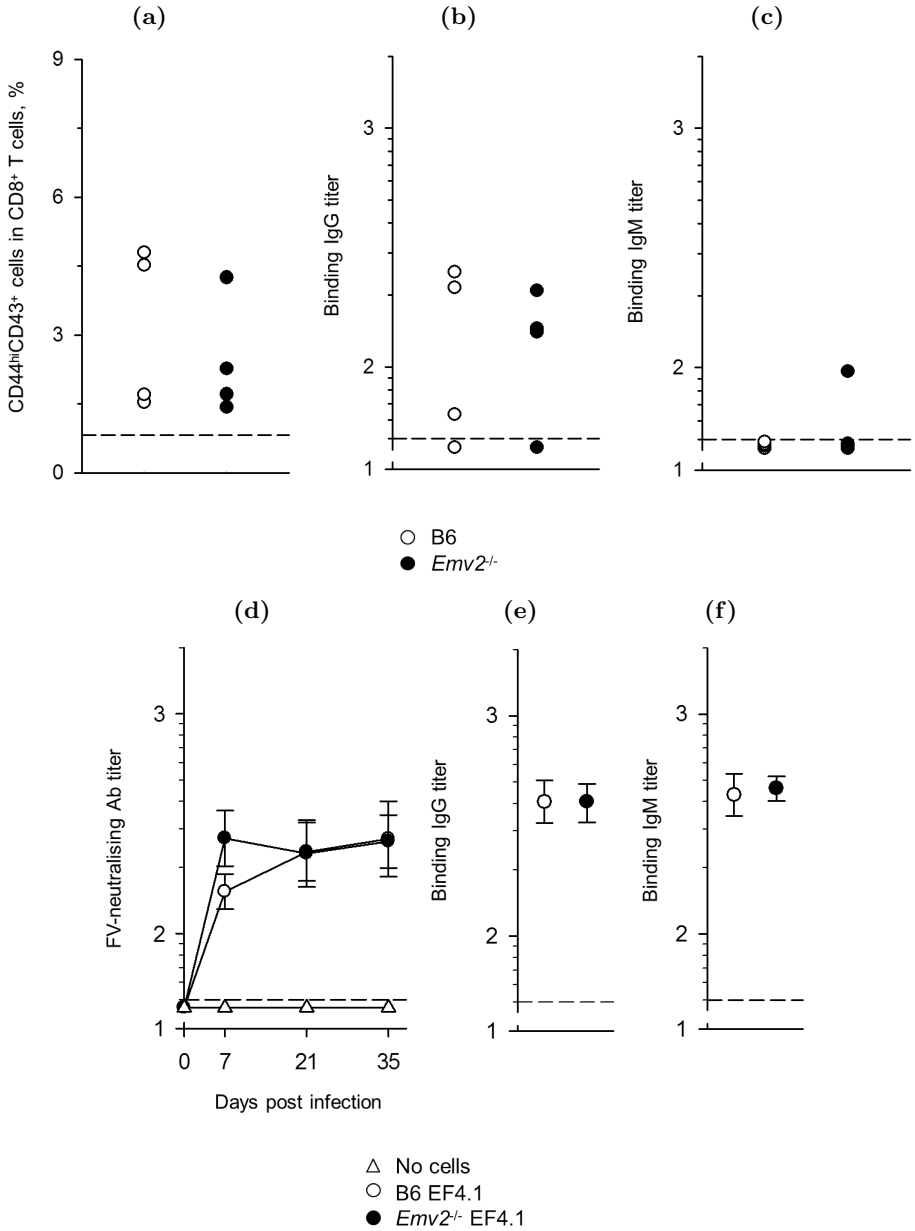
Mean percentage ±SEM of FV-infected erythroid precursors (glyco-Gag$^+$Ter119$^+$) in whole spleens of B6 and *Emv2*$^{-/-}$ mice at specified time points after FV inoculation. Points represent 8–19 mice per group. p=0.04 for day 7 comparison. Dashed line indicates detection limit.

ated through transfer of B6 EF4.1 or *Emv2*$^{-/-}$ EF4.1 cells to *Tcra*$^{-/-}$ mice and subsequent FV infection. In this setting, any differences in antibody production must result from functionality of the donor T cells. Although there was a slight increase in FV-neutralising antibody titre at day 7 post infection in hosts receiving *Emv2*$^{-/-}$ EF4.1 cells (Fig. 3.7d), this difference was not significant. Both sets of donor cells also similarly induced binding IgG and IgM responses (Figs. 3.7e and 3.7f).

In the absence of apparent differences in CD4$^+$ T cell 'helper' functionality, it was likely that the contrast observed in FV susceptibility between B6 and *Emv2*$^{-/-}$ mice was due to controls mediated directly by CD4$^+$ effector T cells. To investigate their functionality directly, EF4.1 CD4$^+$ T cells were isolated from either *Emv2*-sufficient or -deficient mice and transferred into the same type of host prior to FV infection. Their relative antiviral capacity was subsequently determined by measuring levels of Ter119$^+$glyco-gag$^+$ cells at day 7.

Both types of donor CD4$^+$ T cells gave almost complete protection from FV infection at the peak of infection and no significant differences were observed (Fig. 3.8a). Wild-type B6 mice are relatively resistant to FV infection, however, and whilst unable to completely clear virus, naturally control the infection within 21 days. It was possible, therefore, that potential differences in the protective capacity of cells from either donor strain were being overlooked, and the experiment was repeated with *Fv2*^{s} congenic mice, which are highly susceptible to infection. In this setting, the protective effects of cells from either donor were greatly reduced, but, importantly, remained equal (Fig. 3.8b).

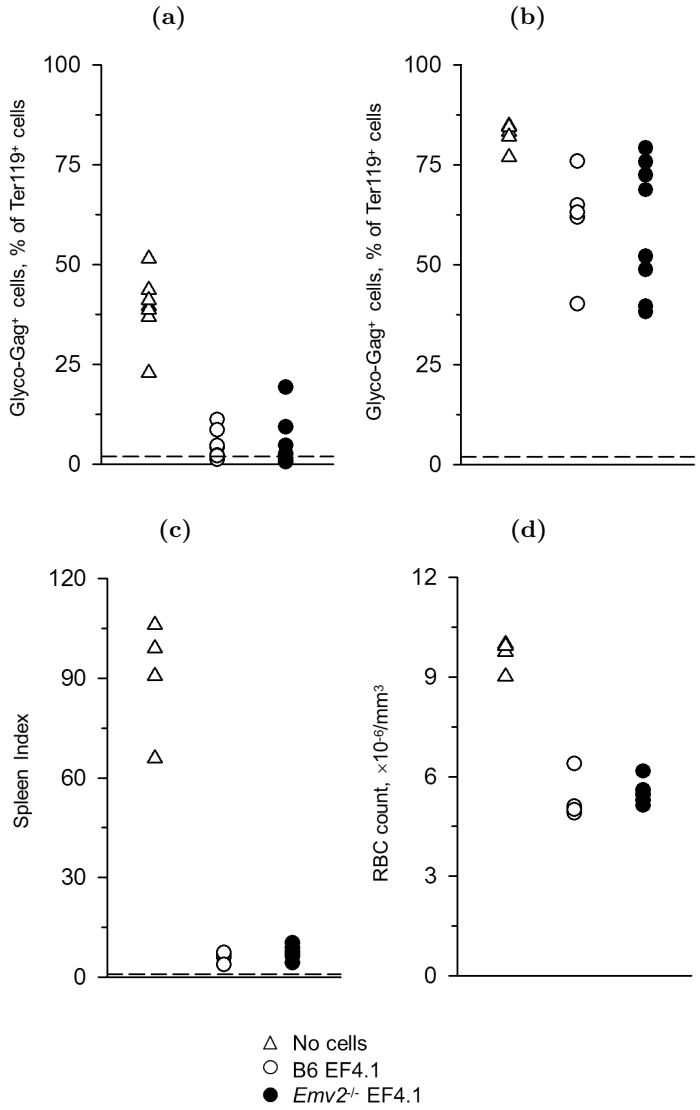
Figure 3.7: Differences in FV infection are not explained by loss of 'helper' functionality



B6 and *Emv2*$^{-/-}$ mice were infected with FV and (a) percentage of CD44$^{\text{hi}}$CD43^{+} cells in total CD8^{+} T cells, and F-MLV-infected cell-binding (b) IgG and (c) IgM antibody titres were determined at day 7 post infection. Each symbol represents an individual mouse. Separately, *Tcra*$^{-/-}$ recipients given purified CD4^{+} T cells, as indicated and inoculated with FV. (d) Titres of FV-neutralising antibodies were determined at the specified points after infection. F-MLV-infected cell-binding (e) IgG and (f) IgM antibodies were determined at day 7 post infection. Plotted are means ±SEM of 11–12 mice per group. Dashed lines indicate detection limits.

Both B6 and *Fv2*^{s} mice still retain functional immune systems prior to the introduction of the donor transgenic cells, however. It therefore remained possible that slight differences in the antiviral effects of the donor cells were being masked or equalised.

Figure 3.8: Equal effector activity in *Emv2*-selected and -non-selected CD4$^+$ T cells



Mice received purified CD4$^+$ T cells, as indicated, and were inoculated with FV for study at day 7 post infection. FACS detection of FV-infected erythroid precursors (glyco-Gag$^+$Ter119$^+$) in whole spleens of (a) B6 (p<0.001 for comparisons of mice receiving cells to the no transfer control) and (b) *Fv2^s* mice (p=0.08 and p=0.003 between no cells and B6 EF4.1 and *Emv2$^{-/-}$* cells, respectively). (c) Measurement of spleen index (spleen weight to body weight ratio) (p=0.001 and p=0.004 between no cells and B6 EF4.1 and *Emv2$^{-/-}$* cells, respectively) and (d) anaemia (p<0.001 for comparisons of mice receiving cells to the no transfer control) in *Rag1$^{-/-}$Fv2^s* mice. Each symbol represents an individual mouse. Dotted lines indicate detection limits.

To this end, *Rag$^{-/-}$Fv2^s* mice, lacking both T and B cells and carrying the *Fv2* susceptibility allele, were used as hosts. The protective capacity of cells in this setting can be established by assessment of the splenomegaly that develops as a result of infection and the anaemia caused by bone marrow pathology. In both of these measures, both

types of donor cells were again comparable (Figs. 3.8c and 3.8d).

Whilst *Emv2* may mediate a large deletion of $env_{122\text{-}141}$L-specific $CD4^+$ T cells, therefore, the 'helper' and effector capacity of both *Emv2*-selected and -non-selected appeared equal.

3.2.4 *Emv2*-mediated selection increases the avidity of the $CD4^+$ T cell response

As *Emv2* expression did not compromise $CD4^+$ T cell-mediated control of FV infection despite causing the deletion of around 35% of potentially $env_{124\text{-}138}$L-responsive cells, it was likely that these cells were not significantly contributing to the immune response in FV infection. To determine why this would be the case, the nature of this population was investigated.

The necessitated cell transfer experiments required the precise distinction of donor and host cells, thus EF4.1 cells selected either in the presence or absence of *Emv2* were transferred into CD45.1/.2 syngeneic recipient mice. Donor cells, identifiable as $CD45.2^+$, could thus be precisely distinguished from host cells, identified as $CD45.1^+CD45.2^+$, by flow cytometry. Significantly higher numbers of donor cells from $Emv2^{-/-}$ EF4.1 mice were seen to respond to FV *in vivo* (Fig. 3.9a), consistent with previous differences identified *in vitro* (Fig. 3.5).

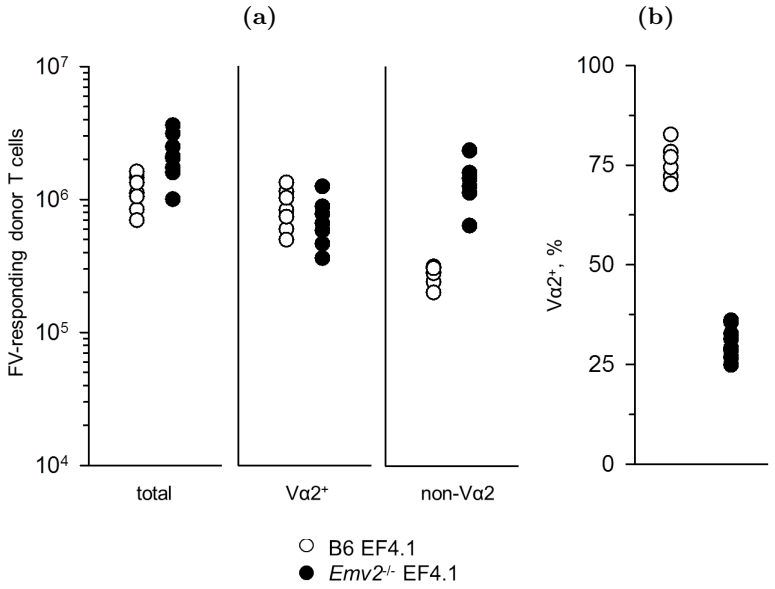
Determination of Vα usage within responding populations indicated comparable numbers of cells from either donor using the endogenous Vα2 (Fig. 3.9a). The significantly higher expansion of donor cells from $Emv2^{-/-}$ EF4.1 mice was, therefore, accounted for by an increase in numbers of low-avidity non-Vα2 $CD4^+$ T cells (Fig. 3.9a). The size of this cell population in fact caused a switch from a predominantly high-avidity ($V\alpha2^+$) response in B6 EF4.1 $CD4^+$ T cells to a predominantly low-avidity (non-Vα2) response in equivalent cells from $Emv2^{-/-}$ EF4.1 donors (Fig. 3.9b).

3.2.5 Presentation of $env_{122\text{-}141}$Y causes the removal of non-Vα2 $CD4^+$ T cells

Differences in development of non-Vα2 $CD4^+$ T cells could be realised by two paths – firstly, control of the proportions of $V\alpha2^+$ and non-Vα2 $CD4^+$ T cells within the naïve population or, secondly, control of their relative avidity to the core epitope of $env_{122\text{-}141}$L, and hence influence over their proliferative capacity.

To investigate this further, EF4.1 $CD4^+$ T cells selected in *Emv2*-sufficient of -deficient mice were stimulated with titrations or $env_{124\text{-}138}$L peptide. No significant effect on the overall avidity at which *Emv2*-selected cells responded to the peptide was observed

Figure 3.9: *Emv2*-mediated selection removes non-Vα2 CD4$^+$ T cells



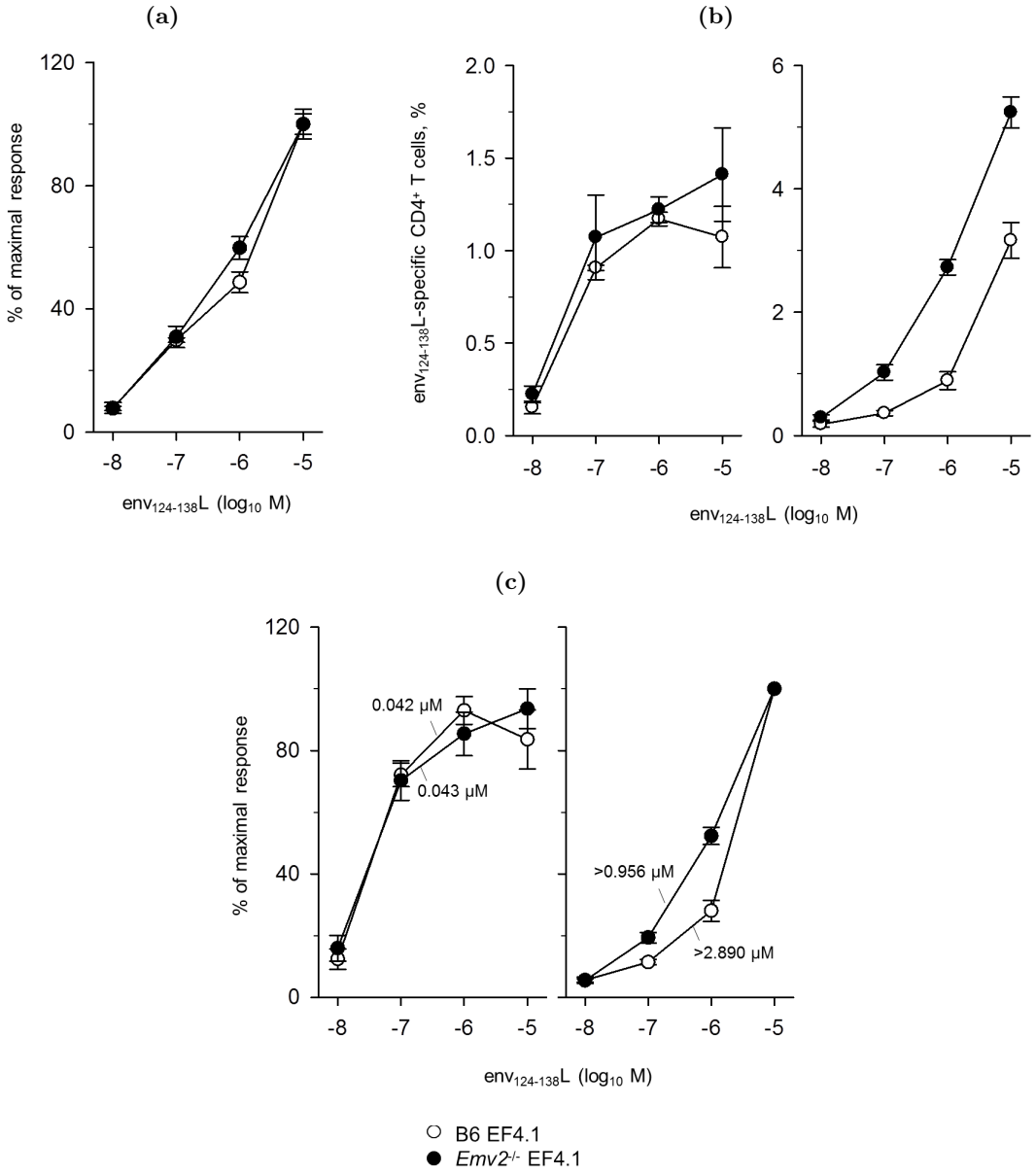
Purified CD4$^+$ T cells from the mice indicated (CD45.2$^+$) were adoptively transferred into CD45.1/.2 syngeneic recipients. Mice were simultaneously inoculated with FV and expansion of donor cells was determined at day 7 post infection. (a) Absolute numbers of total responding (determined as CD69$^+$), Vα2$^+$ and non-Vα2 donor cells (p=0.009 between total and p<0.001 between non-Vα2 cells). (b) relative frequency of Vα2$^+$ cells within donor populations (p<0.001 between donors). Each symbol represents an individual mouse.

(Fig. 3.10a). As seen *in vivo* (Fig. 3.9a), the numbers of Vα2$^+$ cells responding within this population was comparable between both types of donor cells (Fig. 3.10b), and, accordingly, avidity for peptide was unchanged (Fig. 3.10c). In comparison, significantly larger proportions of non-Vα2 cells from *Emv2*$^{-/-}$ mice responded to stimulation (Fig. 3.10b). Correspondingly, the avidity of responding CD4$^+$ T cells was 3-fold higher when selected in *Emv2*-deficient mice, although remained substantially lower than Vα2$^+$ cells from either donor (Fig. 3.10c).

Taken together, the increase in total frequency and partial increase in env$_{124\text{-}138}$L avidity of non-Vα2 cells in *Emv2*$^{-/-}$ EF4.1 mice likely account for the preferential expansion of this population upon FV infection (Fig. 3.9b). The predominance of the non-Vα2 cell population was separately confirmed at a range of peptide concentrations for both peripheral CD4$^+$ T cells (Fig. 3.11a) and CD4$^+$CD8$^-$ thymocytes (Fig. 3.11b). Thus, *Emv2* mediated a significant thymic selection event that significantly altered clonal composition of env$_{124\text{-}138}$L-reactive CD4$^+$ T cells.

The preferential deletion of non-Vα2 CD4$^+$ T cells would necessitate that these cells are capable of reacting to env$_{124\text{-}138}$Y. Percentage of env$_{124\text{-}138}$Y-responsive Vα2$^+$ CD4$^+$ T cells were similar whether *Emv2*-selected or -non-selected (Fig. 3.12a) and has

Figure 3.10: Avidity for env$_{124\text{-}138}$L is unchanged by *Emv2*-mediated selection

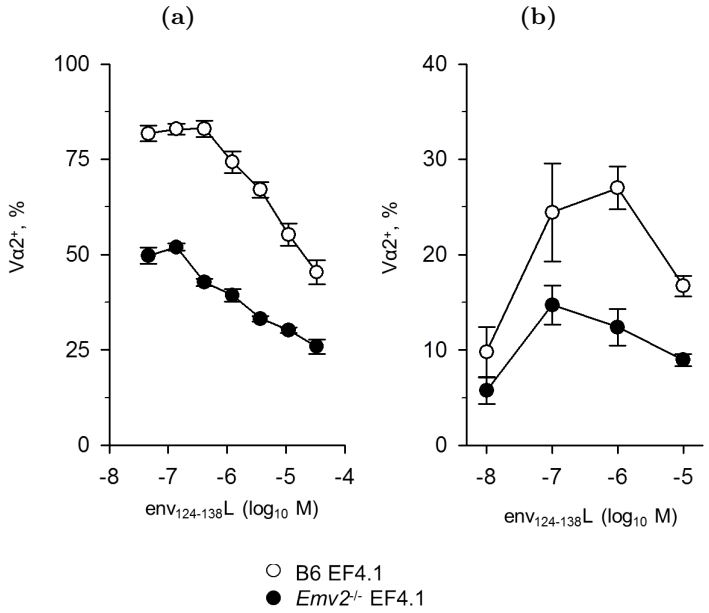


(a) Response of indicated CD4$^+$ T cells to a titration of env$_{124\text{-}138}$L peptide. (b) frequency of responding (determined as CD69$^+$) and (c) functional avidity of Vα2$^+$ (left) and non-Vα2 CD4$^+$ (right) T cells across this titration. p=0.002 between non-Vα2 cells at the highest concentration in (b). Numbers in (c) represent the ED$_{50}$. Data are means ±SEM of 9–12 mice per group from 3 experiments.

similarly low avidity for the peptide (Fig. 3.12b), only responding at the highest concentration of peptide. In comparison, non-Vα2 cells from *Emv2*$^{-/-}$ EF4.1 mice showed an over 70-fold increase in avidity for env$_{124\text{-}138}$Y (Fig. 3.12b) and a corresponding 3.5-fold increase in the percentage of responding cells (Fig. 3.12a).

Thus, selection by *Emv2* did not impact Vα2$^+$ CD4$^+$ T cells due to their low avidity

Figure 3.11: Deletion of non-Vα2 env$_{124\text{-}138}$L-reactive cells is a thymic event



Frequency of Vα2$^+$ CD4$^+$ (CD8$^-$) T cells from the (a) spleens and (b) thymi of the indicated mice responding (determined as CD69$^+$) to a titration of env$_{124\text{-}138}$L peptide. Data in (a) are means ±SEM of 9–12 mice per group from 3 experiments. Data in (b) are the means ±SEM of 8–10 mice per group, p=0.002 for 10^{-6} M comparison.

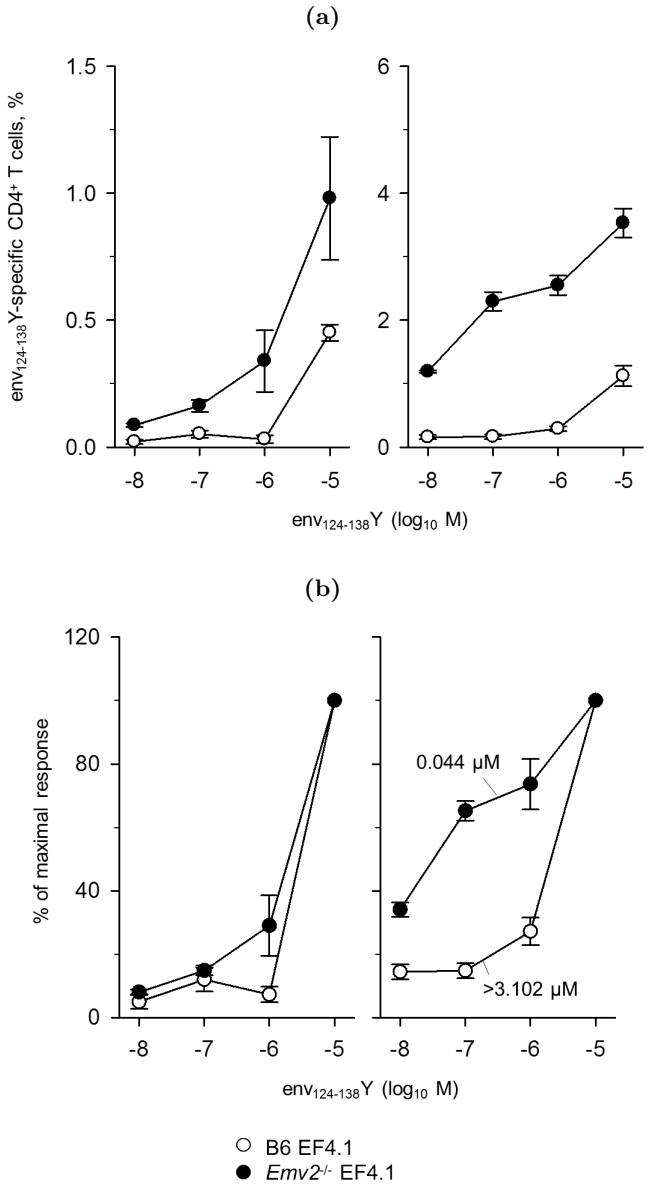
for env$_{124\text{-}138}$Y. However, non-Vα2 cells display a greatly increased cross-reactivity, including to env$_{124\text{-}138}$Y, and are hence deleted as thymocytes in *Emv2*-sufficient mice.

3.2.6 *Emv2* shapes the depth of the env$_{124\text{-}138}$L-responsive CD4$^+$ T cell repertoire

Whilst selection by *Emv2* might not impact the high-avidity (Vα2$^+$) response to FV infection *in vivo* or strongly affect the avidity of these cells to env$_{124\text{-}138}$L peptide, a significant difference in the development of non-Vα2 cells was noted. These CD4$^+$ T cells were significantly less responsive to env$_{124\text{-}138}$Y peptide stimulation, a reduction in repertoire depth that may potentially impact control of viral escape mutants. The magnitude of the change in overall repertoire depth was, therefore, investigated.

Reactivity to a further mutation of the core epitope, env$_{124\text{-}138}$YS, was firstly assessed. The LT-128,129-YS mutant represents a functional sequence derived from the co-opted retroviral *env* at the *Fv4* locus found in some strains, but importantly, has different amino acids at two of the three points identified as contacting the TCR (see Fig. 3.2). Whilst very few Vα2$^+$ or non-Vα2 CD4$^+$ T cells responded to env$_{124\text{-}138}$YS stimulation, development in the absence of selection by *Emv2* produced a 7-fold increase in

Figure 3.12: Avidity for env$_{124\text{-}138}$Y in *Emv2*-selected and -non-selected EF4.1 CD4$^+$ T cells

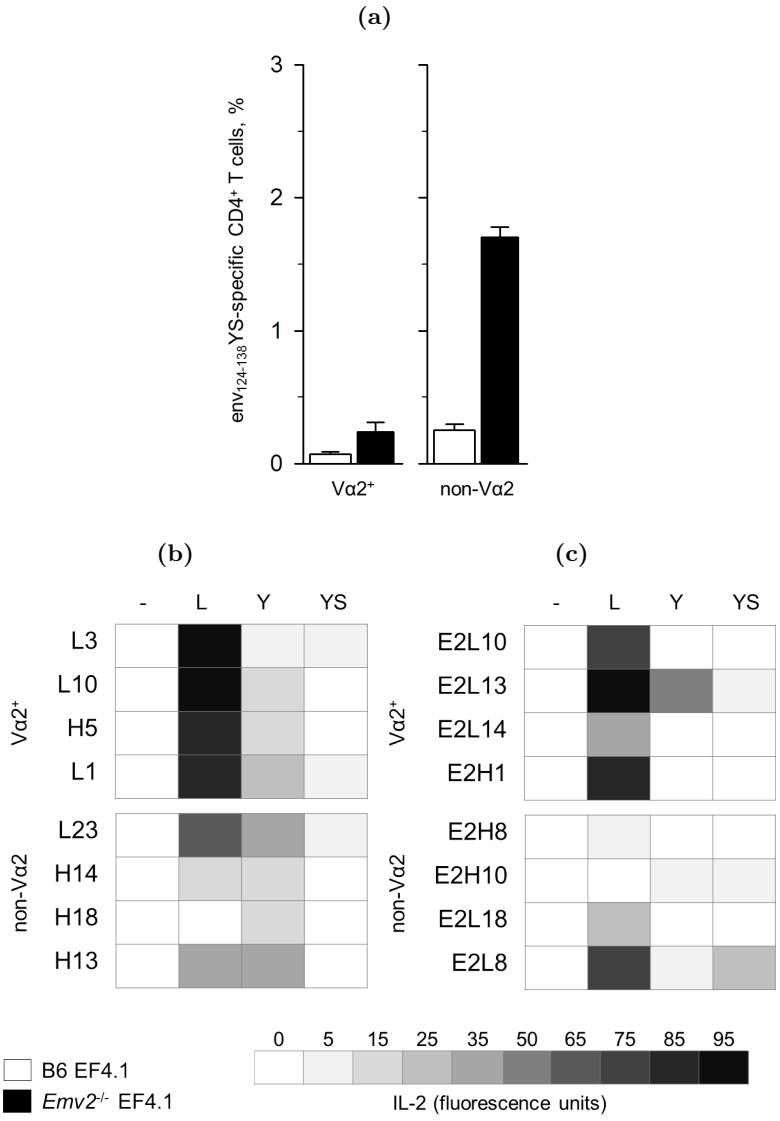


Response of indicated CD4$^+$ T cells to a titration of env$_{124\text{-}138}$Y peptide. (a) frequency of responding (determined as CD69$^+$) and (b) functional avidity of Vα2$^+$ (left) and non-Vα2 CD4$^+$ (right) T cells across this titration. p=0.001 between non-Vα2 cells at the highest concentration in (a). Numbers in (b) represent the ED$_{50}$. Data are means ±SEM of 9–12 mice per group from 3 experiments.

the magnitude of the low-avidity response (Fig. 3.13a).

Previous work had suggested that B6 EF4.1-derived Vα2$^+$ CD4$^+$ T cell hybridomas had the ability to respond to env$_{124\text{-}138}$Y, but that this response was impaired in comparison to when stimulated with the cognate antigen (Fig. 3.4b). To extend this finding to env$_{124\text{-}138}$YS responsiveness and to include *Emv2*-selected TCRs, *Emv2*$^{-/-}$ EF4.1-derived hybridomas were generated and stimulated with a variety of peptides.

Figure 3.13: Selection by *Emv2* reduces the depth of the TCR repertoire



(a) Frequency of $env_{124\text{-}138}$YS-reactive $CD4^+$ T cells (determined as $CD69^+$) isolated from the indicated mice. Cells were stimulated with 10^{-5} M peptide. Data are means ±SEM of 9 mice per group from 3 experiments. $p<0.001$ for non-Vα2 comparison. (b) and (c) show responsiveness (determined by IL-2 production) of hybridomas derived from B6 EF4.1 and *Emv2*$^{-/-}$ EF4.1 mice, respectively. Hybridomas were stimulated with $5x10^{-1}$ M $env_{124\text{-}138}$L, Y, or YS peptide, or incubated in the absence of peptide (–).

Hybridomas were generated via stimulation with $env_{124\text{-}138}$L, thus of all V$\alpha2^+$ hybridomas, none were seen to respond to $env_{124\text{-}138}$Y or $env_{124\text{-}138}$YS more strongly than to their cognate antigen (Figs. 3.13b and 3.13c). Equally, the responsiveness of non-Vα2 hybridomas to mutant peptides was not significantly greater than that of V$\alpha2^+$ hybridomas derived from the same strain (Figs. 3.13b and 3.13c). Thus, whilst hybridomas generated against $env_{124\text{-}138}$L had some limited reactivity to mutant pep-

tides, the majority of non-Vα2 cells developing in the absence of *Emv2*-mediated selection were likely to be distinct from those displaying env$_{124\text{-}138}$L-reactivity. This also indicated that selection by *Emv2* did not reduce the apparent cross-reactivity of TCRs, and that non-Vα2 hybridomas did not have inherent increases in peptide cross-reactivity.

Interestingly, separate work by colleagues cloning and sequencing expressed endogenous Vα genes suggested all non-Vα2 hybridomas from both B6 EF4.1 and *Emv2*$^{-/-}$ EF4.1 mice were members of the *Trav9* family, encoding Vα3, although only a subset could be stained with the anti-Vα3.2 antibody RR3-16 (Table 3.1).

Table 3.1: Vα3 usage in EF4.1 hybridomas

Clone	Vα3.2 staining	TRAV	TRAJ	AA junction
H13	+	9D-4*04	31*01	CAVEGGNNRIFF
H14	−	9N-4*01	17*01	CALSNSAGNKLTF
H18	+	9N-2*01	31*01	CVLSRNSNNRIFF
L23	−	9N-4*01	27*01	CALDNTNTGKLTF
E2H8	−	9N-4*01	17*01	CALSNSAGNKLTF
E2H10	−	9N-4*01	21*01	CALSGPNYNVLYF
E2L8	−	9N-2*01	16*01	CVLSGSSSGQKLVF
E2L18	+	9N-3*01	58*01	CAVRQGTGSKLSF

To further investigate the ability of specific TCRs to tolerate mutations at each of the three residues identified as contacting the env$_{122\text{-}141}$L peptide, individual hybridomas were screened with peptides mutated at each of these positions through the range of all natural amino acids (Figs. 3.14 and 3.15). Amino acids that elicited at least 40% of the maximal determined response for each hybridoma individually were listed in order of the magnitude of the induced response (Fig. 3.16).

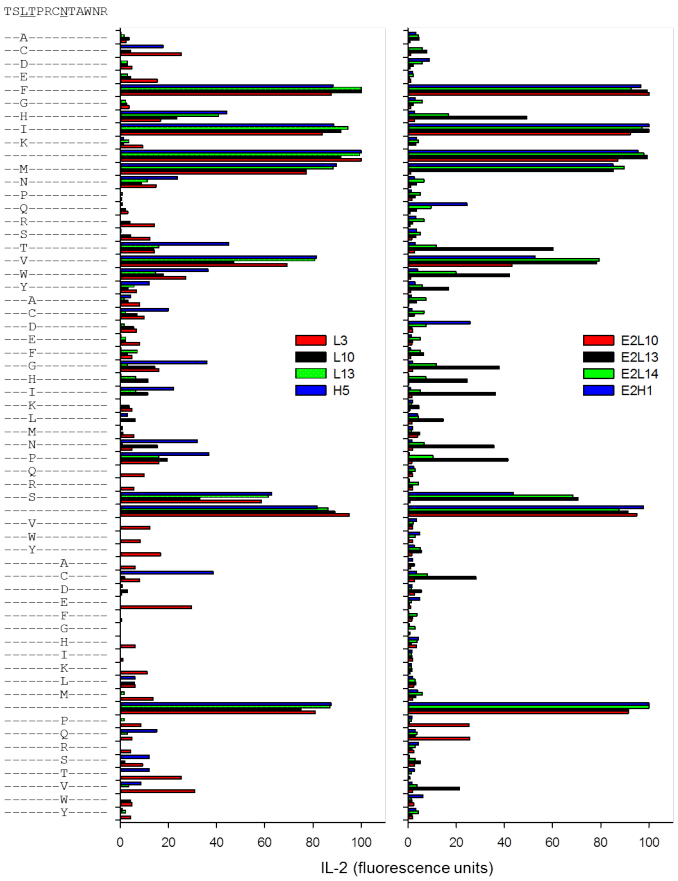
All Vα2^{+} hybridomas, whether *Emv2*-selected or -non-selected, displayed strong preference for the amino acids representing the un-mutated env$_{124\text{-}138}$L peptide, responding most strongly to leucine at position 128, threonine at position 129, and asparagine at position 133 (Figs. 3.16a to 3.16c). Similar hydrophobic amino acids were accepted at position 128 for all Vα2^{+} hybridomas, followed by the smaller hydrophobic amino acid valine, although none reacted significantly to tyrosine at this position. Whilst some similar amino acids could be recognised at position 128, Vα2^{+} hybridomas were less able to respond to mutations at positions 129 and 133 (Figs. 3.16a to 3.16c).

Non-Vα2 hybridomas stimulated with the same range of peptides displayed significantly different amino acid preference to Vα2$^+$ hybridomas at position 128, with only one of eight preferring leucine (Fig. 3.16a). Notably, all non-Vα2 hybridomas responding to leucine (E2L8, H13, and L23) also prefered threonine at position 129 and asparagine at position 133 (Figs. 3.16a to 3.16c), representing the index sequence. Where clones did not respond to leucine at this position, reactivity could be enhanced through further changes at position 133 (Fig. 3.16c).

Interestingly, valine and the closely related isoleucine were highly preferred by the majority of both *Emv2*-selected and -non-selected non-Vα2 hybridomas at position 128 (Fig. 3.16a), but two *Emv2*-selected TCRs showed highly heterogenous reactivity, responding to polar, non-polar, hydrophobic, and hydrophilic amino acids of a variety of sizes.

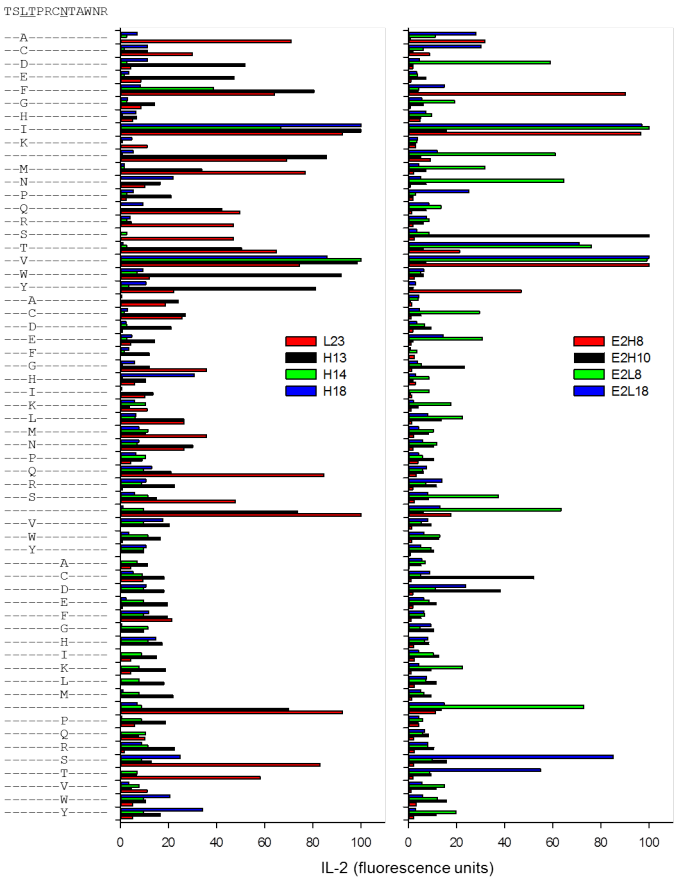
Together, these results confirmed the preference of Vα2$^+$ TCRs for leucine at position 128 and suggested that these were unaffected by *Emv2*-mediated thymic selection, with no changes in cross-reactivity or repertoire depth being seen. Non-Vα2 hybridomas displayed noticeably different amino acid preferences at this position, and selection by *Emv2* appeared to increase reactivity, suggesting a potential switch in TCR contact point, likely to position 127.

Figure 3.14: TCR cross-reactivity in *Emv2*-selected and -non-selected Vα2$^+$ hybridomas



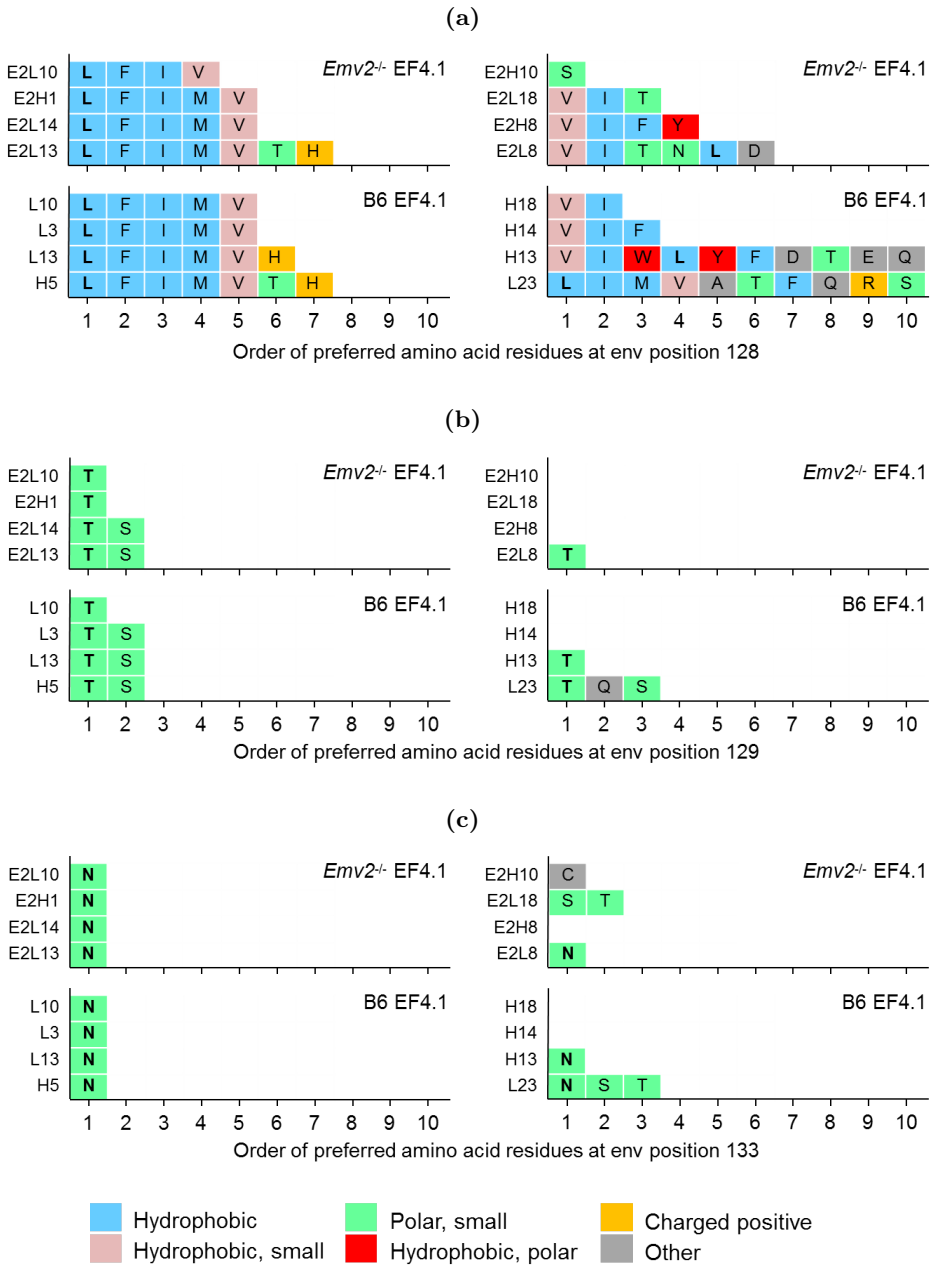
B6 EF4.1- (left) and *Emv2*$^{-/-}$ EF4.1-derived (right) hybridoma lines stimulated with a library of 5x10^{-6} M peptides in which the three positions contacting the TCR are individually replaced with all possible natural amino acids. Responses are measured through IL-2 production. Data are the means of triplicate cultures.

Figure 3.15: TCR cross-reactivity in *Emv2*-selected and -non-selected non-Vα2 hybridomas



B6 EF4.1- (left) and *Emv2*$^{-/-}$ EF4.1-derived (right) hybridoma lines stimulated with a library of 5x10^{-6} M peptides in which the three positions contacting the TCR are individually replaced with all possible natural amino acids. Responses are measured through IL-2 production. Data are the means of triplicate cultures.

Figure 3.16: Order of amino acid preference at defined TCR contact residues



B6 EF4.1- and *Emv2*$^{-/-}$ EF4.1-derived hybridoma lines stimulated with a library of peptides (5x10^{-6} M) in which the three positions contacting the TCR are individually replaced with all possible natural amino acids. Responses are measured through IL-2 production. Data are amino acids substitutions at each position that elicited >40% of the maximal response, listed in decreasing order of IL-2 production.

3.3 Discussion

The processes of somatic recombination that generate variable lymphocyte antigen receptors can produce massive diversity (see Section 1.2.2), a functionality that is vital to the effective recognition and control of a diverse array of pathogens. However the nature of these events is inherently random and as such receptors must be screened for functionality and auto-reactivity prior to release of developing cells into the periphery. Positive and negative selection of thymocytes relies on the large-scale presentation of self-peptides in order to ensure sufficient compatibility with self MHC molecules, whilst preventing excessive reactivity (see Section 3.1.3). The process of negative selection is thought to decrease the frequency, avidity, and cross-reactivity of TCRs recognising peptide epitopes sharing homology with self peptides and to encourage antigen specific responses.

Despite contributing far larger proportions of mammalian genomes than generally appreciated, the impact of REs and ERVs on the development of the immune response has not been thoroughly investigated. Previous work in the literature has suggested the involvement of self peptide epitopes derived from endogenous MLVs in the positive selection and enhancement of peripheral survival of $CD4^+$ T cells with specificity for an unrelated moth cytochrome C peptide restricted by *H2-*E^k (Ebert, Jiang, Xie, Li & Davis, 2009). Thus whilst epitopes derived from ERVs have been shown to be presented in complex with MHC, their potential impact on immunity to exogenous retroviral infection has been largely unstudied.

The sole eMLV of B6 mice, *Emv2*, shares strong homology with the immunodominant *H2-*A^b-restricted epitope of F-MLV, the helper component of FV. FV represents a well-characterised murine retroviral infection model and much previous work has resulted in the generation of the B6 EF4.1 transgenic mouse, carrying a fixed TCRβ chain derived from a F-MLV antigen, $env_{121\text{-}144}$L. Together, these components allowed an unparalleled opportunity to determine the contribution of an ERV to the immune response to retroviral infection, and investigation of the underlying controls surrounding the high-avidity response to F-MLV seen in the commonly used B6 strain.

This work has outlined the role of *Emv2* in deletion of around 35% of $env_{124\text{-}138}$L-reactive $CD4^+$ T cells from the potential responding polyclonal population seen in *Emv2*$^{-/-}$ EF4.1 mice. Further, selection by *Emv2* removed $CD4^+$ T cells carrying TCRs capable of responding to a variety of env mutants and natural variants, theoretically impairing the ability of responding cells to detect escape mutants.

Interestingly, however, negative selection in this way promoted the formation of a high-avidity response to F-MLV in *Emv2* sufficient mice, skewing the clonal composition in favour of Vα2 $CD4^+$ T cells by the preferential deletion of low-avidity clones. Whilst

negative selection removed a significant fraction of potentially F-MLV-reactive CD4$^+$ T cells, therefore, assessment of effector functionality within the remaining population suggested that antiviral activity was unaffected. This finding suggested that Vα2$^+$ and non-Vα2 CD4$^+$ T cells contribute unequally to the immune response to FV infection, with the majority of control being exerted through the high-avidity response.

Although these data do not show an effect on antibody responses to F-MLV at early time points, the high similarity of *Emv2* to F-MLV also has the potential to impact these responses. This work has focussed on differences in control of FV at the peak of viral infection, before the formation of a strong antibody response, and hence does not exclude the possibility of antibody-mediated effects at later time points.

3.3.1 Derived work

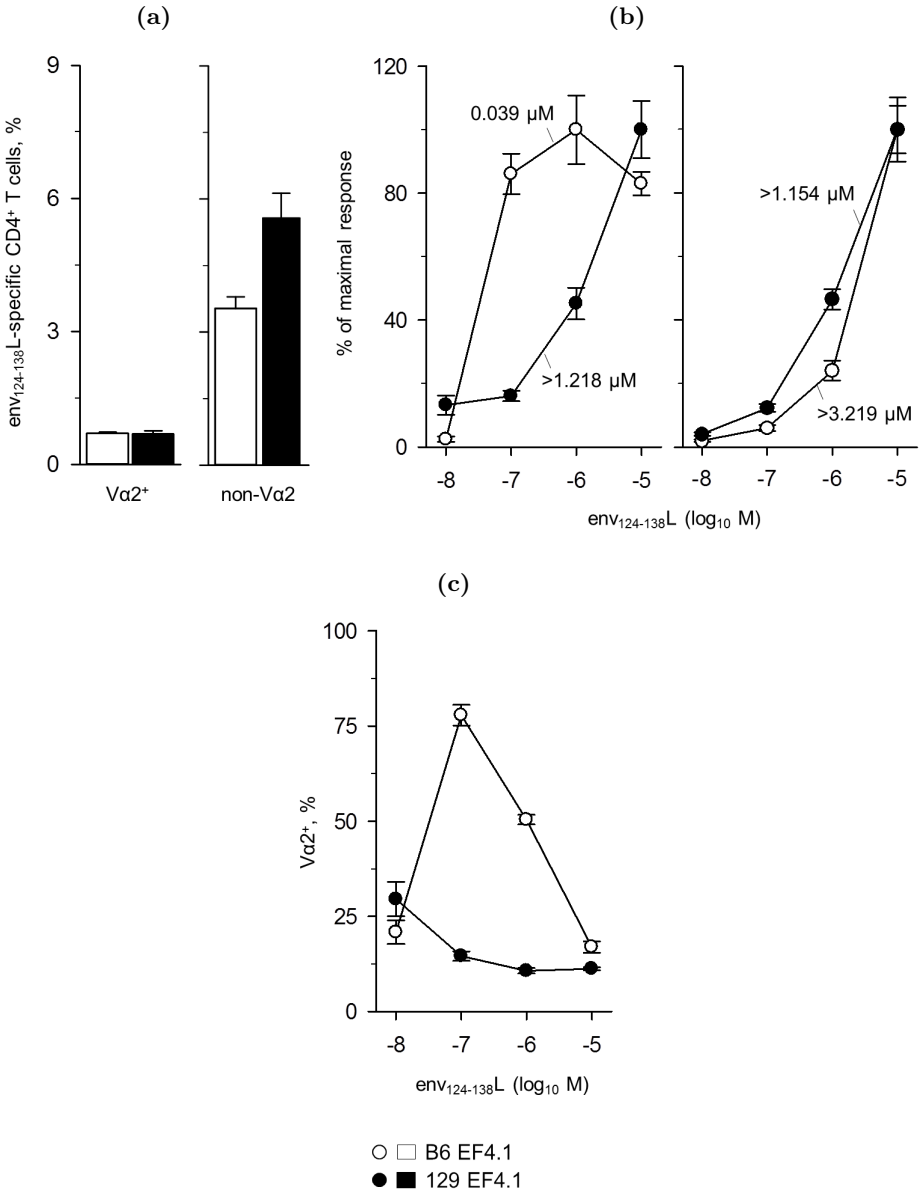
H2-A^b, as noted previously (Section 3.1.2), has been identified as vital to a strong response to F-MLV, allowing antibody class switching and effective immunisation (Miyazawa, Nishio & Chesebro, 1988; Hasenkrug & Chesebro, 1997). However, both the closely-related B6 and B10 strains used in much early research both carry *Emv2*, thus complicating conclusions regarding the absolute contribution of *H2-A*^{b} alone.

Separate work has since investigated this relationship using a variety of strains and crosses with 129S8 mice, which whilst carrying *H2-A*^{b}, have no endogenous eMLVs (Jenkins et al., 1982). Similarly to *Emv2*$^{-/-}$ EF4.1 mice, 129S8 EF4.1 have comparable frequencies of responding Vα2$^+$ CD4$^+$ T cells to B6 EF4.1, but greater proportions of responding non-Vα2 cells (Fig. 3.17a). Surprisingly, however, the avidity of Vα2$^+$ CD4$^+$ T cells from 129S8 EF4.1 mice for env$_{124\text{-}138}$L was greatly reduced in comparison to equivalent cells from B6 EF4.1 and *Emv2*$^{-/-}$ EF4.1 mice (Fig. 3.17b), and caused a predominantly low-avidity response across a range of peptide concentrations (Fig. 3.17c).

These data implied that a large contribution towards the high-avidity response of B6 came from outside of either *H2* or eMLV loci. Further separate study within the group has identified the role of the endogenous *Trav* loci in formation of this response. The F1 generation of a cross of *Tcra*$^{-/-}$ and 129S8 EF4.1 mice, only able to generate Vα chains from *Trav* loci inherited from the 129S8 parent, were notably unable to form a high-avidity response to env$_{124\text{-}138}$L stimulation (Fig. 3.18a) and the responding population was dominated by low-avidity (non-Vα2) clones (Fig. 3.18b).

The high-avidity response of B6 background mice consequently relies on three separate factors. *H2-A*^{b} is required for the presentation of env$_{121\text{-}141}$L, but a high-avidity response cannot be generated without the necessary *Trav* genes encoded at the *Tcra* locus. Separately, the endogenous eMLV *Emv2* is responsible for the negative selection

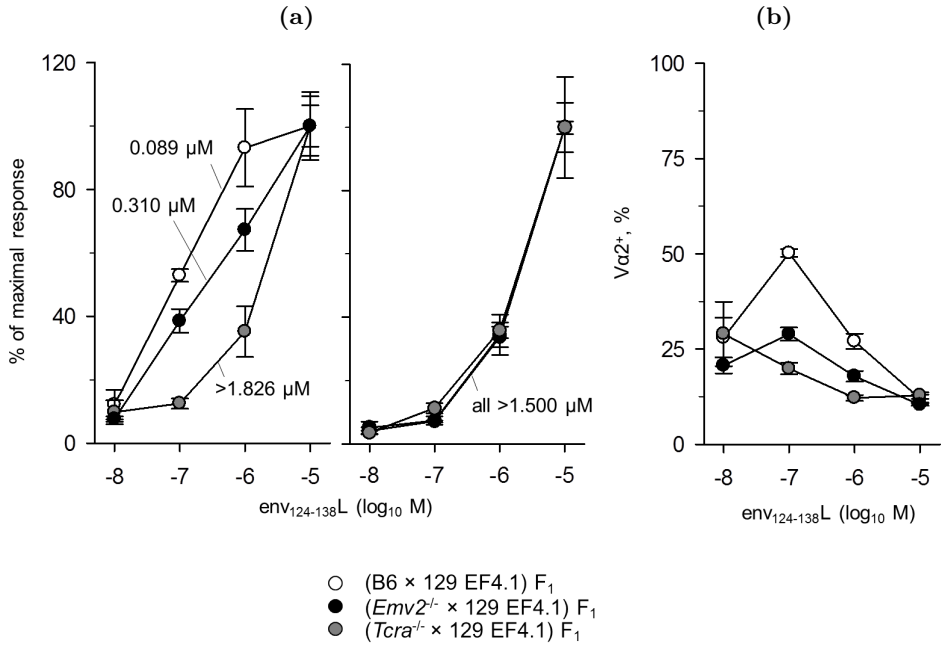
Figure 3.17: The contribution of *H2-A^b* to high-avidity F-MLV responses



(a) Frequency of env$_{124\text{-}138}$L-reactive Vα2$^+$ and non-Vα2 CD4$^+$ T cells from the strains indicated when incubated with 10^{-5} M peptide (p=0.016 between non-Vα2 cells). (b) Functional avidity of the indicated Vα2$^+$ (left) and non-Vα2 (right) CD4$^+$ T cells to stimulation with a titration of peptide, and (c) frequency of Vα2$^+$ CD4$^+$ T cells across these concentrations of peptide. Responsiveness was determined by CD69 expression. Numbers in (b) represent ED$_{50}$. Data are means ±SEM of 4–8 mice from 3 experiments.

of a significant proportion of low-avidity clones, such that the high-avidity response can dominate the response to F-MLV in B6 background mice.

Figure 3.18: The contribution of *Trav* to high-avidity F-MLV responses



(a) Functional avidity and (b) frequency of Vα2$^+$ CD4$^+$ T cells responding (determined as CD69$^+$) to stimulation with a titration of env$_{124\text{-}138}$L peptide. Numbers in (a) represnet ED$_{50}$. p<0.001 for comparison of (B6 x 129 EF4.1)F$_1$ to both other strains for 10^{-7} M in (b). Data are means ±SEM of 4–8 mice per group from 3 experiments.

3.3.2 Implications of this work

A single related study has previously investigated the impact of an endogenous eMLV, *Emv1* of BALB/c mice, on the rejection of CT26 tumour cells (McWilliams et al., 2008). *Emv1* could be shown to mediate deletion and impairment of avidity of env-reactive CD4$^+$ T cells in an expression-dependent manner, a conclusion that would, perhaps, be expected in the light of a traditional view of negative selection of thymocytes.

In comparison, this work, whilst mirroring the deletion of a large proportion of env-reactive thymocytes, has also revealed the counter-intuitive requirement of this selection for the formation of high-avidity responses to F-MLV in B6 background mice. This conclusion provides an important proof of principle, showing that negative selection in this manner cannot be expected to impair the immune response in every case, and highlights the complex and multifactorial nature of thymocyte selection.

The ability to form effective immune responses can be mediated by comparatively few, or indeed single, self peptide epitopes derived from a potential pool of several thousand presented to developing thymocytes at any one time. Separate work by colleagues derived from this study has emphasised the involvement of TCR V region allelic variation

and MHC haplotypes in thymic selection and mimics previously outlined results from studies of the human response to Epstein-Barr virus (EBV). In this case, polymorphism in *TRBV9* has been linked to a high-avidity response to the EBNA-1 protein when in combination with a specific MHC class I molecule, HLA-B*3501 (Gras et al., 2010).

Although there is much debate surrounding the involvement of REs and ERVs in human disease (see Section 1.1.4), the possible interaction of epitopes derived from these elements with development of immune responses represents a separate and fascinating research prospect.

Chapter 4

Immune control of retroelement expression

4.1 Introduction

4.1.1 Activating retroelement expression

Mechanisms of ERV activation have been long-studied and various radiological (Lieberman & H. S. Kaplan, 1959; W. P. Rowe et al., 1971), chemical (Lowy, Rowe, Teich & Hartley, 1971) and mitogenic factors (Moroni & Schumann, 1975; Greenberger, Phillips, Stephenson & Aaronson, 1975) have been characterised that are capable of promoting endogenous retroviral particle formation from untransformed cells.

Related research focussed on HIV has suggested that viral replication can be both positively and negatively regulated by a variety of cellular pathways induced by cytokine signalling (Fauci, 1993). These concepts can also be applied to REs, and ERV expression has been shown to be induced by factors related to immune activation. Transcripts initiating from HERV-H integrations have been detected upon T cell activation *in vitro*, for example, and are likely induced in an autocrine manner (Kelleher, Wilkinson, Freeman, Mager & Gelfand, 1996). Further, HERV-W, -K family, and -H RNA products can be seen on treatment of macrophages with phorbol-12-myristate-13-acetate (PMA) and lipopolysaccharide (LPS) (Johnston et al., 2001), and tumour necrosis factor (TNF)α and IL-1α and β have been seen to induce HERV-R expression in cell cultures (Katsumata et al., 1999). Interestingly, several factors, including IL-1β, TNFα, and LPS, have also been noted to have similar roles in the induction of endogenous MLV transcripts in murine cells (Itoh et al., 1992; Baudino, Yoshinobu, Dunand-Sauthier, Evans & Izui, 2010).

The dynamic nature of RE and ERV expression is highlighted by comparison of expression profiles between tissues (Seifarth, Frank et al., 2005; Ahn & H.-S. Kim, 2009). The most abundantly and globally expressed families were HERV-E, -F, -W, and ERV9 within class I elements, and HML-2, -4, and -6 within class II (HERV-K family) elements. Broadly, expression potential was seen to correlate with estimated element age, and includes the majority of elements highlighted above.

4.1.2 Controlling retroelement expression

As discussed previously (see 1.1.4), large-scale epigenetic and transcriptional controls are placed on the expression of REs within the mammalian genome. Further post-transcriptional controls may also be active in the management of REs. APOBEC family members, for example, have been implicated in the control of LINE retrotransposition (MacDuff et al., 2009; Wissing et al., 2011), although the primary role of these factors is thought to remain in mutagenesis of sequences prior to their integrase-mediated insertion (Jern, Stoye & Coffin, 2007; Armitage et al., 2008).

Whilst many cell-intrinsic controls may be outlined, the potential roles of the innate and adaptive immune responses on RE expression are largely unexplored. Equally, the possible impacts of confounding factors, such as infection, are relatively understudied areas. For example, whilst the impact of heterologous infection on HIV expression has been well characterised and many viruses are known to stimulate expression *in vitro* (Rosenberg & Fauci, 1989), this knowledge has only recently been applied to ERVs. Encouragingly, many parallels can be noted, and HIV infection itself has been suggested to increase HERV-K family expression (R. Contreras-Galindo, López, Vélez & Yamamura, 2007; R. Contreras-Galindo, M. H. Kaplan, A. C. Contreras-Galindo et al., 2012). Further, herpes simplex virus (HSV)-1 and influenza A virus (IAV) infection have been proposed to elevate HERV-W expression (W. J. Lee, Kwun, Kim & Jang, 2003; Nellåker, Yao et al., 2006) and interactions have been noted with Epstein-Barr virus (EBV) infection (H. Gross et al., 2011).

Mammals have long-standing relationships with a large range of microbial species, with interactions ranging from mutualism to parasitism, and noticeable effects on RE expression are unlikely to be confined to viral infections. The gastrointestinal tract plays host to the largest of these interactions, both in terms of total surface area (250-400 m^2 in humans) and in the numbers involved (estimated at 10^{14} bacterial cells). Reaching densities of 10^{11}-10^{12} cells/g of luminal content, the total size of the gut microbiota is an order of magnitude larger than the total number of cells in the human body (Guarner & Malagelada, 2003). Thus, the impacts of dysbiosis can be severe, and the ecology of the gastrointestinal tract is a large area of current research.

TLR4, responsible for sensing bacterial LPS, has previously been seen to affect recombination and selection of MMTVs within mice (Jude et al., 2003). Indeed successful vertical transmission of MMTVs has subsequently been suggested to depend on the presence of the microbiota (Kane et al., 2011), and a following study implicated gut flora in increased poliovirus and reovirus pathogenesis (Kuss et al., 2011).

4.1.3 The microbiome in health and disease – structure and function

Large-scale sequencing efforts have recently facilitated the study of microbial diversity across multiple individuals and body sites (Arumugam et al., 2011; The Human Microbiome Project Consortium, 2012). Systematic sampling of human gastrointestinal microbial communities would require invasive collection procedures, however, and hence current understanding of the gut microbiota largely centres around study of colonic communities. Even with this caveat, these studies have shown significant variation in the composition between individuals, with three main clusters being revealed by principal component analysis, each dominated by a different bacterial genus (*Bacteroides*, *Prevotella*, or *Ruminococcus*). Each grouping displayed differing metabolic specialisms,

but bacterial composition was, for the most part, not seen to be significantly impacted by age, host nationality (and likely dietary differences), or variables such as body mass index (Arumugam et al., 2011).

Differences noted in bacterial enzymatic pathways link to the suggested relevance of the gut microbiome in overall host metabolism and dietary energy retrieval. A major function of the colonic bacteria is proposed to be in the 'secondary digestion' of dietary carbohydrates, unabsorbed sugars, and alcohols (Guarner & Malagelada, 2003), as well as in metabolism of the lumen mucus (Arumugam et al., 2011) and absorption of dietary calcium, magnesium, and iron (Nicholson et al., 2012). Changes in host metabolic requirements, for example in pregnancy, have accordingly been associated with corresponding changes in composition of the microbiota (Koren et al., 2012).

Microbial composition has been linked for some time with development of the gastrointestinal tract, including epithelial cell differentiation (Gordon, Hooper, McNevin, Wong & Bry, 1997) and crypt cell production (Alam, Midtvedt & Uribe, 1994). Given the intrinsic exposure to microbial products, it is perhaps unsurprising that associations have also been made to the development of the immune system. Upon exposure to microbial products, intra-epithelial lymphocyte populations expand and promote antibody production in gut-associated lymphoid tissues (GALT) (Guarner & Malagelada, 2003).

At least 80% of human plasma cells are located in the GALT and produce more IgA than all other immunoglobulin isotypes combined (Fagarasan, Kawamoto, Kanagawa & Suzuki, 2010), much of which is secreted into the gut lumen and can be seen coating bacteria in stool samples (van der Waaij, Limburg, Mesander & van der Waaij, 1996). IgA secretion is suggested to prevent bacterial translocation through neutralisation (see Section 1.2.4) (Brenchley et al., 2006) and to clear translocated microbial products (Reid et al., 1997), to mediate tolerance to intestinal microbial communities (D. A. Peterson, McNulty, Guruge & Gordon, 2007), and to regulate the expansion of certain immunogenic species (Suzuki et al., 2004). A variety of other impacts have also been characterised, including control of immune homeostasis through polarisation of $CD4^+$ T cells to a Th1, Th17, or Treg phenotype (Mazmanian, Liu, Tzianabos & Kasper, 2005; Ivanov et al., 2009; Geuking et al., 2011; Atarashi et al., 2011), and control of natural killer T cell (NKT) populations (Olszak et al., 2012).

4.1.4 Aims

Initial work reported in this thesis has centred on the impact of the endogenous eMLV *Emv2* on the development of the immune response, a little-studied area of research. A converse control, whereby the immune system might act to control RE or ERV expression, is a tempting possibility and is poorly addressed in the literature.

This section investigates the relevance of the adaptive immune system in the control of REs and uses the advantages offered by knockout and transgenic mouse systems to dissect various cellular and humoral requirements.

The gut microbiota represents a large area of current study, but the field is in relative infancy and its potential significance in mouse experimentation (and its application to humans) is likely yet to be fully appreciated. Thus the possibility that the gut microbiota, similarly to MMTV pathogenesis, influences the expression of REs and ERVs is a tantalising prospect. This link was be investigated alongside the relevance of the adaptive immune system.

4.2 Results

4.2.1 eMLV transcripts are increased in immunodeficiency

Mice deficient for *recombination activating gene 1* (*Rag1*) are unable to perform the somatic V(D)J recombination events necessary for the development of variable antigen receptors on both B and T cells. In the absence of either the TCR or BCR, appropriately, cells are unable to mature and development is blocked at an early stage. Whilst these mice lack an adaptive immune response, the innate immune system is not known to be compromised, and these mice thus represent a useful tool to investigate the impact of the adaptive immune system at a high level.

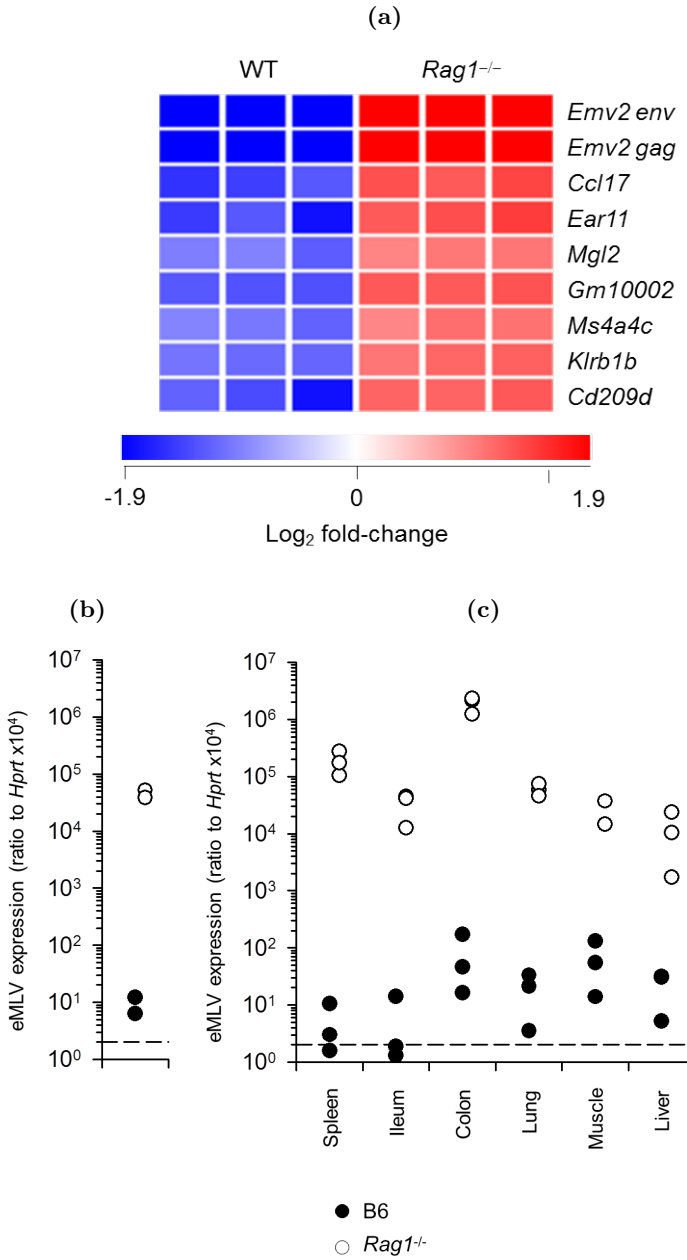
Transcriptional profiling of FACS-sorted peritoneal macrophages, an invariant cell type between wild-type B6 and *Rag1*$^{-/-}$ mice, was performed with Affymetrix Mouse Gene 1.0 ST microarrays. Surprisingly, the two transcripts with the highest fold-increase in macrophages derived from *Rag1*$^{-/-}$ mice were found to correspond to eMLV *env* and *gag* transcripts (transcription clusters 10,582,545 and 10,582,549, respectively, Fig. 4.1a). At the B6 genome level, both transcripts can be matched to *Emv2* (see Section 3.1.5), with the most highly upregulated transcript, *Mela* (*Melanoma antigen*), corresponding to the *Emv2* spliced *env* mRNA product.

This expression was confirmed by qRT-PCR for eMLV spliced *env* in the same cell type (Fig. 4.1b), with around a 4-log increase in expression being seen in *Rag1*$^{-/-}$ samples. To investigate this in greater detail and to determine how widespread the increased eMLV expression was in immunodeficient mice, seven tissues were sampled and expression evaluated by qRT-PCR. Although expression of eMLV varied between tissues, *Rag1*$^{-/-}$ mice displayed consistently higher levels of eMLV expression than immunocompetent B6 controls (Fig. 4.1c).

Detailed analysis of the microarray profiling data using *REquest* (see Section 2.18), a program to mine RE expression from commercial microarray data, revealed further changes in the expression of ERVs. 10 transcription clusters found significantly upregulated ($\bar{x}$ 1.9-fold) in *Rag1*$^{-/-}$ peritoneal macrophages were found to correspond to endogenous MMTVs, and 4 to endogenous MLVs ($\bar{x}$ 1.7-fold upregulated), although probes did not distinguish between xenotropic, polytropic and modified polytropic (x/p/mpMLV) host tropisms.

These data gave an apparent link to wider changes in ERV expression in immunodeficiency and were thus investigated in greater detail. qRT-PCR primers reporting the expression of a variety of ERV families were firstly tested and compared in *Rag1*$^{-/-}$ and control B6 spleen samples (Fig. 4.2a). Whilst there was a trend towards increased expression of x/p/mpMLVs within *Rag1*$^{-/-}$ mice, only the difference in eMLV expression

Figure 4.1: eMLV expression is elevated in *Rag1*$^{-/-}$ mice

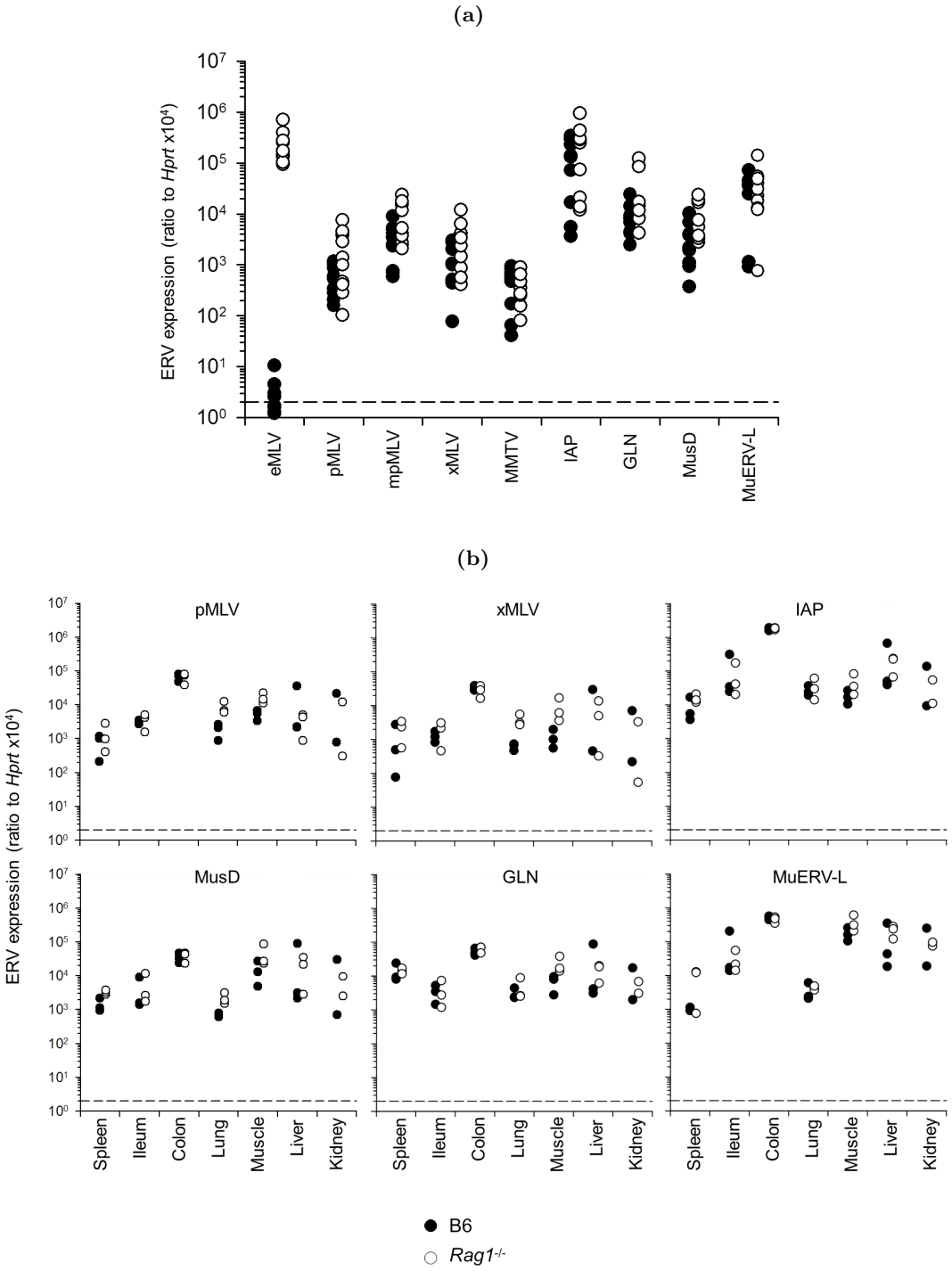


Peritoneal macrophages were isolated and FACS-sorted from B6 and *Rag1*$^{-/-}$ mice and transcription profiled by Affymetrix Mouse 1.0 ST microarray. (a) >4-fold upregulated genes in *Rag1*$^{-/-}$ compared to B6 for triplicate assays using cells isolated from 40 mice per group. (b) qRT-PCR with primers for eMLV spliced *env* in peritoneal macrophages. p=0.024 between groups. Symbols represent groups of 20 mice. (c) qRT-PCR for eMLV spliced *env* for the indicated tissues from B6 and *Rag1*$^{-/-}$ mice. p=0.02 for spleen, p=0.032 for ileum, p=0.004 for colon, p=0.001 for lung, p=0.016 for muscle, and p=0.009 for kidney. Symbols in (c) represent individual mice. Dashed lines represent the detection limit.

was significant. The expression of these ERV families was further compared across the previously obtained tissue samples. Significant upregulation of pMLV expression was

noted in the lungs and quadriceps muscle tissue of *Rag1*$^{-/-}$ mice, along with xMLV and MusD element expression in the lungs (Fig. 4.2b).

Figure 4.2: ERV expression across B6 and *Rag1*$^{-/-}$ tissues



(a) qRT-PCR for the indicated ERVs within spleens of B6 and *Rag1*$^{-/-}$ mice. $p<0.001$ for eMLV. (b) qRT-PCR for the indicated ERV for a range of tissues from B6 and *Rag1*$^{-/-}$ mice. $p=0.033$ and $p=0.035$ for pMLV in the lungs and muscles, respectively, $p=0.024$ for xMLV in the lungs, and $p=0.037$ for MusD in the lungs. Each symbol represents an individual mouse. Dashed lines denote the limit of detection.

Thus, a lack of T and B lymphocytes in *Rag1*$^{-/-}$ mice could be associated with large increases in eMLV expression across a range of tissues. Although noticeably more modest, increases in the expression of other ERV families were also seen.

4.2.2 eMLV transcripts are increased in antibody deficiency

Use of *Rag1*$^{-/-}$ mice did not allow distinction of effects mediated by either T of B cells, thus the exact cause of elevated eMLV expression was dissected further with a number of more selectively lymphocyte deficient mice.

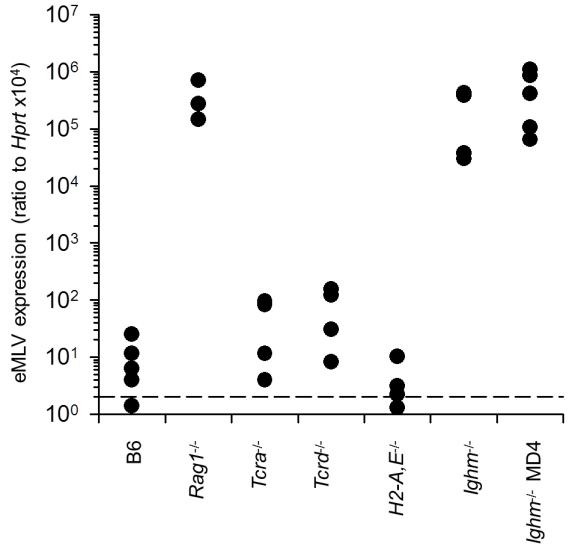
Tcra$^{-/-}$ and *Tcrd*$^{-/-}$ mice lack $\alpha\beta$ and $\gamma\delta$ T lymphocytes due to targeted disruption of the TCRα and δ chains, respectively. *H2-A,E*$^{-/-}$ mice are unable to present peptide on MHC class II molecules, and consequently have greatly impacted CD4^{+} T cell development. Around 5% of CD4^{+} T cells remain in these mice, but display a greatly disturbed phenotype and are unable to contribute to germinal centre reactions. In this way, both *Tcra*$^{-/-}$ and *H2-A,E*$^{-/-}$ display an inability to produce class-switched antibodies through a lack of CD4^{+} T cell function. *Ighm*$^{-/-}$ mice have a targeted disruption of the immunoglobulin heavy chain, preventing production of functional BCRs and hence blocking B cell development. *Ighm*$^{-/-}$ MD4 mice are transgenic for a Hen-egg lysozyme (HEL)-specific BCR, allowing the development of B cells displaying a single specificity.

Of these strains, only *Ighm*$^{-/-}$ and *Ighm*$^{-/-}$ MD4 mice displayed equivalent levels of eMLV expression to *Rag1*$^{-/-}$ (Fig. 4.3). *Tcra*$^{-/-}$, *Tcrd*$^{-/-}$, and *H2-A,E*$^{-/-}$ mice, in spite of varying T cell deficiencies and inabilities to produce class-switched antibodies, all maintained low levels of eMLV expression that were equivalent to wild-type B6 (Fig. 4.3).

To extend these findings to other antibody deficient mice, *Tlr7*$^{-/-}$ and *Myd88*$^{-/-}$ were studied. TLR signalling has been associated with the control of B cell responses, and several studies have linked deficiencies with impaired responses to immunisation and even retroviral infection (Browne & Littman, 2009; Browne, 2011). However, detailed characterisation of these mice has identified significant global deficiencies in levels of natural antibodies (Pasare & Medzhitov, 2005; Demaria et al., 2010), complicating these conclusions. Both *Tlr7*$^{-/-}$ and *Myd88*$^{-/-}$ mice displayed high levels of eMLV spliced *env* mRNA, suggesting the lowered levels of natural antibodies in these mice are insufficient to control expression (Fig. 4.4a). This result was mirrored in tissue samples collected from specific pathogen-free (SPF) mice housed at the Centre d'Immunologie de Marseille-Luminy (CIML) (Fig. 4.4b).

Together, these data suggested a crucial requirement for B cells, manifested through their provision of specific antibodies, in the control of eMLV expression in immun-

Figure 4.3: eMLV expression in selectively immunodeficient mice



qRT-PCR for eMLV spliced *env* in spleens from the indicated lymphocyte deficient strains. $p<0.001$ between B6 and *Rag1*$^{-/-}$, *Ighm*$^{-/-}$, and *Ighm*$^{-/-}$ MD4 mice. Each point represents an individual mouse. Dashed line indicates the limit of detection.

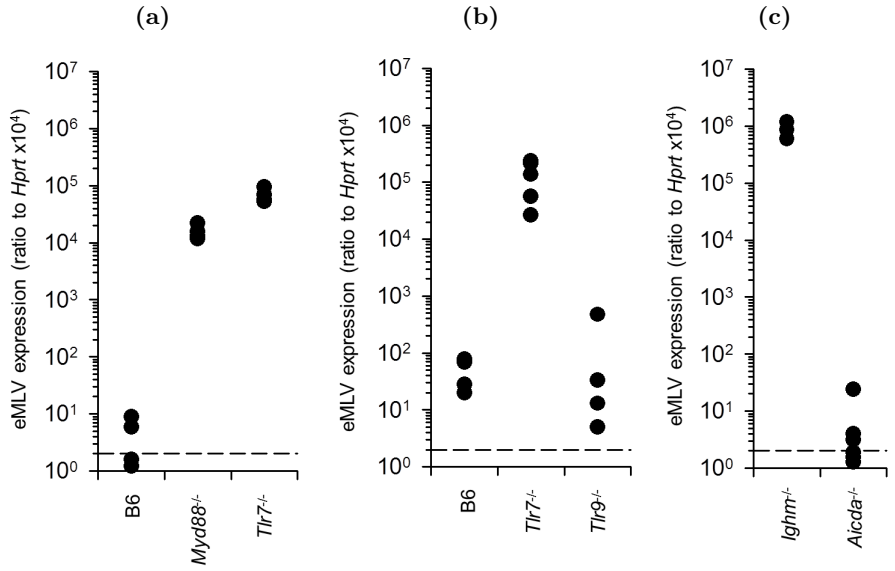
odeficiency. In this way, varyingly antibody deficient mice displayed elevated eMLV expression levels. Given the control of eMLV expression in *Tcra*$^{-/-}$, *Tcrd*$^{-/-}$, and *H2-A,E*$^{-/-}$ mice, this strongly suggested that these antibodies could be produced in the absence of CD4^{+} T cell help. To confirm these data and to define the specific immunoglobulins required in the control of eMLV expression, SPF activation-induced cytidine deaminase (AID)-deficient *Aicda*$^{-/-}$ mice were tested at the NCI Center for Cancer Research (NCI). These mice are unable to recombine gene regions necessary for production of antibodies other than IgM.

In agreement with the low eMLV expression seen in the absence of CD4^{+} T cell help, *Aicda*$^{-/-}$ mice were able to control eMLV levels (Fig. 4.4c). Thus, IgM alone appeared both necessary and sufficient for the control of eMLV expression.

4.2.3 Cell-surface MLV glycoprotein can be stained in antibody deficient mice

To further gauge the global levels of eMLV expression in high-expressing strains, varying cell types were assayed by FACS using the 83A25 antibody (Evans, Morrison et al., 1990) to stain MLV surface (SU) glycoprotein. With few exceptions, this neutralising rat-IgG2a antibody, derived from immunisation of a rat with polytropic F-MLV, reacts strongly with MLVs of all tropisms (Evans, Morrison et al., 1990). Cell types unaffected by *Rag1*-deficiency were examined in these mice, whilst T and B cell MLV SU levels

Figure 4.4: eMLV expression in varying antibody deficiencies



qRT-PCR for eMLV spliced *env* mRNA from the spleens of the indicated strains. (a) $p<0.001$ between B6 and either *Myd88*$^{-/-}$ or *Tlr7*$^{-/-}$ mice. (b) $p=0.003$ between B6 and *Tlr7*$^{-/-}$. (c) $p=0.024$ between *Ighm*$^{-/-}$ and *Aicda*$^{-/-}$. Each symbol represents an individual mouse. Dashed lines denote detection limits.

were assayed using *Ighm*$^{-/-}$ MD4 mice.

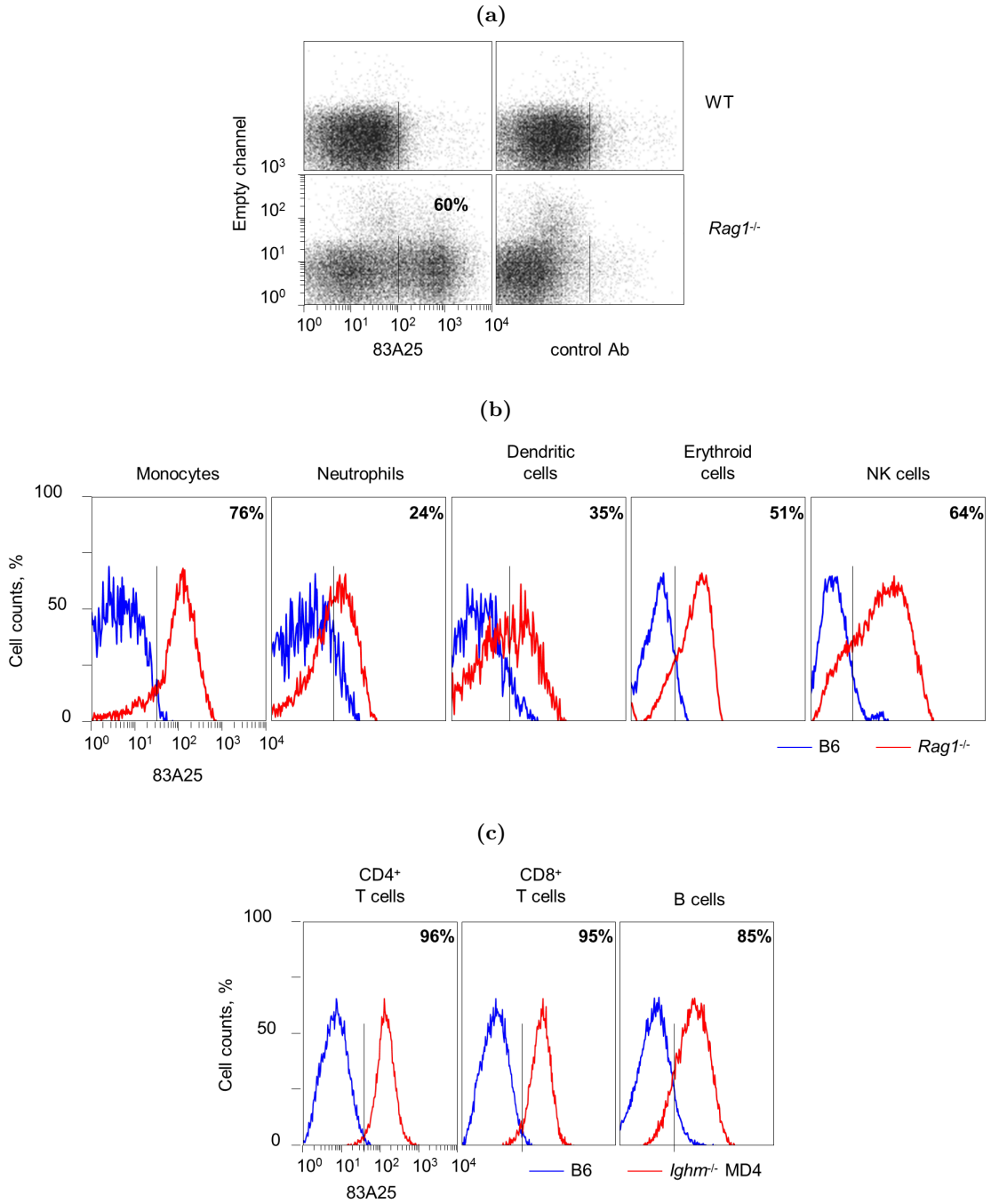
Determined as a whole, 60–80% of splenocytes from *Rag1*$^{-/-}$ mice stained positively with 83A25 (Fig. 4.5a), with only background levels of staining being seen in control animals. Levels of staining varied across cell types within individual *Rag1*$^{-/-}$ mice but remained consistently higher than the equivalent cell type from B6 controls (Fig. 4.5b). CD4^{+}, CD8^{+} and B cells from *Ighm*$^{-/-}$ MD4 were also seen to stain to a high level (Fig. 4.5c), showing that expression could also be seen in lymphocytes.

Not only is the expression of eMLV (and potentially other endogenous MLVs) increased in antibody deficiency, therefore, but this expression can be linked to formation of protein products that are processed and transported, allowing staining on the cell surface. Further, increased levels of staining can be seen in all cell types studied from *Rag1*$^{-/-}$ or *Ighm*$^{-/-}$ MD4 mice, suggesting a wide-ranging, or indeed global increase in expression in these mice.

4.2.4 Infectious virus can be isolated from *Rag1*$^{-/-}$ mice

Such high levels of eMLV mRNA expression, coupled with the large range of tissues and cells staining with 83A25, was partially suggestive of an infection scenario rather than dysregulation of endogenous MLVs. *Emv2*, the only endogenous eMLV of B6

Figure 4.5: MLV SU levels across cell types



FACS analysis of MLV SU expression using the 83A25 antibody. (a) 83A25 staining in whole spleen suspension from B6 and *Rag1*$^{-/-}$ mice. MLV SU expression in the indicated cell types of (b) *Rag1*$^{-/-}$ and (c) *Ighm*$^{-/-}$ MD4 mice. Numbers in figures show percentage positivity. Plots are representative of 4 mice per group.

background mice, whilst replication incompetent and restricted by the action of *Fv1*^{b} in B6 mice (see Section 3.1.5), is known to have recombined in several settings to produce infectious exogenous virus. In this way, *Emv2*-derived sequences are thought to have contributed almost totally to MelARV (Hayashi et al., 1992; Pothlichet, Man-

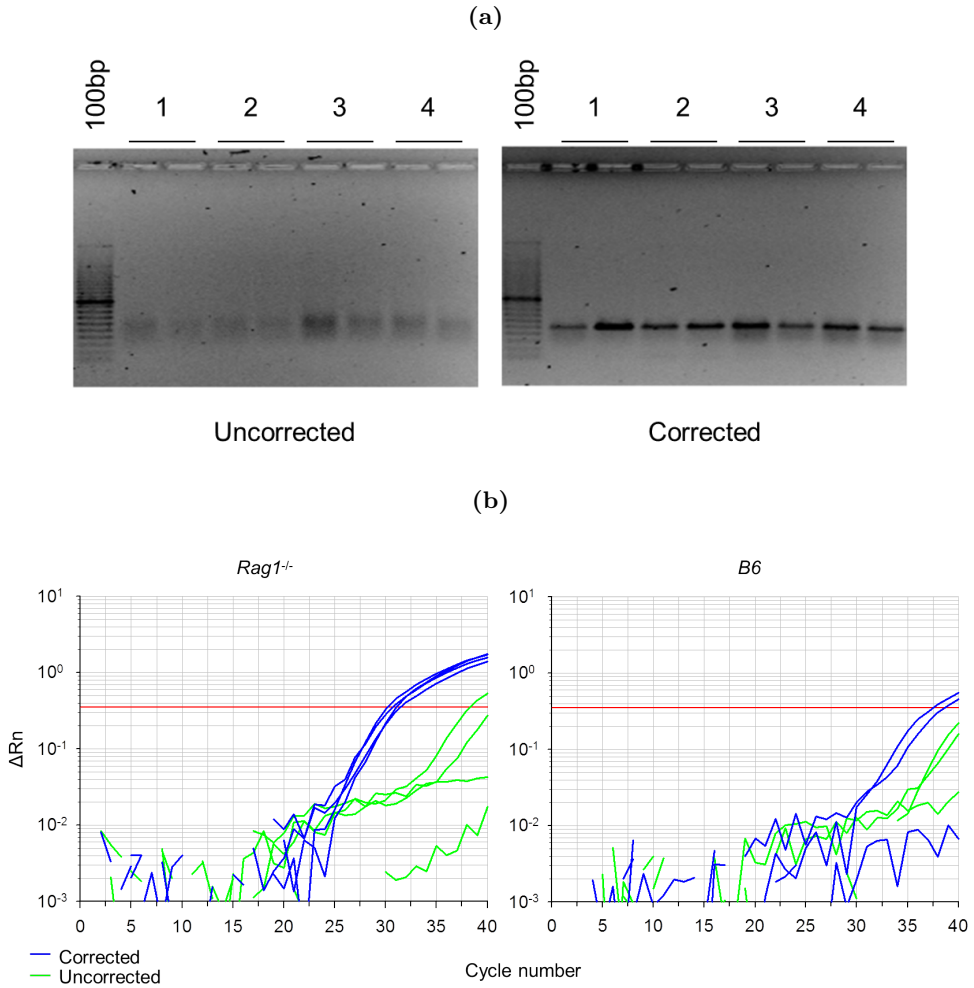
geney & Heidmann, 2006; M. Li et al., 1999), but also to exogenous eMLVs causing myeloid leukeamias in BXH-2 mice (Turcotte et al., 2005) and to the BM5eco ecotropic helper component of the LP-BM5 murine AIDS (MAIDS) retroviral complex (Chattopadhyay, Sengupta, Fredrickson, Morse III & Hartley, 1991; Coppola et al., 1995). Further, provided the eMLV *env* was retained in a recombination event, the spliced *env* primers used would not discriminate between functional and non-functional transcripts derived from either novel integrations of replication competent exogenous virus, or*Emv2*, respectively.

Production of replication competent eMLVs would firstly necessitate expression and co-packaging of two endogenous retroviral sequences within a single cell, followed by recombination during reverse transcription in a target cell. In the case of *Emv2*, either a single large recombination event could correct the G-3576-C mutation and convert the capsid from N- to B-tropism, or multiple events would be required to fulfil the same function. To initially test the possibility that infectious eMLVs derived from *Emv2* were present in *Rag1*$^{-/-}$ mice, primers were designed to distinguish between transcripts containing G or C at position 3576. cDNA was produced using a reverse transcription reaction initiated with a separate primer specific to ecotropic *env*, and PCR (Fig. 4.6a), followed by qRT-PCR (Fig. 4.6b), suggested that the majority of these transcripts were corrected at position 3576.

This finding implied the presence of infectious eMLV within these mice, thus the presence of infectious virus was directly tested through incubation of sera with permissive *Mus dunni* cells. Remarkably, sera isolated from both young (6 weeks) and older (25 weeks) *Rag1*$^{-/-}$ mice, but not age-matched control B6 mice, were confirmed to contain infectious exogenous MLVs, termed *Rag1*$^{-/-}$ mouse-associated retroviruses (RARVs). Three RARVs (2–4) were isolated from 6 week old *Rag1*$^{-/-}$ serum samples and a further four (5–8) from 25 week old mice. These RARVs were capable of replicating within murine cell lines *in vitro* and, on sequencing of the region surrounding the *Emv2* point mutation, indeed showed reversion at position 3576 (Fig. 4.7a), as previously suggested. Conversion to B-tropism was assessed through determination of replicative potential in B- and N-3T3 cells, and can be expressed as a ratio of B:N tropism (Figs. 4.7b and 4.7c). All seven RARVs were seen to display B-tropism, implying a likely recombination event or events within the *Emv2 gag*.

To assess the means by which tropism was altered in the viral isolates, all seven RARVs were sequenced and analysis of amino acid sequences from the viral capsid confirmed that all RARVs were B-tropic (Kozak & Chakraborti, 1996) (Fig. 4.8a). Interestingly, sequence comparison suggested that RARVs 2–4, derived from young *Rag1*$^{-/-}$ mice, were highly similar, but that sequences diverged substantially with increasing age. This was supported by phylogenetic analysis, with viruses clustering according to the

Figure 4.6: Relative quantitation of corrected and non-corrected eMLV transcripts

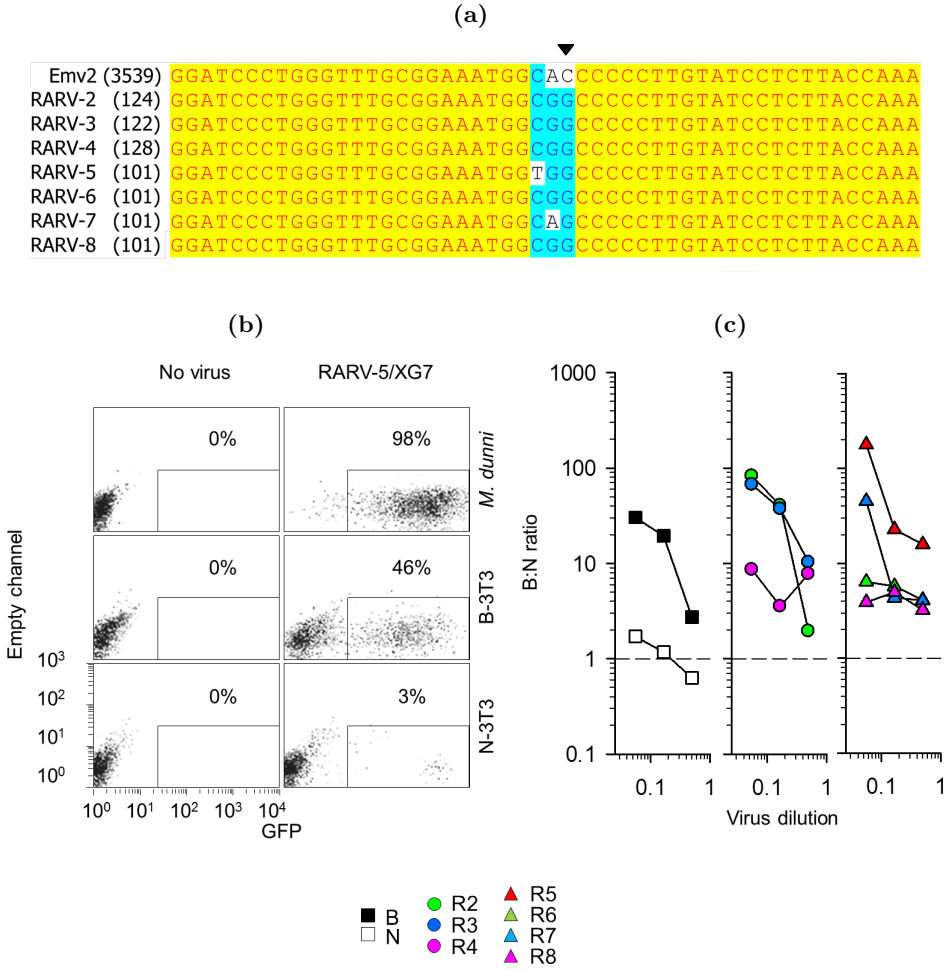


(a) PCR and (b) qRT-PCR comparing corrected and uncorrected eMLV transcripts. eMLV mRNA was reverse transcribed using a single reverse primer within the ecotropic *env* and primers produced at the site of the *Emv2* point mutation. 4 individual mice (duplicates prepared with separate reverse transcription kits in (a)) are shown.

age of the mouse from which they were isolated (Fig. 4.8b).

Bootscanning against endogenous MLV sequences obtained from the literature was performed to define the recombination events necessary for reversion of the point mutation and for conversion to B-tropism, as well as to identify the cause of divergence over time. The same pattern of recombination events was seen in RARVs 2–4, highlighting their similarity and suggesting the existence of a single exogenous infectious virus within the colony (Fig. 4.9). Recombination with *Xmv41* or *Xmv43* was responsible for the correction of the G-3576-C inactivating mutation (Fig. 4.9). Supporting this finding, *Xmv43*, otherwise known as *Bxv1*, has previously been identified as an active recombination partner in a variety of settings (Stoye, Moroni & Coffin, 1991; Hook et al.,

Figure 4.7: Point mutation correction and altered tropism of RARV isolates



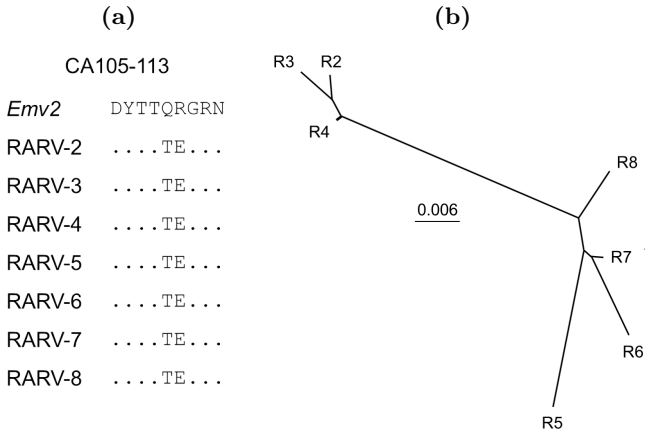
(a) Sequence alignment of the area surrounding the *Emv2* point mutation from RARVs 2–8. Arrow indicates site of the *Emv2* point mutation. (b) Detection of infectious MLVs isolated from the plasma of a *Rag1*$^{-/-}$ mouse. Presence of an infectious virus in the culture allows pseudotyping of a GFP-expressing retroviral vector and subsequent quantitation using a FACS-based assay. (c) *Fv1* tropism determination of isolated RARVs, shown as the ratio of infectivity on B- and N-3T3 cells. B- and N-tropic F-MLV are shown for comparison.

2002).

Further recombinations, likely involving the same xenotropic partner, were also revealed by the bootscanning, although their relevance has not been investigated and does not impact areas previously defined as defining viral tropism. As such, a separate recombination involving one of *Xmv8*, *Xmv10*, or *Xmv13* was required for conversion from N- to B-tropism, and is again seen in all viral isolates (Fig. 4.9).

As suggested by the separate grouping of viral isolates from young and older *Rag1*$^{-/-}$ mice, RARVs 5–8 all contained further recombination events (Fig. 4.9). All events

Figure 4.8: Sequence analysis of RARV isolates



(a) Amino acids at positions 105–113 of the capsid, determining viral tropism. Dots indicate identity. (b) Phylogenetic tree of RARV isolates. Scale represents the probability of base substitution per site.

involved pMLVs endogenous to the B6 genome and consisted of varying substitutions of *env* and the neighbouring region of *pol* (Fig. 4.9). Further alterations to *pro-pol* were made in half of the isolates, although not necessarily with the same partner involved in the *env* recombination (Fig. 4.9).

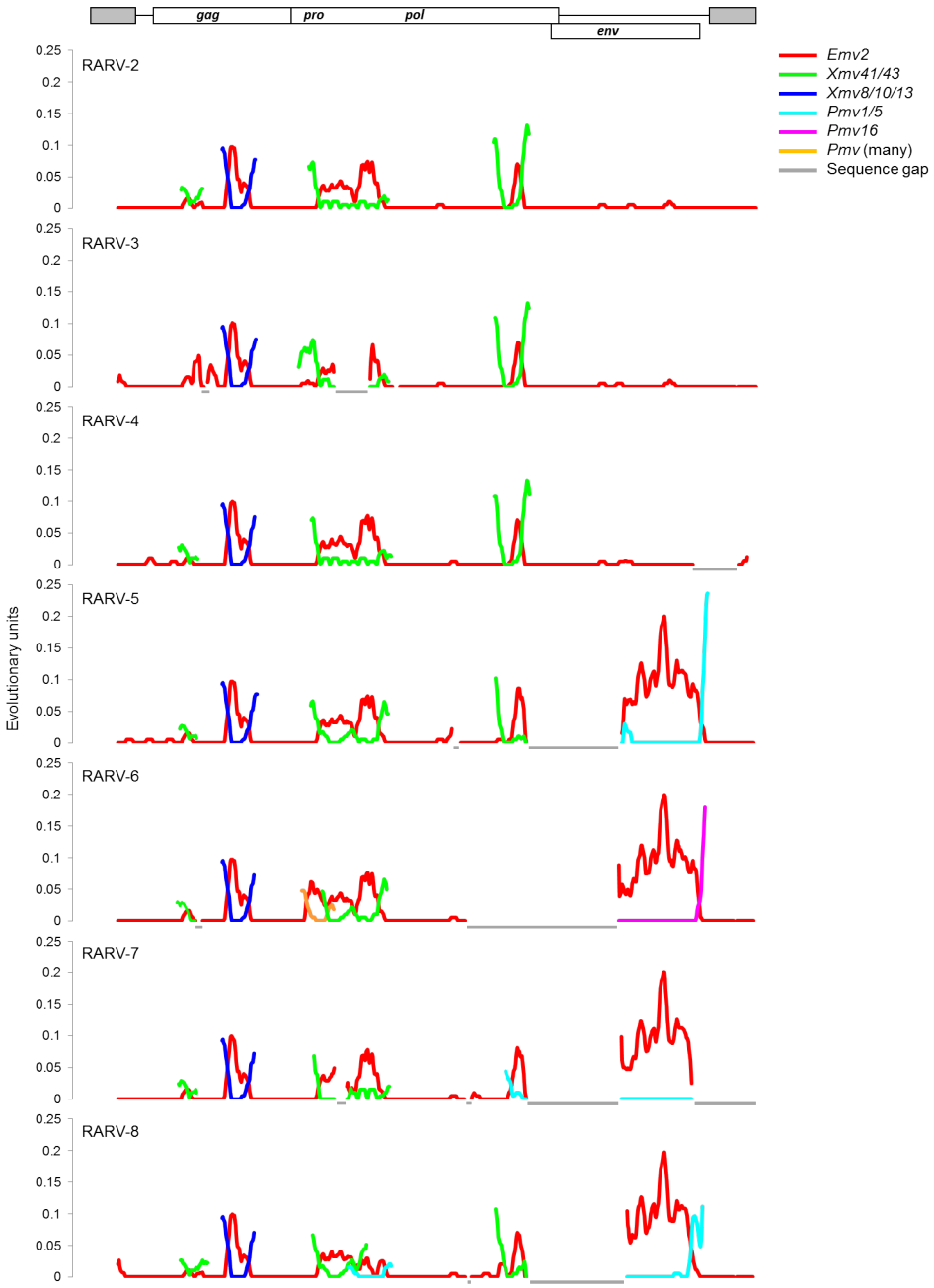
Thus, the *Rag1*$^{-/-}$ colony appeared to harbour an infectious eMLV derived from a recombination event involving *Emv2* as the primary donor and a further two endogenous MLVs. Importantly, both the sequence backbone and the regions of recombination could be matched to MLVs endogenous to the B6 genome by bootscanning, ruling out the potential of 'contamination' of the breeding line with an exogenous virus. Over time, and apparently within each mouse, independent recombination events switch the ecotropic *env* of *Emv2* for an *env* with a polytropic host range. Consistent with this hypothesis, the majority of transcription of ecotropic *gag-pol* mRNA appears to derive from G-3576-C-corrected viruses, and 4-fold higher levels of eMLV *env* than eMLV *gag* can be seen in analysis of the microarray data from peritoneal macrophages. eMLV *env* is shared between *Emv2* and early RARVs, whereas eMLV *gag* is only found completely in *Emv2* itself and is substantially altered in all RARVs isolated.

4.2.5 RARV infection occurs by vertical transmission to neonates

Analysis of RARV sequencing data suggested that a single infectious exogenous retrovirus, represented by RARVs 2–4, existed in the *Rag1*$^{-/-}$ colony held in the NIMR SPF breeding facilities. Subsequent to infection, *de novo* changes appear to occur within each mouse to result in the formation of infectious virus with a polytropic *env* sequence.

To confirm this theory, *Emv2*$^{-/-}$ null congenics generated for the previous study (see

Figure 4.9: Bootscan of RARV isolates against endogenous B6 MLVs

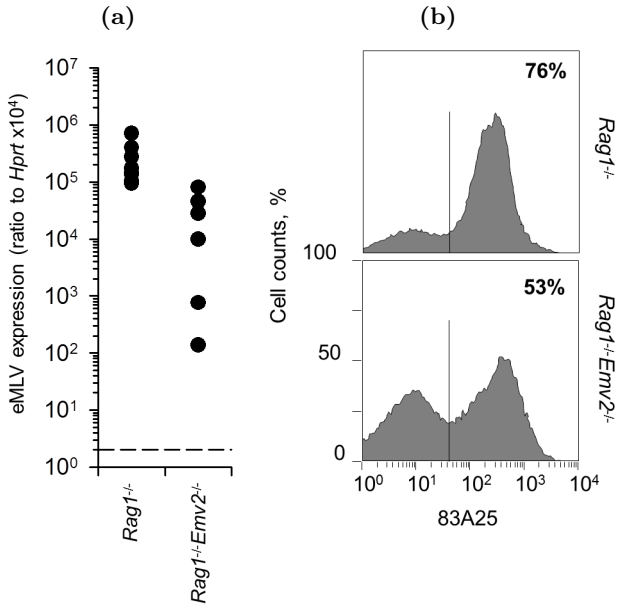


Bootscanning of RARV sequences against MLVs endogenous to the B6 genome. Scale represents evolutionary distance. Where sequence diverges from that of *Emv2*, suggestive of a recombination event, the best matching sequence or sequences, where similarity does not allow distinction, are shown.

Chapter 3) were used for a series of crosses. A colony was firstly established through use of an *Emv2*$^{-/-}$ male and a *Rag1*$^{-/-}$ female, followed by the intercrossing of selected F_1 progeny to produce a homozygous *Rag1*$^{-/-}$*Emv2*$^{-/-}$ line. Using this approach to remove *Emv2*, any eMLV detected must derive from infectious virus passed vertically from one of the parents. Both eMLV spliced *env* expression (Fig. 4.10a) and elevated

levels of MLV SU product (Fig. 4.10b) were seen in these mice, supporting the presence of an infectious exogenous virus.

Figure 4.10: eMLV are present in the absence of *Emv2*

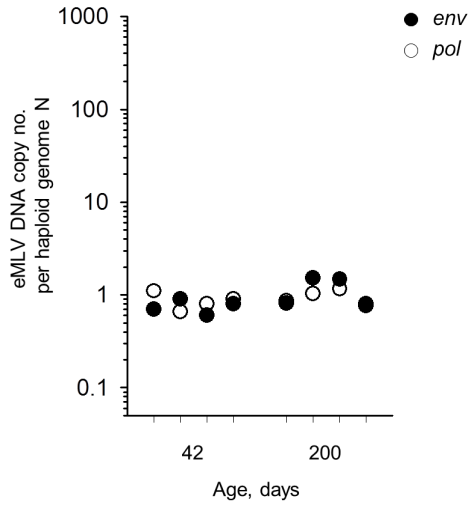


(a) qRT-PCR for eMLV spliced *env* in *Rag1*$^{-/-}$ and *Rag1*$^{-/-}$*Emv2*$^{-/-}$ mice. eMLV present in the latter must derive from an infectious virus. Each point represents a single mouse. (b) FACS staining of MLV SU in the same strains. Data are representative of 9 mice per group.

The *Rag1*$^{-/-}$ colony used within NIMR, originally obtained from The Jackson Laboratory (JAX), has been maintained at the institute for 45 filial generations (F_{45}). All mice obtained from an external source are re-derived by implantation of embryos into pseudopregnant female recipients prior to introduction into the specific pathogen free (SPF) animal facilities. This process ensures the removal of adventitious agents, including retroviruses that are not carried in the germ-line. At an unknown point since the re-derivation of the *Rag1*$^{-/-}$ strain, therefore, an infectious eMLV has been generated *de novo* through the previously described recombination events. This raised the possibility of further germ-line eMLV integrations contributing to the expression seen in this strain, but also to expression within *Rag1*$^{-/-}$*Emv2*$^{-/-}$ mice.

Determination of germ-line eMLV copy number was performed by quantitative PCR, using the single-copy *Ifnar1* gene as a reference for sample concentration. To account for the possibility of both ecotropic and polytropic RARVs being integrated, both eMLV *pol* and *env* primers were used and subsequently compared. Despite an estimated sensitivity of 0.0003 eMLV copies per haploid genome, determined with the use of *Emv2*$^{-/-}$ mice, no significant increases in eMLV copy number (above *Emv2*, at 1 copy per haploid genome) were seen in the germ-line of multiple *Rag1*$^{-/-}$ mice (Fig. 4.11).

Figure 4.11: RARV integrations are absent from the germ-line



eMLV *env* and *pol* copies per haploid genome in *Rag1*$^{-/-}$ spleen cells. Columns represent individual mice. 1 copy (*Emv2*) would be expected per haploid genome in the absence of infectious virus.

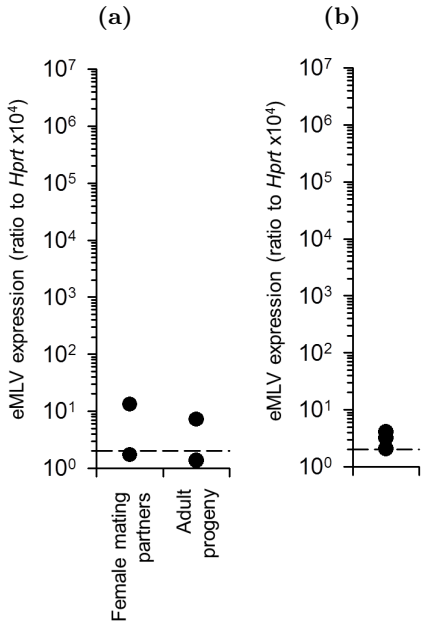
4.2.6 The implications of RARV infection

These data argued against the integration and germ-line inheritance of RARVs, further suggesting the presence of an infectious virus within the colony. Neonatal models of infection generally centre around breast-milk transmission and have long been studied for both MMTVs and MLVs (Law, 1962; Buffett, Grace Jr, DiBerardino & Mirand, 1969; Gardner, Chiri, Dougherty, Casagrande & Estes, 1979). To directly assess the point and route of RARV infection, a number of further crosses were made. These gave no evidence for sexual transmission of RARVs from a *Rag1*$^{-/-}$*Emv2*$^{-/-}$ male to two *Emv2*$^{-/-}$ females or to the adult progeny of the matings (Fig. 4.12a). Further, progeny from the reverse cross, a mating of *Rag1*$^{-/-}$*Emv2*$^{-/-}$ females with an *Emv2*$^{-/-}$ male, did not express high levels of eMLV at birth (Fig. 4.12b). Together, this suggested that RARV transmission occurred neonatally rather than *in utero*.

The infectious potential of RARVs, determined as eMLV DNA copy number, had previously been assessed in *Rag1*$^{-/-}$ mice (Fig. 4.11). Being sufficient for *Emv2*, however, it was possible that a certain level of interference, such as through blockade of *mCAT-1*, was being observed in these mice. To fully assess the the infectious potential of RARVs, therefore, eMLV copy number was also determined in *Rag1*$^{-/-}$*Emv2*$^{-/-}$ mice. Any DNA copies of eMLV present in these mice, in the absence of *Emv2*, must result from RARV infection.

Around 1 integration per haploid genome (2 integrations per somatic cell) was seen in these mice, with non-significant, but slight increases in copy number with increasing

Figure 4.12: RARV transmission



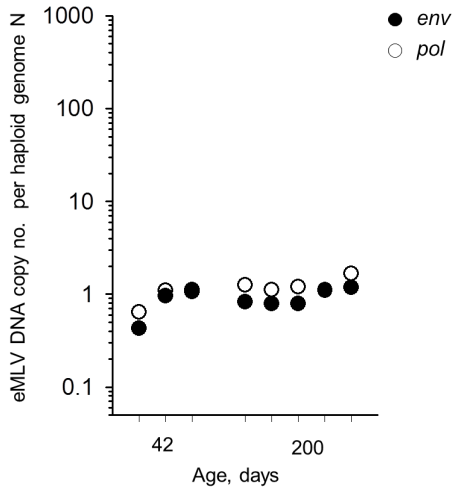
qRT-PCR assessment of eMLV spliced *env*. (a) A RARV-positive *Rag1*$^{-/-}$*Emv2*$^{-/-}$ male was mated with two RARV-negative *Emv2*$^{-/-}$ females and expression assessed 8 weeks after mating and in 6-week old progeny of the matings. (b) Two RARV-positive *Rag1*$^{-/-}$*Emv2*$^{-/-}$ females were mated with a RARV-negative *Emv2*$^{-/-}$ male and expression assessed in the livers of newborn (1 day) progeny. Each point represents a single mouse.

age (Fig. 4.13). Infectious potential was altered considerably in the absence of *Emv2*, therefore.

The level of infection noted here has great mutagenic potential over time. To determine the impacts of this burden on disease and tumourigenesis, cohorts of *Rag1*$^{-/-}$, *Rag1*$^{-/-}$*Emv2*$^{-/-}$, and control B6 mice were aged for over 12 months. Starting from around 180 days (6 months), significant numbers of both *Rag1*$^{-/-}$ and *Rag1*$^{-/-}$*Emv2*$^{-/-}$ mice showed signs of morbidity, characterised by weight loss, loss of condition, unresponsiveness, and partial paralysis. Examination of these mice revealed tumours, frequently associated with anaemia, in all cases (Fig. 4.14a). At the point of termination of the experiment, at day 380 (12.5 months), incidence of tumourigenesis in *Rag1*$^{-/-}$ mice was 67% (Fig. 4.14b), although an additional 5% showed signs of sub-clinical tumours upon internal examination. Tumour incidence occurred at an equal rate in the cohort of *Rag1*$^{-/-}$*Emv2*$^{-/-}$ mice, but was not observed in age-matched B6 wild-type controls.

Determination of the clonal origin of tumours from *Rag1*$^{-/-}$ mice indicated that cells from a variety of lineages could contribute to disease (Fig. 4.15a). Professional histological analysis indicated sheets of monomorphic cells of lymphoblastic character,

Figure 4.13: Potential for infection of *Rag1*$^{-/-}$*Emv2*$^{-/-}$ mice



eMLV *env* and *pol* copies per haploid genome in *Rag1*$^{-/-}$*Emv2*$^{-/-}$ spleen cells. Columns represent individual mice. All eMLV copies derive from infectious virus.

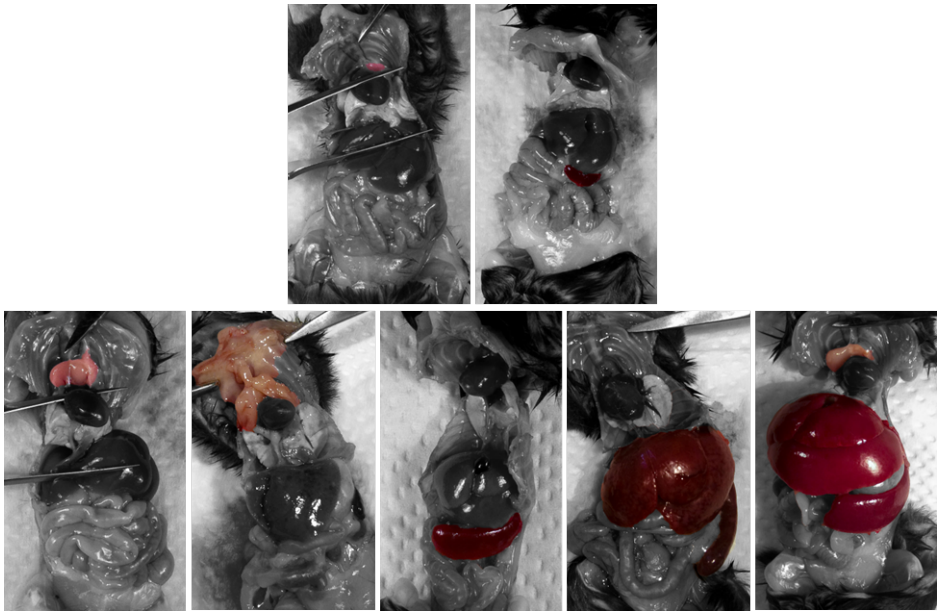
with negligible intervening stroma, frequent apoptoses and a high rate of mitosis, fitting with a retrovirus-induced leukaemia (Fig. 4.15b). In each case, a single MLV SU-expressing cell type was seen to contribute the majority of cells for each tumour sample studied (Fig. 4.15c). Consistent with the apparent clonal origin of cells within individual tumours, distinct and significant chromosomal aberrations were detected by array comparative genome hybridisation (Fig. 4.15d).

Supporting the retroviral origin of tumours in these mice, abundant extracellular C-type retroviral particles were detectable by transmission electron microscopy within tumour samples, but not within a spleen sample from a healthy *Rag1*$^{-/-}$ control (Figs. 4.16a and 4.16b). Significant increases in RARV integrations were also noted though analysis of eMLV DNA copy number (Fig. 4.16c). Interestingly, tumour 6 displayed elevated eMLV *pol* but not *env* transcripts, suggesting that where the majority of tumours appeared to be caused by ecotropic RARVs, this individual case originated from infection with a RARV containing an *env* of a different tropism. Together, this is indicative of copious amplification of RARVs within tumours.

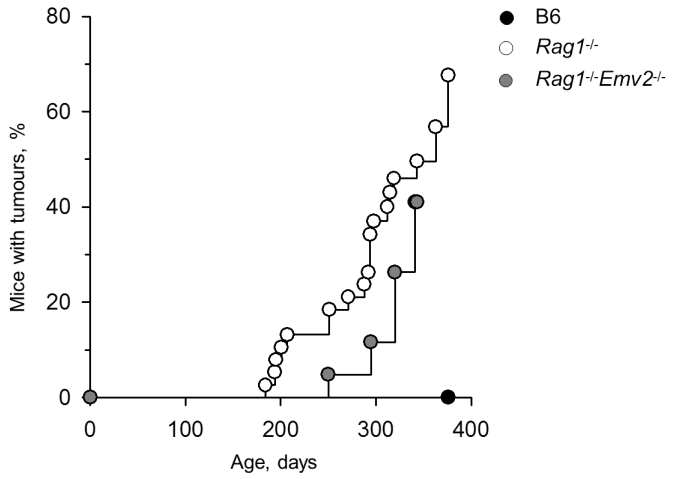
Overall, these data suggest a model of vertical inheritance of infectious eMLVs within antibody deficient mouse lines. Infectious eMLVs appear to be originally generated through the recombination of various MLVs endogenous to the B6 genome, resulting in the restoration of *Emv2* infectivity. Within each mouse, further individual recombination events can be seen that result in formation of virus with a polytopic *env*. The presence of infectious MLVs leads to retroviremia and, over time, results in oncogenesis in a manner similar to that well-defined in mouse strains carrying infectious germ-line MLVs (Stoye, Moroni & Coffin, 1991).

Figure 4.14: Tumour incidence in aged *Rag1*$^{-/-}$ and *Rag1*$^{-/-}$*Emv2*$^{-/-}$ mice

(a)

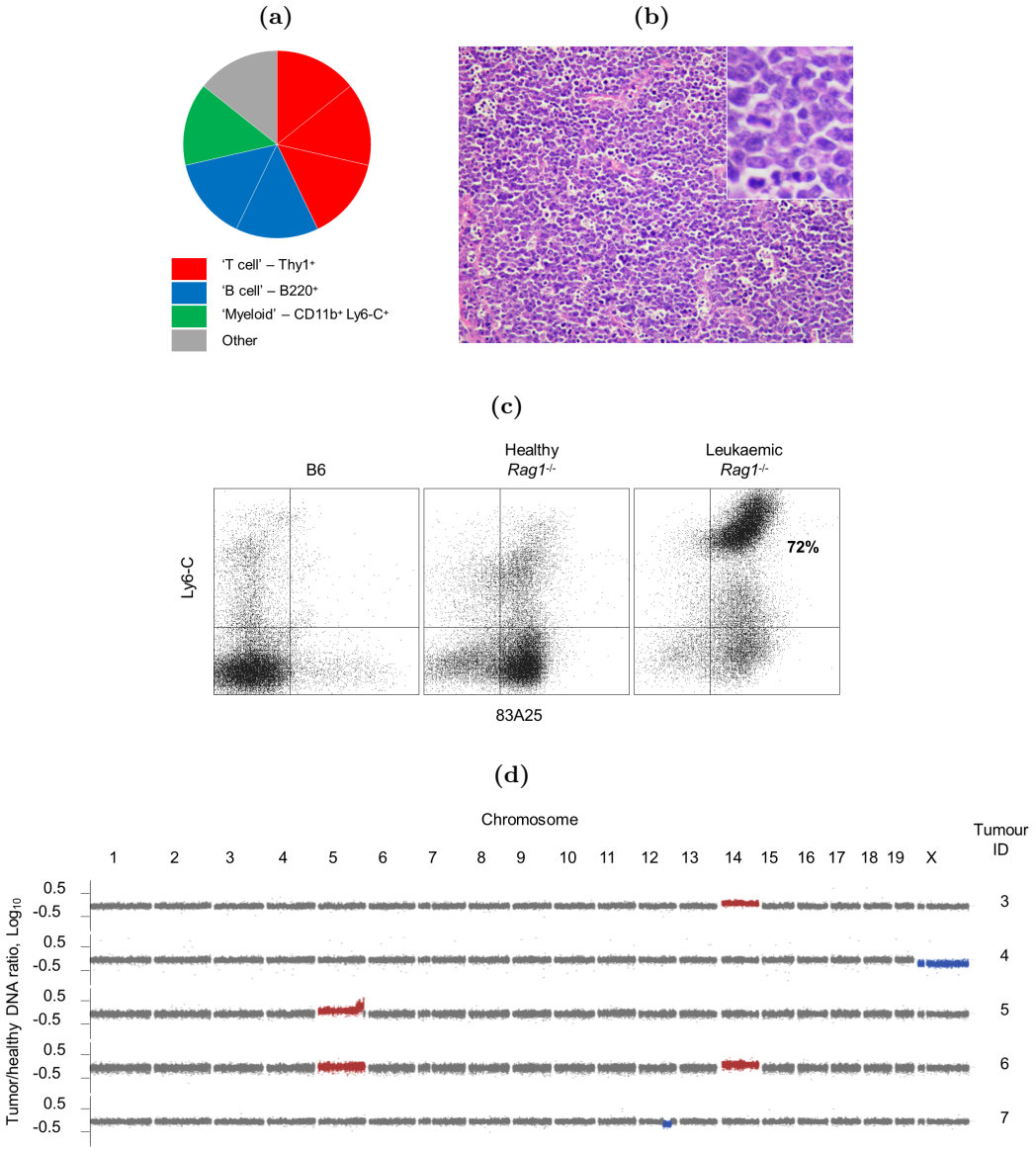


(b)



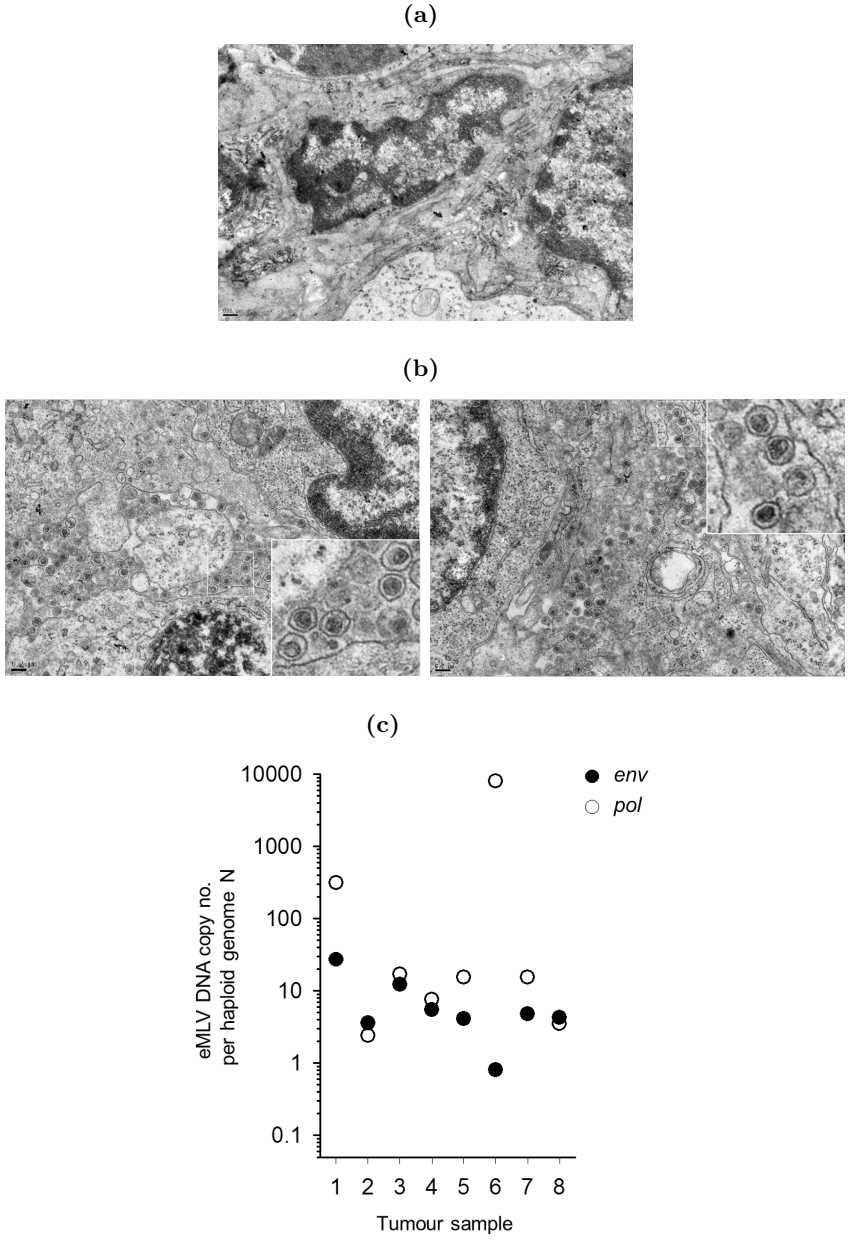
(a) Selectively coloured images of culled *Rag1*$^{-/-}$ mice. Top images show healthy mice, with the thymus (left) and spleen (right) highlighted to show reduced size in these mice in comparison to wild-type B6. Below are images of *Rag1*$^{-/-}$ mice presenting with leukaemias. Images show a thymic tumour, a thymic tumour with metastasis into the rib cage, a splenic tumour, a splenic and hepatic tumour, and a leukaemia impacting all three organs, respectively. (b) Tumour incidence in cohorts of B6 (36 mice), *Rag1*$^{-/-}$ (38 mice, $p<0.001$ to B6), and *Rag1*$^{-/-}$*Emv2*$^{-/-}$ (23 mice, $p<0.001$ to B6).

Figure 4.15: Characterisation of tumours from aged *Rag1*$^{-/-}$ mice



(a) FACS analysis of cell suspensions from *Rag1*$^{-/-}$ tumours to determine cellular origin. Cells representing >70% of total were taken as the predominant cell type. (b) Histological examination of a lymph node lymphoma from a leukaemic *Rag1*$^{-/-}$ mouse. Evident are sheets of monomorphic medium to large round cells with lymphoblastic character. Cells have distinct borders but little intervening stroma, minimal peripheral eosinophilic cytoplasm, and round nuclei with marginated clumped chromatin and conspicuous basophilic nucleoli. Frequent apoptoses and tangible body macrophages can be noted, and approximately five mitoses per high power field (insert). (c) FACS analysis of MLV SU expression in a *Rag1*$^{-/-}$ tumour of monocytic (Ly6-C^{+}) origin. (d) Array comparative genomic hybridization of DNA from individual splenic lymphomas from *Rag1*$^{-/-}$ mice. Spleen cells from a healthy *Rag1*$^{-/-}$ mouse were used in the comparison. Red and blue colours indicate chromosomal duplications and deletions, respectively.

Figure 4.16: RARV infection in tumours from aged *Rag1*$^{-/-}$ mice



Transmission electron micrographs from (a) a healthy *Rag1*$^{-/-}$ control and (b) a lymph node lymphoma from a leukaemic *Rag1*$^{-/-}$ mouse. No viral particles are visible in the healthy control, whereas numerous MLV-like retroviral particles can be noted in the extracellular spaces of the leukeamic tissue (inserts). (c) eMLV *pol* and *env* DNA copy number analysis for splenic lymphomas from individual *Rag1*$^{-/-}$ mice.

4.2.7 Antibodies are required for the control of systemic microbial products

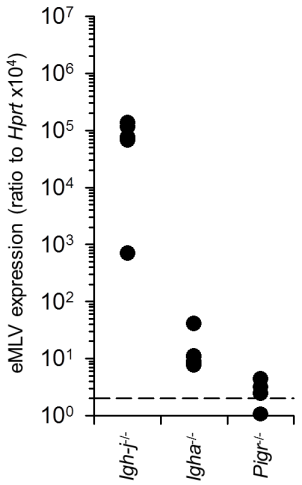
Although previous data had indicated a requirement for antigen-specific antibodies, the target of these was undetermined. Microbial products have a long-established role

in the promotion of MLV expression, both *in vivo* and *in vitro* (Greenberger et al., 1975; Moroni & Schumann, 1975; Moroni, Schumann, Robert-Guroff, Suter & Martin, 1975; Stoye & Moroni, 1983). In fact, *Xmv43* itself, a recombination partner possibly involved in the generation of RARVs in *Rag1*$^{-/-}$ mice, is known to be inducible by bacterial LPS (Stoye & Moroni, 1985).

Interestingly, serum LPS is known to be elevated in B cell deficient mice (Reid et al., 1997; Bourgeois, Hao, Rajewsky, Potocnik & Stockinger, 2008; Shulzhenko et al., 2011) and indeed expression of many ERV families, including eMLV, was most highly elevated in the colons of both B6 and *Rag1*$^{-/-}$ mice (Figs. 4.1c and 4.2b). It was thus possible that increased ERV expression, giving increased likelihood of genome co-packaging and subsequent recombination, was influenced by the presence of microbial products.

Control of microbes and microbial products may occur in either the gut lumen or, subsequent to translocation, within the systemic circulation. IgA has a well defined role in both settings (see Section 4.1.3), but previous data indicated that IgM alone appeared necessary and sufficient for the control of eMLV expression (Fig. 4.4c). As well as IgA, however, pentameric IgM is also excreted into the gut lumen, thus the exact requirement and action of both IgA and IgM was further investigated. Mice specifically deficient for IgA (*Igha*$^{-/-}$) or the polymeric immunoglobulin receptor (pIgR, *Pigr*$^{-/-}$), required for transcytosis of dimeric IgA and pentameric IgM across mucosal surfaces, were obtained from the the HCI Rodent Centre (RCHCI, ETH Zürich) and analysed for eMLV expression.

Figure 4.17: The affect of IgA deficiency and impaired immunoglobulin transcytosis on eMLV expression

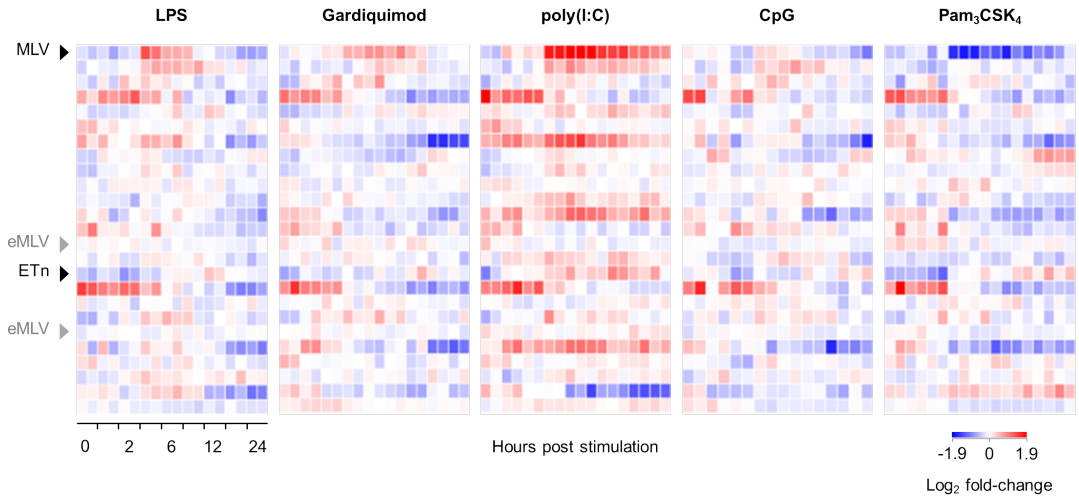


qRT-PCR for eMLV spliced *env* in spleens from the indicated strains from the RCHCI SPF animal facility. Each point represents an individual mouse and the dashed line shows the limit of detection. $p<0.001$ between *Igh-j*$^{-/-}$ and *Pigr*$^{-/-}$ mice.

As expected from the absence of a requirement for $CD4^+$ T cell help and the specific requirement outlined for IgM, *Igha*$^{-/-}$ mice controlled eMLV expression in comparison to B cell deficient positive controls (Fig. 4.17). Interestingly, *Pigr*$^{-/-}$ mice also controlled expression, suggesting that the required specific IgM antibodies were required for control of a systemic antigen (Fig. 4.17).

To investigate in greater detail the influence of microbial products on REs and ERVs, *REquest* was again used to identify probesets reporting their expression from a publicly available microarray dataset (Amit et al., 2009). The transcriptional profile of primary mouse dendritic cells (DCs) was assessed over a time course following treatment with Pam_3CSK_4, a synthetic bacterial lipoprotein mimic, polyinosine-polycytidylic acid (poly(I:C)), a double-stranded RNA, LPS, an outer membrane protein from Gram-negative bacteria, gardiquimod, a small-molecule agonist, and CpG, a synthetic single-stranded DNA. Respectively, these are agonists of TLR1/2, TLR3, TLR4, TLR7, and TLR9.

Figure 4.18: RE and ERV expression in mouse dendritic cells following TLR stimulation

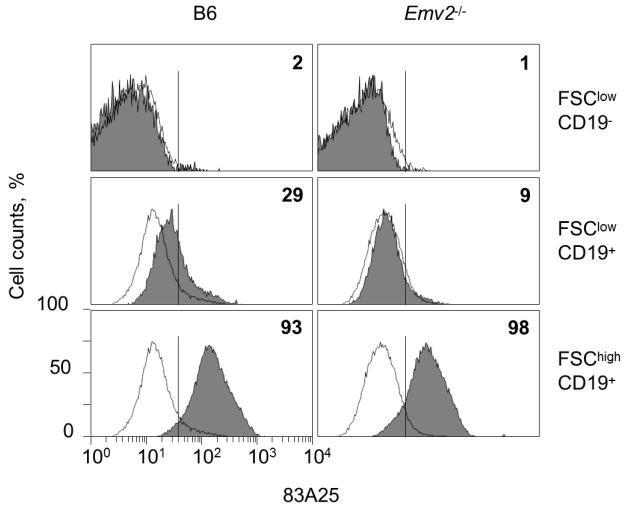


Heatmap of probesets found to correspond to REs and ERVs with *REquest* from the Affymetrix Mouse Genome 430A microarray. Data were downloaded from a public database (EBI Array Express E-GEOD-17721). Data are derived from a timecourse dataset of B6 BMDCs stimulated with various TLR agonists. Black arrows denote probesets that are significantly upregulated (p<0.05) by more than 2-fold by at least one stimulus. Grey arrows denote probesets corresponding to *Emv2*.

Expression of non-ecotropic endogenous MLVs was significantly increased on treatment with LPS and poly(I:C), yet suppressed by Pam_3CSK_4 (Fig. 4.18). This microarray probeset does not report the expression of eMLVs, which were not seen to be significantly affected by these treatments (Fig. 4.18). Early transposon (ETn) elements were also significantly induced by poly(I:C) treatment. The induction of endogenous MLV

expression, not inclusive of eMLVs, was confirmed through independent work undertaken *in vitro*. B6 and *Emv2*$^{-/-}$ splenocytes were stimulated with LPS and were shown to display comparable increases in MLV SU staining (Fig. 4.19).

Figure 4.19: MLV SU staining on LPS treatment of splenocytes



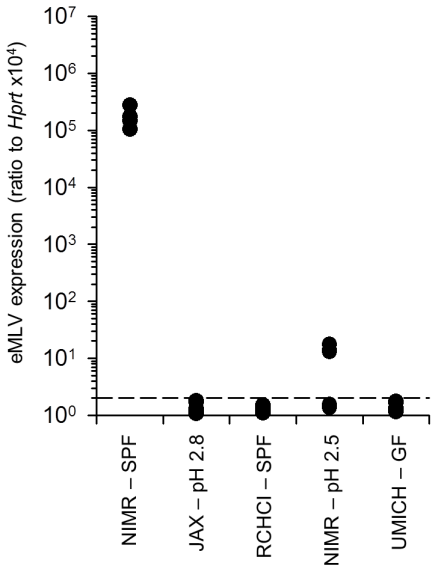
FACS analysis of MLV SU expression in B6 or *Emv2*$^{-/-}$ splenocytes prior to (open histograms) and after 48 hours of stimulation with 10 μg/ml LPS (grey histograms). Cells are grouped according to their forward scatter (FSC) profile, with blasting cells having the highest expression of MLV SU. Numbers represent percentages of cells staining positively with 83A25. Data are representative of 2 donors analysed in duplicate.

Thus, microbial products were seen to be capable of inducing endogenous non-ecotropic MLV expression, in agreement with previous data suggesting elevated expression in the colons of both wild-type B6 and *Rag1*$^{-/-}$ mice. In a setting of IgM deficiency, resulting in the ineffectual neutralisation of circulating systemic microbial products, this potentially ensues in the increased expression of various recombination partners that can be involved in the reversion of the *Emv2* point mutation and tropism alteration. If specific antibodies are indeed directed against microbial products, rather than emerging infectious MLVs, then in the absence of microbial products, high eMLV expression would not be noted.

To assess this directly, *Rag1*$^{-/-}$ were firstly assessed from a separate SPF breeding colony within NIMR. Two main breeding units supply either high-use strains or low-use and mutant lines for use within the institute, and can be separated in regards the water offered breeding pairs. All *Rag1*$^{-/-}$ mice analysed thus far, bred within NIMR-B1, were offered neutral, doubly-filtered, ultra-violet (UV)-irradiated water. A separate *Rag1*$^{-/-}$ line within the second facility, NIMR-B2, was offered acidified (pH 2.5) water – a frequent husbandry choice that reduces bacterial colonisation of the gastrointestinal

tract and subsequent translocation of microbial products into the systemic circulation (Wu et al., 2006).

Figure 4.20: The impact of gut microbiota on eMLV expression in *Rag1*$^{-/-}$ mice



qRT-PCR for eMLV spliced *env* in *Rag1*$^{-/-}$ mice obtained from the indicated facilities and husbandry conditions. Each symbol represents an individual mouse. Dashed line denotes the limit of detection. $p<0.016$ between NIMR SPF mice, housed in NIMR-B1, and all other groups.

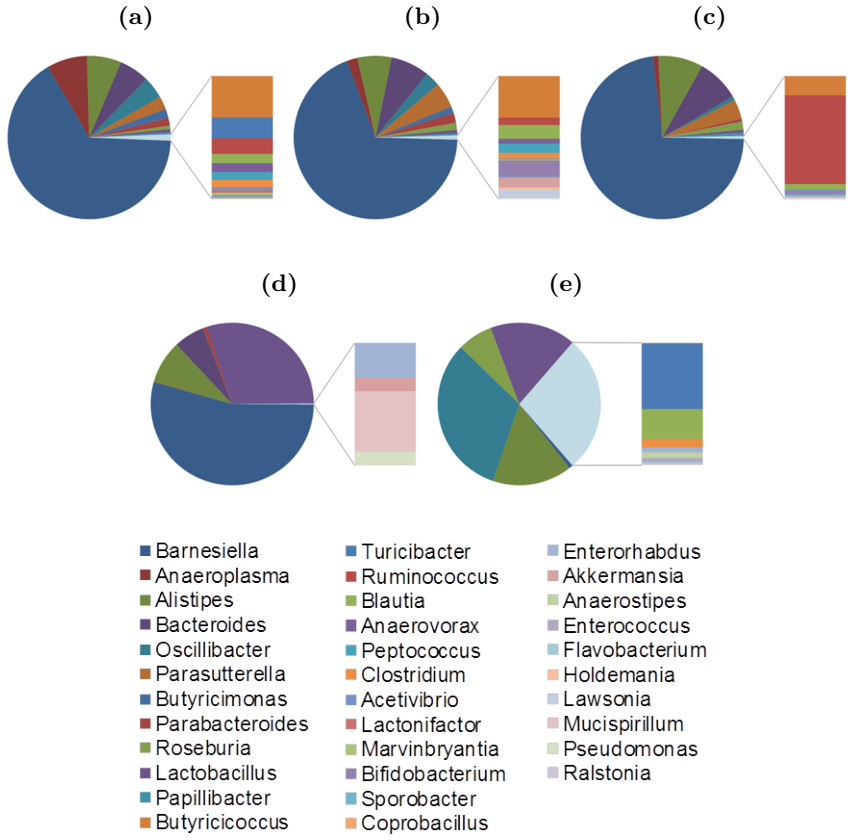
Notably, *Rag1*$^{-/-}$ mice offered acidified water had significantly lower eMLV expression than those offered neutral pH water (Fig. 4.20). This finding supported the involvement of microbial products from the gut microbiota in the generation of infectious eMLV within *Rag1*$^{-/-}$ mice. This was further confirmed by obtaining samples from the same strain from JAX, where mice were again offered acidified water (pH 2.8). Again, low eMLV expression was observed in this situation (Fig. 4.20).

To confirm that this impact was indeed related to the microbiota within these mice, rather than a secondary effect of acidified water, further *Rag1*$^{-/-}$ mice were obtained from the University of Michigan (UMICH) germ-free breeding facility and from RCHCI. UMICH GF mice are maintained, following sterile re-derivation, on autoclaved neutral water. Mice bred within RCHCI are maintained within individually ventilated cages (IVCs) from entry and, being used for various microbiota-related experiments, have an exceptionally low intestinal microbial diversity. Encouragingly, both sets of samples revealed low eMLV expression (Fig. 4.20).

Reduction in eMLV expression could have been due to lack of a specific bacterial species in *Rag1*$^{-/-}$ mice offered acidified water. To determine the likihood of this and to assess the level of change imposed on the gut microbiome by water acidification, DNA was

isolated from bacteria obtained from stool samples from the colon and the conserved bacterial 16S ribosomal RNA region was deep-sequenced. This was undertaken for *Rag1*$^{-/-}$ offered either neutral or acidified water at NIMR, as well as wild-type B6 controls offered neutral pH water, and for both *Rag1*$^{-/-}$ and B6 stool samples from RCHCI.

Figure 4.21: 16S sequencing of the gut microbiota



Comparison of colonic bacterial genera between (a) B6 and (b) *Rag1*$^{-/-}$ mice from NIMR-B1, offered neutral water, (c) *Rag1*$^{-/-}$ from NIMR-B2, offered acidified water, and (d) B6 and (e) *Rag1*$^{-/-}$ mice housed at RCHCI within IVCs and offered neutral water. Genera were assigned from 16S sequencing of stool samples. Data are averages of 4 mice per group.

Table 4.1: Bacterial diversity within the colon

	NIMR **B6** pH 7	NIMR ***Rag1*$^{-/-}$** pH 7	NIMR ***Rag1*$^{-/-}$** pH 2.5	RCHCI **B6** pH 7	RCHCI ***Rag1*$^{-/-}$** pH 7
Shannon Diversity Index (H')	1.31	1.22	0.97	1.14	1.42
Species Richness	21	22	15	11	11

All bacterial genera identified in significant numbers within B6 and *Rag1*$^{-/-}$ mice offered neutral water were present in comparable proportions, with both microbiomes being dominated by *Barnesiella* species (Figs. 4.21a and 4.21b). Equally, both species richness and diversity measures were relatively similar, although slightly reduced diversity was noted in *Rag1*$^{-/-}$ mice (Table 4.1). As would be expected, acidified water reduced the species richness and diversity of the colonic microbiome considerably (Fig. 4.21c and Table 4.1). Supporting the low eMLV expression in *Rag1*$^{-/-}$ mice obtained from RCHCI, 16S sequencing revealed exceptionally low species richness (Figs. 4.21d and 4.21e and Table 4.1).

Bacteria from the colon were chosen for 16S sequencing in this instance due to the previous identification of elevated ERV expression within this tissue. Thus, these results do not exclude greater differences in bacterial colonisation of different areas of the gastrointestinal tract, or the presence of unidentified bacterial species within the colon, but potentially support a theory that in the absence of systemic IgM, *Rag1*$^{-/-}$ mice are unable to control the translocation of universal microbial products from the gut lumen, rather than being colonised by a separate species that induces ERV expression. This suggestion would be supported by previous work in the literature that suggests the presence of elevated serum LPS, and a corresponding transcriptional signature, in B cell-deficient mice in the absence of large differences in the composition of the gut microbiota (Shulzhenko et al., 2011).

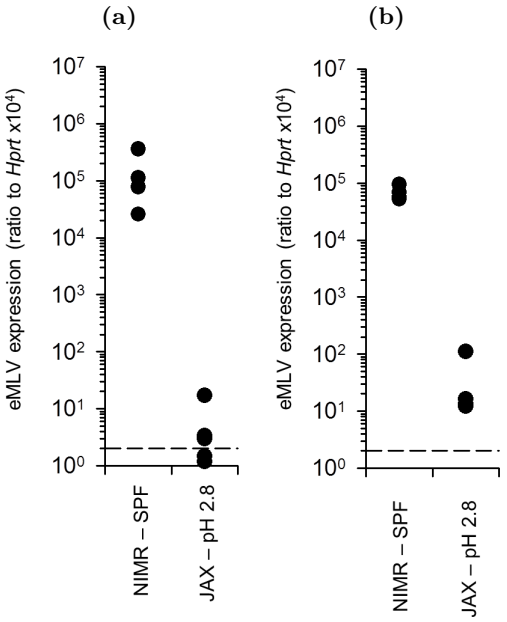
As evidenced by the differences in colonic bacterial composition of mice from NIMR and RCHCI, as well as the differences between strains maintained within RCHCI, the heritable and variable nature of the microbiome may have a large and unappreciated impact in a diverse range of settings. To confirm that the differences in eMLV expression observed in *Rag1*$^{-/-}$ were related to the microbiome, and that this could be extended to other strains previously observed to mirror their eMLV levels, further samples were requested from JAX. *Ighm*$^{-/-}$ and *Tlr7*$^{-/-}$ mice were both observed to have lower eMLV expression than the same strains offered neutral pH water and maintained at NIMR (Figs. 4.22a and 4.22b).

Thus, the processes resulting in the formation of infectious eMLVs within antibody-deficient mice not only can be seen in varying strains and breeding facilities, but can be linked to the hygiene status of the unit concerned.

4.2.8 The frequency of RARV generation

Interestingly, samples from *Myd88*$^{-/-}$ mice bred at JAX displayed a range of eMLV expression levels, suggestive of the current and continued emergence and establishment

Figure 4.22: The impact of gut microbiota on eMLV expression



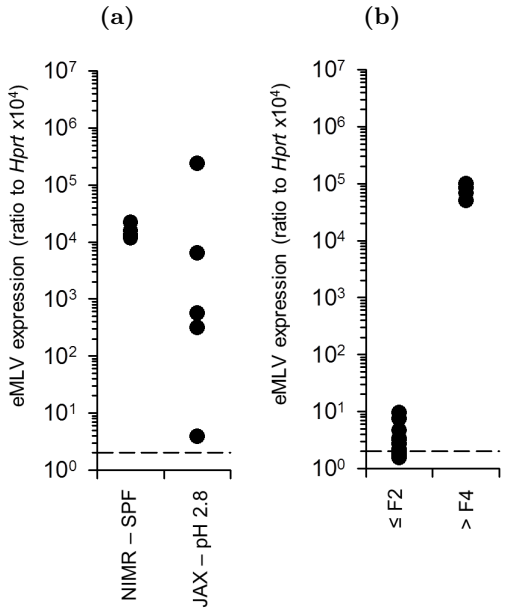
qRT-PCR for eMLV spliced *env* in spleens from (a) *Ighm*$^{-/-}$ and (b) *Tlr7*$^{-/-}$ mice housed at NIMR and offered neutral water or at JAX. p=0.005 for *Ighm*$^{-/-}$, p=0.029 for *Tlr7*$^{-/-}$ comparisons. Points represent individual mice. The dashed lines indicate the limit of detection of the assay.

of eMLVs within this colony (Fig. 4.23a). Once generated, vertical inheritance of infectious MLVs is unlikely to be impacted by acidification of the water supply and is likely to affect the entire colony over time. Thus, the presence of a neutral pH water SPF microbiota, whilst promoting RARV emergence in many cases, is not absolutely required for the eventual generation of infectious virus.

It was thus of interest to determine the probability and frequency of RARV generation over time. Given that the *Rag1*$^{-/-}$ line at NIMR has been maintained for 45 filial generations, however, and that the presence of an infectious MLV had not previously been identified in the line, a separate setting would be necessary to achieve this aim. Firstly, *Rag1*$^{-/-}$ mice bred on acidified water and having low eMLV expression were transferred to neutral water and followed for 120 days. At the end of this period, no mice were seen to have developed eMLV expression above background levels (data not shown).

The probability of RARV generation within a single mouse might be relatively low, therefore, potentially taking several generations of breeding prior to eventual emergence. Fortuitously, two re-derivations of *Ighm*$^{-/-}$ mice at NIMR have been conducted recently, allowing the study of eMLV generation over time. Progeny of mice re-derived and offered neutral pH water retained low eMLV expression for at least 2 filial gen-

Figure 4.23: The emergence of infectious eMLVs



qRT-PCR for eMLV spliced *env* in spleens from (a) *Myd88*$^{-/-}$ mice housed at NIMR and offered neutral water or at JAX and (b) *Igmh*$^{-/-}$ mice from two independently re-derived colonies at NIMR, one at F_2 on neutral pH water ($\leq F_2$) and one after F_4 on neutral pH water followed by F_{12} on acidified water ($> F_4$). Points represent individual mice. The dashed lines indicate the limit of detection of the assay.

erations (Fig. 4.23b). In contrast, a colony re-derived and maintained on neutral pH water for 4 filial generations, before being offered acidified water for a further 12 generations, until sampling, exhibited uniformly high eMLV expression, suggestive of RARV emergence (Fig. 4.23b). Enquiry of the breeding history of *Myd88*$^{-/-}$ mice maintained at JAX revealed the mice sampled had been maintained for 5 filial generations, fitting with the suggested time frame.

Thus, although the probability of RARV generation within a single mouse might be low, and potentially relies on the early establishment of a suitable microbiota, over time the cumulative probability of generation of infectious eMLV rises. Thus, in separate settings, the same or similar virological processes result in the correction of the *Emv2* point mutation, tropism conversion, and *de novo* generation of an infectious eMLV.

4.3 Discussion

These results present a novel link between antibody deficiency and the expression and 'resurrection' of ERVs. The B6 mouse, the most popular inbred strain used globally, whilst harbouring no infectious ERVs in the germ-line, contains the necessary components for the *de novo* creation and emergence of infectious MLVs. As sterile re-derivation of strains into SPF animal facilities prevents the carriage of adventitious agents, and expression can be separately seen in various antibody deficient mice obtained from a variety of sources globally, the likelihood of a contaminating eMLV is low.

A proposed model for the generation of infectious inherited ecotropic RARVs, production of polytropic variants and subsequent tumourigenesis is shown in Fig. 4.24. Although this schematic may represent the virological events leading to resurrection of a virus in the single strain from which viruses were sequenced in this study, differing recombinations are likely to occur between strains and facilities. Indeed, sequence data from collaborators at the NCI Center for Cancer Research differ greatly from sequences presented here (data not shown) and further suggest that viruses are generated in an independent manner.

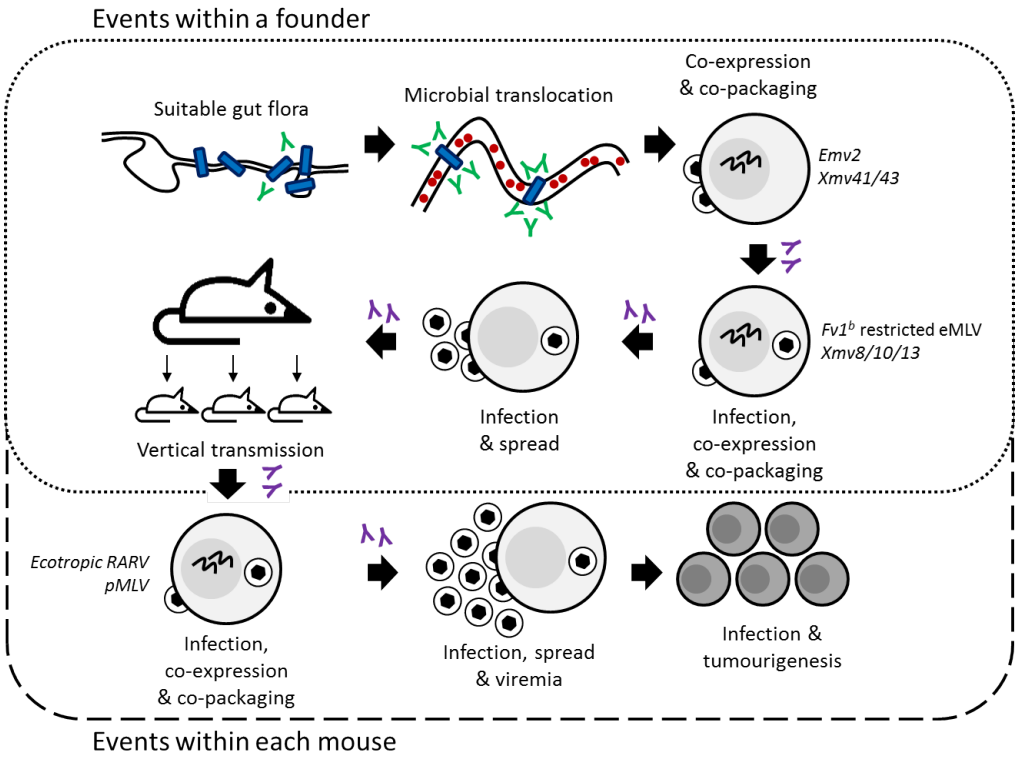
Initial sequencing work revealed a number of recombination events between *Emv2* and *Xmv41/43*, of which only one would traditionally be seen as necessary to correct the G-3576-C point mutation of *Emv2*. The virological processes involved in the correction and tropism alteration of *Emv2*, along with the potential identification of further interacting areas of the MLV genome, are thus a fascinating subject for further study. Similarly, the nature, diversity, and size of the viral swarm within each mouse is a largely uncharted area.

Whilst these data have outlined a strong requirement for systemic IgM, the precise target of these antibodies and the means by which microbial products potentially activate endogenous MLV recombination partners require further investigation. By their nature, antibody deficient mice also lack antiviral antibodies that would likely act after the generation of an infectious eMLV. In the absence of control of infection and prevention of vertical transmission, an infectious virus can establish vertical transmission in a colony and cause disease within individual mice.

Although there is a strong link between a complex gut microbiota, as influenced by varying animal husbandry conditions, and eventual emergence of infectious RARVs, this would equally need to be studied in detail to determine the exact complexity and nature of a microbiome required for this process. In likelihood, this will require the gnotobiotic colonisation of germ-free mice, followed by their continued observation over successive generations.

The characterisation of bacterial diversity from the colon did not reveal a single bac-

Figure 4.24: A model for RARV generation and disease in antibody-deficient mice



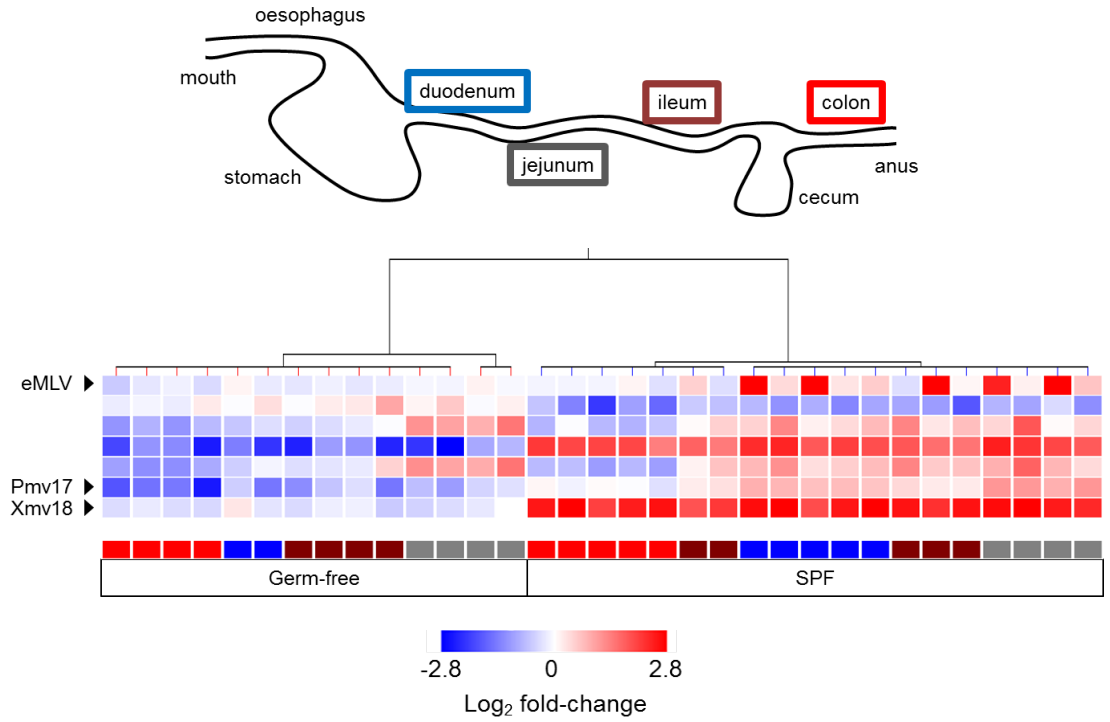
A hypothesised model of the processes surrounding RARV generation. Green antibodies denote specific IgM, involved in the control of microbial products, whilst purple antibodies represent points of potential control by anti-retroviral antibodies.

terial genus likely to be responsible for the induction of ERVs in *Rag1*$^{-/-}$ that was not present in either control B6 or *Rag1*$^{-/-}$ mice offered acidified water. However, variable assignation percentages of 16S sequencing reads and the potential for miss-assignation means that this cannot be formally excluded as a possibility. The colon was chosen for study here as it contains the largest mass of bacteria per unit of luminal content and could be shown to have high eMLV expression by qRT-PCR; however, variability in other areas of the gastrointestinal tract may have been missed.

To determine the likelihood of this possibility, a publicly-available microarray dataset studying expression profiles of varying regions of gut tissue (Larsson et al., 2012) in both germ-free and SPF B6 mice was analysed with *REquest*. Encouragingly, tissues from germ-free mice clustered separately to those from SPF conditions and had generally lower expression values (Fig. 4.25). In SPF conditions, both *Xmv18* and *Pmv17* could be uniquely identified as being upregulated, with strongest expression being seen in the duodenum and jejunum (Fig. 4.25). Of the four tissues studied, the colon had the lowest expression of all REs and ERVs in both germ-free and SPF mice (Fig. 4.25). Thus, there remains a strong case to further investigate the effect of microbiota throughout the

gastrointestinal tract on RE and ERV expression, including the potential interactions with endogenous MLVs.

Figure 4.25: ERV expression throughout the gastrointestinal tract



Heatmap of probesets found to correspond to REs and ERVs with *REquest* from the Affymetrix Mouse Gene 1.0 ST microarray. Conditions are ordered by unsupervised clustering. Data were downloaded from a public database (E-GEOD-17438). Shown are probesets significantly upregulated ($p<0.05$) by more than 2-fold in at least one condition. Probesets corresponding to individually-identifiable endogenous MLVs are highlighted.

At present, only B6 background mice have been studied for the emergence of infectious RARVs. Both *Emv1* and *Emv3* are present in many strains in frequent laboratory use, however (Table 1.2), and the effects of antibody deficiency and the influence of the microbiota have not been investigated in these situations.

Implications of this work

The existence of infectious eMLVs in antibody-deficient B6 background mouse strains is a potentially large confounding factor in both previous and further immunological and virological studies. Historically, for example, much Friend virus literature has been affected by the contamination of viral stocks with lactate dehydrogenase-elevating virus (LDV). FV/LDV co-infection drastically alters the progression and outcome of FV infection (Marques et al., 2008; Miyazawa, Tsuji-Kawahara & Kanari, 2008; Pike

et al., 2009) and significantly impacts the immune response. The potential existence of RARVs in mice used for studies of retroviral immunology, and the unknown impact of co-infection, now needs to be taken into account and investigated in these settings.

Further, the inheritance of infectious viruses neonatally complicates the breeding of knock-out and other mutant mouse lines. Lines generated with use of males carrying the desired genotype may avoid the vertical transmission of viruses, whilst those generated in the complementary manner would likely transmit viruses to progeny strains. Increasingly, littermate controls are seen as preferable to the use of separate litters of control mice in experimental models. Even here, however, the existence of a vertically inherited infectious MLV complicates interpretation. For example, where a mouse heterozygous for a gene associated with antibody-deficiency (e.g. a *Rag1*$^{+/-}$ mouse) may be used to represent 'wild-type' in a litter otherwise homozygous knock-out (e.g. *Rag1*$^{-/-}$), these would possess the necessary antiviral antibodies to likely control infection resulting from inherited virus and prevent disease, where littermates would succumb to infection and resulting pathogenesis.

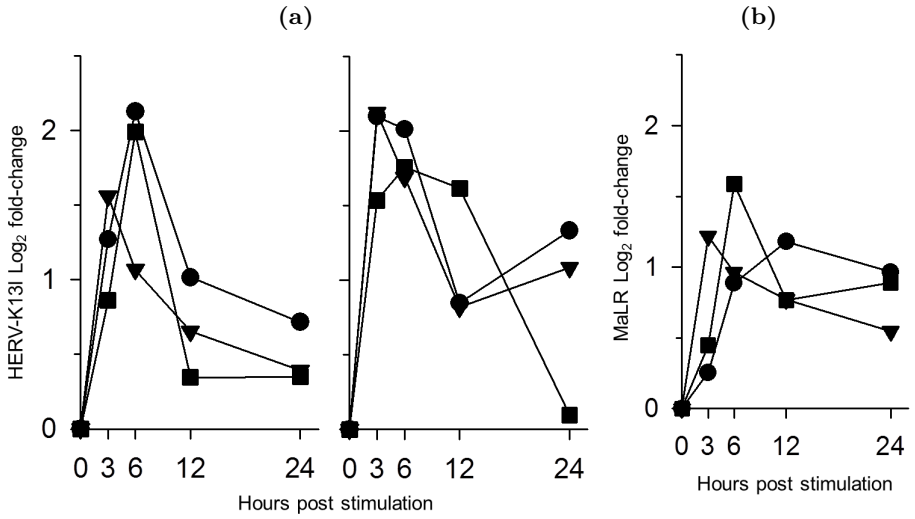
The potential influence of the gut microbiome further complicates comparisons of data obtained under varying husbandry conditions and between research centres. Even within a single institute, mutant mouse lines with specific husbandry requirements or reduced fecundity are frequently offered acidified water where standard stocks may be treated in a different manner. This further complicates comparisons of data obtained through the use of mutant mouse lines and, indeed, much early work involved in this project had to be reinterpreted or repeated on discovery of the implications of acidified water treatment.

Whilst all known human ERVs are non-functional, a proof-of-principle study was able to construct an infectious virus from a consensus of HERV-K family sequences (Y. N. Lee & Bieniasz, 2007). Thus, although well-studied in murine models, an infectious and oncogenic potential of human ERVs has not been observed. Nevertheless, other REs have been shown to be active in humans and have been associated with disease in some cases (Iskow et al., 2010; E. Lee et al., 2012).

To investigate briefly if human ERVs and REs might be controlled in a similar manner to those of mice, a human microarray dataset characterising LPS and R848 (a TLR 7/8 agonist) stimulation of monocyte-derived dendritic cells from 3 healthy donors (Cavalieri et al., 2010) was studied with *REquest*. Responses differed between cells derived from each donor, but encouragingly, significant changes in expression could be noted for probes corresponding to ERV-9-like elements (data not shown), HERV-K13-like elements (Fig. 4.26a), and mammalian apparent long terminal repeat retrotransposon (MaLR) elements (Fig. 4.26b).

These exploratory data open a large area of further study and may have great potential

Figure 4.26: ERV expression in human cells stimulated with TLR agonists



Transcriptional profiling of human monocyte-derived dendritic cells from 3 healthy donors in a publicly-available microarray dataset (EBI Array Express EMTAB-448). Probes corresponding to REs and ERVs within the Illumina HT-12 expression microarray were determined with *REquest* and those significantly upregulated (p<0.05) more than 2-fold determined. (a) LPS (left) and R848 (right) stimulation causing increased expression of HERV-K13l. (b) R848 stimulation causing increased MaLR expression. Each line represents an individual donor.

relevance in human disease and treatment options. Paralleling the results seen in mice, increased risk of lymphoma has previously been associated with antibody-deficiency in humans (Park, Li, Hagan, Maddox & Abraham, 2008), as well as infection and inflammation (Trinchieri, 2012). A potential link between REs and ERVs and microbial products is thus an intriguing research area and could link to factors such as antibiotic use or conditions of excessive hygiene. Increased adiposity, for example, has been correlated with antibiotic exposure in both humans and mice (Cho et al., 2012; Trasande et al., 2012), yet changes such as these are unlikely to be found in isolation.

Chapter 5

Conclusion

Repetitive sequences, including REs and ERVs, comprise significant proportions of mammalian genomes. These elements have accumulated throughout evolution and whilst the majority are now unable to replicate, comparatively little research has centred around their impacts on the host.

The work detailed here has focussed on the interaction of ERVs with the immune system. Firstly, the single endogenous eMLV of B6 mice, *Emv2*, was seen to have a significant affect on the development of thymocytes with potential reactivity in FV infection. In spite of the large-scale deletion mediated by *Emv2*, however, this had the effect of increasing overall proportions of high-avidity $CD4^+$ T cells and is partly responsible for the strong response to FV in this strain.

Secondly, the means by which the immune system impacts ERV expression was determined. A requirement for specific antibodies could be identified in the control of emergence of infectious virus within mouse colonies. Interestingly, the phenomenon of 'resurrection' of defective ERVs could be shown in multiple antibody-deficient stains and across a further five animal facilities, impacting many commonly-used strains and having significant implications for mouse experimentation. Whilst not examined exhaustively, the data presented further suggest an intriguing link to microbial stimulation in the *de novo* emergence of infectious virus within these strains.

The murine model represents a versatile tool in biological research and the work detailed in this thesis has exploited various congenic, transgenic and knock-out strains generated over many years. Whilst there is significant variability in experimental systems defined in mice or humans, there are also significant similarities (Su et al., 2002; Zheng-Bradley, Rung, Parkinson & Brazma, 2010). In this way, whilst the compositions of human and murine REs differ considerably, the overall proportions of these elements within the two genomes is relatively similar. Thus whilst a considerable effort will be required to translate these findings and to determine their possible relevance in human disease, this is an exciting research opportunity.

References

Adachi, O., Kawai, T., Takeda, K., Matsumoto, M., Tsutsui, H., Sakagami, M., ... Akira, S. (1998). Targeted Disruption of the MyD88 Gene Results in Loss of IL-1- and IL-18-Mediated Function. *Immunity*, *9*, 143–150.

Agrawal, A., Eastman, Q. M. & Schatz, D. G. (1998). Implication of transpostion mediated by V(D)J-recombination proteins RAG1 and RAG2 for origins of of antigen-specific immunity. *Nature*, *394*, 744–751.

Ahn, K. & Kim, H.-S. (2009). Structural and quantitative expression analyses of HERV gene family in human tissues. *Molecules and cells*, *28*, 99–103.

Alam, M., Midtvedt, T. & Uribe, A. (1994). Differential Cell Kinetics in the Ileum and Colon of Germfree Rats. *Scandinavian journal of gastroenterology*, *29*, 445–451.

Albritton, L. M., Tseng, L., Scadden, D. & Cunningham, J. M. (1989). A Putative Murine Ecotropic Retrovirus Receptor Gene Encodes a Multiple Membrane-Spanning Protein and Confers Susceptibility to Virus Infection. *Cell*, *57*, 659–666.

Ambrosi, A., Cattoglio, C. & Di Serio, C. (2008). Retroviral integration process in the human genome: is it really non-random? A new statistical approach. *Plos computational biology*, *4*(8), e1000144.

Amit, I., Garber, M., Chevrier, N., Leite, A. P., Donner, Y., Eisenhaure, T., ... Regev, A. (2009). Unbiased reconstruction of a mammalian transcriptional network mediating pathogen responses. *Science*, *326*, 257–263.

Amsen, D., Blander, J. M., Lee, G. R., Tanigaki, K., Honjo, T. & Flavell, R. A. (2004). Instruction of distinct CD4 T helper cell fates by different notch ligands on antigen-presenting cells. *Cell*, *117*, 515–526.

Anderson, M. S., Venanzi, E. S., Klein, L., Chen, Z., Berzins, S. P., Turley, S. J., ... Mathis, D. (2002). Projection of an immunological self shadow within the thymus by the Aire protein. *Science*, *298*, 1395–1401.

Antunes, I., Tolaini, M., Kissenpfennig, A., Iwashiro, M., Kuribayashi, K., Malissen, B., ... Kassiotis, G. (2008). Retrovirus-specificity of regulatory T cells is neither present nor required in preventing retrovirus-induced bone marrow immune pathology. *Immunity*, *29*(5), 782–794.

Appay, V., Zaunders, J. J., Papagno, L., Sutton, J., Jaramillo, A., Waters, A., ... Kelleher, A. D. (2002). Characterization of CD4+ CTLs Ex Vivo. *Journal of immunology*, *168*, 5954–5958.

Arjan-Odedra, S., Swanson, C. M., Sherer, N. M., Wolinsky, S. M. & Malim, M. H. (2012). Endogenous MOV10 inhibits the retrotransposition of endogenous retroelements but not the replication of exogenous retroviruses. *Retrovirology*, *9*, 53.

Armitage, A. E., Katzourakis, A., de Oliveira, T., Welch, J. J., Belshaw, R., Bishop, K. N., ... Iversen, A. K. N. (2008). Conserved footprints of APOBEC3G on Hypermutated human immunodeficiency virus type 1 and human endogenous retrovirus HERV-K(HML2) sequences. *Journal of virology*, *82*(17), 8743–8761.

Arumugam, M., Raes, J., Pelletier, E., Le Paslier, D., Yamada, T., Mende, D. R., ... Bork, P. (2011). Enterotypes of the human gut microbiome. *Nature*, *473*, 174–180.

Atarashi, K., Tanoue, T., Shima, T., Imaoka, A., Kuwahara, T., Momose, Y., ... Honda, K. (2011). Induction of colonic regulatory T cells by indigenous Clostridium species. *Science*, *331*, 337–341.

Baba, K., Nakaya, Y., Shojima, T., Muroi, Y., Kizaki, K., Hashizume, K., ... Miyazawa, T. (2011). Identification of novel endogenous betaretroviruses which are transcribed in the bovine placenta. *Journal of virology*, *85*(3), 1237–1245.

Baillie, J. K., Barnett, M. W., Upton, K. R., Gerhardt, D. J., Richmond, T. A., De Sapio, F., ... Faulkner, G. J. (2011). Somatic retrotransposition alters the genetic landscape of the human brain. *Nature*, *479*, 534–537.

Baltimore, D. (1970). Viral RNA-dependent DNA Polymerase. *Nature*, *226*, 1209–1211.

Barbosa-Morais, N. L., Dunning, M. J., Samarajiwa, S. A., Darot, J. F. J., Ritchie, M. E., Lynch, A. G. & Tavaré, S. (2010). A re-annotation pipeline for Illumina BeadArrays : improving the interpretation of gene expression data. *Nucleic acids research*, *38*(3), e17.

Barreiro, L. B., Ben-Ali, M., Quach, H., Laval, G., Patin, E., Pickrell, J. K., ... Quintana-Murci, L. (2009). Evolutionary Dynamics of Human Toll-Like Receptors and Their Different Contributions to Host Defense. *Plos genetics*, *5*(7), e1000562.

Baudino, L., Yoshinobu, K., Dunand-Sauthier, I., Evans, L. H. & Izui, S. (2010). TLR-mediated up-regulation of serum retroviral gp70 is controlled by the Sgp loci of lupus-prone mice. *Journal of autoimmunity*, *35*, 153–159.

Beck, C. R., Collier, P., Macfarlane, C., Malig, M., Kidd, J. M., Eichler, E. E., ... Moran, J. V. (2010). LINE-1 retrotransposition activity in human genomes. *Cell*, *141*, 1159–1170.

Belshaw, R., Dawson, A. L. A., Woolven-Allen, J., Redding, J., Burt, A. & Tristem, M. (2005). Genomewide Screening Reveals High Levels of Insertional Polymorphism in the Human Endogenous Retrovirus Family HERV-K (HML2): Implications for Present-Day Activity. *Journal of virology*, *79*(19), 12507–12514.

Belshaw, R., Pereira, V., Katzourakis, A., Talbot, G., Paces, J., Burt, A. & Tristem, M. (2004). Long-term reinfection of the human genome by endogenous retroviruses. *Proceedings of the national academy of sciences of the united states of america*, *101*(14), 4894–4899.

Belyi, V. A., Levine, A. J. & Skalka, A. M. (2010). Sequences from ancestral single-stranded DNA viruses in vertebrate genomes: the parvoviridae and circoviridae are more than 40 to 50 million years old. *Journal of virology*, *84*(23), 12458–12462.

Bennett, E. A., Keller, H., Mills, R. E., Schmidt, S., Moran, J. V., Weichenrieder, O. & Devine, S. E. (2008). Active Alu retrotransposons in the human genome. *Genome research*, *18*, 1875–1883.

Bernhard, W. (1958). Electron Microscopy of Tumor Cells and Tumor Viruses: A Review. *Cancer research*, *18*, 491–509.

Best, S., Le Tissier, P., Towers, G. & Stoye, J. P. (1996). Positional cloning of the mouse retrovirus restriction gene Fv1. *Nature*, *382*, 826–829.

Blaise, S., de Parseval, N., Bénit, L. & Heidmann, T. (2003). Genomewide screening for fusogenic human endogenous retrovirus envelopes identifies syncytin 2, a gene conserved on primate evolution. *Proceedings of the national academy of sciences of the united states of america*, *100*(22), 13013–13018.

Bock, M., Bishop, K. N., Towers, G. & Stoye, J. P. (2000). Use of a Transient Assay for Studying the Genetic Determinants of Fv1 Restriction. *Journal of virology*, *74*(16), 7422–7430.

Bonnet, F., Lewden, C., May, T., Heripret, L., Jougla, E., Bevilacqua, S., ... The Mortalité 2000 Study Group. (2005). Opportunistic infections as causes of death in HIV-infected patients in the HAART era in France. *Scandinavian journal of infectious diseases*, *37*, 482–487.

Bourc'his, D. & Bestor, T. H. (2004). Meiotic catastrophe and retrotransposon reactivation in male germ cells lacking Dnmt3L. *Nature*, *431*, 96–99.

Bourgeois, C., Hao, Z., Rajewsky, K., Potocnik, A. J. & Stockinger, B. (2008). Ablation of thymic export causes accelerated decay of naïve CD4 T cells in the periphery because of activation by environmental antigen. *Proceedings of the national academy of sciences of the united states of america*, *105*(25), 8691–8696.

Bousso, P., Casrouge, A., Altman, J. D., Haury, M., Kanellopoulos, J., Abastado, J.-P. & Kourilsky, P. (1998). Individual Variations in the Murine T Cell Response to a Specific Peptide Reflect Variability in Naive Repertoires. *Immunity*, *9*, 169–178.

Brenchley, J. M., Price, D. a., Schacker, T. W., Asher, T. E., Silvestri, G., Rao, S., ... Douek, D. C. (2006). Microbial translocation is a cause of systemic immune activation in chronic HIV infection. *Nature medicine*, *12*(12), 1365–1371.

Bridgeman, J. S., Sewell, A. K., Miles, J. J., Price, D. A. & Cole, D. K. (2011). Structural and biophysical determinants of $\alpha\beta$ T-cell antigen recognition. *Immunology*, *135*, 9–18.

Browne, E. P. (2011). Toll-like receptor 7 controls the anti-retroviral germinal center response. *Plos pathogens*, *7*(10), e1002293.

Browne, E. P. & Littman, D. R. (2009). Myd88 is required for an antibody response to retroviral infection. *Plos pathogens*, *5*(2), e1000298.

Bruton, O. C. (1952). Agammaglobulinemia. *Pediatrics*, *9*, 722–728.

Buffett, R. F., Grace Jr, J. T., DiBerardino, L. A. & Mirand, E. A. (1969). Vertical Transmission of Murine Leukemia Virus. *Cancer research*, *29*, 588–595.

Buzdin, A., Kovalskaya-Alexandrova, E., Gogvadze, E. & Sverdlov, E. (2006). At least 50% of human-specific HERV-K (HML-2) long terminal repeats serve in vivo as

active promoters for host nonrepetitive DNA transcription. *Journal of virology*, *80*(21), 10752–10762.

Carson, R. T., Vignali, K. M., Woodland, D. L. & Vignali, D. A. A. (1997). T Cell Receptor Recognition of MHC Class II – Bound Peptide Flanking Residues Enhances Immunogenicity and Results in Altered TCR V Region Usage. *Immunity*, *7*, 387–399.

Casanova, J.-L., Romero, P., Widmann, C., Kourilsky, P. & Maryanski, J. L. (1991). T cell receptor genes in a series of class I major histocompatibility complex-restricted cytotoxic T lymphocyte clones specific for a Plasmodium berghei nonapeptide: implications for T cell allelic exclusion and antigen-specific repertoire. *Journal of experimental medicine*, *174*, 1371–1383.

Cavalieri, D., Rivero, D., Beltrame, L., Buschow, S. I., Calura, E., Rizzetto, L., ... Austyn, J. M. (2010). DC-ATLAS: a systems biology resource to dissect receptor specific signal transduction in dendritic cells. *Immunome research*, *6*, 10.

Changolkar, L. N., Singh, G. & Pehrson, J. R. (2008). macroH2A1-Dependent Silencing of Endogenous Murine Leukemia Viruses. *Molecular and cellular biology*, *28*(6), 2059–2065.

Chattopadhyay, S. K., Sengupta, D. N., Fredrickson, T. N., Morse III, H. C. & Hartley, J. W. (1991). Characteristics and Contributions of Defective, Ecotropic, and Mink Cell Focus-Inducing Viruses Involved in a Retrovirus-Induced Immunodeficiency Syndrome of Mice. *Journal of virology*, *65*(8), 4232–4241.

Chen, J., Trounstine, M., Alt, F. W., Young, F., Kurahara, C., Loring, J. F. & Huszar, D. (1993). Immunoglobulin gene rearrangement in B cell deficient mice generated by targeted deletion of the JH locus. *International immunology*, *5*(6), 647–656.

Cheng, J., Kapranov, P., Drenkow, J., Dike, S., Brubaker, S., Patel, S., ... Gingeras, T. R. (2005). Transcriptional maps of 10 human chromosomes at 5-nucleotide resolution. *Science*, *308*, 1149–1154.

Cherwinski, H. M., Schumacher, J. H., Brown, K. D. & Mosmann, T. R. (1987). Two types of mouse helper T cell clone III. Further differences in lymphokine synthesis between Th1 and Th2 clones revealed by RNA hybridization, functionally monospecific bioassays, and monoclonal antibodies. *Journal of experimental medicine*, *166*, 1229–1244.

Chesebro, B., Bloom, M., Wehrly, K. & Nishio, J. (1979). Persistence of infectious Friend virus in spleens of mice after spontaneous recovery from virus-induced erythroleukemia. *Journal of virology*, *32*(3), 832–837.

Chesebro, B., Britt, W., Evans, L. H., Wehrly, K., Nishio, J. & Cloyd, M. (1983, May). Characterization of monoclonal antibodies reactive with murine leukemia viruses: use in analysis of strains of friend MCF and Friend ecotropic murine leukemia virus. *Virology*, *127*, 134–148.

Cho, I., Yamanishi, S., Cox, L., Methé, B. A., Zavadil, J., Li, K., ... Blaser, M. J. (2012). Antibiotics in early life alter the murine colonic microbiome and adiposity. *Nature*, *488*, 621–626.

Clark, S. P. & Mak, T. W. (1984). Fluidity of a Retrovirus Genome. *Journal of virology*, *50*(3), 759–765.

Cole, J. R., Wang, Q., Cardenas, E., Fish, J., Chai, B., Farris, R. J., ... Tiedje, J. M. (2009). The Ribosomal Database Project: improved alignments and new tools for rRNA analysis. *Nucleic acids research*, *37*, D141–D145.

Conley, A. B., Piriyapongsa, J. & Jordan, I. K. (2008). Retroviral promoters in the human genome. *Bioinformatics*, *24*(14), 1563–1567.

Conteduca, G., Ferrera, F., Pastorino, L., Fenoglio, D., Negrini, S., Sormani, M. P., ... Filaci, G. (2010). The role of AIRE polymorphisms in melanoma. *Clinical immunology*, *136*(1), 96–104.

Contreras-Galindo, R., Kaplan, M. H., Contreras-Galindo, A. C., Gonzalez-Hernandez, M. J., Ferlenghi, I., Giusti, F., ... Markovitz, D. M. (2012). Characterization of human endogenous retroviral elements in the blood of HIV-1-infected individuals. *Journal of virology*, *86*(1), 262–276.

Contreras-Galindo, R., Kaplan, M. H., Leissner, P., Verjat, T., Ferlenghi, I., Bagnoli, F., ... Markovitz, D. M. (2008). Human endogenous retrovirus K (HML-2) elements in the plasma of people with lymphoma and breast cancer. *Journal of virology*, *82*(19), 9329–9336.

Contreras-Galindo, R., López, P., Vélez, R. & Yamamura, Y. (2007). HIV-1 infection increases the expression of human endogenous retroviruses type K (HERV-K) in vitro. *Aids research and human retroviruses*, *23*(1), 116–122.

Coppola, M. A., Lam, T. M., Strawbridge, R. R. & Green, W. R. (1995). Recognition of endogenous ecotropic murine leukaemia viruses by anti-AKR/Gross virus cytotoxic T lymphocytes (CTL): epitope variation in a CTL-resistant virus. *Journal of general virology*, *76*, 635–641.

Cornelis, G., Heidmann, O., Bernard-Stoecklin, S., Reynaud, K., Véron, G., Mulot, B., ... Heidmann, T. (2012). Ancestral capture of syncytin-Car1 , a fusogenic endogenous retroviral envelope gene involved in placentation and conserved in Carnivora. *Proceedings of the national academy of sciences of the united states of america*, *109*(7), E432–E441.

Cosgrove, D., Gray, D., Dierich, A., Kaufman, J., Lemeur, M., Benoist, C. & Mathis, D. (1991, September). Mice lacking MHC class II molecules. *Cell*, *66*, 1051–1066.

Crawford, L. V. & Crawford, E. M. (1961). The Properties of Rous Sarcoma Virus Purifed by Density Gradient Centrifugation. *Virology*, *13*, 227–232.

Crow, Y. J., Hayward, B. E., Parmar, R., Robins, P., Leitch, A., Ali, M., ... Lindahl, T. (2006). Mutations in the gene encoding the 3'-5' DNA exonuclease TREX1 cause Aicardi-Goutières syndrome at the AGS1 locus. *Nature genetics*, *38*(8), 917–920.

Cunningham, D. A., Dos Santos Cruz, G. J., Fernández-Suárez, X. M., Whittam, A. J., Herring, C., Copeman, L., ... Langford, G. A. (2004). Activation of primary porcine endothelial cells induces release of porcine endogenous retroviruses. *Transplantation*, *77*(7), 1071–1079.

De Palma, M., Montini, E., Santoni de Sio, F. R., Benedicenti, F., Gentile, A., Medico, E. & Naldini, L. (2005). Promoter trapping reveals significant differences in integration site selection between MLV and HIV vectors in primary hematopoietic cells. *Blood*, *105*, 2307–2315.

de Harven, E. & Friend, C. (1960). Further Electron Microscope Studies of a Mouse Leukemia Induced by Cell-Free Filtrates. *Journal of biophysical and biochemical cytology*, *7*(4), 747–752.

Déjardin, J., Bompard-Maréchal, G., Audit, M., Hope, T. J., Sitbon, M. & Mougel, M. (2000). A novel subgenomic murine leukemia virus RNA transcript results from alternative splicing. *Journal of virology*, *74*(8), 3709–3714.

Demaria, O., Pagni, P. P., Traub, S., de Gassart, A., Branzk, N., Murphy, A. J., ... Alexopoulou, L. (2010). TLR8 deficiency leads to autoimmunity in mice. *Journal of clinical investigation*, *120*(10), 3651–3662.

Derbinski, J., Gäbler, J., Brors, B., Tierling, S., Jonnakuty, S., Hergenhahn, M., ... Kyewski, B. (2005). Promiscuous gene expression in thymic epithelial cells is regulated at multiple levels. *Journal of experimental medicine*, *202*(1), 33–45.

DeRisi, J., Penland, L., Brown, P. O., Bittner, M. L., Meltzer, P. S., Ray, M., ... Trent, J. M. (1996). Use of a cDNA microarray to analyse gene expression patterns in human cancer. *Nature genetics*, *14*, 457–460.

Di Nicuolo, G., D'Alessandro, A., Andria, B., Scuderi, V., Scognamiglio, M., Tammaro, A., ... Chamuleau, R. A. F. M. (2010). Long-term absence of porcine endogenous retrovirus infection in chronically immunosuppressed patients after treatment with the porcine cell-based Academic Mecal Center bioartificial liver. *Xenotransplantation*, *17*, 431–439.

Donahue, R. E., Kessler, S. W., Bodine, D., McDonagh, K., Dunbar, C., Goodman, S., ... Nienhuis, A. W. (1992). Helper virus induced T cell lymphoma in nonhuman primates after retroviral mediated gene transfer. *Journal of experimental medicine*, *176*, 1125–1135.

Doolan, D. L., Southwood, S., Freilich, D. A., Sidney, J., Graber, N. L., Shatney, L., ... Sette, A. (2003). Identification of Plasmodium falciparum antigens by antigenic analysis of genomic and proteomic data. *Proceedings of the national academy of sciences of the united states of america*, *100*(17), 9952–9957.

Duhl, D. M. J., Vrieling, H., Miller, K. A., Wolff, G. L. & Barsh, G. S. (1994). Neomorphic agouti mutations in obese yellow mice. *Nature genetics*, *8*, 59–65.

Dunham, I., Shimizu, N., Roe, B. A. & Chissoe, S. (1999). The DNA sequence of human chromosome 22. *Nature*, *402*, 489–496.

Dunlap, K. a., Palmarini, M., Varela, M., Burghardt, R. C., Hayashi, K., Farmer, J. L. & Spencer, T. E. (2006). Endogenous retroviruses regulate periimplantation placental growth and differentiation. *Proceedings of the national academy of sciences of the united states of america*, *103*(39), 14390–14395.

Dupressoir, A., Marceau, G., Vernochet, C., Bénit, L., Kanellopoulos, C., Sapin, V. & Heidmann, T. (2005). Syncytin-A and syncytin-B, two fusogenic placenta-specific murine envelope genes of retroviral origin conserved in Muridae. *Proceedings of the national academy of sciences of the united states of america*, *102*(3), 725–730.

Dupressoir, A., Vernochet, C., Bawa, O., Harper, F., Pierron, G., Opolon, P. & Heidmann, T. (2009). Syncytin-A knockout mice demonstrate the critical role in placentation of a fusogenic, endogenous retrovirus-derived, envelope gene. *Proceedings of the national academy of sciences of the united states of america*, *106*(29), 12127–12132.

Ebert, P. J. R., Jiang, S., Xie, J., Li, Q.-J. & Davis, M. M. (2009). An endogenous positively selecting peptide enhances mature T cell responses and becomes an autoantigen in the absence of microRNA miR-181a. *Nature immunology*, *10*(11), 1162–1169.

Elleder, D., Kim, O., Padhi, A., Bankert, J. G., Simeonov, I., Schuster, S. C., ... Poss, M. (2012). Polymorphic integrations of an endogenous gammaretrovirus in the mule deer genome. *Journal of virology*, *86*(5), 2787–2796.

Ellermann, V. & Bang, O. (1908). Experimentelle Leukämie bei Hühnern. *Zentralblatt fur bakteriologie*, *46*, 595–609.

Evans, L. H., Alamgir, A. S. M., Owens, N., Weber, N., Virtaneva, K., Barbian, K., ... Rosenke, K. (2009). Mobilization of endogenous retroviruses in mice after infection with an exogenous retrovirus. *Journal of virology*, *83*(6), 2429–2435.

Evans, L. H., Morrison, R. P., Malik, F. G., Portis, J. & Britt, W. J. (1990). A neutralizable epitope common to the envelope glycoproteins of ecotropic, polytropic, xenotropic, and amphotropic murine leukemia viruses. *Journal of virology*, *64*(12), 6176–6183.

Fagarasan, S., Kawamoto, S., Kanagawa, O. & Suzuki, K. (2010). Adaptive immune regulation in the gut: T cell-dependent and T cell-independent IgA synthesis. *Annual review of immunology*, *28*, 243–273.

Fauci, A. S. (1993). Multifactorial nature of Human Immunodeficieincy Virus disease: implications for therapy. *Science*, *262*, 1011–1018.

Fernandez-Botran, R., Sanders, V. M., Mosmann, T. R. & Vitetta, E. S. (1988). Lymphokine-mediated regulation of the proliferative response of clones of T helper 1 and T helper 2 cells. *Journal of experimental medicine*, *168*, 543–558.

Fodor, S. P. A., Read, J. L., Pirrung, M. C., Stryer, L., Lu, A. T. & Solas, D. (1991). Light-Directed, Spatially Addressable Parallel Chemical Synthesis. *Science*, *251*, 767–773.

Frank, O., Verbeke, C., Schwarz, N., Mayer, J., Fabarius, A., Hehlmann, R., ... Seifarth, W. (2008). Variable transcriptional activity of endogenous retroviruses in human breast cancer. *Journal of virology*, *82*(4), 1808–1818.

Frankel, W. N., Stoye, J. P., Taylor, B. A. & Coffin, J. M. (1990). A linkage map of endogenous murine leukemia proviruses. *Genetics*, *124*, 221–36.

Friend, C. (1957). Cell-free transmission in adult swiss mice of a disease having the character of a leukemia. *Journal of experimental medicine*, *105*(4), 307–318.

Friend, C. (1959). Immunological relationships of a filterable agent causing a leukemia in adult mice. I. The neutralization of infectivity by specific antiserum. *Journal of experimental medicine*, *109*(2), 217–228.

Ganz, T. (2003). Defensins: antimicrobial peptides of innate immunity. *Nature reviews immunology*, *3*, 710–720.

Gardner, M. B., Chiri, A., Dougherty, M. F., Casagrande, J. & Estes, J. D. (1979). Congenital Transmission of Murine Leukemia Virus From Wild Mice Prone to the Development of Lymphoma and Paralysis. *Journal of the national cancer institute*, *62*(1), 63–70.

Geuking, M. B., Cahenzli, J., Lawson, M. A. E., Ng, D. C. K., Slack, E., Hapfelmeier, S., ... Macpherson, A. J. (2011). Intestinal bacterial colonization induces mutualistic regulatory T cell responses. *Immunity*, *34*, 794–806.

Goodchild, N. L., Wilkinson, D. A. & Mager, D. L. (1993). Recent Evolutionary Expansion of a Subfamily of RTVL-H Human Endogenous Retrovirus-like Elements. *Virology*, *196*, 778–788.

Goodnow, C. C., Crosbie, J., Adelstein, S., Lavoie, T. B., Smith-Gill, S. J., Brink, R. A., ... Basten, A. (1988). Altered immunoglobulin expression and functional silencing of self-reactive B lymphocytes in transgenic mice. *Nature*, *334*, 676–682.

Gordon, J. I., Hooper, L. V., McNevin, M. S., Wong, M. & Bry, L. (1997). Epithelial cell growth and differentiation. III. Promoting diversity in the intestine: conversations between the microflora, epithelium and diffuse GALT. *Americal journal of physiology gastrointestinal and liver physiology*, *273*, G565–G570.

Gras, S., Chen, Z., Miles, J. J., Liu, Y. C., Bell, M. J., Sullivan, L. C., ... Burrows, S. R. (2010). Allelic polymorphism in the T cell receptor and its impact on immune responses. *Journal of experimental medicine*, *207*(7), 1555–1567.

Greenberger, J. S., Phillips, S. M., Stephenson, J. R. & Aaronson, S. A. (1975). Induction of mouse type-C RNA virus by lipopolysaccharide. *Journal of immunology*, *115*(1), 317–320.

Gross, H., Barth, S., Pfuhl, T., Willnecker, V., Spurk, A., Gurtsevitch, V., ... Grässer, F. A. (2011). The NP9 protein encoded by the human endogenous retrovirus HERV-K(HML-2) negatively regulates gene activation of the Epstein-Barr virus nuclear antigen 2 (EBNA2). *International journal of cancer*, *129*(5), 1105–1115.

Gross, L. (1951). "Spontaneous" leukemia developing in C3H mice following inoculation in infancy, with AK-leukemic extracts, or AK-embryos. *Proceedings of the society for experimental biology and medicine*, *76*(1), 27–32.

Gross, L. (1957). Development and serial cell free passage of a highly potent strain of mouse leukemia virus. *Proceedings of the society for experimental biology and medicine*, *94*(4), 767–771.

Guallar, D., Pérez-Palacios, R., Climent, M., Martínez-Abadía, I., Larraga, A., Fernández-Juan, M., ... Schoorlemmer, J. (2012). Expression of endogenous retroviruses is negatively regulated by the pluripotency marker Rex1/Zfp42. *Nucleic acids research*, *40*(18), 8993–9007.

Guarner, F. & Malagelada, J.-R. (2003). Gut flora in health and disease. *The lancet*, *360*, 512–519.

Harper, A. L., Sudol, M. & Katzman, M. (2003). An Amino Acid in the Central Catalytic Domain of Three Retroviral Integrases That Affects Target Site Selection in Nonviral DNA. *Journal of virology*, *77*(6), 3838–3845.

Harriman, G. R., Bogue, M., Rogers, P., Finegold, M., Pacheco, S., Bradley, A., ... Mbawuike, I. N. (1999). Targeted Deletion of the IgA Constant Region in Mice Leads to IgA Deficiency with Alterations in Expression of Other Ig Isotypes. *Journal of immunology*, *162*, 2521–2529.

Hasenkrug, K. J., Brooks, D. M. & Dittmer, U. (1998). Critical Role for CD4+ T Cells in Controlling Retrovirus Replication and Spread in Persistently Infected Mice. *Journal of virology*, *72*(8), 6559–6564.

Hasenkrug, K. J. & Chesebro, B. (1997). Immunity to retroviral infection: The Friend virus model. *Proceedings of the national academy of sciences of the united states of america*, *94*, 7811–7816.

Hayashi, H., Matsubara, H., Yokota, T., Kuwabara, I., Kanno, M., Koseki, H., ... Taniguchi, M. (1992). Molecular cloning and characterization of the gene encoding mouse melanoma antigen by cDNA library transfection. *Journal of immunology*, *149*(4), 1223–1229.

He, Z., Wu, L., Li, X., Fields, M. W. & Zhou, J. (2005). Empirical Establishment of Oligonucleotide Probe Design Criteria. *Applied and environmental microbiology*, *71*(7), 3753–3760.

Heidmann, O., Vernochet, C., Dupressoir, A. & Heidmann, T. (2009). Identification of an endogenous retroviral envelope gene with fusogenic activity and placenta-specific expression in the rabbit: a new "syncytin" in a third order of mammals. *Retrovirology*, *6*, 107.

Hemmi, H., Kaisho, T., Takeuchi, O., Sato, S., Sanjo, H., Hoshino, K., ... Akira, S. (2002). Small anti-viral compounds activate immune cells via the TLR7 MyD88 – dependent signaling pathway. *Nature immunology*, *3*(2), 196–200.

Hemmi, H., Takeuchi, O., Kawai, T., Kaisho, T., Sato, S., Sanjo, H., ... Akira, S. (2000). A Toll-like receptor recognizes bacterial DNA. *Nature*, *408*, 740–745.

Herbst, H., Sauter, M. & Mueller-Lantzsch, N. (1996). Expression of human endogenous retrovirus K elements in germ cell and trophoblastic tumors. *American journal of pathology*, *149*(5), 1727–1735.

Hoffmann, J. A., Kafatos, F. C., Janeway, C. A. & Ezekowitz, R. A. B. (1999). Phylogenetic Perspectives in Innate Immunity. *Science*, *284*, 1313–1318.

Hook, L. M., Jude, B. A., Ter-Grigorov, V. S., Hartley, J. W., Morse III, H. C., Trainin, Z., ... Golovkina, T. V. (2002). Characterization of a Novel Murine Retrovirus Mixture That Facilitates Hematopoiesis. *Journal of virology*, *76*(23), 12112–12122.

Horie, M., Honda, T., Suzuki, Y., Kobayashi, Y., Daito, T., Oshida, T., ... Tomonaga, K. (2010). Endogenous non-retroviral RNA virus elements in mammalian genomes. *Nature*, *463*, 84–87.

Horie, M. & Tomonaga, K. (2011). Non-retroviral fossils in vertebrate genomes. *Viruses*, *3*, 1836–1848.

Hou, B., Reizis, B. & Defranco, A. L. (2008). Toll-like Receptors Activate Innate and Adaptive Immunity by using Dendritic Cell-Intrinsic and -Extrinsic Mechanisms. *Immunity*, *29*, 272–282.

Howe, S. J., Mansour, M. R., Schwarzwaelder, K., Bartholomae, C., Hubank, M., Kempski, H., ... Thrasher, A. J. (2008). Insertional mutagenesis combined with acquired somatic mutations causes leukemogenesis following gene therapy of SCID-X1 patients. *Journal of clinical investigation*, *118*(9), 3143–3150.

Hozumi, N. & Tonegawa, S. (1976). Evidence for somatic rearrangement of immunoglobulin genes coding for variable and constant regions. *Proceedings of the national academy of sciences of the united states of america*, *73*(10), 3628–3632.

Huang, A. Y. C., Gulden, P. H., Woods, A. S., Thomas, M. C., Tong, C. D., Wang, W., ... Jaffee, E. M. (1996). The immunodominant major histocompatibility complex class I-restricted antigen of a murine colon tumor derives from an endogenous retroviral gene product. *Proceedings of the national academy of sciences of the united states of america*, *93*, 9730–9735.

Huang, C. R. L., Schneider, A. M., Lu, Y., Niranjan, T., Shen, P., Robinson, M. A., ... Burns, K. H. (2010). Mobile interspersed repeats are major structural variants in the human genome. *Cell*, *141*, 1171–1182.

Hughes, J. F. & Coffin, J. M. (2001). Evidence for genomic rearrangements mediated by human endogenous retroviruses during primate evolution. *Nature genetics*, *29*, 487–489.

Huseby, E. S., Crawford, F., White, J., Kappler, J. & Marrack, P. (2003). Negative selection imparts peptide specificity to the mature T cell repertoire. *Proceedings*

of the national academy of sciences of the united states of america, *100*(20), 11565–11570.

Hutchison, K. W., Copeland, N. G. & Jenkins, N. A. (1984). Dilute-coat-color locus of mice: nucleotide sequence analysis of the d+2J and d+Ha revertant alleles. *Molecular and cellular biology*, *4*(12), 2899–2904.

Ikeda, H. & Sugimura, H. (1989). Fv-4 resistance gene: a truncated endogenous murine leukemia virus with ecotropic interference properties. *Journal of virology*, *63*(12), 5405–5412.

Ikeda, H., Laigret, F., Martin, M. A. & Repaske, R. O. Y. (1985). Characterization of a Molecularly Cloned Retroviral Sequence Associated with Fv-4 Resistance. *Journal of virology*, *55*(3), 768–777.

International Human Genome Sequencing Consortium. (2001). Initial sequencing and analysis of the human genome. *Nature*, *409*, 860–921.

Iskow, R. C., McCabe, M. T., Mills, R. E., Torene, S., Pittard, W. S., Neuwald, A. F., ... Devine, S. E. (2010). Natural mutagenesis of human genomes by endogenous retrotransposons. *Cell*, *141*, 1253–1261.

Itoh, Y., Maruyama, N., Kitamura, M., Shirasawa, T., Shigemoto, K. & Koike, T. (1992). Induction of endogenous retroviral gene product (SU) as an acute-phase protein by IL-6 in murine hepatocytes. *Clinical and experimental immunology*, *88*, 356–359.

Itohara, S., Mombaerts, P., Lafaille, J., Iacomini, J., Nelson, A., Clarke, A. R., ... Tonegawa, S. (1993). T Cell Receptor Delta Gene Mutant Mice: Independent Generation of Alpha Beta T cells and programmed rearrangements of Gamma Delta TCR genes. *Cell*, *72*, 337–348.

Ivanov, I. I., Atarashi, K., Manel, N., Brodie, E. L., Shima, T., Karaoz, U., ... Littman, D. R. (2009). Induction of intestinal Th17 cells by segmented filamentous bacteria. *Cell*, *139*, 485–498.

Iwashiro, M., Kondo, T., Shimizu, T., Yamagishi, H., Takahashi, K., Masuda, T., ... Ishimoto, A. (1993). Multiplicity of Virus-Encoded Helper T-Cell Epitopes Expressed on FBL-3 Tumor Cells. *Journal of virology*, *67*(8), 4533–4542.

Iwashiro, M., Peterson, K., Messer, R. J., Stromnes, I. M. & Hasenkrug, K. J. (2001). CD4+ T Cells and Gamma Interferon in the Long-Term Control of Persistent Friend Retrovirus Infection. *Journal of virology*, *75*(1), 52–60.

Jacobson, S., Richert, J. R., Biddison, W. E., Satinsky, A., Hartzman, R. J. & McFarland, H. F. (1984). Measles virus-specific T4+ human cytotoxic T cell clones are restricted by class II HLA antigens. *Journal of immunology*, *133*(2), 754–757.

Jalloh, A., Jalloh, M. & Matsuoka, H. (2009). T-cell epitope polymorphisms of the Plasmodium falciparum circumsporozoite protein among field isolates from Sierra Leone: age-dependent haplotype distribution? *Malaria journal*, *8*, 120.

Jalloh, A., van Thien, H., Ferreira, M. U., Ohashi, J., Matsuoka, H., Kanbe, T., ... Kawamoto, F. (2006). Sequence Variation in the T-Cell Epitopes of the Plasmodium falciparum Circumsporozoite Protein among Field Isolates Is Temporally Stable: a 5-Year Longitudinal Study in Southern Vietnam. *Journal of clinical microbiology*, *44*(4), 1229–1235.

Jellison, E. R., Kim, S.-K. & Welsh, R. M. (2005). MHC Class II-restricted killing in vivo during viral infection. *Journal of immunology*, *174*, 614–618.

Jenkins, N. A., Copeland, N. G., Taylor, B. A. & Lee, B. K. (1981). Dilute (d) coat colour mutation of DBA/2J mice is associated with the site of integration of an ecotropic MuLV genome. *Nature*, *293*, 370–374.

Jenkins, N. A., Copeland, N. G., Taylor, B. A. & Lee, B. K. (1982). Organization, distribution, and stability of endogenous ecotropic murine leukemia virus DNA sequences in chromosomes of Mus musculus. *Journal of virology*, *43*(1), 26–36.

Jern, P. & Coffin, J. M. (2008, January). Effects of retroviruses on host genome function. *Annual review of genetics*, *42*, 709–732.

Jern, P., Stoye, J. P. & Coffin, J. M. (2007). Role of APOBEC3 in genetic diversity among endogenous murine leukemia viruses. *Plos genetics*, *3*, e183.

Johnston, J. B., Silva, C., Holden, J., Warren, K. G., Clark, A. W. & Power, C. (2001). Monocyte Activation and Differentiation Augment Human Endogenous Retrovirus Expression: Implications for Inflammatory Brain Diseases. *Annals of neurology*, *50*(4), 434–442.

Jordan, I. K., Rogozin, I. B., Glazko, G. V. & Koonin, E. V. (2003). Origin of a substantial fraction of human regulatory sequences from transposable elements. *Trends in genetics*, *19*(2), 68–72.

Jude, B. A., Pobezinskaya, Y., Bishop, J., Parke, S., Medzhitov, R. M., Chervonsky, A. V. & Golovkina, T. V. (2003). Subversion of the innate immune system by a retrovirus. *Nature immunology*, *4*(6), 573–578.

Juriloff, D. M., Harris, M. J., Dewell, S. L., Brown, C. J., Mager, D. L., Gagnier, L. & Mah, D. G. (2005). Investigations of the Genomic Region that Contains the clf1 Mutation, a Causal Gene in Multifactorial Cleft Lip and Palate in Mice. *Birth defects research (part a): clinical and molecular teratology*, *73*, 103–113.

Jurka, J., Kapitonov, V. V., Pavlicek, A., Klonowski, P., Kohany, O. & Walichiewicz, J. (2005, January). Repbase Update, a database of eukaryotic repetitive elements. *Cytogenetic and genome research*, *110*, 462–467.

Kane, M., Case, L. K., Kopaskie, K., Kozlova, A., MacDearmid, C., Chervonsky, A. V. & Golovkina, T. V. (2011). Successful transmission of a retrovirus depends on the commensal microbiota. *Science*, *334*, 245–249.

Kano, H., Godoy, I., Courtney, C., Vetter, M. R., Gerton, G. L., Ostertag, E. M. & Kazazian, H. H. (2009). L1 retrotransposition occurs mainly in embryogenesis and creates somatic mosaicism. *Genes and development*, *23*, 1303–1312.

Karimi, M. M., Goyal, P., Maksakova, I. A., Bilenky, M., Leung, D., Tang, J. X., ... Lorincz, M. C. (2011). DNA methylation and SETDB1/H3K9me3 regulate predominantly distinct sets of genes, retroelements, and chimeric transcripts in mESCs. *Cell stem cell*, *8*, 676–687.

Katoh, K., Misawa, K., Kuma, K.-i. & Miyata, T. (2002). MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic acids research*, *30*(14), 3059–3066.

Katsumata, K., Ikeda, H., Sato, M., Ishizu, A., Kawarada, Y., Kato, H., ... Yoshiki, T. (1999). Cytokine Regulation of env Gene Expression of Human Endogenous Retrovirus-R in Human Vascular Endothelial Cells. *Clinical immunology*, *93*(1), 75–80.

Kelleher, C. A., Wilkinson, D. A., Freeman, J. D., Mager, D. L. & Gelfand, E. W. (1996). Expression of novel transposon-containing mRNAs in human T cells. *Journal of general virology*, *77*, 1101–1110.

Kershaw, M. H., Hsu, C., Mondesire, W., Parker, L. L., Wang, G., Overwijk, W. W., ... Hwu, P. (2001). Immunization against endogenous retroviral tumor-associated antigens. *Cancer research*, *61*, 7920–7924.

Kidd, J. M., Cooper, G. M., Donahue, W. F., Hayden, H. S., Sampas, N., Graves, T., ... Eichler, E. E. (2008). Mapping and sequencing of structural variation from eight human genomes. *Nature*, *453*, 56–64.

Kidd, J. M., Graves, T., Newman, T. L., Fulton, R., Hayden, H. S., Malig, M., ... Eichler, E. E. (2010). A Human Genome Structural Variation Sequencing Resource Reveals Insights into Mutational Mechanisms. *Cell*, *143*, 837–847.

Kim, J., Woods, A., Becker-Dunn, E. & Bottomly, K. (1985). Distinct functional phenotypes of cloned Ia-restricted helper T cells. *Journal of experimental medicine*, *162*, 188–201.

Kitamura, D., Roes, J., Kühn, R. & Rajewsky, K. (1991). A B cell-deficient mouse by targeted disruption of the membrane exon of the immunoglobulin μ chain gene. *Nature*, *350*, 423–426.

Kohany, O., Gentles, A. J., Hankus, L. & Jurka, J. (2006, January). Annotation, submission and screening of repetitive elements in Repbase: RepbaseSubmitter and Censor. *Bmc bioinformatics*, *7*, 474.

Konkel, M. K., Wang, J., Liang, P. & Batzer, M. A. (2007). Identification and characterization of novel polymorphic LINE-1 insertions through comparison of two human genome sequence assemblies. *Gene*, *390*, 28–38.

Koren, O., Goodrich, J. K., Cullender, T. C., Spor, A., Laitinen, K., Bäckhed, H. K., ... Ley, R. E. (2012). Host Remodeling of the Gut Microbiome and Metabolic Changes during Pregnancy. *Cell*, *150*, 470–480.

Kozak, C. A. & Chakraborti, A. (1996). Single Amino Acid Changes in the Murine Leukemia Virus Capsid Protein Gene Define the Target of Fv1 Resistance. *Virology*, *225*, 300–305.

Kozak, C. A. & O'Neill, R. R. (1987). Diverse wild mouse origins of xenotropic, mink cell focus-forming, and two types of ecotropic proviral genes. *Journal of virology*, *61*(10), 3082–3088.

Kozak, C. A. & Rowe, W. P. (1982). Genetic mapping of ecotropic murine leukemia virus-inducing loci in six inbred strains. *Journal of experimental medicine*, *155*, 524–534.

Kramerov, D. A. & Vassetzky, N. S. (2011). SINEs. *Wiley interdisiplinary reviews rna*, *2*, 772–786.

Kuramoto, T., Nakanishi, S., Ochiai, M., Nakagama, H., Voigt, B. & Serikawa, T. (2012). Origins of Albino and Hooded Rats: Implications from Molecular Genetic Analysis across Modern Laboratory Rat Strains. *Plos one*, *7*(8), e43059.

Kuss, S. K., Best, G. T., Etheredge, C. A., Pruijssers, A. J., Frierson, J. M., Hooper, L. V., ... Pfeiffer, J. K. (2011, October). Intestinal microbiota promote enteric virus replication and systemic pathogenesis. *Science*, *334*, 249–252.

Lamprecht, B., Walter, K., Kreher, S., Kumar, R., Hummel, M., Lenze, D., ... Mathas, S. (2010). Derepression of an endogenous long terminal repeat activates the CSF1R proto-oncogene in human lymphoma. *Nature medicine*, *16*(5), 571–579.

Larsson, E., Tremaroli, V., Lee, Y. S., Koren, O., Nookaew, I., Fricker, A., ... Bäckhed, F. (2012). Analysis of gut microbial regulation of host gene expression along the length of the gut and regulation of gut microbial ecology through MyD88. *Gut*, *61*, 1124–1131.

Lavie, L., Kitova, M., Maldener, E., Meese, E. & Mayer, J. (2005). CpG Methylation Directly Regulates Transcriptional Activity of the Human Endogenous Retrovirus Family HERV-K(HML-2). *Journal of virology*, *79*(2), 876–883.

Law, L. W. (1962). Influence of Foster-Nursing on Virus-Induced and Spontaneous Leukemia in Mice. *Proceedings of the society for experimental biology and medicine*, *111*(3), 615–623.

Lee, E., Iskow, R. C., Yang, L., Gokcumen, O., Hasely, P., Luquette III, L. J., ... The Cancer Genome Atlas Research Network. (2012). Landscape of Somatic Retrotransposition in Human Cancers. *Science*, *337*, 967–971.

Lee, W. J., Kwun, H. J., Kim, H. S. & Jang, K. L. (2003). Activation of the Human Endogenous Retrovirus W Long Terminal Repeat by Herpes Simplex Virus Type 1 Immediate Early Protein 1. *Molecules and cells*, *15*(1), 75–80.

Lee, Y. N. & Bieniasz, P. D. (2007). Reconstitution of an infectious human endogenous retrovirus. *Plos pathogens*, *3*(1), e10.

Lewinski, M. K., Yamashita, M., Emerman, M., Ciuffi, A., Marshall, H., Crawford, G., ... Bushman, F. D. (2006). Retroviral DNA integration: viral and cellular determinants of target-site selection. *Plos pathogens*, *2*(6), e60.

Li, J.-P., D'Andrea, A. D., Lodish, H. F. & Baltimore, D. (1990). Activation of cell growth by binding of Friend spleen focus-forming virus gp55 glycoprotein to the erythropoietin receptor. *Nature*, *343*, 762–764.

Li, M., Huang, X., Zhu, Z. & Gorelik, E. (1999). Sequence and insertion sites of murine melanoma-associated retrovirus. *Journal of virology*, *73*(11), 9178–9186.

Lieberman, M. & Kaplan, H. S. (1959). Leukemogenic activity of filtrates from radiation-induced lymphoid tumors of mice. *Science*, *130*, 387–388.

Limjoco, T. I., Dickie, P., Ikeda, H. & Silver, J. (1993). Transgenic Fv-4 mice resistant to Friend virus. *Journal of virology*, *67*(7), 4163–4168.

Lockhart, D. J., Dong, H., Byrne, M. C., Follettie, M. T., Gallo, M. V., Chee, M. S., ... Brown, E. L. (1996). Expression monitoring by hybridization to high-density oligonucleotide arrays. *Nature biotechnology*, *14*, 1675–1680.

Lötscher, M., Recher, M., Lang, K. S., Navarini, A., Hunziker, L., Santimaria, R., ... Zinkernagel, R. M. (2007, January). Induced prion protein controls immune-activated retroviruses in the mouse spleen. *Plos one*, *2*(11), e1158.

Lowy, D. R., Rowe, W. P., Teich, N. & Hartley, J. W. (1971). Murine Leukemia Virus: High-Frequency Activation in vitro by 5-Iododeoxyuridine and 5- Bromodeoxyuridine. *Science*, *174*, 155–156.

Ludewig, B., Ochsenbein, A. F., Odermatt, B., Paulin, D., Hengartner, H. & Zinkernagel, R. M. (2000). Immunotherapy with dendritic cells directed against tumor antigens shared with normal host cells results in severe autoimmune disease. *Journal of experimental medicine*, *191*(5), 795–804.

Lukacher, A. E., Morrison, L. A., Braciale, V. L., Malissen, B. & Braciale, T. J. (1985). Expression of specific cytolytic activity by H-2l region-restricted, influenza virus-specific T lymphocyte clones. *Journal of experimental medicine*, *162*, 171–187.

Lund, J. M., Alexopoulou, L., Sato, A., Karow, M., Adams, N. C., Gale, N. W., ... Flavell, R. A. (2004). Recognition of single-stranded RNA viruses by Toll-like receptor 7. *Proceedings of the national academy of sciences of the united states of america*, *101*(15), 5598–5603.

MacDuff, D. A., Demorest, Z. L. & Harris, R. S. (2009). AID can restrict L1 retrotransposition suggesting a dual role in innate and adaptive immunity. *Nucleic acids research*, *37*(6), 1854–1867.

Macfarlan, T. S., Gifford, W. D., Agarwal, S., Driscoll, S., Lettieri, K., Wang, J., ... Pfaff, S. L. (2011). Endogenous retroviruses and neighboring genes are coordinately repressed by LSD1/KDM1A. *Genes and development*, *25*(6), 594–607.

Macfarlan, T. S., Gifford, W. D., Driscoll, S., Lettieri, K., Rowe, H. M., Bonanomi, D., ... Pfaff, S. L. (2012). Embryonic stem cell potency fluctuates with endogenous retrovirus activity. *Nature*, *487*, 57–63.

Magiorkinis, G., Gifford, R. J., Katzourakis, A., De Ranter, J. & Belshaw, R. (2012). Env-less endogenous retroviruses are genomic superspreaders. *Proceedings of the national academy of sciences of the united states of america*, *109*(19), 7385–7390.

Malik, H. S., Henikoff, S. & Eickbush, T. H. (2000). Poised for Contagion : Evolutionary Origins of the Infectious Abilities of Invertebrate Retroviruses. *Genome research*, *10*, 1307–1318.

Marques, R., Antunes, I., Eksmond, U., Stoye, J. P., Hasenkrug, K. J. & Kassiotis, G. (2008). B lymphocyte activation by coinfection prevents immune control of friend virus infection. *Journal of immunology*, *181*, 3432–3440.

Martin, D. P., Lemey, P., Lott, M., Moulton, V., Posada, D. & Lefeuvre, P. (2010). RDP3: a flexible and fast computer program for analyzing recombination. *Bioinformatics*, *26*(19), 2462–2463.

Martin, M. A., Bryan, T., Rasheedt, S. & Khan, A. S. (1981). Identification and cloning of endogenous retroviral sequences present in human DNA. *Proceedings of the national academy of sciences of the united states of america*, *78*(8), 4892–4896.

Maryanski, J. L., Attuil, V., Hamrouni, A., Mutin, M., Rossi, M., Aublin, A. & Bucher, P. (2001). Individuality of Ag-Selected and Preimmune TCR Repertoires. *Immunologic research*, *23*(1), 75–84.

Matsui, T., Leung, D., Miyashita, H., Maksakova, I. A., Miyachi, H., Kimura, H., ... Shinkai, Y. (2010). Proviral silencing in embryonic stem cells requires the histone methyltransferase ESET. *Nature*, *464*, 927–931.

Mazmanian, S. K., Liu, C. H., Tzianabos, A. O. & Kasper, D. L. (2005). An immunomodulatory molecule of symbiotic bacteria directs maturation of the host immune system. *Cell*, *122*, 107–118.

McCarthy, E. M. & McDonald, J. F. (2004). Long terminal repeat retrotransposons of Mus musculus. *Genome biology*, *5*, R14.

McClintock, B. (1950). The origin and behaviour of mutable loci in maize. *Proceedings of the national academy of sciences of the united states of america*, *36*, 344–355.

McCubrey, J. & Risser, R. (1982). Genetic interactions in the spontaneous production of endogenous murine leukemia virus in low leukemic mouse strains. *Journal of experimental medicine*, *156*, 337–349.

McWilliams, J. A., Sullivan, R. T., Jordan, K. R., McMahan, R. H., Kemmler, C. B., McDuffie, M. & Slansky, J. E. (2008). Age-dependent tolerance to an endogenous tumor-associated antigen. *Vaccine*, *26*, 1863–1873.

Medstrand, P., van de Lagemaat, L. N. & Mager, D. L. (2002). Retroelement Distributions in the Human Genome: Variations Associated With Age and Proximity to Genes. *Genome research*, *12*, 1483–1495.

Mi, S., Lee, X., Li, X.-p., Veldman, G. M., Finnerty, H., Racie, L., ... McCoy, J. M. (2000). Syncytin is a captive retroviral envelope protein involved in human placental morphogenesis. *Nature*, *403*, 785–789.

Miller, D. G., Edwards, R. H. & Miller, A. D. (1994). Cloning of the cellular receptor for amphotropic murine retroviruses reveals homology to that for gibbon ape leukemia virus. *Proceedings of the national academy of sciences of the united states of america*, *91*, 78–82.

Miyazawa, M., Nishio, J. & Chesebro, B. (1988). Genetic control of T cell responsiveness to the Friend murine leukemia virus envelope antigen. Identification of Class II Loci of the H-2 as Immune Response Genes. *Journal of experimental medicine*, *168*, 1587–1605.

Miyazawa, M., Tsuji-Kawahara, S. & Kanari, Y. (2008). Host genetic factors that control immune responses to retrovirus infections. *Vaccine*, *26*, 2981–2996.

Mombaerts, P., Iacomini, J., Johnson, R. S., Herrup, K., Tonegawa, S. & Papaioannou, V. E. (1992, March). RAG-1-deficient mice have no mature B and T lymphocytes. *Cell*, *68*, 869–877.

Moreau-Gachelin, F., Tavitian, A. & Tambourin, P. (1988). Spi-1 is a putative oncogene in virally induced murine erythroleukaemias. *Nature*, *331*, 277–280.

Moroni, C. & Schumann, G. (1975). Lipopolysaccharide induces C-type virus in short term culture of BALB/c spleen cells. *Nature*, *254*, 60–61.

Moroni, C., Schumann, G., Robert-Guroff, M., Suter, E. R. & Martin, D. (1975). Induction of endogenous murine C-type virus in spleen cell cultures treated with mitogens and 5-bromo-2'-deoxyuridine. *Proceedings of the national academy of sciences of the united states of america*, *72*(2), 535–538.

Mosmann, T. R., Cherwinski, H., Bond, M. W., Giedlin, M. A. & Coffman, R. L. (1986). Two types of murine helper T cell clone I. Definition according to proflies of lymphokine activities and secreted proteins. *Journal of immunology*, *136*(7), 2348–2357.

Mouse Genome Sequencing Consortium. (2002). Initial sequencing and comparative analysis of the mouse genome. *Nature*, *420*, 520–562.

Mowat, M., Cheng, A., Kimura, N., Bernstein, A. & Benchimol, S. (1985). Rearrangements of the cellular p53 gene in erythroleukaemic cells transformed by Friend virus. *Nature*, *314*, 633–636.

Mühlbock, O. (1955). Note on a new inbred mouse strain GR/A. *European journal of cancer*, *1*, 123–124.

Muramatsu, M., Kinoshita, K., Fagarasan, S., Yamada, S., Shinkai, Y. & Honjo, T. (2000). Class Switch Recombination and Hypermutation Require Activation-Induced Cytidine Deaminase (AID), a Potential RNA Editing Enzyme. *Cell*, *102*, 553–563.

Nair, S. R., Zelinskyy, G., Schimmer, S., Gerlach, N., Kassiotis, G. & Dittmer, U. (2010). Mechanisms of control of acute Friend virus infection by CD4+ T helper

cells and their functional impairment by regulatory T cells. *Journal of general virology*, *91*, 440–451.

Nekrutenko, A. & Li, W.-H. (2001). Transposable elements are found in a large number of human protein-coding genes. *Trends in genetics*, *17*(11), 619–621.

Nellåker, C., Keane, T. M., Yalcin, B., Wong, K., Agam, A., Belgard, T. G., ... Ponting, C. P. (2012). The genomic landscape shaped by selection on transposable elements across 18 mouse strains. *Genome biology*, *13*, R45.

Nellåker, C., Yao, Y., Jones-Brando, L., Mallet, F., Yolken, R. H. & Karlsson, H. (2006). Transactivation of elements in the human endogenous retrovirus W family by viral infection. *Retrovirology*, *3*, 44.

Nicholson, J. K., Holmes, E., KInross, J., Burcelin, R., Gibson, G., Jia, W. & Pettersson, S. (2012). Host-Gut Microbiota Metabolic Interactions. *Science*, *336*, 1262–1267.

Nicklas, W., Baneux, P., Boot, R., Decelle, T., Deeny, A. A., Fumanelli, M. & Illgen-Wilcke, B. (2002). Recommendations for the health monitoring of rodent and rabbit colonies in breeding and experimental units. *Laboratory animals*, *36*, 20–42.

Nikolich-Zugich, J., Slifka, M. K. & Messaoudi, I. (2004). The many important facets of T-cell repertoire diversity. *Nature reviews immunology*, *4*, 123–132.

Oie, H. K., Gazdar, A. F., Lalley, P. A., Russell, E. K., Minna, J. D., DeLarco, J., ... Francke, U. (1978). Mouse chromosome 5 codes for ecotropic murine leukaemia virus cell-surface receptor. *Nature*, *274*, 60–62.

Okonechnikov, K., Golosova, O. & Fursov, M. (2012). Unipro UGENE: a unified bioinformatics toolkit. *Bioinformatics*, *28*(8), 1166–1167.

Olszak, T., An, D., Zeissig, S., Vera, M. P., Richter, J., Franke, A., ... Blumberg, R. S. (2012). Microbial exposure during early life has persistent effects on Natural Killer T cell function. *Science*, *336*, 489–493.

Park, M. A., Li, J. T., Hagan, J. B., Maddox, D. E. & Abraham, R. S. (2008). Common variable immunodeficiency: a new look at an old disease. *Lancet*, *372*, 489–502.

Pasare, C. & Medzhitov, R. (2005). Control of B-cell responses by Toll-like receptors. *Nature*, *438*, 364–368.

Persons, D. A., Paulson, R. F., Loyd, M. R., Herley, M. T., Bodner, S. M., Bernstein, A., ... Ney, P. A. (1999). Fv2 encodes a truncated form of the Stk receptor tyrosine kinase. *Nature genetics*, *23*, 159–165.

Peterson, D. A., McNulty, N. P., Guruge, J. L. & Gordon, J. I. (2007). IgA response to symbiotic bacteria as a mediator of gut homeostasis. *Cell host and microbe*, *2*(5), 328–339.

Philpott, K. L., Viney, J. L., Kay, G., Rastan, S., Gardiner, E. M., Chae, S., ... Owen, M. J. (1992, June). Lymphoid development in mice congenitally lacking T cell receptor alpha beta-expressing cells. *Science*, *256*, 1448–1452.

Pike, R., Filby, A., Ploquin, M. J.-Y., Eksmond, U., Marques, R., Antunes, I., ... Kassiotis, G. (2009). Race between retroviral spread and CD4+ T-cell response determines the outcome of acute Friend virus infection. *Journal of virology*, *83*(21), 11211–11222.

Ploquin, M. J.-Y., Eksmond, U. & Kassiotis, G. (2011). B cells and TCR avidity determine distinct functions of CD4+ T cells in retroviral infection. *Journal of immunology*, *187*, 3321–3330.

Pothlichet, J., Mangeney, M. & Heidmann, T. (2006). Mobility and integration sites of a murine C57BL/6 melanoma endogenous retrovirus involved in tumor progression in vivo. *International journal of cancer*, *119*, 1869–1877.

Pruitt, K. D., Tatusova, T. & Maglott, D. R. (2007). NCBI reference sequences (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins. *Nucleic acids research*, *35*, D61–D65.

Purcell, D. F. J., Broscius, C. M., Vanin, E. F., Buckler, C. E., Nienhuis, A. W. & Martin, M. A. (1996). An array of murine leukemia virus-related elements is transmitted and expressed in a primate recipient of retroviral gene transfer. *Journal of virology*, *70*(2), 887–897.

Rasmussen, M. H., Ballarín-González, B., Liu, J., Lassen, L. B., Füchtbauer, A., Füchtbauer, E.-M., ... Pedersen, F. S. (2010). Antisense transcription in gammaretroviruses as a mechanism of insertional activation of host genes. *Journal of virology*, *84*(8), 3780–3788.

Rast, J. P., Smith, L. C., Loza-Coll, M., Hibino, T. & Litman, G. W. (2006). Genomic Insights into the Immune System of the Sea Urchin. *Science*, *314*, 952–956.

Rauscher, F. J. (1962). A Virus-Induced Disease of Mice Characterized by Erythrocytopoiesis and Lymphoid Leukemia. *Journal of the national cancer institute*, *29*(3), 515–543.

Rearden, A., Magnet, A., Kudo, S. & Fukuda, M. (1993). Glycophorin B and glycophorin E genes arose from the glycophorin A ancestral gene via two duplications during primate evolution. *The journal of biological chemistry*, *268*(3), 2260–2267.

Reichmann, J., Crichton, J. H., Madej, M. J., Taggart, M., Gautier, P., Garcia-Perez, J. L., ... Adams, I. R. (2012). Microarray Analysis of LTR Retrotransposon Silencing Identifies Hdac1 as a Regulator of Retrotransposon Expression in Mouse Embryonic Stem Cells. *Plos computational biology*, *8*(4), e1002486.

Reid, R. R., Prodeus, A. P., Khan, W., Hsu, T., Rosen, F. S. & Carroll, M. C. (1997). Endotoxin Shock in Antibody-Deficient Mice. *Journal of immunology*, *159*, 970–975.

Repaske, R., Steele, P. E., O'Neill, R., Rabson, A. B. & Martin, M. A. (1985). Nucleotide Sequence of a Full-Length Human Endogenous Retroviral Segment. *Journal of virology*, *54*(3), 764–772.

Ribet, D., Harper, F., Dupressoir, A., Dewannieux, M., Pierron, G. & Heidmann, T. (2008, April). An infectious progenitor for the murine IAP retrotransposon: emergence of an intracellular genetic parasite from an ancient retrovirus. *Genome research*, *18*, 597–609.

Roe, T., Reynolds, T. C., Yu, G. & Brown, P. O. (1993). Integration of murine leukemia virus DNA depends on mitosis. *Embo journal*, *12*(5), 2099–2108.

Rosato, A., Dalla Santa, S., Zoso, A., Giacomelli, S., Milan, G., Macino, B., ... Zanovello, P. (2003). The cytotoxic T-lymphocyte response against a poorly immunogenic mammary adenocarcinoma is focused on a single immunodominant class I epitope derived from the gp70 Env product of an endogenous retrovirus. *Cancer research*, *63*, 2158–2163.

Rosenberg, Z. F. & Fauci, A. S. (1989). Induction of Expression of HIV in Latently Chronically Infected Cells. *Aids research and human retroviruses*, *5*(1), 711–713.

Rous, P. (1910). A Transmissible Avian Neoplasm (Sarcoma of the Common Fowl). *Journal of experimental medicine*, *12*, 696–705.

Rous, P. (1911). A Sarcoma of the Fowl Transmissible by an Agent Separable from the Tumor Cells. *Journal of experimental medicine*, *13*, 397–411.

Rowe, H. M., Jakobsson, J., Mesnard, D., Rougemont, J., Reynard, S., Aktas, T., ... Trono, D. (2010). KAP1 controls endogenous retroviruses in embryonic stem cells. *Nature*, *463*, 237–40.

Rowe, W. P., Hartley, J. W., Lander, M. R., Pugh, W. E. & Teich, N. (1971). Noninfectiuous AKR Mouse Embyro Cell Lines in Which Each Cell Has the Capacity to be Activated to Produce Infectious Murine Leukemia Virus. *Virology*, *46*, 866–876.

Rudensky, A. Y., Preston-Hurlburt, P., Hong, S.-C., Barlow, A. & Janeway, C. A. J. (1991). Sequence analysis of peptides bound to MHC class II molecules. *Nature*, *353*, 622–627.

Samuelson, L. C., Phillips, R. S. & Swanberg, L. J. (1996). Amylase gene structures in primates: retroposon insertions and promoter evolution. *Molecular biology and evolution*, *13*(6), 767–779.

Sauter, M., Roemer, K., Best, B., Afting, M., Schommer, S., Seitz, G., ... Mueller-Lanzsch, N. (1996). Specificity of antibodies directed against Env protein of human endogenous retroviruses in patients with germ cell tumors. *Cancer research*, *56*, 4362–4365.

Schaller, T., Ocwieja, K. E., Rasaiyaah, J., Price, A. J., Brady, T. L., Roth, S. L., ... Towers, G. J. (2011). HIV-1 capsid-cyclophilin interactions determine nuclear import pathway, integration targeting and replication efficiency. *Plos pathogens*, *7*(12), e1002439.

Schena, M., Shalon, D., Davis, R. W. & Brown, P. O. (1995). Quantitative Monitoring of Gene Expression Patterns with a Complementary DNA Microarray. *Science*, *270*, 467–470.

Schwartz, A., Chan, D. C., Brown, L. G., Alagappan, R., Pettay, D., Disteche, C., ... Page, D. C. (1998). Reconstructing hominid Y evolution: X-homologous block, created by X-Y transposition, was disrupted by Yp inversion through LINE-LINE recombination. *Human molecular genetics*, *7*(1), 1–11.

Seifarth, W., Frank, O., Zeilfelder, U., Spiess, B., Greenwood, A. D. & Leib-Mosch, C. (2005). Comprehensive Analysis of Human Endogenous Retrovirus Transcriptional Activity in Human Tissues with a Retrovirus-Specific Microarray. *Journal of virology*, *79*(1), 341–352.

Seifarth, W., Spiess, B., Zeilfelder, U., Speth, C., Hehlmann, R. & Leib-Mösch, C. (2003). Assessment of retroviral activity using a universal retrovirus chip. *Journal of virological methods*, *112*, 79–91.

Shih, C.-C., Stoye, J. P. & Coffin, J. M. (1988). Highly preferred targets for retrovirus integration. *Cell*, *53*, 531–537.

Shimizu, T., Uenishi, H., Teramura, Y., Iwashiro, M., Kuribayashi, K., Tamamura, H., ... Yamagishi, H. (1994). Fine Structure of a Virus-Encoded Helper T-Cell Epitope Expressed on FBL-3 Tumor Cells. *Journal of virology*, *68*(12), 7704–7708.

Shinnick, T. M., Lerner, R. A. & Sutcliffe, J. G. (1981). Nucleotide sequence of Moloney murine leukaemia virus. *Nature*, *293*, 543–548.

Shulzhenko, N., Morgun, A., Hsiao, W., Battle, M., Yao, M., Gavrilova, O., ... Matzinger, P. (2011). Crosstalk between B lymphocytes, microbiota and the intestinal epithelium governs immunity versus metabolism in the gut. *Nature medicine*, *17*(12), 1585–1593.

Slansky, J. E., Rattis, F. M., Boyd, L. F., Fahmy, T., Jaffee, E. M., Schneck, J. P., ... Pardoll, D. M. (2000). Enhanced antigen-specific antitumor immunity with altered peptide ligands that stabilize the MHC-peptide-TCR complex. *Immunity*, *13*, 529–538.

Smit, A. F. A., Hubley, R. & Green, P. (1996). RepeatMasker.

Snášel, J., Rosenberg, I., Pačes, O. & Pichová, I. (2009). Mapping of HIV-1 integrase preferences for target site selection with various oligonucleotides. *Archives of biochemistry and biophysics*, *488*(2), 153–162.

Soll, S. J., Neil, S. J. D. & Bieniasz, P. D. (2010). Identification of a receptor for an extinct virus. *Proceedings of the national academy of sciences of the united states of america*, *107*, 19496–19501.

Specht, J. M., Wang, G., Do, M. T., Lam, J. S., Royal, R. E., Reeves, M. E., ... Hwu, P. (1997). Dendritic cells retrovirally transduced with a model antigen gene are therapeutically effective against established pulmonary metastases. *Journal of experimental medicine*, *186*(8), 1213–1221.

Spencer, T. E., Mura, M., Gray, C. A., Griebel, P. J. & Palmarini, M. (2003). Receptor Usage and Fetal Expression of Ovine Endogenous Betaretroviruses: Implications

for Coevolution of Endogenous and Exogenous Retroviruses. *Journal of virology*, *77*(1), 749–753.

Stetson, D. B., Ko, J. S., Heidmann, T. & Medzhitov, R. (2008). Trex1 prevents cell-intrinsic initiation of autoimmunity. *Cell*, *134*, 587–598.

Stocking, C. & Kozak, C. A. (2008). Murine endogenous retroviruses. *Cellular and molecular life sciences*, *65*, 3383–3398.

Stoye, J. P. & Coffin, J. M. (1987). The four classes of endogenous murine leukemia virus: structural relationships and potential for recombination. *Journal of virology*, *61*(9), 2659–2669.

Stoye, J. P. & Coffin, J. M. (1988). Polymorphism of murine endogenous proviruses revealed by using virus class-specific oligonucleotide probes. *Journal of virology*, *62*(1), 168–175.

Stoye, J. P., Fenner, S., Greenoak, G. E., Moran, C. & Coffin, J. M. (1988). Role of endogenous retroviruses as mutagens: the hairless mutation of mice. *Cell*, *54*, 383–391.

Stoye, J. P. & Moroni, C. (1983). Endogenous retrovirus expression in stimulated murine lymphocytes. Identification of a new locus controlling mitogen induction of a defective virus. *Journal of experimental medicine*, *157*, 1660–1674.

Stoye, J. P. & Moroni, C. (1985). Proliferation and differentiation requirements for the induction of two retroviral loci during B-cell activation. *Journal of general virology*, *66*, 109–120.

Stoye, J. P., Moroni, C. & Coffin, J. M. (1991). Virological events leading to spontaneous AKR thymomas. *Journal of virology*, *65*(3), 1273–1285.

Su, A. I., Cooke, M. P., Ching, K. A., Hakak, Y., Walker, J. R., Wiltshire, T., ... Hogenesch, J. B. (2002). Large-scale analysis of the human and mouse transcriptomes. *Proceedings of the national academy of sciences of the united states of america*, *99*(7), 4465–4470.

Suzuki, K., Meek, B., Doi, Y., Muramatsu, M., Chiba, T., Honjo, T. & Fagarasan, S. (2004). Aberrant expansion of segmented filamentous bacteria in IgA-deficient gut. *Proceedings of the national academy of sciences of the united states of america*, *101*(7), 1981–1986.

Tailor, C. S., Nouri, A., Lee, C. G., Kozak, C. A. & Kabat, D. (1999). Cloning and characterization of a cell surface receptor for xenotropic and polytropic murine leukemia viruses. *Proceedings of the national academy of sciences of the united states of america*, *96*, 927–932.

Takahashi, Y., Harashima, N., Kajigaya, S., Yokoyama, H., Cherkasova, E., McCoy, J. P., ... Childs, R. W. (2008). Regression of human kidney cancer following allogeneic stem cell transplantation is associated with recognition of an HERV-E antigen by T cells. *Journal of clinical investigation*, *118*(3), 1099–1109.

Tanaka-Taya, K., Sashihara, J., Kurahashi, H., Amo, K., Miyagawa, H., Kondo, K., ... Yamanishi, K. (2004). Human herpesvirus 6 (HHV-6) is transmitted from parent to child in an integrated form and characterization of cases with chromosomally integrated HHV-6 DNA. *Journal of medical virology*, *73*, 465–473.

Taniguchi, R. T. & Anderson, M. S. (2010). The role of Aire in clonal selection. *Immunology and cell biology*, *89*, 40–44.

Tarlinton, R. E., Meers, J. & Young, P. R. (2006). Retroviral invasion of the koala genome. *Nature*, *442*, 79–81.

Temin, H. M. (1963). The Effects of Actinomycin D on Growth of Rous Sarcoma Virus in Vitro. *Virology*, *20*, 577–582.

Temin, H. M. (1964). Nature of the provirus of Rous sarcoma. *Journal of the national cancer institute*, *17*, 557–570.

Temin, H. M. (1980). Origin of retroviruses from cellular moveable genetic elements. *Cell*, *21*, 599–600.

Temin, H. M. & Mizutani, S. (1970). RNA-dependent DNA Polymerase in Virions of Rous Sarcoma Virus. *Nature*, *226*, 1211–1213.

The ENCODE Project Consortium. (2007). Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. *Nature*, *447*, 799–816.

The FANTOM Consortium & RIKEN Genome Exploration Research Group and Genome Science Group. (2005). The transcriptional landscape of the mammalian genome. *Science*, *309*, 1559–1563.

The Human Microbiome Project Consortium. (2012). Structure, function and diversity of the healthy human microbiome. *Nature*, *486*, 207–214.

Thomas, J. H. & Schneider, S. (2011). Coevolution of retroelements and tandem zinc finger genes. *Genome research*, *21*(11), 1800–1812.

Thomson, S. J. P., Goh, F. G., Banks, H., Krausgruber, T., Kotenko, S. V., Foxwell, B. M. J. & Udalova, I. A. (2009). The role of transposable elements in the regulation of IFN-lambda1 gene expression. *Proceedings of the national academy of sciences of the united states of america*, *106*(28), 11564–11569.

Tipper, C. H., Bencsics, C. E. & Coffin, J. M. (2005). Characterization of hortulanus endogenous murine leukemia virus, an endogenous provirus that encodes an infectious murine leukemia virus of a novel subgroup. *Journal of virology*, *79*(13), 8316–8329.

Trasande, L., Blustein, J., Liu, M., Corwin, E., Cox, L. M. & Blaser, M. J. (2012). Infant antibiotic exposures and early-life body mass. *International journal of obesity*, In.

Trinchieri, G. (2012). Cancer and Inflammation: An Old Intuition with Rapidly Evolving New Concepts. *Annual review of immunology*, *30*, 677–706.

Tristem, M. (2000). Identification and Characterization of Novel Human Endogenous Retrovirus Families by Phylogenetic Screening of the Human Genome Mapping Project Database. *Journal of virology*, *74*(8), 3715–3730.

Turcotte, K., Gauthier, S., Tuite, A., Mullick, A., Malo, D. & Gros, P. (2005). A mutation in the Icsbp1 gene causes susceptibility to infection and a chronic myeloid leukemia-like syndrome in BXH-2 mice. *Journal of experimental medicine*, *201*(6), 881–890.

Turner, S. J., Doherty, P. C., McCluskey, J. & Rossjohn, J. (2006). Structural determinants of T-cell receptor bias in immunity. *Nature reviews immunology*, *6*, 883–894.

Uren, T. K., Johansen, F.-e., Wijburg, O. L. C., Koentgen, F., Brandtzaeg, P. & Strugnell, R. A. (2003). Role of the Polymeric Ig Receptor in Mucosal B Cell Homeostasis. *Journal of immunology*, *170*, 2531–2539.

Vallée, H. & Carré, H. (1904). Sur la nature infectieuse de l'anémie du cheval. *Comptes rendus de l'académie des sciences*, *139*, 331–333.

van de Lagemaat, L. N., Landry, J.-R., Mager, D. L. & Medstrand, P. (2003). Transposable elements in mammals promote regulatory variation and diversification of genes with specialized functions. *Trends in genetics*, *19*(10), 530–536.

van der Waaij, L. A., Limburg, P. C., Mesander, G. & van der Waaij, D. (1996). In vivo IgA coating of anaerobic bacteria in human faeces. *Gut*, *38*, 348–354.

van Zeijl, M., Johann, S. V., Closst, E., Cunninghamt, J., Eddy, R., Shows, T. B. & O'Hara, B. (1994). A human amphotropic retrovirus receptor is a second member of the gibbon ape leukemia virus receptor family. *Proceedings of the national academy of sciences of the united states of america*, *91*, 1168–1172.

Venables, P. J. W., Brookes, S. M., Griffiths, D., Weiss, R. A. & Boyd, M. T. (1995). Abundance of an endogenous retroviral envelope protein in placental trophoblasts suggests a biological function. *Virology*, *211*, 589–592.

Wang, H., Kavanaugh, M. P., North, R. A. & Kabat, D. (1991). Cell-surface receptor for ecotropic murine retroviruses is a basic amino-acid transporter. *Nature*, *352*, 729–731.

Wang-Johanning, F., Radvanyi, L., Rycaj, K., Plummer, J. B., Yan, P., Sastry, K. J., ... Johanning, G. L. (2008). Human endogenous retrovirus K triggers an antigen-specific immune response in breast cancer patients. *Cancer research*, *68*(14), 5869–5877.

Weigert, M., Gatmaitan, L., Loh, E., Schilling, J. & Hood, L. (1978). Rearrangement of genetic information may produce immunoglobulin diversity. *Nature*, *276*, 785–790.

Weiss, R. A. (2006, January). The discovery of endogenous retroviruses. *Retrovirology*, *3*, 67.

Weiss, R. A., Friis, R. R., Katz, E. & Vogt, P. K. (1971). Induction of Avian Tumor Viruses in Normal Cells by Physical and Chemical Carcinogens. *Virology*, *46*, 920–938.

Wentzensen, N., Coy, J. F., Knaebel, H.-P., Linnebacher, M., Wilz, B., Gebert, J. & von Knebel Doeberitz, M. (2007). Expression of an endogenous retroviral sequence from the HERV-H group in gastrointestinal cancers. *International journal of cancer*, *121*(7), 1417–1423.

White, H. D., Roeder, D. A. & Green, W. R. (1994). An immunodominant Kb-restricted peptide from the p15E transmembrane protein of endogenous ecotropic murine leukemia virus (MuLV) AKR623 that restores susceptibility of a tumor line to anti-AKR/Gross MuLV cytotoxic T lymphocytes. *Journal of virology*, *68*(2), 897–904.

Wilson, C. A., Wong, S., Muller, J., Davidson, C. E., Rose, T. M. & Burd, P. (1998). Type C retrovirus released from porcine primary peripheral blood mononuclear cells infects human cells. *Journal of virology*, *72*(4), 3082–3087.

Wing, K. & Sakaguchi, S. (2010). Regulatory T cells exert checks and balances on self tolerance and autoimmunity. *Nature immunology*, *11*(1), 7–13.

Wissing, S., Montano, M., Garcia-Perez, J. L., Moran, J. V. & Greene, W. C. (2011). Endogenous APOBEC3B restricts LINE-1 retrotransposition in transformed cells and human embryonic stem cells. *Journal of biological chemistry*, *286*(42), 36427–36437.

Wu, L., Kohler, J. E., Zaborina, O., Akash, G., Musch, M. W., Chang, E. B. & Alverdy, J. C. (2006). Chronic Acid Water Feeding Protects Mice Against Lethal Gut-derived Sepsis due to Pseudomonas aeruginosa. *Current issues in interstinal microbiology*, *7*, 19–28.

Xiong, Y. & Eickbush, T. H. (1990). Origin and evolution of retroelements based upon their reverse transcriptase sequences. *Embo journal*, *9*(10), 3353–3362.

Yalcin, B., Wong, K., Agam, A., Goodson, M., Keane, T. M., Gan, X., ... Flint, J. (2011). Sequence-based characterization of structural variation in the mouse genome. *Nature*, *477*, 326–329.

Yan, Y., Buckler-White, A., Wollenberg, K. & Kozak, C. A. (2009). Origin, antiviral function and evidence for positive selection of the gammaretrovirus restriction gene Fv1 in the genus Mus. *Proceedings of the national academy of sciences of the united states of america*, *106*(9), 3259–3263.

Yang, Y.-L., Guo, L., Xu, S., Holland, C. A., Kitamura, T., Hunter, K. & Cunningham, J. M. (1999). Receptors for polytropic and xenotropic mouse leukaemia viruses encoded by a single gene at Rmc1. *Nature genetics*, *21*, 216–219.

Yoshiki, T., Mellors, R. C., Strand, M. & August, J. T. (1974). The viral envelope glycoprotein of murine leukemia virus and the pathogenesis of immune complex glom-

erulonephritis of New Zealand mice. *Journal of experimental medicine*, *140*(4), 1011–1027.

Yoshinobu, K., Baudino, L., Santiago-Raber, M.-L., Morito, N., Dunand-Sauthier, I., Morley, B. J., ... Izui, S. (2009, June). Selective up-regulation of intact, but not defective env RNAs of endogenous modified polytropic retrovirus by the Sgp3 locus of lupus-prone mice. *Journal of immunology*, *182*(12), 8094–8103.

Young, G. R., Kassiotis, G. & Stoye, J. P. (2012). Emv2, the only endogenous ecotropic murine leukemia virus of C57BL/6J mice. *Retrovirology*, *9*, 23.

Young, G. R., Ploquin, M. J.-Y., Eksmond, U., Wadwa, M., Stoye, J. P. & Kassiotis, G. (2012). Negative Selection by an Endogenous Retrovirus Promotes a Higher-Avidity CD4+ T Cell Response to Retroviral Infection. *Plos pathogens*, *8*(5), e1002709.

Zhang, X.-B., Beard, B. C., Trobridge, G. D., Wood, B. L., Sale, G. E., Sud, R., ... Kiem, H.-P. (2008). High incidence of leukemia in large animals after stem cell gene therapy with a HOXB4-expressing retroviral vector. *Journal of clinical investigation*, *118*(4), 1502–1510.

Zheng, Y. & Rudensky, A. Y. (2007). Foxp3 in control of the regulatory T cell lineage. *Nature immunology*, *8*(5), 457–462.

Zheng-Bradley, X., Rung, J., Parkinson, H. & Brazma, A. (2010). Large scale comparison of global gene expression patterns in human and mouse. *Genome biology*, *11*, R124.

Zhu, Y., Dai, J., Fuerst, P. G. & Voytas, D. F. (2003). Controlling integration specificity of a yeast retrotransposon. *Proceedings of the national academy of sciences of the united states of america*, *100*(10), 5891–5895.

Publications

Young, G. R., Ploquin, M. J.-Y., Eksmond, U., Wadwa, M., Stoye, J. P. & Kassiotis, G. (2012). Negative selection by an endogenous retrovirus promotes a higher-avidity $CD4^+$ T cell response to retroviral infection. *PLoS Pathogens*, *8* (5), e1002709.

Young, G. R., Eksmond, U., Salcedo, R., Alexopoulou, L., Stoye, J. P. & Kassiotis, G. (2012). Resurrection of endogenous retroviruses in antibody-deficient mice. *Nature*, in press.

OPEN ACCESS Freely available online PLoS PATHOGENS

Negative Selection by an Endogenous Retrovirus Promotes a Higher-Avidity CD4$^+$ T Cell Response to Retroviral Infection

George R. Young[1], Mickaël J.-Y. Ploquin[1¤a], Urszula Eksmond[1], Munisch Wadwa[1¤b], Jonathan P. Stoye[2], George Kassiotis[1]*

1 Division of Immunoregulation, MRC National Institute for Medical Research, London, United Kingdom, **2** Division of Virology, MRC National Institute for Medical Research, London, United Kingdom

Abstract

Effective T cell responses can decisively influence the outcome of retroviral infection. However, what constitutes protective T cell responses or determines the ability of the host to mount such responses is incompletely understood. Here we studied the requirements for development and induction of CD4$^+$ T cells that were essential for immunity to Friend virus (FV) infection of mice, according to their TCR avidity for an FV-derived epitope. We showed that a self peptide, encoded by an endogenous retrovirus, negatively selected a significant fraction of polyclonal FV-specific CD4$^+$ T cells and diminished the response to FV infection. Surprisingly, however, CD4$^+$ T cell-mediated antiviral activity was fully preserved. Detailed repertoire analysis revealed that clones with low avidity for FV-derived peptides were more cross-reactive with self peptides and were consequently preferentially deleted. Negative selection of low-avidity FV-reactive CD4$^+$ T cells was responsible for the dominance of high-avidity clones in the response to FV infection, suggesting that protection against the primary infecting virus was mediated exclusively by high-avidity CD4$^+$ T cells. Thus, although negative selection reduced the size and cross-reactivity of the available FV-reactive naïve CD4$^+$ T cell repertoire, it increased the overall avidity of the repertoire that responded to infection. These findings demonstrate that self proteins expressed by replication-defective endogenous retroviruses can heavily influence the formation of the TCR repertoire reactive with exogenous retroviruses and determine the avidity of the response to retroviral infection. Given the overabundance of endogenous retroviruses in the human genome, these findings also suggest that endogenous retroviral proteins, presented by products of highly polymorphic *HLA* alleles, may shape the human TCR repertoire that reacts with exogenous retroviruses or other infecting pathogens, leading to interindividual heterogeneity.

Citation: Young GR, Ploquin MJ-Y, Eksmond U, Wadwa M, Stoye JP, et al. (2012) Negative Selection by an Endogenous Retrovirus Promotes a Higher-Avidity CD4$^+$ T Cell Response to Retroviral Infection. PLoS Pathog 8(5): e1002709. doi:10.1371/journal.ppat.1002709

Editor: Susan R. Ross, University of Pennsylvania School of Medicine, United States of America

Received January 3, 2012; **Accepted** April 4, 2012; **Published** May 10, 2012

Funding: This work was supported by the UK Medical Research Council (U117581330). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: gkassio@nimr.mrc.ac.uk

¤a Current address: Institut Pasteur, Unité de Régulation des Infections Rétrovirales, Paris, France.
¤b Current address: Institute of Medical Microbiology, University Hospital, University Duisburg-Essen, Essen, Germany.

Introduction

Adaptive immunity to viral infection relies on appropriate activation of T cells by complexes of viral peptides with MHC molecules. The host *MHC* haplotype, the nature of the viral peptide targeted and the T cell receptor (TCR) repertoire of responding T cells are three interconnected parameters that play a decisive role in the outcome of infection. Indeed, the *MHC* is the predominant genetic factor affecting susceptibility to many infectious diseases [1–4]. For example, the *HLA* locus shows the strongest genetic association with control of HIV infection, with certain *HLA* alleles having been consistently found to confer a protective advantage [3,5,6]. Although the precise underlying mechanism is not completely understood, T cell responses restricted by protective *HLA/MHC* alleles often comprise narrower TCR repertoires, dominated by public TCR sequences, and exhibit higher magnitude, avidity or depth, and thus greater contribution to HIV or SIV control, than those restricted by non-protective *HLA/MHC* alleles [7–9].

Bias in the use of certain TCRVα (Vα) or TCRVβ (Vβ) chains has been frequently observed in the T cell response to several antigenic epitopes, and public T cell responses with identical or similar TCRs have been found to dominate the response of different individuals to a given epitope. Skewed TCR usage has often correlated with higher functional avidity to a given antigenic epitope, and, in diverse systems, also translated into more efficient protection against infection [10–12]. Despite the potential importance in cellular immunity to infection, however, the mechanisms by which TCR biases (and particularly high-avidity T cell responses to viral infections) are generated and maintained remains incompletely understood. The mechanisms leading to bias in the T cell response will vary considerably depending on the antigenic peptide and MHC combination. TCR repertoire bias can be generated during thymic selection, leaving only certain Vα

Author Summary

Our immune systems defend against viral infection. However, the immune response to a virus often differs substantially between individuals, as does the outcome of infection. The antiviral immune response relies on recognition of viral proteins by T lymphocytes using T cell antigen receptors (TCRs). TCRs are created randomly in an individual and each T lymphocyte has a different TCR. T lymphocytes with TCRs that recognize our own (self) proteins are removed during development, to prevent autoimmunity. Our cells can also make proteins that belong to endogenous retroviruses (ERVs), relics of ancestral retroviral infection that accumulated during evolution to constitute a large proportion of our genomes. The impact of ERVs on lymphocyte development and the subsequent influence on antiviral immunity is incompletely understood. Here, we use a mouse model to investigate this link and show that the T lymphocyte response to exogenous retrovirus infection is heavily influenced by an ERV. Surprisingly, we find that ERVs do not always have a negative impact on immunity, and in our model they improve the sensitivity with which T lymphocytes react to retroviral infection. This work may thus provide a basis for the understanding of the heterogeneity in immunity to retroviral infections in genetically diverse individuals.

or Vβ chains able to respond to a given antigen [10]. It can also be generated at the initiation of the immune response, where clones using particular Vα or Vβ chains may have a recruitment or proliferative advantage and can quickly dominate the response [10]. Lastly, bias can also be generated during chronic viral infection either due to preferential maintenance of certain T cell clones or differential margins for cross-reactivity with viral escape mutations [10] or by prior or concurrent infection with heterologous viruses, sharing cross-reactive epitopes [13].

We have previously described the TCRβ-transgenic strain EF4.1, which generates increased frequencies of $CD4^+$ T cells reactive with the H2-A^b-restricted $env_{122\text{-}141}$ epitope of Friend murine leukemia virus (F-MLV) [14]. Virus-specific EF4.1 $CD4^+$ T cells show bias in the use of endogenous Vα2 chains in their response to infection with Friend virus (FV), a retroviral complex of F-MLV and spleen focus-forming virus (SFFV) [14,15]. Use of Vα2 chains by virus-specific $CD4^+$ T cells creates higher avidity for the same epitope than use of other Vα chains, and although they represent a minority in the naïve repertoire, high-avidity Vα2 T cells become the dominant subset at the peak of the response [15]. Here we have examined the potential mechanisms underlying the formation of TCR repertoire diversity in this system, which might be responsible for the high-avidity response to FV infection. We have found that a thymic selection event was necessary for the dominance of Vα2 virus-specific $CD4^+$ T cells during the response to FV infection. Selection of virus-specific $CD4^+$ T cells was mediated by a self peptide encoded by an endogenous retrovirus with substantial similarity to F-MLV. Unexpectedly, despite deleting a sizeable fraction of virus-specific $CD4^+$ T cells, negative selection by this endogenous retrovirus was necessary for a predominantly high-avidity response to FV infection.

Results

Higher functional avidity of Vα2 F-MLV $env_{122\text{-}141}$-specific $CD4^+$ T cells

On average, 4% of EF4.1 $CD4^+$ T cells in virus-naïve mice react with the $env_{122\text{-}141}$ peptide, of which approximately 25% stain positive with the anti-Vα2 monoclonal antibody B20.1 [14,15]. Vα2 env-specific $CD4^+$ T cells were previously [14,15] found to be >30-fold more sensitive than non-Vα2 T cells to stimulation with a 20-mer $env_{122\text{-}141}$ peptide spanning the core $env_{128\text{-}134}$ epitope [16]. This higher avidity of Vα2 $CD4^+$ T cells was not due to recognition of the core epitope-flanking residues by this family of Vα chains, as has been described for other TCR – epitope combinations [17], since it was maintained against a series of N-terminal truncated peptide epitopes (Figure S1A). Thus, Vα2 $CD4^+$ T cells would recognize with higher avidity all the nested peptides of variable lengths likely to be generated during *in vivo* processing of env [18].

To examine whether the polyclonal Vα2 $CD4^+$ T cell population displayed higher affinity for F-MLV env-derived epitopes even at the clonal level, we generated hybridoma cell lines from primary EF4.1 $CD4^+$ T cells stimulated *in vitro* with either a low (10^{-7} M) or a high (10^{-5} M) peptide dose. In agreement with our previous findings [14,15], 71% (20/28) and 30% (9/30) of hybridoma cell lines derived from $CD4^+$ T cells stimulated with the low or high peptide dose, respectively, were $V\alpha2^+$. Similarly to primary EF4.1 $CD4^+$ T cells, randomly selected Vα2 T cell hybridomas were more sensitive to stimulation with all the peptides tested than non-Vα2 ones, irrespective of whether a high or low peptide dose was used for their generation (Figure S1B). Thus, the higher avidity of Vα2 $CD4^+$ T cells for F-MLV env-derived epitopes was also observed at the level of individual clones.

To assess whether low-avidity F-MLV env-reactive $CD4^+$ T cells were characterized by expression of any particular family of endogenous Vα chains, we screened $env_{122\text{-}141}$-specific non-Vα2 $CD4^+$ T cells for expression of *Trav* transcripts encoding different Vα families. Although this analysis indicated enrichment for *Trav9* expression (encoding Vα3), only a small percentage of $env_{122\text{-}141}$-reactive non-Vα2 $CD4^+$ T cells stained positive with the anti-Vα3.2 monoclonal antibody RR3-16 (*unpublished data*), and only 2 out of 4 F-MLV env-reactive non-Vα2 T cell hybridomas were positive for Vα3.2 (Table S1). However, Vα3.2 is used preferentially in $CD8^+$ T cells in B6 mice, whereas the other three of the four expressed Vα3 family members are preferentially expressed in $CD4^+$ T cells [19]. It was therefore possible that $env_{122\text{-}141}$-reactive non-Vα2 $CD4^+$ T cells that did not stain positive with the RR3-16 antibody were also expressing Vα3. Indeed, cloning and sequencing of expressed endogenous *Trav* genes from theses hybridomas revealed that they were all members of the *Trav9* family (Table S1). Thus, similarly to selective usage of Vα2 chains in high-avidity cells, low-avidity F-MLV $env_{122\text{-}141}$-reactive $CD4^+$ T cells selectively used Vα3 chains. However, in the absence of a Vα3-specific antibody that can detect all Vα3 family members, these cells were referred to here as non-Vα2 cells.

Lastly, we tested whether biased use of Vα2 chains also characterized the response of non-transgenic $CD4^+$ T cells to F-MLV env. $CD4^+$ T cells from wild-type (wt) C57BL/6 (B6) mice 7 days post FV infection were stained with an $env_{123\text{-}141}$-presenting tetramer (A^b-$env_{123\text{-}141}$). In comparison with a control tetramer, staining with A^b-$env_{123\text{-}141}$ tetramer identified a measurable population of $env_{122\text{-}141}$-specific $CD4^+$ T cells in all infected mice (Figure 1A), in agreement with published data [20,21]. FV infection had no impact on the frequency of Vα2 cells in naïve ($CD44^{lo}$) and total memory ($CD44^{hi}$) $CD4^+$ T cells (15% and 12%, respectively), with minimal variation between individual mice (Figure 1B). In contrast, the frequency of Vα2 cells in A^b-$env_{123\text{-}141}$ $tetramer^+$ $CD4^+$ T cells varied considerably between 4% and 23%. These results revealed substantial deviation in Vα2 usage in A^b-

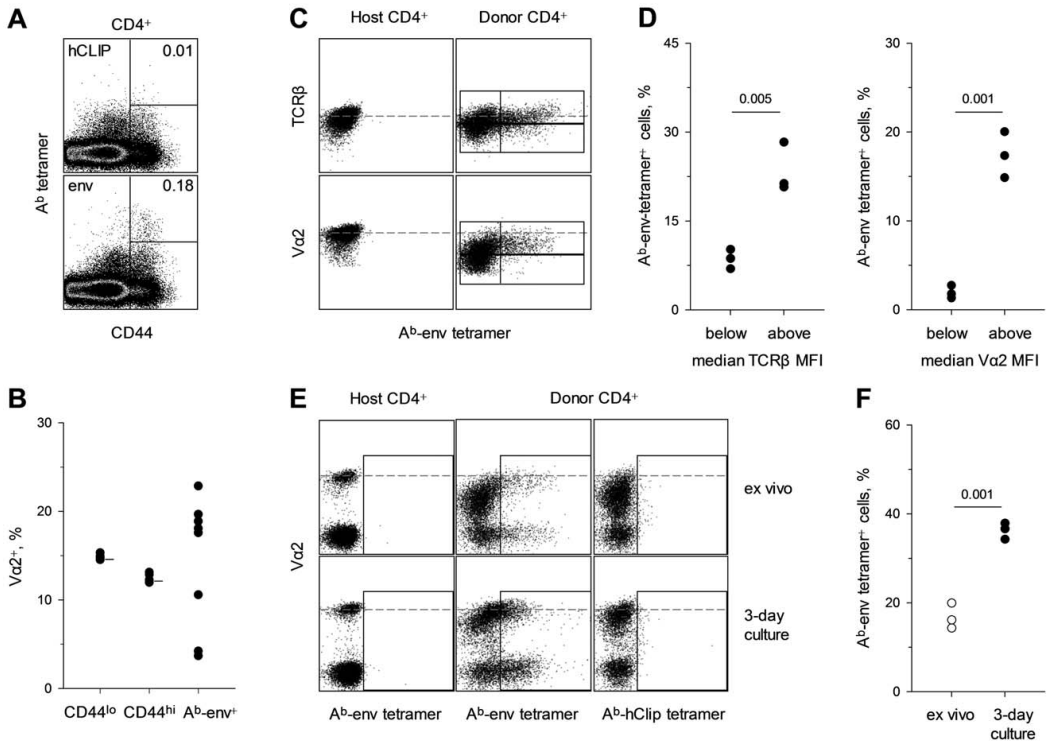


Figure 1. Detection of env-specific CD4$^+$ T cells by A^b-env$_{123\text{-}141}$ tetramer eclipsed by antigen-induced TCR downregulation. (A) A^b-hCLIP (control) or A^b-env$_{123\text{-}141}$ tetramer staining in total CD4$^+$ T cells isolated from the spleen of wild-type B6 mice 7 days post FV infection. Plots are representative of 7 mice. (B) Frequency of Vα2 cells in either bulk naïve (CD44^{lo}), bulk memory (CD44^{hi}) or A^b- env$_{123\text{-}141}$ tetramer$^+$ CD4$^+$ T cells from the same FV infected mice. Horizontal short lines in naïve and memory subsets denote the mean frequency of Vα2 cells in the same populations from uninfected mice. Each symbol represents an individual mouse. (C–F) CD45.1$^+$ EF4.1 CD4$^+$ T cells were adoptively transferred into wild-type B6 recipients that were infected with FV the same day. (C) A^b-env$_{123\text{-}141}$ tetramer staining in host (CD45.1$^-$) or donor (CD45.1$^+$) CD4$^+$ T cells according to TCRα or TCRβ staining. Gates in donor CD4$^+$ T cells are set around the median TCRα and TCRβ staining, respectively. (D) Percentage of A^b-env$_{123\text{-}141}$ tetramer$^+$ cells in donor CD4$^+$ T cells with TCRβ (*left*) or TCRα (*right*) staining below or above the median. (E) A^b-hCLIP or A^b-env$_{123\text{-}141}$ tetramer staining in host or donor CD4$^+$ T cells from the same recipients assessed directly *ex vivo* (*top*) or following 3-day *in vitro* culture in the absence of antigenic stimulation (*bottom*). (F) Percentage of A^b-env$_{123\text{-}141}$ tetramer$^+$ cells in donor CD4$^+$ T cells before and after *in vitro* culture. In (D) and (F) each symbol represents an individual mouse from one of two experiments.
doi:10.1371/journal.ppat.1002709.g001

env$_{123\text{-}141}$ tetramer$^+$ CD4$^+$ T cells from the same usage in total CD4$^+$ T cells, but also indicated substantial heterogeneity. However, this particular tetramer is known to bind only some env$_{122\text{-}141}$-specific T cell clones but not others [14,22]. Furthermore, at the peak of their response, env$_{122\text{-}141}$-specific CD4$^+$ T cell reversibly downregulate up to 70% of their surface TCR [15,23], which could prevent tetramer binding. Indeed, combining adoptive transfer of EF4.1 CD4$^+$ T cells and tetramer staining revealed that A^b-env$_{123\text{-}141}$ tetramer staining was restricted to env$_{122\text{-}141}$-reactive CD4$^+$ T cells with above-average TCR levels, independently of Vα usage, and TCR re-expression improved tetramer staining (Figure 1C–F). Thus, detection of env$_{122\text{-}141}$-reactive CD4$^+$ T cells by A^b-env$_{123\text{-}141}$ tetramer staining was eclipsed by *in vivo* antigen-induced TCR downregulation. Collectively, these results both validated and necessitated the use of env$_{122\text{-}141}$-reactive EF4.1 CD4$^+$ T cells that can be indelibly identified, independently of A^b-env$_{123\text{-}141}$ tetramer binding, to study the requirements for induction of a high-avidity CD4$^+$ T cell response to F-MLV env.

Deletion of F-MLV env$_{122\text{-}141}$-specific CD4$^+$ T cells by *Emv2*-encoded env

F-MLV env$_{122\text{-}141}$-reactive CD4$^+$ T cells in EF4.1 mice have a naïve phenotype [14], and it was therefore likely that the F-MLV env$_{122\text{-}141}$-reactive TCR repertoire and associated avidity differences were the result of thymic selection events. We searched the mouse proteome for the presence of self-derived epitopes with homology to F-MLV env$_{122\text{-}141}$. This approach identified the single-copy endogenous ecotropic MLV at the *Emv2* locus [24]. *Emv2* shares 80% homology with F-MLV at the DNA sequence level, and although it represents a full-length provirus, it is unable to produce infectious particles due to a single inactivating point-mutation in the *pol* gene [24]. Nevertheless, *Emv2* has full potential for *env* expression, and, importantly, the env$_{122\text{-}141}$ epitope is almost identical between *Emv2* and F-MLV, with the exception of a Y instead of an L at position 128 (Figure S2A). For this reason, *Emv2* and F-MLV env-derived epitopes were referred to here as env$_{122\text{-}141}$Y and env$_{122\text{-}141}$L, respectively. Position 128, together with 129 and 133, have been previously mapped as important

contact residues for the SB14-31 TCR (Figure S2B), which was the donor of the TCRβ chain transgene used in EF4.1 mice [16].

We next investigated whether or not *Emv2* could be involved in T cell selection. *In vitro* stimulation with the $env_{124\text{-}138}$Y epitope activated a fraction of EF4.1 $CD4^+$ T cells, which was however smaller than the fraction activated by the $env_{124\text{-}138}$L epitope (Figure 2A). As EF4.1 mice generate a polyclonal TCR repertoire, it was unclear whether the same $CD4^+$ T cells could respond to both epitopes. However, analysis of $env_{124\text{-}138}$L-reactive T cell hybridomas revealed the same TCR could be activated by both $env_{124\text{-}138}$L and $env_{124\text{-}138}$Y epitopes, albeit less potently by the latter peptide (Figure 2B). Thus, F-MLV $env_{124\text{-}138}$-reactive TCRs have the potential to recognize *Emv2* env. This analysis also revealed that *Emv2* was not causing complete tolerance of either $env_{124\text{-}138}$L or $env_{124\text{-}138}$Y epitopes. We then confirmed that *Emv2* was expressed in primary and secondary lymphoid organs. Using primers specific to the spliced *env* mRNA that could distinguish between genuine transcripts and contaminating genomic DNA, *Emv2* was found to be expressed at low levels in the majority of mice tested (Figure 2C). This low level of expression was further confirmed by comparison with a newly-generated B6 congenic strain lacking *Emv2* (Figure S3). To evaluate the extent of *Emv2*-mediated deletion of env-reactive $CD4^+$ T cells more directly, we generated B6-$Emv2^{-/-}$ EF4.1 mice and compared them with *Emv2*-expressing EF4.1 mice. *Emv2*-deficient EF4.1 mice contained a significantly higher frequency of $env_{124\text{-}138}$L-reactive $CD4^+$ T cells than *Emv2*-sufficient EF4.1 mice, with *Emv2*, when present, being responsible for the deletion of approximately 35% of these cells (Figure 2D). Thus, albeit low, expression of *Emv2* in B6 mice significantly impacted on the frequency of $env_{124\text{-}138}$L-reactive EF4.1 $CD4^+$ T cells.

The *Emv2*-selected $CD4^+$ T cell repertoire retains full antiviral activity

Emv2-mediated deletion of a proportion of $env_{124\text{-}138}$L-reactive EF4.1 $CD4^+$ T cells suggested that *Emv2* may impinge on resistance to FV infection. We therefore examined the possible effect of *Emv2* expression on FV control. Firstly, we infected non-transgenic B6 and *Emv2*-deficient B6 mice and measured the levels of infected cells in the spleen. B6 mice are relatively resistant to FV infection due to genetic resistance at the *Fv2* locus and due to mounting a strong adaptive immune response [1,25], resulting in control of the infection by the third week. Percentages of FV-infected (glyco-Gag^+) erythroid precursor (nucleated $Ter119^+$) cells were significantly lower in B6-$Emv2^{-/-}$ mice than in wt counterparts at day 7 of infection (Figure 3A). Nevertheless, wt B6 mice effectively controlled FV infection to levels comparable with those in B6-$Emv2^{-/-}$ mice by the second week of infection (Figure 3A). Thus, *Emv2* deficiency did not extensively increase the natural resistance of B6 mice to FV infection.

The modest increase in resistance to FV infection in B6-$Emv2^{-/-}$ mice suggested that this low *Emv2* expression was immunologically relevant, but did not indicate if any arm of the adaptive immune response was affected. We thus measured the FV-specific $CD4^+$ T cell, $CD8^+$ T cell and antibody responses in these mice. In contrast to the MHC class II-restricted $env_{122\text{-}141}$L epitope, the FV-derived MHC class I-restricted epitopes that have been described do not share extensive homology or cross-reactivity with those derived from *Emv2* [26–28]. We examined the $CD8^+$ T cell response to FV by measuring numbers of activated $CD44^{hi}CD43^+CD8^+$ T cells, irrespective of antigen specificity, in the spleens of B6 and B6-$Emv2^{-/-}$ mice 7 days post FV infection (Figure S4A). The two types of host showed comparable expansion of $CD44^{hi}CD43^+CD8^+$ T cells, suggesting they mounted a $CD8^+$ T cell response of similar magnitude (Figure S4A). We further measured the $CD8^+$ T cell response to the immunodominant D^b-restricted epitope from the leader sequence $gPr80gag_{85\text{-}93}$ encoded by the F-MuLV *gag* gene [28]. $CD8^+$ T cells specific to the $gPr80gag_{85\text{-}93}$ epitope display strong bias for the use of Vα3.2 and Vβ5.2 in combination, which allows their identification by flow cytometry [29]. Expectedly, FV infection led to an increase in the percentage of $V\alpha3.2^+V\beta5.2^+$ cells in antigen-experienced ($CD44^{hi}$), but not naïve ($CD44^{lo}$) $CD8^+$ T cells (Figure S4B). However, this expansion of $V\alpha3.2^+V\beta5.2^+CD44^{hi}CD8^+$ T cells was comparable in B6 and B6-$Emv2^{-/-}$ mice 7 days post FV infection (Figure S4B).

We next examined the FV-specific antibody response of B6 and B6-$Emv2^{-/-}$ hosts. As FV-neutralizing antibodies are not readily

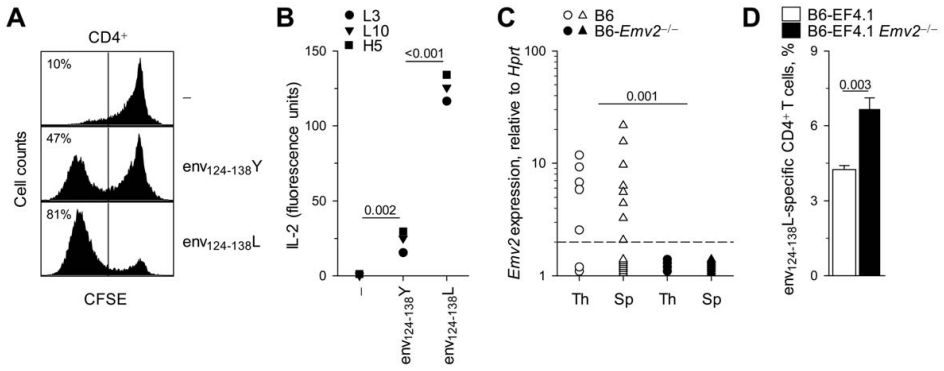


Figure 2. *Emv2* selects against a fraction of $env_{124\text{-}138}$L-specific $CD4^+$ T cells. (A) Dilution of prior CFSE label by primary EF4.1 $CD4^+$ T cells incubated for three days *in vitro* in the absence of peptide stimulation (-) or in the presence of 10^{-5} M $env_{124\text{-}138}$L or $env_{124\text{-}138}$Y peptides. Numbers within the plots denote the percentage of $CFSE^-$ cells and are representative of 4 mice per condition. (B) IL-2 production by three $env_{124\text{-}138}$L-specific hybridoma T cell lines in response to *in vitro* stimulation with the same peptides at 5×10^{-6} M. (C) *Emv2* transcription, relative to *Hprt* transcription, in thymi (Th) and spleens (Sp) of wild-type B6 or *Emv2*-deficient B6 mice (B6-$Emv2^{-/-}$). Each symbol is an individual mouse. The dashed line represents the limit of detection. (D) Frequency of $env_{124\text{-}138}$L-specific cells in primary $CD4^+$ T cells from B6 (B6-EF4.1) or *Emv2*-deficient B6 (B6-EF4.1 $Emv2^{-/-}$) EF4.1 mice. Data are the means ± SEM ($n=9$) from 3 experiments.
doi:10.1371/journal.ppat.1002709.g002

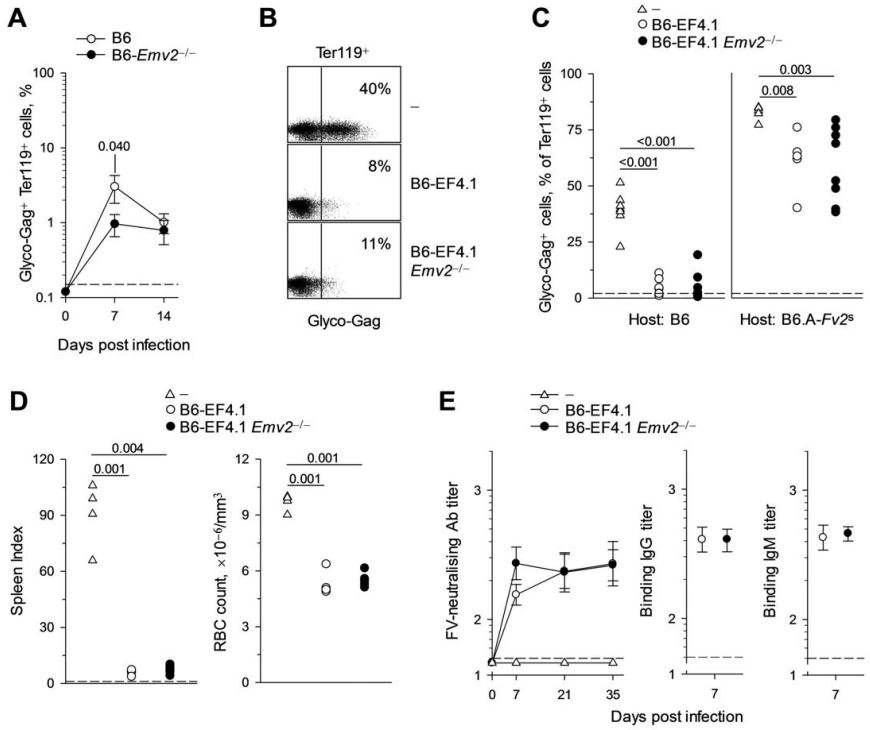


Figure 3. *Emv2*-selected CD4$^+$ T cells retain full antiviral activity. (A) Mean frequency (± SEM, $n = 8$–19) of FV-infected (glyco-Gag$^+$) Ter119$^+$ cells in the spleens of FV-infected B6 or *Emv2*-deficient B6 mice (B6-*Emv2*$^{-/-}$). (B–C) CD4$^+$ T cells isolated from either B6 (B6-EF4.1) or *Emv2*-deficient B6 (B6-EF4.1 *Emv2*$^{-/-}$) EF4.1 mice were adoptively transferred into B6 or B6.A-*Fv2*^{s} recipients that were infected with FV the same day and analyzed 7 days later. (B) Flow cytometric example of FV-infected Ter119$^+$ cells from B6 recipients and (C) frequency of FV-infected cells in Ter119$^+$ cells from the spleens of B6 or B6.A-*Fv2*^{s} recipients of CD4$^+$ T cells. Control B6 and B6.A-*Fv2*^{s} mice that received no CD4$^+$ T cells (-) are also included. Each symbol is an individual mouse. (D) Spleen index (*left*) and RBC count (*right*) of B6-*Rag1*$^{-/-}$*Fv2*^{s} mice that were infected with FV and either received the same day CD4$^+$ T cells isolated from either B6 (B6-EF4.1) or *Emv2*-deficient B6 (B6-EF4.1 *Emv2*$^{-/-}$) EF4.1 mice or no cells (-). Each symbol is an individual mouse analyzed 3 weeks post infection. (E) Titers of FV-neutralizing antibodies during the course of FV infection (*left*) and titers of F-MLV-infected cell-binding IgG (*middle*) and IgM (*right*) 7 days post FV infection, in the sera of B6-*Tcra*$^{-/-}$ mice that either received CD4$^+$ T cells isolated from either B6 (B6-EF4.1) or *Emv2*-deficient B6 (B6-EF4.1 *Emv2*$^{-/-}$) EF4.1 mice or no cells (-) the day of the infection. Dashed lines represent the limit of detection. Data are the means ± SEM ($n = 11$–12) from 2 experiments.
doi:10.1371/journal.ppat.1002709.g003

detected in B6 mice on day 7 post FV infection [25], we measured titers of antibodies that were able to bind F-MLV-infected cells (Figure S4C). On day 7 post FV infection all mice produced both IgG and IgM F-MLV-infected cell-binding antibodies that could be measured by flow cytometry (*unpublished data*). However, titers of these antibodies were low at this early time-point and in a proportion of FV-infected mice they were below 50, a value that we set as the detection limit (Figure S4C). Importantly, serum titers of both IgG and IgM F-MLV-infected cell-binding antibodies were similar between B6 and B6-*Emv2*$^{-/-}$ hosts (Figure S4C).

Lastly, the frequency of A^b-env$_{123\text{-}141}$ tetramer$^+$ CD4$^+$ T cells as well as the frequency of Vα2 cells within this population was highly variable between individual mice and as a result not statistically different between groups of B6 and B6-*Emv2*$^{-/-}$ hosts on day 7 post FV infection (Figure S4D). However, staining with the A^b-env$_{123\text{-}141}$ tetramer may have underestimated the frequency of env$_{122\text{-}141}$L-reactive CD4$^+$ T cells (Figure 1) and it was also possible that env$_{122\text{-}141}$L-reactive CD4$^+$ T cells selected in the presence or absence of *Emv2* expressed TCRs with distinct A^b-env$_{123\text{-}141}$ tetramer-binding properties. Furthermore, virus-specific CD4$^+$ T cells can mediate both direct and indirect protection against FV infection [15,23,30], and env$_{122\text{-}141}$-specific CD4$^+$ T cells have been shown to mediate direct cytotoxic activity [31]. It was thus uncertain whether weakened immunity in *Emv2*-expressing mice was directly linked to a potentially less effective CD4$^+$ T cell response. We therefore examined the effect of *Emv2* expression on the FV-specific CD4$^+$ T cell response functionally and directly. To this end, equal numbers of *Emv2*-selected or -nonselected EF4.1 CD4$^+$ T cells were transferred into the same type of host. This approach ensured that only donor EF4.1 CD4$^+$ T cells differed with respect to exposure to *Emv2*. Surprisingly, the two types of donor CD4$^+$ T cells provided comparable and almost complete protection of wild-type B6 hosts, at the peak of FV replication on day 7 post infection (Figure 3B, C). To rule out that differences in antiviral activity between the two types of donor CD4$^+$ T cells were not missed because this activity was already maximal, we have additionally used B6.A-*Fv2*^{s} hosts, expressing the susceptibility allele at the *Fv2* locus, which confers susceptibility to FV infection by enhancing proliferation of infected erythroid

precursors [25]. The two types of donor CD4$^+$ T cells provided significant, suboptimal and, importantly, comparable protection in B6.A-*Fv2*^{s} hosts, at the peak of FV replication and expansion of infected erythroid precursors on day 7 post infection in this strain (Figure 3C). Thus, *Emv2*-mediated selection did not impair the antiviral activity of CD4$^+$ T cells exerted in wt hosts. To further examine direct CD4$^+$ T cell-mediated protection we transferred equal numbers of *Emv2*-selected or -nonselected EF4.1 CD4$^+$ T cells into T and B cell-deficient *Rag1*$^{-/-}$ *Fv2*^{s} hosts. Both types of donor CD4$^+$ T cells were similarly protective against severe FV-induced splenomegaly (Figure 3D) that otherwise develops in these hosts [15]. In addition, the two types of donor CD4$^+$ T cells caused comparable levels of anemia in these T and B cell -deficient hosts (Figure 3D), which results from bone marrow pathology [14]. Lastly, FV-neutralizing antibodies were similarly and efficiently induced in T cell-deficient *Tcra*$^{-/-}$ hosts by transfer of either type of donor CD4$^+$ T cells, although they were slightly, but not significantly higher in hosts of *Emv2*-nonselected CD4$^+$ T cells on day 7 post infection (Figure 3E). Nevertheless, at this time-point, the two types of donor CD4$^+$ T cells induced comparable titers of IgG or IgM antibodies that were able to bind F-MLV-infected cells, which also included antibodies potentially mediating antibody-dependent cell-mediated cytotoxicity (Figure 3E). Collectively, these results demonstrated that despite selecting against a significant fraction of env$_{124\text{-}138}$L-reactive CD4$^+$ T cells, *Emv2* expression did not compromise CD4$^+$ T cell function against FV infection.

Selection by *Emv2* promotes a higher-avidity response to F-MLV

Retention of full CD4$^+$ T cell-mediated antiviral activity, despite deletion of over a third of env$_{124\text{-}138}$L-reactive CD4$^+$ T cells in *Emv2*-expressing mice, suggested that the deleted cells were not contributing to immunity against FV infection. We therefore assessed the impact of *Emv2* expression on both the magnitude and composition of the CD4$^+$ T cell response to FV. Equal numbers of EF4.1 CD4$^+$ T cells from either *Emv2*-sufficient or -deficient donor B6 mice, positive for CD45.2 (encoded by the *Ptprc*^{2} allele), were adoptively transferred into *Ptprc*$^{1/2}$ syngeneic B6 recipients that were positive for both CD45.1 and CD45.2. Recipient mice were infected with FV on the day of T cell transfer and FV-responding donor CD4$^+$ T cells were identified as CD44^{hi}CD45.2$^+$CD45.1$^-$ cells (Figure S5). Consistent with increased precursor frequency in *Emv2*-deficient donor mice, significantly higher numbers of total responding CD4$^+$ T cells could be recovered at the peak of the response from secondary recipients that received *Emv2*-nonselected than those that received *Emv2*-selected donor CD4$^+$ T cells (Figure 4A). Notably, the two types of donor CD4$^+$ T cells generated comparable numbers of high-avidity responding CD4$^+$ T cells, and the numerical increase in total numbers of responding CD4$^+$ T cells from *Emv2*-nonselected donors was due to significantly higher expansion of low-avidity non-Vα2 responding CD4$^+$ T cells from these donors in comparison with the expansion of non-Vα2 responding CD4$^+$ T cells from *Emv2*-selected donors (Figure 4A). As a result, peak expansion of *Emv2*-selected CD4$^+$ T cells was dominated by high-avidity Vα2 CD4$^+$ T cells, whereas that of *Emv2*-nonselected CD4$^+$ T cells was dominated by low-avidity non-Vα2 CD4$^+$ T cells (Figure 4B, C). Thus, *Emv2* expression converted a predominantly low-avidity response to FV to a predominantly high-avidity response.

Emv2-encoded env preferentially deletes non-Vα2 CD4$^+$ T cells

The shift from a predominantly Vα2 response of *Emv2*-selected CD4$^+$ T cells to a predominantly non-Vα2 response of *Emv2*-nonselected CD4$^+$ T cells could be the result of *Emv2*-induced modulation of either the relative frequency in the naïve repertoire of the two subsets of env$_{124\text{-}138}$L-reactive CD4$^+$ T cells, or their relative avidity for env$_{124\text{-}138}$L (or both). We first measured the overall functional avidity to env$_{124\text{-}138}$L of EF4.1 CD4$^+$ T cells selected with or without *Emv2* as an indicator of potential avidity repertoire changes. Surprisingly, although the presence of *Emv2* reduced the precursor frequency of env$_{124\text{-}138}$L-reactive CD4$^+$ T cells, it had no effect on the avidity with which they responded to env$_{124\text{-}138}$L stimulation (Figure 5A). This result suggested that *Emv2*-mediated selection either affected high- and low-avidity cells similarly, or that potential loss of higher-avidity T cells was compensated by an increase in average avidity of the remaining T cells. To examine whether *Emv2* preferentially selected against high-avidity env$_{124\text{-}138}$L-reactive cells, we measured their frequency separately in either Vα2 or non-Vα2 CD4$^+$ T cells from EF4.1 mice. Notably, *Emv2* expression significantly reduced the frequency of non-Vα2, but not Vα2 env$_{124\text{-}138}$L-reactive cells in EF4.1 CD4$^+$ T cells (Figure 5B), indicating that it only selected against

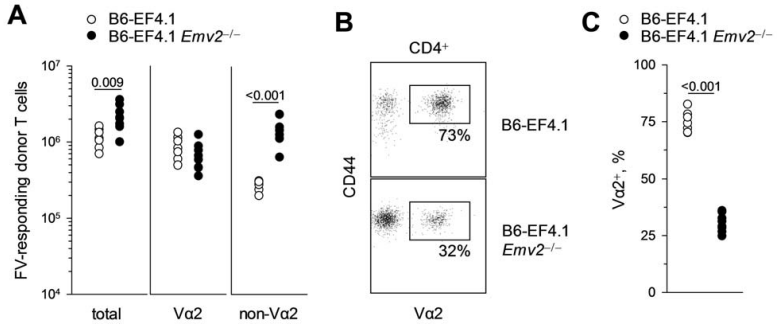


Figure 4. *Emv2*-selected CD4$^+$ T cells mount a predominantly high-avidity response. (A–C) CD45.2$^+$ (*Ptprc*$^{2/2}$) CD4$^+$ T cells isolated from either B6 (B6-EF4.1) or *Emv2*-deficient B6 (B6-EF4.1 *Emv2*$^{-/-}$) EF4.1 donor mice were adoptively transferred into *Ptprc*$^{1/2}$ B6 recipients that were infected with FV the same day and analyzed 7 days later. (A) Absolute number of total, Vα2 or non-Vα2 FV-responding (CD44^{hi}) donor (CD45.2$^+$CD45.1$^-$) CD4$^+$ T cells isolated from the spleens of recipient mice according to donor type. (B) Flow cytometric example and (C) frequency of high-avidity Vα2 cells in responding CD4$^+$ T cells according to donor type. In (A) and (C) each symbol is an individual mouse.
doi:10.1371/journal.ppat.1002709.g004

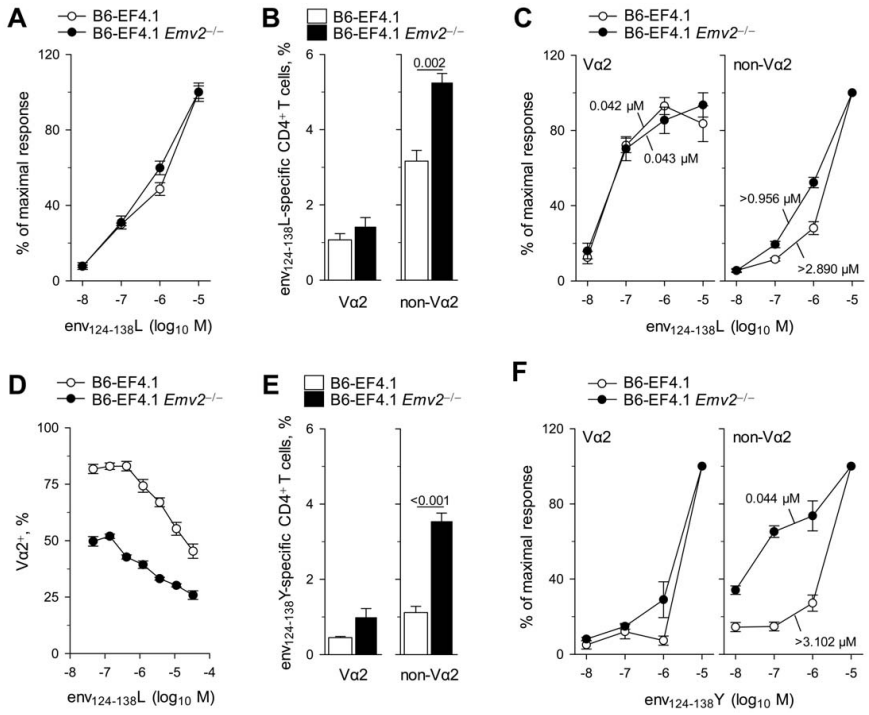


Figure 5. *Emv2* preferentially selects against non-Vα2 env-specific CD4$^+$ T cells. (A) Dose-response to env$_{124\text{-}138}$L stimulation of CD4$^+$ T cells isolated from either B6 (B6-EF4.1) or *Emv2*-deficient B6 (B6-EF4.1 *Emv2*$^{-/-}$) EF4.1 mice. (B) Frequency of env$_{124\text{-}138}$L-specific cells in Vα2 or non-Vα2 primary CD4$^+$ T cells from the same donors. (C) Functional avidity of *Emv2*-selected (B6-EF4.1) or -nonselected (B6-EF4.1 *Emv2*$^{-/-}$) EF4.1 CD4$^+$ T cells for env$_{124\text{-}138}$L. (D) Frequency of Vα2 cells in env$_{124\text{-}138}$L-specific CD4$^+$ T cells from the same donors as a function of peptide concentration. (E) Frequency of env$_{124\text{-}138}$Y-specific cells in Vα2 or non-Vα2 primary CD4$^+$ T cells from the same donors. (F) Functional avidity of *Emv2*-selected (B6-EF4.1) or -nonselected (B6-EF4.1 *Emv2*$^{-/-}$) EF4.1 CD4$^+$ T cells for env$_{124\text{-}138}$Y. Numbers in (C) and (F) represent the ED$_{50}$. Data in (A–F) are the means ± SEM ($n = 9$–12) of 18-hr stimulations from 3 experiments.
doi:10.1371/journal.ppat.1002709.g005

non-Vα2 CD4$^+$ T cells. Correspondingly, the avidity of Vα2 CD4$^+$ T cells to env$_{124\text{-}138}$L was not altered by *Emv2* expression, whereas the avidity of non-Vα2 CD4$^+$ T cells was 3-fold higher in the absence than in the presence of *Emv2* (Figure 5C). Nevertheless, non-Vα2 CD4$^+$ T cells from *Emv2*-deficient mice still displayed lower avidity than Vα2 CD4$^+$ T cells from either *Emv2*-deficient or -sufficient mice (Figure 5C).

To test whether *Emv2*-mediated changes in the frequency and avidity for env$_{124\text{-}138}$L of non-Vα2 CD4$^+$ T cells could account for the dominance of this subset in the *in vivo* response to FV of *Emv2*-nonselected CD4$^+$ T cells, we examined the *in vitro* response of *Emv2*-selected or -nonselected primary naïve EF4.1 CD4$^+$ T cells to env$_{124\text{-}138}$L stimulation. As a result of differences in initial frequency and functional avidity between virus-naïve Vα2 and non-Vα2 env$_{122\text{-}141}$-specific cells, the composition of the responding population varied according to the amount of env$_{122\text{-}141}$ presentation [14] and Vα2 T cells dominated the response at doses lower than 10^{-7} M (Figure 5D). Importantly, this percentage of Vα2 cells was significantly lower at all peptide doses in CD4$^+$ T cells selected in the absence than in the presence of *Emv2* (Figure 5D), demonstrating that selection by this single provirus heavily influenced the clonal composition of env$_{124\text{-}138}$L-reactive CD4$^+$ T cells, in favor of high-avidity cells. We further confirmed that this effect of *Emv2* expression of reducing the overall frequency of env$_{124\text{-}138}$L-reactive cells, but significantly increasing the percentage of high-avidity Vα2 cells in the env$_{124\text{-}138}$L-reactive population was already evident in CD4$^+$CD8$^-$ thymocytes (Figure S6), consistent with a thymic, rather than peripheral event.

Preferential deletion by *Emv2* of non-Vα2 CD4$^+$ T cells, which had low avidity for F-MLV env$_{124\text{-}138}$L raised the possibility that these cells may have been cross-reactive with *Emv2*-encoded env$_{124\text{-}138}$Y. Indeed, lack of *Emv2* expression in EF4.1 mice had a small, non-significant effect on env$_{124\text{-}138}$Y-reactive Vα2 CD4$^+$ T cells, but caused a significant 3.5-fold increase in the frequency of env$_{124\text{-}138}$Y-reactive non-Vα2 CD4$^+$ T cells (Figure 5E). Furthermore, Vα2 CD4$^+$ T cells from either *Emv2*-deficient or -sufficient EF4.1 mice could only react with env$_{124\text{-}138}$Y at the highest dose of 10^{-5} M, whereas non-Vα2 CD4$^+$ T cells from *Emv2*- deficient mice were markedly more sensitive to env$_{124\text{-}138}$Y than those from *Emv2*-sufficient mice (and as sensitive as Vα2 CD4$^+$ T cells to env$_{124\text{-}138}$L) (Figure 5F). Together, these findings indicated that *Emv2* expression was not affecting env$_{124\text{-}138}$L-reactive Vα2 CD4$^+$ T cells because they displayed low avidity for env$_{124\text{-}138}$Y, but was deleting a significant proportion of non-Vα2 CD4$^+$ T cells that could react with either env$_{124\text{-}138}$L or env$_{124\text{-}138}$Y.

Shaping of env-reactive CD4$^+$ T cell repertoire depth by *Emv2*

Although EF4.1 CD4$^+$ T cells selected by *Emv2* mounted high-avidity responses to the index env$_{124\text{-}138}$L sequence *in vitro*, and to FV infection *in vivo*, and retained full antiviral activity, counter-selection of env$_{124\text{-}138}$Y-reactive clones indicated that this repertoire would be less able to respond to viral escape mutations, and especially to an L128Y mutation. To extend these findings, we used another variant of env, which differed from F-MLV env in two of the three putative TCR-binding residues. This variant has Y and S in positions 128 and 129, respectively (referred to as env$_{124\text{-}138}$YS) and is a naturally-occurring functional form of ecotropic env, encoded by the *Fv4* locus in certain strains and species of mouse, other than the B6 strain [32,33]. Again, a very small fraction of EF4.1 Vα2 CD4$^+$ T cells could react to env$_{124\text{-}138}$YS, regardless of the presence or absence of *Emv2* (Figure 6A). In contrast, lack of *Emv2* led to a 7-fold increase in the frequency of env$_{124\text{-}138}$YS-reactive EF4.1 non-Vα2 CD4$^+$ T cells, which now made a sizable fraction (Figure 6A). Thus, non-Vα2 CD4$^+$ T cells from *Emv2*-deficient EF4.1 mice could react with the index sequence and the two env variants and with high avidity to env$_{124\text{-}138}$Y, suggesting that *Emv2*-mediated selection significantly reduced the ability of CD4$^+$ T cells, at the population level, to recognize these env variants.

This analysis of polyclonal cells from *Emv2*-deficient EF4.1 mice did not reveal whether the same T cell could react to all three env variants or if env$_{124\text{-}138}$L-, env$_{124\text{-}138}$Y- and env$_{124\text{-}138}$YS-reactive non-Vα2 CD4$^+$ T cells were distinct. We therefore tested the reactivity of hybridoma cell lines generated from env$_{124\text{-}138}$L-reactive EF4.1 CD4$^+$ T cells that developed either in the presence or the absence of *Emv2* expression to other env variants. Similarly to non-Vα2 CD4$^+$ T cell hybridomas from *Emv2*-sufficient donors, all 4 non-Vα2 CD4$^+$ T cell hybridomas tested from *Emv2*-deficient donors used members of the TCRVα3 family (encoded by the *Trav9* gene family; Table S2). Notably, neither *Emv2*-selected nor -nonselected env$_{124\text{-}138}$L-reactive non-Vα2 CD4$^+$ T cell hybridomas responded to env$_{124\text{-}138}$Y more potently than Vα2 CD4$^+$ T cell hybridomas from the same donor strain, and only 1 out of 4 had a measureable response to env$_{124\text{-}138}$YS (Figure 6B, C). These findings suggested that the env$_{124\text{-}138}$L-reactive non-Vα2 CD4$^+$ T cells that developed in *Emv2*-deficient EF4.1 mice were largely distinct from env$_{124\text{-}138}$Y- and env$_{124\text{-}138}$YS-reactive T cells in the same mice. They also indicated that env$_{124\text{-}138}$L-reactive non-Vα2 CD4$^+$ T cells were not inherently more cross-reactive than env$_{124\text{-}138}$L-reactive Vα2 CD4$^+$ T cells at the clonal level.

To gain a more detailed view of the depth, defined here as the ability to tolerate epitope mutations, of env$_{124\text{-}138}$L-reactive Vα2 or non-Vα2 TCRs, we screened *Emv2*-selected or -nonselected env$_{124\text{-}138}$L-reactive T cell hybridomas for reactivity against a library of env$_{126\text{-}138}$ peptide variants carrying all possible single mutations in each of the amino acid residues in positions 128, 129 and 133 (Figure S7). Amino acids that elicited at least 40% of the maximal response were listed in the order they were preferred by the individual TCRs (Figure 7). All Vα2 T cell hybridomas displayed strong preference for L at position 128 and also recognized similar amino acids with hydrophobic side chains, namely F, I, M and V, but not the less hydrophobic Y (Figure 7A). Vα2 T cell hybridomas also showed strong preference and specificity for the amino acid residues of the index sequence against which they were derived, T or highly similar S at position 129, and N at position 133 (Figure 7B, C). Overall, the depth of Vα2 T cell hybridomas was highly homogeneous and unaffected by *Emv2* expression. In contrast to Vα2 T cell hybridomas, and as expected by their low avidity for the index env$_{124\text{-}138}$L sequence, none of the non-Vα2 T cell hybridomas derived from *Emv2*-deficient mice displayed strong preference for L at position 128 (Figure 7A). The latter hybridomas did, however, respond strongly to env variants with a different amino acid residue at this position, most frequently V or I, or in the case of clone E2H10 the unrelated S (Figure 7A). Non-Vα2 T cell hybridomas selected by *Emv2* were also heterogeneous, with two clones showing similar preference and specificity for V or I, and two other clones showing much wider reactivity to at least 10 different amino acid residues, including L (Figure 7A). Furthermore, the low reactivity to the index env$_{124\text{-}138}$L sequence of two of the four non-Vα2 T cell hybridomas derived from *Emv2*-deficient mice, but not those derived from *Emv2*-sufficient mice, could be enhanced by substitutions at another position (C for clone E2H10, and S or T for clone E2L18, instead of N at position 133) (Figure 7C). Non-Vα2 T cell hybridomas that could recognize L at position 128 also preferred the amino acid residue of the index env$_{124\text{-}138}$L sequence at the two other positions (T and N for positions 129 and 133, respectively) (Figure 7B, C). Collectively, these results confirmed the differential preference for L at position 128 between Vα2 and

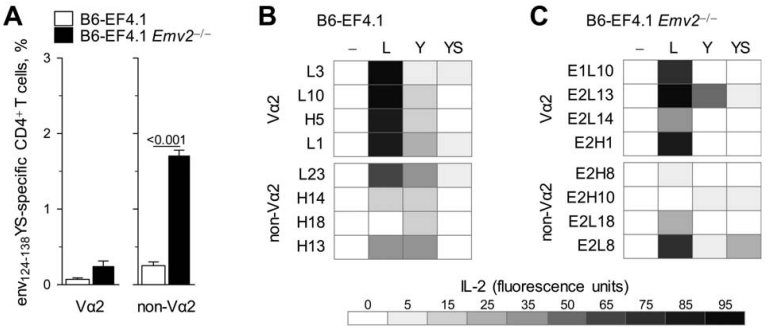


Figure 6. Cross-reactivity of individual Vα2 or non-Vα2 CD4$^+$ T cells. (A) Frequency of env$_{124\text{-}138}$YS-reactive cells in Vα2 or non-Vα2 CD4$^+$ T cells isolated from either B6 (B6-EF4.1) or *Emv2*-deficient B6 (B6-EF4.1 *Emv2*$^{-/-}$) EF4.1 mice. Data are the means ± SEM ($n = 9$) of 18-hr stimulations from 3 experiments. (B–C) IL-2 production in response to stimulation with 5×10^{-6} M env$_{124\text{-}138}$L (L), env$_{124\text{-}138}$Y (Y) or env$_{124\text{-}138}$YS (YS) in comparison with the absence of peptide stimulation (-) of Vα2 or non-Vα2 env$_{124\text{-}138}$L-reactive hybridoma T cell lines derived from *Emv2*$^{+/+}$ (B) or *Emv2*$^{-/-}$ (C) EF4.1 mice.

doi:10.1371/journal.ppat.1002709.g006

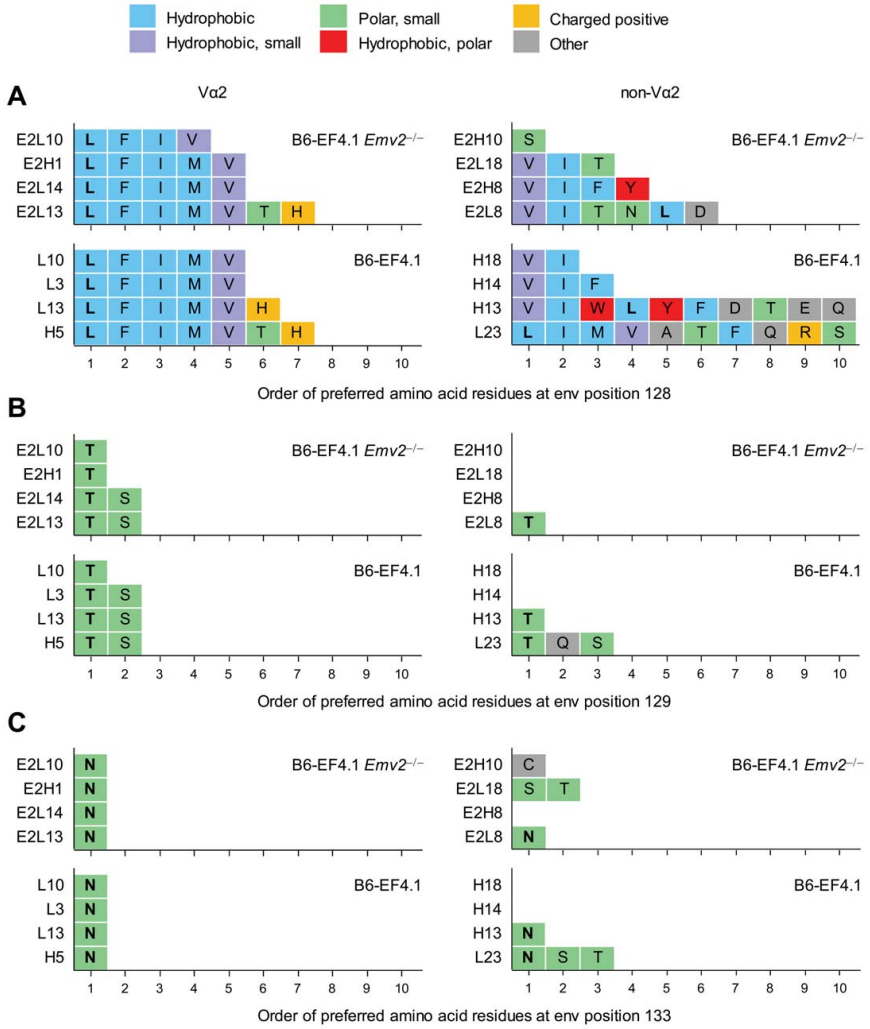


Figure 7. Depth of Vα2 or non-Vα2 env-specific CD4$^+$ T cell repertoires. (A–C) Vα2 or non-Vα env$_{124\text{-}138}$L-reactive hybridoma T cell lines were derived from *Emv2*$^{+/+}$ (B6-EF4.1) or *Emv2*$^{-/-}$ (B6-EF4.1 *Emv2*$^{-/-}$) EF4.1 mice and tested for reactivity against a library of env$_{126\text{-}138}$ peptide epitopes. The amino acid residues in positions 128 (A), 129 (B) and 133 (C) that elicited at least 40% of the maximal response are listed in the order of preference by the individual clones.
doi:10.1371/journal.ppat.1002709.g007

non-Vα2 T cell hybridomas and further suggested that selection by *Emv2* enriched the non-Vα2 repertoire for clones with relative indifference for this position.

Genetic contribution to a high-avidity env$_{124\text{-}138}$L-reactive CD4$^+$ T cell repertoire

Analysis of the env-reactive CD4$^+$ T cell repertoire in B6 mice revealed a clear effect of *Emv2*-mediated selection. However, in addition to *Emv2*, the presence of numerous other endogenous retroviruses could affect the formation of the env$_{124\text{-}138}$L-reactive CD4$^+$ T cell repertoire, even if their primary amino acid sequence is not closely homologous with that of F-MLV env. Furthermore, the functional avidity of env$_{124\text{-}138}$L-reactive CD4$^+$ T cells could also be affected by additional genetic determinants other than endogenous retroviruses. To address this question we generated congenic EF4.1 mice on the 129S8 background. 129S8 mice share the same MHC class II allele with B6 mice (H2-A^b), thus allowing restriction of env$_{124\text{-}138}$L-specific EF4.1 CD4$^+$ T cells. However, they do differ substantially with respect to the composition of endogenous retroviruses and, importantly, 129S8 mice are naturally devoid of endogenous ecotropic MLVs [34,35]. Similar frequency of env$_{124\text{-}138}$L-reactive Vα2 CD4$^+$ T cells developed in B6 and 129S8 EF4.1 mice (Figure 8A). In contrast, the frequency of env$_{124\text{-}138}$L-reactive non-Vα2 CD4$^+$ T cells was significantly higher on the 129S8 than on the B6 background (Figure 8A), and was comparable with that on the *Emv2*-deficient B6 background

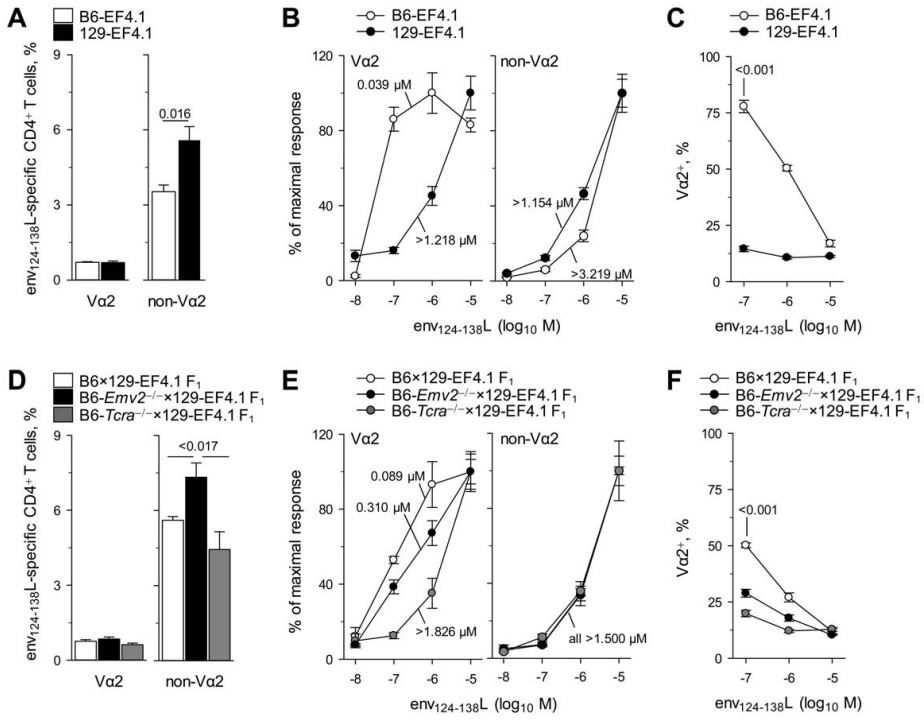


Figure 8. Genetic contribution to a high-avidity env-reactive CD4$^+$ T cell repertoire. (A) Frequency of env$_{124\text{-}138}$L- reactive cells in Vα2 or non-Vα2 primary CD4$^+$ T cells isolated from either B6 (B6-EF4.1) or 129S8 (129S8-EF4.1) EF4.1 mice. (B) Functional avidity of env$_{124\text{-}138}$L-reactive Vα2 or non-Vα2 primary CD4$^+$ T cells from the same donors in A. (C) Frequency of Vα2 cells in env$_{124\text{-}138}$L-reactive CD4$^+$ T cells from the same donors in A as a function of peptide concentration. (D) Frequency of env$_{124\text{-}138}$L- reactive cells in Vα2 or non-Vα2 primary CD4$^+$ T cells isolated from either B6×129S8-EF4.1 F$_1$, B6-*Emv2*$^{-/-}$×129S8-EF4.1 F$_1$ or B6-*Tcra*$^{-/-}$×129S8-EF4.1 F$_1$, EF4.1 mice. (E) Functional avidity of env$_{124\text{-}138}$L-reactive Vα2 or non-Vα2 primary CD4$^+$ T cells from the same donors in D. (F) Frequency of Vα2 cells in env$_{124\text{-}138}$L-reactive CD4$^+$ T cells from the same donors in D as a function of peptide concentration. Numbers in (B) and (E) represent the ED$_{50}$. In (C) and (F) the CD4$^+$ T cell response elicited by the last peptide dose (10^{-8} M) was too small to allow accurate measurement of the frequency of Va2 cells and was therefore omitted. Data in (A–F) are the means ± SEM (n = 4–8) of 18-hr stimulations from 3 experiments.
doi:10.1371/journal.ppat.1002709.g008

(Figure 5B), as was their functional avidity (Figure 8B). This finding indicated that deletion of env$_{124\text{-}138}$L-reactive non-Vα2 CD4$^+$ T cells in B6, but not in B6-*Emv2*$^{-/-}$ or 129S8 mice was mediated primarily by *Emv2*. Surprisingly, however, the functional avidity of env$_{124\text{-}138}$L-reactive Vα2 CD4$^+$ T cells in 129S8 mice was very much reduced in comparison with that of Vα2 CD4$^+$ T cells in B6 mice (Figure 8B), and was as low as that of low-avidity non-Vα2 CD4$^+$ T cells. As, a result of differences in frequency and functional avidity, the env$_{124\text{-}138}$L-specific response of 129S8 mice was dominated by non-Vα2 CD4$^+$ T cells at all peptide doses, in contrast to that of B6 mice, which was dominated by Vα2 CD4$^+$ T cells at low peptide doses (Figure 8C).

To further explore the origin of high-avidity env$_{124\text{-}138}$L-reactive Vα2 CD4$^+$ T cells in B6, but not in 129S8 mice, we tested the response of a series of B6×129S8-EF4.1 F$_1$ mice. In comparison with B6×129S8-EF4.1 F$_1$ mice, which inherited *Emv2* from the B6 parent, B6-*Emv2*$^{-/-}$×129S8-EF4.1 F$_1$ mice, which lacked ecotropic MLVs, had elevated frequencies of env$_{124\text{-}138}$L-reactive non-Vα2 CD4$^+$ T cells, whereas frequencies of env$_{124\text{-}138}$L-reactive Vα2 CD4$^+$ T cells were similar (Figure 8D). These results confirmed that elevated frequencies of env$_{124\text{-}138}$L-reactive non-Vα2 CD4$^+$ T cells in 129S8 mice were indeed due to lack of *Emv2*-mediated selection. Interestingly, both B6×129S8-EF4.1 and B6-*Emv2*$^{-/-}$×129S8-EF4.1 F$_1$ mice generated env$_{124\text{-}138}$L-reactive Vα2 CD4$^+$ T cells with higher avidity than those of 129S8 mice (Figure 8E), suggesting that a genetic contribution of the B6 parent, other than *Emv2*, was necessary for the development of high-avidity env$_{124\text{-}138}$L-reactive Vα2 CD4$^+$ T cells. To assess whether this genetic contribution arose from polymorphisms in the *Trav* locus itself, we tested B6-*Tcra*$^{-/-}$×129S8-EF4.1 F$_1$ mice, which inherited *Emv2* from the B6 parent, but could generate endogenous Vα chains only from the locus inherited from the 129S8 parent. The presence of *Emv2* in B6-*Tcra*$^{-/-}$×129S8-EF4.1 F$_1$ mice had the predicable effect on the frequency of env$_{124\text{-}138}$L-reactive non-Vα2 CD4$^+$ T cells (Figure 8D), which displayed comparably low avidity in all three F$_1$ strains tested (Figure 8E). Surprisingly, however, env$_{124\text{-}138}$L-reactive Vα2 CD4$^+$ T cells that had developed in B6-*Tcra*$^{-/-}$×129S8-EF4.1 F$_1$ mice were also low-avidity, which was comparable with that of Vα2 CD4$^+$ T cells in 129S8 mice (Figure 8E), suggesting that the ability of B6 mice to generate high-avidity env$_{124\text{-}138}$L-reactive Vα2 CD4$^+$ T cells was germline-encoded. Consequently, the env$_{124\text{-}138}$L-specific response of B6×129S8-EF4.1 F$_1$ mice, but not of isogenic mice lacking either *Emv2* or the B6-origin *Trav*, was dominated by Vα2 CD4$^+$ T

cells at low peptide doses (Figure 8F). The peak percentage of Vα2 $CD4^+$ T cells in the $env_{124\text{-}138}$L-reactive population was lower in B6×129S8-EF4.1 F_1 mice than in B6 mice, as the former were expressing endogenous Vα chains from both parental *Trav* loci. Thus, the combined effect of *Emv2* on the frequency of non-Vα2 T cells and of *Trav* on the avidity of Vα2 T cells was necessary for the dominance of high-avidity Vα2 $CD4^+$ T cells in the response to $env_{124\text{-}138}$L.

Discussion

As a result of the combinatorial process that creates TCRs, their specificity is random and has to undergo selection. Thymic positive and negative selection of developing T cells ensures that mature T cells in the periphery have a functional TCR and minimal reactivity to self proteins, respectively [36]. Negative selection is thought to decrease the frequency, avidity and cross-reactivity of the developing TCR repertoire specific to foreign epitopes that may be similar to self-derived epitopes presented in the thymus [36] and promote peptide specificity [37]. Here we used a well-characterized molecular system to show that negative selection by a defined self peptide from *Emv2* env indeed decreased the frequency in the naïve $CD4^+$ T cell repertoire of clones specific to a range of foreign env epitopes, thus reducing the magnitude of the $CD4^+$ T cell response to all env epitope variants. However, negative selection counter-intuitively also promoted the avidity of the $CD4^+$ T cell response to F-MLV env by shifting the clonal composition of responding $CD4^+$ T cells in favor of high-avidity cells.

$CD4^+$ T cells play a central coordinating role in the orchestration of adaptive immunity to infection, and may also mediate direct antiviral activity. Recent studies in diverse systems have indicated an essential role for the $CD4^+$ T cell response in the control of retroviral infection [15,38–42]. We have previously shown that protection of wt mice against acute FV infection is proportional to the frequency of virus-specific $CD4^+$ T cells [23]. Surprisingly, we found that although negative selection significantly reduced both the precursor frequency and peak expansion of F-MLV env-specific $CD4^+$ T cells, it did not compromise $CD4^+$ T cell-mediated antiviral activity. This finding suggested that not all virus-specific $CD4^+$ T cells were equal in their ability to mediate antiviral functions. Indeed, negative selection by *Emv2* env affected $CD4^+$ T cells with low avidity for F-MLV env, but not those with high avidity for the same epitope. Preservation of full antiviral activity in the *Emv2*-selected $CD4^+$ T cell repertoire therefore indicated that this activity is primarily, if not exclusively, exerted by high-avidity $CD4^+$ T cells.

High-avidity virus-specific $CD4^+$ T cells may be superior in certain direct antiviral or indirect helper functions than low-avidity ones, but there may also be important exceptions. High-avidity $CD4^+$ T cells responding to FV infection have been reported to show enhanced *ex vivo* production of IFN-γ and IL-21 cytokines and reduced expression of PD-1 inhibitory receptor [15] than low-avidity counterparts, properties that may contribute to superior antiviral activity. However, T follicular helper (Tfh) differentiation and function were previously found to be similar between high- and low-avidity virus-specific $CD4^+$ T cells [15], suggesting that provision of T cell help for the production of virus-neutralizing antibodies may be more sensitive to the frequency of virus-specific $CD4^+$ T cells, rather than their avidity. However, in addition to the frequency of virus-specific $CD4^+$ T cells, the virus-specific antibody response is also proportional to the frequency of rare antigen-specific B cells. Thus, when availability of T cell help is abundant, the virus-specific antibody response may be limited by the frequency of antigen-specific B cells and additional T cell help would not be expected to enhance antibody production. Consistent with this idea, adoptive transfer of virus-specific EF4.1 $CD4^+$ T cells into wt B6 mice did not accelerate the virus-neutralizing antibody response [23]. In addition to an effect of *Emv2* on the availability of T cell help for the FV-specific antibody response, *Emv2* could in principle also directly affect the development of virus-specific B cells [43]. Although we observed comparably low FV-specific antibody responses between B6 and B6-*Emv2*$^{-/-}$ mice at the peak of FV infection, our results did not exclude a potential direct effect of *Emv2* on FV-specific B cell and antibody responses at later time-points, when these responses are fully induced. Indeed, *Emv2*-encoded env shares 79% amino acid identity with F-MLV env and it is therefore possible that *Emv2* expression, especially when upregulated, might affect the FV-specific antibody response.

As previously shown, high-avidity F-MLV $env_{122\text{-}141}$L-specific Vα2 $CD4^+$ T cells are a minority subset in the naïve repertoire and only dominate the immune response to FV as a result of their preferential expansion during infection [15]. We have now found that for this ability of high-avidity F-MLV $env_{122\text{-}141}$L-specific Vα2 $CD4^+$ T cells to dominate the peak response, negative selection by *Emv2* of at least some of the competitor low-avidity F-MLV $env_{122\text{-}141}$L-specific non-Vα2 $CD4^+$ T cells is necessary. These findings indicate that even subtle thymic events can have profound effects on the induction of an effective T cell response to retroviral infection. Recently, a comprehensive theoretical study has indicated that *HLA* class I alleles that associated with control of HIV infection, such as *HLA-B*5701*, sample far fewer self peptides than other *HLA* alleles [5]. As a result of less stringent negative selection, a higher frequency of $CD8^+$ T cells restricted by these protective alleles were predicted to recognize viral peptide epitopes and to cross-react with variants of the targeted epitopes [5].

Our results with a single self peptide provide further experimental confirmation of negative selection reducing both the precursor frequency and cross-reactivity of env-specific $CD4^+$ T cells, although in this case the effect on cross-reactivity was more pronounced at the population, rather than the single-cell level. These results also suggest that from the thousands of self peptides that can mediate thymic selection of retrovirus-specific T cells, the main effects may be mediated by only a few self peptides. Moreover, self peptides with such strong influence may also be polymorphic between different individuals, which might contribute to the partial association of *HLA* polymorphisms with virus control [3,5,6].

In addition to polymorphisms at the *MHC/HLA* locus or of self peptides mediating thymic selection, the *Trav/TRAV* and *Trbv/TRBV* loci may also display allelic sequence variation. A polymorphism in the *TRBV9* gene has been shown to affect TCR affinity for and functional recognition of an HLA-B*3501-restricted epitope from the EBNA-1 protein of Epstein-Barr virus (EBV), leading to a public T cell response dominated by the high-affinity variant [44]. Similarly, we found that the ability of Vα2 chains to confer high avidity for $env_{122\text{-}141}$L in EF4.1 mice seems to be germline-encoded, as only Vα2 chains encoded by the B6, but not the 129 *Trav* locus had this ability. It is tempting to speculate that amino acid residues unique to the B6-germline *Trav14*-encoded Vα2 chains participate in recognition of the strongly interacting L (or a limited set of amino acids with similar properties) at env position 128. Notably, the $CD8^+$ T cell response to an HLA-B8-restricted epitope from the latent antigen EBNA 3A of EBV uses almost exclusively identical Vα and Vβ, as well as other TCR-region sequences, and comprehensive structural studies have shown that a unique amino acid residue in the

germline-encoded complementarity-determining region 2 (CDR2) of the preferred Vα chain, encoded by *TRAV26-2*, is critically required for binding to a residue from the peptide epitope [45]. Despite the vast number of somatically-generated random TCRs that can arise during T cell development, these studies highlight the potential for germline-encoded residues to provide exquisite specificity and competitive advantage to the TCRs that carry them.

In addition to likely representing the best-fit for recognition of A^b-restricted $env_{122\text{-}141}L$, the dominance of Vα2 EF4.1 $CD4^+$ T cells could also result from preferential pairing of the transgenic TCRβ chain with Vα2 chains in general. This is unlikely to be the case as the usage of Vα2 cells was not increased in either total or $env_{122\text{-}141}L$-reactive EF4.1 $CD4^+$ T cells, and indeed in the $env_{122\text{-}141}L$-reactive preimmune repertoire clones using other Vα chains were at least 3 times more frequent than those using Vα2. However, although non-Vα2 $env_{122\text{-}141}L$-reactive $CD4^+$ T cells were still the majority in *Emv2*-expressing mice, their ability to participate in the response to FV and compete with $env_{122\text{-}141}L$-reactive Vα2 $CD4^+$ T cells was severely compromised by *Emv2*. Thus, the dominance of Vα2 $CD4^+$ T cells in the response to FV infection can be seen as a combination of germline-encoded advantage in A^b-restricted $env_{122\text{-}141}L$ recognition conferred to Vα2 $CD4^+$ T cells and of *Emv2*-mediated self-tolerance of other non-Vα2 $CD4^+$ T cells capable of recognizing A^b-restricted $env_{122\text{-}141}L$.

One important novel insight of the current study is the proof of principle that negative selection is not necessarily always impairing high-avidity T cell responses. By counter-selecting some cross-reactive $CD4^+$ T cells, negatively selecting self peptides have the ability to significantly enhance the avidity for the response to at least some epitope variants. Higher precursor frequency and cross-reactivity with emerging epitope variants seem to be the best correlates for an effective cytotoxic $CD8^+$ T cell response [5]. Whether higher avidity for the primary infecting epitope, rather than cross-reactivity with epitope variants better describes an effective $CD4^+$ T cell response to retroviral infection needs to be further addressed.

It should be noted that differences in avidity for antigen in this system were defined functionally. Indeed, Vα2 $env_{122\text{-}141}L$-specific primary $CD4^+$ T cells or hybridomas reacted to much lower concentrations of $env_{122\text{-}141}L$ peptide stimulation *in vitro* than their non-Vα2 counterparts. Furthermore, this higher sensitivity translated to higher *in vivo* expansion and increased potential for cytokine production [15]. It is currently unclear whether differences in functional avidity between Vα2 and non-Vα2 $env_{122\text{-}141}L$-specific $CD4^+$ T cells resulted from overall higher affinity of individual TCRs of these polyclonal populations for the peptide-MHC class II complex. Although dissociation kinetics between TCRs and peptide-MHC class II tetramers are often informative with respect to the biochemical affinity of these TCRs, they may not be universally useful. For example, the available $env_{123\text{-}141}$-A^b tetramer (A^b-env) is known to bind only some $env_{124\text{-}138}L$-specific $CD4^+$ T cell clones but not others, irrespective of their functional avidity or Vα usage [14,22]. Therefore, this reagent could not be used to access the biochemical affinity of all $env_{124\text{-}138}L$-specific $CD4^+$ T cells in the polyclonal repertoire. Furthermore, identification of antigen-specific cells using a sensitive two-dimensional binding assay has recently demonstrated that the affinity of many $CD4^+$ T cells that participate in the response to two separate antigens is below detection with peptide-MHC class II tetramers [46]. Thus, peptide-MHC class II tetramers may generally only detect some but not all antigen-specific $CD4^+$ T cells. In addition, such detection is conditional on expression of sufficient TCR levels. Indeed, we have found that the extensive, antigen-induced downregulation of their TCR *in vivo*, eclipses detection with the A^b-$env_{123\text{-}141}$ tetramer of even the $env_{122\text{-}141}L$-reactive $CD4^+$ T cells that could otherwise bind this reagent. Similar observations have been recently made with peptide-MHC class I tetramer staining of virus-specific effector $CD8^+$ T cells [47], suggesting that the inability of peptide-MHC multimers to identify antigen-specific effector T cells that have downregulated their TCRs may be a general problem for T cells restricted by both classes of MHC molecules.

Negative selection ensures minimal reactivity of developing thymocytes to self proteins. However, endogenous retroviruses are a large constituent of mammalian genomes and thus represent a potentially large pool of self proteins able to mediate selection, both positive and negative. Self peptides encoded by endogenous MLVs have been shown to mediate positive selection of $CD4^+$ T cells with specificity for an unrelated H2-E^k-restricted moth cytochrome C peptide, and to enhance the response of mature $CD4^+$ T cells with this specificity in the periphery [48]. We found that *Emv2* was expressed at very low levels in the thymus of B6 mice, in agreement with a previous report [49], and was undetectable by qRT-PCR in some of the mice. It should be noted, however, that the qRT-PCR method employed was specific only for the spliced *env* mRNA that is transcribed by *Emv2*. This was chosen to eliminate the possibility of detecting contaminating genomic DNA or viral genomic RNA, but may underestimate the total amount of spliced and unspliced mRNA that leads to the production of other viral proteins. Nevertheless, as demonstrated by its effect on thymic development, this low level of *Emv2* expression was clearly functional.

Endogenous retroviruses have been known for many years to cause a range of different diseases in mice, including cancer, immunodeficiency and autoimmunity, although a similar causal effect in humans has been questioned [50]. Immune reactivity to endogenous retroviruses has been amply demonstrated in mice where is has been strongly associated with the development of spontaneous autoimmune conditions [51,52]. Interestingly, immune reactivity to endogenous retroviruses has also been frequently observed in humans during infection, inflammation, autoimmunity and cancer [50,53–56]. Expression of human endogenous retroviruses, as well as $CD8^+$ T cell responses against their antigens, have been documented in HIV infection [57,58]. Furthermore, a whole-genome association study has suggested that part of the effect of the protective *HLA-B*5701* allele during the asymptomatic period of HIV infection may be mediated by a linked human endogenous retrovirus at the same locus [59]. Human endogenous retroviral antigens have also been reported to serve as targets for $CD8^+$ T cell-mediated rejection of cancer cells [60]. It might be evident from the studies in humans and the results of the current study that peptide epitopes encoded by endogenous retroviruses have a strong influence on T cell thymic selection and may also participate in the shaping of the peripheral T cell response. It is also clear that endogenous retroviruses do not always cause immunological tolerance, and although their activation in infected or transformed cells may provide a non-mutable target for immune attack, activation of endogenous retroviruses may also trigger inflammatory or autoimmune phenomena frequently associated with infection and cancer. Further study of endogenous retrovirus regulation during infection, autoimmunity or cancer, and of the immune responsiveness to them should shed more light into their pathogenic potential.

Materials and Methods

Ethics statement

All animal experiments were approved by the ethical committee of the NIMR, and conducted according to local guidelines and

UK Home Office regulations under the Animals Scientific Procedures Act 1986 (ASPA).

Mice

Inbred C57BL/6J (B6), A/J and B6.SJL-*Ptprc*^{a} *Pepc3*^{b}/BoyJ (CD45.1^{+} B6) mice were originally obtained from The Jackson Laboratory (Bar Harbor, Maine, USA) and were subsequently maintained at NIMR animal facilities. Inbred 129S8/SvEvNimrJ (129S8) mice were developed from an 129/Sv substrain, maintained at NIMR animal facilities, and were subsequently deposited at The Jackson Laboratory. The B6 TCRβ-transgenic strain EF4.1, expressing a transgenic TCRβ chain from a T cell clone specific to F-MuLV env$_{122\text{-}141}$ presented by H2-A^{b}, has been described [14]. 129S8-congenic EF4.1 mice were generated by serial backcrossing of B6-EF4.1 mice for 10 nuclear generations onto the 129S8 genetic background. B6-backcrossed Rag1-deficient (*Rag1*$^{-/-}$) mice [61] and T cell receptor α-deficient (*Tcra*$^{-/-}$) mice [62] were also maintained at NIMR animal facilities. *Fv2*^{s}-congenic B6 (*Fv2*^{s}) and *Rag1*$^{-/-}$ (*Fv2*^{s} *Rag1*$^{-/-}$) mice have been previously described [25]. *Emv2*-deficient (*Emv2*$^{-/-}$) B6 mice were created by introducing the *Emv2* integration site of chromosome 8 from the A/J strain, which lacks this proviral integration, by serial backcrossing for at least 12 nuclear generations onto the B6 genetic background. Lack of *Emv2* was validated by PCR for both the *D8Mit49* microsatellite marker close to the locus that detects polymorphisms in A/J (*Emv2*$^{-}$) and B6 (*Emv2*$^{+}$) strains of mice (*D8Mit49* forward 5′-TCTGTGCATGGCTGTGTATG-3′ and *D8Mit49* reverse 5′-TGGTGTGCTGCTGATGCT-3′), and also for the actual integration site using three primers, two of which were flanking the integration site (forward 5′-ACCCACTAAGTAACCCAG-GCTGCCTCAGCT-3′ and reverse 5′-GACCAGAATAGAAA-GACGTTCAAGTGAGCT-3′) and one located in the *Emv2* LTR (5′-ATCAGCTCGCTTCTCGCTTCTGTACCCGCG-3′) (Figure S3).

In vitro T cell activation

Spleen or lymph node single-cell suspensions were prepared from EF4.1 mice and 5×10^{5} cells per well were stimulated in 96-well plates with the indicated amount of env peptide variants. The frequency of env-reactive cells in stimulated CD4^{+} T cells was defined as the frequency of cells that responded to 18-hr stimulation, before cell division or death had occurred, by upregulating CD69 expression. Correct identification of env-reactive CD4^{+} T cells by CD69 upregulation was confirmed in control experiments by co-staining for CD154 (CD40L) expression in stimulated T cells. Both antibodies were obtained from eBiosciences. For assessment of T cell activation on day 3, cells were labeled with CFSE before stimulation and responding cells were identified by CFSE dilution.

Hybridoma cell line generation and stimulation

Single-cell suspensions were prepared from spleens and lymph nodes from *Emv2*-sufficient or -deficient EF4.1 mice and stimulated *in vitro* with 10^{-7} M or 10^{-5} M env$_{122\text{-}141}$L peptide and 4 ng/ml recombinant human IL-2 for 4 days. CD4^{+} T cells were subsequently purified from stimulated cultures using immunomagnetic positive selection (StemCell Technologies, Vancouver, BC, Canada) and fused to TCRαβ-negative BW5147 thymoma cells to produce hybridoma cell lines. Established hybridoma cell lines were stimulated with a range of env peptide variants presented by dendritic cells. Dendritic cells were obtained from cultures of bone marrow cells isolated from B6 mice and supplemented with granulocyte macrophage colony-stimulating factor (GM-CSF). GM-CSF was obtained from culture supernatant of ×63 cells transfected with mouse *Csf2* and was used at 1:10 dilution. Bone marrow cells were culture in these conditions for 7 days, at which point they consisted of 50–70% dendritic cells. These cells were then used to stimulate hybridoma cells at a ratio of 5×10^{4} dendritic cells to 1×10^{5} hybridoma cells, for 18 hrs, in the presence or absence of env peptide variants. Dendritic cell-hybridoma cell co-cultures were plated in flat-bottom 96-well plates in 200 µl final volume. The concentration of peptides used is indicated in individual figures and figure legends. In additional experiments peritoneal macrophages were also used as antigen-presenting cells with results comparable to the use of dendritic cells. Macrophages were isolated from B6 mice following plating of the peritoneal cavity exudate cells for 1 hr and washing off the non-adherent fraction. Env-specific responses were assessed by measuring the amount of IL-2 secreted in co-culture supernatants using an AlamarBlue (Invitrogen, Carlsbad, CA, USA)-based CTLL-2 assay.

Tra gene usage

Trav and *Traj* usage by T cell hybridomas was probed by staining with an anti-Vα2 (clone B20.1) or anti-Vα3.2 (clone RR3-16) monoclonal antibodies, and by reverse transcription (RT)-PCR amplification and sequencing of expressed *Trav* genes, using previously described primers [63]. *Trav* and *Traj* segment identification and alignment, and confirmation of productive rearrangements were performed on the International Immunogenetics Information System website (http://www.imgt.org).

Viruses and infections

The FV used in this study was a retroviral complex of a replication-competent B-tropic F-MuLV and a replication-defective polycythemia-inducing spleen focus-forming virus (SFFVp). Stocks were propagated *in vivo* and prepared as 10% w/v homogenate from the spleen of 12-day infected BALB/c mice. Mice received an inoculum of ~1,000 spleen focus-forming units of FV. All viral stocks were free of Sendai virus, Murine hepatitis virus, Parvoviruses 1 and 2, Reovirus 3, Theiler's murine encephalomyelitis virus, Murine rotavirus, Ectromelia virus, Murine cytomegalovirus, K virus, Polyomavirus, Hantaan virus, Murine norovirus, Lymphocytic choriomeningitis virus, Murine adenoviruses FL and K87, and Lactate dehydrogenase-elevating virus. Virus inocula were injected via the tail vein in 0.1 ml of phosphate-buffered saline. FV-infected cells were detected by flow cytometry using surface staining for the glycosylated product of the viral *gag* gene (glyco-Gag), using the matrix (MA)-specific monoclonal antibody 34 (mouse IgG2b), followed by an anti-mouse IgG2b-FITC secondary reagent (BD, San Jose, CA, USA). For the assessment of anemia, mice were bled by a small incision of the tail vein and blood was collected into heparinized capillary tubes. Complete blood counts were measured on a VetScan HMII hematology analyzer (Abaxis, CA, USA), following the manufacturer's instructions. RBC counts of uninfected mice were ~9.95×10^{6} per mm^{3} of blood. FV-induced splenomegaly in infected mice was expressed as spleen index, which is the ratio of the weight of the spleen (in mg) to the weight of the rest of the body (in g).

FV-neutralizing and F-MLV-infected cell-binding antibody assays

Serum titers of FV-neutralizing antibodies were measured as previously described [25]. The dilution of serum which resulted in 75% neutralization was taken as the neutralizing titer. Serum titers of F-MLV-infected cell-binding antibodies were determined by

flow cytometry following primary staining of F-MLV-infected *Mus dunni* cells with serial dilutions of serum samples and secondary staining with fluorescently labeled anti-mouse IgG1 (clone A85-1), anti-mouse IgG2a/c (clone R19-15), anti-mouse IgG2b (clone R12-3) or anti-mouse IgM (clone R6-60.2) antibodies (BD). B6 mice express the IgG2c isotype, which may not be efficiently detected by anti-IgG2a reagents [64]. Although the R19-15 monoclonal antibody has higher affinity for IgG2a, it can be effectively used for detection of IgG2c. This was confirmed by staining of F-MLV-infected *Mus dunni* cells that were first incubated with serum from FV-infected mice, with the anti-IgG1 or anti-IgG2a/c or anti-IgG2b reagents separately (Figure S8A). The three reagents used separately resulted in comparable staining intensity, which allowed us to use all three IgG subclass-specific antibodies in combination. For IgG titers, F-MLV-infected *Mus dunni* cells were first incubated with serum samples and then with anti-IgG1, anti-IgG2a/c and anti-IgG2b antibodies mixed together. Serum samples were 2-fold serially diluted, starting from an initial dilution of 1:50. The last positive serum dilution resulting in staining intensity at least twice the background level was taken as the binding titer (Figure S8B).

T cell purification and adoptive transfer

Single-cell suspensions were prepared from the spleens and lymph nodes of donor CD45.2$^+$ EF4.1 mice by mechanical disruption. Spleen suspensions were treated with ammonium chloride for erythrocyte lysis. CD4$^+$ T cells were enriched using immunomagnetic positive selection (StemCell Technologies) according to the manufacturer's instructions. Purity of the isolated CD4$^+$ T-cell population was routinely higher than 92%. A total of approximately 1×10^6 EF4.1 CD4$^+$ T cells were injected in B6-congenic CD45.1$^+$CD45.2$^+$ recipients via the tail vein in 0.1 ml of air-buffered Iscove's Modified Dulbecco's Media. When adoptive transfer of CD4$^+$ T cells was combined with FV infection, purified CD4$^+$ T cells and virus stocks were injected separately into recipient mice within a 24 hour-period.

Flow cytometry

Spleen-cell suspensions were stained with directly-conjugated antibodies to surface markers, obtained from eBiosciences (San Diego, CA, USA), CALTAG/Invitrogen, BD Biosciences (San Jose, CA, USA) or BioLegend (San Diego, CA, USA). Donor-type env-specific CD4$^+$ T cells were identified as CD44^{hi}CD45.2$^+$CD45.1$^-$CD4$^+$ cells. Four- and 8-color cytometry were performed on FACSCalibur (BD Biosciences) and CyAn (Dako, Fort Collins, CO) flow cytometers, respectively, and analyzed with FlowJo v8.7 (Tree Star Inc., Ashland, OR, USA) or Summit v4.3 (Dako) analysis software, respectively.

Emv2 expression by quantitative reverse transcription (qRT)-PCR

Total RNA was extracted from whole organs using TRI-reagent (Sigma-Aldrich, St. Louis, US) according to the manufacturer's instructions, precipitated with isopropanol and washed in 75% ethanol before being dissolved in water. DNase digestion and cleanup was performed with the RNeasy Mini Kit (Qiagen, Hilden, Germany) and cDNA produced with the high capacity reverse transcription kit (Applied Biosystems, Carlsbad, US) with an added RNase inhibitor (Promega Biosciences, Madison, US). A final clean-up was performed with the QIAquick PCR purification kit (Qiagen). Level of expression of *Emv2* RNA was determined by qRT-PCR using DNA Master SYBR Green I kit (Roche, Mannheim, Germany) and the ABI Prism 7000 or 7900HT Detection System (TaqMan, Applied Biosystems, Foster City, CA). The following primers were used for the amplification of target transcripts: *Hprt*: forward 5′-TTGTATACCTAATCATTATGCCGAG-3′ and reverse 5′- CATCTCGAGCAAGTCTTTTCA-3′; *Emv2*: forward 5′-CCAGGGACCACCGACCCACCGT-3′ and reverse 5′-TAGTCGGTCCCGGTAGGCCTCG-3′. *Emv2*-specific primers amplified only the spliced form of *env* mRNA, thus eliminating the possibility of residual genomic DNA or RNA contamination contributing to *Emv2* signal. The housekeeping gene *Hprt* was used to normalize the Critical Threshold (C_T) values for *Emv2*. Analysis was conducted with the ΔC_T method [65] and *Emv2* expression corresponding to an *Emv2* C_T value of 40 (the total number of amplification cycles used) was set at 1 arbitrary unit. A theoretical detection limit of 2 arbitrary units was also used, which represents the detectable *Emv2* signal in the penultimate cycle of amplification.

Statistical analysis

Statistical comparisons were made using SigmaPlot 12.0 (Systat Software Inc., Germany). Parametric comparisons of normally-distributed values that satisfied the variance criteria were made by unpaired Student's *t*-tests. Linear percentages of FV-infected cells, spleen indices and nAb titers, which did not pass the variance test, were compared with non-parametric two-tailed Mann-Whitney Rank Sum or Wilcoxon Signed Rank tests.

Accession numbers

- *Cd4* cluster of differentiation 4 antigen [*Mus musculus*]; Gene ID: 12504; Protein ID: NP_038516.1
- *Rag1* recombination activating gene 1 [*Mus musculus*]; Gene ID: 19373; Protein ID: NP_033045.2
- *Fv2* Friend virus susceptibility 2 [*Mus musculus*]; Gene ID: 19882; Protein ID: NP_033100.1
- *Tcra* T cell receptor alpha chain [*Mus musculus*]; Gene ID: 21473
- *Tcrb* T cell receptor beta chain [*Mus musculus*]; Gene ID: 21577
- *Emv2* endogenous ecotropic MuLV 2 [*Mus musculus*]; Gene ID: 111372
- *env* envelope protein [*Friend murine leukemia virus*]; Gene ID: 1491875; Protein ID: NP_040334.1

Supporting Information

Figure S1 Effect of N-terminal epitope length on TCR recognition by primary and hybridoma EF4.1 envL-specific CD4$^+$ T cells. (A) Frequency of CD69$^+$ cells in Vα2 or non-Vα2 CD4$^+$ T cells (expressed as percentage of the maximal response elicited by the env$_{122\text{-}141}$L peptide), following 18-hr *in vitro* stimulation of spleen cell suspensions from EF4.1 mice with the indicated range of N-terminal truncated envL peptides. (B) IL-2 production in the supernatant of hybridoma cells lines established from Vα2 or non-Vα2 env-specific EF4.1 CD4$^+$ T clones following 24-hr *in vitro* stimulation with the indicated range of N-terminal truncated envL peptides. Data are pooled from 3 separate experiments.
(PDF)

Figure S2 Sequence and TCR SB14-31 contact residues of F-MLV- and *Emv2*-encoded env$_{123\text{-}140}$. (A) Amino acid sequence, in single-letter code, of env$_{123\text{-}140}$ encoded by either F-MLV or *Emv2*. Differences in sequence are indicated by red color. (B) Important contact residues for the SB14-31 TCR (indicated in

red) and for H2-A^b (indicated in blue) in F-MLV-encoded $env_{123-140}$L. Numbers underneath amino acid residues correspond to amino acid positions in env.
(PDF)

Figure S3 Chromosomal location of *Emv2* in B6 mice and screening of *Emv2*$^{-/-}$ B6 mice. *Emv2* is integrated near the telomere of Chromosome 8 of B6 mice in reverse orientation relative to the forward strand, between the *Tubb3* and *Def8* genes (Search for *Mela* on http://www.ncbi.nlm.nih.gov/mapview), and it is absent from A/J mice. Lack of *Emv2* on *Emv2*$^{-/-}$ congenic B6 mice is shown by PCR for the actual integration site (*red arrows*) or for the polymorphic D8Mit49 microsatellite marker that is further telomeric with respect to *Emv2* (*not shown on map*).
(PDF)

Figure S4 Effect of *Emv2* on the endogenous CD4$^+$ T cell, CD8$^+$ T cell and antibody responses. B6 and B6-*Emv2*$^{-/-}$ mice were infected with FV and their adaptive responses were measured 7 days later. T cell responses were measured in cells isolated from the spleens and antibody responses from the sera of these mice. (A) Percentage of CD44^{hi}CD43$^+$ cells in total CD8$^+$ T cells. (B) Percentage of Vα3.2$^+$Vβ5.2$^+$ cells in either CD44^{hi} (*left*) or CD44^{lo} (*right*) CD8$^+$ T cells. The dashes horizontal line represents the same frequency in uninfected control mice. The dashed horizontal lines in (A) and (B) represent the depicted frequencies in uninfected control mice. (C) Serum titers of F-MLV-infected cell-binding IgG (*left*) and IgM (*right*). Dashed lines represent the limit of detection. (D) Percentage of A^b-$env_{123-141}$ tetramer$^+$ cells in total CD4$^+$ T cells. Horizontal short lines denote the median frequencies and the dashed line denotes the median frequency of A^b-hCLIP (control) tetramer$^+$ cells in the same populations. (E) Percentage of Vα2 cells in A^b-$env_{123-141}$ tetramer$^+$ CD4$^+$ T cells from the same mice. The dashed horizontal line represents the frequency of Vα2 cells in total CD4$^+$ T cells from the same mice. In (A) to (E) each symbol represents an individual mouse.
(PDF)

Figure S5 Gating strategy for the identification of env-specific donor CD4$^+$ T cells. CD45.2$^+$ (*Ptprc*$^{2/2}$) EF4.1 CD4$^+$ T cells (10^6) were adoptively transferred into wild-type *Ptprc*$^{1/2}$ B6 recipients that were infected with FV the same day. Host cells were identified as CD45.1 CD45.2 double-positive whereas donor cells were CD45.2 single-positive.
(PDF)

Figure S6 Effect of *Emv2* on the frequency and composition of $env_{124-138}$L-specific CD4$^+$ thymocytes. Thymocytes from *Emv2*$^{+/+}$ or *Emv2*$^{-/-}$ EF4.1 mice were stimulated for 18 hrs *in vitro* with the indicated amount of env_{124-13}L peptide presented by bone marrow-derived dendritic cells and responding cells were identified by upregulation of CD69 expression. Frequency of responding (CD69$^+$) cells in gated CD4$^+$ thymocytes (*left*) and frequency of Vα2 cells in env_{124-13}L-specific cells (*right*) is shown, with $p<0.006$ and $p<0.002$, respectively, for 10^{-6} M peptide concentration. Results are the means ± SEM ($n = 8–10$).
(PDF)

Figure S7 Depth of env epitope recognition by *Emv2*-selected and -nonselected T cell hybridomas. Vα2 or non-Va2 (Vα3) $env_{124-138}$L-reactive T cell hybridomas were established from *Emv2*$^{+/+}$ or *Emv2*$^{-/-}$ EF4.1 mice and tested for reactivity against a library of $env_{126-138}$ peptide epitopes (at 5×10^{-6} M concentration), in which positions 128, 129 and 133 were individually replaced by all natural amino acids. The response of each clone was measured by secretion of IL-2 and is expressed as a percentage of the maximal response obtained with the most potent variant. Results are the means of triplicate cultures.
(PDF)

Figure S8 Determining titers of F-MLV-infected cell-binding antibodies. (A) *Mus dunni* cells, chronically infected with F-MLV, were stained with an 1:50 dilution of pooled serum samples from wt B6 mice that were infected with FV 35 days earlier (serum sample) or with a dilution of the anti-gp70 of F-MLV monoclonal antibody 48 (mAb 48; IgG2a) that gives similar *in vitro* FV neutralization as 1:50 dilution of the serum sample. Cells stained with the serum sample were subsequently stained with anti-IgG1, anti-IgG2a/c or anti-IgG2c antibodies separately or with all three antibodies mixed together, and cells stained with mAb 48 were stained with anti-IgG2a/c. Black-filled histograms show staining with both primary and secondary antibodies, and gray-shaded histograms show staining with the secondary antibody only. Note comparable staining intensity between all three IgG subclass-specific antibodies used separately as secondary reagents in combination with the serum sample and also in comparison with the mAb 48. (B) Example of titer determination. Experimental serum samples were serially diluted 2-fold (starting from 1:50) and used for staining F-MLV-infected *Mus dunni* cells. The median fluorescent intensity (MFI) of co-staining with all three IgG subclass-specific antibodies at the same time is plotted against the serum dilution. The horizontal dashed line represents the MFI of unstained cells. Data were fitted to a sigmoidal curve. The red lines connect the MFI that is twice the background level and the serum dilution that results in that MFI. The inverse of the serum dilution that results in an MFI at least twice the background level was taken as the titer. This was preferred over titer determination based on half the maximal response, as in none of the samples were an obvious maximum (plateau of the curve) reached.
(PDF)

Table S1 Va3.2 expression (determined by FACS), *Trav* and *Traj* segment usage, and amino acid residues at the VJ junction of $env_{124-138}$L-reactive hybridoma T cell lines generated from *Emv2*$^{+/+}$ EF4.1 mice.
(PDF)

Table S2 Va3.2 expression (determined by FACS), *Trav* and *Traj* segment usage, and amino acid residues at the VJ junction of $env_{124-138}$L-reactive hybridoma T cell lines generated from *Emv2*$^{-/-}$ EF4.1 mice.
(PDF)

Acknowledgments

We wish to thank Kim Hasenkrug for critical reading of the manuscript. We are grateful for assistance from the Division of Biological Services and the Flow Cytometry Facility at NIMR.

Author Contributions

Conceived and designed the experiments: GK JPS GRY. Performed the experiments: GRY MJYP UE MW. Analyzed the data: GRY MJYP UE MW GK. Wrote the paper: GK JPS GRY.

References

1. Hasenkrug KJ, Chesebro B (1997) Immunity to retroviral infection: The Friend virus model. Proc Natl Acad Sci U S A 94: 7811–7816.
2. Miyazawa M, Tsuji-Kawahara S, Kanari Y (2008) Host genetic factors that control immune responses to retrovirus infections. Vaccine 26: 2981–2996.

3. McMichael AJ, Borrow P, Tomaras GD, Goonetilleke N, Haynes BF (2010) The immune response during acute HIV-1 infection: clues for vaccine development. Nat Rev Immunol 10: 11–23.
4. Handunnetthi L, Ramagopalan SV, Ebers GC, Knight JC (2010) Regulation of major histocompatibility complex class II gene expression, genetic variation and disease. Genes Immun 11: 99–112.
5. Kosmrlj A, Read EL, Qi Y, Allen TM, Altfeld M, et al. (2010) Effects of thymic selection of the T-cell repertoire on HLA class I-associated control of HIV infection. Nature 465: 350–354.
6. The International HIV Controllers Study (2010) The Major Genetic Determinants of HIV-1 Control Affect HLA Class I Peptide Presentation. Science 330: 1551–1557.
7. Berger CT, Frahm N, Price DA, Mothe B, Ghebremichael M, et al. (2011) High-Functional-Avidity Cytotoxic T Lymphocyte Responses to HLA-B-Restricted Gag-Derived Epitopes Associated with Relative HIV Control. J Virol 85: 9334–9345.
8. Iglesias MC, Almeida JR, Fastenackels Sn, van Bockel DJ, Hashimoto M, et al. (2011) Escape from highly effective public CD8+ T-cell clonotypes by HIV. Blood 118: 2138–2149.
9. Price DA, Asher TE, Wilson NA, Nason MC, Brenchley JM, et al. (2009) Public clonotype usage identifies protective Gag-specific CD8+ T cell responses in SIV infection. J Exp Med 206: 923–936.
10. Turner SJ, Doherty PC, McCluskey J, Rossjohn J (2006) Structural determinants of T-cell receptor bias in immunity. Nat Rev Immunol 6: 883–894.
11. Bridgeman JS, Sewell AK, Miles JJ, Price DA, Cole DK (2011) Structural and biophysical determinants of alphabeta T-cell antigen recognition. Immunology 135: 9–18.
12. Davenport MP, Price DA, McMichael AJ (2007) The T cell repertoire in infection and vaccination: implications for control of persistent viruses. Curr Opin Immunol 19: 294–300.
13. Welsh RM, Che JW, Brehm MA, Selin LK (2010) Heterologous immunity between viruses. Immunol Rev 235: 244–266.
14. Antunes I, Tolaini M, Kissenpfennig A, Iwashiro M, Kuribayashi K, et al. (2008) Retrovirus-specificity of regulatory T cells is neither present nor required in preventing retrovirus-induced bone marrow immune pathology. Immunity 29: 782–794.
15. Ploquin MJ, Eksmond U, Kassiotis G (2011) B cells and TCR avidity determine distinct functions of CD4+ T cells in retroviral infection. J Immunol 187: 3321–3330.
16. Shimizu T, Uenishi H, Teramura Y, Iwashiro M, Kuribayashi K, et al. (1994) Fine structure of a virus-encoded helper T-cell epitope expressed on FBL-3 tumor cells. J Virol 68: 7704–7708.
17. Carson RT, Vignali KM, Woodland DL, Vignali DA (1997) T cell receptor recognition of MHC class II-bound peptide flanking residues enhances immunogenicity and results in altered TCR V region usage. Immunity 7: 387–399.
18. Rudensky AY, Preston-Hurlburt P, Hong SC, Barlow A, Janeway CA, Jr. (1991) Sequence analysis of peptides bound to MHC class II molecules. Nature 353: 622–627.
19. Sim BC, Wung JL, Gascoigne NRJ (1998) Polymorphism within a TCRAV family influences the repertoire through Class I/II restriction. J Immunol 160: 1204–1211.
20. Robertson SJ, Ammann CG, Messer RJ, Carmody AB, Myers L, et al. (2008) Suppression of acute anti-friend virus CD8+ T-cell responses by coinfection with lactate dehydrogenase-elevating virus. J Virol 82: 408–418.
21. Nair SR, Zelinskyy G, Schimmer S, Gerlach N, Kassiotis G, et al. (2010) Mechanisms of control of acute Friend virus infection by CD4+ T helper cells and their functional impairment by regulatory T cells. J Gen Virol 91: 440–451.
22. Schepers K, Toebes M, Sotthewes G, Vyth-Dreese FA, Dellemijn TA, et al. (2002) Differential kinetics of antigen-specific CD4+ and CD8+ T cell responses in the regression of retrovirus-induced sarcomas. J Immunol 169: 3191–3199.
23. Pike R, Filby A, Ploquin MJ, Eksmond U, Marques R, et al. (2009) Race between retroviral spread and CD4+ T-cell response determines the outcome of acute Friend virus infection. J Virol 83: 11211–11222.
24. King SR, Berson BJ, Risser R (1988) Mechanism of interaction between endogenous ecotropic murine leukemia viruses in (BALB/c X C57BL/6) hybrid cells. Virology 162: 1–11.
25. Marques R, Antunes I, Eksmond U, Stoye J, Hasenkrug K, et al. (2008) B lymphocyte activation by coinfection prevents immune control of friend virus infection. J Immunol 181: 3432–3440.
26. Green WR (1999) Cytotoxic T lymphocytes to endogenous mouse retroviruses and mechanisms of retroviral escape. Immunol Rev 168: 271–286.
27. Ruan KS, Lilly F (1991) Identification of an epitope encoded in the env gene of Friend murine leukemia virus recognized by anti-Friend virus cytotoxic T lymphocytes. Virology 181: 91–100.
28. Chen W, Qin H, Chesebro B, Cheever MA (1996) Identification of a gag-encoded cytotoxic T-lymphocyte epitope from FBL- 3 leukemia shared by Friend, Moloney, and Rauscher murine leukemia virus-induced tumors. J Virol 70: 7773–7782.
29. Brawand P, Biasi G, Horvath C, Cerottini JC, MacDonald HR (1998) Flow-Microfluorometric Monitoring of Oligoclonal CD8+ T Cell Responses to an Immunodominant Moloney Leukemia Virus-Encoded Epitope In Vivo. J Immunol 160: 1659–1665.
30. Hasenkrug KJ, Brooks DM, Dittmer U (1998) Critical Role for CD4+ T Cells in Controlling Retrovirus Replication and Spread in Persistently Infected Mice. J Virol 72: 6559–6564.
31. Iwanami N, Niwa A, Yasutomi Y, Tabata N, Miyazawa M (2001) Role of natural killer cells in resistance against friend retrovirus-induced leukemia. J Virol 75: 3152–3163.
32. Goff SP (2004) Retrovirus Restriction Factors. Molecular Cell 16: 849–859.
33. Nethe M, Berkhout B, van der Kuyl AC (2005) Retroviral superinfection resistance. Retrovirology 2: 52.
34. Lowy DR, Chattopadhyay SK, Teich NM, Rowe WP, Levine AS (1974) AKR murine leukemia virus genome: frequency of sequences in DNA of high-, low-, and non-virus-yielding mouse strains. Proc Natl Acad Sci U S A 71: 3555–3559.
35. Chattopadhyay SK, Lander MR, Rands E, Lowy DR (1980) Structure of endogenous murine leukemia virus DNA in mouse genomes. Proc Natl Acad Sci U S A 77: 5774–5778.
36. Jenkins MK, Chu HH, McLachlan JB, Moon JJ (2010) On the composition of the preimmune repertoire of T cells specific for Peptide-major histocompatibility complex ligands. Annu Rev Immunol 28: 275–294.
37. Huseby ES, Crawford F, White J, Kappler J, Marrack P (2003) Negative selection imparts peptide specificity to the mature T cell repertoire. Proc Natl Acad Sci U S A 100: 11565–11570.
38. Virgin HW, Walker BD (2010) Immunology and the elusive AIDS vaccine. Nature 464: 224–231.
39. Chevalier MF, Julg B, Pyo A, Flanders M, Ranasinghe S, et al. (2011) HIV-1-Specific Interleukin-21+ CD4+ T Cell Responses Contribute to Durable Viral Control through the Modulation of HIV-Specific CD8+ T Cell Function. J Virol 85: 733–741.
40. Yue FY, Lo C, Sakhdari A, Lee EY, Kovacs CM, et al. (2010) HIV-Specific IL-21 Producing CD4+ T Cells Are Induced in Acute and Chronic Progressive HIV Infection and Are Associated with Relative Viral Control. J Immunol 185: 498–506.
41. Vingert B, Perez-Patrigeon S, Jeannin P, Lambotte O, Boufassa F, et al. (2010) HIV controller CD4+ T cells respond to minimal amounts of gag antigen due to high TCR avidity. PLoS Pathog 6: e1000780.
42. Ortiz AM, Klatt NR, Li B, Yi Y, Tabb B, et al. (2011) Depletion of CD4+ T cells abrogates post-peak decline of viremia in SIV-infected rhesus macaques. J Clin Invest 121: 4433–4445.
43. Portis JL (1994) Endogenous retroviral envelope antigens recognized by B lymphocytes during graft-versus-host reaction. Tohoku J Exp Med 173: 83–89.
44. Gras S, Chen Z, Miles JJ, Liu YC, Bell MJ, et al. (2010) Allelic polymorphism in the T cell receptor and its impact on immune responses. J Exp Med 207: 1555–1567.
45. Kjer-Nielsen L, Clements CS, Purcell AW, Brooks AG, Whisstock JC, et al. (2003) A structural basis for the selection of dominant alphabeta T cell receptors in antiviral immunity. Immunity 18: 53–64.
46. Sabatino JJ, Jr., Huang J, Zhu C, Evavold BD (2011) High prevalence of low affinity peptide-MHC II tetramer-negative effectors during polyclonal CD4+ T cell responses. J Exp Med 208: 81–90.
47. Munitic I, Decaluwe H, Evaristo C, Lemos S, Wlodarczyk M, et al. (2009) Epitope specificity and relative clonal abundance do not affect CD8 differentiation patterns during lymphocytic choriomeningitis virus infection. J Virol 83: 11795–11807.
48. Ebert PJR, Jiang S, Xie J, Li QJ, Davis MM (2009) An endogenous positively selecting peptide enhances mature T cell responses and becomes an autoantigen in the absence of microRNA miR-181a. Nat Immunol 10: 1162–1169.
49. Pothlichet J, Mangeney M, Heidmann T (2006) Mobility and integration sites of a murine C57BL/6 melanoma endogenous retrovirus involved in tumor progression in vivo. Int J Cancer 119: 1869–1877.
50. Stoye JP (1999) The pathogenic potential of endogenous retroviruses: a sceptical view. Trends Microbiol 7: 430.
51. Baudino L, Yoshinobu K, Morito N, Santiago-Raber ML, Izui S (2010) Role of endogenous retroviruses in murine SLE. Autoimmun Rev 10: 27–34.
52. Nakagawa K, Harrison LC (1996) The potential roles of endogenous retroviruses in autoimmunity. Immunol Rev. pp 193–236.
53. Christensen T (2005) Association of human endogenous retroviruses with multiple sclerosis and possible interactions with herpes viruses. Rev Med Virol 15: 179–211.
54. Perl A, Nagy G, Koncz A, Gergely P, Fernandez D, et al. (2008) Molecular mimicry and immunomodulation by the HRES-1 endogenous retrovirus in SLE. Autoimmunity 41: 287–297.
55. Kleiman A, Senyuta N, Tryakin A, Sauter M, Karseladze A, et al. (2004) HERV-K(HML-2) GAG/ENV antibodies as indicator for therapy effect in patients with germ cell tumors. Int J Cancer 110: 459–461.
56. Humer J, Waltenberger A, Grassauer A, Kurz M, Valencak J, et al. (2006) Identification of a melanoma marker derived from melanoma-associated endogenous retroviruses. Cancer Res 66: 1658–1663.
57. Contreras-Galindo R, Kaplan MH, Markovitz DM, Lorenzo E, Yamamura Y (2006) Detection of HERV-K(HML-2) viral RNA in plasma of HIV type 1-infected individuals. AIDS Res Hum Retroviruses 22: 979–984.
58. Garrison KE, Jones RB, Meiklejohn DA, Anwar N, Ndhlovu LC, et al. (2007) T Cell Responses to Human Endogenous Retroviruses in HIV-1 Infection. PLoS Pathog 3: e165.

59. Fellay J, Shianna KV, Ge D, Colombo S, Ledergerber B, et al. (2007) A Whole-Genome Association Study of Major Determinants for Host Control of HIV-1. Science 317: 944–947.
60. Takahashi Y, Harashima N, Kajigaya S, Yokoyama H, Cherkasova E, et al. (2008) Regression of human kidney cancer following allogeneic stem cell transplantation is associated with recognition of an HERV-E antigen by T cells. J Clin Invest 118: 1099–1109.
61. Mombaerts P, Iacomini J, Johnson RS, Herrup K, Tonegawa S, et al. (1992) RAG-1-deficient mice have no mature B and T lymphocytes. Cell 68: 869–877.
62. Philpott KL, Viney JL, Kay G, Rastan S, Gardiner EM, et al. (1992) Lymphoid development in mice congenitally lacking T cell receptor alpha beta-expressing cells. Science 256: 1448–1452.
63. Casanova JL, Romero P, Widmann C, Kourilsky P, Maryanski JL (1991) T cell receptor genes in a series of class I major histocompatibility complex-restricted cytotoxic T lymphocyte clones specific for a Plasmodium berghei nonapeptide: implications for T cell allelic exclusion and antigen-specific repertoire. J Exp Med 174: 1371–1383.
64. Martin RM, Brady JL, Lew AM (1998) The need for IgG2c specific antiserum when isotyping antibodies from C57BL/6 and NOD mice. J Immunol Methods 212: 187–192.
65. Livak KJ, Schmittgen TD (2001) Analysis of relative gene expression data using real-time quantitative PCR and the 2(−Delta Delta C(T)) Method. Methods 25: 402–408.