Genes and Immunity (2012), 1–12

www.nature.com/gene

ORIGINAL ARTICLE

Human B-cell ontogeny in humanized NOD/SCID γc^{null} mice generates a diverse yet auto/poly- and HIV-1-reactive antibody repertoire

H Chang[1,2,3,5], S Biswas[1,2,5], AS Tallarico[1,2], PTN Sarkis[1,2], S Geng[1,2,4], MM Panditrao[1,2], Q Zhu[1,2] and WA Marasco[1,2]

Characterization of the human antibody (Ab) repertoire in mouse models of the human immune system is essential to establish their relevance in translational studies. Single human B cells were sorted from bone marrow and periphery of humanized NOD/SCID γc^{null} (hNSG) mice at 8–10 months post engraftment with human cord blood-derived CD34$^+$ stem cells. Human IG variable heavy (V_H) and kappa ($V_κ$) genes were amplified, cognate V_H–$V_κ$ gene-pairs assembled as single-chain variable fragment-Fc Abs (scFvFcs) and functional studies were performed. Although overall distribution of V_H genes approximated the normal human Ab repertoire, analysis of the V_H-third complementarity-determining regions in the mature B-cell subset demonstrated an increase in length and positive charges, suggesting autoimmune characteristics. Additionally, >70% of $V_κ$ sequences utilized $V_κ$4-1, a germline gene associated with autoimmunity. The mature B-cell subset-derived scFvFcs displayed the highest frequency of autoreactivity and polyspecificity, suggesting defects in checkpoint control mechanisms. Furthermore, these scFvFcs demonstrated binding to recombinant HIV envelope corroborating previous observations of poly/autoreactivity in anti-HIVgp140 Abs. These data lend support to the hypothesis that anti-HIV broadly neutralizing antibodies may be derived from auto/polyspecific Abs that escaped immune elimination and that the hNSG mouse could provide a new experimental platform for studying the origin of anti-HIV-neutralizing Ab responses.

Genes and Immunity advance online publication, 17 May 2012; doi:10.1038/gene.2012.16

Keywords: humanized mouse; single B cell; antibody repertoire; autoreactive; checkpoint control

INTRODUCTION

Advancement in high-throughput screening techniques has led to the recent discovery of several highly potent broadly neutralizing antibodies (BnAbs) against HIV[1–4] and Influenza A[5] recovered from peripheral blood (PB)-derived B cells of infected individuals; however, the occurrence of these BnAbs is extremely rare. This has spurred a renewed interest in 'rational vaccine design' where it may be feasible to analyze the patient's antibodyome in order to obtain insight into the ontogeny of the BnAbs.[6] This information may then be used to design candidate vaccines in order to improve BnAb responses.[4]

Given the cost and ethical constraints of using human subjects for investigative vaccine studies, there is a growing need for a predictive and surrogate system to study human Ab evolution at the single-cell level. 'Humanized' mouse models are increasingly being used to study human immunity, developmental and disease processes.[7,8] Newer mouse models deficient in the expression of the interleukin-2 receptor γ-chain (γc^{null}), including NOD/SCID γc^{null} (NSG), BALB/c-Rag2$^{-/-}$ γc^{null} and H2^d- Rag2$^{-/-}$ γc^{null} mice, support the development of a multilineage human hemato-lymphoid system following transplantation with fetal or adult hematopoietic stem cells (HSCs). Additionally, these engrafted γc^{null} mice exhibit normal life spans, unlike previous models, thus enabling long-term studies.[9] In spite of these favorable advances, the adaptive Ab responses of these animals are weak with barely detectable secondary responses including class switching and affinity maturation.[7] Growth-factor supplementation with human BLyS and T-cell cytokines in order to support growth and differentiation of the transplanted cells has resulted in only marginal improvement.[10,11] Treatment of these mice with human cytokines and other costimulatory/growth factors delivered by a variety of techniques are being actively investigated to further improve human immune system development.[12]

Clonal diversity and immune tolerance are two major cornerstones of an effective Ab response that must also be considered in evaluating these mice as a relevant platform system to study human Ab responses. Several laboratories have studied immune repertoire complexity in hNSG mice by TCR CDR3 spectratyping,[9] BCR V_H-third complementarity-determining regions (H-CDR3s) immunoscope analysis[13] and multiplex PCR of variable (V)–J rearrangements of TCRβ and H-CDR3[14] and have concluded that both repertoires show levels of diversity comparable to those of humans. In addition, there have been two reports that analyzed the diversity of the immunoglobulin (IG) repertoire with a focus on only the variable heavy (V_H)4 family in NOD/SCID and NOD/SCID/β2 m^{null} mice.[15,16] However, a systematic study in which the diversity of the human B-cell repertoire is analyzed via genetic and functional analysis of the V, diversity (D) and joining (J) gene

[1]Department of Cancer Immunology and AIDS, Dana-Farber Cancer Institute, Boston, MA, USA and [2]Department of Medicine, Harvard Medical School, Boston, MA, USA. Correspondence: Dr WA Marasco, Department of Cancer Immunology and AIDS, Dana-Farber Cancer Institute, JFB824, Boston, MA 02215, USA.
Email: wayne_marasco@dfci.harvard.edu
[3]Current address: Department of Biochemistry, School of Basic Medicine, Hebei Medical University, Shijiazhuang, China.
[4]Current address: Tian Guang Shi Biotechnology Co. LTD, Beijing, China.
[5]These authors contributed equally to this work.
Received 3 January 2012; revised 15 March 2012; accepted 4 April 2012

segments of the IG heavy- and light-chain genes has not been performed in hNSG mice. Analysis of immune tolerance in hNSG mice by evaluation of the physiological checkpoint control mechanisms that are normally operative during B-cell development[17,18] has also not been reported (see Mouquet *et al.*[17] for schema).

In the present study, analysis of V_H and kappa (V_κ) gene arrangements in hNSG-derived single human B cells sorted at different developmental stages was performed. Nucleotide and amino-acid sequence analysis of the heavy chain genes indicated the presence of a diverse Ab repertoire; however, characterization of H-CDR3 regions and a specific restriction in the V_κ repertoire suggested an autoreactive potential. This was further confirmed by the functional studies where single-chain variable fragment-Fc Abs (scFvFcs) cloned from single B cells were found to exhibit binding to self-antigens. Intriguingly, many autoreactive clones also displayed affinity for HIV-1 envelope protein gp140 (HIV-1gp140). These data lend support to the contemporary hypothesis that anti-HIV BnAbs may be derived from auto/polyspecific Abs that escaped immune elimination.[19,20] Thus, the defects in immune tolerance in these hNSG mice may provide a unique model to study the development of anti-HIV BnAbs, where ancestral origins of Abs can be determined and the existing HIV-reactive Ab clones can be modulated via experimental infection or immunization.

RESULTS

B-cell development in long-term engrafted hNSG mice

Significant numbers (>10% of total lymphocytes) of human $CD45^+$ cells could be detected in the PB of hNSG mice between 1–2 months post HSC injection; however, total human $CD3^+$ T cells remained low (<5% of total human $CD45^+$ cells) during that time. At 6 months post HSC injection, the human $CD45^+$ frequency in the PB of the hNSG mice was $43.87 \pm 10.11\%$ ($n = 8$) of total lymphocytes, of which $41.49 \pm 22.75\%$ stained for hCD3 (T cells) and $53.28 \pm 21.5\%$ for hCD19 (B cells). Of the total $CD19^+$ human B cells, $53.64 \pm 12.69\%$ expressed the CD5 surface antigen, which was higher than the reported level of circulating $CD19^+$ $CD5^+$ cells (~15%) found in humans.[21] Profiles of the different $CD19^+$ B-cell subpopulations in the hNSG mice at >6 months post HSC transplantation at the time of single B-cell sorting are shown in Figure 1. B cells at all major stages of the developmental pathway (early immature, immature, new emigrant and mature naïve) were detected, but $CD27^+$ B cells, associated with a memory phenotype,[22] were notably absent. Overall, the data demonstrate that the human B-cell developmental pathway is recapitulated in the hNSG mice; however, the memory B-cell compartment is not or only sparsely populated.

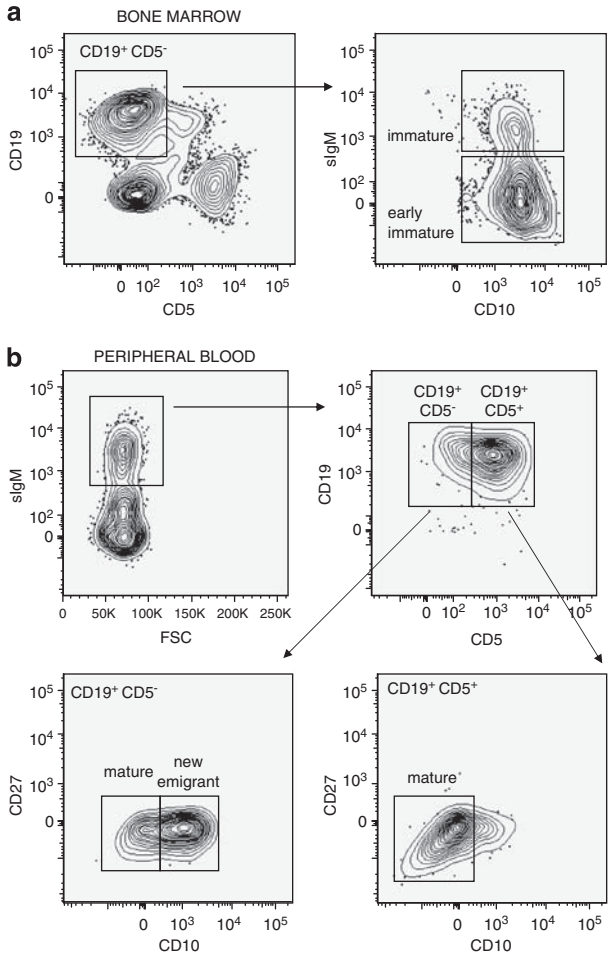


Figure 1. Sorting criteria of human B-cell subsets from humanized NOD/SCID γc^{null} (hNSG) mice. (**a**) Single-cell suspension was prepared from the bone marrow of a representative single hNSG mouse and cells in the lymphoid gate (not shown) were further gated for $CD19^+CD5^-$ expression (left). Cells were then plotted for surface IgM (sIgM) and CD10 expression (right), and single B cells from the early immature ($sIgM^-CD10^+$) and immature ($sIgM^+$ $CD10^+$) subsets were sorted. (**b**) PBMCs were prepared from three hNSG mice, all transplanted with the same donor and were gated for lymphoid population (not shown). These lymphocytes were gated on IgM^+ cells (middle left) and further characterized as $CD19^+$ $CD5^-$ or $CD19^+CD5^+$. Sorting was performed in the $CD19^+CD5^-$ gate to obtain single B cells from the new emigrant ($CD10^+CD27^-$) and mature ($CD10^-CD27^-$) B-cell subsets (lower left). Similarly, the $CD5^+CD19^+$-gated cells were single sorted to obtain $CD5^+$ mature B cells ($CD10^-CD27^-$) (lower right). Single-cell sortings were performed between 8–10 months post engraftment of the hNSG mice. All Abs used were specific for human cell-surface markers.

Analysis of human V_H sequences

Table 1 summarizes the total data set analyzed in the present study. Single B cells from different B-cell subpopulations were sorted and IG gene sequences were obtained as described in Materials and methods. In total, 257 and 264 sequences unique for human IG heavy and light chains, respectively, were obtained and further analyzed. Intriguingly, in the first round PCR reactions to amplify V_κ and lambda (V_λ) light-chain genes, only those sequences that corresponded to V_κ light-chain gene segments were recovered. PCR-mediated primer bias was ruled out by using the same primer mix to amplify heavy- and light-chain genes from single-sorted mature B cells from a normal human donor. From a total of 192 single-sorted human PB B cells, 137 V_H, 44 V_κ and 33 V_λ unique and productive gene segments could be amplified (Supplementary Figure 1). The reason for this observed exclusivity of the V_κ genes is currently under investigation and may include factors, such as V_λ transcript instability, greater V_λ variability and V_κ gene rearrangement bias in the hNSG mice.

Individual IG sequences were aligned with the International Immunogenetics Information System (IMGT) database and the overall germline usage was determined. Human V_H gene segments are divided into seven families, named V_H1–7 with 51 known functional genes and >40% of the genes representing the V_H3 family (IMGT). V_H sequences from all, except V_H2 and V_H6, families were represented in the hNSG mice, and the sequence usage from different V_H gene families across the B-cell subpopulations is shown in Figure 2, upper panel. In comparison with the human V_H repertoire found in normal adults,[23,24] gene segments corresponding to the V_H3 family were most frequently utilized

followed by V_H4 in most of the B-cell subpopulations. An exception was observed in new emigrant B cells where the V_H3 usage (17.4%) was marginally lower than that of V_H4 (22.5%). Additionally, V_H7 gene segments were detected at a frequency higher than normally found (~2% in the human V_H repertoire, IMGT), particularly in the early immature stage of B-cell development where it was ~12% of the total V_H sequences (Supplementary Table 1). No other significant differences were found in the distribution of the V_H gene families across the different B-cell subsets. These observations suggest that the V_H repertoire in the hNSG mouse is quite diverse with 36 out of 51 known functional genes found at least once in the various subpopulations (Supplementary Table 1). Among the individual gene segments, V_H4-34, V_H3-30 and V_H1-2 were utilized most often. V_H4-34 was most frequently observed in the repertoire of all the B-cell subsets, except for the immature subpopulation where V_H3-30 had the highest occurrence.

Analysis of V_H rearrangement and H-CDR3 composition

The complexity of the IG repertoire is enhanced by gene rearrangement of V_H with D_H and J_H segments and $V_{\kappa/\lambda}$ with $J_{\kappa/\lambda}$ segment for the light chain. Therefore, the utilization of the D_H and J_H gene segments across the different B-cell subpopulations was analyzed next (Figure 2, middle and lower panels, respectively). The D_H genes are classified into seven families, D_H1–D_H7. As reported for humans,[25] the usage of the D_H3 family in the hNSG mice was significantly more frequent than the other families ($P<0.005$, Exact Fisher–Freeman–Halton's test). An increase in the usage of D_H1 and D_H2 segments along with the $CD5^-$ B-cell developmental pathway was observed (early immature < immature < new emigrant < mature). There were differences between the mature $CD5^-$ and $CD5^+$ B cells in terms of D_H gene-segment usage, particularly with regards to the utilization of D_H2 (lower in $CD5^+$ B cells), D_H3 (higher in $CD5^+$ B cells) and D_H5 gene segments (absent in $CD5^+$ B cells). Rearrangement of D_H gene segments may occur either through inversion or deletion processes, which determines the D_H reading frame utilized. An analysis of the usage of the D_H reading frame indicated a significantly higher utilization of reading frame 2, found on an average of 51% of all the sequences across all B-cell subsets in the hNSG mice ($P<0.005$, Fisher's exact test, data not shown). This draws a parallel to that observed for normal human D_H gene rearrangements where deletion was found to be the favored

Table 1. Numeric representation of the human IG repertoire analyzed from humanized NSG mice

Source	*Human $CD5^-$/$CD5^+$ B-cell subsets*[a]	*Number of cells sorted*	*Number of unique IG sequences*			*Number of scFvfc*
			VH	*Vκ*	*Cognate VH–Vκ pair*[b]	
Bone marrow	Early immature	72	60	49	44	31
	Immature	72	39	48	34	27
Peripheral blood	New emigrant	96	50	51	42	36
	Mature	96	59	64	57	57
	Mature ($CD5^+$)	72	49	52	42	25
Total		408	257	264	219	176

[a]Except the 'mature ($CD5^+$)' B cells that were sorted following gating on a $CD19^+CD5^+$ cell population, all the other human B-cell subsets were sorted on a $CD19^+CD5^-$ gate (see Figure 1). [b]Inclusive of numbers indicated for individual V_H and V_κ sequences.

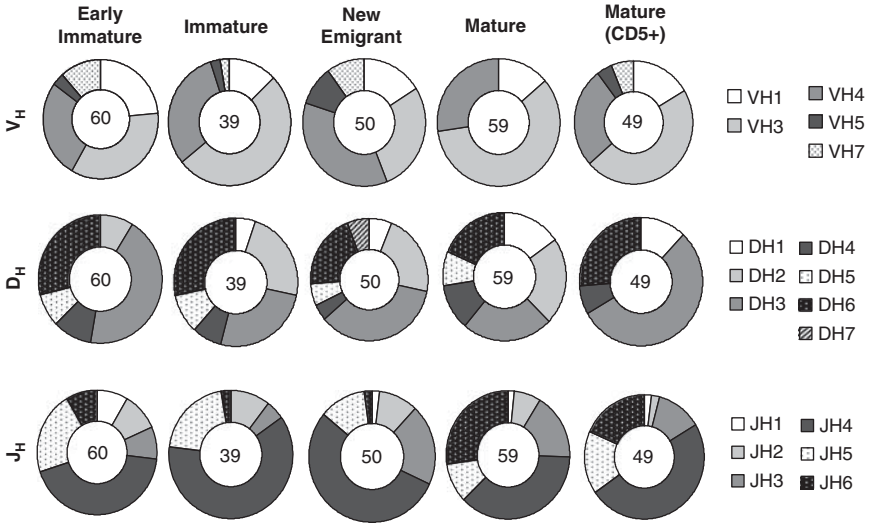
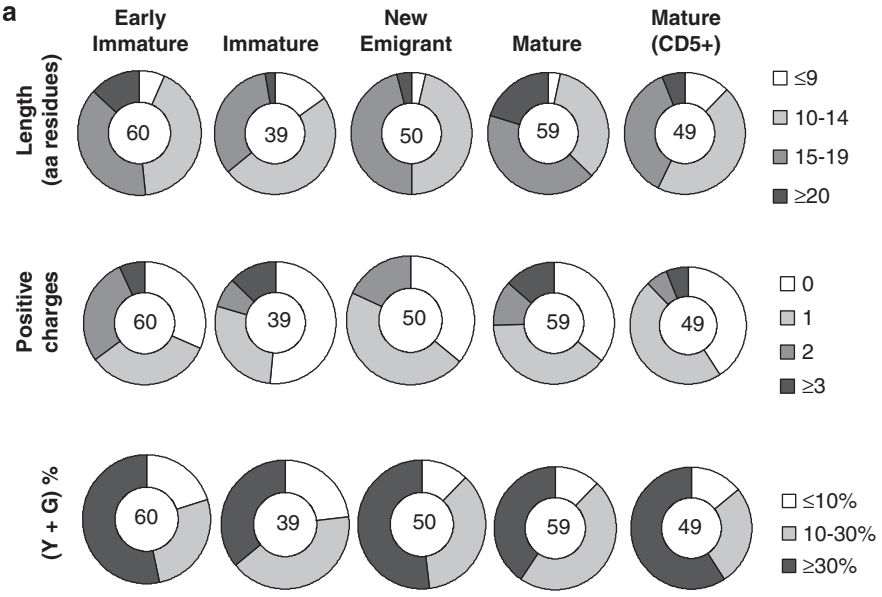


Figure 2. Analysis of the human IG V_H repertoire in hNSG mice. The frequency of utilization of the different V_H gene families (upper panel), D_H (middle) and J_H (lower) gene families across the different B-cell subpopulations (indicated above pie charts) is shown. The number in the center of each pie chart indicates the total sample size for the particular B-cell subset. Any significant difference in the overall usage of specific gene families in each B-cell population was assessed by Fisher's test. Usage of V_H3 family was significantly higher in all cell subsets ($P<0.005$), except the new emigrant B cells. Sequence identities and categorizations into different V_H gene families were performed using the IMGT database.



b

N-Nucleotide insertions (bp)	Early Immature	Immature	New Emigrant	Mature	Mature (CD5+)
N1 (V-D)	7.8	6.4	7.6	7.3	5.7
N2 (D-J)	6.7	5.4	4.4	8.0	4.7
JH exonuclease activity (bp)					
Median	3.28	3.33	3.08	3.98	4.29
Range	0-13	0-13	0-9	0-14	0-13

Figure 3. Composition of the third complementarity region in the human IG heavy chain (H-CDR3) regions. (**a**) The pie charts indicate length (amino-acid residues) (upper panel), total positive-charge content (middle) and percentage composition of total tyrosine and glycine residues in each H-CDR3 region analyzed (lower). The different B-cell subpopulations from which the samples were derived are indicated above each pie chart. The number in the center of each pie chart indicates the total sample size for the particular B-cell subset. (**b**) Average N-nucleotide insertions and J_H exonuclease activity across the different B-cell subsets are listed. All the analyses were performed using the IMGT database.

process along with a lower frequency of the use of reading frame 3.[25] The distribution of J_H genes clearly showed a higher bias for J_H4 usage ($P<0.005$, Fisher's exact test). No significant difference was observed between the mature B-cell subsets ($P>0.005$, Fisher's exact test). However, distribution of J_H6 usage compared with all other J_H segments (except J_H3) was found to be significantly higher in the mature B-cell subsets ($CD5^+$ and $CD5^-$) compared with the BM-derived B cells (early immature and immature B cells) ($P<0.005$, Fisher's exact test).

The H-CDR3 region is considered to be an important determinant that imparts Ab specificity as well as affinity. The overall composition of H-CDR3 regions in terms of length, positive-charge content, N-nucleotide insertions and exonuclease activity was analyzed (Figure 3). Extremely long (>20 amino acids) H-CDR3 regions were more frequent in the $CD5^-$ mature B-cell subset, which may be explained by the higher utilization of the J_H6 segment by this group of cells that results in longer CDR3.[25] Additionally, this same B-cell subset carried more cells with 2+ and 3+ positively charged residues in the H-CDR3 compared with the $CD5^+$ B subset. Although increases in length and positive charges in the H-CDR3 region are expected in the IG repertoire from BM-derived immature B-cell populations, these are less frequently observed when the B cells mature and migrate to the periphery. Thus, the mature B-cell subset in the hNSG mice with longer and increased positive charges in their H-CDR3 regions is a major and previously unrecognized difference from their counterpart in normal humans. No significant difference in N-insertions and J_H exonuclease activity between the B-cell subpopulations was observed ($P<0.005$, Fisher's exact test) (Figure 3b). A bias in tyrosine and glycine composition in the H-CDR3 regions (40–50% of total residues) that is observed in normal humans [26] was recapitulated in the hNSG mice (Figure 3a).

Presence of hypermutated V_H sequences

Hypermutation in the V_H sequences was analyzed by comparing against the closest germline sequences, and the B-cell subpopulations were classified as those with identical or highly homologous (99–100% homology), moderately mutated (94–99%) and highly mutated (≤94%) sequences as described earlier for the analysis of human IG V_H repertoire[24] (Table 2). Accumulation of hypermutated V_H sequences (≤99%) was significantly lower in the mature B-cell subsets in the periphery compared with the earlier stages in development in the bone marrow ($P<0.005$, Fishers' exact test). In contrast to an earlier report,[24] where IgM^+CD5^+ human B cells were found to accumulate significantly higher number of mutations compared with the IgM^+CD5^- B cells, we did not observe any significant difference between the $CD5^-$ and $CD5^+$ mature B-cell subsets ($P<0.005$, Fisher's exact test). Moreover, the V_H repertoire in the peripheral B cells was notably devoid of highly mutated sequences, for example, displaying <94% homology to germline sequences. Analysis of the mutations in the framework

Table 2. Distribution of hypermutated V_H genes in hNSG mice

B-cell subpopulations	*Homology to germline sequence*								
	100–99%			*99–94%*			*⩽94%*		
	# (%) Sequences[a]	*nt/aa mutations*[b]		*# (%) Sequences*	*nt/aa mutations*		*# (%) Sequences*	*nt/aa mutations*	
		FR	*CDR*		*FR*	*CDR*		*FR*	*CDR*
Early immature	36 (60%)	0.46/0.32	0.07/0.05	20 (33%)	1.77/1.14	1.18/0.79	4 (7%)	0.44/0.23	0.53/0.30
Immature	26 (67%)	0.68/0.38	0.16/0.08	8 (21%)	0.97/0.57	0.51/0.32	5 (13%)	1.46/0.86	1.14/0.54
New emigrant	33 (66%)	0.30/0.22	0.18/0.08	17 (34%)	1.56/1.14	1.14/0.72	0 (0%)	—	—
Mature	50 (85%)	0.60/0.38	0.17/0.10	9 (15%)	0.36/0.24	0.14/0.05	0 (0%)	—	—
Mature ($CD5^+$)	41 (84%)	0.38/0.23	1.18/0.79	8 (16%)	0.68/0.49	0.47/0.30	0 (0%)	—	—

Abbreviations: aa, amino acid; CDR, complementarity-determining regions; FR, framework region; nt, nucleotide. [a]Total number and corresponding percentage (in parentheses) of total sequences per B-cell subset. [b]Total number of nucleotide/amino-acid changes in the framework region (FR) and H-CDR3 (CDR) regions per V_H sequence.

regions and H-CDR3, both at the nucleotide and amino-acid residue levels, revealed the occurrence of both silent as well as replacement mutations (Table 2).

Potential V_H replacement in hNSG mice

The cryptic recombination signal sequence (TACTGTG)[27] at the 3′-end of rearranged V_H genes allowed the detection of potential V_H replacement products. The cryptic recombination signal sequence is the binding motif for the recombinase-activating gene family of enzymes that mediate V_H replacement, a process in which a secondary arrangement of an upstream V_H sequence occurs in an already formed V_H–D_H–J_H gene.[27] Remarkably, 13 potential V_H replacements in the 257 V_H sequences were detected (Table 3). This secondary rearrangement process was not confined within the early stages of B-cell development and did not only involve V_H4 genes as reported earlier.[28,29] In order to confirm the validity of the sequences scored as the products of V_H replacement, a similar scan was performed on the D_H–J_H junction of the V_H sequences. One major criteria of such replacement products, as shown in Table 3, is that they often encode highly charged amino-acid residues (R/D/E).[30] No such sequences were identified in the D_H–J_H region of the V_H genes.

Analysis of human V_κ sequences and cognate V_H–V_κ pairs

Figure 4a presents an analysis of the V_κ sequences recovered from the different human B-cell subsets obtained in the hNSG mice. The human V_κ locus contains 44 functional gene segments divided into five gene families: $V_\kappa1$–$V_\kappa5$ (IMGT). Only a single member is known for each of the $V_\kappa4$ and $V_\kappa5$ families, V_κ 4-1 and V_κ 5-2, respectively (IMGT). At least 24 unique V_κ gene segments, representing four out of five V_κ gene families in the B-cell subpopulations, were detected with a distinct bias towards the utilization of $V_\kappa4$-1 (Supplementary Table 2). Notably, the $V_\kappa4$-1 gene segment is located most proximally in the human V_κ locus on chromosome 2; however, the significance of this location with regards to the overutilization of this gene in the hNSG mice is presently unclear. $V_\kappa5$ family genes remained undetected in the hNSG mice, which may be attributed to the limited sample size. An analysis of the J_κ sequences demonstrated the utilization of all known J_κ segments in the human IG repertoire ($J_\kappa1$–$J_\kappa5$) with an overrepresentation of the $J_\kappa4$ gene segments across all B-cell subpopulations (Figure 4a).

V_H–V_κ pairs derived from single B cells were interrogated for the frequency of pairing of individual heavy and light chains across different cell subsets (Figure 4b), and a complete list of the pairings is shown in Supplementary Figure 2. Given that each V_H–V_κ pair represents a unique Ab, the profile of heavy- and light-chain pairing across the different B-cell subsets can provide information on Ab evolution, clonality, editing and other physiological processes that shape the IG repertoire. Interestingly, as represented in Figure 4b, a significant decrease in V_H-pairing with $V_\kappa4$-1 in mature B cells ($CD5^+$ and $CD5^-$) was found when compared with BM-derived B-cell subsets ($P<0.005$, Fisher's exact test). Given that the usage of $V_\kappa4$-1 has been associated with autoreactivity and anti-DNA auto-Abs,[31] the decreasing frequency of $V_\kappa4$-1 usage with increasing B-cell maturation suggests that checkpoint control mechanisms, which are associated with peripheral tolerance, are at least partially intact.

IG levels and auto/polyreactive Abs in hNSG mice

Total human IgM and IgG concentrations in hNSG mouse serum ranged 15–50 and 20–90 $\mu g\,ml^{-1}$, respectively, which were ~10–50- and ~100–200-fold lower than normal human serum IgM and IgG levels, respectively. By comparison, in newborns, the Ab levels approximate to that found in the adult human, albeit they increase with time, plateauing at ~1 year for IgM and ~5–6 years for IgG.[32] It has also been shown that in human bone marrow transplants, serum IG levels return within normal levels by day 90 post transplant.[33]

The physiological presence of auto/polyreactive Abs was next tested by analyzing the binding of serum IgM and IgG Abs using enzyme-linked immunosorbent assay (ELISA) to a panel of antigens, including cardiolipin, insulin, single-stranded DNA, double-stranded DNA and lipopolysaccharide, and the overall reactivity was compared with normal human sera. As shown in Figure 5, the hNSG mice sera showed significantly higher binding levels over that of normal human sera to all antigens tested, and this auto/polyreactivity was almost entirely restricted to the IgM fraction ($P<0.005$, Mann–Whitney test). An increased level of binding to lipopolysaccharide by both normal human IgG and IgM may be attributed to 'natural Abs' present in the serum.[34]

To investigate this auto/polyreactivity in more detail, additional studies were performed on recombinant Abs recovered from single B cells. As shown in Table 1, 176/219 cognate V_H–V_κ pairings, representing all different human B-cell subsets, were successfully assembled into scFvs (80% efficiency), expressed as scFvFcs and purified for functional studies. To determine the autoreactivity of these Abs, HEp-2 (human epidermoid cancer cell) ELISA, which is a standard clinical assay for detection of anti-nuclear Abs (ANAs), was performed. As shown in Figure 6a, autoreactive ANAs were detected in all B-cell subsets. Circa 50% of

Table 3. Potential V_H replacement analysis of the human IG sequences from various B-cell subsets[a]

Sequence	*V_H Donor family*	*V_H Sequence*	*V-D (N1)*	*D_H*	*D_H*	*J_H*	*R_{Fam}*
Early immature							
NE14-1	4–59	5′-TACTGTGcgagag-3′	5′-CTTCCT**TAGATA**ATAGTGAACACCC-3′ (SL**D**NSEHP)	5′-TATTGTGGTGGTGACTGCTATT-3′	2–21	2	1–45
NE31	*3–13*	5′-TACTGTGcaagag-3′	5′-GACCC**CAAAA**TATAGA-3′ (GP**Q**N**I**D)	5′-TGGCTACGGTGACTACGTCGG-3′	4–17	4	3–43
Immature							
NI47-1	4–61	5′-TACTGTGcgagag-3′	5′-CTTCCTT**AGATA**GTAGTGAACACCC-3′ (SL**D**SSEHP)	5′-TATTGTGGTGGTGACTGCTATT-3′	2–21	2	1–45
New emigrant							
NN37-2	*3–11*	5′-TACTGTGcgagaa-3′	5′-T**CTAGA**AG-3′ (PL**E**)	5′-GGGGGAG-3′	3–16	4	3–72
NN70	4–39	5′-TACTGTGcgag-3′	**CACGGA**CCAAGGGGGGCGA-3′ (T**D**QGGR)	5′-ATGATAGTAGTGGTTAT-3′	3–22	4	2–26
NN80	*4–34*	5′-TACTGTGcgagag-3′	5′-CTTCCTT**AGATA**ATAGTGAACACCC-3′ (SL**D**NSEHP)	5′-TATTGTGGTGGTGACTGCTATT-3′	2–21	2	3–43
Mature							
NM10-1	4–31	5′-TACTGTGcgagaga-3′	5′-T**CACGG**CCACTCACGCGCCGCTAGTCC-3′ (**HG**HSRAASP)	5′-GGGTATAGCAGCAGCTGGTAC-3′	6–13	3	2–26
NM48-4	*3–23*	5′-TTCAATC**CGGATT**TGCCCAGCATTCCCGGCG-3′ (Q**SGF**AQHSRR)	5′-ATGGACCGGGG-3′	3–10	3	2–26
NM54-1	*1–2*	5′-TACTGTGcgagag-3′	5′-GCGTG**GGCAG**TC-3′ (V**GS**H)	5′-ATTACTATGATAGTAGTGGTTA-3′	3–22	4	1–58
NM7-2	*3–33*	5′-TACTGTGcgagaga-3′	5′-TTTCA**GGCAGA**GGTTG-3′ (F**RQR**L)	5′-GTGGCTGGTACCAG-3′	6–19	5	1–58
Mature ($CD5^+$)							
N1M26	4–34	5′-TACTGTGcgaga-3′	5′-CGGAACTCAA**AGAAAA**-3′ (RNS**KK**N)	5′-ATTACTATGGTTCGGGGAG-3′	3–10	6	4–28
N1M43-1	3–30	5′-TACTGTGcgaaag-3′	5′-CTC**AGAAA**-3′ (**QK**)	5′-GGTTCGGGGAGTTATTATA-3′	3–10	3	3–23 4-28
N1M51-1	7–4–1	5′-TACTGTGcgagag-3′	5′-T**CACAC**GTA-3′ (**VTS**)		ND	4	2–5

Abbreviations: D, diversity; J, joining; V, variable. Italics in the V_H donor column represent gene segments that are downstream to the recipient and thus correspond to a potential non-conventional V_H replacement process. [a]Nucleotide sequence nomenclature used in the table is from IMGT. The listed sequences start from the first nucleotide of the cryptic recombination signal sequence (indicated in upper case), through the V–D junction into the D_H region. The V_H donor family, D_H family, J_H family and replacement (recipient) family (R_{Fam}) designations are listed as well. The bold and underlined nucleotides and amino acids in the V–D (N1) region match the 3′-end of the V_H germline of the replacement family and the charged amino acids are shown in red. V_H replacement sequences represent 5% of the total number of Abs analyzed.

the scFvFcs from all B-cell subsets were scored as autoreactive, with the exception of the $CD5^-$ mature B-cell subset where autoreactivity was observed in a significantly higher percentage of the Abs (80.7%) compared with the other B-cell subsets ($P<0.005$, Mann–Whitney test). The reactivity of all the 176 scFvFcs against each antigen is shown in Figure 6b. There was a significant increase in the median reactivity levels of antigen binding in the mature B-cell subsets when compared with that observed for the B cells at earlier stages in development ($P<0.005$, Mann–Whitney test). No significant difference was observed between the $CD5^+$ and $CD5^-$ mature B-cell subsets. All the polyreactive Abs also scored positive in the ANA assay. Additionally, scFvFcs demonstrating high-level reactivity to one antigen also bound strongly to other antigens. Thus, the results of both the serological and single B-cell-derived binding studies provide evidence for the retention of a substantial pool of auto/polyreactive B-cell clones, centrally in the bone marrow and particularly in the periphery, suggesting that the first two checkpoint control steps (between early immature to immature and new emigrant to mature) involved in the clearance of autoreactive clones are impaired.[17] ELISA-binding studies using a comparable mouse cell line to test scFvFc autoreactivity to mouse ANA antigens may provide additional information on autoreactivity, given the development of these human B cells in the murine host.

A total of 22 scFvFcs, shown in Figure 6b, were scored based on their absolute meso scale discovery (MSD) values (>3-fold background as obtained from reactivity of buffer alone to antigen-coated wells) as low, medium and highly polyreactive based on the relative reactivity of 4E10 to the same antigens (<20%, ⩾20–60% and ⩾60% of 4E10 scFvFc reactivity, respectively). Of these 22 scFvFcs, 2 were from the early immature B-cell subset, 1 from immature, 2 from new emigrant, 13 from the $CD5^-$ mature B cells and 4 from $CD5^+$ B cells. To identify any common features between the 22 polyreactive scFvFcs in terms of sequence composition, the respective V_H and V_κ families and H-CDR3 length and positive-charge composition were compared (Supplementary Table 3); however, except the fact that the majority of the light-chain sequences corresponded to V_κ4-1, no other obvious similarities was observed.

Polyreactivity extends to HIV-1gp140 binding

Previous studies have reported that several anti-HIV-1 envelope-directed BnAbs, including 2F5 and 4E10, show promiscuous binding to multiple auto-antigens, which raises the hypothesis that their rare appearance may be because of escape from immune tolerance.[35,36] A more recent study demonstrated that 75% of all anti-HIV Abs isolated from infected individuals showed polyreactivity against single-stranded DNA, double-stranded DNA, cardiolipin, lipopolysaccharide and keyhole limpet hemocyanin using ELISA.[19] Given the auto/polyreactive binding profiles above, both serum and the hNSG-derived scFvFcs were tested for reactivity to HIV-1 envelope. As shown in Figure 7a, sera samples from hNSG mice displayed significantly higher reactivity to HIV gp140 compared with human serum controls with most of the reactivity restricted to the IgM fraction. Given that the IgG content in normal human serum is 100–200-fold higher than that in hNSG sera, the apparently higher reactivity of normal human IgG is most likely a concentration effect rather than antigen specificity.

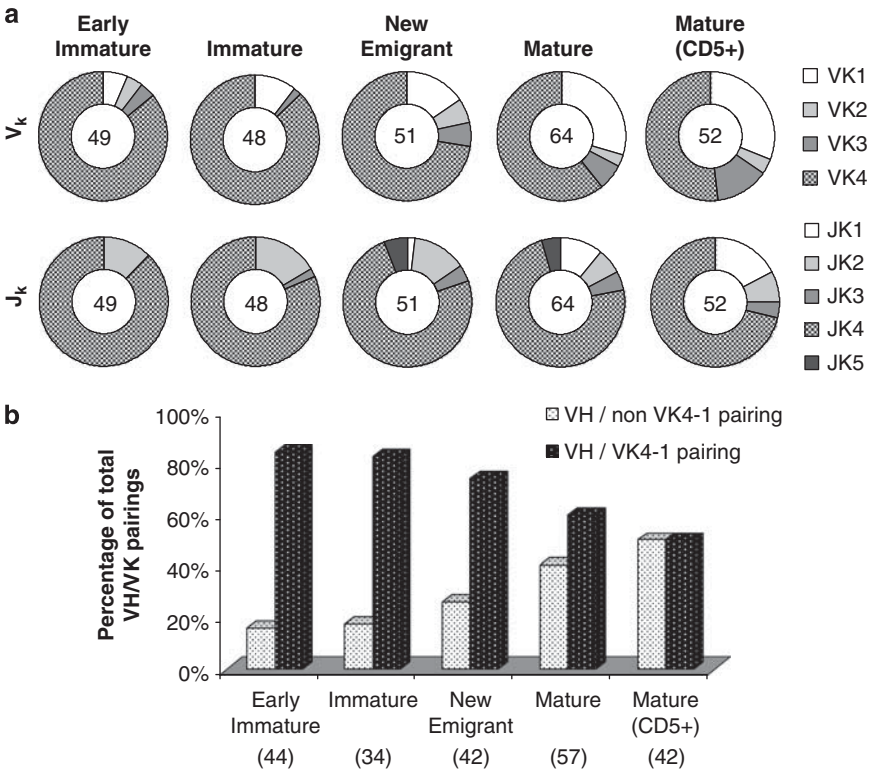


Figure 4. Analysis of the V_κ repertoire and V_H–V_κ pairing in human B cells from hNSG mice. (**a**) The frequency of utilization of the different V_κ (upper panel) and J_κ (lower) gene families across the different B-cell subpopulations, as indicated above each pie chart is shown. The number in the center of each pie chart indicates the total sample size for the particular B-cell subset. (**b**) V_H–V_k pairs were classified into two groups—V_H/non V_κ4-1 pairings and V_H/V_κ4-1 pairings, their percentages over total number of V_H–V_k pairing in a particular B-cell subset were calculated and plotted as shown. A significant decrease in the incidence of V_H/V_κ4-1 pairings (or conversely, an increase in V_H/non V_κ4-1 pairings) was observed in the mature B-cell subsets ($CD5^-$ and $CD5^+$ subsets, combined) in comparison with the bone marrow-derived B cell (early immature and immature subsets, combined) (Fisher's test, $P<0.005$).

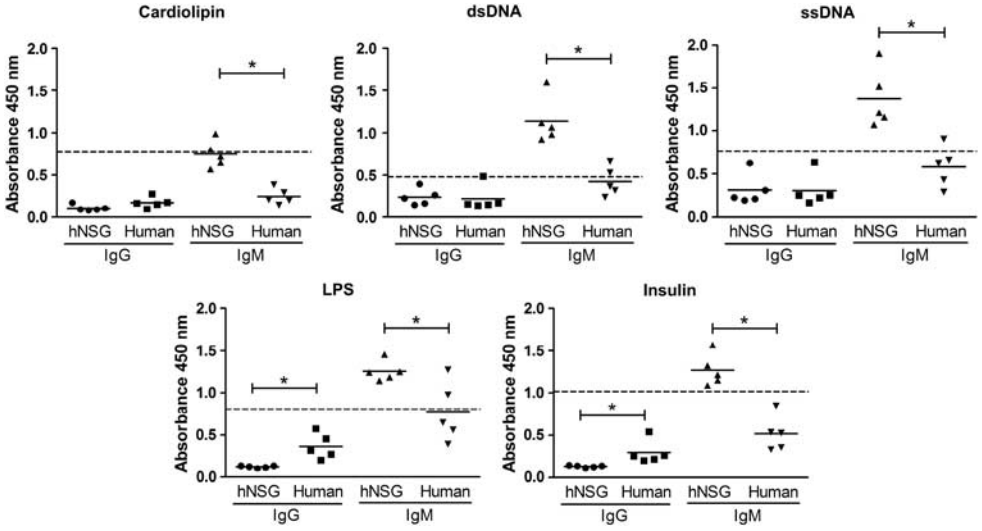


Figure 5. Polyreactivity in the sera of hNSG mice. Sera from five hNSG mice at 10 months post engraftment and five normal human subjects were tested by ELISA for binding to different antigens as listed above each graph. Antigens were coated on 96-well plates, and each serum sample was tested in two sets of duplicates, each set scored for human IgM- or IgG (X-axis)-mediated antigen binding using the appropriate detection Ab, that is, anti-human IgM and anti-human IgG, respectively, both conjugated to HRP. All serum samples were used at 1:100 dilution. Total human IgM and IgG content were 10–50- and 100–200-fold higher respectively, in the human sera compared with the hNSG sera. Sera from unengrafted hNSG mice showed no reactivity (not shown). Each dot represents an individual hNSG mouse or human subject as indicated on the X-axis, and a scatter plot along with the mean for each group is shown. The horizontal dashed line represents the level of reactivity by 4E10 IgG (5 $\mu g\,ml^{-1}$) used as a positive control. Significant difference in reactivity between data sets as indicated (*$P<0.005$) was calculated using the Mann–Whitney test.

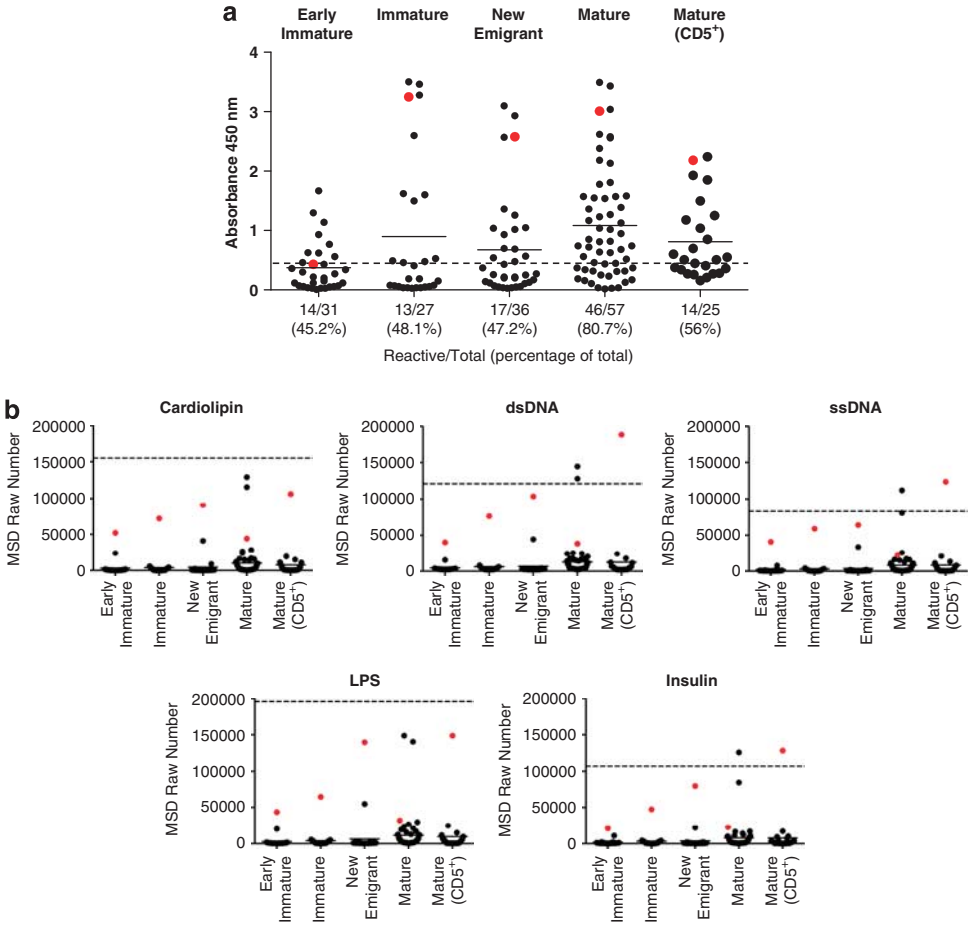


Figure 6. Auto and polyreactivity analysis of the hNSG-derived scFvFcs. (**a**) Hep-2 ELISA was performed with scFvFcs assembled and expressed from each B-cell subset and purified from 293FT cell culture supernatants by protein A chromatography. ELISA was performed using commercially available QUANTA Lite ANA ELISA plates. The assay was performed twice, once with 50 $\mu g\,ml^{-1}$ and repeated with 25 $\mu g\,ml^{-1}$ of the scFvFcs. The horizontal dashed line represents the low positive control (manufacturer provided) cutoff and the numbers below each scatter plot represent the total number and percentage of the scFvFcs those were scored as reactive in the Hep-2 assay. Each dot represents a single scFvFc and the horizontal bar represents mean reactivity for each cell subset. A single scFvFc determined to be polyreactive (see Figure 6b) in each B-cell subset has been color-coded (red dot) for correlation between autoreactivity and polyspecificity. Autoreactivity was significantly higher in the mature B cell (median reactivity in mature vs early immature/immature/new emigrant and mature $CD5^+$ vs early immature/immature/new emigrant, $P<0.005$). (**b**) Polyreactivity of the scFvFcs isolated from the different B-cell subsets (indicated on the X-axis) was tested using the MSD platform. 384-well plates were coated with the different antigens and scFvFcs were tested at a concentration of 5 $\mu g\,ml^{-1}$, with all samples tested in duplicates. The mean value in MSD raw numbers (Y-axis) as obtained in the MSD sector imager was plotted. 4E10 scFvFc used as a positive control is represented as the horizontal dashed line in each plot and readings obtained with PBS binding to antigen-coated wells was considered as background. Each dot represents the binding of individual scFvFcs and the horizontal bar represents mean reactivity for each cell subset. A single scFvFc in each B-cell subset was color-coded (red dot) for correlation of polyspecificity across different antigens. Polyspecificity was significantly higher in the mature B-cell subsets (median reactivity in mature vs early immature/immature/new emigrant and mature $CD5^+$ vs early immature/immature/new emigrant, $P<0.005$). Differences in median reactivity between the scFvFcs of each B-cell subset were calculated using the Mann–Whitney test.

Furthermore, in a similar pattern to that described earlier, the scFvFcs derived from the mature B cells displayed significantly higher binding to HIV-1 envelope protein compared with the scFvFcs derived from B cells earlier in the developmental pathway (Figure 7b).

Kinetic analysis of two random scFvFcs displaying high (N1M10-1, isolated from $CD5^+$ B-cell subset) and medium reactivity (NM10-1, isolated from $CD5^-$ mature B cells) for binding to HIV-1gp140 trimer protein was performed using Biolayer Interferometry, and k_{on}, k_{off} and K_D values of the scFvFcs (b12 and 4E10 IgG used as positive controls) were measured. The sensorgrams of each Ab (which were immobilized on protein A sensors) binding to a series of different concentrations of HIV envelope protein in solution are shown in Figure 7c. The association rates (k_{on}) of the scFvFcs were comparable to that of b12 and 4E10 ($k_{on}(Ms^{-1})$ values were 4.69×10^6 and 5.54×10^5 for N1M10-1 and NM10-1, respectively, whereas it was 2.89×10^5 and 2.55×10^5 for b12 and 4E10, respectively); however, the scFvFcs dissociated from the target protein at a faster rate (k_{off} (s^{-1}) values were 2.68×10^{-2} and 2.40×10^{-2} for N1M10-1 and NM10-1, respectively, whereas it was 4.19×10^{-5} and 1.42×10^{-4} for b12 and 4E10, respectively). Intriguingly, the equilibrium dissociation constant (K_D, in M) values of the scFvFcs obtained from naïve hNSG mice closely approximated to those of the HIV BnAbs, with K_D values of 5.72×10^{-9} for both N1M10-1 and NM10-1, and 1.45×10^{-10} and 5.56×10^{-10} for b12 and 4E10, respectively. Although these Abs displayed substantial affinity to HIV envelope protein, they did not show neutralizing activity to HIV-1 or compete with target-binding of b12 or 4E10 (data not shown), most likely because of their fast dissociation rates.

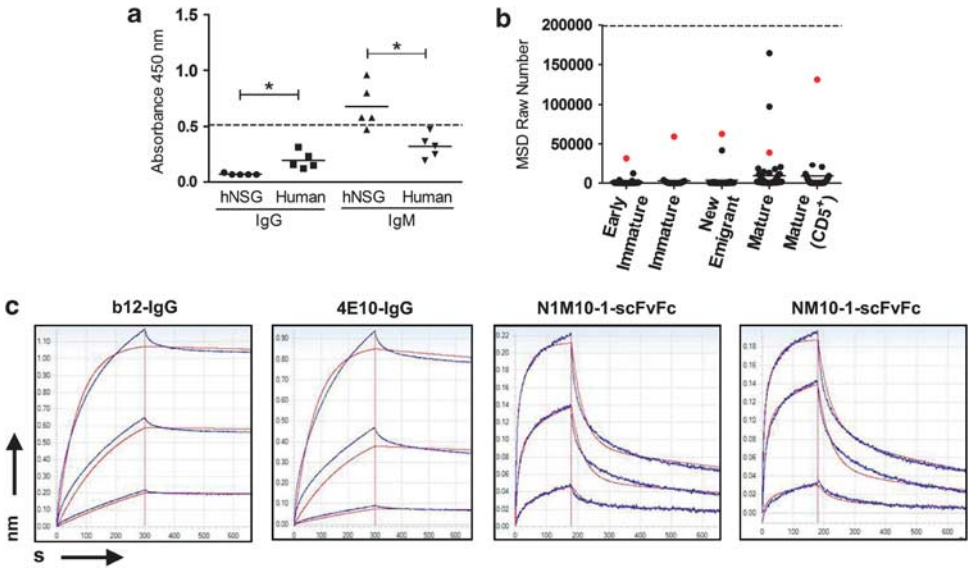


Figure 7. Binding of hNSG Abs to recombinant HIV-1gp140. (**a**) Sera from five hNSG mice and five normal human subjects were tested by ELISA for binding to recombinant HIV-1gp140 coated at a concentration of 5 μg ml^{-1} per well. The experiment was performed twice with each serum sample tested in duplicate sets as earlier, (see Figure 5). All serum samples were used at 1:100 dilution. The horizontal dashed line represent the level of 4E10 IgG (5 μg ml^{-1}) reactivity, used as a positive control. Significant differences in reactivity between data sets as indicated (*$P<0.005$) was calculated using the Mann–Whitney test. (**b**) HIV-1gp140 reactivity of the scFvFcs isolated from the different B-cell subsets (indicated on the X-axis) was tested using the MSD platform. HIV envelope protein was coated at a concentration of 5 μg ml^{-1} per well. All samples were tested in duplicates and the mean value in MSD raw numbers (Y-axis) as obtained in the MSD Sector Imager was plotted. 4E10 scFvFc was used as positive control (represented as the horizontal dashed line) and readings obtained with PBS binding to antigen-coated wells were considered as background. Each dot represents the binding of an individual scFvFc, and the horizontal bar represents mean reactivity for each cell subset. The same scFvFc in each B-cell subset has been color-coded (red dot) as in Figure 6. Reactivity to HIV-1gp140 was significantly higher in the mature B-cell subsets (median reactivity in mature vs early immature/immature/new emigrant and mature CD5$^+$ vs early immature/immature/new emigrant, $P<0.005$, Mann-Whitney unpaired *t*-test). (**c**) Affinity measurements of the hNSG-derived scFvFcs to HIV-1gp140 were obtained using Biolayer interferometry. Representative sensorgrams of two HIV BnABs (b12-IgG and 4E10 IgG) and two randomly selected polyreactive scFvFcs (NM10-1 and N1M10-1, color-coded in Figure 6) binding to different concentrations of recombinant HIV-1gp140 trimer are shown. Abs were loaded on protein A sensors, which were dipped into recombinant HIV-1gp140 in solution to determine association followed by dissociation in PBS. Traces were obtained for Ab binding to a range of HIV-1gp140 trimer concentrations. Dissociation was performed for up to 1800 s, but steady state was attained by 600 s, which is shown here. Binding responses to <10 nM HIV envelope protein were negative in case of the scFvFcs; therefore, only the binding traces obtained for 50, 16.6 and 5.6 nM of HIV-1gp140 are shown. Experiments were repeated at least three times with different lots of proteins. Affinity measurement values for each Ab are described in Results.

Auto/polyreactivity is still present in 'humanized' BLT and GTL mice

To address the question of whether T cells affect the Ab repertoire as suggested earlier,[37] similar studies were performed with humanized BLT and GTL mice. Although both the BLT[38] and GTL mice display a rapid development of human T cells as a result of co-engraftment of human thymic tissue in addition to CD34$^+$ stem cells, preimmune serum Ab levels of human IgM and IgG were not improved over that observed for the hNSG mice. Following similar methods as described for the hNSG mice, single human B-cell sorting of mature splenic B cells (CD19$^+$ IgM$^+$CD10$^-$), Ab gene recovery and scFvFc assembly were performed from a single BLT and GTL mouse 16–20 weeks post tissue transplantation. A total of 59 and 37 scFvFcs were assembled from mature human B cells isolated from the spleen, respectively. The gene usage profile for heavy- and light-chain V regions was comparable to that observed for the hNSG mice, including the lack of somatic hypermutations in the V_H sequences, sole recovery of V_κ genes and an overutilization of V_κ4-1 (data not shown). A majority of the scFvFcs, ~75% and ~78% from the BLT and GTL mice, respectively, also displayed auto- and polyreactivity as shown in Supplementary Figure 3. These results suggest that the polyreactive IG repertoire is not exclusive to the hNSG model, and it is likely that other 'humanized' mouse models may share common deficiencies in peripheral tolerance mechanisms.

DISCUSSION

In this study, we performed an in-depth analyses and characterization of the human Ab repertoire in the humanized NSG mouse model. The hNSG mice supported the development of a multilineage human hematopoietic system following transplantation with human HSCs, and at 8–10 months post transplantation, phenotypic characterization indicated a normal B-cell developmental pathway with the immature B cells residing predominantly in the bone marrow and the mature cells in the periphery. However, several deviations were noted, some of which have been reported earlier, for example, slow development of human T cells[9] and a higher than normal frequency of CD19$^+$CD5$^+$ B cells in the periphery.[38,39] Additionally, the current work provides evidence for the first time that the human Ab repertoire that develops in the hNSG and other humanized mouse models is largely auto- and polyreactive.

The human V_H diversity in the hNSG mice appears to closely approximate to that of the normal adult IG repertoire in terms of germline gene usage, but with an unexpectedly high level of positively charged, long H-CDR3s in the mature CD5$^-$ population

that displayed increased auto/polyreactivity. There was also an abnormal high contribution to the rearranged V_H gene pool from the $CD5^+$ B cells in the periphery (Figure 6a). It has been previously shown that increased numbers of $CD5^+$ B cells are found in cord blood, which display auto/polyreactivity.[40,41] It is doubtful that the source of human stem cells induces the development of $CD5^+$ B cells in the humanized mice, because $CD5^+$ B cells have also been shown to develop in the hNSG mice when engrafted with stem cells derived from the adult bone marrow or mobilized PB.[39] In early life, a restricted V_H repertoire resulting from limited IG diversity has been observed, which coincides with the observed autoreactivity in the B-cell clones isolated from fetal liver and cord blood.[42,43] In addition, in fetal liver B cells, there is skewing in the distribution of V_H genes that are chromosomally located closer to the D_H–J_H–C_H locus.[44] Recently, next-generation sequencing of expressed Ab repertoires from human cord blood cells have shown comparable frequencies of V_H germline gene usage to those present in adult IgM repertoire except for the high frequency of V_H1-2 germline gene that was preferentially expressed in the cord blood cells. As expected, lower degree of somatic mutation in the CDR and framework regions was observed in the cord blood cells.[45] Although our genetic analyses show both similarities and differences compared with the different V_H repertoires discussed above, the auto/polyreactivity of the hNSG Abs set their functional maturity at an early stage.

The observations of a large pool of auto/polyreactive Abs in the humanized mice made in this study raise two inter-related questions: (1) why are the autoreactive B-cell clones retained in the periphery and (2) are the auto/polyreactive Abs in the humanized mice akin to normal circulating 'natural Abs' found in both men and mice?[46] During normal course of B-cell development, a majority of the self-reactive B-cell clones are eliminated by checkpoint control mechanisms as characterized by Wardemann *et al.*[18] A 'central' checkpoint occurs in the BM during the development of immature B cells from the previous early immature stage and a 'peripheral' checkpoint between the new emigrant and the mature phase of B-cell development. A third checkpoint has also been predicted that eliminates residual autoAbs from the IgM^+ memory B cells.[47] These checkpoint controls result in a significant reduction in H-CDR3 length and positive-charge content from B-cell clones isolated at the immature and mature naïve stages when compared with the corresponding early immature and new emigrant B-cell stages. Our data support that the peripheral checkpoint control mechanism in the hNSG mice may be impaired, given the significant increase in the number of mature B cells carrying long and highly charged H-CDR3 regions. The factors regulating these checkpoint control steps have not been identified, and human B-cell ontogeny in a xenogeneic environment adds to the complexity. Autoreactive B cells are thought to be selected for elimination by their reactivity to self-antigens by deletion and receptor editing. In the hNSG BM, human B-cell development conceivably occurs by selection against both donor (human) and self-antigens (mouse), and additionally, the distribution of these antigens may be varied in different compartments (for example, bone marrow vs periphery). This may result in improperly 'educated' human B cells, particularly in the periphery where self-antigens predominate and the peripheral checkpoint control mechanism, designed to eliminate autoreactive clones, could be more impaired. Despite the apparent autoreactivity in the periphery, these mice did not display any apparent signs of autoimmunity, that is, symptoms of graft-versus-host disease, for example, anorexia, wasting, and so on.

We verified to a great extent, that the predominance of V_κ sequences, particularly V_κ4-1 gene segment, as the exclusive light-chain partner in all assembled Abs was not a PCR-induced bias. The fact that cloned and serum Abs alike displayed autoreactivity and polyspecific responses to multiple antigens provides a strong correlation with V_κ4-1 prevalence. These results also illustrate an apparent defect or insufficiency in one of the major mechanisms of silencing self-reactive B cells; light-chain receptor editing, which is mostly mediated by V_λ genes.[48] Polyreactive Abs have also been detected in the IgG^+ memory B-cell pools from normal human subjects and these have been described as a by-product of extensive somatic hypermutation occurring during antigen-induced B-cell differentiation.[47] However, this observation does not appear relevant to the current study, as neither a memory phenotype nor a high level of hypermutation was observed in all the B-cell clones studied.

Natural Abs secreted by B cells are predominantly of the IgM subclass, but are also represented by IgG and IgA isotypes. These Abs have been characterized extensively in mice[49] and are present during early human life.[50] A distinct B-cell subset, called B-1 cells, characterized by surface expression of the CD5 marker has been found to be the major producer of natural Abs in mice. In addition to displaying low-affinity interactions with self-antigens, these Abs have also been found to have a major role in early protective responses against pathogens.[34] Further studies remain to be performed to understand the nature of the auto/polyreactive human Abs that develop in the hNSG mice under pathogen-free conditions and whether the $CD5^+$ human B cells represent the murine equivalent.

A number of anti-HIV envelope BnAbs have shown characteristics of auto/polyreactive Abs and thus, it has been proposed that the rare emergence of such Abs in the vaccinated or infected host may be due to physiological tolerance mechanisms that are operative during B-cell development.[51] Recently, it was also shown that HIV envelope reactive Abs isolated from 'elite controllers' of infection were polyreactive and that heteroligation to gp120 and self-antigen can increase the apparent affinity of the Abs for HIV.[19] The preponderance of peripheral auto/polyreactive Abs in naïve and germ-free hNSG mice that also show binding to HIV envelope proteins may provide a model in which the generation and origins of protective Ab responses against HIV can be investigated. Indeed, because many of these Abs already display binding to the HIV envelope, their potency could be further enhanced by experimental infection or immunization. Therefore, the hNSG and other similar mouse models may provide a novel experimental system in which the breakdown in physiological tolerance mechanisms can be exploited to further investigate the role that auto/polyreactivity may have in the development and evolution of HIV BnAbs. This knowledge will be of critical importance to current and future HIV/AIDS vaccine efforts.

MATERIALS AND METHODS

Construction of humanized mice

Male NSG mice, 6 weeks of age (Jackson Laboratory, Bar Harbor, ME, USA), were housed under BSL-2 conditions at the Animal Research Facility, Dana-Farber Cancer Institute (Boston, MA, USA). Mice received autoclaved food and Baytril (fluoroquinolone)-treated water. $CD34^+$ HSC were isolated from human umbilical cord blood, obtained from Brigham and Women's Hospital (Boston, MA, USA), using immunomagnetic column purification techniques (MiniMACS) per the manufacturer's protocol (Miltentyi, Auburn, CA, USA). At $\sim$8 weeks of age, mice were sublethally irradiated with 325 cGy (Gammacell-40, Best Theratronics, Ottawa, Ontario, Canada) and then injected with 2.5×10^5 HSC resuspended in 200 μl of phosphate-buffered saline (PBS) via the tail vein. NOD/SCID-thy/liv (also called BLT) and NSG-thy/liv (GTL) mice were also constructed as previously described.[38] Briefly, 6–8-week-old NOD/SCID or NSG female mice were 'humanized' following implantation of human fetal liver and thymus tissues under the kidney capsule along with an intravenous delivery of fetal liver-derived autologous $CD34^+$ stem cells. All the fetal tissues (17–20 weeks of gestational age) were obtained from Advanced Bioscience Resources, Alameda, CA, USA. Tissues were also screened for the presence of HIV-1 and -2 and Hepatitis B virus and determined to be negative.

Mice were bled monthly via the mandibular route, and engraftment levels were determined by flow cytometry analysis of human CD45$^+$ cells in the PB. All the studies involving human tissues were approved by the Institutional Review Boards at both Dana-Farber Cancer Institute and Brigham and Women's Hospital.

Flow cytometry

Analysis of human immune reconstitution following CD34$^+$ HSC delivery was performed by flow cytometry (Supplementary Methods). Cell isolation techniques and criteria for single-cell sorting from different human B-cell subsets were adopted from a previous report[52] with modifications as described in Supplementary Methods. Single-cell sorting from NSG mice was performed on early immature and immature B-cell subsets from the bone marrow (BM) and new emigrant/transitional and mature B-cell subsets (both CD5$^-$ and CD5$^+$) from the PB.

Single-cell reverse transcriptase PCR and analysis of IG heavy- and light-chain gene segments

Reverse transcriptase PCR reactions were performed as described earlier (Supplementary Methods).[18,52]

Expression and purification of scFvFc

Cognate V_H- and light-chain gene segments amplified from single B cells were assembled following cloning into a human IgG1-Fc-expressing vector, transiently transfected in 293FT cells (Invitrogen, Grand Island, NY, USA). The expressed scFvFcs were purified as described in Supplementary Methods.

Detection of self (auto)-reactive and polyspecific Abs

QUANTA Lite ANA (anti-nuclear antibody) ELISA (INOVA Diagnostics, San Diego, CA, USA) was used to test self-reactivity of the Abs. 4E10 scFvFc and sera from patients and healthy individuals (manufacturer provided) were used as the positive or negative controls in this assay. Purified antibodies were tested at 50 $\mu g\,ml^{-1}$ and reactive samples were further confirmed at 25 $\mu g\,ml^{-1}$. The cutoff at OD450 for positive reactivity was calculated at $\geqslant$80% of the absorbance reading obtained for the manufacturer provided low positive control.

The electrochemiluminescent-based MSD platform (Gaithersburg, MD, USA) was used to evaluate the auto- and polyreactivity of hNSG-derived scFvFcs. Human single- and double-stranded DNA, recombinant insulin, cardiolipin, lipopolysaccharide (Sigma, St. Louis, MO, USA) at 10 $\mu g\,ml^{-1}$ and recombinant HIV-1 gp140-trimer from YU2 strain[38] at 5 $\mu g\,ml^{-1}$ were used as polyreactive antigens in this assay. Each well in MSD 384-well High Bind MULTI-ARRAY plates was coated with 5 μl of polyreactive antigens in PBS (cardiolipin in 20% alcohol) at 4°C overnight. The plates were blocked with 35 μl of 10% fetal bovine serum/PBS for 1 h with shaking at 800 r.p.m. on the microplate shaker. After washing with 35 μl PBST, purified scFvFcs were added at 5 $\mu g\,ml^{-1}$ and the plates were incubated at room temperature for 2 h with shaking. 4E10 scFvFc was used as a positive control. Goat anti-human IgG-Fc Ab with SULFO-Tag (15 μl at 1 $\mu g\,ml^{-1}$ in 1% fetal bovine serum/PBS) was used as the detection Ab, and the plates were read on the MSD Sector Imager 2400 according to the manufacturer's protocol. Routine ELISA (Supplementary Methods) was also performed to analyze auto/polyreactivity in hNSG and normal human serum samples. 4E10 IgG (NIH, AIDS Research and Reference Reagent Program) was used as a positive control to estimate the level of autoreactivity.

Affinity measurements by Biolayer Interferometry

Kinetic analysis of the hNSG-derived scFvFcs, b12-IgG and 4E10 IgG was performed by measuring binding affinity to recombinant HIV-1gp140 trimeric protein. Analysis was performed by Biolayer Interferometry using the Octet Red instrument (ForteBio, Menlo Park, CA, USA) (Supplementary Methods).

Statistical analysis

To analyze differences in the distribution of various gene segments across different B-cell subsets, non-parametric statistical methods were used because the data were not normally distributed. Differences were assessed for statistical significance by the Fisher–Freeman–Halton's test. Monte Carlo estimation (10 000 samples) was used to approximate the exact tests. Significant differences in the H-CDR3 length and N1 and N2 insertion across B-cell subsets were determined by fitting one-way analysis of variance using SAS PROC GLM (SAS, Cary, NC, USA). Differences in the utilization of individual gene segments were analyzed by one sample binomial test. Differences in utilization of particular gene segments between mature and immature B-cell subsets were analyzed by using a two-sample test for equality of proportions with continuity correction. Mann–Whitney *U*-tests (two-tailed) (Graph Pad, La Jolla, CA, USA) were performed to determine statistically significant differences between median values of each ELISA data set. *P*-values <0.005 were considered significant.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

ACKNOWLEDGEMENTS

We thank the staff at the Animal Research Facility at Dana-Farber Cancer Institute (DFCI) for excellent maintenance and mice handling and Michael Waring at the Flow Cytometry core facility at Ragon Institute, Massachusetts General Hospital, Charlestown, MA, USA, for assistance with single B-cell sorting. Biostatistical computations performed by Amany Awad and the DFCI Biostatistical Core service (Sandra Lee and Yang Feng), technical guidance from Islay Campbell, ForteBio and assistance from Raymond Moniz of the Marasco Laboratory with the preparation of figures are appreciated. The HIV envelope reactive monoclonal antibodies (4E10 from Hermann Katinger and b12 from Dennis Burton) were obtained through the AIDS Research and Reference Reagent Program, Division of AIDS, NIAID, NIH (Germantown, MD, USA). This work was partially supported by NIH grants R21 AI091557 and UO1 AI0070343 to WAM Financial support provided by the National Foundation for Cancer Research to the Center for Therapeutic Antibody Engineering at DFCI (NFCR-CTAE) is acknowledged. Research support by the Center For AIDS Research at Harvard University (CFAR-HU) is also acknowledged.

REFERENCES

1 Pietzsch J, Scheid JF, Mouquet H, Klein F, Seaman MS, Jankovic M *et al.* Human anti-HIV-neutralizing antibodies frequently target a conserved epitope essential for viral fitness. *J Exp Med* 2010; **207**: 1995–2002.
2 Walker LM, Huber M, Doores KJ, Falkowska E, Pejchal R, Julien JP *et al.* Broad neutralization coverage of HIV by multiple highly potent antibodies. *Nature* 2011; **477**: 466–470.
3 Wu X, Yang ZY, Li Y, Hogerkorp CM, Schief WR, Seaman MS *et al.* Rational design of envelope identifies broadly neutralizing human monoclonal antibodies to HIV-1. *Science* 2010; **329**: 856–861.
4 Wu X, Zhou T, Zhu J, Zhang B, Georgiev I, Wang C *et al.* Focused evolution of HIV-1 neutralizing antibodies revealed by structures and deep sequencing. *Science* 2011; **333**: 1593–1602.
5 Corti D, Voss J, Gamblin SJ, Codoni G, Macagno A, Jarrossay D *et al.* A neutralizing antibody selected from plasma cells that binds to group 1 and group 2 influenza A hemagglutinins. *Science* 2011; **333**: 850–856.
6 Dimitrov JD, Kazatchkine MD, Kaveri SV, Lacroix-Desmazes S. 'Rational vaccine design' for HIV should take into account the adaptive potential of polyreactive antibodies. *PLoS pathogens* 2011; **7**: e1002095.
7 Manz MG, Di Santo JP. Renaissance for mouse models of human hematopoiesis and immunobiology. *Nat Immunol* 2009; **10**: 1039–1042.
8 Shultz LD, Ishikawa F, Greiner DL. Humanized mice in translational biomedical research. *Nat Rev Immunol* 2007; **7**: 118–130.
9 Shultz LD, Lyons BL, Burzenski LM, Gott B, Chen X, Chaleff S *et al.* Human lymphoid and myeloid cell development in NOD/LtSz-scid IL2R gamma null mice engrafted with mobilized human hemopoietic stem cells. *J Immunol* 2005; **174**: 6477–6489.
10 Schmidt MR, Appel MC, Giassi LJ, Greiner DL, Shultz LD, Woodland RT. Human BLyS facilitates engraftment of human PBL derived B cells in immunodeficient mice. *PloS one* 2008; **3**: e3192.
11 Huntington ND, Alves NL, Legrand N, Lim A, Strick-Marchand H, Mention JJ *et al.* IL-15 transpresentation promotes both human T-cell reconstitution and T-cell-dependent antibody responses *in vivo*. *Proc Natl Acad Sci USA* 2011; **108**: 6217–6222.
12 Willinger T, Rongvaux A, Strowig T, Manz MG, Flavell RA. Improving human hemato-lymphoid-system mice by cytokine knock-in gene replacement. *Trends Immunol* 2011; **32**: 321–327.
13 Becker PD, Legrand N, van Geelen CM, Noerder M, Huntington ND, Lim A *et al.* Generation of human antigen-specific monoclonal IgM antibodies using vaccinated 'human immune system' mice. *PloS one* 2010; **5**: e13137.

14 Marodon G, Desjardins D, Mercey L, Baillou C, Parent P, Manuel M *et al.* High diversity of the immune repertoire in humanized NOD.SCID.gamma c−/− mice. *Eur J Immunol* 2009; **39**: 2136–2145.
15 Kolar GR, Yokota T, Rossi MI, Nath SK, Capra JD. Human fetal, cord blood, and adult lymphocyte progenitors have similar potential for generating B cells with a diverse immunoglobulin repertoire. *Blood* 2004; **104**: 2981–2987.
16 Rossi MI, Medina KL, Garrett K, Kolar G, Comp PC, Shultz LD *et al.* Relatively normal human lymphopoiesis but rapid turnover of newly formed B cells in transplanted nonobese diabetic/SCID mice. *J Immunol* 2001; **167**: 3033–3042.
17 Mouquet H, Nussenzweig MC. Polyreactive antibodies in adaptive immune responses to viruses. *Cell Mol Life Sci* 2012; **69**: 1435–1445.
18 Wardemann H, Yurasov S, Schaefer A, Young JW, Meffre E, Nussenzweig MC. Predominant autoantibody production by early human B cell precursors. *Science* 2003; **301**: 1374–1377.
19 Mouquet H, Scheid JF, Zoller MJ, Krogsgaard M, Ott RG, Shukair S *et al.* Polyreactivity increases the apparent affinity of anti-HIV antibodies by heteroligation. *Nature* 2010; **467**: 591–595.
20 Haynes BF, Moody MA, Verkoczy L, Kelsoe G, Alam SM. Antibody polyspecificity and neutralization of HIV-1: a hypothesis. *Hum Antibodies* 2005; **14**: 59–67.
21 Schettino EW, Chai SK, Kasaian MT, Schroeder Jr. HW, Casali P. VHDJH gene sequences and antigen reactivity of monoclonal antibodies produced by human B-1 cells: evidence for somatic selection. *J Immunol* 1997; **158**: 2477–2489.
22 Agematsu K, Hokibara S, Nagumo H, Komiyama A. CD27: a memory B-cell marker. *Immunol Today* 2000; **21**: 204–206.
23 Brezinschek HP, Brezinschek RI, Lipsky PE. Analysis of the heavy chain repertoire of human peripheral B cells using single-cell polymerase chain reaction *J Immunol* 1995; **155**: 190–202.
24 Brezinschek HP, Foster SJ, Brezinschek RI, Dorner T, Domiati-Saad R, Lipsky PE. Analysis of the human VH gene repertoire. Differential effects of selection and somatic hypermutation on human peripheral CD5(+)/IgM+ and CD5(−)/IgM+ B cells. *J Clin Invest* 1997; **99**: 2488–2501.
25 Schroeder Jr HW. Similarity and divergence in the development and expression of the mouse and human antibody repertoires. *Dev Comp Immunol* 2006; **30**: 119–135.
26 Zemlin M, Klinger M, Link J, Zemlin C, Bauer K, Engler JA *et al.* Expressed murine and human CDR-H3 intervals of equal length exhibit distinct repertoires that differ in their amino acid composition and predicted range of structures. *J Mol Biol* 2003; **334**: 733–749.
27 Chen C, Nagy Z, Prak EL, Weigert M. Immunoglobulin heavy chain gene replacement: a mechanism of receptor editing. *Immunity* 1995; **3**: 747–755.
28 Longo NS, Grundy GJ, Lee J, Gellert M, Lipsky PE. An activation-induced cytidine deaminase-independent mechanism of secondary VH gene rearrangement in preimmune human B cells. *J Immunol* 2008; **181**: 7825–7834.
29 Zhang Z. VH replacement in mice and humans. *Trends Immunol* 2007; **28**: 132–137.
30 Zhang Z, Zemlin M, Wang YH, Munfus D, Huye LE, Findley HW *et al.* Contribution of Vh gene replacement to the primary B cell repertoire. *Immunity* 2003; **19**: 21–31.
31 Meffre E, Schaefer A, Wardemann H, Wilson P, Davis E, Nussenzweig MC. Surrogate light chain expressing human peripheral B cells produce self-reactive antibodies. *J Exp Med* 2004; **199**: 145–150.
32 Suzuki I, Milner EC, Glas AM, Hufnagle WO, Rao SP, Pfister L *et al.* Immunoglobulin heavy chain variable region gene usage in bone marrow transplant recipients: lack of somatic mutation indicates a maturational arrest. *Blood* 1996; **87**: 1873–1880.
33 Manz RA, Hauser AE, Hiepe F, Radbruch A. Maintenance of serum antibody levels. *Annu Rev Immunol* 2005; **23**: 367–386.
34 Ochsenbein AF, Fehr T, Lutz C, Suter M, Brombacher F, Hengartner H *et al.* Control of early viral and bacterial distribution and disease by natural antibodies. *Science* 1999; **286**: 2156–2159.
35 Alam SM, McAdams M, Boren D, Rak M, Scearce RM, Gao F *et al.* The role of antibody polyspecificity and lipid reactivity in binding of broadly neutralizing anti-HIV-1 envelope human monoclonal antibodies 2F5 and 4E10 to glycoprotein 41 membrane proximal envelope epitopes. *J Immunol* 2007; **178**: 4424–4435.
36 Haynes BF, Fleming St J, Clair EW, Katinger H, Stiegler G, Kunert R *et al.* Cardiolipin polyspecific autoreactivity in two broadly neutralizing HIV-1 antibodies. *Science* 2005; **308**: 1906–1908.
37 Meffre E, Salmon JE. Autoantibody selection and production in early human life. *J Clin Invest* 2007; **117**: 598–601.
38 Biswas S, Chang H, Sarkis PT, Fikrig E, Zhu Q, Marasco WA. Humoral immune responses in humanized BLT mice immunized with West Nile virus and HIV-1 envelope proteins are largely mediated via human CD5(+) B cells. *Immunology* 2011; **134**: 419–433.
39 Matsumura T, Kametani Y, Ando K, Hirano Y, Katano I, Ito R *et al.* Functional CD5+ B cells develop predominantly in the spleen of NOD/SCID/gammac(null) (NOG) mice transplanted either with human umbilical cord blood, bone marrow, or mobilized peripheral blood CD34+ cells. *Exp Hematol* 2003; **31**: 789–797.
40 Mackenzie LE, Mageed RA, Youinou PY, Yuksel B, Jefferis R, Lydyard PM. Repertoire of CD5+ and CD5- cord blood B cells: specificity and expression of VH I and VH III associated idiotopes. *Clin Exp Immunol* 1992; **88**: 107–111.
41 Paavonen T, Quartey-Papafio R, Delves PJ, Mackenzie L, Lund T, Youinou P *et al.* CD5 mRNA expression and auto-antibody production in early human B cells immortalized by EBV. *Scand J Immunol* 1990; **31**: 269–274.
42 Lydyard PM, Quartey-Papafio R, Broker B, Mackenzie L, Jouquan J, Blaschek MA *et al.* The antibody repertoire of early human B cells. I. High frequency of autoreactivity and polyreactivity. *Scand J Immunol* 1990; **31**: 33–43.
43 Lydyard PM, Quartey-Papafio RP, Broker BM, MacKenzie L, Hay FC, Youinou PY *et al.* The antibody repertoire of early human B cells. III. Expression of cross-reactive idiotopes characteristic of certain rheumatoid factors and identifying V kappa III, VHI, and VHIII gene family products. *Scand J Immunol* 1990; **32**: 709–716.
44 Schroeder Jr HW, Mortari F, Shiokawa S, Kirkham PM, Elgavish RA, Bertrand 3rd FE. Developmental regulation of the human antibody repertoire. *Ann NY Acad Sci* 1995; **764**: 242–260.
45 Prabakaran P, Chen W, Singarayan MG, Stewart CC, Streaker E, Feng Y *et al.* Expressed antibody repertoires in human cord blood cells: 454 sequencing and IMGT/HighV-QUEST analysis of germline gene usage, junctional diversity, and somatic mutations. *Immunogenetics* 2012; **64**: 337–350.
46 Coutinho A, Kazatchkine MD, Avrameas S. Natural autoantibodies. *Curr Opin Immunol* 1995; **7**: 812–818.
47 Tiller T, Tsuiji M, Yurasov S, Velinzon K, Nussenzweig MC, Wardemann H. Autoreactivity in human IgG+ memory B cells. *Immunity* 2007; **26**: 205–213.
48 Wardemann H, Hammersen J, Nussenzweig MC. Human autoantibody silencing by immunoglobulin light chains. *J Exp Med* 2004; **200**: 191–199.
49 Baumgarth N. The double life of a B-1 cell: self-reactivity selects for protective effector functions. *Nat Rev Immunol* 2011; **11**: 34–46.
50 Merbl Y, Zucker-Toledano M, Quintana FJ, Cohen IR. Newborn humans manifest autoantibodies to defined self molecules detected by antigen microarray informatics. *J Clin Invest* 2007; **117**: 712–718.
51 Haynes BF, Nicely NI, Alam SM. HIV-1 autoreactive antibodies: are they good or bad for HIV-1 prevention? *Nat Struct Mol Biol* 2010; **17**: 543–545.
52 Tiller T, Meffre E, Yurasov S, Tsuiji M, Nussenzweig MC, Wardemann H. Efficient generation of monoclonal antibodies from single human B cells by single cell RT-PCR and expression vector cloning. *J Immunol Methods* 2008; **329**: 112–124.

Supplementary Information accompanies the paper on Genes and Immunity website (http://www.nature.com/gene)