



## Trials and Tribulations with VH Replacement

Wenzhao Meng, Sahana Jayaraman, Zhang Bochao, Gregory W. Schwartz, Robert Daber, Uri Hershberg, Alfred Garfall, Christopher S Carlson and Eline T. Luning\_Prak

Journal Name:	Frontiers in Immunology
ISSN:	1664-3224
Article type:	Hypothesis & Theory Article
Received on:	29 Oct 2013
Accepted on:	07 Jan 2014
Provisional PDF published on:	07 Jan 2014
Frontiers website link:	<a href="http://www.frontiersin.org">www.frontiersin.org</a>
Citation:	Meng W, Jayaraman S, Bochao Z, Schwartz GW, Daber R, Hershberg U, Garfall A, Carlson CS and Luning_prak ET(2014) Trials and Tribulations with VH Replacement. <i>N/A</i> 5:10. doi:10.3389/fimmu.2014.00010
Article URL:	<a href="http://www.frontiersin.org/Journal/Abstract.aspx?s=1176&amp;name=b%20cell%20biology&amp;ART_DOI=10.3389/fimmu.2014.00010">http://www.frontiersin.org/Journal/Abstract.aspx?s=1176&amp;name=b%20cell%20biology&amp;ART_DOI=10.3389/fimmu.2014.00010</a> (If clicking on the link doesn't work, try copying and pasting it into your browser.)
Copyright statement:	© 2014 Meng, Jayaraman, Bochao, Schwartz, Daber, Hershberg, Garfall, Carlson and Luning_prak. This is an open-access article distributed under the terms of the <a href="http://creativecommons.org/licenses/by/4.0/">Creative Commons Attribution License (CC BY)</a> . The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

This Provisional PDF corresponds to the article as it appeared upon acceptance, after rigorous peer-review. Fully formatted PDF and full text (HTML) versions will be made available soon.

## **Trials and Tribulations with VH Replacement**

W. Meng\*, S. Jayaraman\*, B. Zhang<sup>†</sup>, G.W. Schwartz<sup>†</sup>, R. Daber\*, U. Hershberg<sup>†></sup>, A. Garfall<sup>&</sup>, Chris Carlson<sup>#</sup> and E. T. Luning Prak<sup>\*^</sup>

\*Department of Pathology and Laboratory Medicine, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA 19104

&Division of Hematology-Oncology, Department of Medicine, Perelman School of Medicine, University of Pennsylvania, PA 19104

<sup>†</sup>Drexel University, School of Biomedical Engineering, Science and Health Systems, Philadelphia, PA 19104

<sup>></sup>Department of Microbiology and Immunology, College of Medicine, Drexel University, Philadelphia, PA 19104

<sup>#</sup>Public Health Sciences Division, Fred Hutchinson Cancer Research Center, Seattle, WA 98109-1024

*<sup>^</sup>corresponding author*

tel. (215) 746-5768

fax. (215) 573-6317

e-mail: [luning@mail.med.upenn.edu](mailto:luning@mail.med.upenn.edu)

### Summary

VH replacement is a type of antibody gene rearrangement in which an upstream heavy chain variable gene segment (VH) invades a pre-existing rearrangement (VDJ). In this Hypothesis and Theory article, we begin by reviewing the mechanism of VH replacement, its developmental timing and its potential biological consequences. Then we explore the hypothesis that specific sequence motifs called footprints reflect VH replacement vs. other processes. We provide a compilation of footprint sequences from different regions of the antibody heavy chain, include data from the literature and from a high throughput sequencing experiment to evaluate the significance of footprint sequences. We conclude by discussing the difficulties of attributing footprints to VH replacement.

### *Keywords*

V(D)J recombination

VH replacement

VH, DH and JH gene segments

Receptor editing

*Abbreviations*

IgH: antibody heavy chain

VH: antibody heavy chain variable gene segment

DH: antibody heavy chain diversity gene segment

JH: antibody heavy chain joining gene segment

VHR: VH replacement

RSS: recombination signal sequence

cRSS: cryptic recombination signal sequence

GVHD: graft versus host disease

SLE: systemic lupus erythematosus

CDR3: third complementarity determining region

footprint 5-mer: pentameric nucleotide sequence attributed to VHR

Article type: Hypothesis and Theory

## Context, definition and potential mechanisms of VH replacement

Antibodies are heterotetrameric proteins comprised of two heavy chains and two light chains that are formed through V(D)J recombination to generate a highly diverse repertoire of antigen binding receptors expressed by B cells. The *recombinase activating gene* encoded proteins, RAG1 and RAG2, target conserved heptamer and nonamers within recombination signal sequences (RSS) to cleave the DNA that flanks recombining gene segments that join together to form the variable regions of antibody heavy and light chains (reviewed in <sup>1</sup>). Typical V(D)J recombination generates a signal joint and a coding joint, and the latter is further diversified at the junction between the recombining gene segments by mechanisms including P-addition, N-addition and exonucleolytic nibbling (reviewed in <sup>2</sup>). Occasionally atypical rearrangements occur, generating hybrid joints, open-and-shut joints or joints between RSSs that ordinarily do not recombine.<sup>2-5</sup>

Antibodies can be further revised and diversified through receptor editing of the light chain, somatic hypermutation, gene conversion and VH replacement. Receptor editing typically involves RAG-dependent leapfrogging rearrangements on the same allele as the defective or autoreactive light chain, rearrangement on other alleles (kappa or lambda) and/or RS deletion (which renders preceding kappa rearrangement non-functional, reviewed in <sup>6</sup>). Somatic hypermutation is DNA point hypermutation carried out by activation induced cytidine deaminase (AID)<sup>7</sup>, and typically signifies a T-cell dependent antibody response. Gene conversion, in which homologous sequences from other V genes are grafted into the functional V gene, is a common method of gene diversification in chickens<sup>8</sup>, rabbits and more recent examples have been described in horses and humans<sup>9</sup>, and appear to be AID-dependent<sup>10</sup>. The final category of antibody gene diversification is VH replacement, which is the focus of this article. Replacement involves the transfer (or invasion) of some or most of another V gene into an existing gene rearrangement.

Darlow and Stott have reviewed the literature on VH replacement and envision two broad mechanistic classes of V replacement.<sup>11</sup> The first, also termed “classical” VH replacement, consists of invasion of an existing VDJ rearrangement by an upstream VH. In classical VH replacement there is RAG-mediated cleavage at a cryptic RSS (cRSS) located in the 3’ end of the previously rearranged VH gene. The cRSS has a DNA sequence that differs from the conventional heptamer that flanks the DH gene segment by one nucleotide, bolded in the sequence that follows: 5’-TACTGTG-3’<sup>12</sup> and is found in

approximately 70% of murine VHs and over 90% of human VHs.<sup>13</sup> Occasionally other heptamers containing the 3' GTG nucleotides can be used, suggesting that the last three nucleotides of the cRSS motif are critical<sup>14, 15</sup>. The TGT within the cRSS is the codon encoding the conserved cysteine at the junction between FR3 and CDR3. The second class of replacement, according to Darlow and Stott, involves the transfer of other sequences of homology between different V genes at different sites, many of which appear to also resemble cRSSs. Examples of this second category of VH replacement have been described in antibodies cloned from single B cells in human tonsils<sup>16</sup>, in antibodies cloned from synovial tissue of patients with rheumatoid arthritis<sup>17</sup> and in antibodies cloned from human mucosa associated lymphoid tissue lymphomas.<sup>18</sup> Alternatively or in addition to RAG mediated rearrangement, replacements in this second category may arise due to AID-mediated homologous recombination events that are unrelated to the putative cRSSs.<sup>11</sup> However, the mechanism of type 2 replacement is far from resolved as recently a non-AID-dependent form of replacement has been described at the kappa locus using human pre-B cell lines.<sup>19</sup> As the molecular mechanism of type 2 replacement remains to be fully elucidated, we will focus the remainder of our analysis in this manuscript on classical VH replacement (which we refer to hereafter as "VH replacement").

During VH replacement, an upstream VH gene invades into the cRSS, replacing all but the last few nucleotides of the previously rearranged VH gene (**Fig. 1a**). The remaining 3' nucleotides of the VH, DH and JH gene segments are retained in the new rearrangement. The extra nucleotides from the 3' end of the previous VH gene are sometimes referred to as a "footprint". Nearly all human VH genes have between 5 and 9 nucleotides in the potential footprint, between the cRSS and the RSS. Most primary RSS rearrangements delete several of these nucleotides from the 3' end, so the potential footprint may not be easily recognizable. Moreover, during VH replacement, additional nucleotides can be deleted, so the footprint from the primary VH can be entirely lost during VH replacement. It is also possible for more than one replacement rearrangement to occur on the same heavy chain allele, a process referred to as "serial" or "successive" VH replacement (**Fig. 1b**).<sup>20</sup> An example of a heavy chain rearrangement with more than one footprint sequence is given in **Fig. 1c**.

#### Demonstration of VH replacement in cell lines and mouse models

VH replacement was initially discovered in two different transformed B cell lines.<sup>21, 12</sup> In both of these early studies, B cells with non-functional heavy chain gene rearrangements (VDJ-) were able to generate functional heavy chains (VDJ+) by undergoing further heavy chain rearrangement into the cRSS.

Continued VH replacement could also convert a functional VDJ+ rearrangement into a non-functional one through the incorporation of an upstream pseudo-VH gene.<sup>12</sup>

The development of antibody heavy chain knock-in mice provided a formal demonstration of VH replacement in B cells *in vivo*. VH replacement was documented in hybridomas derived from the 3H9 heavy chain knock-in mouse.<sup>13</sup> VH replacement and invasion of upstream DH gene also occurred in a knock-in for the T15 heavy chain.<sup>15</sup> Furthermore, B cells from quasi-monoclonal mice (which have an anti-(4-hydroxy-3-nitrophenyl) acetyl (NP) heavy chain knock in and can only produce  $\lambda$  light chains, due to homozygous engineered  $\kappa$  deficiency) can lose reactivity to NP by VH replacement. Strikingly, most secreted antibodies in the quasi-monoclonal mouse appear to arise through VH replacement.<sup>22</sup> VH replacement was also observed in mice that were genetically engineered to contain two non-productively rearranged heavy chain alleles. In these VDJ-/VDJ- mice, antibody heavy chains were generated via VH replacement in a RAG-dependent manner (crossing the VDJ-/VDJ- mice onto a RAG2 deficient background resulted in a failure to generate IgM+ B cells).<sup>23</sup> The ability of RAG1 and RAG2 to bind to the cRSS was also demonstrated by electrophoretic mobility shift assays using VH4-34 cRSS vs. consensus 12-RSS sequences.<sup>24</sup>

In all of the preceding mouse models, VH replacement conferred greater diversity or functionality upon the B cell repertoire (i.e. there was selective pressure that favored VH replacement). In contrast, when VH replacement was compared with conventional rearrangement, using a mouse model with an out of frame VDJ rearrangement (VDJ-) that was knocked into the heavy chain locus, conventional rearrangement on the other heavy chain allele occurred far more frequently.<sup>25</sup> Similarly, in the 56R anti-dsDNA heavy chain knock-in mouse, receptor editing was far more efficient in B cells that were heterozygous rather than homozygous for 56R.<sup>26</sup> One caveat to the 56R study was that cells that had undergone VH replacement on one allele but were still left with a functional copy of the DNA-reactive 56R heavy chain on the other allele could be counter-selected.

#### VH replacement in bone marrow B cells

To gain further insight into the mechanism of VH replacement, studies were performed in mice to determine its developmental timing. Several studies suggest that V<sub>H</sub> replacement occurs at or near the time of conventional antibody heavy chain gene rearrangement. The junctions of IgH sequences with evidence

of V<sub>H</sub> replacement in IgH knock-in mice usually contain N-additions.<sup>13</sup> Terminal deoxynucleotidyl transferase (TdT), the enzyme that carries out N-addition, is typically expressed at highest levels during H chain rearrangement in pro-B and large cycling pre-B cells.<sup>27</sup> Therefore, the presence of N-additions provides indirect evidence that V<sub>H</sub> replacement occurred at the time when TdT was active and therefore probably took place in pro-B or early pre-B cells. Further evidence in support of V<sub>H</sub> replacement in early stage B cells includes ligation-mediated PCR to measure DNA breaks at the heavy chain locus, which occurred at the highest levels in pro-B cells.<sup>28</sup> These studies suggest that V<sub>H</sub> replacement is either occurring in cells where IgH rearrangement has not yet shut down (failed allelic exclusion) or is driven by pre-BCR rather than BCR signaling, since only the former receptor is expressed at the pre-B cell stage of development.

With respect to pre-BCR signaling (reviewed in <sup>29</sup>), it is noteworthy that surrogate light chain knock-out mice have autoreactive antibodies with long CDR3 sequences.<sup>30</sup> One potential explanation for this result is that, in the absence of surrogate light chain, the pre-BCR does not assemble and turn off heavy chain rearrangement. Without a heavy chain rearrangement stop signal, there may be higher frequencies of V<sub>H</sub> replacement, leading to CDR3 elongation. However, an alternative possibility is that peripheral selection of B cells with long CDR3 sequences is relaxed in the lymphopenic setting that arises due to inefficient primary B cell production in surrogate light chain knock out mice. It is known that in the absence of normal numbers of peripheral B cells, the level of the B cell survival factor B<sub>Ly</sub>S (also known as BAFF) increases, since B cells are the primary consumers of B<sub>Ly</sub>S. It is also known that the stringency of B cell selection can be reduced when B<sub>Ly</sub>S levels are increased.<sup>31, 32</sup>

#### V<sub>H</sub> replacement in peripheral B cells

Some studies suggest that V<sub>H</sub> replacement could occur in more mature B cell subsets. For example, there are data implicating BCR signaling in V<sub>H</sub> replacement in the EU12 human B cell line, which phenotypically resembles IgM<sup>+</sup>, CD10<sup>+</sup>, CD24<sup>high</sup> cells. In these cells, BCR crosslinking promotes V<sub>H</sub> replacement and, conversely, Syk and Src kinase inhibitors inhibit V<sub>H</sub> replacement.<sup>33</sup> While some of the kinase inhibition experiments could also be influencing mechanisms that operate at earlier stages of B cell development, the BCR crosslinking experiment suggests that BCR signaling could promote V<sub>H</sub> replacement in more mature B cells. Furthermore, ligation-mediated PCR experiments documented double-stranded DNA breaks at V<sub>H</sub>3 cRSS sites in human immature (IgM<sup>+</sup>, CD27<sup>-</sup>, CD10<sup>+</sup>) and mature naïve (IgM<sup>+</sup>, CD27<sup>-</sup>, CD10<sup>-</sup>)

circulating B cells, also suggesting that VH replacement may not be limited to immature B cells.<sup>33</sup>

Chronic graft versus host disease (GVHD) is one of the most intriguing examples in which VH replacement could be occurring in more mature B cells.<sup>34</sup> B6 mice injected with I-A incompatible T cells from bm12 mice develop chronic GVHD and produce a spectrum of autoantibodies that resembles those found in systemic autoimmune conditions such as SLE.<sup>35</sup> When anti-dsDNA heavy chain knock-in mice such as 3H9 and 56R are used, GVHD occurs and the production of anti-nuclear antibodies is enhanced.<sup>34</sup> But the remarkable finding is that among IgG antibodies, a large fraction does not use the knocked in heavy chain.<sup>34</sup> Although this unexpected skewing away from the 56R H chain could be the result of selective pressures on the minority population of H chain edited B cells that emerge from the bone marrow, it is not at all obvious how this selection could operate to disfavor the transgene, and why its effects would be largely confined to IgG and not IgM. It is possible that the transgene was revised (by further gene rearrangement) in the periphery, either because it was inactivated by somatic mutation<sup>36</sup>, or because the stimulus afforded by cGVH re-induced the rearrangement machinery. An alternative explanation is that the 56R transgene bearing cells are disfavored during primary B cell maturation because they recognize DNA and this self-reactivity causes them to be anergized (this would predict that 56R+ cells would be over-represented amongst IgM rather than IgG B cells). Consistent with the possibility of anergy, most B cells expressing the IgM allotype of the 56R transgene have low levels of IgM.<sup>37-39</sup>

#### What are the consequences and potential functions of VH replacement?

VH replacement allows a B cell with an inadequate pre-BCR or an autoreactive BCR to swap out the existing heavy chain and replace it with a different heavy chain. But why would this be useful? One possibility is that VH replacement increases the odds of generating a functional antibody. Producing a functional antibody is rather difficult<sup>40</sup>: many rearrangements are out of frame, VH pseudogenes outnumber functional VH genes, many newly generated antibodies are autoreactive<sup>41</sup>, some combinations may be sequestered inside the cells<sup>37</sup> and some H and L chain combinations may not pair well with each other. VH replacement may also facilitate the use of a wider array of upstream VH genes. By giving cells with defective antibody rearrangements a chance at revising those antibodies, perhaps the efficiency of primary B cell generation is greatly improved.



On the other hand, a seemingly diametrically opposed consequence of VH replacement is the potential generation of multireactive antibodies. VH replacement can sometimes result in the retention of a “footprint” that is comprised of DNA sequences downstream of the cryptic heptamer of the invaded VH gene (**Fig. 1**). Because the cRSS is typically positioned further from the WGXXG motif in the JH segment than the 5’RSS of a DH segment, VH replacement is likely to produce longer CDR3 segments than primary rearrangements. Not surprisingly, longer CDR3 sequences have a higher proportion of footprints (**Fig. 2**), but this does not guarantee that all long CDR3 are the product of VH replacement. 78% of the potential footprint regions in functional human VH genes contain an arginine codon, so footprint-containing sequences often also harbor a larger number of charged residues. Longer CDR3s have been associated with greater multireactivity, and such multireactive B cells are normally counter-selected as B cells mature during normal B cell development<sup>41</sup>. RA patients have antibodies with unusual CDR3 sequences in their synovium<sup>17</sup> and we have seen CDR3 sequences in patients with SLE that have regions of sequence homology that could arise due to VH replacement. For example, **Fig. 1c** shows a rearrangement from an expanded B cell clone in a patient with SLE that appears to contain two footprint sequences (highlighted in yellow). Autoimmune-prone strains of mice have elongated CDR3s, although many of these may arise through mechanisms other than VH replacement, such as D-D fusion<sup>42, 43</sup>. All of these findings beg the question of whether such “multireactivity” serves a useful function. Is multireactivity protective, particularly in the context of an innate immune response? Or could multireactive antibodies be useful in clearing debris that might be inflammatory if left to accumulate? It is intriguing in this regard that some multireactive IgM antibodies such as the famous T15 idiotype, which binds phosphorylcholine<sup>44</sup>, also have anti-inflammatory properties<sup>45</sup>.

It is possible that there is no simple single answer to the function of VH replacement, if it has one at all. It would certainly seem that the biological consequences of VH replacement depend upon the developmental context in which the rearrangement occurs. If replacement occurs centrally, as is likely to occur in wild type strains of mice such as B6<sup>39, 46</sup>, it could serve as a tolerance mechanism (receptor editing) or as means of increasing the efficiency of primary B cell generation. It might also generate a portion of the primary antibody repertoire that has special functional properties such as multireactivity. Conversely, if it occurs peripherally, as might arise in dysregulated states of immune activation such as GVHD<sup>47</sup>, perhaps autoimmunity results.

### VH replacement in pre-B cell ALL

Given the abundance of findings linking VH replacement to pro- or pre-B cell development discussed above, it is not surprising that the initial demonstrations of VH replacement occurred in transformed pre-B cell lines. More recently, VH replacement has been demonstrated to be a major contributor to clonal evolution in precursor B cell acute lymphoblastic leukemia (B-ALL)<sup>48 49</sup>. In B-ALL, there is presumably a large clone of cells "frozen" in the pre-B cell stage. The recombinase machinery remains active in at least some of these cells and can drive VH replacement. It is instructive to review the early work in the murine pre-B cell line NFS5, in which VH replacement was found to alter not only the productive but also the non-productively rearranged allele.<sup>12</sup> Thus assays where one attempts to define a clone based upon its predicted "conservation" of other immunoglobulin gene rearrangements (such as the other H chain allele) within the same cell are not necessarily reliable or easy to interpret. The potential for VH replacement to contribute to intraclonal diversification is highly relevant to the design and interpretation of assays for minimal residual disease monitoring that employ quantitative PCR with probes or primers for clone-specific junctional sequences<sup>50</sup> or, more recently, high-throughput sequencing of heavy chain CDR3.<sup>51</sup> Such studies must take VH replacement and other forms of intraclonal diversification into account.

### Analysis of VH replacement footprints

The most convincing demonstrations of VH replacement are those in which a precursor-product relationship can be documented. For example, if the precursor VH gene is known and then additional B cells can be found to share most of the 3' side of the CDR3 (the same DH-JH junction), but have a different VH gene, this can be compelling, as in B-ALL or in mouse models with heavy chain knock-ins. In contrast, the analysis of VH replacement in a physiologic and fully diversified immune repertoire has by necessity focused on indirect evidence, namely the enumeration of footprints, which are potential traces of previous VDJ rearrangements in IgH sequence data. In mice, footprints are readily observed in constrained immune repertoires (for example,<sup>13, 15, 22</sup>). Footprints are also observed in humans.<sup>52</sup> However, a fundamental issue with footprint analysis in humans is one of specificity of attribution: does the footprint arise due to VH replacement or is it due to some other form of junctional diversification or skewing in the rearrangement process? Or does it occur by chance?

To investigate the hypothesis that footprint sequences are due to the process of VH replacement, we sequenced antibody heavy chain

rearrangements from peripheral blood B cells of a healthy human adult subject, following an IRB-approved protocol. We identified 42,221 unique sequences from this sample, which we analyzed for VH footprints using a sliding window method (see **electronic supplement** for further details). All of the potential footprints arising from sequences at the 3' ends of the germline VH genes are listed in **Table 1**. In accordance with their conventional description in the literature<sup>40</sup>, we required the footprint to be least 5 nucleotides long (we hereafter refer to these sequences as footprint 5-mers). If footprint 5-mers are due to VH replacement, they will have specific characteristics in antibody repertoire data, indicated by the tests described below.

Test 1: VH replacement footprints should be located in the 5' end of the CDR3 sequence. One way to distinguish bona fide footprints from other sources of sequence variation is to compare the number of footprints in the junction between VH and DH (referred to as N1) to the number in the junction between DH and JH (referred to as N2). Footprints arising via VH replacement should occur in N1 rather than N2 because the cryptic heptamer is located in N1. However, as shown in **Fig. 3a**, there is a roughly bimodal distribution of footprint 5-mers. Even though we excluded the most common footprints that were found in the germline DH gene segments (**Table 2**), there were still plenty of footprint sequences in N1, DH and N2. In **Fig. 3b** we took the analysis one step further and removed some more of the common footprint sequences that are found not only in the germline DH gene segments but also in JH6. This resulted in more skewing towards N1, but a large proportion of the footprints were still outside of N1. In fact, not only were footprint 5-mers found in DH and JH, but they were also found in other parts of the VH gene. **Table 3** lists the positions of all of the footprint 5-mers found amongst the germline VH alleles listed in the IMGT database.

Test 2: VH replacement footprints should be more frequent in upstream VH genes than downstream VH genes and absent from VH6-1. Another requirement for a footprint to be consistent with VH replacement, is that the invading VH must be upstream of the VH that donated the footprint. Unfortunately, the recipient VH is often difficult to define because many VHs have the same or very similar footprints (see **Table 1**). However a more straightforward test of whether a footprint 5-mer represents the product of VH replacement is to evaluate the frequency of footprints in different VH rearrangements. In particular, the 3' most VH gene (VH6-1, in humans), when rearranged, should not exhibit VH replacement footprints as there is no downstream VH that it can invade. Conversely, VH genes that are situated in the 5' end of the locus should have

higher frequencies of footprints than 3' VH genes, if VH replacement is frequent. Yet the overall frequency of footprint 5-mers was similar amongst unique sequences in all of the most commonly used VHs including VH6-1 (**Fig. 4**). The frequency of footprints was also not significantly higher in out of frame (unselected) vs. in-frame rearrangements (**Fig. 4a**).

We also performed this analysis using immunoglobulin analysis tool (IgAT) software<sup>41</sup> and observed that the frequency of footprints was not reduced in VH6-1 when compared to other VHs (**Fig. 4b**). Lower VH footprint frequencies were observed overall because footprints in the 3' end of the CDR3 are excluded by the IgAT program.<sup>41</sup> One intriguing feature of the IgAT data was that, unlike our footprint analysis that captured 5-mers at both N1 and N2, when only N1 was analyzed, some VHs, including VH6-1, had higher footprint frequencies than others. Since VH6-1 cannot have any footprints due to VH replacement, we conclude that many footprint 5-mers that are found in the CDR3 do not arise by VH replacement.

The simplest explanation is that the great majority of 5-mer sequences found throughout the CDR3 resemble footprints by chance. The frequency of footprint 5-mers in the entire CDR3 was highly correlated with the length of the CDR3 (**Fig. 2**). The ability to generate a replacement footprint by chance may be underappreciated. In a completely random DNA sequence with equal proportions of A, T, G and C bases, the chance of finding a specific 5-mer sequence is 1/1024 (or ~ 0.001). However, there are at least 50 different footprint-derived 5-mer sequences amongst human VH genes (**Table 3**), increasing the odds to 50/1024 (~5%). But this calculation ignores the number of different positions along the VDJ rearrangement where the footprint might be detected and on how many variants of the footprint are permitted. If a CDR3 sequence is 60 nucleotides long, that means that there are 12 completely non-overlapping sequences that have a length of 5 nucleotides, bringing the minimum likelihood of detection of at least a single footprint in that sequence up to 46% ( $1-(1-0.05)^{12}$ ) or 1-Pr(not getting any 5-mers in the 60 bp CDR3 sequence). If the base composition of the DNA is non-uniform or the entire CDR3 sequence is surveyed or if sequences with mutations are permitted (for example those matching in 4 out of 5 bp), the chances of detecting a footprint increase even further.

We also wondered why some VH genes had higher footprint frequencies in N1 than others (**Fig. 4b**), as this finding is not similar to what one would expect by random chance. We wondered if the real VH replacement events were hiding somewhere in a large pile of non-VH replacement footprints. A high "false

positive" rate of footprint 5-mers could come about because of sequencing errors. Alternatively or in addition, it may be easy to create false VH replacement 5-mer sequences in primary VDJ rearrangements through a combination of N-addition, nibbling (or sequencing deletion) and the 3' sequence of the VH. For example AAAGA could become AAGA or AGA.

It may be worthwhile to develop a better computational approach for detecting VH replacement footprints with greater specificity for VH replacement. The IgAT software already eliminated footprints that match the germline VH sequence exactly, but this is insufficient, given the high frequency of footprints in N1 of VH6-1. Further specificity might be achievable if one were to limit the detection of footprint 5-mers to sequences that are *unlikely* to arise through a single nucleotide change (for example, those that arise by deletion that converts a non-matching 6-mer to a matching 5-mer, or mutation of a non-matching 5-mer to a footprint 5-mer or N-addition of a nucleotide adjoining a 4-mer to create a footprint 5-mer). An alternative approach is to require that there be two footprint-like sequences in tandem. Either or both of these methods might increase specificity, but could also reduce sensitivity of footprint detection. The validity of either approach would need to be tested further using validated data sets in which VH replacement events are known to have or have not occurred.

We also considered the possibility that footprint 5-mers may frequently arise through some mechanism other than VH replacement. We considered two potential alternative mechanisms-- 1) microhomology mediated joining and 2) cleavage, nibbling and rejoining at the cryptic heptamer.

Alternative Theory 1: Microhomology mediated joining. We considered the possibility that footprints at N1 were arising primarily due to microhomology-mediated joining of similar sequences between the VH and the DH segments. If microhomology-mediated joining were common, one might expect that VHs that share the same 5-mers with DHs are more likely to rearrange, but as shown in **Fig. 5**, this is probably not usually the case. DH5-12 (open bars in **Fig. 5**), which has three footprint 5-mers, does not appear to be used more frequently in rearrangements involving VHs that contain the same 5-mers such as VH2-26 (red arrow). Rather than being skewed towards particular VHs with matching or similar footprints, the frequency of rearrangements of different VHs to DH5-12 rearrangements resembled overall VH usage (**Fig. 5**, closed bars). While this analysis is very preliminary and only focused on a single DH, it suggests that microhomology mediated joining, based upon shared sequences between VH and DH, is not a frequent mechanism for generating footprint 5-mers.

Alternative Theory 2: Cleavage, nibbling and rejoining at the cryptic heptamer in VH. We wondered if there could be cleavage at the cryptic heptamer, followed by exonucleolytic nibbling and re-sealing at the site of the break, without full-blown rearrangement (**Fig. S1** illustrates this idea for a VH6-1 rearrangement). Note that this type of rearrangement product would not involve VH replacement, but would have the result of diversifying the 3' end of the VH in the primary rearrangement product, altering the primary amino acid sequence and/or the reading frame of the rearrangement. This hypothesis makes predictions regarding the sequence characteristics that would be more or less amenable to this type of atypical open and shut joint.<sup>2</sup> For example, one would expect that if most footprint 5-mers at N1 arise by this mechanism that the frequency of footprint 5-mers would be very low in VH genes that lack cryptic heptamers. Furthermore, one would expect that the 5' footprint 5-mer seen in most rearrangements would resemble the 3' end of the germline sequence of the same VH gene that is present in the rearrangement.

Preliminary Conclusions and Caveats from Footprint Analysis. Taken together, these data suggest that many if not most footprint sequences arise by some mechanism(s) other than VH replacement. But there are some caveats to this analysis. First, these data were only obtained on one healthy adult. It is possible that footprints may differ in other individuals or in a minority of individuals. In addition, different findings might occur in individuals with immunologic disorders such as SLE or neoplastic conditions such as B-ALL. Furthermore, only B cells from the peripheral blood were analyzed. It is conceivable that B cell populations with extensive VH replacement reside elsewhere in the body, particularly within the bone marrow. Finally, as discussed above, it is possible that some of the VH replacement footprints that were identified are due to sequencing errors. We tried to protect against this artifact by selectively analyzing unique sequences that are present in at least two copies. But even with this precaution, there are still likely to be many sequencing errors.

#### Conclusions:

VH replacement exchanges the VH within a pre-existing VDJ rearrangement with an upstream VH gene, while preserving most of the original CDR3 sequence. It also sometimes results in the retention of a footprint sequence in the VH gene that was invaded. The result of VH replacement is an alteration in the specificity or functional status of the antibody. But the mechanistic consequence of that alteration is unclear. Is it to diversify the

repertoire once a good CDR3 sequence has been found? Or is it to reduce autoreactivity or generate some form of protective multireactivity? Or is it simply a means by which B cells with non-productive rearrangements on one or both alleles have another shot at creating a productive rearrangement? In humans, the analysis of VH replacement is confounded by not having a means of definitively identifying the precursor rearrangement. Rather, the analysis of VH replacement in humans is accomplished indirectly through footprint analysis, but as demonstrated herein, footprints may arise for reasons other than VH replacement. Thus, while VH replacement certainly occurs, footprint analysis is a poor measure of its frequency because of the high rate of false positives and an unknown rate of false negatives. Nevertheless, it is possible that footprints may provide other insights into the mechanisms of V(D)J recombination and its potentially aberrant regulation in disease states. With the advent of high throughput sequencing studies, further analysis of antibody heavy chain gene rearrangements for VH replacement and other mechanisms of CDR3 diversification promise to be illuminating.

Acknowledgments:

We acknowledge the Arthritis Foundation (WM), the Lupus Research Institute (ELP), the Hematology Translational Center of Excellence at the Perelman School of Medicine of the University of Pennsylvania, NIH T32 CA009615 (AG), the Institute for Translational Medicine and Therapeutics at the University of Pennsylvania (AG) and NIH R25-DK066028 (SJ) for funding this work.

## Figure Legends:

**Fig. 1a. VH replacement:** an upstream VH gene invades by rearranging into a pre-existing rearrangement. RAG cleaves the conventional recombination signal sequence (black triangle) of the invading VH (light blue VH gene) and cleaves at a cryptic heptamer sequence (cRSS, dashed white triangle) of the invaded VH gene (yellow VH gene). The resulting rearrangement is shown on the second line of the diagram and includes the DH and JH genes of the previous rearrangement and the new VH gene. Also included in the VH replacement product is a remnant or "footprint" of the preceding VH gene (denoted by a yellow box). Often the products of VH replacement exhibit CDR3 elongation, due to the retention of the footprint sequence (the CDR3 is indicated by the bar under the sequence and the added length of the new CDR3 sequence including the footprint (in red) is indicated by the black bar below the sequence.) **Fig. 1b) Serial VH replacement.** The same conventions are used in Fig. 1a and a longer CDR3 is generated, via the accumulation of footprint sequences. In both panels, boxes denote exons, lines introns, triangles RSS and the rearrangement is indicated by dashed black lines. This diagram is not drawn to scale. **Fig. 1c. Long CDR3 sequence with possible VH replacement(s).** Shown is the nucleotide sequence of an expanded B cell clone that was recovered from peripheral blood DNA of a patient with systemic lupus erythematosus (SLE) that reveals a 91 nucleotide CDR3. Kowal et al. described an anti-dsDNA H chain sequence comprised of a VH3-N-DH2-2-JH6, which has similar features to this junction, although it was shorter<sup>53</sup>. Sequences in black font match the corresponding germline gene segments. Red font denotes possible N-additions and yellow shading highlights potential footprint sequences. Dashes indicate regions where sequences do not overlap. FR3 = framework region 3; CDR3 = third complementarity determining sequence; FR4 = framework region 4.

**Fig. 2a. Longer CDR3 sequences have more footprints.** Plotted are the average numbers of footprint 5-mers per sequence. Sequences are averaged at each CDR3 size by the number of sequences that have CDR3s of that particular length. Blue dots are in-frame (IF) rearrangements and red dots are out of frame (OF) rearrangements. We find a positive linear correlation between the length of the CDR3 and the number of footprints ( $r^2=0.89$  for IF,  $r^2=0.7$  for OF). The line describing this relationship has a slope of  $0.06 \pm 0.008$  for OF (red line) and  $0.05 \pm 0.008$  for IF rearrangements (blue line). **Insets (Fig. 2b and Fig. 2c)** are two examples of the entire distribution of footprint numbers at two positions - position 56 (red circle OF) and position 69 (blue circle IF). Black stems indicate the numbers of footprints and red lines represent the fit of a Poisson distribution ( $\lambda = 1.57$  and  $2.74$  respectively).



**Fig. 3: Positions of Footprint 5-mers.** **Fig. 3a.** Frequency distribution of all footprint 5-mers out of total unique sequences (n=42,221) plotted against the normalized CDR3 position. The CDR3 is herein defined to begin at the conserved CAR amino acid sequence (TGT GCG AGA nucleotide sequence) within the 3' end of VH and end at the conserved W (TGG nucleotide sequence) that is immediately upstream of the first conserved glycine, GGC nucleotide sequence) within the JH. 'TGGAG' is excluded as it is found in many alleles of all DH genes. The position of the footprint is defined by where the footprint starts within the CDR3. For example, if a footprint occupies nucleotides 12-16 of the CDR3, it will be plotted at position 12. CDR3 lengths were normalized to a scale of 1-100 using  $p1 = p/(L*100)$  where p is the position of the 5-mer in the real CDR3 sequence, L is the length of the CDR3 sequence and p1 is the normalized position. Normalized positions are rounded to the nearest integer. **Fig. 3b.** Frequency distribution of all footprint 5-mers plotted against the normalized CDR3 position, corrected for footprint 5-mers in the germline JH6 gene. 'ATGGA', 'TACGG' and 'CATGG' were excluded when found in the 3' end of the CDR3 within the JH6 gene, as these 5-mers are found in the germline JH6 sequence. Footprint 5-mers found in other JHs (see table 2) are not counted in either Fig. 3a or Fig. 3b because they are all located outside CDR3 region of JH.

**Fig. 4a. Percentage of rearrangements with footprints for the ten most frequent VH rearrangements and VH6-1.** The percentages of rearrangements that contain at least one footprint 5-mer in either (or both) the 5' end or the 3' end of the rearrangement are shown for unique rearrangements for each of the ten most common VH genes and for VH6-1. CDR3 sequences are defined as described in the legend to Fig. 4 and normalized to an arbitrary length of 100. The 5' end (which almost always contains N1) is defined as the first 20% of the sequence and the 3' end is defined as the last 20% of the sequence (which almost always contains N2). VH6-1 is the most 3' VH gene and cannot contain footprints that are due to VH replacement. Black bars denote unique rearrangements that are in frame (IF VH), white bars denote out of frame rearrangements (OF VH) and grey bars indicate total unique rearrangements (total VH). **Fig. 4b: Footprint Frequency of In-Frame VH rearrangements in N1 using IgAT Software.** The same unique in-frame (IF) VH rearrangements as shown in Fig. 4a, were analyzed for the presence of one or more footprints using IgAT software.<sup>2</sup> Plotted are the frequencies of unique IF VH rearrangements that have one or more footprints in the N1 region.

**Fig. 5. VH usage is similar in all unique rearrangements and in rearrangements with DH5-12.** VH usage is shown for total unique sequences (closed bars, n=42,221) and for those with DH5-12 (open bars, n=1,029). Sequences containing the DH5-12 gene segment in the CDR3 region were recognized using IMGT high V quest analysis, and further analysis on VH usage was performed in-house (see text). The red arrow points to VH2-26, which

contains some of the same footprint 5-mers as DH5-12, but does not appear to be used more frequently in rearrangements that include DH5-12.

## Table Legends

**Table 1: Footprint sequences in the 3' end of human germline VH genes and alleles.** 273 functional VH genes, including alleles and sequences designated as open reading frames, were downloaded from the IMGT database<sup>54</sup> and manually scanned for footprints. A footprint is defined by the nucleotide sequence following the cryptic recombination signal sequence (cRSS), TACTGTG, at the 3' end of each VH, and is listed in the left column of the table. The footprint 5-mer variants that were used to scan the sequences are included in parentheses. The right column lists the VH genes and alleles. An asterisk in the VH name refers to a specific allele. If all alleles of a particular VH have the same footprint, allele names are omitted. Overlapping footprints are listed for some footprints in the sub-column on the right. Footprints in red font are also found in germline DH and JH gene segments. Sequences with ambiguous nucleotide designations (N-nucleotides), are highlighted in yellow. The following VH genes do not have a cRSS: VH1-18\*2, VH1-69\*3, VH1-69\*7, VH1-c, VH1-f\*2, VH2-5\*2, VH2-5\*3, VH2-70\*2, VH2-70\*3, VH2-70\*4, VH2-70\*5, VH2-70\*6, VH2-70\*7, VH2-70\*8, VH2-70\*13, VH3-15, VH3-20, VH3-25\*4, VH3-49, VH3-72\*2, VH3-73\*1, VH3-73\*2, VH3-d, VH3/OR16-13, VH3/OR16-6\*2, VH4-30-4\*3, VH4-30-4\*4, VH4-31\*6, VH4-31\*7, VH4-31\*8, VH4-31\*9, VH4-34\*3, VH4-34\*6, VH4-34\*7, VH4-34\*11, VH4-39\*3, VH4-39\*4, VH4-4\*1, VH4-4\*3, VH4-4\*4, VH4-4\*5, VH4-61\*4, VH5-51\*5, VH7-4-1\*3.

**Table 2. Footprint sequences in DH and JH alleles.** 34 germline functional human DH alleles and 13 JH alleles were downloaded from the IMGT database.<sup>54</sup> Each allele was scanned for all of the possible 5 nucleotide footprint motifs listed in **Table 1**. Sequences containing 5 nucleotide footprints are given in red font. In some cases the region matches more than one possible footprint (for example, D5-12\*01 contains three different footprints: GGATA, GATAT and ATATA).

**Table 3: The number of footprints found in various regions of human VH genes.** 234 functional germline human VH alleles were downloaded from the IMGT database.<sup>54</sup> For each footprint, the number of times that footprint was found in the VH alleles was recorded. The first column lists the footprint, the second column lists the number of alleles in which the footprint is found. Some alleles contain more than one footprint. The remaining columns indicate how many footprints were found in each of the corresponding regions of the V region. FR = framework; CDR = complementarity determining region. The columns named FR and CDR provide the total number of times that a particular footprint is found in the FR or the CDR.

## Electronic Supplement: Sequencing and data analysis methods

Target enrichment was performed using a dual step, PCR based amplicon capture. Round 1 captured VDJ rearrangements from peripheral blood B cell DNA using multiplex primer cocktails that were modified to include a universal illumina adapter. Individual sample barcodes and illumina sequencing adapters were introduced during the second round of PCR. Equimolar amounts of each sample library were subsequently sequenced on the Illumina Mi-seq using 2 by 200 base pair reads.

Sequences were filtered for length, presence of barcodes, trimmed and analyzed for their nearest unique VH, DH, JH and CDR3 translated sequences using both our own alignment tools and the IMGT high throughput server.<sup>1</sup> To remove the baseline noise of sequencing errors, unique sequences were defined as those that differed by at least 3 nt from any other sequence. We further removed all sequences that had only a single copy in the dataset. This led to the identification of 42,221 unique heavy chain sequences. Taking this dataset of unique sequences, we searched for all the instances of perfect matches of all of the putative footprints or footprint sequence derived pentamers ("footprint 5-mers"). For comparison, in some instances sequences were further analyzed using IgAT (see text).<sup>2</sup>

**Fig. S1: Model for Atypical Open and Shut Joints.** Footprint 5-mers can be generated by cleaving at the cryptic heptamer, followed by preferential trimming by exonucleolytic nibbling at the 3' end of the double strand break. Shown is the generation of a footprint sequence in a VH6-1 rearrangement. The cryptic heptamer is indicated by a dashed triangle, colored squares indicate the VH, DH and JH gene segments (not drawn to scale). According to this model, there is cleavage at the cryptic heptamer, followed by nibbling at the 3' end of the break (red wavy line), leading to selective loss of the A residue. The third line of the figure shows the repaired rearrangement, in which the A residue is missing from the rearrangement.

### **References (for supplement only):**

1. Alamyar, E., Duroux, P., Lefranc, M.P. & Giudicelli, V. IMGT(®) tools for the nucleotide analysis of immunoglobulin (IG) and T cell receptor (TR) V-(D)-J repertoires, polymorphisms, and IG mutations: IMGT/V-QUEST and IMGT/HighV-QUEST for NGS. *Methods Mol Biol.* **882**:569-604 (2012).
2. Rogosch, T., Kerzel, S., Hoi, K.H., Zhang, Z., Maier, R.F., Ippolito, G.C. & Zemlin, M. Immunoglobulin Analysis Tool: a novel tool for the analysis of human and mouse heavy and light chain transcripts. *Front. Immun.* **3**:176. (2012).

## References

1. Jung, D., Giallourakis, C., Mostoslavsky, R. & Alt, F.W. Mechanism and control of V(D)J recombination at the immunoglobulin heavy chain locus. *Annu Rev Immunol* **24**, 541-570 (2006).
2. Gellert, M. V(D)J recombination: RAG proteins, repair factors, and regulation. *Annual review of biochemistry* **71**, 101-132 (2002).
3. Vinocur, J.M., Fesnak, A.D., Liu, Y., Charan, D. & Prak, E.T. Violations of the 12/23 rule at the mouse immunoglobulin kappa locus, including V kappa-V kappa rearrangement. *Molecular immunology* **46**, 2183-2189 (2009).
4. Bassing, C.H. *et al.* Recombination signal sequences restrict chromosomal V(D)J recombination beyond the 12/23 rule. *Nature* **405**, 583-586 (2000).
5. Briney, B.S., Willis, J.R., Hicar, M.D., Thomas, J.W., 2nd & Crowe, J.E., Jr. Frequency and genetic characterization of V(DD)J recombinants in the human peripheral blood antibody repertoire. *Immunology* **137**, 56-64 (2012).
6. Luning Prak, E.T., Monestier, M. & Eisenberg, R.A. B cell receptor editing in tolerance and autoimmunity. *Annals of the New York Academy of Sciences* **1217**, 96-121 (2011).
7. Muramatsu, M. *et al.* Class switch recombination and hypermutation require activation-induced cytidine deaminase (AID), a potential RNA editing enzyme. *Cell* **102**, 553-563 (2000).
8. Reynaud, C.A., Garcia, C., Hein, W.R. & Weill, J.C. Hypermutation generating the sheep immunoglobulin repertoire is an antigen-independent process. *Cell* **80**, 115-125 (1995).
9. Darlow, J.M. & Stott, D.I. Gene conversion in human rearranged immunoglobulin genes. *Immunogenetics* **58**, 511-522 (2006).
10. Duvvuri, B. & Wu, G.E. Gene Conversion-Like Events in the Diversification of Human Rearranged IGHV3-23\*01 Gene Sequences. *Frontiers in immunology* **3**, 158 (2012).
11. Darlow, J.M. & Stott, D.I. V(H) replacement in rearranged immunoglobulin genes. *Immunology* **114**, 155-165 (2005).
12. Kleinfeld, R.W. & Weigert, M.G. Analysis of VH gene replacement events in a B cell lymphoma. *J Immunol* **142**, 4475-4482 (1989).
13. Chen, C., Nagy, Z., Prak, E.L. & Weigert, M. Immunoglobulin heavy chain gene replacement: a mechanism of receptor editing. *Immunity* **3**, 747-755 (1995).
14. Golub, R. *et al.* VH gene replacement in thymocytes. *J Immunol* **166**, 855-860 (2001).
15. Taki, S., Schwenk, F. & Rajewsky, K. Rearrangement of upstream DH and VH genes to a rearranged immunoglobulin variable region gene inserted into the DQ52-JH region of the immunoglobulin heavy chain locus. *Eur J Immunol* **25**, 1888-1896 (1995).
16. Wilson, P.C. *et al.* Receptor revision of immunoglobulin heavy chain variable region genes in normal human B lymphocytes. *The Journal of experimental medicine* **191**, 1881-1894 (2000).

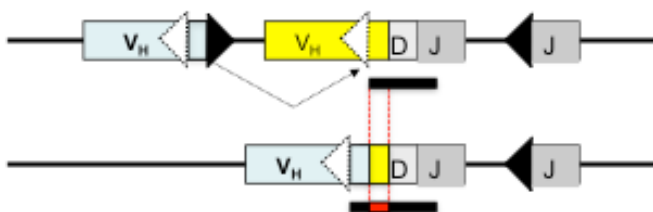
17. Itoh, K. *et al.* Immunoglobulin heavy chain variable region gene replacement As a mechanism for receptor revision in rheumatoid arthritis synovial tissue B lymphocytes. *The Journal of experimental medicine* **192**, 1151-1164 (2000).
18. Lenze, D. *et al.* Receptor revision of immunoglobulin heavy chain genes in human MALT lymphomas. *Molecular pathology : MP* **56**, 249-255 (2003).
19. Ouled-Haddou, H. *et al.* Characterization Of A New V Gene Replacement In The Absence Of Activation-Induced Cytidine Deaminase And Its Contribution To Human Bcr Diversity. *Immunology* (2013).
20. Zhang, Z. *et al.* Molecular mechanism of serial VH gene replacement. *Annals of the New York Academy of Sciences* **987**, 270-273 (2003).
21. Reth, M., Gehrman, P., Petrac, E. & Wiese, P. A novel VH to VHDJH joining mechanism in heavy-chain-negative (null) pre-B cells results in heavy-chain production. *Nature* **322**, 840-842 (1986).
22. Cascalho, M., Wong, J. & Wabl, M. VH gene replacement in hyperselected B cells of the quasimonoclonal mouse. *J Immunol* **159**, 5795-5801 (1997).
23. Lutz, J., Muller, W. & Jack, H.M. VH replacement rescues progenitor B cells with two nonproductive VDJ alleles. *J Immunol* **177**, 7007-7014 (2006).
24. Rahman, N.S., Godderz, L.J., Stray, S.J., Capra, J.D. & Rodgers, K.K. DNA cleavage of a cryptic recombination signal sequence by RAG1 and RAG2. Implications for partial V(H) gene replacement. *The Journal of biological chemistry* **281**, 12370-12380 (2006).
25. Koralov, S.B., Novobrantseva, T.I., Konigsmann, J., Ehlich, A. & Rajewsky, K. Antibody repertoires generated by VH replacement and direct VH to JH joining. *Immunity* **25**, 43-53 (2006).
26. Liu, Y. *et al.* Lupus susceptibility genes may breach tolerance to DNA by impairing receptor editing of nuclear antigen-reactive B cells. *J Immunol* **179**, 1340-1352 (2007).
27. Park, Y.H. & Osmond, D.G. Dynamics of early B lymphocyte precursor cells in mouse bone marrow: proliferation of cells containing terminal deoxynucleotidyl transferase. *Eur J Immunol* **19**, 2139-2144 (1989).
28. Davila, M. *et al.* Multiple, conserved cryptic recombination signals in VH gene segments: detection of cleavage products only in pro B cells. *The Journal of experimental medicine* **204**, 3195-3208 (2007).
29. Vettermann, C., Herrmann, K. & Jack, H.M. Powered by pairing: the surrogate light chain amplifies immunoglobulin heavy chain signaling and pre-selects the antibody repertoire. *Semin Immunol* **18**, 44-55 (2006).
30. Keenan, R.A. *et al.* Censoring of autoreactive B cell development by the pre-B cell receptor. *Science* **321**, 696-699 (2008).
31. Ota, M. *et al.* Regulation of the B cell receptor repertoire and self-reactivity by BAFF. *J Immunol* **185**, 4128-4136 (2010).
32. Hondowicz, B.D. *et al.* The role of BLYS/BLYS receptors in anti-chromatin B cell regulation. *Int Immunol* **19**, 465-475 (2007).
33. Liu, J. *et al.* Regulation of VH Replacement by B Cell Receptor-Mediated Signaling in Human Immature B Cells. *J Immunol* **190**, 5559-5566 (2013).

34. Sekiguchi, D.R., Eisenberg, R.A. & Weigert, M. Secondary heavy chain rearrangement: a mechanism for generating anti-double-stranded DNA B cells. *The Journal of experimental medicine* **197**, 27-39 (2003).
35. Eisenberg, R. The chronic graft-versus-host model of systemic autoimmunity. *Curr Dir Autoimmun* **6**, 228-244 (2003).
36. Brard, F., Shannon, M., Prak, E.L., Litwin, S. & Weigert, M. Somatic mutation and light chain rearrangement generate autoimmunity in anti-single-stranded DNA transgenic MRL/lpr mice. *The Journal of experimental medicine* **190**, 691-704 (1999).
37. Khan, S.N. *et al.* Editing and escape from editing in anti-DNA B cells. *Proc Natl Acad Sci U S A* **105**, 3861-3866 (2008).
38. Sekiguchi, D.R. *et al.* Development and selection of edited B cells in B6.56R mice. *J Immunol* **176**, 6879-6887 (2006).
39. Yunk, L., Meng, W., Cohen, P.L., Eisenberg, R.A. & Luning Prak, E.T. Antibodies in a heavy chain knock-in mouse exhibit characteristics of early heavy chain rearrangement. *J Immunol* **183**, 452-461 (2009).
40. Coleclough, C., Perry, R.P., Karjalainen, K. & Weigert, M. Aberrant rearrangements contribute significantly to the allelic exclusion of immunoglobulin gene expression. *Nature* **290**, 372-378 (1981).
41. Wardemann, H. *et al.* Predominant autoantibody production by early human B cell precursors. *Science* **301**, 1374-1377 (2003).
42. Klonowski, K.D., Primiano, L.L. & Monestier, M. Atypical VH-D-JH rearrangements in newborn autoimmune MRL mice. *J Immunol* **162**, 1566-1572 (1999).
43. Klonowski, K.D. & Monestier, M. Heavy chain revision in MRL mice: a potential mechanism for the development of autoreactive B cell precursors. *J Immunol* **165**, 4487-4493 (2000).
44. Chen, C., Bruderer, U. & Rittenberg, M.B. The developmental patterns of B cell precursors distinguishing between environmental and nonenvironmental forms of phosphocholine. *Cell Immunol* **143**, 378-388 (1992).
45. Binder, C.J. *et al.* Pneumococcal vaccination decreases atherosclerotic lesion formation: molecular mimicry between *Streptococcus pneumoniae* and oxidized LDL. *Nat Med* **9**, 736-743 (2003).
46. Watson, L.C. *et al.* Paucity of V-D-D-J rearrangements and VH replacement events in lupus prone and nonautoimmune TdT<sup>-/-</sup> and TdT<sup>+/+</sup> mice. *J Immunol* **177**, 1120-1128 (2006).
47. Eisenberg, R.A. Secondary receptor editing in the generation of autoimmunity. *Autoimmun Rev* **11**, 787-789 (2012).
48. Choi, Y. *et al.* Clonal evolution in B-lineage acute lymphoblastic leukemia by contemporaneous VH-VH gene replacements and VH-DJH gene rearrangements. *Blood* **87**, 2506-2512 (1996).
49. Gawad, C. *et al.* Massive evolution of the immunoglobulin heavy chain locus in children with B precursor acute lymphoblastic leukemia. *Blood* **120**, 4407-4417 (2012).

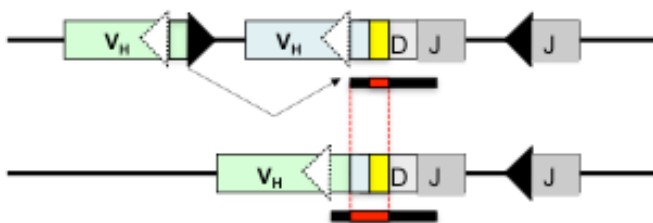
50. van der Velden, V.H. *et al.* Detection of minimal residual disease in hematologic malignancies by real-time quantitative PCR: principles, approaches, and laboratory aspects. *Leukemia* **17**, 1013-1034 (2003).
51. Faham, M. *et al.* Deep-sequencing approach for minimal residual disease detection in acute lymphoblastic leukemia. *Blood* **120**, 5173-5180 (2012).
52. Zhang, Z. *et al.* Contribution of Vh gene replacement to the primary B cell repertoire. *Immunity* **19**, 21-31 (2003).
53. Kowal, C., Weinstein, A. & Diamond, B. Molecular mimicry between bacterial and self antigen in a patient with systemic lupus erythematosus. *Eur J Immunol* **29**, 1901-1911 (1999).
54. Lefranc, M.P. *et al.* IMGT, the international ImMunoGeneTics information system. *Nucleic Acids Res* **37**, D1006-1012 (2009).



**Fig. 1a: VH replacement**



**Fig. 1b: serial VH replacement**



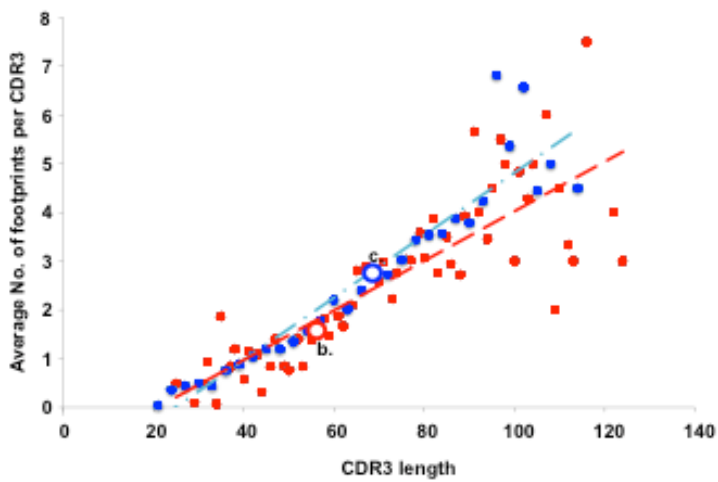
**Fig. 1c: Long CDR3 sequence with possible VH replacement(s).**

```

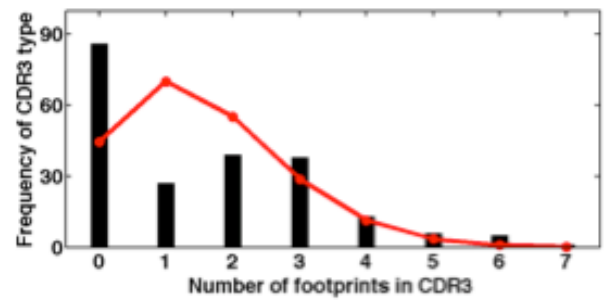
1 -----FR3-----><-----CDR3----- 60
S1E ACACGGCCGTGTATTACTGTGCAC AAAGCCC AAAG CAAATATATTGTAGTAGTACC
VH3-23 ACACGGCCGTGTATTACTGTGC AAAG
D2-2 -----AGGATATTGTAGTAGTACC
JH6 -----

61 -----CDR3-----><-----FR4----- 120
S1E AGCTGCCACOGGATACCTCTTTATTACTACTACTACTACGGTATGGACGTCTGGGGCCAA
VH3-23 AGCTGCCACOGGATACCTCTTTATTACTACTACTACTACGGTATGGACGTCTGGGGCCAA
D2-2 AGCTGCTATACC-----
JH6 -----ATTACTACTACTACTACGGTATGGACGTCTGGGGCCAA
    
```

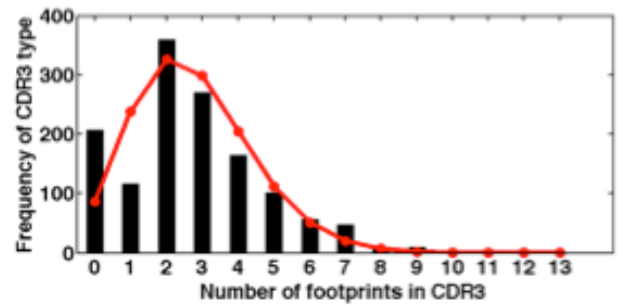
**Fig. 2a:** Longer CDR3 sequences have more footprints



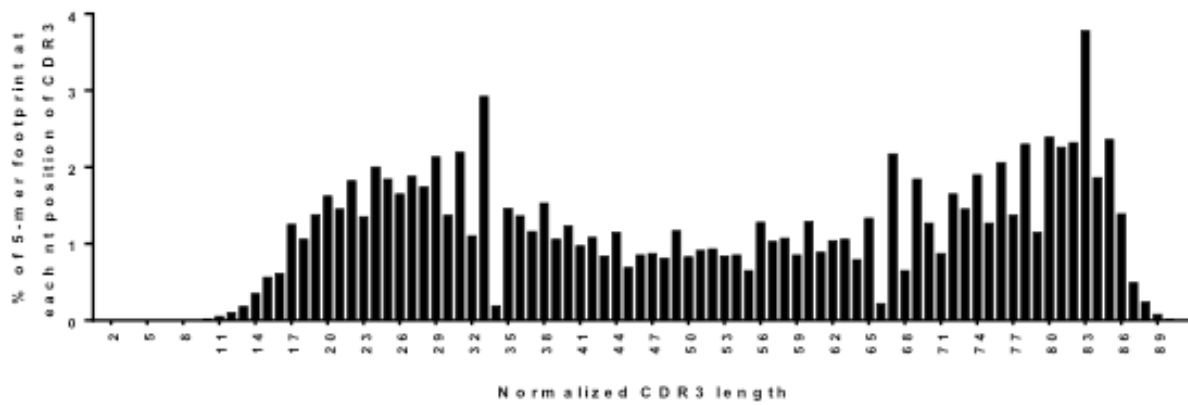
**Fig. 2b:** Footprint frequency distribution in sequences with normalized CDR3 length = 56.



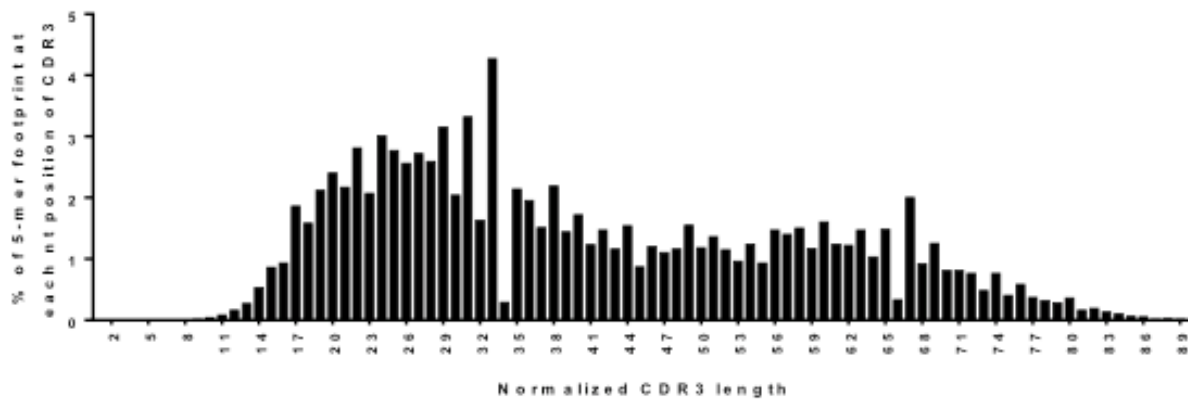
**Fig. 2c:** Footprint frequency distribution in sequences with normalized CDR3 length = 69.



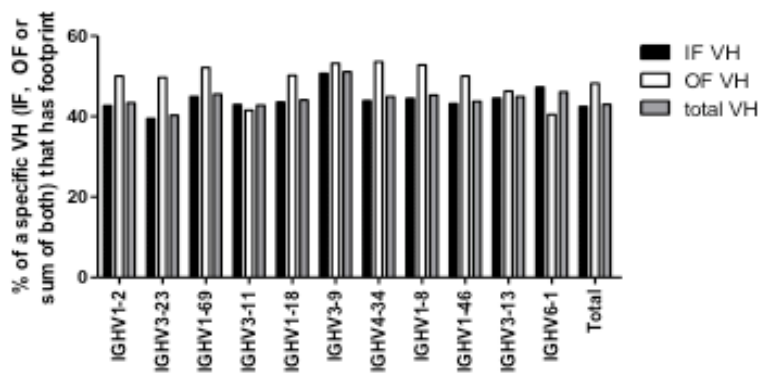
**Fig. 3a:** Position of footprint 5-mers in all unique rearrangements.



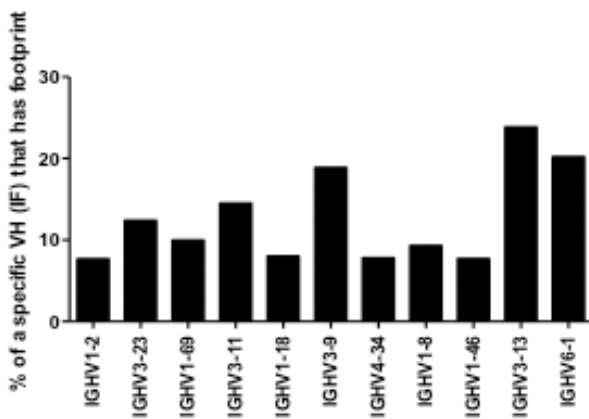
**Fig. 3b:** Position of footprint 5-mers in all unique sequences, corrected for footprints in JH6.



**Fig. 4a:** Percentage of rearrangements with footprints throughout the CDR3 for the ten most frequent VH rearrangements and VH6-1.



**Fig. 4b:** Percentage of in-frame rearrangements with footprints in N1 (V to D junction) for the ten most frequent VH rearrangements and VH6-1.



**Fig. 5** VH usage in all unique vs. DH5-12 containing rearrangements

