



HAL
open science

Independent recruitment of Igh alleles in V(D)J recombination.

Clara F Alves-Pereira, Raquel de Freitas, Telma Lopes, Rui Gardner, Filipa Marta, Paulo Vieira, Vasco M Barreto

► **To cite this version:**

Clara F Alves-Pereira, Raquel de Freitas, Telma Lopes, Rui Gardner, Filipa Marta, et al.. Independent recruitment of Igh alleles in V(D)J recombination.. Nature Communications, 2014, 5, pp.5623. 10.1038/ncomms6623 . pasteur-01164712

HAL Id: pasteur-01164712

<https://pasteur.hal.science/pasteur-01164712>

Submitted on 17 Jun 2015

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

ARTICLE

Received 14 Jul 2014 | Accepted 20 Oct 2014 | Published 17 Dec 2014

DOI: 10.1038/ncomms6623

OPEN

Independent recruitment of *Igh* alleles in V(D)J recombination

Clara F. Alves-Pereira¹, Raquel de Freitas¹, Telma Lopes², Rui Gardner², Filipa Marta¹, Paulo Vieira^{3,4} & Vasco M. Barreto¹

How the vast majority of B cells express only one of the two alleles at their immunoglobulin loci remains a biological puzzle. Here, in mice reconstituted with a single haematopoietic stem cell, we demonstrate that each of the two immunoglobulin heavy chain (*Igh*) alleles has a similar probability to be the first to undergo V_H to DJ_H rearrangement. We also observe this similar probability in clones from multipotent and common lymphoid precursors. The extreme biases in the expression of the alleles that we find in more differentiated subsets are mostly due to constraints imposed by early rearrangements. Our data demonstrate that each of the two *Igh* alleles in a B cell behaves independently of the other, up to the moment when a successful rearrangement in one allele triggers a feedback mechanism that prevents further recombination.

¹Epigenetics and Soma Laboratory, Instituto Gulbenkian de Ciência, Rua da Quinta Grande, nº 6, 2780-156 Oeiras, Portugal. ²Flow Cytometry Laboratory UIC, Instituto Gulbenkian de Ciência, Rua da Quinta Grande, nº 6, 2780-156 Oeiras, Portugal. ³Unité Lymphopoïèse, Institut Pasteur, 25, Rue du Dr Roux, 75724 Paris, France. ⁴INSERM U668, F-75015 Paris, France. Correspondence and requests for materials should be addressed to V.M.B. (email: vbarreto@igc.gulbenkian.pt).

Each B cell expresses immunoglobulin heavy chains at its surface, typically encoded by only one of the two alleles¹. The question of why this phenomenon of allelic exclusion occurs has generated a number of non-exclusive and untested answers. For instance, it could be a solution to produce one or a limited set of antigen receptors per cell, as postulated by the clonal selection theory², or a safety mechanism to decrease the chances of detrimental translocations³, since antigen receptor genes undergo somatic recombination⁴. How allelic exclusion arises would seem to be more tractable; however, this is an ongoing question since the 1960s. Mostly because of the imprecise junction of the V_H , D_H and J_H gene segments that are found in the immunoglobulin heavy chain (*Igh*) locus, only a minor proportion of the rearrangements will be 'productive' or $V_HDJ_H^+$ (encoding a full-length chain), which intrinsically ensures some level of allelic exclusion. However, for the murine *Igh* locus, a minimal stochastic model of allelic exclusion based exclusively on the outcome of the rearrangement process⁵ is insufficient to account for the degree of exclusion observed at the surface of the cells⁶.

The D_H to J_H rearrangement step occurs in both *Igh* alleles⁷. This event has recently been proposed to mark the DJ_H junction for secondary rearrangement, that is, the V_H to DJ_H stage⁸, and once a productive secondary rearrangement takes place in one of the alleles of the pro-B cell, expression of the pre-B-cell receptor (pre-BCR) at the surface signals a series of events leading to the shutdown of recombination at the second allele. This feedback mechanism, first suggested for the immunoglobulin kappa locus⁹, and subsequently elaborated for the *Igh* locus¹⁰, has been independently confirmed in a number of ways. Notably, a μ H chain encoded by a transgene or a knocked-in allele inhibits V_H to DJ_H rearrangement at the endogenous *Igh* locus only if expressed at the surface, as part of the pre-BCR^{11–13}. In addition, in cells with genotypic inclusion (with both alleles rearranged productively) only one of the chains pairs efficiently with the surrogate light chain¹⁴ that also forms the pre-BCR and the frequency of cells with genotypic inclusion ($V_HDJ_H^+/V_HDJ_H^+$) increases in the absence of the surrogate light chain¹⁵ or when the signalling from an intact pre-BCR is perturbed by mutations in one of its downstream targets¹⁶. More recently, two additional feedback mechanisms have been proposed, one triggered by an Ataxia Telangiectasia Mutated (ATM)-dependent sensing of the rearrangement at the first *Igh* allele³ and the other by a stable μ H mRNA with a premature stop codon that cannot be translated into a μ H chain¹⁷. Overall, feedback inhibition and the rearrangement of the second allele in its absence largely explain the proportion of $V_HDJ_H^+/DJ_H$ and $V_HDJ_H^+/V_HDJ_H^+$ IgM⁺ B cells¹⁸.

The events initially leading to V_H to DJ_H rearrangement that render the V_H regions 'accessible' for rearrangement include the relocalization of the *Igh* locus away from the nuclear periphery in pro-B cells¹⁹ and long-range interactions mediated by looping of the *Igh* domains^{20,21} that depend on regulatory regions, namely the one located in the V_H to D_H intergenic region²². However, it is unclear whether a specific mechanism ensures that each cell only tests one V_H to DJ_H rearrangement at a time. The acquisition of competence to undergo rearrangement by the *Igh* locus in the course of B-cell differentiation has been described at the molecular level (for example, ref. 23); however, typically these studies are carried out in bulk cultures, averaging out any existing differences between the alleles at the single-cell level. In turn, imaging-assisted studies (for example, refs 19–21,24) can discriminate between the alleles at the single-cell level, but measure features that are only correlates of $V_H(D_H)J_H$ accessibility, which are not always confirmed by different techniques²⁵, and provide mere snapshots of a process that is essentially dynamic.

Here we performed a clonal analysis of *Igh* allelic exclusion, and provide conclusive evidence that the allelic exclusion of immunoglobulin receptor loci differs from X-chromosome inactivation, as no stable epigenetic mark is propagated until pro-B cells start rearranging. More importantly, we also show that the *Igh* alleles rearrange independently from each other in the majority of clones derived from common lymphoid precursors.

Results

Clonal analysis tracks V(D)J recombination. Clonal analysis is typically used to investigate the commitment of precursor cells to a given lineage or set of lineages. We have adapted this technique to test whether an *Igh* allele from a precursor cell is committed to recombine a V_H segment before the other allele as the cell divides and enters the lymphoid lineage. As shown in Fig. 1, although in a bulk culture it is not possible to conclude whether the alleles from cells heterozygous for the *Igh* locus ($Igh^{a/b}$) are differently committed for rearrangement, in isolated clones the commitment and non-commitment scenarios produce sharply different outcomes. In the case of commitment, all the cells expressing the μ chain from the first allele to rearrange (in Fig. 1a, IgM^a) would have the second allele in the non- V_H -rearranged DJ_H configuration and the remaining IgM⁺ cells (in Fig. 1a, IgM^b) would have the second allele in the V_HDJ_H configuration. In contrast, if commitment does not occur (Fig. 1b), the IgM^a and IgM^b pools of cells would not differ in the status of rearrangement of the silenced allele, since each pool would have the same proportion of IgM⁺ cells with a unique V_H to DJ_H rearrangement versus cells that were rescued by the second V_H to DJ_H rearrangement. Thus, D_H to J_H and V_H to DJ_H junctions function as a natural reporter system, since the signature junction sequences somatically introduced in the *Igh* locus help to ascertain when rearrangement occurred in each clone as the cells divide.

Both *Igh* alleles can rearrange first in clones derived from HSC.

The main feature of the *Igh* locus and other antigen receptor loci is V(D)J recombination; however, these genes are frequently included in a growing group of autosomal genes displaying what has been called random monoallelic expression²⁶. Unlike loci under parental origin-specific imprinting, in these genes roughly 50% of the cells express the maternal allele and the remaining 50% express the paternal allele. Since the parallel with X-chromosome inactivation has been made in several reports^{27–29} and haematopoietic stem cells (HSCs) show X-chromosome inactivation, we first asked whether in mice with single-HSC-derived haematopoietic systems, the emerging IgM^a and IgM^b populations would also show signs of a commitment that is already present in the HSC and that is then stably propagated.

Cells heterozygous for a pan-leukocitary marker polymorphism (CD45.1/45.2), the *Igh* allotypes ($Igh^{a/b}$) and the X-linked gene *Foxp3* and its *gfp*-fused knock-in (*Foxp3^{gfp/wt}*) were single-cell-sorted from the HSC-long-term repopulation potential-enriched Lin[−] Rho[−] CD45^{int} Hoe[−] side population (SP) pool of C57BL/6J female bone marrows (BMs; see Methods; Supplementary Fig. 1). Sublethally irradiated B6-*Rag2*^{−/−} was then single-cell-reconstituted and monitored over time to detect animals showing chimerism in the blood. The percentage of reconstituted animals was 4.8–22.9%. This low efficiency is in the range of what has been described³⁰, and further increases the likelihood of obtaining a collection mainly composed of animals with monoclonal haematopoietic systems. Animals with chimerism ($N=15$) were euthanized within 8–18 weeks from the single-cell transplant, analysed by flow cytometry for the presence of lymphoid and myeloid cells and in some cases used to produce secondary

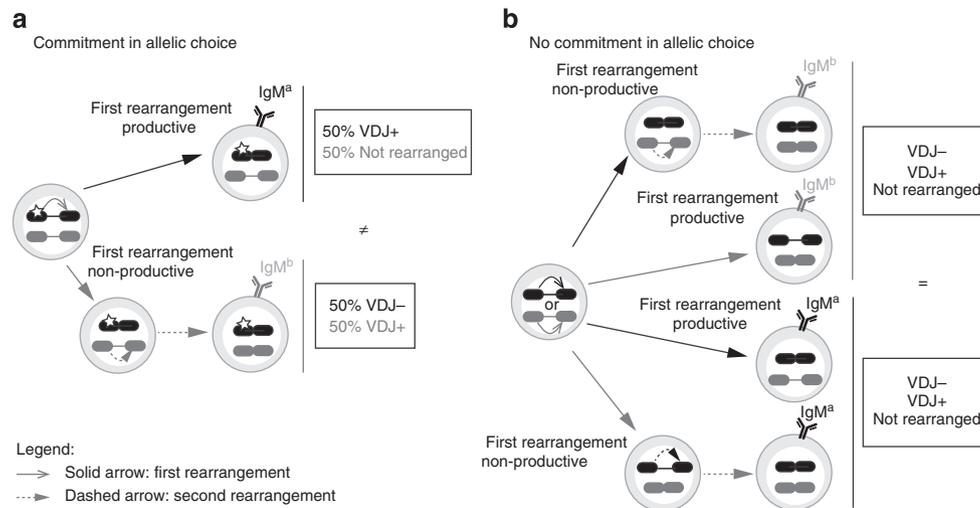


Figure 1 | Rationale. The commitment and no commitment models of *Igh* allelic choice predict a distinctive V_H to DJ_H rearrangement status in the two IgM^a and IgM^b fractions of B-cell clonal populations. **(a)** In a B-cell clone that is pre-determined to rearrange first a given allele (represented by a black locus with a white star), all the cells rearrange the marked allele first and the second allele (grey locus) remains non-rearranged, unless the previous rearrangement is non-productive ($V_H DJ_H^-$). The resulting clone contains two populations of B cells: one that expresses the product of the first rearrangement (a productive rearrangement, $V_H DJ_H^+$) and retains 50% of the alleles non-rearranged (IgM^a population in the scheme) and another that expresses the product of the second rearrangement (IgM^b population in the scheme) and has both alleles rearranged. **(b)** The alternative model (no commitment) assumes that both alleles are equally accessible, each with equal chance of rearranging first. This condition renders the IgM^a and IgM^b populations of a cell clone undistinguishable at the structure of the silent *Igh* allele, since both populations will have a similar frequency of silent alleles in the $V_H DJ_H$ and DJ_H configurations. The present scheme assumes the ordered model of $V(D)J$ rearrangement⁷, and represents the V_H to DJ_H rearrangement only (the D-to-J rearrangement is biallelic and has occurred before the developmental stage illustrated). In a sorted population of IgM^+ cells, the frequency of non-rearranged alleles can be measured by quantifying the V_H-D_H intergenic fragment retention within the *Igh* locus.

recipients, attesting to their long-term repopulation potential (Table 1). The collection analysed for allelic exclusion comprised animals that were long-term and short-term reconstituted.

Female donor cells heterozygous both for the X-linked *Foxp3* and *Igh* loci were used to compare in the same experiment X-chromosome inactivation with *Igh* allelic exclusion. FOXP3 is a transcription factor expressed by regulatory T cells (Tregs). Owing to X-chromosome inactivation, in heterozygous animals for a knock-in green fluorescent protein (GFP) fusion protein; *Foxp3^{3gfp/wt}*³¹ only 50% of Tregs are green (Fig. 2a). Control animals reconstituted with 1,000 HSC-enriched SP cells reproduced this balanced bimodal distribution of the GFP signal in Tregs (Fig. 2a, Table 1), which became biased when less cells were used in the reconstitution and strictly unimodal (0% GFP or 100% GFP) in monoclonal haematopoietic systems. We confirmed that each individual HSC has one inactivated X-chromosome and that inactivation is a stable epigenetic property, functioning here as an internal control for the mono-clonality of the reconstitutions.

We then analysed *Igh* allelic exclusion in the single HSC-reconstituted animals. We consistently found roughly equal numbers of IgM^a and IgM^b B cells per animal (Fig. 2a and Supplementary Table 2), similar to what is found in polyclonal haematopoietic systems⁶. This observation is in sharp contrast to the X-chromosome inactivation data and suggests the lack of stable allelic commitment in the *Igh*; however, as explained above, the definitive read-out relies on the examination of the rearrangement status of the non-expressed allele. Thus, a quantitative PCR assay was designed to measure the fraction of alleles that retain the V_H-D_H intergenic region, which is excised as an episome during V_H to DJ_H recombination and is diluted out as the pre-BCR-expressing cell divides and progresses in development (Fig. 2b). The assay reports the expected differences in the retention of the V_H-D_H intergenic fragment

in *Rag2^{-/-}* pro-B cells (100%) and splenic B cells from wild-type (WT) animals (roughly 60% of cells in $V_H DJ_H / DJ_H$ configuration and 40% in the $V_H DJ_H / V_H DJ_H$ configuration³², which equals 30% of retention), *Jht^{+/-}* IgM^+ cells (the knockout allele cannot rearrange and therefore all IgM^+ cells are in the $V_H DJ_H$ /germline configuration³³, which equals 50% intergenic fragment retention) and *Atm^{-/-}* (which have an intergenic fragment retention below that of WT animals³). If allelic commitment exists before rearrangement, then the isolated IgM^a and IgM^b subpopulations of a given clone must differ drastically in the amount of the V_H-D_H intergenic fragment retained (50 versus 0%); however, we did not detect any difference in the four animals analysed (Fig. 2c). Similarly, if allelic pre-commitment exists, the cells would carry a productive rearrangement in the pre-committed allele (either IgM^a or IgM^b) together with no V_H rearrangement in the other, while the cells expressing the other allele would have roughly the same number of productive and non-productive rearrangements. Since the $V_H DJ_H^+ / V_H DJ_H^-$ ratio may change depending on the V_H family, we matched the type of families selected for analysis per isolated fraction, focusing on the most prominent families: J558 and 7183. We did not find any striking difference in the frequency of $V_H DJ_H^+$ and $V_H DJ_H^-$ rearrangements of IgM^a and IgM^b -isolated splenic B cells from the monoclonally reconstituted animals (Fig. 2d). Taken together, these data conclusively show that, contrary to X-chromosome inactivation, there is no stable propagation of an allelic commitment to rearrange first the V_H segment in the clones that emerge from a single HSC and differentiate *in vivo* into pro-B cells.

$IgM^{a/b}$ expression becomes skewed in clones derived from CLP. We next asked whether *Igh* allelic commitment for V_H to DJ_H rearrangement would emerge at some stage during the lineage

differentiation to pro-B cells, by applying the previous clonal approach to progressively more differentiated precursor cells. Given the low expansion potential of more differentiated precursors compared with the HSC, we did not reconstitute mice with single B-cell precursors, and used instead a cell culture system. Single cells from pools enriched for HSCs, multipotent progenitors (MPPs), common lymphoid precursors (CLPs) or Pro-B cells were plated under stimulation conditions that are routinely used to mimic the development that occurs in the BM³⁴. As expected, the B-cell-plating efficiency increased from HSCs to CLPs (see Methods; Supplementary Fig. 4), while the

time required to detect surface IgM expression decreased. In Fig. 3a, we plot the percentage of IgM^a-expressing cells (relative to total IgM⁺) per clone growing from seeded HSCs, MPPs and CLPs. As expected, within each group the distribution of IgM^a- and IgM^b-expressing cells is similar. However, the dispersion of each of these percentages is much more pronounced in the clones emerging from CLPs than in the clones resulting from the HSCs and the MMP progenitors (CLP versus MPP, $P = 4.5 \times 10^{-12}$ by F-test; CLP versus HSC, $P = 1.1 \times 10^{-5}$ by F-test). Notably, the collection of the CLP-derived clones includes examples of clones completely skewed to IgM^a or IgM^b (Fig. 3b). Given the

Table 1 | Single HSC-reconstituted animals and representative controls.

Sample (no.)	No. injected cells	Graft age (weeks)	(%) CD45.1 ⁺ spleen	(%) CD45.1 ⁺ LN	(%) CD45.1 ⁺ BM	(%) CD45.1 ⁺ thymus	FOXP3 ⁻ GFP ⁻	FOXP3 ⁻ GFP ⁺	
1.	4.1_2.7M	1	10.8	26.6	14.6	1.6	2.9	0.0	100.0
2.	4.2_2.7M	1	10.8	18.6	32.5	2.4	1.8	0.0	100.0
3.	5.1_3.5M	1	14.0	71.0	27.8	73.6	—	0.6	99.4
4.	5.2_3.5M	1	14.0	24.9	89.9	10.9	46.4	99.6	0.4
5.	8.1_3.7M	1	14.8	59.9	69.5	2.3	0.1	98.3	1.7
6.	4.6_4.0M	1	16.0	41.6	12.4	2.3	3.0	0.0	100.0
7.	4.3_4.5M	1	18.0	15.4	1.6	1.3	0.1	—	—
8.	4.5_4.5M	1	18.0	23.9	72.3	1.6	3.7	100.0	0.0
9.	8.2_5.4M	1	21.6	39.3	3.1	17.8	2.6	0.0	100.0
10.	8.3_5.4M	1	21.6	35.7	—	3.2	6.3	—	—
11.	8.4_5.4M	1	21.6	32.7	55.1	4.1	32.6	0.6	99.4
12.	8.5_5.4M	1	21.6	—	59.8	6.2	18.8	99.6	0.4
13.	8.6_5.4M	1	21.6	19.4	—	7.2	7.6	—	—
14.	8.7_5.4M	1	21.6	37.5	—	6.4	19.2	—	—
15.	5.4_6.9M	1	27.4	76.7	82.1	13.6	8.6	99.8	0.2
16.	4.1_7.9M_2nd_a	1	31.6	1.6	2.1	0.6	30.8	0.0	100.0
17.	4.1_7.9M_2nd_b	1	31.6	8.1	50.7	1.4	1.4	0.6	99.4
18.	5.1_7.4M_2nd_a	1	29.6	37.1	68.1	2.6	6.1	0.8	99.2
19.	c4.7_2.2M	6	8.8	—	73.1	—	12.5	99.5	0.5
20.	c4.4_4.5M	6	18.0	51.5	25.0	13.0	0.2	100.0	0.0
21.	c4.12_7.9M	12	31.6	18.2	22.3	0.3	0.3	—	—
22.	c8.49_3.7M	<50	14.8	78.9	77.6	14.3	72.8	97.4	2.6
22.	c4.6_2.2M	50	8.8	27.0	78.9	—	57.3	6.5	93.5
23.	c8.50_3.7M	50	14.8	25.7	82.9	12.1	0.2	13.4	86.6
24.	c5.1000_2.7M	1,000	10.8	69.7	19.4	13.9	69.6	42.4	57.6
25.	B6.Foxp3 ^{wt/wt}	GFP ^{-/-} Control	—	—	—	—	—	99.9	0.1
26.	B6.Foxp3 ^{gfp/gfp}	GFP ^{+/+} Control	—	—	—	—	—	0.2	99.8
27.	B6.Foxp3 ^{gfp/wt}	Donor Control	—	—	—	—	—	53.5	46.5

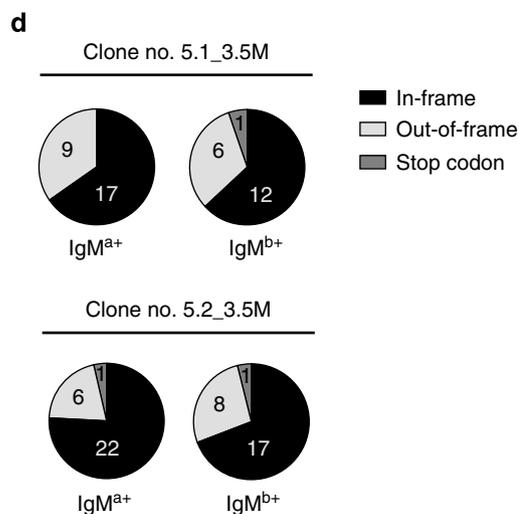
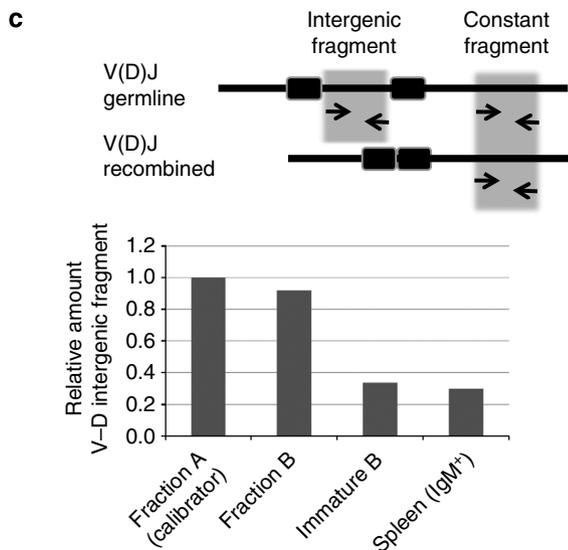
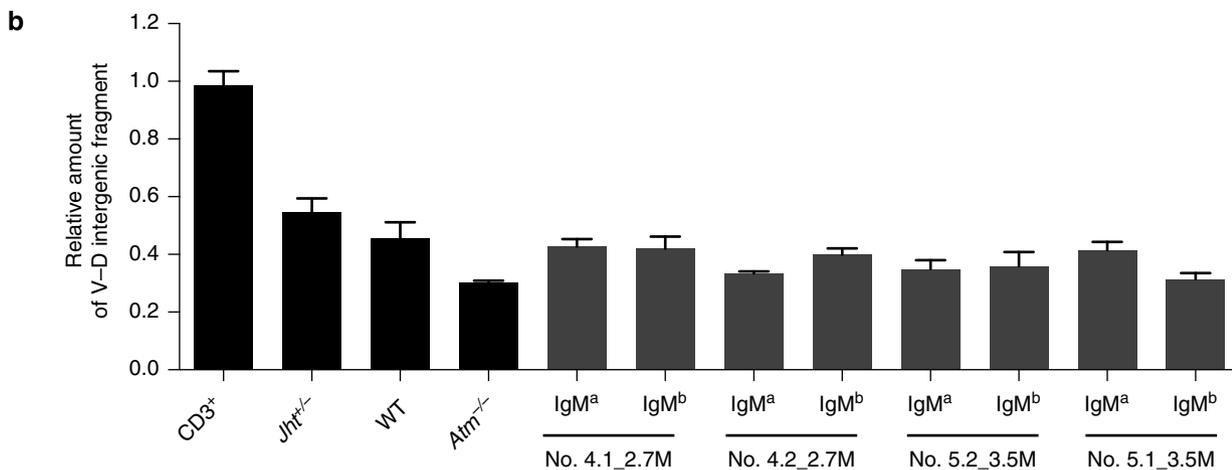
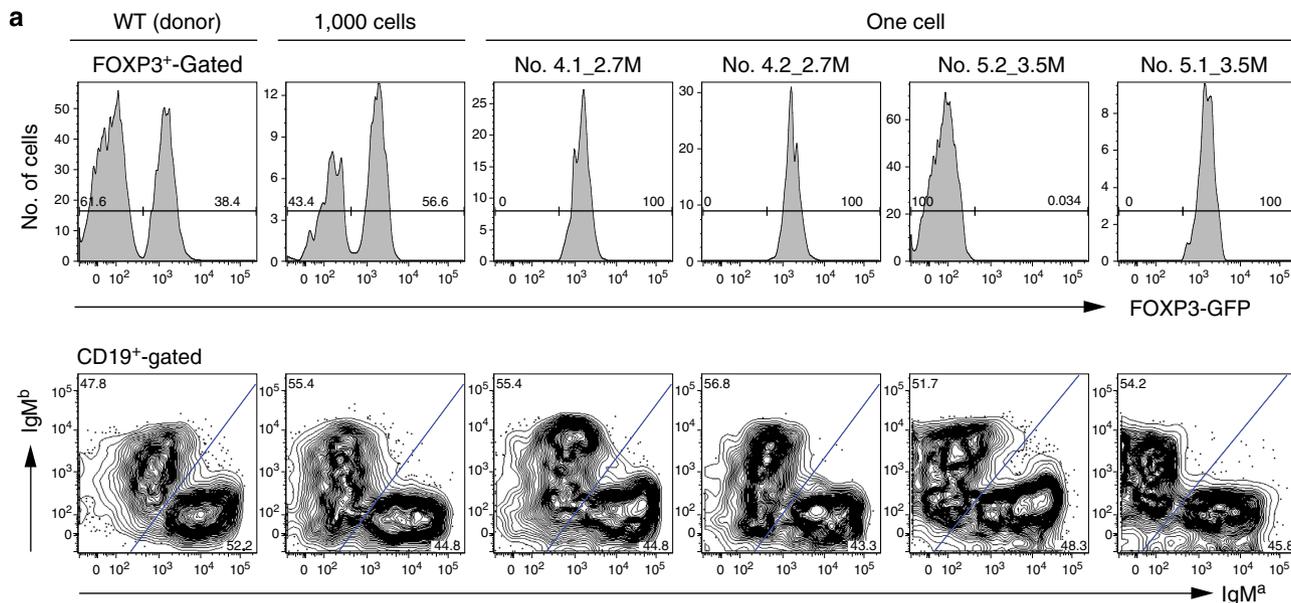
BM, bone marrow; CWBM, clonal whole bone marrow; GFP, green fluorescent protein; HSC, haematopoietic stem cell; LN, lymph node.

Entries 1–15 correspond to single-cell-injected mice, 16–18 are secondary recipients of 5,000 WBM cells of clone no. 4.1_2.7 M (line 1) and no. 5.1_3.5 M (line 3), lines 19–24 correspond to animals injected with more than one cell, and animals 25–27 are non-reconstituted controls. Graft age is the time in weeks at which the sample was collected after the injection of the single cell. Frequency of CD45.1⁺ cells in the spleen, LN, BM and thymus is presented if available. The last two columns (Foxp3 – GFP^{+/–}) show the frequency of LN CD45.1⁺ Foxp3⁺ cells that were staining positive or negative for Foxp3 – GFP⁺ by flow cytometry (representative histograms are shown in Fig. 2a). The animal no. 8.1_3.7 M was excluded from further analysis because a small population of GFP⁺ cells were observed in addition to GFP[–] Foxp3⁺ cells. On the basis of these data, we estimate that the percentage of false monoclonal reconstitutions within the selected pool of animals with unimodal distributions of GFP in the regulatory T-cell population is below 5%.

Figure 2 | Direct test for a stable *Igh* epigenetic commitment in HSC that dictates which allele undergoes V_H to DJ_H rearrangement first in the B lineage. (a) The expression of FOXP3GFP in lymph node (LN)-derived regulatory T cells is bimodal in the non-clonal populations, and unimodal in mice reconstituted with a single HSC (top histograms), but the expression of the *Igh* alleles is balanced in the splenic B cells of both the non-clonal and single-cell-reconstituted mice (bottom contour plots). Top histograms represent GFP expression profiles of CD45.1⁺ CD4⁺ FOXP3⁺-gated LN cells, and bottom contour plots show the CD45.1⁺ CD19⁺-gated splenic B-cell surface IgM^a/IgM^b profiles of the same clonal and non-clonal samples. (b) Genomic quantification with qPCR of *in vivo*-differentiated B-cell clones. The retention of the V_H - D_H intergenic fragment in IgM^a and IgM^b-sorted splenic populations from each of the single-HSC clones—the same samples shown in a—is similar to each other, and in the same range of IgM⁺-sorted cells from non-clonal populations. The qPCR on DNA from IgM⁺ cells purified from *Jht*^{+/-} mice (that can only rearrange one of the *Igh* alleles) and *Atm*^{-/-} mice (which have a higher frequency of V_HDJ_H/V_HDJ_H cells than the WT^{3,22}) produced the expected results. Error bars are the s.e.m. of three independent runs of the same samples in different input amounts. (c) The reduction in the amount of the V_H - D_H intergenic fragment from the fractions A and B of pro-B cells to immature B cells shows that the assay is reporting the changes in the structure of the *Igh* locus throughout development. (d) Comparison of the frequency of productive and non-productive V_HDJ_H rearrangements detected in the splenic IgM^a and IgM^b-sorted pools from clone no. 5.1_3.5 M ($P = 1.00$, Fisher's exact test) and no.5.2_3.5 M ($P = 0.55$, Fisher's exact test).

heterogeneous nature of the CLP populations, we decided to further dissect them to identify the earliest population containing progenitors with potential for the highly biased expression of

IgM^a or IgM^b. The surface marker Ly6d (Fig. 3c) identifies one population (Ly6d⁻) that conserves the full lymphoid potential and one population (Ly6d⁺) that is to date the earliest B-cell



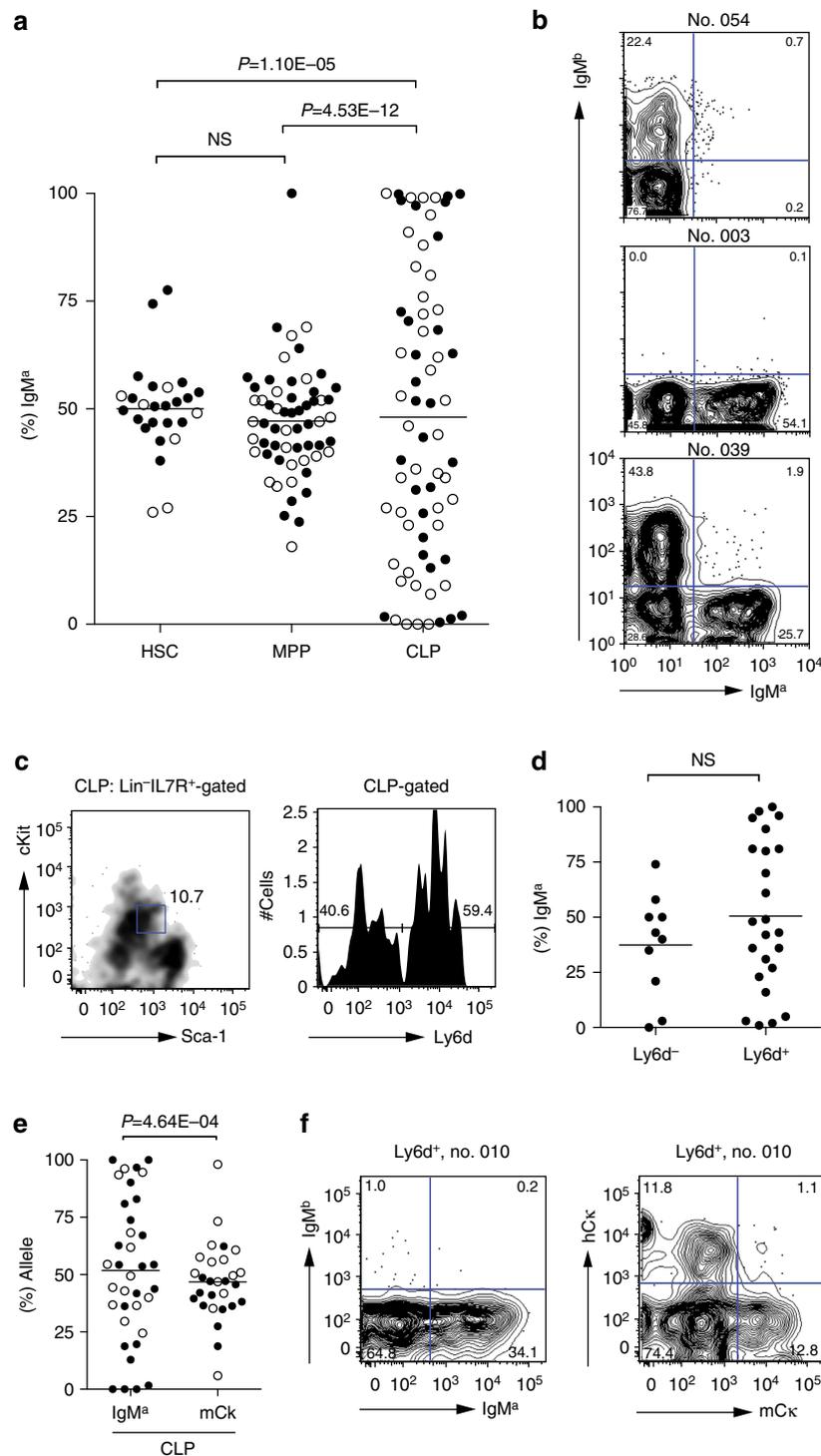


Figure 3 | IgM^{a/b} expression in CLP-derived clones. As the developmental stage of the isolated precursor cell approaches the stage when recombination typically occurs, the distribution of IgM^a/IgM^b-expressing cells in the resulting B-cell clones differentiated *in vitro* becomes biased. The graphs show *P* values from an F-test. **(a)** IgM^a percentage (IgM^a/[IgM^a + IgM^b]) obtained from flow cytometric analysis of B-cell clones differentiated in culture from HSCs, MPPs and CLPs; each dot represents a clone, open and closed dots distinguish two independent assays. **(b)** Representative examples of flow cytometric profiles of CLP-derived clones presenting extreme bias in the expression of IgM^a or IgM^b (clone no. 054 and no. 003) and a mild biased/balanced (clone no. 039). **(c)** Left: density plot shows the Lin⁻IL7R⁺-gated BM population, staining cKit^{int} and Sca-1^{int}, thus identified as the CLP population. Right: histogram shows the Ly6d-staining profile in the CLP-gated population. **(d)** Percentages of IgM^a-expressing cells in clones derived from CLPs partitioned according to the presence of the Ly6d marker³⁵. The percentage of IgM-expressing cell clones emerging per CLP.Ly6d⁻ plated cell was 2%, while Ly6d⁺ cells render IgM⁺ B cell clones at least 10 × more frequently. **(e)** Distribution of the percentages of heavy chain alleles (IgM^a/[IgM^a + IgM^b]) and light chain alleles (hCκ/[hCκ + mCκ]) in CLP Ly6d⁺-derived clones from the same experiment. **(f)** CLP Ly6d⁺ clone no. 010 flow cytometric profile for the immunoglobulin heavy chain alleles (IgM^a/IgM^b) and light chain alleles (hCκ/mCκ).

progenitor³⁵. We could confirm that, under conditions promoting B-cell differentiation, the B-cell-plating efficiency of the Lyd6⁻ CLP-derived clones was at least 10 × lower than in Lyd6⁺ CLP-derived clones (in one experiment, 15 wells presented colonies out of 960 single-cell Lyd6⁻ CLP-plated cells, while 68 wells presented colonies out of 384 single-cell-plated Lyd6⁺ CLP-plated cells). We observed that the highly skewed clones are more frequent in the Lyd6⁺ pool (Fig. 3d), although the difference is not statistically significant. We have further differentiated B-cell clones from whole B220⁺CD43⁺ pro-B cells and from the fraction B and found that, although the lower expansion potential of pro-B cells originated smaller clones, CLP- and B220⁺CD43⁺-derived clones have a similar frequency of skewed clones (Supplementary Fig. 4c). Fraction B presented an even lower expansion potential, which precluded further molecular analysis. We have also compared *Igh* and *Igk* loci with respect to the allelic bias observed in *in vitro* CLP-derived clones from mice heterozygous for both *Igh* and light chain immunoglobulin loci (B6.Cg-*Igh*^{a/b}, *hCk/mCk*). The *Igh* locus typically rearranges before the kappa light chain and usually four to six cell divisions occur between the rearrangement of these immunoglobulin genes^{36,37}. As observed in Fig. 3e,f, the bias observed for the *Igh* is not reproduced for the *Igk* ($P = 0.00046$ by F-test) presumably because the number of cells rearranging the *Igk* alleles independently is much higher than the number of cells rearranging the *Igh* alleles within the clone. The balanced ratio of the *Igk*-expressing cells also suggests that the *Igk* locus is not pre-committed in the CLP stage of B-cell development. Taken together, these data show that both the probability of finding clones with allelic bias and the degree of bias increase as the stage at which the cells are seeded gets closer to the time of V(D)J rearrangement.

Independent recruitment of *Igh* alleles in CLP-derived clones.

Different scenarios could explain the distribution of the IgM^a and IgM^b percentages in CLP-derived clones (Fig. 3a). In the first scenario, the clones would be synchronized for V(D)J rearrangement and in each clone there would be a wave of IgM^a (or IgM^b)-expressing cells followed by a partially overlapping second wave of IgM^b (or IgM^a) cells. In this scenario, consistent with an asymmetric allele choice for rearrangement within each clone, the time point at which the clones were analysed would only be a snapshot, failing to capture the ongoing dynamics in the cultures. We have sampled the cultures at different time points and found no clear evidence for the existence of such waves in the majority of clones (Supplementary Fig. 5). The remaining scenarios can be distinguished by the status of rearrangement of the silent allele in the IgM^a and IgM^b populations of each clone. This status was estimated with the real-time quantitative PCR (qPCR) of the *V_H-D_H* intergenic fragment and also by sequencing *V_HDJ_H* and *DJ_H* rearrangements from the sorted populations, taking into account polymorphisms downstream of *J_H1*, *J_H2* and *J_H3* segments that enable us to assign rearrangements to the BALB/c (*Igh*^a) or C57BL/6 (*Igh*^b) allotypes. We first examined the CLP-derived clones with both the IgM^a and IgM^b pools clearly discernible and balanced with the help of flow cytometry (Fig. 4a,b). In the genome of these IgM⁺ populations, the percentage of *V_H-D_H* intergenic fragment retention when compared with the Rag2^{-/-} genome was typically between 40% (similar to the mean retention observed in non-clonal *in vitro*-derived IgM⁺ cells, Fig. 4a) and 50% (the percentage in the genome of non-clonal *in vitro*-derived IgM⁺ cells of *Jht*^{+/-} mice). More importantly, within the same clone, the IgM^a and IgM^b genomes retain approximately the same percentage of the *V_H-D_H* intergenic fragment, which means that each allele can undergo *V_H* to *DJ_H* rearrangement first (Figs 1 and 4a).

Furthermore, we sequenced *DJ_H* and *V_HDJ_H* rearrangements and found diverse *V_H-DJ_H* junctions as well as diverse *D_H-J_H* junctions, and a limited number of these *D_H-J_H* junctions contributing to the diverse *V_HDJ_H* rearrangements per clone. This shows that the Lyd6⁺ single-cell-sorted cells are devoid of *D_H-J_H* rearrangements before culture, and suggests that in some clones we are testing subclonal expansions and tracking the choice to rearrange shortly before the decision to undergo *V_H* to *DJ_H* rearrangement. Representatively, in clone Xp40, the same two *DJ_H* junctions (one originating from the *Igh*^b allele and the other from the *Igh*^a allele) prevail in the genome of the IgM^a, IgM^b and IgM⁻ pools (Table 2). The *Igh*^b *DJ_H* junction is shown to be predominantly contributing to the diverse *V_HDJ_H* rearrangements of the IgM^b pool and in IgM^a pool the same *DJ_H* junction is part of non-productive (*V_HDJ_H*⁻) and the same junction was detected in the germline-*DJ_H* configuration, while the dominant *Igh*^a *DJ_H* junction was detected as part of productive (*V_HDJ_H*⁺) rearrangements in the IgM^a cell pool and in non-productive and *DJ_H* rearrangements in the IgM^b pool. Different *DJ_H* junctions are detected in all three fractions of the clone but were not detected as part of *V_HDJ_H* rearrangements. From the analysis of rearrangements we conclude that the culture system mimics the features of V(D)J recombination *in vivo* and that considerable intraclonal diversity is generated in our clones, extending previous observations made in human B-cell lines³⁸. More importantly, we observed that any of the alleles can undergo *V_H* to *DJ_H* rearrangement first within each clone.

Founder effects in highly skewed clones. The most striking behaviour of the CLP-derived clones is the extreme bias displayed by some clones (Figs 3a,b and 5). This pattern could emerge from a pre-determined choice to rearrange a given allele first, if the culture conditions do not favour the second attempt to rearrange *V_H* to *DJ_H*. The comparison between the amount of *V_H-D_H* intergenic fragment in IgM⁺ non-clonal cell populations differentiated *in vivo* and *in vitro* shows that in culture conditions there is indeed a minor increase in the retention of the *V_H-D_H* intergenic fragment (41.6 ± 5.8% versus 34.1 ± 4.9%), but a sizable proportion of cells have two rearrangements (Supplementary Fig. 5). Thus, it is difficult to explain the extreme skewing observed in these clones solely by attributing a fixed choice in the allele to undergo *V_H* to *DJ_H* rearrangement. Next, the status of rearrangement of the expressed and silent alleles in the clones was investigated. In the 18 IgM⁺ populations of the CLP-derived clones with both IgM^a and IgM^b clearly distinguishable (Fig. 4a), no example was found of a population with a *V_H-D_H* intergenic fragment below 30%. However, in the 10 IgM⁺ populations from CLP-derived clones with extreme biases, three populations have a percentage of *V_H-D_H* intergenic fragment below 30% ($P = 0.037$, Fisher's exact test). Interestingly, two of those clones were essentially depleted of the *V_H-D_H* intergenic fragment (Fig. 5a), showing that these populations are mostly composed of cells that also have the silenced allele in the *V_HDJ_H* configuration. This observation suggests that the extreme bias in this subset of clones results from a rearrangement occurring early on in the culture, that supposedly is non-productive and permanently silences the allele. In the case of the populations without a *V_H-D_H* intergenic fragment, such rearrangement could be a *V_HDJ_H*⁻ or a *V_HDJ_H*⁺ encoding a μ chain that cannot form the pre-BCR¹⁴; in the case of the populations with a sizable *V_H-D_H* intergenic fragment, such rearrangements could be *DJ_H* junctions using the reading frame 3 that frequently contains premature stop codons³⁹. For the CLP-derived B-cell clone Xp29, which does not retain the *V_H-D_H* intergenic fragment (Fig. 5a), sequencing data support the hypothesis of a non-productive rearrangement-imposed founder

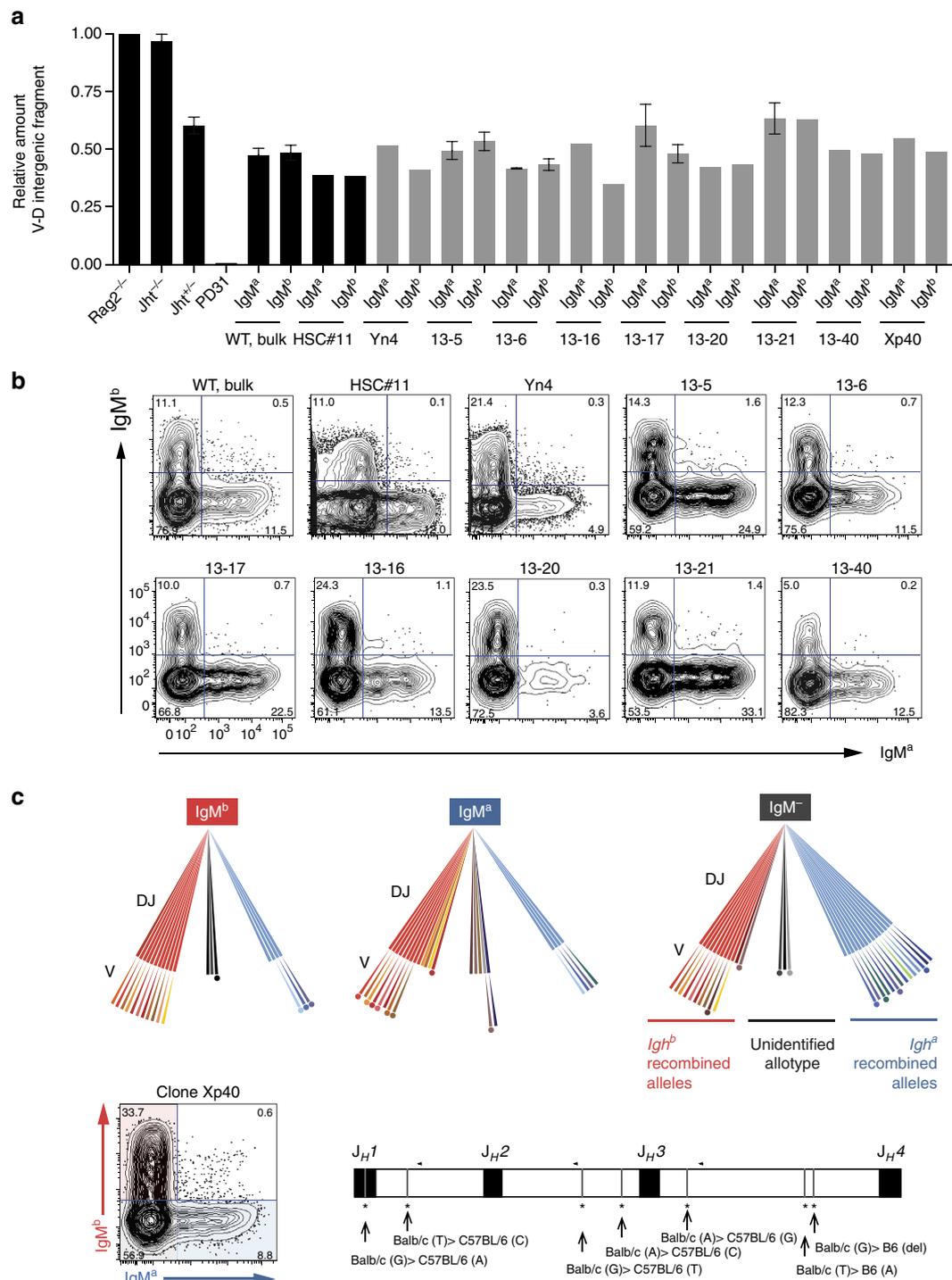


Figure 4 | Rearrangement status of the silent *Igh* allele in CLP-derived clones. The status of rearrangement of the silent allele in IgM^a and IgM^b -sorted cells from *in vitro* CLP-derived B-cell clones shows independent allele recruitment for rearrangement. **(a)** V_H - D_H intergenic fragment retention quantification data are shown in two blocks: in black, a battery of control profiles: *Rag2*^{-/-} and *Jht*^{+/-} CD19⁺ cells, that do not rearrange the *Igh* locus, *Jht*^{+/-} as in Fig. 2c, the PD31 cell line, which lacks the V_H - D_H intergenic fragment⁵⁵, and additional WT controls, all differentiated *in vitro*. In grey, data from pairs of IgM^a - and IgM^b -sorted samples from mildly biased/balanced clones. Sample Yn4 is a CLP-Ly6d⁻-derived clone, and the remaining samples are CLP-Ly6d⁺-derived. Error bars denote s.e.m. **(b)** Flow cytometric profiles of the WT bulk control and clones represented in the graph. **(c)** Top panel: schematic representation of the diversity of rearrangements obtained by sequencing V_HDJ_H and DJ_H rearrangements of IgM^b , IgM^a and IgM^- populations of the clone Xp40 (see Table 2 for a detailed description). Legend: long brush strokes represent D-J joints and short ones the V_H - D_H joints. Sequences of Igh^a , Igh^b and unknown allotype are represented by cold, warm and greyish colours, respectively. A long stroke not followed by a short stroke represents a DJ_H rearrangement. A dot represents a non-productive rearrangement. Different colours in longer strokes represent diverse D-to-J joints as different colours in shorter strokes represent diverse V_H to DJ_H joints. The same colours represent the same DJ_H joint and identical VDJ_H joints are not represented. Left lower panel: clone Xp40 flow cytometric profile as in **b**. Right lower panel: map showing polymorphisms in the J_H cluster of Igh^a (Balb/c) and Igh^b (C57BL/6) allotypes, the J_H segment position and relative positions of 3'- J_H1 , - J_H2 and - J_H3 primers (arrowheads, J_H4 primer position not shown in map).

Table 2 | Sequencing V_HDJ_H and DJ_H junctions of clone Xp40.

ID*	V_HDJ_H (CDR3)/ DJ_H junction (5'-3') [†]	V_H	D_H	RF [‡]	J_H	A [§]	F	R
Xp40, IgM^{b+}								
V_HDJ_H								
2.1.	tgt gca aga <u>aga gat gaa aac</u> tac ttt gac tac tgg	J558.26.116	SP2.9	1	2	b	p	
2.2.	tgt gca aga <u>ggg agg gaa aac</u> tac ttt gac tac tgg	V23	SP2.9	1	2	b	p	
2.3.	tgt gcc aga <u>agc ctc tat gaa aac</u> tac ttt gac tac tgg	Q52.2.4	SP2.9	1	2	b	p	
2.4.	tgt gcc aga <u>aat tgg gaa aac</u> tac ttt gac tac tgg	Q52.9.29	SP2.9	1	2	b	p	
2.5.	tgt gca aga <u>agg ggc tat gaa aac</u> tac ttt gac tac tgg	36-60.8.74	SP2.9	1	2	b	p	
2.6.	tgt gca aga <u>tft gaa aac</u> tac ttt gac tac tgg	36-60.8.74	SP2.9	1	2	b	p	
2.7.	tgt aca aga <u>ggc tat gaa aac</u> tac ttt gac tac tgg	7183.14.25	SP2.9	1	2	b	p	
2.8.	tgt gca aga <u>ggt cgg gaa aac</u> tac ttt gac tac tgg	7183.21b	SP2.9	1	2	b	p	
2.9.	tgc aca ggg <u>cta gat gaa aac</u> tac ttt gac tac tgg	J606.1.79	SP2.9	1	2	b	p	
2.10.	tgc aca ggc <u>ctg gat gaa aac</u> tac ttt gac tac tgg	J606.1.79	SP2.9	1	2	b	p	
11.1.	tgt gca aga <u>gaa gac ccc ctc ccg</u> ttt gct tac tgg	J558.84.190	?	—	3	b	p	2
1.1.	tgt gcc aga aat act acg <u>g.t ccc</u> gcc tgg ttt gct tac tgg	Q52.a2.4	FL16.2	2	3	a	np	2
1.2.	tga gca <u>cat tac tac ggt ccc</u> gcc tgg ttt gct tac tgg	7183.a2.3	FL16.2	1	3	a	np	2
1.3.	tgt acc agg <u>gca cta cgg. t ccc</u> gcc tgg ttt gct tac tgg	J606.a6.127	FL16.2	3	3	a	np	2
DJ_H								
1.	tt cat tac tac <u>ggt ccc</u> gcc tgg ttt gct tac tgg	—	FL16.2	1	3	a	—	8
3.	tt tat tac tac <u>ggt agt agc</u> tat gct atg gac tac tgg	—	FL16.1	1	4	?	—	
4.	tt tat tac tac <u>ggt agt agc</u> tat tac tat gct atg gac tac tgg	—	FL16.1	1	4	?	—	
5.	t tta tta cta <u>cgg tag tag</u> cat tac tat gct atg gac tac tgg	—	FL16.1	3	4	?	stp	
Xp40, IgM^{a+}								
V_HDJ_H								
2.11.	tgt gat cta <u>tga aa. ac</u> tac ttt gac tac tgg	J558.21pg.111	SP2.9	3	2	b	np	
2.12.	tgt gca tga <u>ggc ggg gaa aac</u> tac ttt gac tac tgg	J558.1.85	SP2.9	1	2	b	np	
2.13.	tgt gca aga <u>agc tga aa. ac</u> tac ttt gac tac tgg	J558.26.116	SP2.9	3	2	b	np	
2.14.	tgt gca aga <u>tgg cta tga aa. ac</u> tac ttt gac tac tgg	J558.72.173	SP2.9	3	2	b	np	
2.15.	tgt gca aga <u>tat gaa aac</u> tac ttt gac tac tgg	X24.1pg.45	SP2.9	1	2	b	p	2
2.16.	tgt gca aga <u>aag gga aac</u> tac ttt gac tac tgg	X24.2.50	?	—	2	b	p	2
2.17.	tgt gca aga <u>ctatga aa. ac</u> tac ttt gac tac tgg	7183.a2.3	SP2.9	3	2	b	np	
2.18.	tgt gca aga <u>gat gtg aaa. ac</u> tac ttt gac tac tgg	7183.21b	SP2.9	2	2	b	np	
1.4.	tgt gca agg <u>aac tac ggt ccc</u> gcc tgg ttt gct tac tgg	J558.47	FL16.2	1	3	a	p	
1.5.	tgt gca aga <u>tgc gga ggt ccc</u> gcc tgg ttt gct tac tgg	124	FL16.2	1	3	a	p	
1.6.	tgt gca aga <u>tcc ctc tac ggt ccc</u> gcc tgg ttt gct tac tgg	J558.7	FL16.2	1	3	a	p	
1.7.	tgt gca aga <u>gaa ggg gac tac ggt ccc</u> gcc tgg ttt gct tac tgg	S107.a3.106	FL16.2	1	3	a	p	2
12.1	tgt gca aga <u>tcc tac</u> tat gat ggt tac <u>cac gct</u> tac tat gct atg gac tac tgg	J558.84.190	SP2.9	1	4	?	p	4
13.1	tgt gca aga <u>cat agt aac. t</u> gct atg gac tac tgg	7183.2.3	SP2.x	1	4	?	np	2
DJ_H								
2.	tc tat <u>gaa aac</u> tac ttt gac tac tgg	—	SP2.9	1	2	b		
14.	tt tat tac tac <u>ggt agt agc cgc</u> ttt gct tac tgg	—	FL16.1	1	3	b		
15.	tt tat tac tac <u>ggt agt agc</u> ttt gct tac tgg	—	FL16.1	1	3	b		
16.	ttt att act acg gta gta <u>gct</u> gcc tgg ttt gct tac tgg	—	FL16.1	2	3	b		
17.	c cta <u>ctafag tag ggg</u> gcc tgg ttt gct tac tgg	—	SP2.x	3	3	b	stp	
18.	t tta tta cta <u>cgg</u> tac tat gct atg gac tac tgg	—	FL16.1	3	4	?		
19.	ttt att act acg gta gta <u>gca</u> tat tac tat gct atg gac tac tgg	—	FL16.1	2	4	?		
20.	tt tat tac tac <u>ggt agt agc</u> tac tat gct atg gac tac tgg	—	FL16.1	1	4	?		
Xp40, IgM⁻								
V_HDJ_H								
1.4.	tgt gca agg <u>aac tac ggt ccc</u> gcc tgg ttt gct tac tgg	J558.17	FL16.2	1	3	a	p	2
1.8.	tgt aca aga <u>gcg att</u> cat tac tac <u>ggt ccc</u> gcc tgg ttt gct tac tgg	J558.40	FL16.2	1	3	a	p	
1.9.	tgt gcc aga <u>gag cat tac tac ggt ccc</u> gcc tgg ttt gct tac tgg	Q52.a27.79	FL16.2	1	3	a	p	3
1.10.	tgt gcc agt <u>cac tac ggt ccc</u> gcc tgg ttt gct tac tgg	Q52.a2.4	FL16.2	1	3	a	p	
1.11.	tgt gca agg <u>ttc att act acg gtc cc.</u> gcc tgg ttt gct tac tgg	36-60.a6.114	FL16.2	2	3	a	np	
1.12.	tgt gca aga <u>tcc cac tac gg[c]</u> gcc tgg ttt gct tac tgg	36-60.a9.121	FL16.2	1	3	a	p	
1.13.	tgt gca <u>atc cat tac tac ggt ccc</u> gcc tgg ttt gct tac tgg	36-60.a9.121	FL16.2	1	3	a	p	2
1.14.	tgt gtc aga <u>cga ggg tac tac ggt ccc</u> gcc ggg ttt gct tac tgg	7183.a19.31	FL16.2	1	3	a	p	2
1.15.	tgt gca aga <u>tta cta cgg tcc c.</u> gcc tgg ttt gct tac tgg	81X	FL16.2	3	3	a	np	
1.16.	tgt acg agg <u>tac ggt ccc</u> gcc tgg ttt gct tac tgg	7183.a33.55	FL16.2	1	3	a	p	
1.17.	tgt gca <u>aaa tca att</u> tac tac <u>ggt ccc</u> gcc tgg ttt gct tac tgg	X24.a1.84	FL16.2	1	3	a	p	3
1.18.	tgt gca aga <u>cgg aat tac tac ggt ccc</u> gcc tgg ttt gct tac tgg	X24.a1.84	FL16.2	1	3	a	p	
1.19.	tgt atg aga <u>tta cta cgg tcc c.</u> gcc tgg ttt gct tac tgg	11.a2.92	FL16.2	3	3	a	np	3
1.20.	tgt atg aga <u>tgg gga cta cgg tcc c.</u> gcc tgg ttt gct tac tgg	11.a2.92	FL16.2	3	3	a	np	
2.19.	tgt gca <u>att tat gaa aac</u> tac ttt gac tac tgg	J558.26.116	SP2.9	1	2	b	p	
2.20.	tgt aca aga <u>tcc tat gaa aac</u> tac ttt gac tac tgg	J558.3.90	SP2.9	1	2	b	p	
2.21.	tgt gca <u>tta tat gaa aac</u> tac ttt gac tac tgg	J558.78.182	SP2.9	1	2	b	p	
2.22.	tgt gcc aga <u>atg aaa. ac</u> tac ttt gac tac tgg	Q52.2.4	SP2.9	2	2	b	np	4

Table 2 (Continued)

ID*	$V_H DJ_H$ (CDR3)/ DJ_H junction (5'-3') [†]	V_H	D_H	RF [‡]	J_H	A [§]	F	R
2.23.	tgt gca aga tat ggg atc tac gaa aac tac ttt gac tac tgg	36-60.8.74	SP2.9	1	2	b	p	
2.24.	tgt gca aga tcc tcc ttc tat gaa aac tac ttt gac tac tgg	36-60.8.74	SP2.9	1	2	b	p	
2.25.	tgt gca agg cac tat gaa aac tac ttt gac tac tgg	7183.7.10	SP2.9	1	2	b	p	
2.26.	tgt acc agg gaa aac tac ttt gac tac tgg	J606.4.82	SP2.9	1	2	b	p	
2.27.	tgc aca ggc ccc ctc tat gaa aac tac ttt gac tac tgg	J606.1.79	SP2.9	1	2	b	p	
2.28.	tgt acc agc tat gaa aac tac ttt gac tac tgg	J606.4.82	SP2.9	1	2	b	p	2
DJ_H								
1.	tt cat tac tac ggt ccc gcc tgg ttt gct tac tgg	—	FL16.2	1	3	a		6
6.	tt tat tac tac ggt agt agt ccc tgg ttt gct tac tgg	—	FL16.1	1	3	b		4
7.	t tta tta cta cga tag agg ttt gct tac tgg	—	FL16.1	3	3	b	stp	
8.	ttt att act acg gta gta gta tac tat gcc atg gac tac tgg	—	FL16.1	2	4	?		2
9.	t tta tta cta cggtag tag cta tat tac tat gct atg gac tac tgg	—	FL16.1	3	4	?	stp	
10.	t tta tta cta cgg tag tag ctt gct atg gac tac tgg	—	FL16.1	3	4	?	stp	

A, allotype; F, functionality prediction; p, productive; np, non-productive; stp, stop codon in DJ_H conformation.

*Each line preceded by a number shows the CDR3 sequences or the $D-J_H$ junctions of B-cell clone-sorted populations (IgM^{a+} or IgM^{b+} genomic fractions) sequenced from amplicons produced using a battery of 5' primers specific for the mouse V_H families or D_H gene segments and 3' J_H -specific primers (see Methods). Numbers are used as sequence identification tags. The first number before the first dot identifies a unique DJ_H junction, and the second number (between dots) identifies a unique $V_H DJ_H$. In the table, a total of 20 different DJ_H junctions are described, but only five were also detected within $V_H DJ_H$ rearrangements ('1', '2', '11', '12' and '13'). Of these DJ_H junctions, two are predominant among all fractions of the clone: an Igh^b -derived DJ_H2 joint and an Igh^b -derived DJ_H3 junction. These recombined independently with multiple V_H -gene segments distributed along the locus: DJ_H sequence number '1' was embedded in $N=20$ $V_H DJ_H$ sequences (1.1-1.20.), and DJ_H sequence number '2', that was embedded in $N=28$ $V_H DJ_H$ sequences (2.1-2.28.). These two particular DJ_H sequences were detected in all three populations of the clone, with the $V_H DJ_H$ configuration presumably corresponding to the expressed allele, and $V_H DJ_H$ and DJ_H corresponding to the non-expressed one. The existence of many independent $V_H DJ_H$ sequences of the non-expressed allele in both IgM^+ fractions is indicative of a recruitment of V_H -gene segments for V_H to DJ_H rearrangement on each allele that is independent of the other allele. The diversity of $V_H DJ_H$ rearrangements sharing the same pair of DJ_H joints, and a number of independent DJ_H joints not detected in $V_H DJ_H$ rearrangements in the IgM^a , IgM^b and IgM^- populations, further suggests a subclonal expansion within the clone.

†The analysis was initially performed by running the IMG2/HIGHV-QUEST⁵⁶, and subsequently the sequences were manually curated. The CDR3 was defined as the starting cysteine at position 104 and ending at the JH 5' invariant tryptophan. Underlined sequences are the nucleotides attributed to D_H sequences; N and P additions are in bold; frame shifts are represented by dots; stop codons are in italics; parenthesis are used to highlight probable PCR errors. This notation applies to all sequencing data tables.

‡F, DH-reading frame.

§a', Igh^a -coding for IgM^a and 'b' designates Igh^b -coding for IgM^b .

||R, the number of times the same sequence was detected (by Sanger sequencing) each time a band was cloned; typically, four colonies were screened per cloned band, and in some relevant cases where no diversity was found, more colonies were sequenced from the same cloned bands (usually in batches of four).

effect: the $V_H DJ_H$ rearrangements detected are diverse for the productive V_H to DJ_H junctions, but limited concerning both the D_H to J_H junctions ($N=2$ detected), and the V_H to DJ_H non-productive junction (unique sequence detected, several times; Fig. 5c and Supplementary Table 3). Thus, this clone appears to have made a first or very few non-productive V_H to DJ_H rearrangements early on in the culture, by chance in the same allele, which was followed by a number of cell divisions and then an extensive rearrangement of the second allele. Clones 13-33, Xp39, no. 5.3 and no. 4.2, which also show absolute biases (Fig. 5b) retain an intermediate amount of V_H-D_H intergenic fragment (Fig. 5a). Sequencing data for clone no. 5.3 (showing a stop codon in the DJ_H joint of one of the alleles) and no. 4.2 (with a DJ_H joint that lacks a highly conserved codon) also confirm the founder effects imposed by rearrangement (Fig. 5c and Supplementary Tables 4 and 5). In clones 13-2, 13-18 Yp62 and Yp4, the silent allele is mostly in the DJ_H configuration (Fig. 5a), which suggests that one of the alleles may have a lower probability to rearrange. Sequencing analysis of the IgM^b -skewed clone 13-18 indeed shows that the IgM^+ population is composed of cells that have rearranged V_H to DJ_H only in the Igh^b allele, leaving Igh^a in the DJ_H configuration (Fig. 5c and Supplementary Table 6). Interestingly, the IgM^- population sequencing analysis shows $V_H DJ_H$ rearrangements in both alleles, with Igh^a rearrangements that are more frequently non-productive (seven non-productive with frequent stop codons out of eight sequences) than the Igh^b ones (one non-productive out of 10 sequences). These results do not exclude that, for a few subclonal expansions, one of the alleles acquires an increased probability of rearrangement. Overall, we conclude that a minority of the CLP-derived clones with extreme skewing may have a relative bias to rearrange predominantly one of the alleles and that the majority of CLP-derived clones with extreme biases is explained by founder effects imposed by rearrangement that introduce a genetic constraint in the ability to rearrange a given allele in the precursor cells from which the IgM^+ cells emerge.

Discussion

Here we provide the first conclusive evidence that the allelic exclusion of immunoglobulin receptor loci differs from X-chromosome inactivation, as no stable epigenetic mark is propagated until pro-B cells start rearranging. More importantly, we also show that the Igh alleles rearrange independently from each other in the majority of clones derived from common lymphoid precursors.

Roughly, the number of cells expressing the maternal or the paternal Igh allele is the same. Similarly, in female mammals each cell inactivates the paternal or the maternal X-chromosome with identical probability. This common feature is frequently emphasized, and it has been hypothesized that autosomal genes under monoallelic expression of one or the other allele are flanked by regions with a high density of long interspersed nuclear elements⁴⁰, the same elements that play a key role in the generation and propagation of the silencing epigenetic marks along the X-chromosome⁴¹. Furthermore, it is known that the alleles replicate asynchronously in parent-of-origin imprinting⁴², other cases of monoallelic expression^{28,43} and X-linked inactivation⁴⁴. In particular, antigen receptor loci were found to establish this pattern early in development, around the time X-inactivation is also established, and then proceed to maintain it²⁸. This led to the proposal that the allele to replicate first retains its epigenetic mark as the cell divides and undergoes differentiation, so that in clonally related developing B cells, it is always the same allele that undergoes V_H to DJ_H rearrangement first²⁸. The analysis of haematopoietic systems resulting from a single female HSC is a direct test of this hypothesis because the HSC population is heterogeneous for the inactivated X-chromosome, and if the Igh locus carries an epigenetic mark that is transmitted to the progeny as stably as the X-chromosome marks, V_H to DJ_H rearrangement would occur first in the same allele in all, or the vast majority, of the pro-B cells. As expected⁴⁵, X-chromosome inactivation is stably transmitted during all cell divisions that occur as a HSC differentiates into a regulatory T

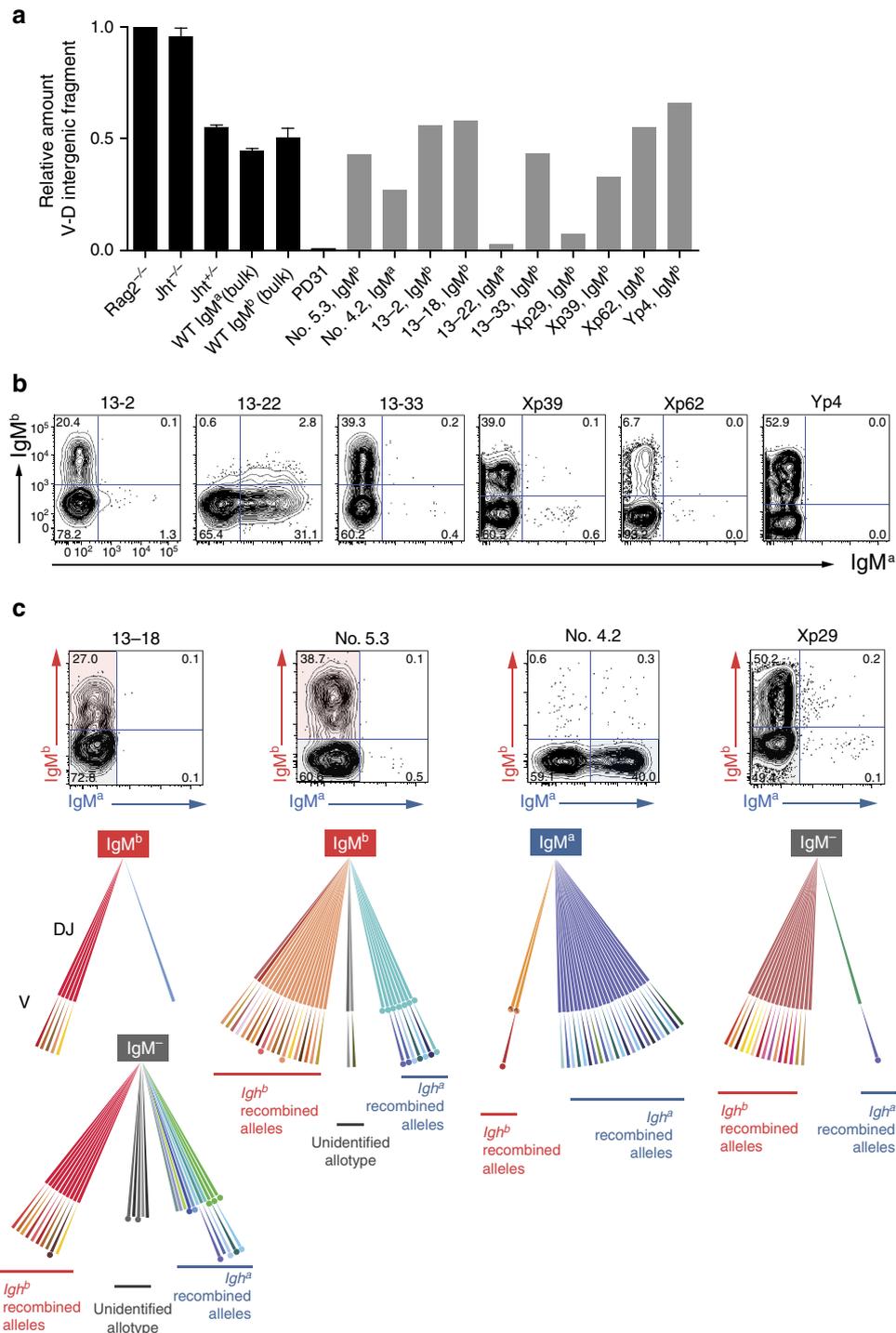


Figure 5 | Rearrangement status of the silent *Igh* allele in highly skewed CLP-derived clones. (a) Quantitative PCR of the *V_H-D* intergenic DNA fragment in IgM⁺ fractions of the *in vitro*-derived B-cell clones expressing only one allele (grey bars) and *in vitro*-derived controls. Error bars denote s.e.m. (b) CD19⁺-gated clones' dot plots showing IgM^a/IgM^b flow cytometric profiles for some of the clones in a. (c) Top panels' dot plots show the remaining clones in a, not shown in b; shaded quadrants in the plots show the fractions that are represented in the tree-like scheme of the lower panels. Each scheme represents the diversity of *V(D)J* rearrangements obtained by sequencing *VDJ_H* and *DJ_H* rearrangements of the clone fractions indicated, as described in Fig. 4. From the represented clones, only clone 13-18 profile (for the IgM⁺ fraction) is compatible with a model of allelic preference. Clone no. 5.3 has a stop codon in the *Igh^a*-derived *DJ_H* joint that prevents IgM^a-productive rearrangements; clone no. 4.2 contains a prohibitive rearrangement in the *Igh^b*-derived *DJ_H* joint (see Supplementary Table 6); clone Xp29 contains a single non-productive *V_HDJ_H* rearrangement in the *Igh^a*-derived allele as inferred by phenotype.

cell; in contrast, from the status of rearrangement of the silent allele in IgM^a- and IgM^b-sorted populations we have found that in cells clonally related to a single HSC, either the maternal or the paternal alleles can rearrange the *V_H* first with similar frequency, thus showing that there is no stable epigenetic mark established

early in development that defines a temporal allelic hierarchy for *V_H* to *DJ_H* rearrangement. Farago *et al.*⁴⁶ have reported the same conclusion for the *Igκ* locus. In their study, multiple HSCs carrying distinguishable *Igκ* alleles (hCκ/mCκ) were transduced with lentiviruses (that integrate the genome of each cell in a

unique manner, thus identifying its clonal progeny), and then used in bulk to reconstitute mice. This technique was instrumental in the past to establish whether cell lineages share a precursor cell⁴⁷; however, we argue that it provides inconclusive results for allelic commitment. Since it is well established that 40–50% of B cells rearrange both *Igk* alleles⁴⁸, the progeny of each precursor cell is inevitably present in both the hCk⁺ and mCk⁺ populations and it follows that the pattern of lentiviral integrations in the isolated hCk⁺ and mCk⁺ populations is expected to be roughly the same, with or without commitment. Instead, from the *in vivo* clonal analysis here described based on the analysis of the status of rearrangement of the silent allele, we conclude that the parallel between the allelic exclusion of *Igh* and X-chromosome inactivation is misleading. It thus appears that either allele from the daughter pro-B cells of a given HSC can undergo V_H to DJ_H first, and that asynchronous replication is not a good predictor of the fate of the alleles, in $V(D)J$ recombination and other contexts. Indeed, it has been shown that replication timing is stable after X-inactivation but switches between alleles before the onset of the inactivation and is independent from the subsequent choice to inactivate the X-chromosome⁴⁹.

The data from CLP-derived clones revealed that both *Igh* alleles have the potential to undergo V_H to DJ_H first, until shortly before the onset of rearrangement. It cannot be completely excluded that the accessibility of the *Igh* locus in culture may differ from the *in vivo* conditions, particularly at the level of the initiation of the *Igh* locus accessibility, nuclear localization of the alleles and overall rearrangement kinetics. However, our culture conditions are consensually accepted as mimicking the B-cell development that occurs *in vivo* in the bone marrow³⁴, allelic exclusion was observed in the flow cytometric profiles of the clones, *in vitro* experiments performed months apart in different laboratories with cytokine supernatants from different batches produced similar results and a direct comparison of rearrangement efficiencies *in vitro* and *in vivo* did not reveal statistically significant differences (Supplementary Fig. 6). Strikingly, most of the clones with an extreme bias in the expression of IgM^a or IgM^b could be explained as a consequence of rearrangement dead ends, that is, those that permanently prevent one allele to be productively expressed at the membrane, giving the other the chance to undergo multiple productive rearrangements within the clone. This explanation of a bias determined by a limiting number of rearranging cells is compatible with the more balanced IgM^a/IgM^b ratios in clones derived from HSCs and MPPs, in which rearrangement starts relatively late in the cultures, after the initial cell has expanded substantially. It also explains the slightly more pronounced IgM^a/IgM^b bias in Ly6d⁺-derived clones compared with the clones derived from the precursor Ly6d⁻, since the former start rearranging in culture earlier than the latter. Finally, it makes sense that, in CLP-derived clones, the allele rearrangement bias observed for the *Igh* alleles is not reproduced by the *Igk* alleles, which typically rearrange after the substantial expansion of cells expressing the pre-BCR^{37,50}. This last result is the only one addressing the *Igk* locus in this work and does not reproduce the results presented in ref. 46. Further analysis is needed to address this issue thoroughly.

From these results, it can be concluded that the mechanism of allelic exclusion of the *Igh* locus differs from the one claimed for the *Igk* locus⁴⁶. In that study, it was proposed that stable epigenetic marks acquired around the time of lymphoid differentiation pre-determine a temporal hierarchy for rearrangement; however, the status of rearrangement of the silenced allele in the clones analysed was not reported, which limits the interpretation of the data. It is nevertheless possible that the *Igh* and the *Igk* loci behave differently, particularly given the organization of the V_H and J_H elements that allows successive

rearrangements in the *Igk* alleles and the existence of D_H elements in the *Igh* locus. Indeed, the *Igh* locus has more structural similarities with the *Tcrb* locus than with the *Igk* locus, namely the ordered rearrangement. Of note, the *Tcrb* alleles were found to independently associate with pericentromeric heterochromatin⁵¹, an interaction that inhibits rearrangement, and this observation is compatible with our conclusions for the *Igh* locus. Furthermore, this model is also the simplest explanation. Unlike the ‘pre-determination’ model, in which the strict temporal hierarchy for recombination implies an *ad hoc* mechanism to render the second allele active for recombination, in the model of independent recruitment the second rearrangement will eventually occur unless the first rearrangement triggers the feedback mechanism that shuts down recombination or the cell dies. The idea of independent recruitment is also compatible with the existence of a proportion of cells with ‘synchronous’ rearrangements, that is, cells in which the second V_H to DJ_H rearrangement occurs in the time window before the feedback mechanism triggered by the formation of the pre-BCR blocks further rearrangement of the *Igh* locus (Supplementary Fig. 7). It is not known what proportion of cells rearrange the alleles ‘synchronously’. However, in the absence of complementary feedback mechanisms mediated by ATM³ and the μ chain messenger RNA¹⁷, which respond to the first rearrangement faster than the feedback triggered by the pre-BCR, there is a measurable, although small, increase in the frequency of allelically included cells. We have confirmed this increase in genomic allelic inclusion for *Atm*^{-/-} cells in our assay (Fig. 2c). Finally, this model of independent rearrangement is not incompatible with the existence of moderate and transient allelic biases in the probability to undergo V_H to DJ_H rearrangement, if each allele behaves independently from the other while becoming accessible to rearrangement, but not from its previous copy in the mother cell. Thus, the simplest interpretation of our data is that in the vast majority of clones the two alleles gain recombination competence in a sufficiently synchronized way for the $V(D)J$ mechanism to reveal its stochastic component. Any case of an initial asymmetry that may persist for a number of divisions and lead to biases in the recombination does not contradict the model. Indeed, we may have identified such possible bias in only one of the clones studied, and these rare cases are expected from a truly independent opening of the alleles (Supplementary Fig. 7).

For over three decades, the explanations of allelic exclusion have essentially differed on the weight given to the stochastic and the regulated events. Using *in vivo* and *in vitro* clonal approaches that are based on the outcome of the rearrangement process in genetically unmanipulated murine cells, we found that an independent recruitment of the *Igh* alleles underlies the mechanism of allelic exclusion of the *Igh* locus. This intrinsic stochastic feature establishes the default mechanism that ensures two rearrangement attempts per cell, unless the first rearrangement triggers the definitive feedback that shuts down further rearrangement.

Methods

Mice. All mice were bred and maintained at the specific pathogen-free animal facilities of the *Instituto Gulbenkian de Ciéncia* (Oeiras, Portugal). BALB/c, C57BL/6J, B6.Rag2^{-/-}, B6.SJL-Ptprc^a (the pan-leukocytic marker gene *Ptprc*^a is also known as *CD45.1* or *Ly5.1*; the C57BL/6J strain carries the *Ptprc*^b allele, also known as *CD45.2* or *Ly5.2*) and B6.Cg-Igh^a were originally received from The Jackson Laboratory (Bar Harbor, ME, USA). Mice with *Foxp3*^{38fp} knock-in³¹ (B6.*Foxp3*^{38fp}, courtesy of Dr Alexander Rudensky—Memorial Sloan-Kettering Cancer Center, New York, USA) were backcrossed with the C57BL/6J strain at our animal facility. Animals were bred to generate female heterozygous B6.Cg-CD45.1/45.2 *Igh*^{a/b} *Foxp3*^{38fp/wt} and used in cell transfer experiments. Female 8–12 week-old-mice were

used as donors. For the *in vitro* cell culture system (BALB/c x C57BL/6J)F1, B6.Cg-Igh^{a/b} or B6;129.Igh^{a/b}Igk^{h/m} mice heterozygous for the *Igh* and/or the *Igk* loci were used (B6;129.Igk^h mice⁵² were courtesy of Dr Michel Nussenzweig (Rockefeller University, NY, USA). Mice with a targeted deletion for the immunoglobulin heavy chain joining region (B6.Igh-Jht^{-/-} mice³³) were courtesy of Dr Jocelyne Demengeot—IGC, Oeiras, Portugal) or ATM knockout mice (B6;129.Atm^{-/-} mice⁵³, courtesy of Dr Almudena Ramiro—CNIO, Madrid, Spain). This research project was reviewed and approved by the Ethics Committee of the Instituto Gulbenkian de Ciência, and by the Portuguese National Entity that regulates the use of laboratory animals (Direcção Geral de Alimentação e Veterinária. All experiments conducted on animals followed the Portuguese (Decreto-Lei n° 113/2013) and European (Directive 2010/63/EU) legislations concerning housing, husbandry and animal welfare.

Cell suspensions. All cells were suspended in Hanks' balanced-salt solution (HBSS, Gibco) supplemented with 5% fetal bovine serum (FBS, Life Technologies). BMs were flushed and single-cell-suspended from the femora and tibiae with a 26-gauge needle of 1-ml syringe, filtered and rinsed with HBSS 5% FBS. Thymic spleens and lymph nodes were dissected and single-cell suspensions obtained using a 70- μ m nylon mesh. Blood was collected from the facial vein in EDTA. Erythrocytes from blood and spleen samples were lysed using an NH₄Cl hypotonic solution at a 9:1 NH₄Cl:cell suspension ratio (no. 07850, STEMCELL Technologies), for 10 min and immediately rinsed at least twice and resuspended in HBSS with 5% FBS.

SP staining. BM cell suspensions were incubated with the FcBlock reagent (anti-CD16/32, in Rat serum—no. 13551, STEMCELL Technologies), labelled with the biotinylated antibodies anti-CD45R/B220, anti-CD19, anti-CD11b/Mac-1, anti-Ly6G/GR-1, anti-pan-NK, anti-Ly-76/TER and anti-CD3 (full list of antibodies and dyes in Supplementary Table 1), and incubated with Streptavidin microbeads (Miltenyi Biotec), for negative selection of lineage-positive cells by immunomagnetic separation using a MACS column (Miltenyi Biotec). Cells were further incubated in serum-free media (StemSpam SFEM, no. 09650, StemCell Technologies) with 20 μ g ml⁻¹ Rhodamine 123 (no. R302, Invitrogen) for 30 min in a water bath at 37°C (in the dark), cooled, rinsed with HBSS, 5% FBS and then stained with Hoechst 33342 (no. H3570, Invitrogen) at a concentration of 2.5 μ g ml⁻¹ diluted in the serum-free media as before, for 90 min at 37°C, in the dark. Cells were then labelled in HBSS with 5% FBS for 30 min on ice, with Streptavidin-APC and anti-CD45-PE and 1 μ g ml⁻¹ Propidium Iodide (no. P3566, Invitrogen). For the long-term reconstituting SP cells gating adjustment, an aliquot of the sample was incubated with 25 mg ml⁻¹ of Verapamil (no. 676777, Calbiochem) before staining with the dyes. Further stainings with different HSC-specific marker combinations were tested, confirming the enrichment of HSCs within the SP gates.

Cell sorting and *in vivo* single-cell repopulation assay. Single Lin⁻ Rho⁻ CD45^{int}Hoe⁻ long-term reconstituting SP cells of adult female mice in the C57BL/6J background, heterozygous for *CD45.1/45.2*, *Foxp3^{gfp/x}* and *Igh^{a/b}*, stained as above described, were sorted using the single-cell deposition unit of a MoFlo (Beckman Coulter) or FACSAria (Becton Dickinson) fluorescence-activated cell sorter machine into the individual wells of Terasaki plates (no. 452256, MicroWell 60-well MiniTray, Nunc Brand, Thermo Fisher Scientific Inc.) preloaded with 15 μ l of serum-free media (StemSpam SFEM, StemCell Technologies). Sorted IgM^a and IgM^b populations had purities above 95% (for most samples the purity was assessed with flow cytometry immediately after the sorting and for a few samples the purity was assessed with flow cytometry and/or semiquantitative PCR analysis of an allele-specific polymorphism in the *VH-D* intergenic region). To check for the purity of the single-cell sorting used in the *in vivo* experiments, each well was examined in a 4°C room using an inverted microscope and the wells containing single cells were used in reconstitutions (Supplementary Fig. 1).

Eight- to twelve-week-old recipient females of the B6.Rag2^{-/-} background received sublethal whole-body γ -irradiation with 600 cGy (Gammacell 2000 Mølsgaard Medical), 4–8 h before an intravenous retro-orbital injection with one or more test cells, as indicated. Recipient animals were analysed routinely every 2 weeks for up to 10 weeks post reconstitution for the presence of chimeric cells in the peripheral blood by staining with anti-CD45.1 and anti-CD45.2 antibodies (Supplementary Table 1). Clonally repopulated animals were killed and cells from BM, spleen, thymus and lymph nodes (LNs) were analysed for the presence of terminally differentiated B cells, T cells and myeloid cells of donor origin, as well as progenitor donor cells (antibodies used are described in Supplementary Table 1). Animals selected for subsequent analysis showed in Tregs *Foxp3^{gfp}* knock-in gene expression from a single X-chromosome. Splenic CD19⁺ IgM^{a+} and IgM^{b+} cells were then sorted on a FACSAria instrument, their DNA extracted and the recombination status of the *Igh* gene was analysed as described below.

All flow cytometric data collection was performed on a FACSCalibur (Becton Dickinson), a FACSAria (Becton Dickinson) or a CyAN ADP (Dako Cytomation, Beckman Coulter). The analysis was performed using the FlowJo software (Tree Star Inc.).

***In vitro* B-cell clonal differentiation from precursors.** BM cell suspensions enriched for haematopoietic progenitors (Lin⁻) were treated as described in ref. 34. Briefly, HSC, MPP and CLP cells were sorted using a FACSAria instrument after staining with anti-CD117(c-Kit), anti-Ly6A/E(Scal-1), anti-CD127(IL7R α) and anti-CD135(FLK2; Supplementary Table 1). Single cells were directly sorted into flat-bottom 96-well plates (Techno Plastic Products) previously seeded with non-irradiated 2.5 \times 10³ OP-9 cells per well, in complete culture media (OPTI-MEM, Invitrogen), supplemented with 10% FBS (Life Technologies) penicillin (50 U ml⁻¹), streptomycin (50 μ g ml⁻¹), β -mercaptoethanol (50 μ M) and Gentamycin (50 μ g ml⁻¹) and cultured in saturating amounts of the following cytokines: interleukin-7 (IL-7), c-Kit ligand and Flt3 ligand. Supernatants of myeloma cell lines transfected with the DNA encoding the above-mentioned cytokines were kindly provided by Dr Ana Cumano, Institut Pasteur, Paris, France.

After 12–14, 17–20 or 34 days of culture of CLPs, MPPs or HSCs, respectively, individual plates were analysed for single colonies. Wells with single colonies were stained with anti-CD19, anti-IgM^a and anti-IgM^b (Supplementary Table 1) analysed with flow cytometry and each IgM⁺ allelic fraction (and in some cases also the IgM⁻ fraction) was sorted in bulk using a FACSAria cell sorter instrument as described above.

DNA samples. Genomic DNA (gDNA) from *in vivo*-differentiated clonal CD19⁺ IgM^a and CD19⁺ IgM^b pools of at least 10⁴ sorted cells was isolated by proteinase K (Sigma) digestion, phenol-chloroform-extracted, precipitated by ethanol and resuspended in TE buffer. The set of *in vitro*-differentiated sorted samples included also CD19⁺ IgM⁻ populations and the numbers of sorted cells were invariably smaller than those of the *ex vivo* samples (at least 2,000 cells); for these *in vitro*-derived samples, gDNA was obtained by incubation with cell digestion medium containing 400 ng μ l⁻¹ proteinase K and 17 μ M sodium dodecyl sulfate (SDS, Sigma) in PBS (Invitrogen) for 1 h at 50°C, followed by enzyme heat inactivation for 30 min at 95°C. gDNA thus obtained was diluted to homogeneous cell-equivalents per μ l, as determined by the number of cells sorted, and used directly in PCR reactions, as described below. Treatment of control samples was matched to that of each set of samples. As controls for *in vivo*-differentiated clones, *ex vivo* splenocytes of (C57BL/6 \times BALB/c)F1 WT animals and mice of the *Rag2^{-/-}*, *Jht^{-/-}*, *Jht^{+/-}* and *Atm^{-/-}* genotype were sorted and treated as the test samples. qDNA from the cells lines 18–81 and PD31 (cell lines kindly provided by Dr Michele Goodhardt, Institut Universitaire d'Hématologie, Paris, France and Dr Patricia Cortes, Icahn Medical Institute, New York, USA) were used as negative controls for the presence of *VH-DH* intergenic fragment.

Real-time quantitative PCR. The frequency of *DJ_H* only rearranged alleles in the IgM^a and IgM^b populations of the same B-cell clones was determined with real-time qPCR, with the test and reference sequences localized in the recombining and non-recombining regions of the *Igh* locus, as shown in Supplementary Fig. 2a. The test primer pairs, VD4FW_5'-GCAGGAAGACCACAGGAGC-3' and VD4RE_5'-AGCCACTGTAGATGAGGGT-3', amplify a ~100-bp fragment herein termed '*VH-D* intergenic fragment' downstream of the last *VH*-gene segment of the *Igh* locus and upstream of the first *DH*-gene segment. The reference primer pairs, CmuFW_5'-GTTCTGTGCCTCCGCTAGC-3' and CmuRE_5'-CTTGGGTGACTGAGCATTTC-3', also amplify a ~100-bp fragment located in the μ -constant region of the *Igh* locus (*C μ*). Each primer pair was included in a different reaction mixture and runs were performed in triplicates or quadruplicates using the 7900HT Fast Real-Time PCR System (Applied Biosystems), in 384-well plates with optical tape sealers (Applied Biosystems). Reactions were performed in 10 μ l final volume containing 1 \times the Power SYBR Green PCR Master Mix (Applied Biosystems) and 0.2 μ M of each primer. Thermocycling conditions were 95°C for 10 min, followed by 45 cycles of 95°C for 15 s and 60°C for 60 s. The melting curve of the finished PCR reaction was systematically used to confirm that a single amplicon was amplified.

Calibration curves were performed using six standard dilutions, in triplicates, of B6.Jht^{+/-} gDNA from *ex vivo*-sorted IgM⁺ cells, ranging from 17 cell equivalents up to 54,240 cell equivalents per reaction (Supplementary Fig. 2b), and test sample concentrations used per run ranged from 200 to 700 cell equivalents per run. No template control reactions were performed using water or a diluted digestion medium solution without template. Beyond replicas, independent repetition of each sample quantification was performed whenever the sample amount was sufficient, in diverse dilutions, but given the low amount of DNA in the majority of samples it was not possible to repeat the PCR on a different day. The SDS software v.2.2 (Applied Biosystems) was used to calculate the quantification cycle (Cq) value. Real-time quantification data were averaged between replicates and relative amount of DNA was calculated with the *C μ* constant fragment quantification as the reference, using the delta Cq method adjusted for PCR efficiency⁵⁴.

***Igh* rearrangement PCR and sequencing.** For the amplification of *DJ_H* and *V_HDJ_H* rearrangements, we used the primer collection and the two-step nested PCR conditions previously described¹⁸. One thousand cell-equivalents per run were used for the first round multiplex PCR of *V_HDJ_H*-rearrangement amplifications in a 100- μ l final master mix volume. For the nested round of PCR,

alternative reverse primers were J_{H1-1} , J_{H2-2} and J_{H3-3} -specific (JH1A—5'-TGTGGCA GATGGCTGACATGGGG-3', JH2A-5'-TCAGCCAGGGCTCCCAATGACC C-3', and JH3A-5'-CAGCAGGCAGAGAGTCCCTGACCC-3'), allowing allotype discrimination upon sequencing (Supplementary Fig. 3). For sequencing, PCR bands were cloned in the pGEM-T Easy VECTOR SYSTEM (Promega) and a number of clones were analysed. Typically, four clones were sequenced per band. Given both the high number of PCR cycles used and the cloning step before sequencing the $V_H D J_H$ amplicons, it was necessary to evaluate the impact in the analysis of PCR/sequencing-introduced errors. This was estimated using the formula ($f_{mut} \times p_{stop} + f_{indel} - f_{mut} \times f_{stop} \times f_{indel}$) $\times L/1,000$, where f_{mut} and f_{indel} stand for the number of PCR/sequencing-introduced point mutations (2.6) and the number of PCR/sequencing-introduced gaps/insertions (0.18) per 1,000 nucleotides, respectively, estimated from alignments of the germline regions (a total of 16,020 bp) downstream of the J_H elements that we have sequenced, p_{stop} is the probability of producing a stop codon from a PCR-introduced point mutation (0.071) taking in consideration the codon usage of the $V_H D J_H$ sequences obtained in this study and L is the average size of the amplicons. If the entire V_H and CDR3 regions ($L = 269$ bp) of the amplicon are considered, the percentage of productive $V_H D J_H$ rearrangements that become non-productive because of PCR errors is 8.9%. This percentage decreases to 2.0% if only a sequence region including the entire CDR3 is used ($L = 60$ bp), which is feasible since the PCR errors in the V_H region can be aligned with the known V_H germline sequences from antigen receptor databases. In any case, the errors introduced by PCR do not affect our conclusions.

References

- Cebra, J. J., Colberg, J. E. & Dray, S. Rabbit lymphoid cells differentiated with respect to alpha-, gamma-, and mu- heavy polypeptide chains and to allotypic markers Aa1 and Aa2. *J. Exp. Med.* **123**, 547–558 (1966).
- Burnet, F. M. A modification of Jerne's theory of antibody production using the concept of clonal selection. *Aust. J. Sci.* **20**, 67–68 (1957).
- Hewitt, S. L. *et al.* RAG-1 and ATM coordinate monoallelic recombination and nuclear positioning of immunoglobulin loci. *Nat. Immunol.* **10**, 655–664 (2009).
- Tonegawa, S. Somatic generation of antibody diversity. *Nature* **302**, 575–581 (1983).
- 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).
- Barreto, V. & Cumano, A. Frequency and characterization of phenotypic Ig heavy chain allelically included IgM-expressing B cells in mice. *J. Immunol.* **164**, 893–899 (2000).
- Alt, F. W. *et al.* Ordered rearrangement of immunoglobulin heavy chain variable region segments. *EMBO J.* **3**, 1209–1219 (1984).
- Selimyan, R. *et al.* Localized DNA demethylation at recombination intermediates during immunoglobulin heavy chain gene assembly. *PLoS Biol.* **11**, e1001475–e1001475 (2013).
- Rose, S. M. S., Kuehl, W. M. W. & Smith, G. P. G. Cloned MPC 11 myeloma cells express two kappa genes: a gene for a complete light chain and a gene for a constant region polypeptide. *Cell* **12**, 453–462 (1977).
- Alt, F. F., Rosenberg, N. N., Lewis, S. S., Thomas, E. E. & Baltimore, D. D. Organization and reorganization of immunoglobulin genes in A-MULV-transformed cells: rearrangement of heavy but not light chain genes. *Cell* **27**, 381–390 (1981).
- Weaver, D., Costantini, F., Imanishi-Kari, T. & Baltimore, D. A transgenic immunoglobulin mu gene prevents rearrangement of endogenous genes. *Cell* **42**, 117–127 (1985).
- Nussenzweig, M. C. *et al.* Allelic exclusion in transgenic mice that express the membrane form of immunoglobulin mu. *Science* **236**, 816–819 (1987).
- Kitamura, D. & Rajewsky, K. Targeted disruption of mu chain membrane exon causes loss of heavy-chain allelic exclusion. *Nature* **356**, 154–156 (1992).
- Boekel, E. T., Melchers, F. & Rolink, A. G. Precursor B cells showing H chain allelic inclusion display allelic exclusion at the level of pre-B cell receptor surface expression. *Immunity* **8**, 199–207 (1998).
- Löffert, D., Ehlich, A., Muller, W. & Rajewsky, K. Surrogate light chain expression is required to establish immunoglobulin heavy chain allelic exclusion during early B cell development. *Immunity* **4**, 133–144 (1996).
- Hauser, J., Grundström, C. & Grundström, T. Allelic exclusion of IgH through inhibition of E2A in a VDJ recombination complex. *J. Immunol.* **192**, 2460–2470 (2014).
- Lutz, J. *et al.* Pro-B cells sense productive immunoglobulin heavy chain rearrangement irrespective of polypeptide production. *Proc. Natl Acad. Sci. USA* **108**, 10644–10649 (2011).
- Ehlich, A., Martin, V., Muller, W. & Rajewsky, K. Analysis of the B-cell progenitor compartment at the level of single cells. *Curr. Biol.* **4**, 573–583 (1994).
- Kosak, S. T. *et al.* Subnuclear compartmentalization of immunoglobulin loci during lymphocyte development. *Science* **296**, 158–162 (2002).
- Sayegh, C. E., Sayegh, C., Jhunjhunwala, S., Riblet, R. & Murre, C. Visualization of looping involving the immunoglobulin heavy-chain locus in developing B cells. *Genes Dev.* **19**, 322–327 (2005).
- Roldán, E. *et al.* Locus 'decontraction' and centromeric recruitment contribute to allelic exclusion of the immunoglobulin heavy-chain gene. *Nat. Immunol.* **6**, 31–41 (2005).
- Guo, C. C. *et al.* CTCF-binding elements mediate control of V(D)J recombination. *Nature* **477**, 424–430 (2011).
- Fuxa, M. *et al.* Pax5 induces V-to-DJ rearrangements and locus contraction of the immunoglobulin heavy-chain gene. *Genes Dev.* **18**, 411–422 (2004).
- Skok, J. A. *et al.* Nonequivalent nuclear location of immunoglobulin alleles in B lymphocytes. *Nat. Immunol.* **2**, 848–854 (2001).
- Holwerda, S. J. B. *et al.* Allelic exclusion of the immunoglobulin heavy chain locus is independent of its nuclear localization in mature B cells. *Nucleic Acids Res.* **41**, 6905–6916 (2013).
- Chess, A. Random and non-random monoallelic expression. *Neuropsychopharmacology* **38**, 55–61 (2012).
- Singh, N. *et al.* Coordination of the random asynchronous replication of autosomal loci. *Nat. Genet.* **33**, 339–341 (2003).
- Mostoslavsky, R. *et al.* Asynchronous replication and allelic exclusion in the immune system. *Nature* **414**, 221–225 (2001).
- Pereira, J. P., Girard, R., Chaby, R., Cumano, A. & Vieira, P. Monoallelic expression of the murine gene encoding Toll-like receptor 4. *Nat. Immunol.* **4**, 464–470 (2003).
- Osawa, M., Hanada, K., Hamada, H. & Nakauchi, H. Long-term lymphohematopoietic reconstitution by a single CD34-low/negative hematopoietic stem cell. *Science* **273**, 242–245 (1996).
- Fontenot, J. D. *et al.* Regulatory T cell lineage specification by the forkhead transcription factor foxp3. *Immunity* **22**, 329–341 (2005).
- Vettermann, C. & Schlissel, M. S. Allelic exclusion of immunoglobulin genes: models and mechanisms. *Immunol. Rev.* **237**, 22–42 (2010).
- Gu, H. H., Zou, Y. R. Y. & Rajewsky, K. K. Independent control of immunoglobulin switch recombination at individual switch regions evidenced through Cre-loxP-mediated gene targeting. *Cell* **73**, 1155–1164 (1993).
- Vieira, P. & Cumano, A. Differentiation of B lymphocytes from hematopoietic stem cells. *Methods Mol. Biol.* **271**, 67–76 (2004).
- Inlay, M. A. *et al.* Ly6d marks the earliest stage of B-cell specification and identifies the branchpoint between B-cell and T-cell development. *Genes Dev.* **23**, 2376–2381 (2009).
- Decker, D. J., Boyle, N. E., Koziol, J. A. & Klinman, N. R. The expression of the Ig H chain repertoire in developing bone marrow B lineage cells. *J. Immunol.* **146**, 350–361 (1991).
- Hess, J. *et al.* Induction of pre-B cell proliferation after de novo synthesis of the pre-B cell receptor. *Proc. Natl Acad. Sci. USA* **98**, 1745–1750 (2001).
- Wang, Y.-H. *et al.* V(D)J recombinatorial repertoire diversification during intracлонаl pro-B to B-cell differentiation. *Blood* **101**, 1030–1037 (2003).
- Gu, H., Kitamura, D. & Rajewsky, K. B cell development regulated by gene rearrangement: arrest of maturation by membrane-bound μ protein and selection of DHelement reading frames. *Cell* **65**, 47–54 (1991).
- Allen, E. *et al.* High concentrations of long interspersed nuclear element sequence distinguish monoallelically expressed genes. *Proc. Natl Acad. Sci. USA* **100**, 9940–9945 (2003).
- Chow, J. C. *et al.* LINE-1 activity in facultative heterochromatin formation during X chromosome inactivation. *Cell* **141**, 956–969 (2010).
- Kitsberg, D. *et al.* Allele-specific replication timing of imprinted gene regions. *Nature* **364**, 459–463 (1993).
- Chess, A., Simon, I., Cedar, H. & Axel, R. Allelic inactivation regulates olfactory receptor gene expression. *Cell* **78**, 823–834 (1994).
- Takagi, N. Differentiation of X chromosomes in early female mouse embryos. *Exp. Cell Res.* **86**, 127–135 (1974).
- Lebon, J. M., Tam, P. P., Singer-Sam, J., Riggs, A. D. & Tan, S. S. Mouse endogenous X-linked genes do not show lineage-specific delayed inactivation during development. *Genet. Res.* **65**, 223–227 (1995).
- Farago, M. *et al.* Clonal allelic predetermination of immunoglobulin- κ rearrangement. *Nature* **490**, 561–565 (2012).
- Keller, G., Paige, C., Gilboa, E. & Wagner, E. F. Expression of a foreign gene in myeloid and lymphoid cells derived from multipotent hematopoietic precursors. *Nature* **318**, 149–154 (1985).
- Mostoslavsky, R. & Alt, F. W. The lingering enigma of the allelic exclusion mechanism. *Cell* **118**, 539–544 (2004).
- Gribnau, J. X chromosome choice occurs independently of asynchronous replication timing. *J. Cell Biol.* **168**, 365–373 (2005).
- Yamagami, T., Boekel, ten, E., Andersson, J., Rolink, A. & Melchers, F. Frequencies of multiple IgL chain gene rearrangements in single normal or kappaL chain-deficient B lineage cells. *Immunity* **11**, 317–327 (1999).

51. Schlimgen, R. J., Reddy, K. L., Singh, H. & Krangel, M. S. Initiation of allelic exclusion by stochastic interaction of Tcrb alleles with repressive nuclear compartments. *Nat. Immunol.* **9**, 802–809 (2008).
52. Casellas, R. R. *et al.* Contribution of receptor editing to the antibody repertoire. *Science* **291**, 1541–1544 (2001).
53. Elson, A. *et al.* Pleiotropic defects in ataxia-telangiectasia protein-deficient mice. *Proc. Natl Acad. Sci. USA* **93**, 13084–13089 (1996).
54. Pfaffl, M. W. A new mathematical model for relative quantification in real-time RT-PCR. *Nucleic Acids Res.* **29**, e45 (2001).
55. Hesse, J. E., Lieber, M. R., Gellert, M. & Mizuuchi, K. Extrachromosomal DNA substrates in pre-B cells undergo inversion or deletion at immunoglobulin V-(D)-J joining signals. *Cell* **49**, 775–783 (1987).
56. 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).

Acknowledgements

We thank Marie Bonnet, Thiago Carvalho, Ana Cumano, Revati Masilamani, Gillian Wu and all the members of our laboratory for the critical reading of the manuscript. C.F.A.P. is supported by an FCT fellowship (SFRH/BD/33217/2007). This work was supported by an FCT grant (PTDC/BIA-BCM/108020/2008).

Author contributions

C.F.A.P. and V.M.B. designed the experiments, analysed the data, discussed the results and wrote the manuscript. C.F.A.P. performed all the experiments. R.de.F and F.M. provided technical assistance. T.L. and R.G. gave support concerning flow cytometry. P.V. advised and gave support to all experiments. V.M.B. supervised all the work.

Additional information

Supplementary Information accompanies this paper at <http://www.nature.com/naturecommunications>

Competing financial interests: The authors declare no competing financial interests.

Reprints and permission information is available online at <http://npg.nature.com/reprintsandpermissions/>

How to cite this article: Pereira, C. F. A. *et al.* Independent recruitment of *Igh* alleles in V(D)J recombination. *Nat. Commun.* 5:5623 doi: 10.1038/ncomms6623 (2014).



This work is licensed under a Creative Commons Attribution 4.0 International License. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>