European Journal of Immunology

SHORT COMMUNICATION

Single-cell based high-throughput sequencing of full-length immunoglobulin heavy and light chain genes

Christian E. Busse[1], Irina Czogiel[2], Peter Braun[3], Peter F. Arndt[2] and Hedda Wardemann[1]

[1] Research Group Molecular Immunology, Max Planck Institute for Infection Biology, Berlin, Germany
[2] Department for Computational Molecular Biology, Max Planck Institute for Molecular Genetics, Berlin, Germany
[3] Department of Molecular Biology, Max Planck Institute for Infection Biology, Berlin, Germany

Single-cell PCR and sequencing of full-length Ig heavy (*Igh*) and *Igk* and *Igl* light chain genes is a powerful tool to measure the diversity of antibody repertoires and allows the functional assessment of B-cell responses through direct Ig gene cloning and the generation of recombinant mAbs. However, the current methodology is not high-throughput compatible. Here we developed a two-dimensional bar-coded primer matrix to combine *Igh* and *Igk*/*Igl* chain gene single-cell PCR with next-generation sequencing for the parallel analysis of the antibody repertoire of over 46 000 individual B cells. Our approach provides full-length *Igh* and corresponding *Igk*/*Igl* chain gene-sequence information and permits the accurate correction of sequencing errors by consensus building. The use of indexed cell sorting for the isolation of single B cells enables the integration of flow cytometry and Ig gene sequence information. The strategy is fully compatible with established protocols for direct antibody gene cloning and expression and therefore advances over previously described high-throughput approaches to assess antibody repertoires at the single-cell level.

Keywords: Antibody repertoire · B cell · Index cell sorting

Additional supporting information may be found in the online version of this article at the publisher's web-site

Introduction

Antibodies are B-cell antigen receptors encoded by two independent Ig heavy (*Igh*) and light (*Igk* or *Igl*) chain genes. Therefore, the accurate assessment of antibody repertoires and clonal relationships between individual B cells requires the combined analysis of *Igh* and *Igk/Igl* genes from the same cell. Conventional single-cell (sc) based antibody repertoire studies use Sanger sequencing of *Igh* and corresponding *Igk/Igl* PCR products to assess V(D)J gene usage, levels of somatic hypermutations, and B-cell clonality [1–5]. This process is highly efficient, unbiased, and has been widely used to assess antibody repertoires in health and disease in mice and humans. Available single-cell PCR (scPCR) protocols provide full-length gene information and are compatible with the direct cloning of *Igh* and corresponding *Igk/Igl* genes [3, 4]. This facilitates the production of recombinant mAbs, which have been

Correspondence: Hedda Wardemann
e-mail: wardemann@mpiib-berlin.mpg.de

© 2013 WILEY-VCH Verlag GmbH & Co. KGaA, Weinheim

used to study B-cell repertoires based on Ig reactivity and function. Applications for scPCR range from the analysis of antibody responses in immune diseases to the identification of protective mAbs against diverse human pathogens [5–11]. However, scPCR is costly and time consuming and therefore limited to a relatively small number of several hundred input cells.

With the advent of next-generation sequencing (NGS) technologies, high-throughput sequencing of *Igh* and to a lesser extent *Igk/Igl* light chains from bulk isolated B cells has been widely used to analyze Ig gene repertoires in mice and humans [12–16]. A major drawback of NGS from bulk populations is the inherent loss of information about the association between *Igh* and *Igk/Igl* genes at cellular level, which prevents the phylogenetic analysis of B-cell responses and the faithful functional assessment of antibody reactivity after Ig gene cloning and in vitro expression of recombinant mAbs. Physical linkage of *Igh* and *Igk/Igl* PCR amplicons before sequencing can overcome this shortcoming, but does not provide full-length Ig gene sequence information due to short NGS read lengths and is therefore not compatible with the direct and high-throughput generation of recombinant mAbs [17]. In addition, NGS is associated with high-error rates. Computational filters can mask sequencing errors but cannot distinguish them from naturally occurring somatic mutations in Ig genes of antigen-experienced B cells.

Here we combined the scPCR-based amplification of *Igh* and *Igk/Igl* transcripts and NGS into a unified pipeline that allows the high-throughput antibody repertoire analysis of thousands of single isolated B cells. The platform is scalable and provides high-quality sequence information by consensus building of multiple reads originating from the same cell. It integrates sequencing and flow cytometry data obtained by indexed cell sorting into a single data set and thereby allows for the retrospective phenotypic characterization of B cells according to their antibody genotype. The strategy provides full-length Ig gene sequences of corresponding *Igh* and *Igk/Igl* genes and is therefore ideally suited for phylogenetic B-cell repertoire analyses and fully compatible with established procedures for antibody cloning and expression [3]. Importantly, the tag matrix can be used universally to barcode primer sets for the single-cell-based amplification of Ig genes from any species including humans [4].

Results

Primer matrix design

The aim of our experimental strategy was to combine the advantages of scPCR and NGS to facilitate the massive parallel analysis of B-cell antibody repertoires (Fig. 1). To perform high-throughput parallel NGS of pooled PCR products without loss of *Igh* and *Igk/Igl* linkage, we designed a primer matrix that incorporates unique identification tags on both ends of the PCR amplicons and thereby allows backtracking of each *Igh* and *Igk/Igl* sequence read to the original cell. The two-dimensional (2D) nature of the primer matrix with distinct row and column tags facilitates the cost-efficient mapping of large numbers of PCR amplicons with a relatively small number of differently tagged primers, i.e. backtracking of NGS reads from $x \cdot y$ PCR amplicons to the original cell per well requires only $x + y$ different primers. Thus for example only 200 different primer tags are sufficient to reliably map reads from 10 000 PCR amplicons back to their respective cell or well of origin. The expected amplicon size (including both linkers) of around 575, 430, and 470 bp, for *Igh*, *Igk*, and *Igl*, respectively, requires sequencing platforms that provide sufficient read lengths to cover the proximal and distal tags. This read length is generated by the 454 pyrosequencing technology. However, NGS is error-prone compared to conventional Sanger sequencing. Therefore, the efficient tag identification and read mapping requires built-in mechanisms to detect and correct sequencing errors. To be able to reliably identify all primer tags after 454 sequencing, we designed a set of 4096 16 bp tags with balanced GC content and lack of homopolymer stretches that permit the detection of up to six and the correction of up to three sequencing errors based on a mutual Hamming distance of six [18]. Next we randomly picked 72 of these tags to design a 2D-tagged primer matrix (24×48) for the amplification of *Igh*, *Igk*, and *Igl* transcripts from 1152 single-isolated murine B cells.

Validation of primer matrix

To test the 2D primer matrix, we isolated individual peritoneal cavity B cells (plate 1), bone marrow plasma cells (plate 2), and B220^{+} B cells from spleen (plate 3) by flow cytometric index cell sorting into 384-well plates (Supporting Information Fig. 1–3). cDNA was prepared in the original sort plates and used to amplify *Igh*, *Igk*, and *Igl* genes following a previously described nested scPCR strategy [3]. The use of tagged matrix primers in all nested PCRs allowed pooling of all *Igh*, *Igk*, and *Igl* amplicons for pyrosequencing. In total, 219 164 reads were obtained after 454 sequencing, of which 91.3% were of sufficient quality to identify both tags (Fig. 2A). Sequence stretches of lower Q-values were associated with homopolymer stretches mainly in *Ighm* and *Igl*-V gene segments (Supporting Information Fig. 4). As expected, all proximal tags were located within the first 10 bps, whereas the distal tags followed a bimodal distribution corresponding to the median amplicon length of 510 and 385 bps, respectively, for *Igh* and *Igk/Igl* (Fig. 2B). In total, 95% of proximal tags and over 85% of distal tag sequences required no error correction whereas correction of one or two errors was sufficient for the unequivocal identification of the remaining 5% proximal and 15% distal tags (Fig. 2C and D). Single or double nucleotide insertions, deletions, or replacements were the most common errors, whereas only few sequences (0.1 and 1.3% of total for proximal and distal tags, respectively) showed combinations of these. The slightly higher error rate in distal than proximal tags was associated with the lower sequencing quality toward the end of the reads, but did not interfere with the reliable tag identification (Fig. 2D).

After error correction, we mapped all reads containing both tags to the original well position (Fig. 2E). On average, 213, 85,

© 2013 WILEY-VCH Verlag GmbH & Co. KGaA, Weinheim

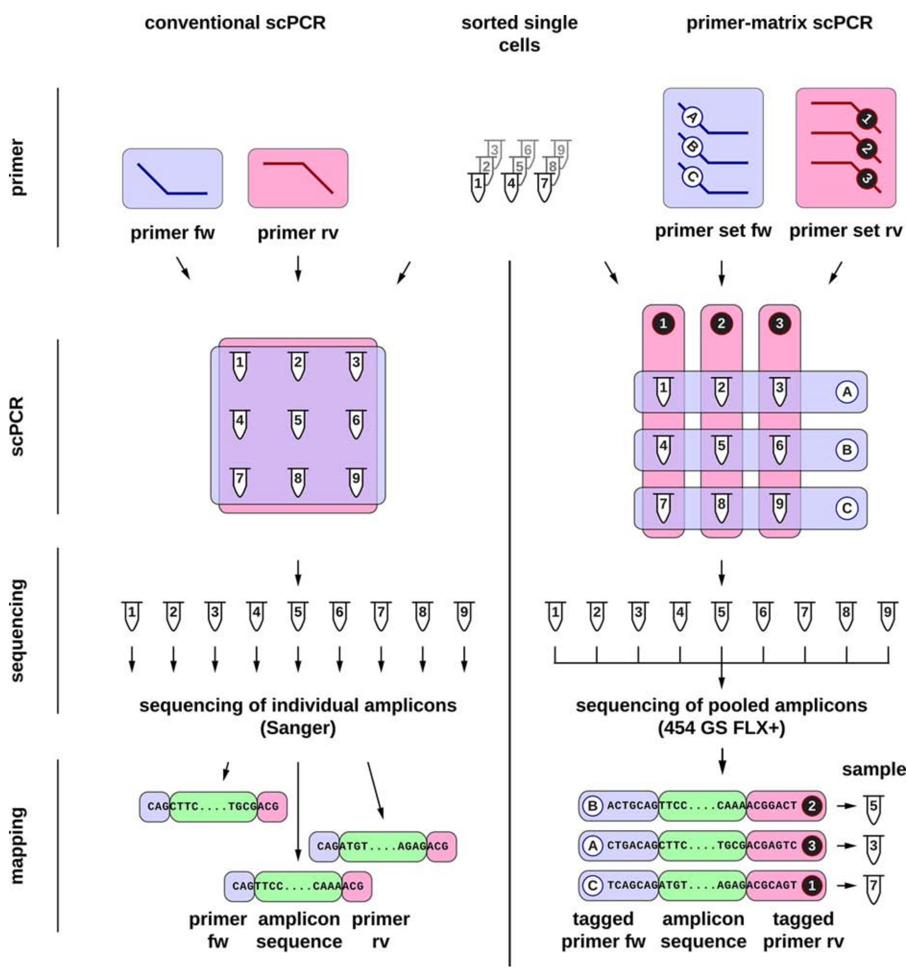

Figure 1. Comparison between conventional and primer-matrix scPCR. Schematic overview of the conventional (left) and primer matrix based (right) Ig gene scPCR and sequencing approach. Not shown is the initial single-cell isolation into 384-well PCR plates by indexed FACS. Primers and scPCR: Conventional scPCR uses identical pairs of primers or primer pools for all samples (left). In contrast, matrix scPCR uses individual forward and reverse primers or primer pools with identical tags in all wells from each column and row (right). Sequencing: Conventional scPCR uses Sanger sequencing of each amplicon (left). Matrix primer-based scPCR allows pooling of all amplicons followed by NGS on a 454 GS FLX+ platform (right). Mapping: Conventional scPCR does not require mapping of sequences, since all PCR products from a given cell per well are sequenced individually (left). For matrix scPCR, 454 sequences are mapped back to the original cell per well based on the respective row and column tag identification (right).

and 38 reads per well were obtained for *Igh*, *Igk*, and *Igl* amplicons, respectively (10–90% range: 43–405 for *Igh*, 23–221 for *Igk*, and 18–75 for *Igl*), and were used for consensus sequence building. Based on the manual analysis of sequences from peritoneal B cells and naïve spleen B cells, which lack Ig gene somatic mutations, we observed that the minimum number of reads necessary for accurate consensus building was five. In total, *Igh* and corresponding *Igk* or *Igl* genes were amplified from 404 out of all 1152 cells. Over 99% of the peritoneal and splenic B cells expressed the *Ighm* constant region, whereas the majority of bone marrow plasma cells were class-switched and expressed *Igha* (56%) or *Ighg2a,b* (5%) isotype transcripts (Supporting Information Table 4). Although plasma cells are expected to contain higher amounts of Ig gene transcripts than peritoneal or naïve spleen B cells, no strong differences in efficiency were observed between the three populations. This suggests that Ig isotype subclass and transcript levels were not major limiting factors during the scPCR and Ig gene sequencing procedure.

To validate the scalability of our approach, we tested a second independent 240 × 192 2D-tagged primer matrix designed for the analysis of up to 4.6×10^4 cells or 10 × 12 384-well plates. The test was performed using three plates of single sorted peritoneal cavity B cells, splenic marginal zone B cells, and mature naïve B cells corresponding to plate position 4, 65, and 81 in the 2D-tagged primer matrix, respectively (Supporting Information Fig. 5). Sequence analysis showed similar read qualities and tag identification efficiencies for this subset and yielded full-length *Igh* and corresponding *Igk/Igl* gene information from 33% (381/1152) of all cells (Supporting Information Fig. 3 and 5).

Thus, we conclude that the designed primer tags and scPCR strategy can be used for the efficient high-throughput NGS-based analysis of antibody repertoires.

Linkage of flow cytometric and sequencing data

All single B cells had been isolated by index cell sorting, a procedure that records the flow cytometric phenotype of each sorted cell. It thereby facilitates direct linkage of phenotype with *Igh* and corresponding *Igk/Igl* genotype data after sequencing. To determine the accuracy of the indexed cell sorting procedure, we made use of the fact that peritoneal B cells frequently express canonical VH11/Vκ9 and VH12/Vκ4 antibodies [19, 20]. Indeed, canonical antibodies originated from $CD5^{+}B220^{low}$ B1a B cells (Fig. 3A). Overlap between genotype and phenotype was also observed for

© 2013 WILEY-VCH Verlag GmbH & Co. KGaA, Weinheim

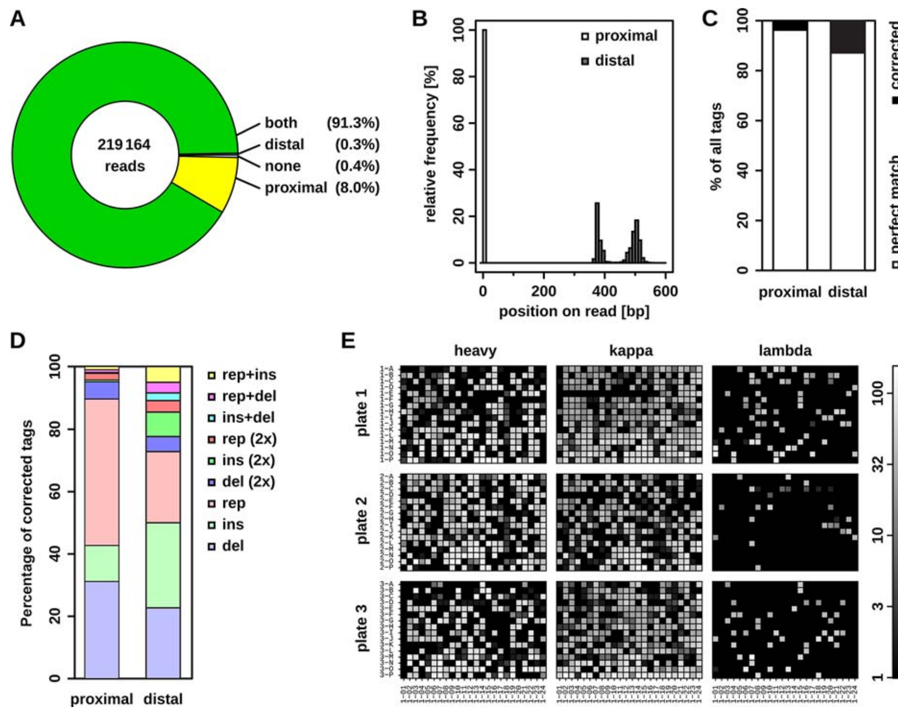

Figure 2. NGS and tag identification for primer matrix 1. (A) Pie diagram shows the frequency of reads with or without identification of the proximal (yellow), distal, or proximal and distal (yellow) tag. (B) The relative frequency of proximal and distal tag positions for combined *Igh*, *Igk*, and *Igl* sequences is shown. (C) The frequency of proximal and distal tags that require error correction compared with that of perfect match tag reads is shown. (D) The frequency of tags with insertions (ins), deletions (del), and replacements (rep) or combinations of these is shown. (E) Log-scaled heat map of the number of *Igh*, *Igk*, and *Igl* reads obtained per cell per well of single sorted C57BL/6 peritoneal cavity B cells (plate 1), bone marrow plasma cells (plate 2), and naïve spleen B cells is shown. Data shown are from a single experiment.

IgA$^+$ and IgG$^+$ bone marrow PCs suggesting that indexed cell sorting provides a powerful tool to integrate phenotype and genotype data at the single-cell level (Fig. 3B).

Concluding remarks

Here we demonstrate that scPCR and NGS can be integrated into an automated platform for the high-throughput analysis of antibody repertoires at the monoclonal level. The 2D-tagged primer matrix allows the unequivocal backtracking of pooled *Igh* and *Igk/Igl* amplicons to the original cell after NGS. The approach is cost efficient due to the 2D nature of the matrix, which allows the use of primers with identical tags in all wells of the same column or row, respectively. The matrix is scalable for the high-throughput analysis of a wide range of cell numbers.

The primers described here have been successfully used in conventional scPCRs to amplify BALB/c and 129 family alleles [21,22] suggesting that the matrix strategy should be directly applicable for single-cell based Ig gene repertoire analyses of these mouse strains (Supporting Information Table 2). In addition, the tags can be used universally to barcode any primer set including primers for scPCR of human Ig genes [4]. Unfortunately, so far no single promiscuous or degenerate primer has been described for the unbiased scPCR amplification of human Ig genes. Standard protocols rely on the usage of primer sets, each of which would have to be tagged for use in a matrix scPCR and NGS sequencing, an approach that will be highly costly. To increase the flexibility of

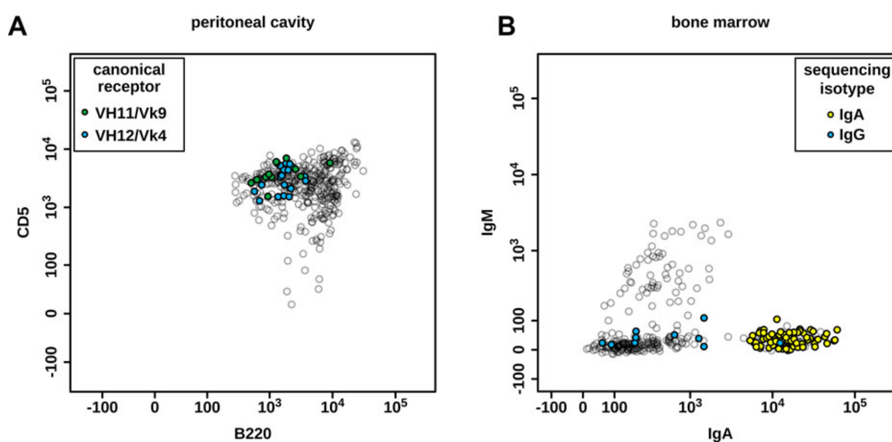

Figure 3. Overlay of sequencing and flow cytometry data. Cell surface markers indicated on the X- and Y-axes were determined by flow cytometry at the time of single-cell sorting. Information derived from *Igh* transcripts is shown by color mark-up as indicated in the legend. Cells were either (A) peritoneal cavity B cells or (B) bone marrow plasma cells. Data shown are from one experiment representative of two performed.

© 2013 WILEY-VCH Verlag GmbH & Co. KGaA, Weinheim

the system, additional tags may be incorporated by tertiary PCRs prior to NGS. For example, PCR products from individual plates that were amplified with the same 2D primer matrix may be barcoded with a secondary tag to discriminate pooled amplicons from individual plates after sequencing.

Until recently, antibody repertoire studies based on 2D tagging at the proximal and distal end of Ig gene amplicons, were not feasible due to read-length limitations of the available NGS platforms. Here we show that 454 pyrosequencing provides sufficient read quality over the length of the amplicon for the efficient and unequivocal identification and backtracking of *Igh* and *Igk/Igl* gene sequences. Whether alternative platforms would perform similarly in terms of read-length and quality remains to be determined [23].

FACS sorting is ideally suited to isolate large numbers of circulating or tissue resident B cells for high-throughput matrix scPCR. However in principle, matrix scPCR is compatible with all cell isolation methods that preserve mRNA and cell integrity to allow for the efficient and reliable amplification of IgH and associated IgL chain transcripts from individual cells, including, for example, laser microdissection of cells from tissue sections.

In contrast to the recently described high-throughput approach developed by DeKosky et al. [17], our strategy preserves a part of the cDNA and amplicon material from each individual cell. It is therefore fully compatible with direct Ig gene cloning and expression, which has been widely used for the functional analysis of the B-cell antibody repertoire in infectious and noninfectious disease settings including autoimmunity and immunodeficiency. In addition, the use of random hexamer primers for cDNA synthesis leaves the option to study and link gene expression patterns of non-Ig genes with Ig gene usage in individual B cells.

The high-throughput screening of B-cell responses may also be used for the longitudinal analysis of repertoire changes over time or to identify protective mAbs against a variety of pathogens as previously shown for HIV, *Plasmodium falciparum*, influenza, and dengue virus [5, 9, 11, 24, 25].

Material and methods

Mice and tissue preparation

Female C57BL/6J mice, age 13–14 weeks, were killed by cervical dislocation. Peritoneal cavity cells were obtained by lavage with 10 mL of PBS + 2% FCS. Bone marrow cells were harvested by briefly grinding both femurs plus 4 mL of PBS + 2% FCS in a mortar and filtering the resulting suspension through a 70 μm sieve. Splenocytes were isolated by passing the whole spleen through a 70 μm cell strainer, followed by erythrocyte lysis using ACT buffer.

Flow cytometric single-cell sorting and liquid handling

Peritoneal cavity, bone marrow, and spleen single-cell suspensions were stained with fluorescently labeled antibodies according to standard methods. A complete list of all reagents and the detailed staining panels are given in (Supporting Information Table 3). Single-cell sorting was performed on an FACSAria II instrument (BD Biosciences, Heidelberg, Germany) using single-cell sort masks and FSC/SSC-based doublet discrimination. Flow cytometry data were evaluated using R 2.14 and the Bioconductor flowCore package [26, 27]. An evaluation copy of the FACS DiVa software version 6.1.3 with enabled index sorting was kindly provided by BD Biosciences, San Jose, CA, USA. Cells were sorted into 384-well PCR plates (4titude, Berlin, Germany) containing 2 μL of sort/RHP-Mix per well (final concentrations: 0.25× PBS; 5 mM DTT; 0.6875% NP-40 v/v; 20 ng/μL random hexamer primers (Roche); 1.875 U/μL RNAsin (Promega, Karlsruhe, Germany)). Plates were sealed and stored at −80°C directly after sorting.

Tag design

The initial set of $\sim 1.68 \times 10^7$ 16 bp tags was generated by concatenating eight dinucleotide blocks of the IUPAC-sequence SW or WS with stable GC content and lack of homopolymers of ≥3 consecutive identical bps. Next, the complete set of tags was filtered for a global pairwise Hamming distance of 6 bps [18]. The remaining 4096 tags were further filtered according to the following criteria: 1. No G at the 5′ terminus to avoid the formation of GG-homodimer with the default 454 GS FLX+ amplicon key (TCAG). 2. At maximum three homodimers. 3. Nonoverlapping sense/antisense regions of at least 4 bps each. 4. Lack of similarity to Ig germline segments, defined as full match that requires at maximum two insertions, deletions, or replacement mutations (note that this filter was only used for the second 240 × 192 matrix). To ensure a homogeneous sequencing behavior of the tags, the resulting list was finally ordered according to sequencing cost (defined as number of individual reagent flows on the pyrosequencing system) and tags requiring 22–25 flows were selected. Since each tag can either be located at the proximal position (as original sequence) or the distal position (as reverse complement sequence) of a read, it is important to note that the flow order of the system results in symmetric behavior of sequencing costs (i.e. $flow_{original} = flow_{revcomp}$ versus a complementary behavior $flow_{original} + flow_{revcomp} = constant$). Of the remaining tags, 72 and 432 were randomly selected and incorporated into the respective primer matrix. All tags are listed in (Supporting Information Table 2). Filtering was performed using Perl scripts, which are available upon request.

Primer matrix

Primers were synthesized at 25 nmol scale and purified via reverse-phase cartridges (Sigma-Aldrich, Taufkirchen, Germany). All primers are given in (Supporting Information Table 2). Of note, sets of three and two different primers with identical tags were used at equimolar ratio for the amplification of *Igh*-C to cover all different isotypes (*Ighm*, *Ighg1–3*, and *Igha*) and for *Igl*-V,

© 2013 WILEY-VCH Verlag GmbH & Co. KGaA, Weinheim

respectively. Single promiscuous primers were used for all other loci (*Igh*-V, *Igk*-V, *Igk*-C, and *Igl*-C).

Single-cell PCRs

scPCRs were performed in a semi-nested fashion following previously described protocols, with one V primer or primer set and two nested constant region primers or primer sets [3]. Primer sequences are listed in (Supporting Information Table 1). All PCRs were scaled down to 384-well plate format and adapted to allow for automation using a Tecan Freedom EVO 200 liquid handling platform. In brief, sort plates were thawed on ice and subjected to a heat step (68°C/60 s). Then 2 μL RT-Mix were added (final concentrations: 2× first-strand buffer; 15 mM DTT; 1.7 mM of each dNTP; 1.125 U/μL RNAsin and 3.5 U/μL SuperScript III). Reverse transcription was performed at 42°C, 5 min; 25°C, 10 min; 50°C, 1 h; 94°C, 5 min. cDNA was diluted with 7 μL nuclease-free water and 3 μL were used as template in each of the three primary PCRs (*Igh*, *Igk*, and *Igl*). Primary nested PCRs were performed in a reaction volume of 10 μL, with the following final concentrations: 1× HotStar PCR buffer; 162.5 nM forward primer; 162.5 nM reverse primer; 250 μM of each dNTP; 25 mU/μL HotStar Taq (Qiagen, Hilden, Germany). Nested PCRs were set up using 1 μL of primary PCR product and tagged primers. All other reagents were used in volumes and concentrations identical to the primary PCR.

454 sequencing

For sequencing, pooled products from secondary *Igh*, *Igk*, and *Igl* PCRs (1 μL from each individual reaction), respectively, were first concentrated using spin columns (Machery–Nagel). Then amplicon DNA of the expected size range (500–600 bp for *Igh* and 375–425 bp for *Igk/Igl*) was purified from 4% agarose gels. The purified (*Igh*:*Igk*:*Igl*) amplicon pools were mixed at 10:10:1 molar ratio and sequenced on a 454 GS FLX+ system using Titanium plus chemistry and eightfold gaskets (24 × 48 matrix 1) or spiked into shotgun sequencing reactions (240 × 192 matrix 2).

Sequence analysis pipeline

For the analysis of the raw 454 sequence data, we created a bioinformatics pipeline to rapidly generate a database that contains high-fidelity sequences for each successfully amplified locus of each sorted B cell and some basic descriptions of each sequence such as CDR3 length, mutation status, and V(D)J usage. As a first step, each read was analyzed with RazerS to map the respective tag position [28] and with NCBI's stand-alone version of IgBLAST to define the V(D)J segments of the Ig gene cargo sequence (http://www.ncbi.nlm.nih.gov/igblast/). All raw reads were divided into high and low quality reads. "Low quality" referred to reads for which RazerS failed to determine both the proximal tag and the distal tag with at most two errors (mismatches, insertions, deletions, and combinations thereof) and/or reads for which IgBLAST failed to assign germline V and J segments. As additional quality check reads that mapped the distal tag to the actual Ig gene sequence rather than the flanking region (before bp 420 for *Igh* or bp 370 for *Igk/Igl*) were marked as low quality. All low-quality reads were discarded and the remaining high-quality reads were sorted according to their well of origin (defined by the respective tag combination and *Igh*, *Igk*, or *Igl*-V segment).

To detect and compensate for sequencing errors, all reads were further sorted according to their V:J combination and aligned using the multiple alignment program MUSCLE [29]. Based on the alignment, a consensus sequence was built by sequentially retaining the most prominent nucleotide at each bp position in the multiple alignment and deleting the position if the "most prominent nucleotide" was a gap.

The resulting error-corrected consensus sequences for the *Igh* and *Igk/Igl* gene of each B cell were analyzed by a combination of IgBLAST, the ordinary nucleotide BLAST, and our own programs to determine V(D)J usage. These programs also allow for the analysis of FWR/CDR and somatic hypermutations. Isotype subclass identification was based on the > = 80 bp nonprimer templated constant region present in each *Igh* amplicon and could reliably distinguish the related *Ighg2a/b* subclasses (91.4% id). Last, *Igh* and *Igk/Igl* sequence data from each individual cell were combined with the respective FACS data and fed into a MySQL database to ensure for user-friendly accessibility. The pipeline is written in (Bio)Perl and (together with all used parameters) is available upon request.

Acknowledgements: The authors thank Bernd Timmermann and Sven Klages (Max Planck Institute for Molecular Genetics, Berlin) for 454 sequencing; Toralf Kaiser (Max Planck Institute for Infection Biology, Berlin) for expert cell sorting, BD Biosciences (Heidelberg, Germany) for making available the index sorting enabled FACS DiVa software ahead of release and Jan-David Manntz (Max Planck Institute for Infection Biology, Berlin) for assistance with the liquid handling platform. The work was supported by the Max Planck Society Strategic Innovations Fund to H.W.

Conflict of interest: The authors declare no financial or commercial conflict of interest.

References

1 **Kantor, A. B., Merrill, C. E., MacKenzie, J. D., Herzenberg, L. A. and Hillson, J. L.**, Development of the antibody repertoire as revealed by single-cell PCR of FACS-sorted B-cell subsets. *Ann. NY Acad. Sci.* 1995. **764**: 224–227.

© 2013 WILEY-VCH Verlag GmbH & Co. KGaA, Weinheim

2 **Ghia, P., ten Boekel, E., Sanz, E., de la Hera, A., Rolink, A. and Melchers, F.,** Ordering of human bone marrow B lymphocyte precursors by single-cell polymerase chain reaction analyses of the rearrangement status of the immunoglobulin H and L chain gene loci. *J. Exp. Med.* 1996. **184**: 2217–2229.

3 **Tiller, T., Busse, C. E. and Wardemann, H.,** Cloning and expression of murine Ig genes from single B cells. *J. Immunol. Methods* 2009. **350**: 183–193.

4 **Tiller, T., Meffre, E., Yurasov, S., Tsuiji, M., Nussenzweig, M. C. and Wardemann, H.,** Efficient generation of monoclonal antibodies from single human B cells by single cell RT-PCR and expression vector cloning. *J. Immunol. Methods* 2008. **329**: 112–124.

5 **Scheid, J. F., Mouquet, H., Feldhahn, N., Seaman, M. S., Velinzon, K., Pietzsch, J., Ott, R. G. et al.,** Broad diversity of neutralizing antibodies isolated from memory B cells in HIV-infected individuals. *Nature* 2009. **458**: 636–640.

6 **Brezinschek, H. P., Foster, S. J., Dorner, T., Brezinschek, R. I. and Lipsky, P. E.,** Pairing of variable heavy and variable kappa chains in individual naive and memory B cells. *J. Immunol.* 1998. **160**: 4762–4767.

7 **Dorner, T., Farner, N. L. and Lipsky, P. E.,** Ig lambda and heavy chain gene usage in early untreated systemic lupus erythematosus suggests intensive B cell stimulation. *J. Immunol.* 1999. **163**: 1027–1036.

8 **Yurasov, S., Hammersen, J., Tiller, T., Tsuiji, M. and Wardemann, H.,** B-cell tolerance checkpoints in healthy humans and patients with systemic lupus erythematosus. *Ann. NY Acad. Sci.* 2005. **1062**: 165–174.

9 **Wrammert, J., Smith, K., Miller, J., Langley, W. A., Kokko, K., Larsen, C., Zheng, N. Y. et al.,** Rapid cloning of high-affinity human monoclonal antibodies against influenza virus. *Nature* 2008. **453**: 667–671.

10 **Tiller, T., Kofer, J., Kreschel, C., Busse, C. E., Riebel, S., Wickert, S., Oden, F. et al.,** Development of self-reactive germinal center B cells and plasma cells in autoimmune Fc gammaRIIB-deficient mice. *J. Exp. Med.* 2010. **207**: 2767–2778.

11 **Muellenbeck, M. F., Ueberheide, B., Amulic, B., Epp, A., Fenyo, D., Busse, C. E., Esen, M. et al.,** Atypical and classical memory B cells produce *Plasmodium falciparum* neutralizing antibodies. *J. Exp. Med.* 2013. **210**: 389–399.

12 **Wu, Y. C., Kipling, D., Leong, H. S., Martin, V., Ademokun, A. A. and Dunn-Walters, D. K.,** High-throughput immunoglobulin repertoire analysis distinguishes between human IgM memory and switched memory B-cell populations. *Blood* 2010. **116**: 1070–1078.

13 **Weinstein, J. A., Jiang, N., White, R. A., 3rd, Fisher, D. S. and Quake, S. R.,** High-throughput sequencing of the zebrafish antibody repertoire. *Science* 2009. **324**: 807–810.

14 **Benichou, J., Ben-Hamo, R., Louzoun, Y. and Efroni, S.,** Rep-Seq: uncovering the immunological repertoire through next-generation sequencing. *Immunology* 2012. **135**: 183–191.

15 **Lindner, C., Wahl, B., Fohse, L., Suerbaum, S., Macpherson, A. J., Prinz, I. and Pabst, O.,** Age, microbiota, and T cells shape diverse individual IgA repertoires in the intestine. *J. Exp. Med.* 2012. **209**: 365–377.

16 **Zhu, J., Ofek, G., Yang, Y., Zhang, B., Louder, M. K., Lu, G., McKee, K. et al.,** Mining the antibodyome for HIV-1-neutralizing antibodies with next-generation sequencing and phylogenetic pairing of heavy/light chains. *Proc. Natl. Acad. Sci. USA* 2013. **110**: 6470–6475.

17 **DeKosky, B. J., Ippolito, G. C., Deschner, R. P., Lavinder, J. J., Wine, Y., Rawlings, B. M., Varadarajan, N. et al.,** High-throughput sequencing of the paired human immunoglobulin heavy and light chain repertoire. *Nat. Biotechnol.* 2013. **31**: 166–169.

18 **Hamady, M., Walker, J. J., Harris, J. K., Gold, N. J. and Knight, R.,** Error-correcting barcoded primers for pyrosequencing hundreds of samples in multiplex. *Nat. Methods* 2008. **5**: 235–237.

19 **Hardy, R. R., Carmack, C. E., Shinton, S. A., Riblet, R. J. and Hayakawa, K.,** A single VH gene is utilized predominantly in anti-BrMRBC hybridomas derived from purified Ly-1 B cells. Definition of the VH11 family. *J. Immunol.* 1989. **142**: 3643–3651.

20 **Pennell, C. A., Mercolino, T. J., Grdina, T. A., Arnold, L. W., Haughton, G. and Clarke, S. H.,** Biased immunoglobulin variable region gene expression by Ly-1 B cells due to clonal selection. *Eur. J. Immunol.* 1989. **19**: 1289–1295.

21 **Kantor, A. B., Merrill, C. E., Herzenberg, L. A. and Hillson, J. L.,** An unbiased analysis of V(H)-D-J(H) sequences from B-1a, B-1b, and conventional B cells. *J. Immunol.* 1997. **158**: 1175–1186.

22 **Cassani, B., Poliani, P. L., Marrella, V., Schena, F., Sauer, A. V., Ravanini, M., Strina, D. et al.,** Homeostatic expansion of autoreactive immunoglobulin-secreting cells in the Rag2 mouse model of Omenn syndrome. *J. Exp. Med.* 2010. **207**: 1525–1540.

23 **Quail, M. A., Smith, M., Coupland, P., Otto, T. D., Harris, S. R., Connor, T. R., Bertoni, A. et al.,** A tale of three next generation sequencing platforms: comparison of Ion Torrent, Pacific Biosciences and Illumina MiSeq sequencers. *BMC Genom.* 2012. **13**: 341.

24 **Scheid, J., Mouquet, H., Feldhahn, N., Walker, B., Pereyra, F., Cutrell, E., Seaman, M. et al.,** A method for identification of HIV gp140 binding memory B cells in human blood. *J. Immunol. Methods* 2009. **343**: 65–67.

25 **Wrammert, J., Onlamoon, N., Akondy, R. S., Perng, G. C., Polsrila, K., Chandele, A., Kwissa, M. et al.,** Rapid and massive virus-specific plasmablast responses during acute dengue virus infection in humans. *J. Virol.* 2012. **86**: 2911–2918.

26 **Hahne, F., LeMeur, N., Brinkman, R. R., Ellis, B., Haaland, P., Sarkar, D., Spidlen, J. et al.,** Flowcore: a Bioconductor package for high throughput flow cytometry. *BMC Bioinform.* 2009. **10**: 106.

27 **Gentleman, R. C., Carey, V. J., Bates, D. M., Bolstad, B., Dettling, M., Dudoit, S., Ellis, B. et al.,** Bioconductor: open software development for computational biology and bioinformatics. *Genome Biol.* 2004. **5**: R80.

28 **Weese, D., Emde, A. K., Rausch, T., Doring, A. and Reinert, K.,** RazerS–fast read mapping with sensitivity control. *Genome Res.* 2009. **19**: 1646–1654.

29 **Edgar, R. C.,** MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* 2004. **32**: 1792–1797.

Abbreviations: **Igh**: Ig heavy · **NGS**: next-generation sequencing · **scPCR**: single-cell PCR

Full correspondence: Hedda Wardemann, Max Planck Institute for Infection Biology, Research Group Molecular Immunology Charitéplatz 1, 10117 Berlin, Germany
Fax: +49-30-28460-363
e-mail: wardemann@mpiib-berlin.mpg.de

Additional correspondence: Dr. Christian Busse, Max Planck Institute for Infection Biology, Research Group Molecular Immunology Charitéplatz 1, 10117 Berlin, Germany.
e-mail: busse@mpiib-berlin.mpg.de

Received: 21/7/2013
Revised: 27/8/2013
Accepted: 19/9/2013
Accepted article online: 24/9/2013

© 2013 WILEY-VCH Verlag GmbH & Co. KGaA, Weinheim