Published in final edited form as:
Nat Biotechnol. 2013 February ; 31(2): 166–169. doi:10.1038/nbt.2492.

High-throughput sequencing of the paired human immunoglobulin heavy and light chain repertoire

Brandon J DeKosky[1], **Gregory C Ippolito**[2], **Ryan P Deschner**[1], **Jason J Lavinder**[3], **Yariv Wine**[1], **Brandon M Rawlings**[1], **Navin Varadarajan**[4], **Claudia Giesecke**[5,6], **Thomas Dörner**[5,6], **Sarah F Andrews**[7], **Patrick C Wilson**[7], **Scott P Hunicke-Smith**[3], **C Grant Willson**[1,8], **Andrew D Ellington**[3,8], and **George Georgiou**[1,2,3,9]

[1]Department of Chemical Engineering, University of Texas at Austin, Austin, Texas, USA

[2]Section of Molecular Genetics and Microbiology, University of Texas at Austin, Austin, Texas, USA

[3]Institute for Cellular and Molecular Biology, University of Texas at Austin, Austin, Texas, USA

[4]Department of Chemical & Biomolecular Engineering, University of Houston, Houston, Texas, USA

[5]Department of Medicine, Rheumatology, and Clinical Immunology, Charité University Medicine, Berlin, Germany

[6]German Rheumatism Research Center (DRFZ), Berlin, Germany

[7]Section of Rheumatology, Department of Medicine, University of Chicago, Chicago, Illinois, USA

[8]Department of Chemistry & Biochemistry, University of Texas at Austin, Austin, Texas, USA

[9]Department of Biomedical Engineering, University of Texas at Austin, Austin, Texas, USA

Abstract

Each B-cell receptor consists of a pair of heavy and light chains. High-throughput sequencing can identify large numbers of heavy- and light-chain variable regions (V_H and V_L) in a given B-cell repertoire, but information about endogenous pairing of heavy and light chains is lost after bulk lysis of B-cell populations. Here we describe a way to retain this pairing information. In our approach, single B cells (>5 × 10^4 capacity per experiment) are deposited in a high-density microwell plate (125 pl/well) and lysed *in situ*. mRNA is then captured on magnetic beads, reverse transcribed and amplified by emulsion V_H:V_L linkage PCR. The linked transcripts are analyzed by Illumina high-throughput sequencing. We validated the fidelity of V_H:V_L pairs identified by this approach and used the method to sequence the repertoire of three human cell subsets—peripheral blood IgG^+ B cells, peripheral plasmablasts isolated after tetanus toxoid immunization and memory B cells isolated after seasonal influenza vaccination.

Correspondence should be addressed to G.G. (gg@che.utexas.edu).

Accession code. Sequence data, SRA: SRA061316.

Note: Supplementary information is available in the online version of the paper.

Author Contributions: B.J.D. and G.G. developed the methodology and designed the experiments. B.J.D., G.C.I. and G.G. wrote the manuscript; B.J.D., G.C.I., R.P.D., J.J.L., Y.W., B.M.R., C.G. and S.F.A. performed the experiments; B.J.D. carried out the bioinformatic analysis; S.P.H.-S. performed Illumina sequencing; G.C.I., N.V., T.D., P.C.W., C.G.W. and A.D.E. helped design experiments; B.J.D., G.C.I., J.J.L., Y.W., S.P.H.-S., A.D.E. and G.G. analyzed the data.

Competing Financial Interests: The authors declare competing financial interests: details are available in the online version of the paper.

Next-generation sequencing of immunoglobulin variable region and T-cell receptor repertoires is providing information that is key to understanding adaptive immune responses and to diagnostic and therapeutic applications[1–6]. However, existing immune repertoire sequencing technologies yield data on only one of the two chains of immune receptors and thus cannot provide information about the identity of immune receptor pairs encoded by individual B or T lymphocytes[2,7,8]. Sequence analysis of V_H:V_L pairs is currently done by microtiterwell sorting of individual B cells followed by single-cell RT-PCR (scRT-PCR) and Sanger sequencing[8–15]; however, at most a few hundred V_H:V_L pairs (a number dwarfed by the enormous size of the human antibody repertoire) are identified by means of scRT-PCR[11–14]. Microfluidic methods for RT-PCR and the sequencing of two or more genes (e.g., using the Fluidigm platform[16]) have so far been limited to only 96 wells per run and require complex, proprietary instrumentation. As a result, comprehensive analysis of paired V_H:V_L gene family usage and somatic hypermutation frequency has been elusive.

Here we report the development, validation and application of an accessible and scalable technology for the high-throughput sequencing of V_H:V_L pairs from individual human B lymphocytes. We used this technology to determine the V_H:V_L repertoires of several human B-cell subpopulations of particular interest, namely peripheral class-switched B cells in healthy individuals, memory B cells after influenza vaccination and antigen-specific plasmablasts after tetanus toxoid (TT) immunization. We applied this technology to discover high-affinity, antigen-specific human antibodies from the sequenced plasmablasts.

In this approach, a population of sorted B cells was deposited by gravity into 125-pl wells molded in polydimethylsiloxane (PDMS) slides (Fig. 1). Each slide contained 1.7×10^5 wells; four slides processed concurrently accommodated 68,000 lymphocytes at a ≥1:10 cell/well occupancy, which gave at least a 95% probability of there being only one cell per well based on Poisson statistics. Poly(dT) magnetic beads with a diameter of 2.8 μm were deposited into the microwells at an average of 55 beads/well and the slides were covered with a dialysis membrane. Subsequently, the membrane-covered slides were incubated with an optimized cell lysis solution containing 1% lithium dodecyl sulfate that resulted in complete cell lysis within <1 min (Supplementary Video 1). The mRNA annealed to the poly(dT) magnetic beads, which were then collected, washed and emulsified with primers, reverse transcriptase and thermostable DNA polymerase to carry out reverse transcription followed by linkage PCR (Supplementary Fig. 1). The two-step capture and amplification process (Fig. 1) was necessary because single-compartment cell lysis followed by RT-PCR has not proven feasible in volumes ≤5 nl because of inhibition of the reverse transcription reaction by cell lysate constituents. In addition, performing V_H: V_L linkage in emulsion droplets containing single cells would necessitate cell entrapment, lysis, reverse transcription and *in situ* –linkage PCR that can only be done in microfluidic devices, a strategy that requires extensive infrastructure and has been reported to have limited throughput at ≤300 cells per run[17]. PCR amplification generated an ∼850–base pair (bp) linked V_H:V_L DNA product composed of (from 5′ to 3′) the N-terminal end of CH_1, the V_H, a linker region, the V_L and the N terminus of C_κ or C_λ (Supplementary Fig. 1). The most informative 500 bp of this fragment, which encompassed the complementarity determining regions (CDR-H3 and CDR-L3), was then sequenced on a long-read next-generation sequencing platform such as the 2 × 250 Illumina MiSeq (which also provided the framework region FR3 and FR4 sequences and constant region N termini amino acid sequences that can be used for isotype assignment). If FR1 to CDR2 region sequences were also desired, the V_H and V_L gene repertoires were analyzed by separate 2 × 250-bp sequencing runs. This latter step was required because of read-length limitations with existing technology; whereas single-molecule sequencing techniques allow for longer reads, the error rate is too high to enable robust classification of V_H:V_L sequences.

We employed the methodology of Figure 1 to determine the V_H:V_L repertoire of three different B-cell populations of relevance to human immunology and antibody discovery. First, we isolated IgG^+ B cells from fresh blood donated by a healthy individual. We spiked 61,000 IgG^+ B cells with immortalized IM-9 lymphoblast cells (to ∼4% of total mixture) that express known V_H and V_L sequences as an internal control. We analyzed these cells in four PDMS slides (6.8×10^5 total wells). After 2 × 250 MiSeq sequencing, we clustered the CDR-H3 regions based on 96% sequence identity, consistent with the established error rate of the MiSeq platform, to determine the number of unique clones recovered from this human sample. A total of 2,716 unique pairs were thus identified (Supplementary Table 1). The spiked IM-9 heavy chain overwhelmingly (78-fold above background) paired with its known light chain. A heat map shows frequencies of pairing between V_H and V_L segments of different germline families in the class-switched IgG^+ cell repertoire (Fig. 2a). A second IgG^+ repertoire analysis was done using B cells from another anonymous individual; this analysis identified 2,248 unique CDR-H3 from 47,000 IgG^+ cells, and the IM-9 control spike again demonstrated high pairing accuracy (125-fold above background; Supplementary Fig. 2 and Supplementary Table 2). Several V gene families (e.g., IGHV7; IGKV5, 6, and 7; IGLV4, 10, and 11) are expressed at very low frequencies in the human immune repertoire[3,18]. We detected V_H:V_L pairs containing these rare families, indicating that this technique can identify rare B-cell clones present at physiological levels together with much more abundant clones (e.g., the much more highly used IGHV3 or IGHV4 families; Fig. 2a and Supplementary Fig. 2). Interestingly, the V_H:V_L germline pairing frequencies were highly correlated between the two individuals (Spearman rank correlation coefficient = 0.804; $P < 10^{-29}$); the most highly transcribed heavy chain genes (IGHV3, IGHV4 and IGHV1 families) paired most frequently with the most highly transcribed light chain genes (IGKV1, IGKV3, IGLV1 and IGLV2 families). However, putative differences in IgG^+ V_H:V_L germline pairing frequencies between the two individuals were also evident.

In a separate experiment, human plasmablasts ($CD19^+CD3^-$ $CD14^-CD38^{++}CD27^{++}CD20^-$) from a healthy volunteer were collected 7 d after TT immunization, sorted for surface antigen binding and then frozen[12]. After thawing, ∼400 recovered cells were spiked with the immortalized ARH-77 cell line as an internal control and seeded onto a single PDMS slide (1.7×10^5 total wells). In this instance, 86 unique primary CDR-H3:CDR-L3 pairs were identified, and the ARH-77 control spike demonstrated high pairing accuracy (Fig. 2b and Supplementary Table 1). We expressed ten of the identified V_H:V_L pairs as IgG proteins in HEK293K cells. As revealed by competitive enzyme-linked immunosorbent assay (ELISA), all ten antibodies showed specificity for TT and bound TT with high affinity (0.1 nM ≤ K_D ≤ 18 nM; Table 1; antibody sequences provided in Supplementary Data). Whereas certain V_H chains can pair promiscuously with multiple V_Ls to yield functional antibodies, it is statistically implausible that 10/10 antibodies could display nM and sub-nM affinities for TT merely as a consequence of fortuitous V_H: V_L pairing. For comparison, 10–15% of antibodies generated by random pairing of V_H genes with a small set of enriched V_L genes were antigen-specific[19,20].

Finally, we compared the V_H:V_L pairings identified using this high-throughput approach to those identified using the established single-cell sorting method[11,21]; this experiment was conducted in a double-blinded manner. Peripheral $CD19^+CD3^-CD27^+CD38^{int}$ memory B cells were isolated from a healthy volunteer 14 d after vaccination with the 2010-2011 trivalent FluVirin influenza vaccine[21]. For the scRT-PCR analysis, 168 single B cells were sorted into four 96-well plates, and 168 RT and 504 nested PCR reactions were carried out individually to separately amplify the V_H and V_L (κ and λ) genes. DNA products were resolved by gel electrophoresis and sequenced to yield a total of 51 V_H:V_L pairs, of which 50 were unique. A separate B-cell aliquot from the same individual was frozen at −80 °C and later thawed and processed using the new high-throughput approach described here.

Two PDMS slides (3.4×10^5 total wells) were used, and the sample was spiked with IM-9 cells to confirm pairing accuracy (Supplementary Table 1). A total of 240 unique CDR-H3:CDR-L3 pairs were recovered (Fig. 2c). Four CDR-H3 sequences detected in the high-throughput pairing set were also observed in the single-cell RT-PCR analysis. A blinded analysis revealed that CDR-H3:CDR-L3 pairs isolated by the two approaches were in complete agreement (Supplementary Table 3, sequences provided in Supplementary Data and repertoire comparisons in Supplementary Fig. 3). Further, the one V_H:V_L pair detected in more than one of the 51 cells analyzed by single-cell RT-PCR was also detected in the aliquot processed by the new high-throughput approach (clone 2D02 was observed in two cells by scRT-PCR, Supplementary Table 3); these findings suggest that this B-cell clone may have undergone a great deal of expansion. The 46 V_H genes that were each observed only once by single-cell RT-PCR but that were not detected in the aliquot processed by our high-throughput approach presumably represent unique or very low abundance B-cell clones, as expected given the great degree of V gene diversity normally found in human peripheral memory B cells.

In these experiments the control cell lines spiked into each aliquot of primary B cells were selected to approximate the levels of heavy chain and light chain transcription in that particular primary B-cell subpopulation. For example, the ARH-77 cell line expressed high levels of heavy chain and light chain transcripts and therefore was spiked into plasmablast populations that also express abundant heavy chain and light chain transcripts; in contrast, the IM-9 B lymphoblast cell line, which expresses lower levels of heavy chain and light chain transcripts, was spiked into memory B-cell populations. Known V_H and V_L sequences from spiked-in control cell lines were used to evaluate the frequency of non-native pairings, that is, the false discovery rate (FDR). The FDR, determined from the mispairing of spiked-in control V_H and V_L chains, was commensurate with the probability of coincident cells per well, which in turn is dictated by cell seeding density and follows Poisson statistics (Online Methods and Supplementary Table 4). The FDR revealed by the mispairing frequency of control cell lines represents the upper bound of the FDR, as the control cell lines were introduced at levels over tenfold higher than the levels at which even a very highly expanded B-cell clone might be present in a biological sample in humans. Although currently V_H:V_L pairing efficiency in memory B-cell populations is relatively modest (Fig. 2 and Supplementary Fig. 2), efforts to further improve efficiency are under way.

The high-throughput V_H:V_L pairing technique described here requires one emulsion RT-PCR reaction, followed by nested PCR, sequencing and bioinformatic analysis. The entire process from B-cell isolation to the generation of V_H:V_L heat maps can be completed by a single investigator in 10 research hours over the course of 4 d (which includes 3 d for gene sequencing). For example, the work required to recover 2,716 unique V_H:V_L pairs from a sample of IgG^+ peripheral B cells (Fig. 2a) was completed by a single researcher in 10 h and cost $550. Analysis of 2,700 cells using established optimal single-cell RT-PCR protocols would have required >10 weeks of effort by an experienced technician and >$25,000 in reagent and sequencing costs[11].

Because only sequences of up to 500 bp can be accurately determined with current Illumina next-generation sequencing technology, our method detects the different antibody clonotypes (antibodies comprising the same CDR-H3:CDR-L3) but cannot yet distinguish somatic variants originating from clonally related B cells that contain upstream mutations between FR1 and CDR2 regions. However, this method distinguished nearly identical but distinct CDR3 regions, as indicated by B-cell clones 2D02 and 3D05, which express light chain CDR3s that differ by only two nucleotides (Supplementary Table 3). Rapid advances in next-generation sequencing read-length and quality will likely enable upstream somatic variant analysis in the near future.

We used PCR amplification primers targeted to the FR1 region of heavy and light chains (primers reported in Supplementary Tables 5–7). In some chronic infections (e.g., HIV), constant antigen exposure can generate antibodies that are highly mutated in all regions including the FR1 region, and therefore amplification with FR1-specific primers can bias the repertoire. In these cases, such bias can be readily circumvented by using primers that anneal to the leader peptide[22].

Finally, we note that the analysis reported here focused on the light chain that was the dominant light chain paired with a particular V_H. There are known, albeit rare, instances in mice where one heavy chain can be found paired with more than one light chain[23]. Bioinformatic analysis might discriminate between biologically relevant V_H:V_L pairs of one heavy chain with multiple light chains and false pairings that might result from multiple cells seeded into the same well of the experimental device; false pairings can be flagged because coincident cells 1 and 2 would yield the products V_H1:V_L2 and V_H2:V_L1 in addition to V_H1:V_L1 and V_H2:V_L2.

This accessible technology for sequencing the paired V_H:V_L repertoire enables rapid interrogation of the immune response and can be applied to investigate B-cell maturation, vaccine efficacy, immune system health and autoimmunity in clinical and research settings. The high-throughput identification and cloning of paired V_H:V_L antigen-specific antibodies from responding B cells may also enable rapid generation of novel diagnostic, therapeutic or prophylactic antibodies.

Online Methods

Preparation of microwell slides

A grid of micropillars (56-μm diameter, 50-μm height) was photolithographically patterned onto a silica wafer using SU-8 photoresist (Fisher Scientific, Pittsburgh, PA) and the silica wafer was used as a mold to print PDMS slides (Sylgard 184, Dow Corning, Midland, MI) with the dimensions of a standard microscope slide and containing ∼170,000 wells each.

Molded PDMS slides were treated in an oxygen plasma chamber for 5 min to generate a hydrophilic surface. The PDMS slides were blocked in 1% BSA for 30 min and washed with deionized water followed by PBS to prepare for cell seeding.

Analysis of memory B-cell V_H:V_L pairings in response to seasonal influenza vaccination

The study was approved by the University of Chicago Institutional Review Board (IRB 09-043-A) and the University of Texas Institutional Review Board (IRB 2012-07-0002). A healthy 30-year-old male was vaccinated with the 2010-2011 trivalent FluVirin influenza vaccine (Novartis) and blood was drawn at day 14 after vaccination after informed consent had been obtained. PBMCs were isolated and resuspended in DMSO/10% FCS for cryopreservation.

Frozen PBMCs were subsequently thawed and cell suspensions were stained in PBS/0.2% BSA with anti-human CD19 (HIB19, BioLegend, San Diego, CA), CD27 (O323, BioLegend), CD38 (HIT2, BioLegend) and CD3 (7D6, Invitrogen, Grand Island, NY). $CD19^+CD3^-CD27^+CD38^{int}$ memory B cells were sorted using a FACSAria II sorter system (BD Biosciences, San Diego, CA). Cells were either cryopreserved in DMSO/10% FCS for subsequent high-throughput V_H: V_L pairing or single-cell sorted into 96-well plates containing RNAse Inhibitor Cocktail (Promega, Madison, WI) and 10 mM Tris-HCl pH 8.0 for single-cell RT-PCR analysis. cDNA was synthesized from single-sorted cells using the Maxima First Strand cDNA Synthesis Kit (Fermentas, Waltham, MA) followed by amplification of the immunoglobulin variable genes using primer sets and PCR conditions

previously described[11]. Variable genes were determined with in-house analysis software using the IMGT search engine[24].

Memory B cells frozen for high-throughput V_H:V_L pairing were thawed and recovered by centrifugation at 250*g* for 10 min. Cells were resuspended in 200 μl RPMI-1640 supplemented with 1× GlutaMAX, 1× non-essential amino acids, 1× sodium pyruvate and 1× penicillin/streptomycin (Life Technologies) and incubated at 37 °C for 13 h in a 96-well plate. Recovered cells were centrifuged again at 250*g* for 10 min and resuspended in 400 μl PBS, and 6 μl were withdrawn for cell counting with a hemocytometer. Approximately 8,800 cells were recovered from frozen stock. Memory B cells were then spiked with ~880 IM-9 cells (ATCC number CCL-159) as an internal control. Cells were resuspended over two PDMS microwell slides (340,000 wells) and allowed to settle into wells by gravity over the course of 5 min with gentle agitation. The cell seeding process has been calculated to be 90% efficient by measuring cell concentration in seeding buffers both pre- and post- cell seeding; thus 8,000 primary cells were analyzed in this experiment. The fraction of cells isolated in the single and multiple cell per well states was calculated using Poisson statistics

$$P(k,\mu)=\frac{\mu^k e^{-\mu}}{k!}$$

where *k* equals the number of cells in a single microwell and μ is the average number of cells per well, so that the 1:39 cell:well ratio used in this experiment corresponds to 98.7% of cells deposited at an occupancy of one cell/well. We resuspended 25 μl of poly(dT) magnetic beads (Invitrogen mRNA Direct Kit) in 50 μl PBS and distributed over each PDMS slide surface, (mean of 55 poly(dT) beads per well). Magnetic beads were allowed to settle into wells by gravity for ~5 min, then a BSA-blocked dialysis membrane (12,000–14,000 MWCO regenerated cellulose, 25-mm flat width, Fisher Scientific) that had been rinsed in PBS was laid over each slide surface, sealing the microwells and trapped cells and beads inside. Excess PBS was removed from the slide and membrane surfaces using a 200 μL pipette. 500 μL of cell lysis solution (500 mM LiCl in 100 mM TRIS buffer (pH 7.5) with 1% lithium dodecyl sulfate, 10 mM EDTA and 5 mM DTT) was applied to the dialysis membranes for 20 min at room temperature. Time-lapse microscopy revealed that all cells are fully lysed within 1 min (Supplementary Video 1). Subsequently the slides were incubated at 4 °C for 10 min at which point a Dynal MPC-S magnet was placed underneath the PDMS microwell device to hold magnetic beads inside the microwells as the dialysis membrane was removed with forceps and discarded. Working quickly, the PDMS slides were sequentially inverted in a Petri dish containing 2 mL of cold lysis solution and the magnet was applied underneath the Petri dish to force the beads out of the microwells. Subsequently, 1 ml aliquots of the lysis solution containing resuspended beads were placed into Eppendorf tubes and beads were pelleted on a Dynal MPC-S magnetic rack and washed once without resuspension using 1 mL per tube of wash buffer 1 (100 mM Tris, pH 7.5, 500 mM LiCl, 1 mM EDTA, 4 °C). Beads were resuspended in wash buffer 1, pelleted and resuspended in wash buffer 2 (20 mM Tris, pH 7.5, 50 mM KCl, 3 mM MgCl) and pelleted again. Finally beads were suspended in 2.85 mL cold RT-PCR mixture (Quanta OneStep Fast, VWR) containing 0.05 wt% BSA (Invitrogen Ultrapure BSA, 50 mg/mL) and primer sets for V_H and V_L linkage amplification (Supplementary Fig. 1 and Supplementary Table 5)[3,25]. The suspension containing the poly(dT) magnetic beads was added dropwise to a stirring IKA dispersing tube (DT-20, VWR) containing 9 mL chilled oil phase (molecular biology grade mineral oil with 4.5% Span-80, 0.4% Tween 80, 0.05% Triton X-100, v/v%, Sigma-Aldrich, St. Louis, MO), and the mixture was agitated for 5 min at low speed. The resulting emulsion was added to 96-well PCR plates with 100 μL emulsion per well and

placed in a thermocycler. The RT step was performed under the following conditions: 30 min at 55 °C, followed by 2 min at 94 °C. PCR amplification was performed under the following conditions: four cycles of 94 °C for 30 s denature, 50 °C for 30 s anneal, 72 °C for 2 min extend; four cycles of 94 °C for 30 s denature, 55 °C for 30 s anneal, 72 °C for 2 min extend; 22 cycles of 94 °C for 30 s denature, 60 °C for 30 s anneal, 72 °C for 2 min extend; then a final extension step for 7 min at 72 °C. After thermal cycling the emulsion was visually inspected to ensure the absence of a bulk water phase, which is a key indicator of emulsion stability. Following visual verification, the emulsion was collected and centrifuged at room temperature for 10 min at 16,000*g*, the mineral oil upper phase was discarded, and 1.5 mL diethyl ether was added to extract the remaining oil phase and break the emulsion. The upper ether layer was discarded, two more ether extractions were performed and residual ether was removed in a SpeedVac for 25 min at room temperature. The aqueous phase was diluted 5:1 in DNA binding buffer and passed through a silica spin column (DNA Clean & Concentrator, Zymo Research, Irvine, CA) to capture the cDNA product. The column was washed twice with 300 μL wash buffer (Zymo Research Corp) and cDNA was eluted into 40 μL nuclease-free water. Finally, a nested PCR amplification was performed (ThermoPol PCR buffer with Taq Polymerase, New England Biosciences, Ipswich, MA) in a total volume of 200 μL using 4 μL of eluted cDNA as template with 400 nM primers (Supplementary Table 6) under the following conditions: 2 min initial denaturation at 94 °C, denaturation at 94 °C for 30 s for 39 cycles, annealing at 62 °C for 30 s and extension at 72 °C for 20 s, final extension at 72 °C for 7 min. The approximately 850 bp linked product was extracted by agarose gel electrophoresis and sequenced using the 2 × 250 paired end MiSeq NextGen platform (Illumina, San Diego, CA).

Analysis of plasmablast V_H:V_L pairings in response to TT vaccination

One female donor underwent booster immunization against TT/diphtheria toxoid (TD, 20 I.E. TT and 2 I.E. diphtheria toxoid, Sanofi Pasteur Merck Sharpe & Dohme GmbH, Leimen, Germany) after informed consent by the Charité Universitätsmedizin Berlin had been obtained (samples were anonymously coded and study approved by the hospital's ethical approval board, number EA1/178/11, and the University of Texas at Austin Institutional Review Board, IRB# 2011-11-0095). At 7 d post TT immunization, EDTA blood was withdrawn and PBMC isolated by density gradient separation as described[26]. PBMCs were stained in PBS/BSA at 4 °C for 15 min with anti-human CD3/CD14-PacB (clones UCHT1 and M5E2, respectively, Becton Dickinson, BD), CD19-PECy7 (clone SJ25C1, BD), CD27-Cy5 (clone 2E4, kind gift from René van Lier, Academic Medical Centre, University of Amsterdam, The Netherlands, labeled at the Deutsches Rheumaforschungszentrum (DRFZ), Berlin), CD20-PacO (clone HI47, Invitrogen), IgD-PerCpCy5.5 (clone L27, BD), CD38-PE (clone HIT2, BD) and TT-Digoxigenin (labeled at the DRFZ) for 15 min at 4 °C. Cells were washed and a second staining was performed with anti-Digoxigenin-FITC (Roche, labeled at the DRFZ) and DAPI was added before sorting. $CD19^+CD3^-CD14^-CD38^{++}CD27^{++}CD20^-TT^+$ plasmablasts were sorted using a FACSAria II sorter system (BD Biosciences). A portion of sorted cells were washed and cryopreserved in DMSO/10%FCS for high-throughput V_H:V_L pairing.

One vial containing approximately 2,000 frozen TT^+ plasmablasts was thawed and recovered by centrifugation at 250*g* for 10 min, for which 20–30% recovery was anticipated[27]. Cells were resuspended in 300 μL RPMI-1640 supplemented with 10% FBS, 1× GlutaMAX, 1× non-essential amino acids, 1× sodium pyruvate and 1× penicillin/streptomycin (all from Life Technologies) and incubated at 37 °C for 13 h in a 96-well plate. Recovered cells were centrifuged again at 250*g* for 10 min and resuspended in 400 μL PBS, and 6 μL were withdrawn for cell counting with a hemocytometer. Cells were spiked with approximately 30 ARH-77 cells as an internal control (ATCC number CRL-1621) and

V_H:V_L transcripts were linked as described above, omitting IgM primers and using a 38-cycle nested PCR; the resulting product was submitted for 2 × 250 MiSeq sequencing. V_H and V_L chains were also amplified individually to obtain full V_H and V_L sequences for antibody expression. Nested PCR product was diluted 1:9 and 0.5 μL were used as template in a PCR reaction with the following conditions: 400 nM primers (Supplementary Table 7), 2 min initial denaturation at 94 °C, denaturation at 94 °C for 30 s for 12 cycles, annealing at 62 °C for 30 s and extension at 72 °C for 15 s, final extension at 72 °C for 7 min. The resulting ~450 bp V_H or ~400 bp V_L products were purified by agarose gel electrophoresis and submitted for 2 × 250 MiSeq sequencing.

Ten V_H and V_L pairs (Supplementary Data, .fna file) were selected from TT^+ plasmablast pairings and cloned into the human IgG expression vectors pMAZ-VH and pMAZ-VL, respectively[28]. 40 μg each of circularized ligation product were co-transfected into HEK293F cells (Invitrogen, NY, USA). Medium was harvested 6 d after transfection by centrifugation and IgG was purified by a protein-A agarose (Pierce, IL, USA) chromatography column.

Antigen affinities were determined by competitive ELISA[29] using different concentrations of IgG in a serial dilution of antigen, ranging from 100 nM to 0.05 nM in the presence of 1% milk in PBS. Plates were coated overnight at 4 °C with 10 μg/mL of TT in 50 mM carbonate buffer, pH 9.6, washed three times in PBST (PBS with 0.1% Tween 20) and blocked with 2% milk in PBS for 2 h at room temperature. Pre-equilibrated samples of IgG with TT antigen were added to the blocked ELISA plate, incubated for 1 h at room temperature, and plates were washed 3× with PBST and incubated with 50 μl of anti-human kappa light chain-HRP secondary antibody (1:5,000, 2% milk in PBS) for ~2 min, 25 °C. Plates were washed 3× with PBST, then 50 μl Ultra TMB substrate (Thermo Scientific, Rockford, IL) was added to each well and incubated at 25 °C for 5 min. Reactions were stopped using equal volume of 1M H_2SO_4 and absorbance was read at 450 nm (BioTek, Winooski, VT). Each competitive ELISA replicate was fit using a four-parameter logistic (4PL) equation, with error represented as the s.d. of 2–3 replicates for each IgG analyzed.

Analysis of IgG^+ class switched cells in healthy donor peripheral blood

PBMC from two different anonymous healthy donors (Texas Gulf Coast Regional Blood Center) were isolated from whole blood using Histopaque-1077 centrifugation gradient (Sigma-Aldrich) according to manufacturer protocols. 3.8×10^8 PBMCs were used as input for cell purification using a IgG^+ Memory B Cell Isolation Kit (Miltenyi Biotec, Auburn, CA). Approximately 68,000 cells (Fig. 2a) or 52,000 cells (Supplementary Fig. 2) were spiked with 4% IM-9 immortalized B lymphoblast cells (ATCC number CCL-159) and seeded onto four or three PMS microwell slides, respectively, at a 1:10 cell/well ratio corresponding to 95.1% of cells isolated as single cells by Poisson distribution. Cell seeding was calculated to be 90% efficient and thus 61,000 cells (Fig. 2a and Supplementary Table 1) or 47,000 cells (Supplementary Fig. 2 and Supplementary Table 2) were analyzed. V_H:V_L chains were paired and amplified as above using 35 nested PCR cycles and sequenced on the Illumina MiSeq platform.

Bioinformatic methods

Raw 2 × 250 MiSeq data were filtered for minimum Phred quality score of 20 over 50% of nucleotides to ensure high read quality in the CDR3-containing region (approximately HC nt 65-115 or LC nt 55-100). Sequence data were submitted to the International ImMunoGeneTics Information System (IMGT) for mapping to germline V(D)J genes[30]. Sequence data were filtered for in-frame V(D)J junctions, and productive V_H and $V_{\kappa,\lambda}$ sequences were paired by Illumina read ID. CDR-H3 nucleotide sequences were extracted

and clustered to 96% nt identity with terminal gaps ignored, to generate a list of unique CDR-H3s in the data set. 96% nt identity cutoff was found to be the optimal cutoff to cluster sequencing error in spiked control clones; the number of unique CDR-H3 sequences and hence the number of unique V genes reported refer to the number of clusters recovered from the sample (Supplementary Table 1 and Supplementary Table 2). Overwhelmingly, the frequency of the top V_L pair for a given V_h was observed to be greater than 90%, and the top V_L pair for each CDR-H3 was used for comparison with single-cell RT-PCR for workflow validation (Supplementary Table 3), consensus sequence generation and paired V_H:V_L expression (Table 1), and V_H:V_L gene family heat maps (Fig. 2 and Supplementary Figs. 2 and 3). In the case of separate V_H and V_L gene amplification for complete antibody sequencing, 2 × 250 bp reads containing the 5′ V gene FR1-CDR2 and 3′ CDR2-FR4 were paired by Illumina read ID and consensus sequences were constructed from reads containing the exact CDR3 of interest.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We thank B. Iverson and M. Pogson for insightful discussions and thoughtful advice, C. Das for assistance with antibody expression, and S. Reddy for help with initial studies. This work was funded by fellowships to B.J.D. from the Hertz Foundation, the University of Texas Donald D. Harrington Foundation and the National Science Foundation, and also by the US National Institutes of Health U19AI057234-09 (P.C.W.), U19AI057234-09 (G.G.), and U54AI057156 (G.G. and A.E.D.).

References

1. Reddy ST, et al. Monoclonal antibodies isolated without screening by analyzing the variable-gene repertoire of plasma cells. Nat Biotechnol. 2010; 28:965–969. [PubMed: 20802495]
2. Wu X, et al. Focused evolution of HIV-1 neutralizing antibodies revealed by structures and deep sequencing. Science. 2011; 333:1593–1602. [PubMed: 21835983]
3. Ippolito GC, et al. Antibody repertoires in humanized NOD-scid-IL2R gamma(null) mice and human B cells reveals human-like diversification and tolerance checkpoints in the mouse. PLoS ONE. 2012; 7:e35497. [PubMed: 22558161]
4. Reddy ST, Georgiou G. Systems analysis of adaptive immunity by utilization of high-throughput technologies. Curr Opin Biotechnol. 2011; 22:584–589. [PubMed: 21570821]
5. Weinstein JA, Jiang N, White RA, Fisher DS, Quake SR. High-throughput sequencing of the zebrafish antibody repertoire. Science. 2009; 324:807–810. [PubMed: 19423829]
6. Benichou J, Ben-Hamo R, Louzoun Y, Efroni S. Rep-Seq: uncovering the immunological repertoire through next-generation sequencing. Immunology. 2012; 135:183–191. [PubMed: 22043864]
7. Fischer N. Sequencing antibody repertoires: the next generation. MAbs. 2011; 3:17–20. [PubMed: 21099370]
8. Wilson PC, Andrews SF. Tools to therapeutically harness the human antibody response. Nat Rev Immunol. 2012; 12:709–719. [PubMed: 23007571]
9. Wardemann H, et al. Predominant autoantibody production by early human B cell precursors. Science. 2003; 301:1374–1377. [PubMed: 12920303]
10. Meijer P, et al. Isolation of human antibody repertoires with preservation of the natural heavy and light chain pairing. J Mol Biol. 2006; 358:764–772. [PubMed: 16563430]
11. Smith K, et al. Rapid generation of fully human monoclonal antibodies specific to a vaccinating antigen. Nat Protoc. 2009; 4:372–384. [PubMed: 19247287]
12. Frölich D, et al. Secondary immunization generates clonally related antigen-specific plasma cells and memory B cells. J Immunol. 2010; 185:3103–3110. [PubMed: 20693426]

13. Tanaka Y, et al. Single-cell analysis of T-cell receptor repertoire of HTLV-1 tax-specific cytotoxic T cells in allogeneic transplant recipients with adult T-cell leukemia/lymphoma. Cancer Res. 2010; 70:6181–6192. [PubMed: 20647322]
14. Scheid JF, et al. Differential regulation of self-reactivity discriminates between IgG(+) human circulating memory B cells and bone marrow plasma cells. Proc Natl Acad Sci USA. 2011; 108:18044–18048. [PubMed: 22025722]
15. Li GM, et al. Pandemic H1N1 influenza vaccine induces a recall response in humans that favors broadly cross-reactive memory B cells. Proc Natl Acad Sci USA. 2012; 109:9047–9052. [PubMed: 22615367]
16. Sanchez-Freire V, Ebert AD, Kalisky T, Quake SR, Wu JC. Microfluidic single-cell real-time PCR for comparative analysis of gene expression patterns. Nat Protoc. 2012; 7:829–838. [PubMed: 22481529]
17. White AK, et al. High-throughput Microfluidic single-cell RT-qPCR. Proc Natl Acad Sci USA. 2011; 108:13999–14004. [PubMed: 21808033]
18. Glanville J, et al. Naive antibody gene-segment frequencies are heritable and unaltered by chronic lymphocyte ablation. Proc Natl Acad Sci USA. 2011; 108:20066–20071. [PubMed: 22123975]
19. Cheung WC, et al. A proteomics approach for the identification and cloning of monoclonal antibodies from serum. Nat Biotechnol. 2012; 30:447–452. [PubMed: 22446692]
20. Sato S, et al. Proteomics-directed cloning of circulating antiviral human monoclonal antibodies. Nat Biotechnol. 2012; 30:1039–1043. [PubMed: 23138294]
21. Wrammert J, et al. Rapid cloning of high-affinity human monoclonal antibodies against influenza virus. Nature. 2008; 453:667–671. [PubMed: 18449194]
22. Scheid JF, et al. Sequence and structural convergence of broad and potent HIV antibodies that mimic CD4 binding. Science. 2011; 333:1633–1637. [PubMed: 21764753]
23. Seidl KJ, et al. Frequent occurrence of identical heavy and light chain Ig rearrangements. Int Immunol. 1997; 9:689–702. [PubMed: 9184914]
24. Ehrenmann F, Kaas Q, Lefranc MP. IMGT/3Dstructure-DB and IMGT/DomainGapAlign: a database and a tool for immunoglobulins or antibodies, T cell receptors, MHC, IgSF and MhcSF. Nucleic Acids Res. 2010; 38:D301–D307. [PubMed: 19900967]
25. Lim TS, et al. V-gene amplification revisited. An optimised procedure for amplification of rearranged human antibody genes of different isotypes. N Biotechnol. 2010; 27:108–117. [PubMed: 20083243]
26. Mei HE, et al. Blood-borne human plasma cells in steady state are derived from mucosal immune responses. Blood. 2009; 113:2461–2469. [PubMed: 18987362]
27. Kyu SY, et al. Frequencies of human influenza-specific antibody secreting cells or plasmablasts post vaccination from fresh and frozen peripheral blood mononuclear cells. J Immunol Methods. 2009; 340:42–47. [PubMed: 18996127]
28. Mazor Y, Barnea I, Keydar I, Benhar I. Antibody internalization studied using a novel IgG binding toxin fusion. J Immunol Methods. 2007; 321:41–59. [PubMed: 17336321]
29. Friguet B, Chaffotte AF, Djavadi-Ohaniance L, Goldberg ME. Measurements of the true affinity constant in solution of antigen-antibody complexes by enzyme-inked immunosorbent assay. J Immunol Methods. 1985; 77:305–319. [PubMed: 3981007]
30. Brochet X, Lefranc MP, Giudicelli V. IMGT/V-QUEST: the highly customized and integrated system for IG and TR standardized V-J and V-D-J sequence analysis. Nucleic Acids Res. 2008; 36:W503–W508. [PubMed: 18503082]

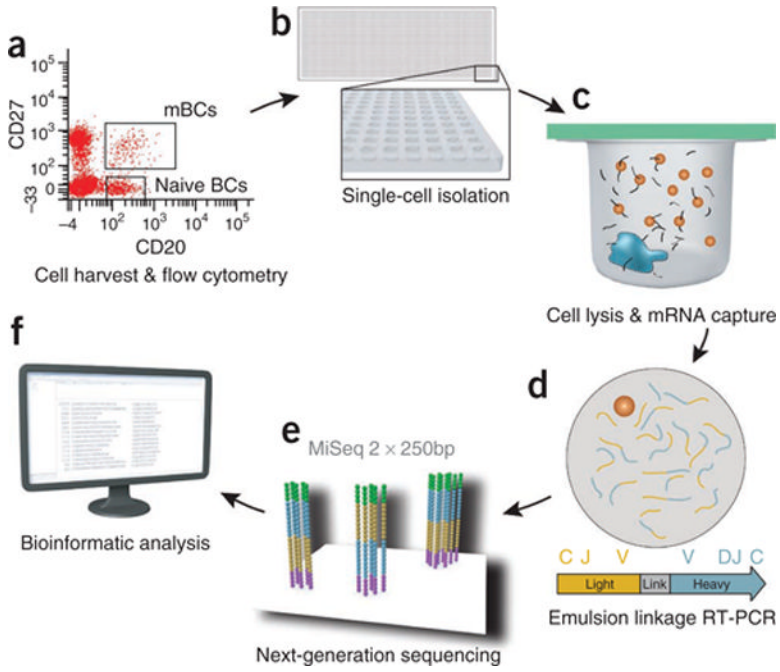

Figure 1.

Overview of the high-throughput methodology for paired V_H:V_L antibody repertoire analysis. (**a**) B-cell populations are sorted for desired phenotype (e.g., mBCs, memory B cells, naive BCs, naive B cells). (**b**) Single cells are isolated by random settling into 125-pl wells (56-μm diameter) printed in polydimethylsiloxane (PDMS) slides the size of a standard microscope slide (1.7×10^5 wells/slide); 2.8-μm poly(dT) microbeads are also added to the wells (average 55 beads/well). (**c**) Wells are sealed with a dialysis membrane and equilibrated with lysis buffer to lyse cells and anneal V_H and V_L mRNAs to poly(dT) beads (blue figure represents a lysed cell, orange circles depict magnetic beads, black lines depict mRNA strands; see Supplementary Video 1). (**d**) Beads are recovered and emulsified for cDNA synthesis and linkage PCR to generate an ~850–base pair V_H:V_L cDNA product (Supplementary Fig. 1). (**e**) Next-generation sequencing is performed to sequence the linked strands. (**f**) Bioinformatic processing is used to analyze the paired V_H:V_L repertoire.

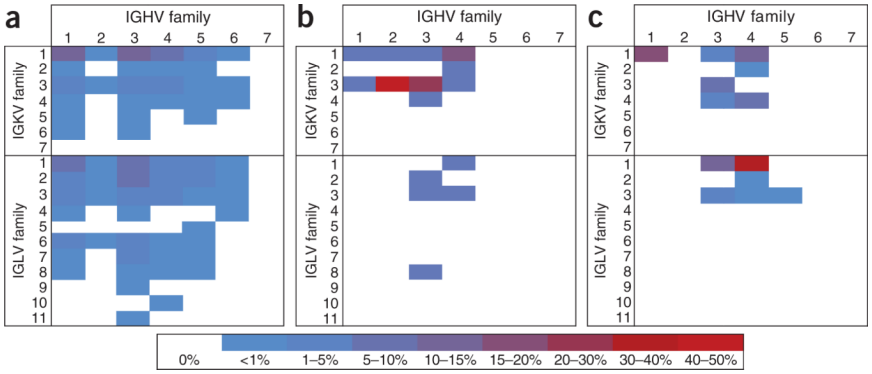

Figure 2.

V_H:V_L gene family usage of unique CDR-H3:CDR-L3 pairs identified by high-throughput sequencing of cell populations from three different individuals in separate experiments using the workflow in Figure 1. (**a**) Healthy donor peripheral IgG^+ B cells (*n* = 2,716 unique CDR3 pairs). (**b**) Peripheral TT–specific plasmablasts, isolated 7 d post-TT immunization ($CD19^+CD3^-CD14^-CD38^{++}CD27^{++}CD20^-TT^+$, *n* = 86 unique pairs). (**c**) Peripheral memory B cells isolated 14 d after influenza vaccination ($CD19^+CD3^-CD27^+CD38^{int}$, *n* = 240 unique pairs). Each panel presents data from an independent experiment obtained from 61,000 fresh B cells (**a**), ∼400 frozen/thawed plasmablasts (**b**) and 8,000 twice frozen/thawed memory B cells (**c**). Color indicates percentage of unique pairs from the sample; repertoire statistics are reported in Supplementary Tables 1 and 4.

Table 1

TT-binding affinities of IgG antibodies sequenced from TT^+ peripheral plasmablasts

Antibody ID	Gene family assignment[a]	Affinity (K_D)
TT1	HV3-HD1-HJ6: KV3-KJ5	1.6 ± 0.1 nM
TT2	HV3-HD3-HJ4: LV3-LJ1	14 ± 3 nM
TT3	HV1-HD2-HJ4: KV3-KJ5	3.6 ± 1.8 nM
TT4	HV2-HD2-HJ4: KV1-KJ1	2.7 ± 0.3 nM
TT5	HV4-HD2-HJ6: KV2-KJ3	18 ± 4 nM
TT6	HV1-HD3-HJ4: KV1-KJ2	0.57 ± 0.03 nM
TT7	HV4-HD3-HJ4: KV1-KJ2	0.46 ± 0.01 nM
TT8	HV3-HD3-HJ4: LV8-LJ3	2.8 ± 0.3 nM
TT9	HV4-HD2-HJ4: KV1-KJ1	0.10 ± 0.01 nM
TT10	HV1-HD3-HJ5: KV3-KJ5	1.6 ± 0.1 nM

Peripheral blood mononuclear cells were isolated from one healthy volunteer 7 d after TT boost immunization and TT-binding $CD19^+CD3^-CD14^-CD38^{++}CD27{+}{+}CD20^-$ cells were sorted and analyzed as in Figure 1. Genes encoding ten of the sequenced V_H:V_L pairs were cloned into an IgG expression vector and expressed transiently in HEK293F cells. TT-binding affinities of the resulting IgG were calculated from competitive ELISA dilution curves.

[a] Each heavy and light chain was distinct (Supplementary Data).