



**HAL**  
open science

## High-throughput single-cell activity-based screening and sequencing of antibodies using droplet microfluidics

Annabelle Gérard, Adam Woolfe, Guillaume Mottet, Marcel Reichen, Carlos Castrillón, Vera Menrath, Sami Ellouze, Adeline Poitou, Raphaël Doineau, Luis Briseño-Roa, et al.

### ► To cite this version:

Annabelle Gérard, Adam Woolfe, Guillaume Mottet, Marcel Reichen, Carlos Castrillón, et al.. High-throughput single-cell activity-based screening and sequencing of antibodies using droplet microfluidics. *Nature Biotechnology*, 2020, 38, pp.715-721. 10.1038/s41587-020-0466-7 . pasteur-02505776

**HAL Id: pasteur-02505776**

**<https://pasteur.hal.science/pasteur-02505776>**

Submitted on 11 Mar 2020

**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 - NonCommercial - NoDerivatives 4.0 International License

# High-throughput single-cell activity-based screening and sequencing of antibodies using droplet microfluidics

Annabelle Gérard<sup>1,‡</sup>, Adam Woolfe<sup>1,¶</sup>, Guillaume Mottet<sup>2,¶</sup>, Marcel Reichen<sup>1,¶</sup>, Carlos Castrillon<sup>2,3,4,¶</sup>, Vera Menrath<sup>1,¶</sup>, Sami Ellouze<sup>1,¶</sup>, Adeline Poitou<sup>1,¶</sup>, Raphaël Doineau<sup>1,3,4,5,¶</sup>, Luis Briseno-Roa<sup>1,6</sup>, Pablo Canales-Herrerias<sup>2,3,7</sup>, Pascaline Mary<sup>8</sup>, Gregory Rose<sup>8</sup>, Charina Ortega<sup>8</sup>, Matthieu Delincé<sup>8</sup>, Sosthene Essono<sup>8</sup>, Bin Jia<sup>9</sup>, Bruno Iannascoli<sup>2</sup>, Odile Richard-LeGoff<sup>2</sup>, Roshan Kumar<sup>8</sup>, Samantha N. Stewart<sup>8</sup>, Yannick Pousse<sup>1</sup>, Bingqing Shen<sup>1</sup>, Kevin Grosselin<sup>1,4</sup>, Baptiste Saudemont<sup>4</sup>, Antoine Sautel-Caillé<sup>4</sup>, Alexei Godina<sup>4</sup>, Scott McNamara<sup>1</sup>, Klaus Eyer<sup>7</sup>, Gaël A. Millot<sup>10</sup>, Jean Baudry<sup>7</sup>, Patrick England<sup>11</sup>, Clément Nizak<sup>4</sup>, Allan Jensen<sup>1,12,‡</sup>, Andrew D. Griffiths<sup>4,‡,\*</sup>, Pierre Bruhns<sup>2,‡,\*</sup> and Colin Brenan<sup>1,8,13,‡,\*</sup>

<sup>1</sup> HiFiBiO SAS, 29 Rue du Faubourg Saint Jacques, 75014 Paris, France;

<sup>2</sup> Unit of Antibodies in Therapy and Pathology, Institut Pasteur, UMR1222 INSERM, F-75724 Paris, France;

<sup>3</sup> École Doctorale Frontières du Vivant (FdV), Centre de Recherches Interdisciplinaires, F-75015 Paris, France;

<sup>4</sup> Laboratoire de Biochimie (LBC), ESPCI Paris, PSL Research University, CNRS UMR8231 Chimie Biologie Innovation, F-75005 Paris, France;

<sup>5</sup> Current address: Institute for Integrative Nanosciences, Leibniz IFW Dresden e.V., 01069 Dresden, Germany;

<sup>6</sup> Current address: Medetia Pharma, Institut Imagine, 24 Boulevard du Montparnasse, 75015, Paris, France;

<sup>7</sup> Laboratoire Colloïdes et Matériaux Divisés (LCMD), ESPCI Paris, PSL Research University, CNRS UMR8231 Chimie Biologie Innovation, F-75005 Paris, France;

<sup>8</sup> HiFiBiO Inc., 237 Putnam Ave, Cambridge, MA, 02139;

<sup>9</sup> Abbvie Bioresearch Center, 100 Research Drive, Worcester, MA, 01605, USA;

<sup>10</sup> Institut Pasteur, Hub Bioinformatique et Biostatistique, C3BI, USR 3756 IP CNRS, F-75724 Paris, France;

<sup>11</sup> Plate-forme de Biophysique Moléculaire, Institut Pasteur, CNRS-UMR 3528, F-75724 Paris, France;

Institut Pasteur, Centre d'Innovation et de Technologie, Plateforme de Biophysique des Macromolécules et de leurs Interactions, F-75724 Paris, France;

<sup>12</sup> Current address: Lundbeck A/S, Ottiliavej 9, 2500 Valby, Denmark;

<sup>13</sup> Current address: 1CellBio Inc., 200 Dexter Ave, Suite 220, Watertown, MA 02472 USA

¶ These authors contributed equally.

‡ Co-senior authorship.

\* Corresponding authors.

Correspondence to: Colin Brenan ([c.brenan@1cell-bio.com](mailto:c.brenan@1cell-bio.com)), Pierre Bruhns ([pierre.bruhns@pasteur.fr](mailto:pierre.bruhns@pasteur.fr)) or Andrew D Griffiths ([andrew.griffiths@espci.fr](mailto:andrew.griffiths@espci.fr)).

## **ABSTRACT**

Mining the antibody repertoire of plasma cells and plasmablasts could enable the discovery of useful antibodies for therapeutic or research purposes<sup>1</sup>. We present a method for high-throughput, single-cell screening of IgG-secreting primary cells to characterize antibody binding to soluble and membrane-bound antigens. *CelliGO* is a droplet microfluidics system that combines high-throughput screening for IgG activity, using fluorescence-based in-droplet single-cell bioassays<sup>2</sup>, with sequencing of paired antibody V genes, using in-droplet single-cell barcoded-reverse transcription. We analyzed IgG repertoire diversity, clonal expansion and somatic hypermutation in cells from mice immunized with a vaccine target, a multifunctional enzyme or a membrane-bound cancer target, and from human B cells activated *ex vivo*. Immunization with these antigens yielded 100-1,000 IgG sequences per mouse. We generated 77 recombinant antibodies from the identified sequences and found that 93% recognized the soluble antigen and 14% the membrane antigen. The approach also allowed recovery of ~450-900 IgG sequences from ~2,200 IgG-secreting activated human memory B cells, demonstrating its versatility.

## MAIN TEXT

The immune system generates a vast repertoire of antibodies in response to infection or immunization that can potentially be explored for diagnostic, therapeutic or research applications. IgGs are the most abundant class of immunoglobulin in circulation. They protect from infections by blocking molecular interactions, and by inducing complement-dependent cytotoxicity (CDC) and antibody-dependent cellular cytotoxicity (ADCC). IgGs are matured during ongoing immune responses, and through this process improve their specificity and affinity for antigen. Their production can also be rapidly recalled following previously-elicited protective immunity in response to pathogen or immunogen exposure<sup>3</sup>. The phenotypic diversity of the target-specific IgG repertoire in an individual underlies protection following vaccination or infection<sup>4</sup>. IgGs are secreted by circulating antibody-secreting cells (plasmablasts) and short-lived or long-lived tissue-resident antibody-secreting cells (plasma cells). Analyzing the activity and sequence of IgGs secreted by plasmablasts and plasma cells is hence of great interest both to investigate their pivotal role in the humoral immune response and to obtain antibodies for therapeutic or other applications<sup>5</sup>.

Until recently, target-specific IgG discovery and repertoire analysis was based on B cell immortalization using hybridoma technology or Epstein-Barr Virus (EBV) transformation, which are relatively inefficient, are biased towards dividing plasmablasts but not fully matured plasma cells, and are limited to specific species. Direct PCR cloning of the V-gene sequences encoding antibody variable (V) regions, from plasma cells has been reported<sup>6, 7</sup> but without prior knowledge of antigen specificity. Recently, next generation sequencing (NGS) approaches have been developed to probe the IgG repertoire more deeply<sup>1, 8</sup>, some even recovering the original pairing of heavy-chain and light-chain variable region ( $V_H$  and  $V_L$ )<sup>9-13</sup>. Antigen-labeled IgG<sup>+</sup> memory B cells and plasmablasts have been isolated using fluorescence-activated cell sorting (FACS) and sequenced<sup>9, 10</sup>, but not antigen-specific plasma cells, given that they lack surface IgG expression. Phenotypic analyses of antibody-secreting cells has evolved from ELISpot<sup>14</sup> to high-throughput assays based on micro/nanowells or pico/nanolitre droplets in microfluidic systems<sup>15-17</sup>, but remains non-adapted or reliant on low-throughput manual micromanipulation for further  $V_H$ - $V_L$  sequencing. *In vitro* display technologies, notably phage and yeast display, have proved to be powerful tools to discover therapeutic antibodies<sup>18</sup>; however, screening is based on binding and rarely on function<sup>18</sup>, and was, until recently<sup>12, 13</sup>, unable to preserve the cognate  $V_H$ - $V_L$  pairing, limiting its utility for the analysis of immune responses. Yeast display has been shown to maintain cognate  $V_H$ - $V_L$

pairing to yield antibodies with higher sensitivity and specificity than random pairing<sup>13</sup>, it involves multiple steps for library construction, antibody fragment display and multiple rounds of screening, and is restricted to soluble antigens<sup>12, 13</sup>. There is, therefore, a need for a system that couples high-throughput single-cell phenotypic screening with paired V<sub>H</sub>-V<sub>L</sub> sequencing of IgG-secreting primary cells for both soluble and membrane-bound antigens in a flexible format that enables direct screening for functional activities<sup>19</sup>.

We reported previously a droplet-based microfluidics technology in which up to 16,000 single IgG-secreting cells are compartmentalized in tens of thousands of 40 pL droplets and analyzed in two-dimensional droplet arrays using a fluorescence relocation-based immunoassay to measure IgG secretion rate and affinity<sup>2</sup>. This method, however, allows neither recovery of specific cells based on the properties of secreted IgGs (the phenotype), nor determination of the sequence of the secreted antibody genes (the genotype).

Here, we describe a droplet-based microfluidics technology, CelliGO™, that enables high-throughput single-cell screening of millions of non-immortalized cells based on the phenotypic properties of secreted IgGs, paired V<sub>H</sub>-V<sub>L</sub> sequencing of antibodies with the desired properties, bioinformatics analysis and rapid expression and *in vitro* antibody characterization (Fig. 1a).

Microfluidic systems allow the generation of picolitre-volume droplets containing single IgG-secreting cells (Fig. 1b left), identification of droplets containing IgGs with the desired activity and sorting of those identified droplets at up to 600 s<sup>-1</sup> (Fig. 1b, middle). This highly flexible droplet-based system allows IgG-secreting cells to be screened based on a diversity of assays including binding to soluble antigens, cross-reactivity or specific binding (using multiple soluble antigens), binding to cell-surface targets<sup>17</sup> (on both bacterial and eukaryotic cells), inhibition of target activity<sup>16</sup>, cellular internalization, opsonization and functional modulation of cellular signaling pathways (Supplementary Fig. 1).

Sorted cells are immediately re-compartmentalized in sub-nanoliter droplets together with a hydrogel bead grafted with barcoded primers<sup>20</sup>, generating droplets containing a single cell and a single bead (Fig. 1b, right). After cell lysis and reverse transcription of V<sub>H</sub> and V<sub>L</sub> mRNAs in the droplets, the cDNAs from each cell carry a unique barcode, allowing cognate

V<sub>H</sub> and V<sub>L</sub> pairs to be identified by NGS. This platform allows coupled phenotype-genotype repertoire characterization and informed selection of multiple target-specific antibody candidates within just 3.5 days of cell harvesting, followed by ~17 days for gene synthesis, cloning, production and validation of the phenotypic properties of the selected IgGs (Fig. 1c).

We applied the platform to characterize repertoires of IgG-secreting cells from mice immunized with two soluble protein antigens, the vaccine target Tetanus Toxoid (TT) and the multifunctional enzyme Neuroleukin/Glucose-6-Phosphate Isomerase (GPI) involved in cancer progression and metastasis<sup>21, 22</sup>, and with cells overexpressing an integral human membrane protein, Tetraspanin-8 (TSPAN8), a potential cancer target involved in tumor proliferation and invasion<sup>23</sup>.

We screened and sequenced IgG-secreting splenocytes from two mice immunized with TT (TT#1 and TT#2), three mice immunized with GPI (GPI#1, GPI#2 and GPI#3) and two mice immunized with TSPAN8-expressing mouse M300.19 cells (TSPAN8#1 and TSPAN8#2), or from a pool of splenocytes from five TT-immunized mice (TT#5-pool) (Online Methods). Splenocytes were harvested from each mouse and B cell lineages enriched 4-6 fold by depletion, resulting in 0.8-10% B220<sup>-int</sup>/CD138<sup>+</sup> plasmablasts and plasma cells (Online Methods, Supplementary Table 1). From 1.5 to 11 million enriched cells were compartmentalized with bioassay reagents in 4-11 million monodisperse 40 pL aqueous droplets, produced at 2,100 droplets s<sup>-1</sup> by hydrodynamic flow focusing<sup>24</sup> with an inert fluorinated carrier oil (Fig. 1b, left).

To detect cells secreting IgGs binding soluble antigens (TT and GPI), splenocytes from TT- or GPI-immunized mice were compartmentalized in droplets following a Poisson distribution with a mean,  $\lambda$ , of 0.2-0.4 cells per droplet to minimize compartmentalization of more than one cell per droplet. After incubating for 60 min at 37°C to allow antibody secretion, antigen-binding was detected using a fluorescent sandwich immunoassay, in which secreted antibodies were captured onto paramagnetic colloidal nanoparticles coated with an anti-mouse kappa light chain (Ig $\kappa$ ) nanobody (V<sub>H</sub>H) in each droplet<sup>2</sup>. Upon application of a magnetic field, the ~1,300 nanoparticles in each droplet formed an elongated aggregate (termed a beadline). Each droplet also contained AlexaFluor647-labelled Fc-specific F(ab')<sub>2</sub>-anti-mouse IgG that relocated onto the beadline if the secreted antibody was an IgG, and AlexaFluor488-

labelled antigen that relocated to the beadline if the IgG had affinity for the antigen, but not otherwise (Fig. 2a).

The spatial distribution of fluorescence in the droplets was analyzed by re-injecting them into a second microfluidic chip where each droplet was scanned with superimposed laser lines and epifluorescence detected using photomultiplier tubes. Secreted IgG and antigen binding were determined from red- and green-fluorescence localization to the beadline, respectively (Online Methods; Fig. 2a-c; Supplementary Fig. 2), and droplets containing a cell of interest sorted by fluorescence-activated dielectrophoretic sorting (Online Methods; Fig. 2d; Supplementary Fig. 3). Compared to approaches relying on the co-encapsulation of a single cell and a single microparticle<sup>25</sup>, the use of beadlines comprising many nanoparticles increases the binding capacity, eliminates the double Poisson statistics governing single cell-single microparticle encapsulation to increase the number of cells assayed, drastically lowers the number of droplets containing no particles, and lowers the risk that the fluorescent signals are not in the focal plane of the detection system<sup>2</sup>. Additionally, droplet scanning with laser lines parallel to beadlines optimizes detection sensitivity as relocalized fluorophores will be maximally excited.

To detect cells secreting IgGs binding to cells overexpressing human TSPAN8, Calcein AM-labeled M300.19 reporter cells expressing TSPAN8 (i.e. the same cell line used for mouse immunizations) were co-compartmentalized in 80 pL droplet with Calcein Violet-labeled B cell-enriched splenocytes with a mean,  $\lambda$ , of 1.6 reporter cells and 1 splenocyte per droplet (Fig. 2e) and incubated for 60 min at 37°C. AlexaFluor647-labelled Fc-specific F(ab')<sub>2</sub>-anti-mouse IgG relocated onto the green fluorescent reporter cell if the secreted IgG bound the antigen, but not otherwise (Fig. 2e-g). The distribution of fluorescence in the droplets was analyzed as above and droplets containing a cell of interest were sorted by fluorescence activated dielectrophoretic sorting (Online Methods; Fig. 2h; Supplementary Fig. 4).

IgG secretion was detected in droplets from 2.7% and 2.8% of the compartmentalized cells from B cell-enriched splenocytes from TT-immunized and GPI-immunized mice respectively. On average, for the TT-immunized mice ~47% of IgG-secreting cells showed detectable TT binding (58% for TT#5-pool) and for the GPI-immunized mice ~28% of IgG-secreting cells showed GPI-binding, which is consistent with the 6-fold higher mean serum titre (EC50) in TT-immunized mice (Supplementary Table 1). Very low frequencies (<0.002%) of antigen-

binding IgGs were detected in splenocytes from mice immunized with a different antigen, naive mice, or B cell deficient (RAG<sup>-/-</sup>) mice (Supplementary Fig. 5).

From ~3,500 to ~22,000 droplets containing secreted IgG with detectable binding to TT, GPI or TSPAN8-overexpressing cells were sorted per run (Online Methods; Supplementary Table 1). The entire compartmentalization, bioassay and sorting process took typically 5.5 hours (Fig. 1c), which contributed to the high viability (79±2%) of cells recovered from sorted droplets (Supplementary Fig. 6a). Recovered cells demonstrated plasma cell/blast morphology with ~56% of sorted cells from TT-sorts secreting IgG (by ELISpot) and most, if not all, of these cells secreted antigen-specific IgG (Supplementary Fig. 6b,c).

To initiate barcoded single-cell reverse transcription, sorted cells were re-compartmentalized in ~10 min at ~250 s<sup>-1</sup> in ~100,000 ~1 nL droplets together with lysis buffer, reverse transcriptase and ~160 pL hydrogel beads (Fig. 1b, right). Each hydrogel bead carried ~10<sup>9</sup> primers for V<sub>H</sub> and V<sub>L</sub> reverse transcription, tagged with a unique barcode generated by split-and-pool synthesis (Online Methods, Supplementary Fig. 7). The injection of close-packed deformable beads (Supplementary Movie)<sup>26</sup> resulted in >79% of droplets containing a single barcoded hydrogel bead and <2.5% containing more than one bead, and led to compartmentalization of ~70% of droplets containing a single cell together with a single bead (Supplementary Fig. 8). In the droplet the cells were lysed and the barcoded primers, released from the beads by photocleavage, were used as primers to reverse transcribe the V<sub>H</sub> and V<sub>L</sub> mRNA (Fig. 3a). The pooled barcoded cDNAs recovered from the droplets were amplified and sequenced using Illumina MiSeq 2x300bp paired-end sequencing to identify cognate V<sub>H</sub>-V<sub>L</sub> pairs carrying the same barcode (Supplementary Fig. 9).

Sequencing data were processed computationally to identify cell-specific barcodes and identify cognate V<sub>H</sub>-V<sub>L</sub> pairs (one pair per barcode) and corrected for errors introduced by PCR or sequencing by creation of consensus sequences (see Online Methods; Supplementary Note; Supplementary Figs. 10-11). Non-redundant V<sub>H</sub>-V<sub>L</sub> pairs were clustered into IgG clonotypes that derive from an ancestral recombination event and have expanded and undergone affinity maturation (see Online Methods). The total number of V<sub>H</sub>-V<sub>L</sub> pairs, unique IgGs and to a lesser extent, the total number of clonotypes correlate positively with serum IgG titer (EC50) in TT and GPI-immunized mice (Supplementary Fig. 12; Supplementary Table 1). For each antigen sort, we saw a high diversity of IgGs covering multiple V-gene families



(Fig. 3b,c; Supplementary Figs. 13-15). Matrix statistical comparison of V-gene usage in heavy and light chains separately (Online Methods) found both significant differences ( $p < 0.05$ ) between samples from mice immunized against different antigens but also between mice immunized with the same antigen, particularly in the case of TT (Supplementary Table 2) compared to GPI or TSPAN8. This suggests higher stochasticity in the immune response of individual mice to TT compared to GPI or TSPAN8. Analyzing the combined  $V_H$ - $V_L$  gene usage, the difference (distance,  $D$ ) was almost maximal (*i.e.*, close to 1) with no or rare common  $V_H$ - $V_L$  pairs identified between mice immunized with different antigens and was lower, with frequent common  $V_H$ - $V_L$  pairs identified, between mice immunized with the same antigen (Supplementary Fig. 16-17). Multidimensional scaling of the distance matrix indicates an antigen-dependent clustering of the combined  $V_H$ - $V_L$  usages, without cluster overlaps in the first two dimensions (Supplementary Fig. 18). Moreover, within IgG clonotypes, constructing phylogenetic trees of natively paired  $V_H$  and  $V_L$  genes revealed evidence of co-evolution during affinity maturation (exemplified for GPI sorts in Fig. 3d,e). An average of 95, 72 and 35 IgG clonotypes were identified for TT-, GPI- and TSPAN8-immunized mice, respectively (per  $1.5 \times 10^6$  encapsulated cells; Supplementary Table 1). The ratio of  $V_H$ - $V_L$  pairs in multi-member clonotypes compared to singletons, the average size of clonally expanded  $V_H$ - $V_L$  clusters, and the fraction of highly expanded clonotypes (containing  $>10$  members), were all highest for TT, intermediate for GPI and lowest for TSPAN8-immunized mice (Fig. 3f; Supplementary Fig. 19a,b). This is consistent with greater clonal expansion being linked to stronger and more specific immune responses but may also reflect difference in the screening bioassays (using purified antigen versus transfected reporter cells).

Screening of cells pooled from five TT-immunized mice identified 219 IgG clonotypes, *i.e.* a  $>2$ -fold increase in diversity compared to a single TT-immunized mouse, suggesting a pooling strategy for maximizing antibody candidate diversity per sort (Supplementary Table 1). For both soluble antigens,  $\geq 96$ -99% of the non-redundant IgG sequences had undergone somatic hypermutation (SHM), compared with only 65% for the membrane antigen, which may be due to lower efficiency of cell-based immunization and/or higher false positive rates (Supplementary Table 1). Across sorts, amino acid changes were more abundant in  $V_H$  than  $V_L$  (Supplementary Figs. 19c, 20), and enriched in CDRs compared to framework regions (Supplementary Fig. 21), as reported for other antigenic specificities<sup>27</sup>. We found no significant differences in the extent of SHMs between expanded and singleton clonotypes in TT, but significantly higher numbers of SHMs in expanded clonotypes in GPI and to a lesser

extent TSPAN8 (Supplementary Fig. 19d) suggesting affinity maturation is a feature of clonally expanded cells, albeit dependent on the antigen.

To demonstrate antigen-binding activity, putative TT-, GPI- and TSPAN8-binding IgGs from one or more representative sorts were cloned and expressed as human IgG1, $\kappa$  isotypes (Online Methods, Supplementary Table 3). From the TT sorts, 42 V<sub>H</sub>-V<sub>L</sub> pairs were expressed. Almost all (39/42; ~93%) bound to TT by ELISA. EC<sub>50</sub> for TT was measured for 27 of these, and ranged from 0.1 nM to 300 nM (median 0.8 nM). From the GPI sorts, nearly all (13/14; ~93%) V<sub>H</sub>-V<sub>L</sub> pairs expressed had measurable affinity for GPI by surface plasmon resonance (SPR) ranging from 0.2 nM to 26 nM (median 0.4 nM) (Fig. 3g). No correlation between the number of mutations and affinity was observed, as described for antibody responses following infection<sup>28, 29</sup>, though four expressed germline V<sub>H</sub>-V<sub>L</sub> pairs (no somatic mutations in either chain) showed no detectable antigen binding (Supplementary Table 3). From the TSPAN8 sorts, 7 out of 21 (33%) V<sub>H</sub>-V<sub>L</sub> pairs demonstrated higher binding to TSPAN8-transfected cells than to non-transfected cells, but among these, only 3 (14%) bound strongly enough to allow the affinity for TSPAN8 to be determined using flow cytometry, with affinities of 374, 397 and 3640 nM (Fig. 3h; Supplementary Fig. 22; Supplementary Table 3). The lower frequency of IgGs showing detectable antigen binding compared to TT and GPI may be due to the immunization strategy as the murine TSPAN8-expressing M300.19 cells used for immunizations in BALB/c mice originate from NIH Swiss outbred mice and should be immunogenic for other membrane antigens, and the fact that the immunization and the in-droplet bioassay were performed using the same transfected cell line. The lower affinities of the anti-TSPAN8 IgGs compared to anti-TT and anti-GPI IgGs may also be due to the immunization strategy: the frequency of somatic hypermutations was lower than for with anti-TT and anti-GPI IgGs, indicating a lower level of affinity maturation (Supplementary Fig. 20). Nevertheless, this shows that CelliGO is capable of discovering low affinity antibodies against challenging membrane protein targets that elicit weak immune responses such as TSPAN8.

The CelliGO pipeline provides the capacity to deeply mine the IgG repertoire by high-throughput screening of millions of IgG-secreting cells based on coupled analysis of phenotype (activity of the secreted IgG) and genotype (the sequence of the IgG), on both soluble (antigen or autoantigen) and membrane-bound targets. From 337 to 1,053 non-redundant IgGs were identified from splenocytes of a single TT-immunized mouse

(Supplementary Table 1) with >90% cloned IgGs binding TT, compared to only 21 using hybridomas and 42 using phage-display<sup>30</sup>, and 86 identified by FACS of TT-stained human peripheral IgG<sup>+</sup> plasmablasts followed by single-cell mRNA capture and NGS<sup>9</sup>. Furthermore, the median affinity of the cloned anti-TT IgGs was 0.8 nM, compared to 3.4 nM, 9.0 nM and 2.15 nM with hybridomas, phage display and FACS plus single-cell mRNA capture and NGS, respectively<sup>9, 30</sup>.

The system can be simply adapted to analyze circulating human memory B cell repertoires, by inducing peripheral B cells *in vitro* to secrete IgG prior to encapsulation in droplets. Human memory B cells enriched from PBMCs from two donors were activated with anti-CD40 monoclonal antibody and CpG oligodeoxynucleotides (CpG) prior to encapsulation. Analysis of three different runs resulted in the identification of ~460-900 paired V<sub>H</sub>-V<sub>L</sub> sequences per run from ~2,300 IgG-secreting cells in each run (Supplementary Fig. 23). Simply changing the primer regions on the barcoded primers on the hydrogel beads allows sequencing of antibodies from different species, or, indeed, transcriptomic profiling of cells (Supplementary Fig. 24). Alternatively, paired V<sub>H</sub>-V<sub>L</sub> sequencing on sorted cells can be performed by single-cell reverse transcription-PCR in microtitre plate wells prior to NGS (Supplementary Fig. 25), which allows for direct cloning into expression vectors of the amplified V<sub>H</sub> and V<sub>L</sub> genes without gene synthesis.

CelliGO is capable of analyzing cells from different species and allows phenotypic screening of IgG repertoires derived from large numbers of plasma cells, plasmablasts and *in vitro* activated memory B cells from different anatomical compartments at high throughput, and using a wide range of assays, including functional and cell-based assays (Supplementary Fig. 1). It should facilitate the analysis of immune responses and the identification of potent antibodies with unique functional activities for therapeutic purposes, including against difficult targets like insoluble membrane antigens.

## **ACKNOWLEDGEMENTS**

We would like to thank Matthew Holsti and Will Somers (Pfizer, BioMedicine Design), Cambridge, MA, USA) for advice on the technology development and critical reading of the manuscript; Robert Nicol (Whitehead Institute, MIT Center for Genome Research, Cambridge, MA, USA) for advice on barcoded sequencing; at Institut Pasteur, Paris, France, Hugo Mouquet for advice on antibody sequence selection, Friederike Jönsson for advice on flow cytometry analysis; Tim Kirk and Sophie Foulon (ESPCI Paris, Paris, France) for their help developing the barcoded hydrogel beads; Samantha N. Stewart (HiFiBio Inc, Boston, MA) for her help producing and analysing recombinant antibodies; the Institut Pierre-Gilles de Gennes (IPGG, Paris, France) for use of clean room facilities and the laser engraver (CII08, Axyslaser).

This work was supported by the French Agence Nationale de la Recherche (ANR-14-CE16-0011 project DROPmAbs), by the Centre d'Innovation et Recherche Technologique (Citech) through the "Institut Carnot Pasteur Microbes & Santé" (ANR 16 CARN 0023-01), by BPI France (OSIRIS and CELLIGO projects) and by the French "Investissements d'Avenir" program via the CELLIGO project and grant agreements ANR-10-NANO-02, ANR-10-IDEX-0001-02 PSL, ANR-10-LABX-31 and ANR-10-EQPX-34. NGS was performed by the ICGex NGS platform of the Institut Curie (Paris) and the Institut de Biologie Intégrative de la Cellule (I2BC) platform (Gif-sur-Yvette) supported by the grants ANR-10-EQPX-03 and ANR-10-INBS-09-08 (France Génomique Consortium), by the Canceropole Ile-de-France and by the SiRIC-Curie program - SiRIC Grant INCa-DGOS- 4654. CC acknowledges financial support from CONCYTEC, Peru. PCH is a scholar in the Pasteur - Paris University (PPU) International PhD program and was also supported by the 'Fondation pour la Recherche Médicale' (FRM; FDT201904008240). KE acknowledges financial support from the 'Swiss National Science Foundation' and the 'The Branco Weiss Fellowship - Society in Science'.

## **CONFLICT OF INTEREST**

A.D. Griffiths and C. Brenan are co-founders of Seven Pines Holding BV, and HiFiBio SAS is a subsidiary of Seven Pines Holding BV. Patents have been filed on some aspects of this work and future development of the platform and the inventors may receive payments related to exploitation of these under their employer's rewards to inventors scheme.

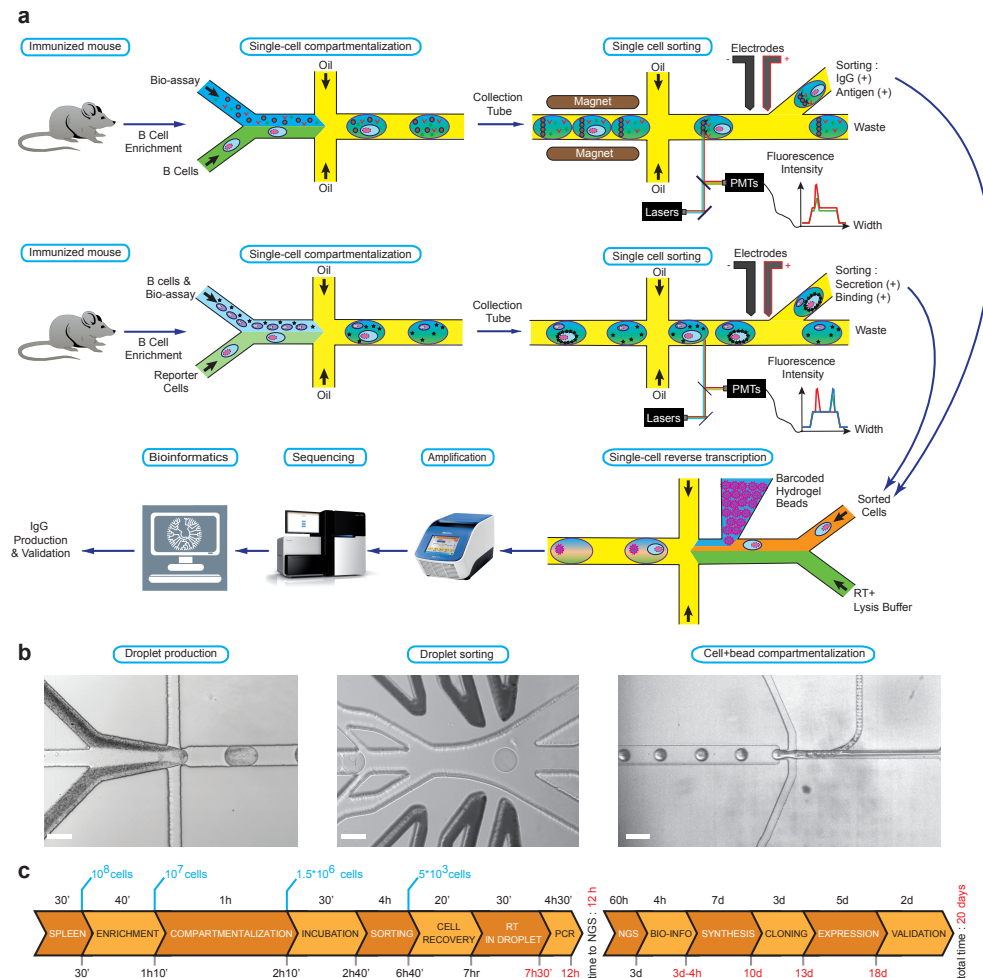
### **AUTHOR CONTRIBUTIONS**

C.B., P.B., A.Gérard, A.J. and A.D.G. designed and supervised the study ; C.B., P.B. and A.D.G. secured funding; L.B.-R., P.C.-H., C.C., M.D., R.D., K.E., S.El., S.Es., A.G., K.G., B.I., B.J., R.K., V.M., P.M., S.M., G.M., C.O., Y.P., A.P., M.R., O.R.-L., G.R., A.S.-C., B.Sa., B.Sh., S.N.S. and A.W. performed the experiments; J.B., C.B., P.B., P.E., A.Gérard, A.D.G., A.J., G.A.M., G.M., C.N. and A.W. analyzed the data; P.B., A.Gérard, A.D.G. and A.W. wrote the paper.

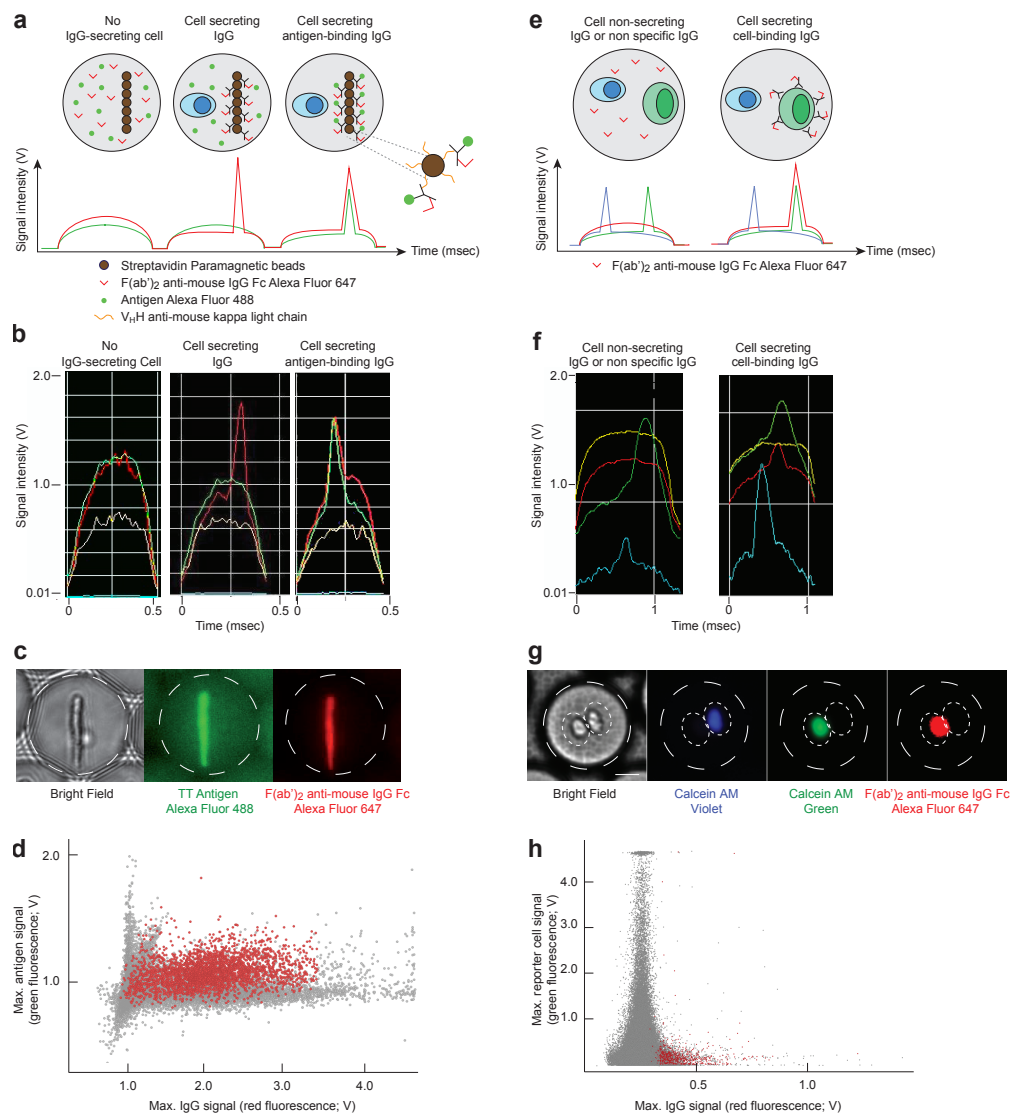
## REFERENCES

1. Georgiou, G. et al. The promise and challenge of high-throughput sequencing of the antibody repertoire. *Nat Biotechnol* **32**, 158-168 (2014).
2. Eyer, K. et al. Single-cell deep phenotyping of IgG-secreting cells for high-resolution immune monitoring. *Nat Biotechnol* **35**, 977-982 (2017).
3. Bannard, O. & Cyster, J.G. Germinal centers: programmed for affinity maturation and antibody diversification. *Curr Opin Immunol* **45**, 21-30 (2017).
4. Gunn, B.M. & Alter, G. Modulating Antibody Functionality in Infectious Disease and Vaccination. *Trends Mol Med* **22**, 969-982 (2016).
5. Nutt, S.L., Hodgkin, P.D., Tarlinton, D.M. & Corcoran, L.M. The generation of antibody-secreting plasma cells. *Nat Rev Immunol* **15**, 160-171 (2015).
6. Meijer, P.J. et al. Isolation of human antibody repertoires with preservation of the natural heavy and light chain pairing. *J Mol Biol* **358**, 764-772 (2006).
7. Di Niro, R. et al. High abundance of plasma cells secreting transglutaminase 2-specific IgA autoantibodies with limited somatic hypermutation in celiac disease intestinal lesions. *Nat Med* **18**, 441-445 (2012).
8. Robinson, W.H. Sequencing the functional antibody repertoire--diagnostic and therapeutic discovery. *Nat Rev Rheumatol* **11**, 171-182 (2015).
9. DeKosky, B.J. et al. High-throughput sequencing of the paired human immunoglobulin heavy and light chain repertoire. *Nat Biotechnol* **31**, 166-169 (2013).
10. Busse, C.E., Czogiel, I., Braun, P., Arndt, P.F. & Wardemann, H. Single-cell based high-throughput sequencing of full-length immunoglobulin heavy and light chain genes. *Eur J Immunol* **44**, 597-603 (2014).
11. McDaniel, J.R., DeKosky, B.J., Tanno, H., Ellington, A.D. & Georgiou, G. Ultra-high-throughput sequencing of the immune receptor repertoire from millions of lymphocytes. *Nature protocols* **11**, 429-442 (2016).
12. Wang, B. et al. Functional interrogation and mining of natively paired human VH:VL antibody repertoires. *Nat Biotechnol* **36**, 152-155 (2018).
13. Adler, A.S. et al. A natively paired antibody library yields drug leads with higher sensitivity and specificity than a randomly paired antibody library. *MAbs* **10**, 431-443 (2018).
14. Czerkinsky, C.C., Nilsson, L.A., Nygren, H., Ouchterlony, O. & Tarkowski, A. A solid-phase enzyme-linked immunospot (ELISPOT) assay for enumeration of specific antibody-secreting cells. *Journal of immunological methods* **65**, 109-121 (1983).
15. Story, C.M. et al. Profiling antibody responses by multiparametric analysis of primary B cells. *Proc Natl Acad Sci U S A* **105**, 17902-17907 (2008).
16. El Debs, B., Utharala, R., Balyasnikova, I.V., Griffiths, A.D. & Merten, C.A. Functional single-cell hybridoma screening using droplet-based microfluidics. *Proceedings of the National Academy of Sciences* **109**, 11570-11575 (2012).
17. Shembekar, N., Hu, H., Eustace, D. & Merten, C.A. Single-Cell Droplet Microfluidic Screening for Antibodies Specifically Binding to Target Cells. *Cell Rep* **22**, 2206-2215 (2018).
18. Bradbury, A.R., Sidhu, S., Dubel, S. & McCafferty, J. Beyond natural antibodies: the power of in vitro display technologies. *Nat Biotechnol* **29**, 245-254 (2011).
19. Fitzgerald, V. & Leonard, P. Single cell screening approaches for antibody discovery. *Methods* **116**, 34-42 (2017).
20. Klein, A.M. et al. Droplet barcoding for single-cell transcriptomics applied to embryonic stem cells. *Cell* **161**, 1187-1201 (2015).
21. Gurney, M.E., Heinrich, S.P., Lee, M.R. & Yin, H.S. Molecular cloning and expression of neuroleukin, a neurotrophic factor for spinal and sensory neurons. *Science* **234**, 566-574 (1986).
22. Liotta, L.A. et al. Tumor cell autocrine motility factor. *Proc Natl Acad Sci U S A* **83**, 3302-3306 (1986).
23. Park, C.S. et al. Therapeutic targeting of tetraspanin8 in epithelial ovarian cancer invasion and metastasis. *Oncogene* **35**, 4540-4548 (2016).
24. Anna, S.L., Bontoux, N. & Stone, H.A. Formation of dispersions using "flow focusing" in microchannels. *Appl Phys Lett* **82**, 364-366 (2003).
25. Mazutis, L. et al. Single-cell analysis and sorting using droplet-based microfluidics. *Nat Protoc* **8**, 870-891 (2013).
26. Abate, A.R., Chen, C.H., Agresti, J.J. & Weitz, D.A. Beating Poisson encapsulation statistics using close-packed ordering. *Lab Chip* **9**, 2628-2631 (2009).
27. Odegard, V.H. & Schatz, D.G. Targeting of somatic hypermutation. *Nat Rev Immunol* **6**, 573-583 (2006).
28. Roost, H.P. et al. Early high-affinity neutralizing anti-viral IgG responses without further overall improvements of affinity. *Proc Natl Acad Sci U S A* **92**, 1257-1261 (1995).
29. Murugan, R. et al. Clonal selection drives protective memory B cell responses in controlled human malaria infection. *Sci Immunol* **3(20)** (2018).
30. Sorouri, M., Fitzsimmons, S.P., Aydanian, A.G., Bennett, S. & Shapiro, M.A. Diversity of the Antibody Response to Tetanus Toxoid: Comparison of Hybridoma Library to Phage Display Library. *PloS one* **9**, e106699-106614 (2014).

## FIGURES



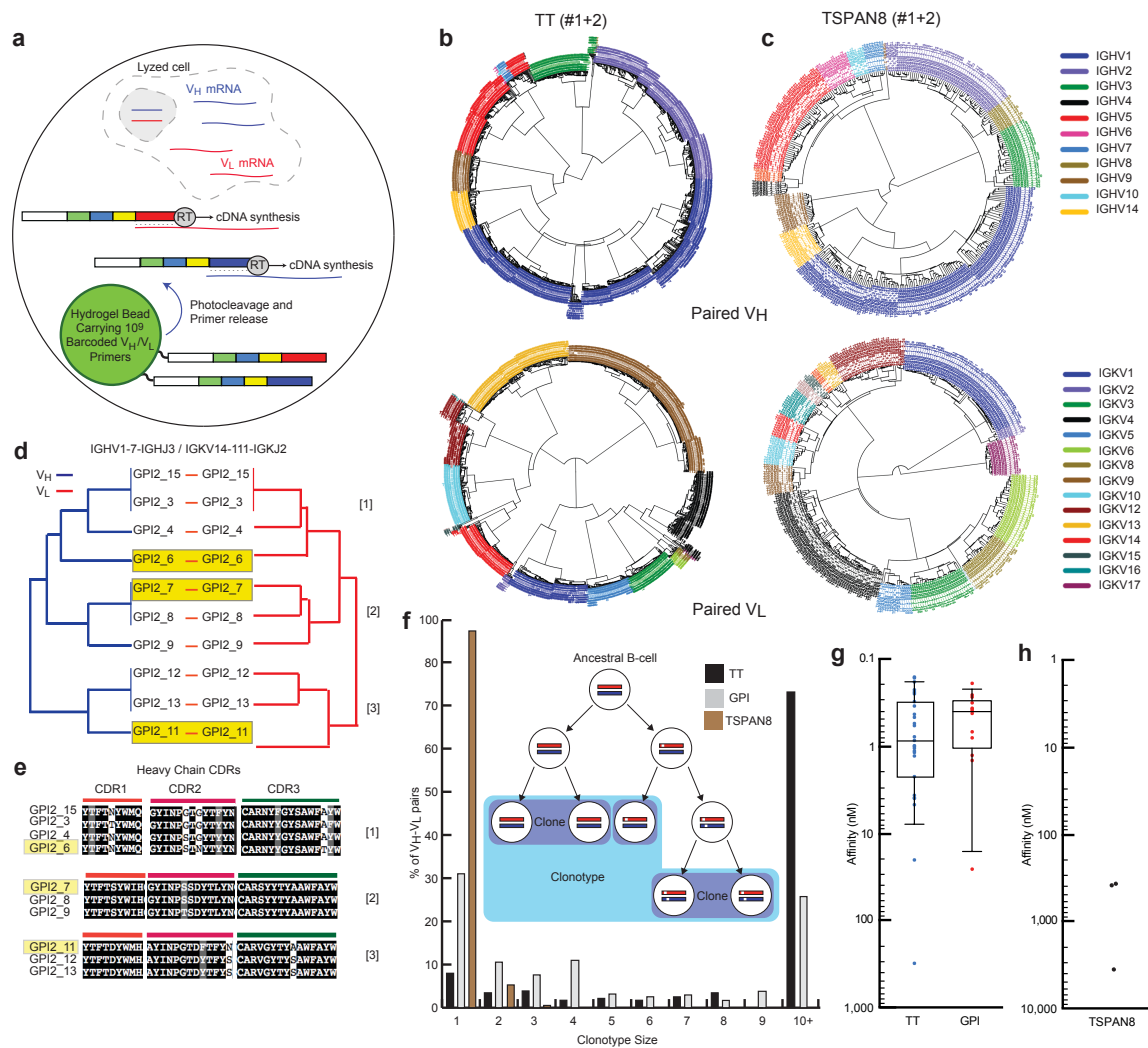
**Figure 1** The CelliGO process. **(a)** Schematic of sorting based on IgG binding to the soluble antigens Tetanus Toxoid (TT) and Glucose-6-Phosphate Isomerase (GPI) (top panel) or to the membrane antigen Tetraspanin-8 (TSPAN8) (middle panel), followed by paired  $V_H$ - $V_L$  sequencing of the sorted cells (bottom panel). See text for details. **(b)** Images of microfluidic operations. From left to right: droplet production, droplet sorting and co-compartmentalization of single-cells with single beads in droplets (see also Supplementary Movie). Scale bars, left and middle 100  $\mu$ m, right 40  $\mu$ m. **(c)** CelliGO timeline. The total time for the entire process from cell harvesting to antibody validation takes 20 days, including only 12h from cell harvesting to NGS.



**Figure 2** Phenotypic screening and sorting of IgG-secreting cells. (a, e) Binding assay in droplets. The droplets are scanned one-by-one as they pass the laser line, producing a 1.7 ms time trace of fluorescence signals in 4 colors (blue, green, orange, red) for each droplet. The distribution of the fluorescent signal of (a) the soluble antigen or (e) the reporter cell expressing the membrane antigen (green fluorescent), and of the secreted IgG (detected by red fluorescent F(ab')<sub>2</sub> anti-IgG Fc) are monitored in real time for each droplet at ~ 600 Hz. A schematic of the time trace of both signals is shown: (a) from left to right, for a drop containing no IgG-secreting cell (both green and red signals are uniformly distributed across the drop), a drop containing a cell secreting an IgG that does not bind the antigen (the secreted IgG is captured on the beads and the red fluorescent F(ab')<sub>2</sub> anti-IgG Fc is relocated to the bead line), and a drop containing a cell secreting an antigen-binding IgG (the secreted IgG is



captured on the beads and both the red fluorescent F(ab')<sub>2</sub> anti-IgG Fc and the green fluorescent antigen are relocated to the beadline), and (e) left, for a drop containing a cell secreting an IgG that does not bind the reporter cell and right, a drop containing a cell secreting an IgG that binds the reporter cell (the secreted IgG is captured on the green fluorescence reporter cell, the red fluorescent F(ab')<sub>2</sub> anti-IgG Fc is relocated to the reporter cell and the red and green peaks overlap). (b, f) Experimental time traces recorded for droplets analyzed at ~ 600 Hz and corresponding to the examples schematized in panels a and e for (b) the TT antigen and (f) TSPAN8<sup>+</sup> M300.19 cells. Droplets are sorted if they display a green peak (green line, (b) relocated antigen or (f) TSPAN8<sup>+</sup> M300.19 cells), and a red peak (red line, relocated antibody) that co-localize. Blue fluorescence (blue line) and orange fluorescence (orange line) signals are used to identify droplets containing (b) purified control antibodies (n=5) or (f) Calcein Violet-labeled splenocytes and control antibodies (n=3), respectively (see Online Methods). (c, g) Bright field and epifluorescence micrographs of a sorted droplet containing a splenocyte secreting (c) a TT-binding IgG (n=10) or (g) a TSPAN8<sup>+</sup> M300.19 cell-binding IgG (n=1). (d, h) Plots of maximum green versus maximum red fluorescence in each droplet containing cells from (d) a TT-immunized mouse or (h) a TSPAN8-immunized mouse. The sorted droplets are indicated as red dots. See Supplementary Fig. 2 for further details of the assay and signal analysis and Supplementary Fig. 3-4 for the full gating strategy.



**Figure 3** Sequencing, expression and characterization of IgGs expressed by sorted cells. **(a)** In-droplet single-cell barcoding of  $V_H$  and  $V_L$  cDNA. Cells from sorted droplets were recovered and re-compartmentalized individually in new droplets in the presence of reverse transcription reagents and a hydrogel bead bearing  $\sim 10^9$  copies of  $V_H$ - $V_L$  RT primers tagged with a bead-specific DNA barcode, which was released from the hydrogel beads by UV photocleavage. After breaking the emulsion, the pooled, barcoded  $V_H$ - $V_L$  cDNAs were amplified and processed for NGS. **(b, c)** Diversity of V-gene families use in paired  $V_H$ - $V_L$  sequences from **(b)** the TT-immunized mice (TT#1+2) and **(c)** the TSPAN8-immunized mice (TSPAN8#1+2). Phylogenies were created using the full-length amino acid sequences of the V regions. Sequences are color-coded by the closest germline V gene family. See also Supplementary Figs. 13-15. **(d)** Cross-phylogeny of 10 unique paired  $V_H$ - $V_L$  sequences from the GPI#2 sort derived from the same recombined germline V(D)J genes, illustrating coevolution of  $V_H$  and  $V_L$  during affinity maturation. **(e)** The sequences of the CDR regions

from the paired  $V_H$ - $V_L$  sequences presented in panel d. One candidate from each cluster was expressed and tested for binding to GPI (shaded yellow boxes): the three IgGs had similar affinity ( $K_d = 0.195$  to  $0.375$  nM). **(f)** Histogram of frequency (percentage) of paired  $V_H$ - $V_L$  sequences versus clonotype size. The clonotype size is equal to the number of non-redundant  $V_H$ - $V_L$  sequences derived from the same ancestral  $V_H$ - $V_L$  germline recombination event (i.e. derived from the same germline V-J genes, with CDR3s of the same length and differing by at least one amino acid across the variable region). *Inset*: Schematic of IgG clustering based on clonal expansion and somatic hypermutation of a single ancestral cell. Germline  $V_H$  and  $V_L$  genes are indicated by red and blue rectangles, and vertical white bars indicate mutations. Dark blue regions define clusters of clones ( $V_H$ - $V_L$  pairs with exactly the same amino acid sequence but associated with different barcodes and therefore derived from different cells), (size = 2, 2 and 1). The number of dark blue regions in a light blue region defines the clonotype size (size = 3). See also Supplementary Table 4. **(g)** Affinities of purified anti-TT IgGs (EC50 by ELISA; n=27) and anti-GPI IgGs ( $K_d$  by bio-layer interferometry; n=13) expressed from cloned genes identified by sequencing of TT-immunized mice and GPI-immunized mice, respectively, tested for binding to the immunogen. Boxes extend from the 25<sup>th</sup> to 75<sup>th</sup> percentile, whiskers for the 10<sup>th</sup> to 90<sup>th</sup> percentile and the line in the middle of the box is the median. **(h)** Affinities of purified anti-TSPAN8 IgGs ( $K_d$  by flow cytometry; n=3) expressed from cloned genes identified by sequencing of TSPAN8-immunized mice. See also Supplementary Table 3.

## ONLINE METHODS

**Immunization.** BALB/c mice (Janvier Labs, age 7-10 weeks at start) were immunized intraperitoneally with 10 µg of formaldehyde-inactivated tetanus toxoid (TT; Statens Serum Institute, Denmark) or 100 µg of rabbit glucose-6-phosphate isomerase (GPI; 89% homologous to mouse GPI; Sigma) emulsified in complete Freund's adjuvant (Sigma F5881) mixed 1:1 with 0.9% (w/v) NaCl (Versylene® Fresenius) for primary immunization (day 0), or emulsified in incomplete Freund's adjuvant (Sigma, F5506) mixed 1:1 with 0.9% (w/v) NaCl for secondary and tertiary immunization (day 14, 42), or only in 0.9% NaCl for boosting (day 46). BALB/c mice were injected intraperitoneally with  $2 \cdot 10^6$  human TSPAN8-transfected m300.19 mouse pre-B cells (originating from the outbred NIH Swiss mouse) in PBS on days 1, 15, 29, 43 and, if titers were estimated to be too low, again on days 57 and 71. Experiments using mice were validated by i) the CETEA ethics committee number 89 (Institut Pasteur, Paris, France) under #2013-0103, and by the French Ministry of Research under agreement #00513.02, and ii) the French ethics committee number 80 (CellVax) under reference APAFIS#6063-2016071213009170 v5.

**Extraction of cells from spleen and B cell enrichment.** Cells were extracted from spleens of mice, harvested at day 46 of the immunization schedule for soluble antigens (4 days after boosting), as described<sup>2</sup>, or at day 61 for the membrane-bound antigen (4 days after boosting). On average 100-200 million cells were obtained per spleen following red blood cell lysis. Cells from the B cell lineage (including IgG-secreting cells) were enriched using the Pan B cell isolation kit II (Miltenyi Biotec) as described<sup>2</sup> and further enriched by depletion with anti-IgM (Miltenyi, Cat#130-095-904) and anti-IgD (Miltenyi, Cat#130-101-899) antibodies. We observed a ~6-fold enrichment of CD138<sup>+</sup> cells by flow cytometry (plasmablast and plasma cells).

**Aqueous phase I: Preparation of cells for compartmentalization in droplets.** Cell suspensions were prepared for compartmentalization in droplets at 4°C as described<sup>2</sup>, and resuspended in DMEM/F12 supplemented with 0.1% Pluronic F68 (LifeTechnologies), 25 mM HEPES pH 7.4 (LifeTechnologies), 5% HyClone Super Low IgG Defined Fetal Bovine serum (GE Healthcare), and 1% Pen/Strep (ThermoFisher), so as to achieve a  $\lambda$  (mean number of cells per droplet) of ~0.3 for B cells in soluble antigen binding assays and of ~0.6 for B cells in cell-based assays.

## **Aqueous phase II: Preparation of bioassay reagents and reporter cells for**

**compartmentalization in droplets.** For the soluble antigens (TT and GPI), paramagnetic nanoparticles (Strep Plus, 300 nm, Ademtech) were coated with biotinylated V<sub>H</sub>H anti-mouse kappa light chain as described<sup>2</sup>, and resuspended in working buffer containing rabbit F(ab')<sub>2</sub> anti-mouse IgG Fc-specific (AlexaFluor647-labeled, Jackson ImmunoResearch) diluted to 75 nM and TT (purified monomer, AlexaFluor488-labeled) or GPI (AlexaFluor488-labeled) diluted to 50 nM, (resulting in final in-droplet concentrations of 37.5 nM anti-mouse IgG antibody and 25 nM antigen, respectively). For the membrane-bound antigen mouse reporter cells M300.19 stably expressing TSPAN8 transfected, pre-labeled with CalceinAM Green, were resuspended as defined above in media containing rabbit F(ab')<sub>2</sub> anti-mouse IgG Fc-specific (AlexaFluor647-labeled, Jackson ImmunoResearch) diluted to 75 nM supplemented with 23.6% v/v Nycodenz and to achieve a  $\lambda$  (mean number of cells per droplet) of  $\sim 1.5$  for reporter cells in cell-based assays.

**Microfluidic chips.** Separate microfluidic chips were used: i) to compartmentalize single cells with bioassay reagents in droplets, or to co-compartmentalize IgG-secreting cells with reporter cells expressing a membrane antigen and bioassay reagents in droplets; ii) to sort droplets by fluorescence-activated dielectrophoresis<sup>31</sup> iii) to produce hydrogel beads, and iv) to compartmentalize single sorted cells with single hydrogel beads (Supplementary Fig. 26). All chips were manufactured by soft-lithography in poly-dimethylsiloxane (PDMS)<sup>32</sup> (Sylgard) as described<sup>25</sup>. Masters were made using one or two layers of SU-8 photoresist (MicroChem) depending on the design. List depth of the photoresist layers for devices i and ii was  $40 \pm 1 \mu\text{m}$  and for devices iii was  $55 \pm 1 \mu\text{m}$ . For device iv, the first layer ( $70\text{-}75 \mu\text{m}$  deep) was for the hydrogel bead inlet and the second layer ( $130\text{-}145 \mu\text{m}$  deep) was for the cell inlet, RT inlet and outlet. Electrodes were prepared by melting a 51In 32.5Bi 16.5Sn alloy (Indium Corporation of America) into the electrode channel<sup>33</sup>.

**Droplet production, collection and incubation.** Aqueous phases I and II were co-flowed and partitioned into droplets using hydrodynamic flow-focusing<sup>24</sup> in dripping mode on a microfluidic chip with a nozzle  $15 \mu\text{m}$  wide,  $40 \mu\text{m}$  deep and  $10 \mu\text{m}$  long (Supplementary Fig. 26a). The continuous phase comprised 2-3% (w/w) 008-FluoroSurfactant (RAN Biotechnologies) in Novec HFE7500 fluorinated oil. The flow rates were adjusted to generate monodisperse droplets of  $40 \pm 2 \text{ pl}$  volume (for assays with soluble antigens), or  $80 \pm 8 \text{ pl}$  (for

cell-based assays with membrane-bound antigen). Shortly after generation, the droplets were collected into a 10 mL hemolysis tube filled with 10 mL Novec HFE7500 fluorinated oil containing 0.1 % (w/w) 008-FluoroSurfactant.

- In the case of the soluble antigens (TT and GPI) the tube was inserted into a ring magnet (Diameter 25/13mm, 7Kg, Magnet-shop) to force the superparamagnetic nanoparticles to form an elongated aggregate (a “beadline”) inside each droplet. For TT sorts only, in addition to the screening emulsion containing splenocytes, 2 control emulsions (positive and negative) were produced, containing 10 nM purified anti-TT7 antibody<sup>2</sup> and 10 nM mouse anti-horseradish peroxidase antibody (Jackson Immuno Research, Cat#223-005-024), respectively; these were differentiated using Sulforodamine B (Sigma Aldrich #S1402-5G) and DY405 (Dyomics #405-00) to optically barcode the droplets in the positive and negative control emulsions, respectively (see Supplementary Fig. 3).

- In the case of the membrane-bound target TSPAN8, enriched B cells were co-compartmentalized in droplets with TSPAN8<sup>+</sup> M300.19 reporter cells. The TSPAN8<sup>+</sup> M300.19 reporter cells were pre-labeled with 300 nM Calcein AM Green (ThermoFisher, Cat#MP03224) and the enriched B cell population was pre-labeled with 300 nM Calcein AM Violet (ThermoFisher, Cat# C34858). In addition to the screening emulsion containing splenocytes, 2 control emulsions - negative and positive - were produced, containing reporter cells only, or reporter cells and 50 nM anti-hTSPAN8 51D3 IgG (Pfizer; see Supplementary Figure 4), respectively. These were differentiated using Sulforodamine B (Sigma Aldrich #S1402-5G) to optically barcode the droplets in the positive and negative control emulsions, respectively (see Supplementary Fig. 4).

**Microfluidic platform.** Droplet fluorescence analysis and sorting was performed on a dedicated droplet microfluidic station, similar to<sup>25</sup>, but including a fixed focus laser line (solid state lasers: 405 nm, 488 nm, 561 nm, 635 nm, Omicron) oriented parallel to the beadline for fluorescence analysis using PMT (Hamamatsu) bandpass filters of 440/40-25 nm, 525/40-25 nm, 593/46-25 nm, 708/75-25 nm.

**Gating strategy for droplet sorting.** The droplets were first gated to eliminate coalesced droplets and retain only droplets of the desired size (Supplementary Fig. 3a, 4a). Using the optical droplet barcoding the negative control droplets, positive control droplets and droplets containing splenocytes were detected (Supplementary Fig. 3b, 4b).

- For the soluble antigens (TT and GPI), fluorescence relocalization to the beadline in droplets was measured by plotting the maximum peak fluorescence signal ( $F_p$ ) in a droplet against the integrated fluorescence signal ( $F_i$ ) from the droplet (Supplementary Fig. 3c-e). Fluorescence relocalization to the beadline results in an increase in the ratio  $F_p/F_i$  (Supplementary Fig. 2, 3c-h). Uncoalesced droplets containing splenocytes were sorted if they satisfied all of the following criteria: i) relocation of antigen (green fluorescence) to the beadline, ii) relocation of the F(ab')<sub>2</sub> anti-mouse IgG Fc (red fluorescence) to the beadline, and iii) colocalization of the green and red fluorescence peak (Supplementary Fig. 3g-i). The colocalization value,  $c$ , was thus bounded between 0 and 1, 1 being the perfect colocalization of the two peaks. Droplets were sorted if  $c > 0.95$ .

- A similar gating strategy was employed for the membrane antigen (TSPAN8) (Supplementary Fig. 4d-i), except that the droplets were first selected for the presence in the drop of both a reporter cell and an enriched B cell based on their fluorescent labeling (Supplementary Fig. 4c). Relocalization of red fluorescence to the reporter cell results in an increase in the ratio  $F_p/F_i$  (Supplementary Fig. 2). Uncoalesced droplets containing splenocytes were sorted if they satisfied all of the following criteria: (i) presence of a reporter cell overexpressing TSPAN8 (green fluorescence) in the droplet, ii) presence of an enriched B cell (violet fluorescence) in the droplet, iii) relocation of the F(ab')<sub>2</sub> anti-mouse IgG Fc (red fluorescence) on the reporter cell, and iv) colocalization of the green and red fluorescence peaks (Supplementary Fig. 4i).

For all antigens the colocalization parameter,  $c$ , was calculated from the time interval in between the peaks in the red and green fluorescence channels,  $t_p$ , and the time interval from the beginning to the end of the droplet,  $t_d$ :  $c = 1 - (t_p/t_d)$ . The colocalization value was thus bounded between 0 and 1, 1 being the perfect colocalization of the two peaks. Droplets were sorted if  $c > 0.95$  for TT and GPI, and  $c > 0.92$  for TSPAN8.

**Droplet sorting and cell recovery.** For the soluble antigens (TT and GPI), droplets were sorted by dielectrophoresis as described<sup>31</sup>, while applying a magnetic field oriented perpendicularly to the droplet flow using permanent magnets (K&J Magnetics). The device (Supplementary Fig. 26b) can sort droplets into up to two 2 bins, by activation of the electrodes above or beneath the channel; however, only one bin was used in this study. The inlet flows,  $Q_{em}$  for the emulsion and  $Q_{oil}$  for the oil, were adjusted to sort droplets at  $600 \text{ s}^{-1}$ : typical parameters for sorting were  $Q_{em} = 50 \text{ } \mu\text{L/h}$  (180 mBar),  $Q_{oil} = 50 \text{ } \mu\text{L/h}$  (180 mBar),  $F$  (the frequency of the sorting pulse) = 10 kHz,  $\tau_{sort}$  (the duration of the sorting pulse) = 2,000

$\mu\text{s}$  and  $U_{\text{sort}}$  (the peak-to-peak voltage applied across the electrodes) = 400 kVp-p. Sorted droplets were collected in a 1.5 mL tube cooled to 4°C, containing an emulsion of 40,000 chinese hamster ovary (CHO) cells used as carrier cells during recovery. Cells were recovered by addition of 100  $\mu\text{L}$  of DMEM/F12 supplemented with 5% serum Low IgG, followed by 100  $\mu\text{L}$  of 1H,1H,2H,2H-Perfluoro-1-octanol (370533, Sigma) gently mixed and centrifuged at 700g for 10min at 4°C to favour complete phase separation.

**ELISpot.** IgG-secreting cells were enumerated after TT sorts and TSPAN8 sorts by ELISpot using the MabTech ELISpotPLUS Kit (MabTech#3825-2AW-Plus) on an AID ELISpot Reader. 96-well ELISpot plates were coated with 0.5 mg/mL anti-IgG mouse mAb (Mabtech#3825-2AW-Plus) or, specifically for cells from TT sorts with 30 ug/mL antigen Tetanus Toxoid (Statens Serum Institute, Denmark), and IgG secretion was detected using 0.5 mg/mL biotinylated anti-IgG mouse mAb (Mabtech#3825-2AW-Plus).

**Labeling and preparation of sorted cells for single-cell sequencing.** Sorted cells and carrier cells were labeled for 20 min at room temperature in the dark with 1  $\mu\text{M}$  Calcein AM (C3099, Thermo). The cells were then washed in 400  $\mu\text{L}$  of 0.1% Pluronic F-68, 25mM Hepes-KOH, 5% v/v low IgG serum, in DMEM/F12, spun for 5 min at 400 g at 4°C, then resuspended in 50  $\mu\text{L}$  of 1x PBS containing 21.82% (v/v) Optiprep density gradient solution (Sigma) and 0.01mg/ml BSA.

**Production of barcoded hydrogel beads.** Polyacrylamide hydrogel beads of 60  $\mu\text{m}$  diameter were produced by polymerization in droplets made using a microfluidic device (Supplementary Fig. 26c) using a method similar to that previously described<sup>20,34</sup>. However, barcoded primers were then added to the beads by split-and-pool synthesis using ligation rather than primer extension (Supplementary Fig. 7a,b). One million beads, each carrying  $\sim 10^9$  copies of a dsDNA oligo with a 5'-overhang, complementary to the first index 5'-overhang sequence, a photo-cleavable site and the T7-SBS12 sequence were distributed into 96-wells of a microtiter plate, each well containing 10  $\mu\text{L}$  of 5  $\mu\text{M}$  double-stranded DNA with a different first index (index A), and a complementary 5'-overhang to the first DNA at one end and a different 5'-overhang at the other end (Supplementary Table 5), and ligated for 15 minutes at 23°C using T7 DNA ligase (NEB) according to the manufacturer's instructions. The hydrogel beads were then pooled, washed as described<sup>34</sup> and re-distributed as above into the wells of a second microtiter plate, each containing a double-stranded DNA with a



different second (index B), a complementary 5'-overhang to index A at one end and a different 5'-overhang at the other end, which was ligated to Index A. Repeating this splitting and pooling process 3 times in total (adding 3 indexes) results in  $96^3$  combinations, which generates  $\sim 10^6$  different barcodes. After adding the last index, the beads were pooled, and a mixture of double-stranded DNA molecules with a complementary 5'-overhang to index C, containing gene-specific primer regions (GSPs) complementary to the regions coding for the mouse  $J_H$  regions and the mouse  $C_k$  domain (Supplementary Fig. 9), were ligated to the barcodes on the beads. The second strand of the primer was then removed by incubating 2 min at 22°C with 300 mM NaOH. Each hydrogel bead ends up with a total of  $\sim 10^9$  primers carrying the same bead-specific barcode. See Supplementary Figure 7 for further details and quality control.

**Single-cell barcoded cDNA synthesis.** Single sorted cells (and CHO carrier cells) were co-compartmentalized in droplets with single barcoded hydrogel beads and lysis and reverse transcription reagents using a microfluidic device (Supplementary Fig. 26d) as described<sup>34</sup> (see Supplementary Fig. 8). Droplets of  $\sim 1$  nL volume were formed at  $250\text{ s}^{-1}$ . The droplets were collected in a 1.5mL tube containing HFE-7500 and 0.1% surfactant, UV photo-cleaved for 90 seconds (OmniCure ac475-365) and incubated at 50°C for cell lysis and cDNA synthesis.

**Sequencing library preparation.** The emulsion containing the barcoded cDNA was broken by adding one volume of 1H,1H,2H,2H-Perfluoro-1-octanol (370533, Sigma). The pooled, barcoded cDNAs were run on a Novex TBE-Urea 6% polyacrylamide gel and cDNAs above 250 bases were excised, purified using a Qiaex II kit (20051, Qiagen), eluted in 10 mM Tris-HCl pH 8.0, further purified with Agencourt RNA CleanUp beads (A63987; Beckman) at 1x ratio (v/v), and eluted in 40  $\mu\text{L}$  DNase/RNase free  $\text{H}_2\text{O}$ . The sequencing library was generated by two-step nested PCR using GoTaq Polymerase (Promega), external (PCR1) and internal (PCR2) reverse primers (T7 primer followed by Illumina TruSeq indexed primer) and external (PCR1) and internal (PCR2) forward primers specific to the  $V_H$  and  $V_k$  leader and framework sequences, as described<sup>35</sup> with minor modifications (Supplementary Fig. 9). Briefly, the PCR2 forward primers described in ref.<sup>35</sup> which have a binding site downstream the first two amino-acids of the framework sequences were replaced by the corresponding PCR1 forward primers during the PCR2 reaction.

**Sequencing.** Final products were sequenced on an Illumina MiSeq (2x300 bp paired-end reads), which allowed sequencing of the entire  $V_H$  and  $V_L$  domain as well as the barcode sequence (Supplementary Fig. 9; Supplementary Table 5).

**Bioinformatic data processing.** A bioinformatics pipeline was developed to allow sequence read trimming, merging, barcode extraction and clustering and antibody sequence characterization and filtering. For all barcode clusters containing productive  $V_H$  and  $V_L$  made up of at least 40 reads each (a threshold based on pairing rates from mixtures of mouse hybridomas, see Supplementary Fig. 10a), the most abundant  $V_H$  and  $V_L$  were considered paired (Supplementary Fig. 11).  $V_H$ - $V_L$  pairs were assumed to derive from the same progenitor rearrangement event if they shared the same V and J gene assignments and had CDR3s of the same length. Somatic mutations were identified by comparing the  $V_H$  and  $V_L$  sequences to the closest germline V-gene. Detailed bioinformatics methods can be found in the Supplementary Information.

**Antibody selection for expression.** From the TT sort, 42  $V_H$ - $V_L$  pairs with the highest numbers of reads were selected for expression. From the GPI sort, 14  $V_H$ - $V_L$  pairs were selected for expression among several  $V_H$ - $J_H$  recombinations representing various  $V_H$ -gene families, associated with (if possible) different  $V_L$ - $J_L$  recombinations, and including the number amino acid differences compared to the closest germline V-gene (1-4, 4-10 and >10 amino acid changes in the V-gene region). Four  $V_H$ - $V_L$  pairs presenting germline configurations were also selected for antibody expression from the GPI sort. From the two TSPAN8 sorts, 20  $V_H$ - $V_L$  pairs were chosen for expression, dividing candidates into three groups, depending on the amino acid differences compared to the closest germline V-gene (<4, 4-9,  $\geq$ 10 changes in the V-region) and in cases of multi-member families choosing only one IgG per family.

**Gene synthesis and cloning of antibody V genes.**  $V_H$  and  $V_L$  gene sequences were codon optimized for expression in human cells using the GeneArt Strings DNA optimization software, and optimized DNA fragments were synthesized by GeneArt (Thermo Fisher Scientific). The synthesized  $V_H$  and  $V_L$  DNA fragments contained a first adapter sequence (5'-ACAGCTACAGGCGCGCACTC-3', corresponding to the V-gene proximal region of the mouse leader sequence) and a second adapter sequence (5'-

GCGTCGACCAAGGGCCCATCG-3' or 5'-CGTACGGTGGCTGCACCATCT-3', overlapping with the V-gene proximal region of the mouse CH1 and C $\kappa$  regions, respectively) to allow cloning of the V-genes. Synthesized DNA fragments were inserted into corresponding pHFB-IgG1 and pHFB- $\kappa$  vectors (HiFiBiO heavy and light chain expression plasmids) in frame with the human IgG1 or kappa constant regions by In-Fusion DNA assembly according to manufacturer's instructions, and transformed into *E. coli* Stellar Chemical Competent Cells (Clontech). Plasmids were purified by mini-prep and analyzed by Sanger sequencing (Wyzer Biosciences).

**Antibody expression.** Correctly cloned V<sub>H</sub>-V<sub>L</sub> pairs were co-transfected into serum-free Expi293F suspension cell cultures (Life Technologies) using ExpiFectamine reagents (Life Technologies) following the manufacturer's instructions. Transfection reactions were incubated for 4 days, at which point supernatants were collected, clarified by centrifugation at 1400 rpm for 15 min at RT, and stored at 4°C. For affinity measurements, IgGs were purified by affinity chromatography using an AKTA pure FPLC instrument (GE Healthcare) and HiTrap Protein G Column (GE Healthcare). After purification, IgGs were desalted with HiTrap Desalting Column (GE Healthcare).

**ELISA assays.** Maxisorp plates were coated with 100  $\mu$ L of 1  $\mu$ g/ml TT or 5  $\mu$ g/ml GPI in PBS overnight at 4°. Plates were blocked with PBS containing 1% BSA. Naive, immune mice serums, IgGs from culture supernatants or control IgG at 1  $\mu$ g/ml were diluted in series. Anti-human Fc-HRP at 0.05  $\mu$ g/mL (Bethyl) was used for detection. The assay was revealed with OPD (ortho-phenylene diamine), stopped with 2M H<sub>2</sub>SO<sub>4</sub>. and absorbance measured at 492 nm.

*Affinities of the recombinant anti-TT antibodies identified by CelliGO:* EC<sub>50</sub>s for anti-TT IgGs were measured using the same protocol as described above.

**Bio-layer interferometry and real-time surface plasmon resonance.** *Affinities of the recombinant anti-GPI antibodies identified by CelliGO:* Bio-layer interferometry with an anti-human IgG sensor on Octet systems (Pall ForteBio) was used to quantify antibody concentrations in supernatants of transfected Expi293F cells, and to measure affinities of the recombinant mouse-human chimeric anti-GPI antibodies. For the latter, biosensors were loaded with 1  $\mu$ g/mL antibody, association and dissociation signals (10 min per step) were measured against 5 different concentrations of GPI (1.85, 5.5, 16.6, 50 and 150 nM), in

duplicate. Curves were retrieved as raw data from the Octet Data Analysis software and processed in Scrubber2 prior to analyzing with the Biaevaluation software. IgG-free buffer and an irrelevant IgG antibody (here an anti-TT IgG) were used as reference and for background subtraction. Fitting curves and  $K_d$  values were obtained using the 1:1 Langmuir model.

*Affinity of serum from mice immunized against TT or immunized against GPI:* Real-time surface plasmon resonance (SPR) on a Biacore T200 instrument (GE Healthcare) was used to measure, at 25°C, the mean affinity of mouse serum against TT or GPI. IgG from mouse serum (diluted 1/500) was captured onto a Streptavidin sensor chip (GE Healthcare) pre-loaded with biotin-conjugated V<sub>H</sub>H anti-IgG-Fc multi species antibodies (ThermoFisher). Purified TT and GPI were injected at concentrations ranging from 7.8 to 1000 nM for 10 min at a rate of 20 µL/min, and dissociation followed for 10 min. For negative controls, TT antigen was injected onto immobilized IgG from anti-GPI sera and vice-versa, and used as a reference surface for calculating specific signals. The association and dissociation profiles were analyzed using the Biacore T200 evaluation software (v 3.0), assuming a 1:1 simple Langmuir interaction mechanism, yielding an average equilibrium dissociation constant ( $K_d$ ) for the anti-TT and anti-GPI IgGs contained in each serum.

**Characterization of anti-TSPAN8 IgG antibodies.** *Specificity:* Parental M300.19 cells and TSPAN8<sup>+</sup> M300.19 cells were incubated with 10 µg/mL recombinant IgG identified from TSPAN8 sorts or with 5 µg/mL anti-TSPAN8 mAb clone 43A6 (positive control) in ice-cold FACS Buffer (Ca<sup>2+</sup>, Mg<sup>2+</sup>-free PBS supplemented with 3% FBS + 0.1% w/v NaN<sub>3</sub>) for 30 min on ice. Cells were washed twice, incubated with goat anti-human IgG-PE (Southern Biotech #1030-09) for 30 min on ice, and resuspended in BD Cytotfix (BD Biosciences #554655) for 30 min on ice, and analyzed on a FACSCalibur flow cytometer (BD). *Affinity:* anti-TSPAN8 IgG antibodies that demonstrated specific binding to TSPAN8<sup>+</sup> M300.19 cells were fluorescently labeled with Alexa Fluor 488, incubated in serial dilutions with TSPAN8<sup>+</sup> M300.19 cells in ice-cold FACS Buffer (Ca<sup>2+</sup>, Mg<sup>2+</sup>-free PBS supplemented with 10% FBS) for 30 min on ice, and cell-bound fluorescence quantified using a MACSQuant flow cytometer (Miltenyi Biotech). The  $K_d$  values were calculated using the BIAevaluation software (Biacore) by fitting the data on antibody-concentration dependence of the fluorescent signal to Equation 1:

$$R = \frac{R_{max} \cdot C}{K_d + C} \quad (1)$$

where the response,  $R$ , is the mean fluorescent signal of the tested IgG minus the mean fluorescent signal of a negative control antibody (rituximab, an anti-human CD20 IgG),  $C$  is the antibody concentration and  $R_{max}$  is the fluorescent signal at infinite concentration.  $K_d$  values were calculated by fitting to the pooled data from triplicate experiments.

**Statistical methods developed for V gene usage comparisons.** Pairwise comparisons of matrices of same dimension (corresponding to usage of  $V_H$  genes,  $V_L$  genes, or  $V_H$ - $V_L$  combinations), as well as metric Multidimensional Scaling (MDS), were performed using Comat, an R tool available at <https://gitlab.pasteur.fr/gmillot/comat>. Comat was run as follows. We first used a normalization method called "proportional matrix equalization by adjustment to minimal frequency". Matrix case frequencies were adjusted so that the total absolute frequencies of the two compared matrices were equal. When  $n_1 > n_2$ , the equation used was:

$$n_{ij1} = n_{ij1} \times \frac{n_2}{n_1}$$

When  $n_2 > n_1$ , the equation used was:

$$n_{ij2} = n_{ij2} \times \frac{n_1}{n_2}$$

With  $n_{ij1}$ , the absolute frequency of matrix 1 case at the row  $i$  and column  $j$ ;  $n_{ij2}$ , the absolute frequency of matrix 2 case at the row  $i$  and column  $j$ ;  $n_1$ , the total absolute frequency of matrix 1 and  $n_2$ , the total absolute frequency of matrix 2. Such normalization was performed to focus the pairwise comparison on the matrix distributions (i.e. matrix proportions), without influence related to  $n_1$  and  $n_2$  differences, which depends on uncontrolled experimental fluctuations. This normalization method loses information, but it provides two matrices of similar experimental sensitivity. Then, the observed distance,  $D_{obs}$ , between the matrices was computed. The distance,  $D$ , between the two matrices was calculated using eq(1):

$$D = \left( \sum_{i=1}^k \sum_{j=1}^c \left| \frac{n_{ij1}}{n_1} - \frac{n_{ij2}}{n_2} \right| \right) / 2 \quad eq(1)$$

with  $k$ , the number of rows and  $c$ , the number of columns. Then, an approximate permutation test was developed. To approximate the distribution of the  $D$  statistic, the  $n_1$  frequencies were randomly spread in the matrix 1, and the same for  $n_2$  in the matrix 2 (here,  $n_1$  and  $n_2$  being equal due to normalization). Then, a theoretical statistic was computed using eq(1). This was repeated  $r = 100,000$  times, which provided a probability distribution of  $D$ , and, thus the probability to obtain the observed  $D_{obs}$  statistic. In a two-sided test, the  $p$  value was:

$$p = P(D \geq D_{obs}) \text{ eq(2)}$$

If  $p = 0$ , then  $p$  was estimated as  $p < 1/r$ , i.e.  $p < 0.00001$ . The same procedure was performed to compare the V genes frequencies (sum of the rows of two matrices of the same dimension). Statistical significance was set to  $P \leq 0.05$ . Type I error was controlled by correcting the  $p$  values according to the Benjamini & Hochberg method ("BH" option in the `p.adjust()` function of R)<sup>36</sup>. MDS results were obtained using the R package `ExPosition`<sup>37</sup>. To generate the 95% confidence intervals shown in the MDS figure, bootstrap of the initial combined  $V_H$ - $V_L$  matrix usage (size  $n$  sampled in the non-empty cases of the matrix) was performed 100 times, MDS of the 100 new distance matrices were added to the original MDS figure, and quantiles 0.025 and 0.975 of the 100 new positions of each initial dot was computed for each dimension.

## DATA AVAILABILITY

Processed sequencing files (quality trimmed, merged and barcode assigned) for all data generated in this study were deposited at Sequence Read Archive (SRA) under accession PRJNA529803.

## METHODS-ONLY REFERENCES

31. Baret, J.C. et al. Fluorescence-activated droplet sorting (FADS): efficient microfluidic cell sorting based on enzymatic activity. *Lab Chip* **9**, 1850-1858 (2009).
32. Duffy, D.C., McDonald, J.C., Schueller, O.J. & Whitesides, G.M. Rapid Prototyping of Microfluidic Systems in Poly(dimethylsiloxane). *Anal Chem* **70**, 4974-4984 (1998).
33. Siegel, A., Bruzewicz, D., Weibel, D. & Whitesides, G. Microsolidics: Fabrication of Three-Dimensional Metallic Microstructures in Poly(dimethylsiloxane). *Adv. Mater.* **19**, 727-733 (2007).
34. Zilionis, R. et al. Single-cell barcoding and sequencing using droplet microfluidics. *Nat Protoc* **12**, 44-73 (2017).
35. Rohatgi, S., Ganju, P. & Sehgal, D. Systematic design and testing of nested (RT-)PCR primers for specific amplification of mouse rearranged/expressed immunoglobulin variable region genes from small number of B cells. *Journal of immunological methods* **339**, 205-219 (2008).
36. R Core Team (R Foundation for Statistical Computing, Vienna, Austria; 2017).
37. Beaton, D., Chin Fatt, C.R. & Abdi, H. An `ExPosition` of multivariate analysis with the singular value decomposition in R. *Computational Statistics & Data Analysis* **72**, 176-189 (2014).