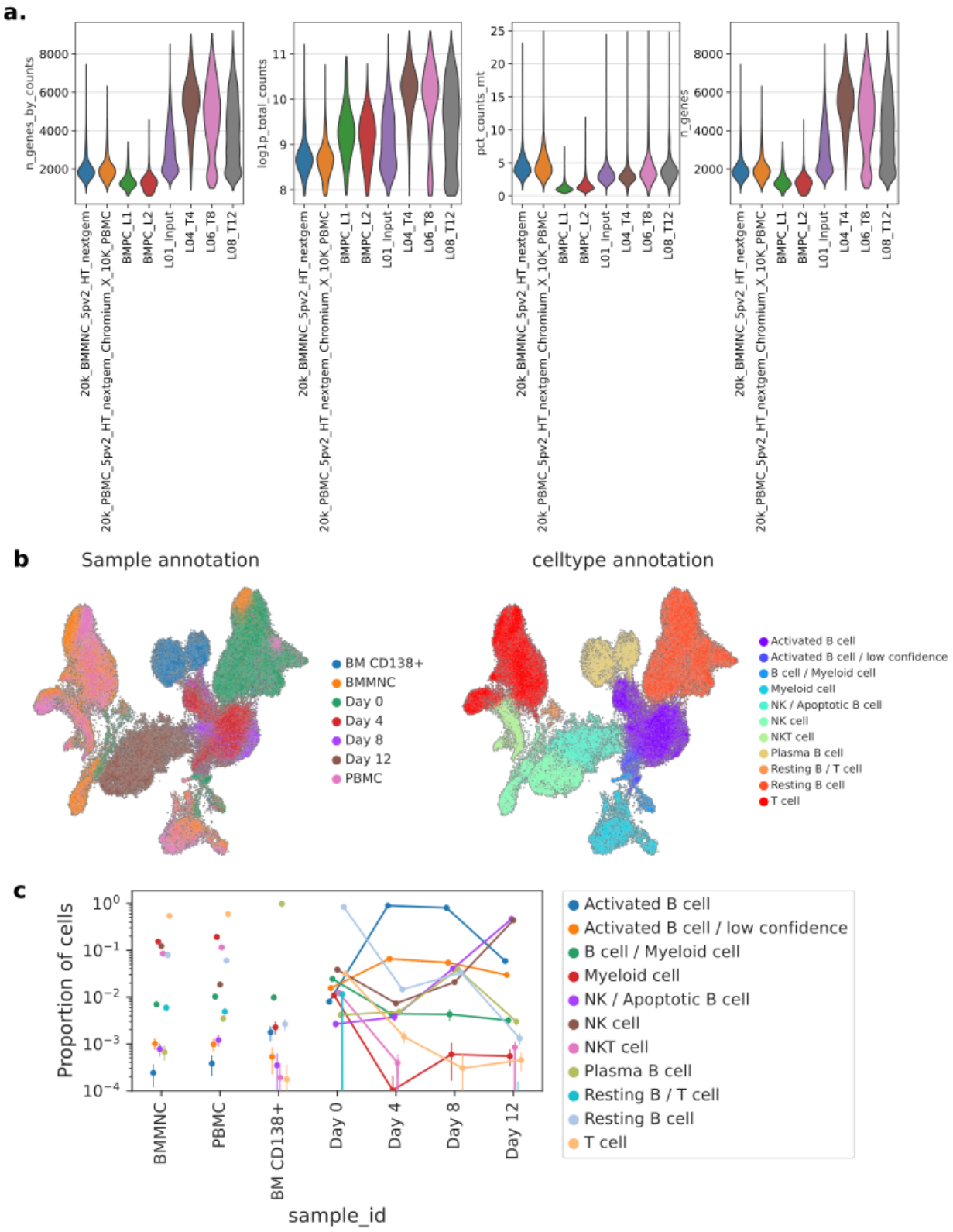
Supplementary Material:


a.
n_genes_by_counts
log1p_total_counts
pct_counts_mt
n_genes
20k_BMMNC_5pv2_HT_nextgem
20k_PBMC_5pv2_HT_nextgem_Chromium_X_10K_PBMC
BMPC_L1
BMPC_L2
L01_Input
L04_T4
L06_T8
L08_T12
b
Sample annotation
BM CD138+
BMMNC
Day 0
Day 4
Day 8
Day 12
PBMC
celltype annotation
Activated B cell
Activated B cell / low confidence
B cell / Myeloid cell
Myeloid cell
NK / Apoptotic B cell
NK cell
NKT cell
Plasma B cell
Resting B / T cell
Resting B cell
T cell
c
Proportion of cells
sample_id
BMMNC
PBMC
BM CD138+
Day 0
Day 4
Day 8
Day 12

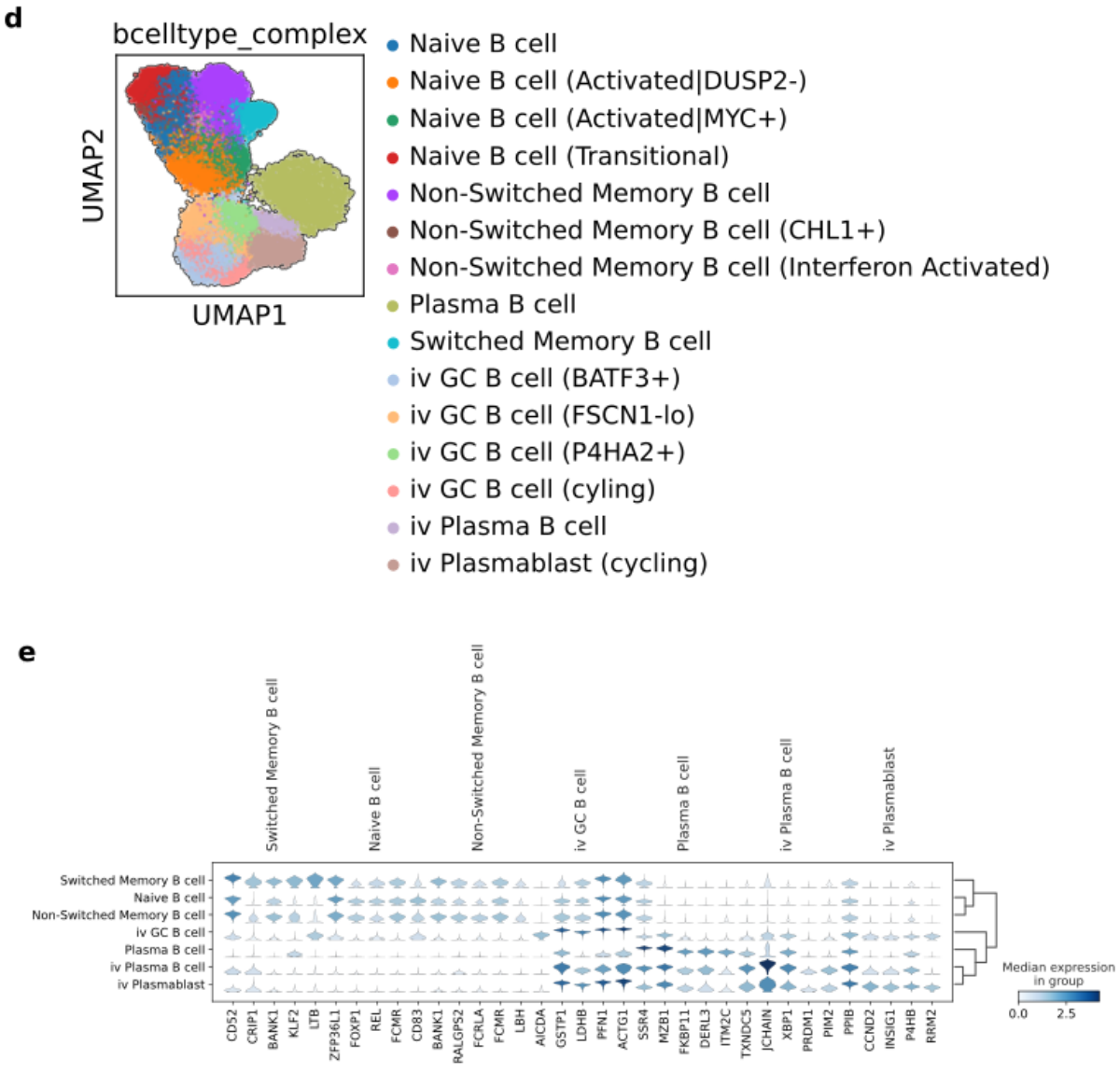


Figure S1. (A) Violin plots of quality metrics for each 10X genomics lane sequenced and/or analyzed (B) UMAPs of all cells in the dataset, colored as shown in legends. Contaminant cells in the B cell purifications cluster with the similar cells in unpurified fractions (C) Pointplot quantifying the proportion of celltypes in each sample_id. (D) Subclustering of B cell type annotations with greater leiden resolution. (E) Differentially expressed genes which define each B cell type.

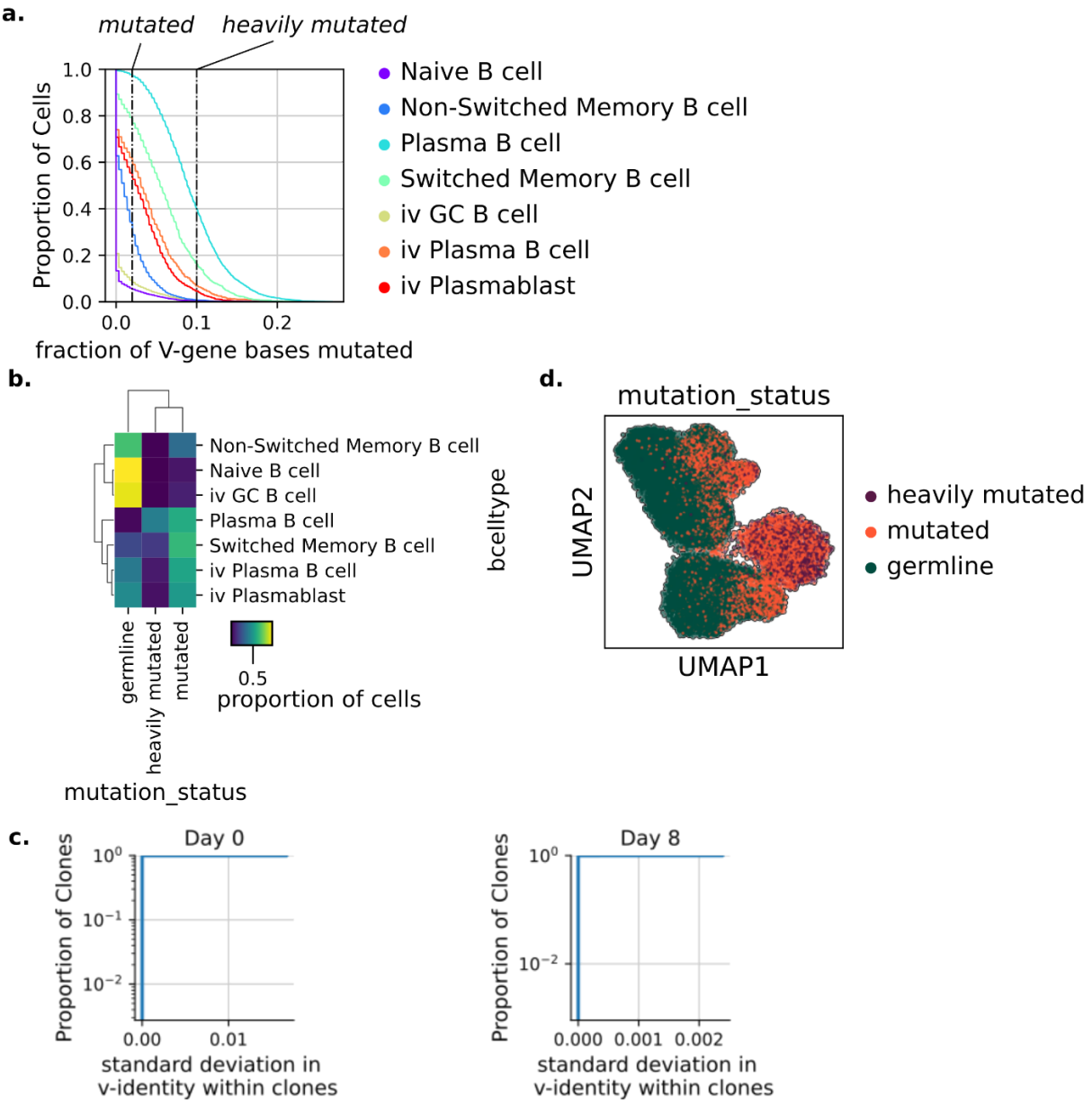


Figure S2. (A) Empirical cumulative distributions of the fraction of V-gene based mutated away from the germline V-gene, for each transcriptomically defined B cell type (B) A confusion matrix showing the concordance between the mutation status based on S2A and each B cell type label based on leiden clustering. (C) Empirical cumulative distributions of the standard deviation in v-identity within clones shows mutations are not collected within clones over the time course (D) UMAP plot colored by the mutation status assigned based on S2A.

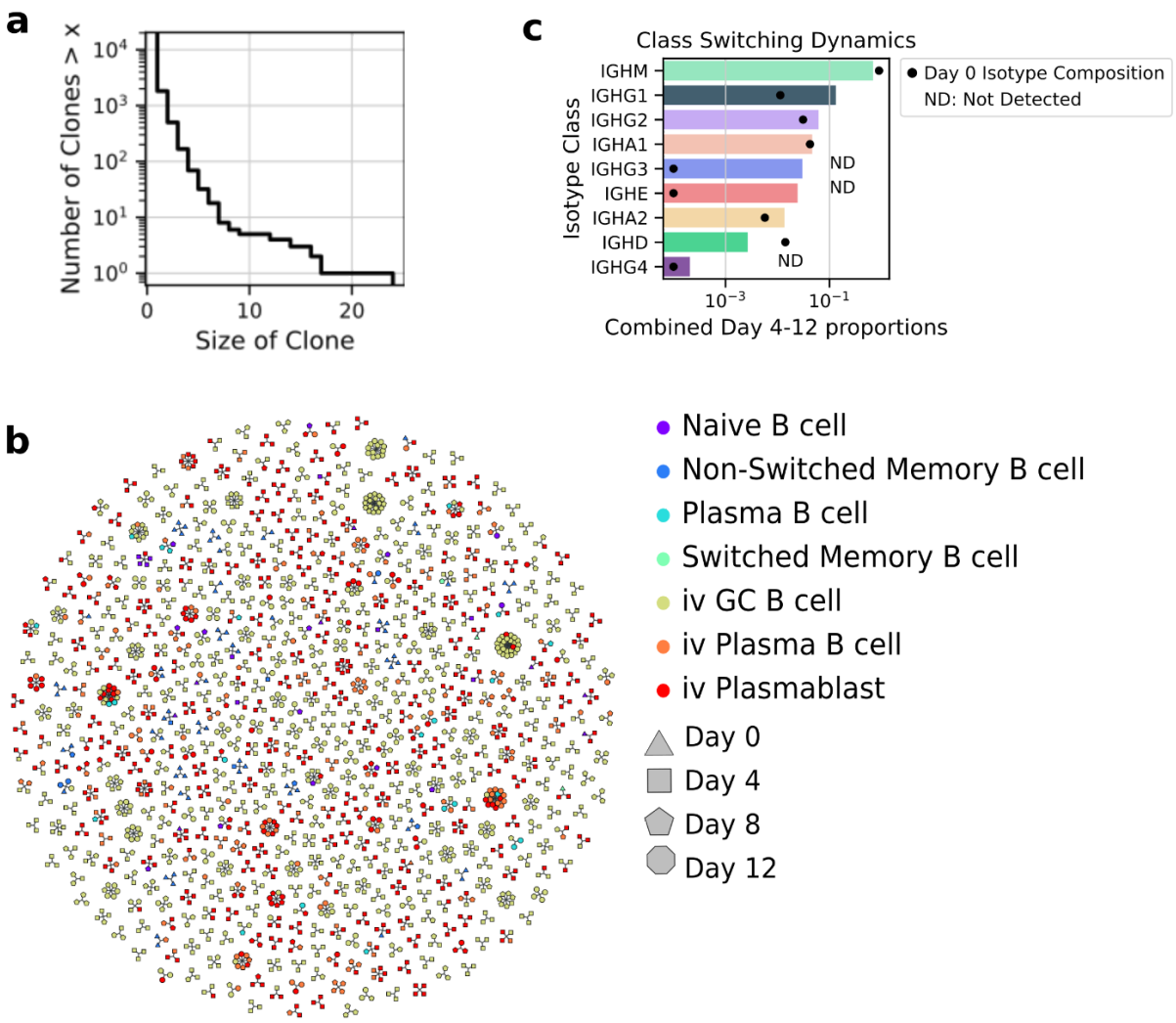


Figure S3. (A) Clone size distribution of the B cell population from the *in vitro* time-course (B) A graph based representation of the persistent clones: clones which were detected in multiple time points. © Class-switching (isotype-switching) dynamics during the culture. ND = Not Detected in the Day 0 population, a pseudo-count is added for visualization purposes.

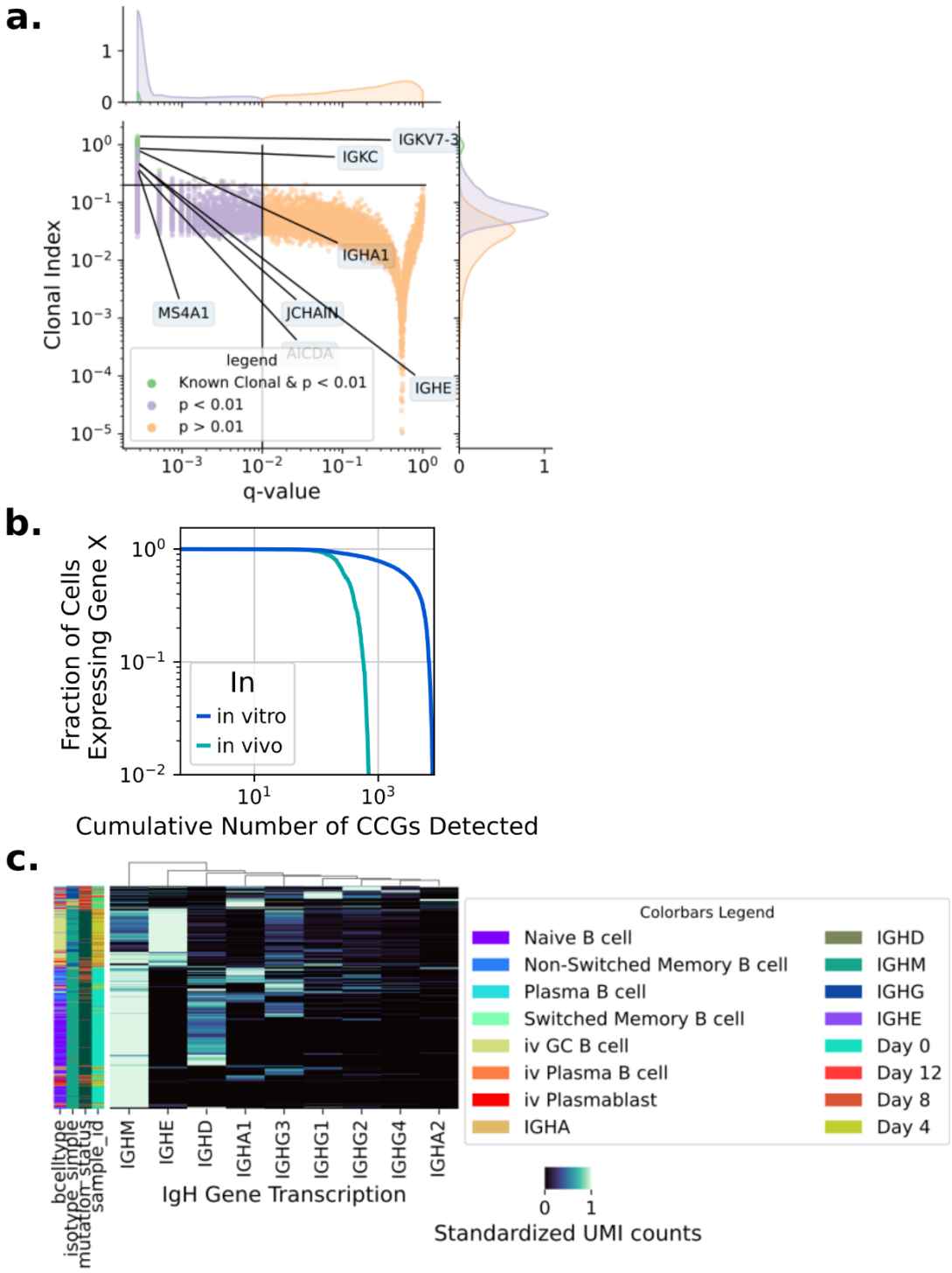


Figure S4. (A) A volcano plot showing results of the transcriptome-wide permutation test only for the *in vivo* (BM CD138+) sample. The q-values are Benjamini-Hochberg corrected p-values and the clonal index is a normalized metric of expression variance described in the methods. Genes of interest are labeled and groups of interest are colored. (B) Empirical cumulative distribution plot of the number of clonal genes detected with q-value < 0.01 by the Fraction of cells expressing a given gene (also known as gene drop-out). (C) Clustermap of cells by transcripts detected at the IgH locus, standardized within each row (cell)

Materials and Methods:

Sample collection

PBMCs were obtained from 2 healthy adult male LRS chambers from the Stanford Blood Center. LRS chamber product was diluted 1:4 in PBS + 2% FBS and PBMCs were isolated using a Ficoll gradient and Red Blood Cell Lysis. Cells were frozen in Cryostor CS10 according to the manufacturer's instructions. Human bone marrow aspirates from a healthy male aged 50-55 were obtained from AllCells. Mononuclear cells were isolated using Ficoll gradient and Red Blood Cell Lysis.

Custom analysis code: https://github.com/michael-swift/seqclone3

B cell purification and culture:

PBMCs were thawed and B cells were purified by negative selection using the StemCell B cell enrichment kit. The resulting cells were measured by flow cytometry to be over 84% pure, which was confirmed by single-cell RNA sequencing. B cells were cultured for 8 hours in B cell media (StemCell) at 37C 5% CO2 at a density of 2 x 10^4 cells per well. After 8 hours B cells were stimulated using the B cell stimulation cocktail according to manufacturer's instructions (Stem Cell). B cell culture wells were thoroughly mixed every 24 hours, to mitigate any spatial effects of the culture on particular B cells. On Day 4, the cells were split into 3 wells and restimulated. Every 24 hours these three wells were pooled, mixed and redistributed into three new wells. On Day 8, the B cells were restimulated and separated into 6 wells, at which point the pooling and mixing was carried out using 6 wells.

Single-cell isolation and sequencing

B cells were washed 2 times in PBS + 1% BSA and stained with cite-seq antibodies according to (Stoeckius et al., 2018). Then, cells were counted and loaded on the Chromium (10X Genomics) at a target loading of 20,000 cells per lane. Reverse transcription and complementary DNA (cDNA) amplification were performed using the Single Cell V(D)J kit V2 (10X Genomics). VDJ and gene expression libraries were prepared from each of the 4 time points. All library preparation was done according to manufacturer's instructions, except for the use of custom constant region primers (Horns et al., 2016) for VDJ enrichment. On Days 8, and 12, cells from culture wells were sorted by propidium iodide exclusion into cold PBS + 2% BSA, before proceeding to antibody staining. Libraries were sequenced using the Illumina NovaSeq platform with paired-end reads of 26 bp and 98 bp.

Preprocessing of single-cell sequence data:

We used snakemake (Mölder et al., 2021) to manage the computational workflow. We CellRanger 6.1 to map, count, and assemble reads from the sequencing libraries. We used the Immcantation docker (Gupta et al., 2015) pipeline and scirpy to reprocess the contigs assembled by CellRanger, distinguish bonafide single cells from multiplets, and assign clonal barcodes, which agreed with our in-house pipeline (Croote et al., 2018). Single B cells were identified by the presence of a single

productive heavy chain and a single productive light chain, yielding a total of 29,703 single B cells for analysis. All other cells were excluded from further analysis.

Single cell sequencing data analysis:

Gene expression analysis of single cells was performed using scanpy (Wolf et al., 2018), and exploratory analysis of the immune receptors with scirpy (Sturm et al., 2020). Briefly, single cell transcriptomes were log-transformed and normalized to counts per 10^4 UMIs. Clusters were identified using the leiden algorithm. These clusters were manually annotated (i.e., labels were assigned) based on expression of marker genes for each cell state. Differentially expressed genes were identified using the wilcoxon-rank sum test adjusted for multiple testing using the Benjamini-Hochberg procedure as implemented in scanpy.

Analysis of mutated vs. germline outcomes:

To explore the cell fate biases of memory vs. naive progenitors, we binned VDJ sequences into heavily mutated, mutated, and germline categories as shown in S2A. We used multiple different levels of reasonable cutoffs and found they did not change our general conclusions. We calculated the ratio of germline over mutated by dividing the number of germline cells by the number of mutated or heavily mutated cells in a given type of category, such which IGH constant region is associated with the cell's VDJ or which b cell type is associated with a VDJ.

Analysis of clonal gene expression:

In general we employed the resampling and permutation methods to estimate confidence intervals and p-values for all clonal effects. We implemented the approach lucidly described here (Horton et al., 2018), and all code implementing these tests will be publicly available on GitHub. The test statistic for genes was the within-clone variance in UMI counts for each gene averaged across all clones. For calculating the clonal index in 4B, we divided the difference in the gene-statistics between true labels and permuted labels by mean transformed UMI count for that gene.

Acknowledgements:

We thank Ivana Cvijovic, Elisabeth Jerison, Derek Croote, Bali Pulendran, Dan Jarosz, and Daria-Mochly Rosen for useful discussions and helpful comments on the manuscript. This work was supported by the National Science Foundation Graduate Research Fellowship Program (to M.A.S.)

References:

Bergthorsdottir, S., Gallagher, A., Jainandunsing, S., Cockayne, D., Sutton, J., Leanderson, T., and Gray, D. (2001). Signals that initiate somatic hypermutation of B cells in vitro. J. Immunol. *166*, 2228–2234. https://doi.org/10.4049/jimmunol.166.4.2228.

Chen, C., Liao, Y., and Peng, G. (2022). Connecting past and present: single-cell lineage tracing. Protein Cell https://doi.org/10.1007/s13238-022-00913-7.

Cheon, H., Kan, A., Prevedello, G., Oostindie, S.C., Dovedi, S.J., Hawkins, E.D., Marchingo, J.M., Heinzel, S., Duffy, K.R., and Hodgkin, P.D. (2021). Cyton2: A model of immune cell population dynamics that includes familial instructional inheritance. Front. Bioinform. *1*. https://doi.org/10.3389/fbinf.2021.723337.

Croote, D., Darmanis, S., Nadeau, K.C., and Quake, S.R. (2018). High-affinity allergen-specific human antibodies cloned from single IgE B cell transcriptomes. Science *362*, 1306–1309. https://doi.org/10.1126/science.aau2599.

Eraslan, G., Drokhlyansky, E., Anand, S., Fiskin, E., Subramanian, A., Slyper, M., Wang, J., Van Wittenberghe, N., Rouhana, J.M., Waldman, J., et al. (2022). Single-nucleus cross-tissue molecular reference maps toward understanding disease gene function. Science *376*, eabl4290. https://doi.org/10.1126/science.abl4290.

Gupta, N.T., Vander Heiden, J.A., Uduman, M., Gadala-Maria, D., Yaari, G., and Kleinstein, S.H. (2015). Change-O: a toolkit for analyzing large-scale B cell immunoglobulin repertoire sequencing data. Bioinformatics *31*, 3356–3358. https://doi.org/10.1093/bioinformatics/btv359.

Hammarlund, E., Thomas, A., Amanna, I.J., Holden, L.A., Slayden, O.D., Park, B., Gao, L., and Slifka, M.K. (2017). Plasma cell survival in the absence of B cell memory. Nat. Commun. *8*, 1781. https://doi.org/10.1038/s41467-017-01901-w.

Hasbold, J., Lyons, A.B., Kehry, M.R., and Hodgkin, P.D. (1998). Cell division number regulates IgG1 and IgE switching of B cells following stimulation by CD40 ligand and IL-4. European Journal of Immunology .

Horns, F., Vollmers, C., Croote, D., Mackey, S.F., Swan, G.E., Dekker, C.L., Davis, M.M., and Quake, S.R. (2016). Lineage tracing of human B cells reveals the in vivo landscape of human antibody class switching. ELife *5*. https://doi.org/10.7554/eLife.16578.

Horton, M.B., Prevedello, G., Marchingo, J.M., Zhou, J.H.S., Duffy, K.R., Heinzel, S., and Hodgkin, P.D. (2018). Multiplexed Division Tracking Dyes for Proliferation-Based Clonal Lineage Tracing. J. Immunol. *201*, 1097–1103. https://doi.org/10.4049/jimmunol.1800481.

Lee, C.G., Kinoshita, K., Arudchandran, A., Cerritelli, S.M., Crouch, R.J., and Honjo, T. (2001). Quantitative regulation of class switch recombination by switch region transcription. J. Exp. Med. *194*, 365–374.

Looney, T.J., Lee, J.-Y., Roskin, K.M., Hoh, R.A., King, J., Glanville, J., Liu, Y., Pham, T.D., Dekker, C.L., Davis, M.M., et al. (2016). Human B-cell isotype switching origins of IgE. J. Allergy Clin. Immunol. *137*, 579-586.e7. https://doi.org/10.1016/j.jaci.2015.07.014.

Lu, R., Neff, N.F., Quake, S.R., and Weissman, I.L. (2011). Tracking single hematopoietic stem cells in vivo using high-throughput sequencing in conjunction with viral genetic barcoding. Nat. Biotechnol. *29*, 928–933. https://doi.org/10.1038/nbt.1977.

McInnes, L., Healy, J., and Melville, J. (2018). UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction. ArXiv https://doi.org/10.48550/arxiv.1802.03426.

Mölder, F., Jablonski, K.P., Letcher, B., Hall, M.B., Tomkins-Tinch, C.H., Sochat, V., Forster, J., Lee, S.,

Twardziok, S.O., Kanitz, A., et al. (2021). Sustainable data analysis with Snakemake. F1000Res. *10*, 33. https://doi.org/10.12688/f1000research.29032.1.

Stoeckius, M., Zheng, S., Houck-Loomis, B., Hao, S., Yeung, B.Z., Mauck, W.M., Smibert, P., and Satija, R. (2018). Cell Hashing with barcoded antibodies enables multiplexing and doublet detection for single cell genomics. Genome Biol. *19*, 224. https://doi.org/10.1186/s13059-018-1603-1.

Sturm, G., Szabo, T., Fotakis, G., Haider, M., Rieder, D., Trajanoski, Z., and Finotello, F. (2020). Scirpy: a Scanpy extension for analyzing single-cell T-cell receptor-sequencing data. Bioinformatics *36*, 4817–4818. https://doi.org/10.1093/bioinformatics/btaa611.

Tabula Sapiens Consortium*, Jones, R.C., Karkanias, J., Krasnow, M.A., Pisco, A.O., Quake, S.R., Salzman, J., Yosef, N., Bulthaup, B., Brown, P., et al. (2022). The Tabula Sapiens: A multiple-organ, single-cell transcriptomic atlas of humans. Science *376*, eabl4896. https://doi.org/10.1126/science.abl4896.

Tangye, S.G., Avery, D.T., Deenick, E.K., and Hodgkin, P.D. (2003). Intrinsic differences in the proliferation of naive and memory human B cells as a mechanism for enhanced secondary immune responses. J. Immunol. *170*, 686–694. https://doi.org/10.4049/jimmunol.170.2.686.

Wang, S.-W., Herriges, M.J., Hurley, K., Kotton, D.N., and Klein, A.M. (2022). CoSpar identifies early cell fate biases from single-cell transcriptomic and lineage information. Nat. Biotechnol. https://doi.org/10.1038/s41587-022-01209-1.

Wang, X., Peticone, C., Kotsopoulou, E., Göttgens, B., and Calero-Nieto, F.J. (2021). Single-cell transcriptome analysis of CAR T-cell products reveals subpopulations, stimulation, and exhaustion signatures. Oncoimmunology *10*, 1866287. https://doi.org/10.1080/2162402X.2020.1866287.

Whitman, C.O. (1878). The Embryology of Clepsine (J.E. Adlard).

Wolf, F.A., Angerer, P., and Theis, F.J. (2018). SCANPY: large-scale single-cell gene expression data analysis. Genome Biol. *19*, 15. https://doi.org/10.1186/s13059-017-1382-0.