Supplementary Information

Multi-omics profiling of primary human B cells activated via distinct stimuli

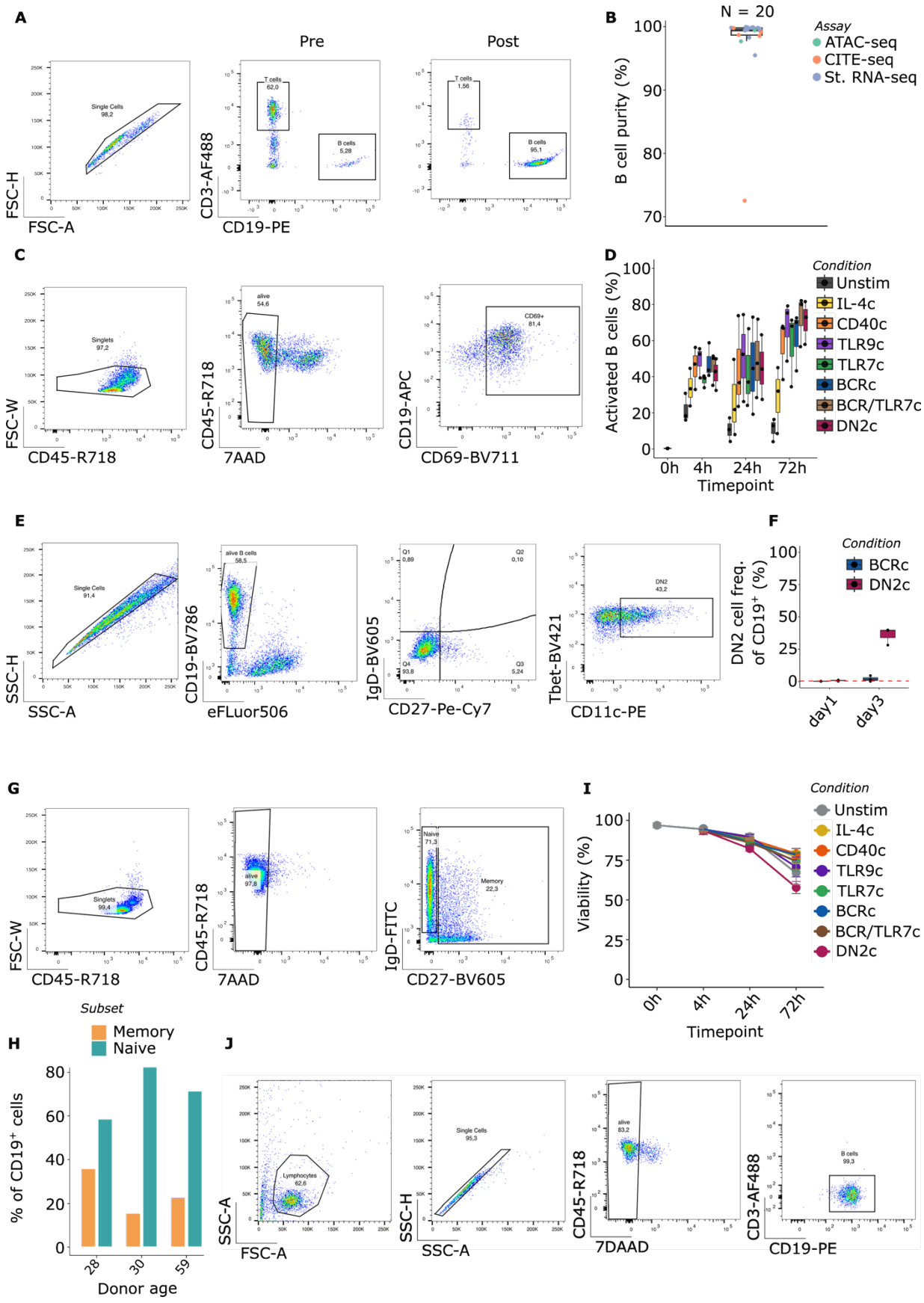
Vitor R. C. Aguiar[1,2,#], Marcella E. Franco[1,2,3,#], Nada Abdel Aziz[1,2,4,5], Daniela Fernandez-Salinas[1,2], Marcos Chiñas[1,2], Mariasilvia Colantuoni[1,2], Qian Xiao[1], Nicolaj Hackert[1,2,6], Yifei Liao[7], Rodrigo Cervantes-Diaz[1,2], Marc Todd[1], Brian Wauford[1], Alex Wactor[1], Vaishali Prahalad[1], Raquel Laza-Briviesca[1], Roxane Darbousset[1], Qiang Wang[1], Scott Jenks[8], Kevin S. Cashman[8], Esther Zumaquero[9], Zhu Zhu[10], Junning Case[10], Paloma Cejas[11], Miguel Gomez[11], Hannah Ainsworth[12,13], Miranda Marion[12], Mehdi Benamar[1,†], Pui Lee[1], Lauren Henderson[1], Margaret H. Chang[1], Kevin Wei[10,14], Henry Long[11], Carl D. Langefeld[12,13], Ashley D. Sanders[3,15,16], Benjamin E. Gewurz[7], Ignacio Sanz[8], Jeffrey A. Sparks[14], Esra Meidan[1], Peter A. Nigrovic[1,14], Maria Gutierrez-Arcelus[1,2,*]

[1] Division of Immunology, Boston Children's Hospital, Harvard Medical School; [2] Program in Medical and Population Genetics, Broad Institute of MIT and Harvard; [3] Max Delbrück Center for Molecular Medicine in the Helmholtz Association, Berlin Institute for Medical System Biology; [4] Biotechnology Department, Faculty of Science, Cairo University; [5] Division of Immunology, Institute of Infectious Diseases and Molecular Medicine, Faculty of Health Sciences, University of Cape Town; [6] Division of Rheumatology, Department of Medicine V, Heidelberg University Hospital; [7] Division of Infectious Diseases, Department of Medicine, Brigham and Women's Hospital, Harvard Medical School; [8] Department of Medicine, Division of Rheumatology, Lowance Center for Human Immunology, Emory University; [9] Department of Microbiology, University of Alabama at Birmingham; [10] Center for Cellular Profiling, Brigham and Women's Hospital, Harvard Medical School; [11] Center for Functional Cancer Epigenetics, Dana-Farber Cancer Institute, Harvard Medical School; [12] Department of Biostatistics and Data Science, Wake Forest School of Medicine; [13] Center for Precision Medicine, Wake Forest School of Medicine; [14] Division of Rheumatology, Inflammation, and Immunity, Brigham and Women's Hospital, Harvard Medical School; [15] Berlin Institute of Health (BIH) Translational Research Unit; [16] Charité – Universitätsmedizin Berlin; [†] Present address: Institute of Regenerative Medicine and Biotherapy, INSERM, University of Montpellier, Montpellier, France.

[#] These authors contributed equally: Vitor R. C. Aguiar, Marcella E. Franco.
Correspondence should be addressed to M. Gutierrez-Arcelus: mgutierr@broadinstitute.org

This supplementary information file includes:

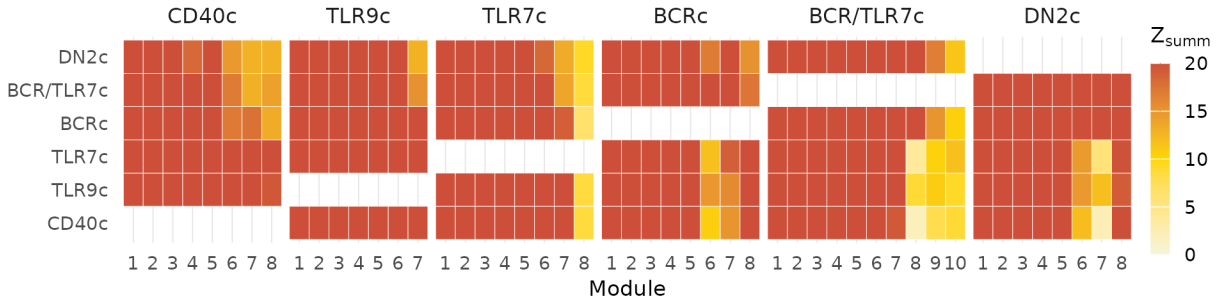
Supplementary Figures 1–16


A
Pre
Post
Single Cells
98,2
T cells
62,0
B cells
5,28
T cells
1,56
B cells
95,1
FSC-H
FSC-A
CD3-AF488
CD19-PE
B
N = 20
B cell purity (%)
Assay
ATAC-seq
CITE-seq
St. RNA-seq
C
Singlets
97,2
alive
54,6
CD69+
81,4
FSC-W
CD45-R718
CD45-R718
7AAD
CD19-APC
CD69-BV711
D
Activated B cells (%)
Condition
Unstim
IL-4c
CD40c
TLR9c
TLR7c
BCRc
BCR/TLR7c
DN2c
0h
4h
24h
72h
Timepoint
E
Single Cells
91,4
alive B cells
58,5
Q1
0,89
Q2
0,10
Q4
93,8
Q3
5,24
DN2
43,2
SSC-H
SSC-A
CD19-BV786
eFLuor506
IgD-BV605
CD27-Pe-Cy7
Tbet-BV421
CD11c-PE
F
Condition
BCRc
DN2c
DN2 cell freq. of CD19+ (%)
day1
day3
G
Singlets
99,4
alive
97,8
Naive
71,3
Memory
22,3
FSC-W
CD45-R718
CD45-R718
7AAD
IgD-FITC
CD27-BV605
I
Condition
Unstim
IL-4c
CD40c
TLR9c
TLR7c
BCRc
BCR/TLR7c
DN2c
Viability (%)
0h
4h
24h
72h
Timepoint
Subset
Memory
Naive
H
% of CD19+ cells
Donor age
J
Lymphocytes
62,6
Single Cells
95,3
alive
83,2
B cells
99,3
SSC-A
FSC-A
SSC-H
SSC-A
CD45-R718
7DAAD
CD3-AF488
CD19-PE

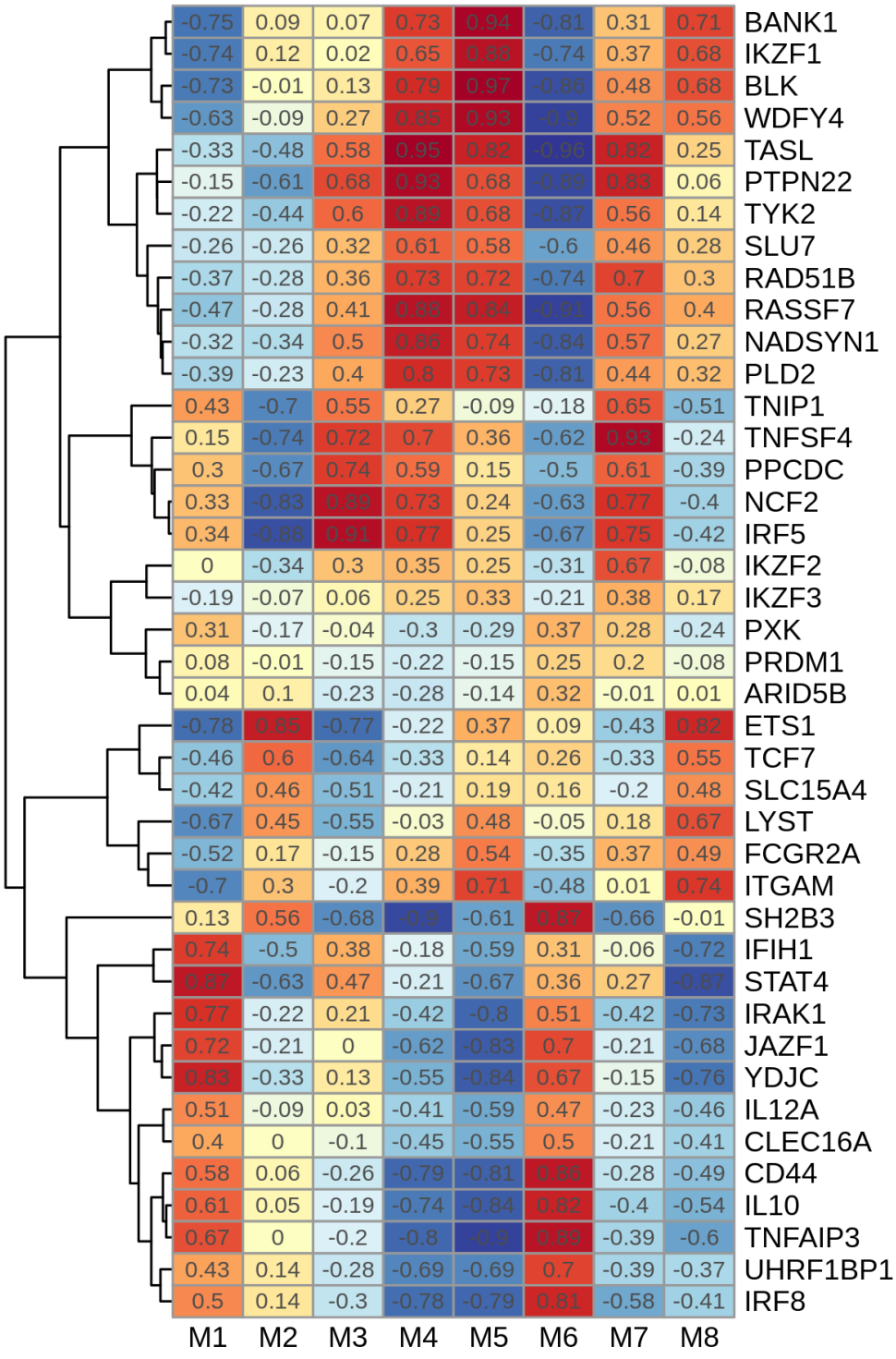
Supplementary Figure 1. (A) Representative flow cytometry plots showing B-cell frequency before (Pre) and after (Post) B-cell enrichment. The gating strategy shown was used to determine B-cell frequency in (B). **(B)** Proportion of B cells after enrichment across assays (n = 20). **(C)** Representative flow cytometry plots showing B-cell activation based on CD69 expression. The gating strategy shown is representative of the DN2 condition at 72 hours post-culture and was used to determine the percentage of activated B cells shown in (D). **(D)** Percentage of activated B cells, based on CD69 expression, across stimulation conditions and time points. **(E)** Gating strategy used to identify DN2 B cells based on surface marker expression at 72 hours post-culture and quantify their frequency in (F). **(F)** Frequency of DN2 B cells at 72 hours post-culture (n = 3). **(G)** Representative flow cytometry plots showing the gating strategy used to identify naïve and memory B-cell populations and determine their proportions in (H). **(H)** Proportion of naïve and memory B cells across donors of different ages. **(I)** B-cell viability, assessed by 7-AAD staining, over time under the indicated culture conditions. **(J)** Flow cytometry gating strategy for low-input RNA-seq and CITE-seq. For low-input RNA-seq, cells were sequentially gated to identify lymphocytes (FSC-A/SSC-A), singlets (FSC-A/FSC-H), and viable cells ($7AAD^-$). For single-cell RNA sequencing, live singlet cells were further gated to select $CD19^+CD3^-$ B cells. Unless otherwise specified, data were generated from independent technical validation experiments performed on three healthy donors. Source data are provided as a Source Data file.


CD40c
M1
MYL12B
SEC61B
PSMB2
PSMA1
PSMA6
protein folding
proteasomal protein catabolic pro...
positive regulation of telomerase...
positive regulation of establishm...
establishment of protein localiza...
M2
TBC1D10C
BLNK
UBA7
CYTH4
AC006129.2
immune response-activating cell s...
immune response-activating signal...
immune response-regulating cell s...
B cell activation
antigen receptor-mediated signali...
M3
CCR7
CFLAR
LMNA
TMEM43
CMTM6
cytokine-mediated signaling pathway
regulation of cytokine-mediated s...
rRNA processing
cellular response to organonitrog...
rRNA metabolic process
M4
HDAC1
CORO1A
CCR2
PPM1K
SYPL1
cholesterol biosynthetic process
secondary alcohol biosynthetic pr...
sterol biosynthetic process
secondary alcohol metabolic process
cholesterol metabolic process
M5
RP11-346D14.1
HELLPAR
LINC00506
RP11-76P17.4
RP11-567P17.5
detection of chemical stimulus in...
sensory perception of smell
sensory perception of chemical st...
detection of chemical stimulus
M6
FOSB
DUSP1
ZBTB16
CH17-3B23.3
LINC00926
response to calcium ion
response to ketone
cellular response to calcium ion
response to cAMP
response to inorganic substance
M7
RPS3A
RPS25
RPS20
RPL39
RPS14
cytoplasmic translation
ribosome biogenesis
ribonucleoprotein complex biogenesis
rRNA processing
rRNA metabolic process
M8
SNRPD2
NDUFA1
TMA7
COX6C
EIF4A2
oxidative phosphorylation
aerobic respiration
cellular respiration
ATP metabolic process
energy derivation by oxidation of...
TLR9c
M1
PSMB2
XRCC6
PSMA2
SEC61B
PSMA6
RNA splicing
RNA splicing, via transesterifica...
mRNA splicing, via spliceosome
mRNA processing
M2
RP11-346D14.1
RP11-490F3.3
MAB21L3
RTCA-AS1
RP11-567P17.5
detection of chemical stimulus in...
sensory perception of smell
sensory perception of chemical st...
detection of chemical stimulus
M3
CORO1A
HDAC1
ARPC1B
IGHG3
GLRX
sterol biosynthetic process
cholesterol biosynthetic process
secondary alcohol biosynthetic pr...
secondary alcohol metabolic process
sterol metabolic process
M4
DDX21
CFLAR
CMTM6
CCR7
BCL2A1
ncRNA metabolic process
ribosome biogenesis
rRNA processing
tRNA aminoacylation for protein t...
rRNA metabolic process
M5
ERP29
PTPRCAP
PTPN6
ACAP1
GSTK1
B cell receptor signaling pathway
B cell activation
regulation of B cell activation
antigen receptor-mediated signali...
activation of immune response
M6
MARCHF1
PRKCB
TP53INP1
TXNIP
FCRL1
regulation of B cell receptor sig...
antigen processing and presentati...
M7
RPS14
RPS20
RPS3A
RPL19
RPL13A
cytoplasmic translation
ribosome biogenesis
ribonucleoprotein complex biogenesis
rRNA processing
rRNA metabolic process
TLR7c
M1
CCT6A
VDAC1
XRCC6
ENO1
PSMA5
protein folding
ribonucleoprotein complex biogenesis
RNA splicing
positive regulation of establishm...
regulation of establishment of pr...
M2
BRK1
ACTB
CFL1
MYL12A
DBI
aerobic respiration
cellular respiration
energy derivation by oxidation of...
generation of precursor metabolit...
ATP metabolic process
M3
RP11-567P17.5
RP11-490F3.3
RP11-346D14.1
RP11-379B18.8
NMNAT3
detection of chemical stimulus in...
sensory perception of smell
sensory perception of chemical st...
detection of chemical stimulus
M4
PTPRCAP
CD79B
CNN2
ARHGDIB
DENND1C
regulation of cell activation
regulation of lymphocyte activation
regulation of leukocyte activation
phagocytosis
immune effector process
M5
NCF1B
METTL7A
MARCHF1
FAM30A
TSPAN13
leukocyte proliferation
lymphocyte proliferation
mononuclear cell proliferation
regulation of B cell receptor sig...
B cell proliferation
M6
BCL2A1
NOP58
PRNP
TNFRSF18
MAT2A
NA
M7
RPS25
RPL39
RPS14
RPL13A
RPL31
cytoplasmic translation
ribosome biogenesis
ribonucleoprotein complex biogenesis
rRNA processing
rRNA metabolic process
M8
AKIRIN1
RALY
ARHGDIA
SDF4
SRCAP
NA
BCRc
M1
SERBP1
PPA1
HSPD1
RAN
PRDX1
ribonucleoprotein complex biogenesis
ribosome biogenesis
rRNA processing
ribonucleoprotein complex assembly
ribonucleoprotein complex subunit...
M2
NDUFV2
ATP5F1A
CFL1
COX8A
MYL6
aerobic respiration
cellular respiration
ATP metabolic process
oxidative phosphorylation
generation of precursor metabolit...
M3
RP11-490F3.3
NMNAT3
MORC4
RP11-346D14.1
RP11-47I22.3
detection of stimulus involved in...
detection of chemical stimulus in...
detection of stimulus
sensory perception of smell
M4
CD79B
PTPRCAP
ARL6IP5
ACAP1
CNN2
leukocyte mediated immunity
immune effector process
B cell receptor signaling pathway
lymphocyte mediated immunity
phagocytosis
M5
FCRL1
LINC00926
RIPOR2
TSC22D3
MARCHF1
regulation of B cell receptor sig...
negative regulation of immune sys...
regulation of antigen receptor-me...
regulation of B cell proliferation
B cell proliferation
M6
ARHGAP18
TYMS
ITGAL
CCNA2
DLGAP5
mitotic spindle assembly
spindle assembly
mitotic spindle organization
nuclear division
spindle organization
M7
MYC
DPH2
SPHK1
MAT2A
CDKN1A
interleukin-17 production
regulation of interleukin-17 prod...
M8
RPL10
RPS25
RPS14
RPL13A
RPL31
cytoplasmic translation
ribosome biogenesis
ribonucleoprotein complex biogenesis
rRNA processing
rRNA metabolic process
logP:
90
60
30
0
Gene Ratio:
0.2
0.5
0.7
1.0
BCR/TLR7c
M1
SERBP1
SLC29A1
CCT5
HSP90AB1
PPA1
ribonucleoprotein complex biogenesis
ribosome biogenesis
rRNA processing
ribonucleoprotein complex assembly
ncRNA metabolic process
M2
LINC00506
RP11-490F3.3
AC003099.2
RP11-624A21.1
RP11-373N22.3
detection of stimulus involved in...
detection of stimulus
detection of chemical stimulus in...
detection of chemical stimulus
sensory perception of smell
M3
ARPC1B
ECH1
CD46
CSRP1
SIGLEC10
fatty acid beta-oxidation
fatty acid oxidation
lipid oxidation
generation of precursor metabolit...
small molecule catabolic process
M4
ATP5PB
ATP5MF
PRDX3
LDHB
NDUFB3
aerobic respiration
ATP metabolic process
oxidative phosphorylation
cellular respiration
generation of precursor metabolit...
M5
ANKZF1
CD79B
SP100
POLD4
TRIM38
B cell receptor signaling pathway
antigen receptor-mediated signali...
phagocytosis
immune response-activating cell s...
immune response-activating signal...
M6
FCRL1
LINC00926
TAPT1-AS1
TSC22D3
NLRP1
NA
M7
MYC
MAT2A
KLHL21
DPH2
PMM2
rRNA processing
rRNA metabolic process
ribosome biogenesis
response to mitochondrial depolar...
negative regulation of cell popul...
M8
RRM2
QPRT
TK1
CDK1
DHFR
mitotic nuclear division
nuclear division
organelle fission
chromosome segregation
sister chromatid segregation
M9
CD55
PPP1R15A
HIVEP2
HGSNAT
LINC02605
NA
M10
SRSF2
RRP1
NOL7
UBE2D1
NFKBIB
NA
DN2c
M1
HSP90AB1
HSPA9
PA2G4
CCT5
SERBP1
ribonucleoprotein complex biogenesis
ribosome biogenesis
rRNA processing
rRNA metabolic process
ncRNA metabolic process
M2
NMNAT3
RP11-490F3.3
LINC00506
RTCA-AS1
RP11-373N22.3
detection of stimulus
detection of stimulus involved in...
detection of chemical stimulus
detection of chemical stimulus in...
sensory perception of smell
M3
ATP5F1A
ALDOA
TUBB
ATP5PB
ATP5MC3
aerobic respiration
ATP metabolic process
cellular respiration
oxidative phosphorylation
generation of precursor metabolit...
M4
RPS6KA1
CORO1A
BTK
SUCLG2
SASH3
B cell receptor signaling pathway
adaptive immune response
leukocyte mediated immunity
B cell activation
positive regulation of B cell act...
M5
UCP2
PPM1K
POLD4
C16orf54
BLK
negative regulation of B cell act...
respiratory burst
immune response-activating cell s...
immune response-activating signal...
immune response-regulating cell s...
M6
RP11-36B6.2
PRNP
AC058791.1
NIPAL4
EIF1
negative regulation of bone resor...
negative regulation of bone remod...
negative regulation of tissue rem...
response to lipopolysaccharide
M7
TK1
DLGAP5
CDCA5
UBE2C
CCNA2
nuclear division
organelle fission
mitotic nuclear division
mitotic sister chromatid segregation
sister chromatid segregation
M8
FCRL1
TSC22D3
ADAM28
PIK3IP1
ITPR1
NA

Supplementary Figure 2. WGCNA co-expression modules, hub genes, and GO enrichments for each activation condition. Weighted Gene Co-expression Network Analysis (WGCNA) was performed separately for each activation condition in the low-input RNA-seq time course (N = 6). For each condition (panels), the left column shows module eigengene expression trajectories over time (the eigengene summarizes a module's expression profile), the middle column shows the top hub genes per module (colored by intra-module connectivity, kIM), and the right column shows enriched Gene Ontology (GO) biological processes per module (over-representation analysis, clusterProfiler).

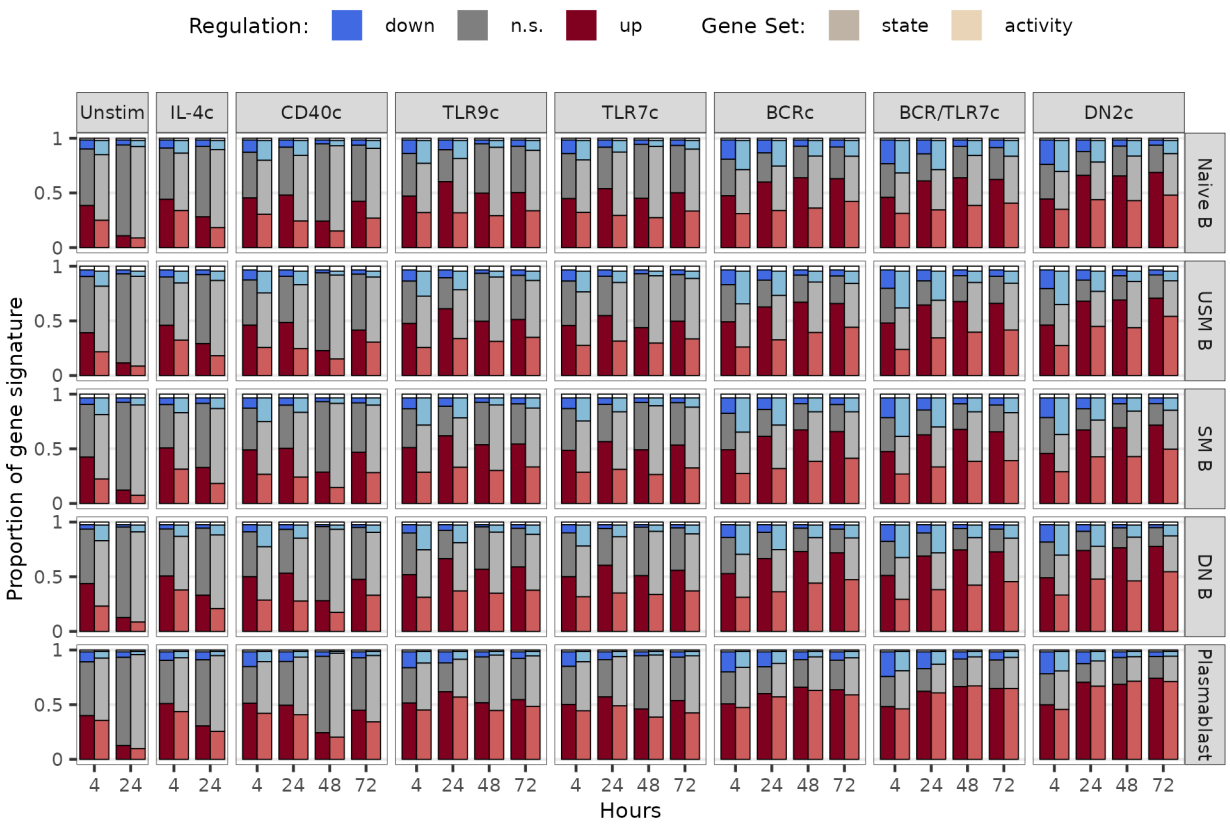


Supplementary Figure 3. Preservation of gene co-expression modules across B cell activation conditions. WGCNA module preservation analysis in the low-input RNA-seq time course (N = 6). The $Z_{summary}$ statistic (a permutation-based composite preservation score) is shown for each module to assess network preservation between each pair of conditions. $Z_{summary} < 2$ indicates unpreserved modules (condition-specific dynamics); $2 < Z_{summary} < 10$, moderate preservation; $Z_{summary} > 10$, strong preservation.

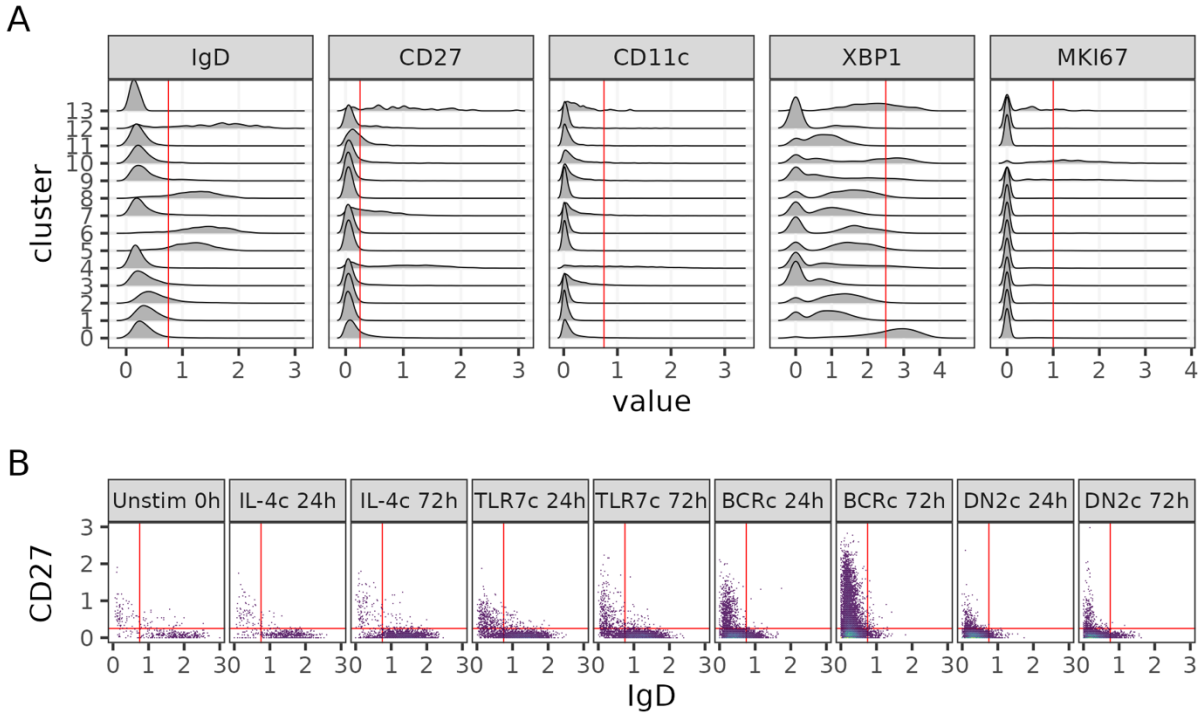


Supplementary Figure 4. WGCNA module membership of SLE-associated genes. Module membership (kME; the correlation of a gene's expression with the module eigengene) for SLE-associated genes across DN2c modules in the low-input RNA-seq data (N = 6). The y-axis lists SLE genes identified by GWAS (Bentham, J., Morris, D., Cunninghame Graham, D. et al. Genetic association analyses implicate aberrant regulation of innate and adaptive immunity genes in the

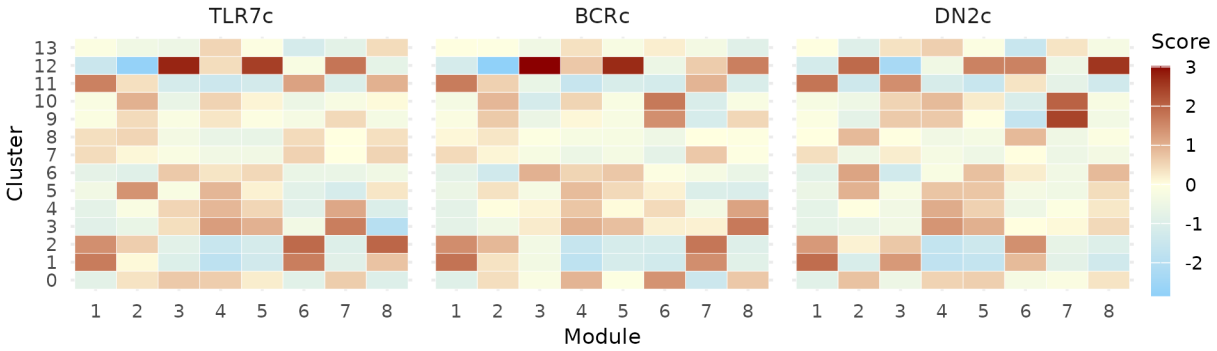
pathogenesis of systemic lupus erythematosus. Nat Genet **47**, 1457–1464 (2015). https://doi.org/10.1038/ng.3434).



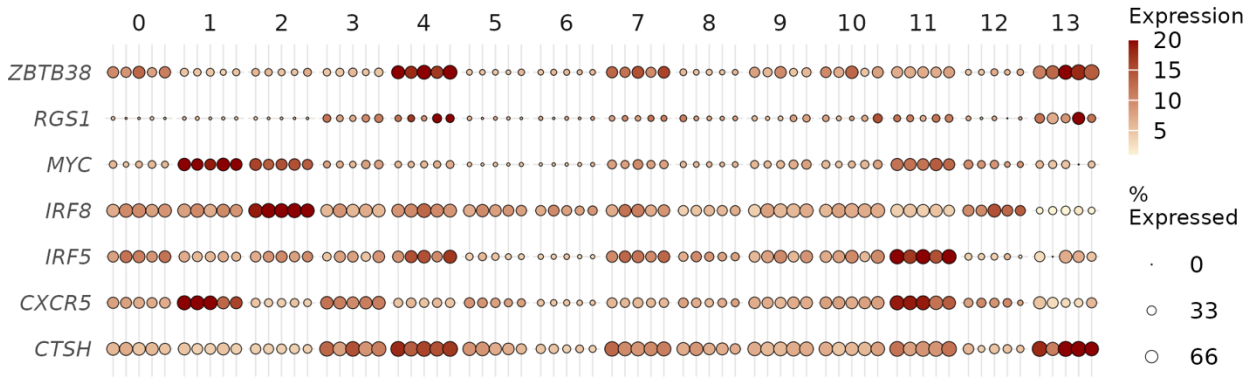
Supplementary Figure 5. Regulation of Nakano et al. B cell gene sets across activation conditions. Proportion of genes from the disease-associated B cell signatures of Nakano et al. (Distinct transcriptome architectures underlying lupus establishment and exacerbation. Cell, 2022; 185, 3375-3389.e21) that are upregulated (red), downregulated (blue), or not significantly changed (gray) at each time point relative to unstimulated cells at 0 h (edgeR, BH FDR < 0.05) in the low-input RNA-seq time course (N = 6). For each time point, the left bar (darker shades) represents the disease-state signature, whereas the right bar (lighter shades) represents the disease-activity signature.



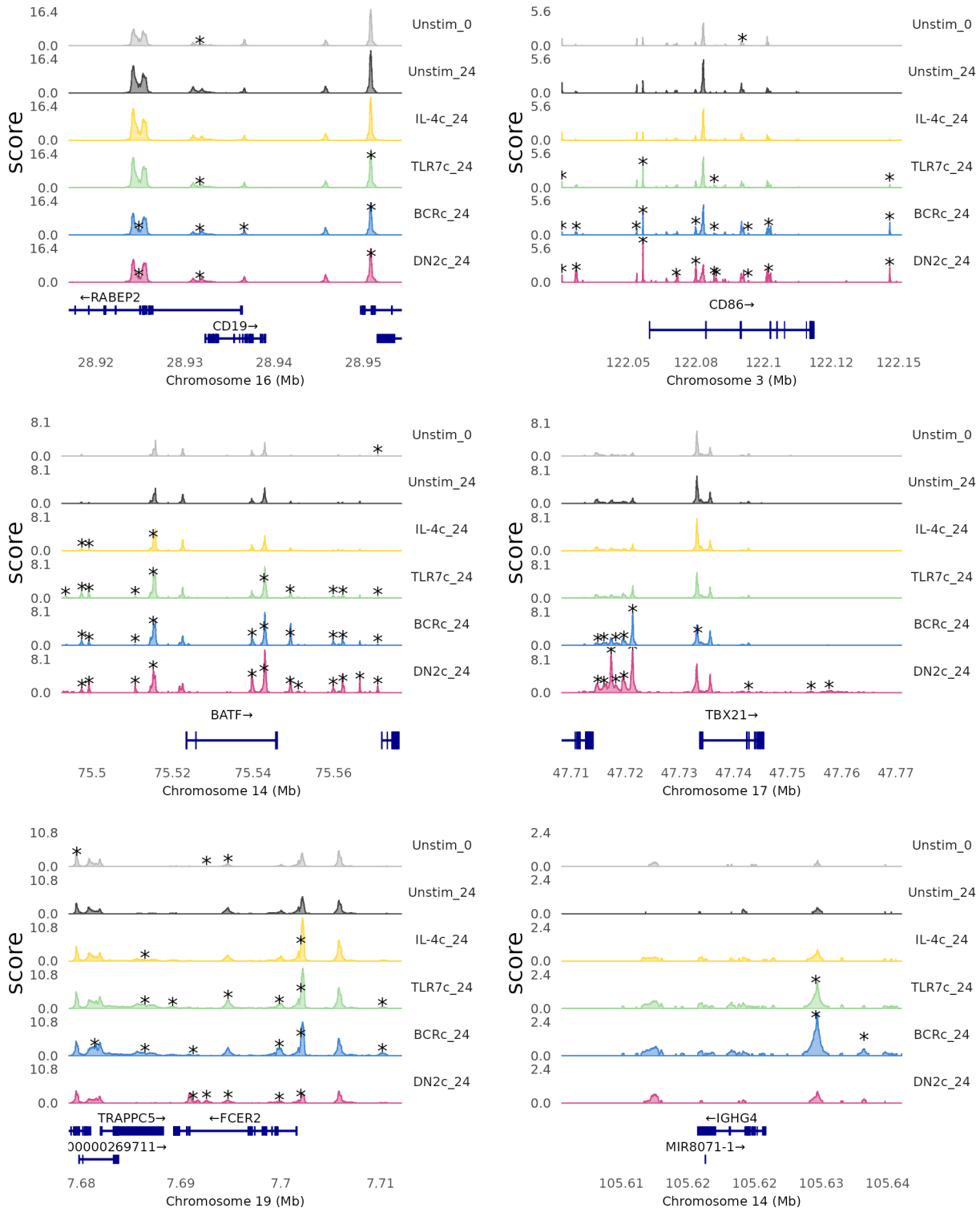
Supplementary Figure 6. Determination of marker thresholds for quantifying B-cell subsets in the CITE-seq data (N = 5). Thresholds used to gate the B cell subsets shown in Fig. 3e. **(A)** Cluster-specific expression distributions of key markers; vertical red lines indicate the specific expression thresholds selected to separate expressing from non-expressing clusters. **(B)** Scatter plot of IgD versus CD27 surface-protein expression; horizontal and vertical red lines correspond to the thresholds defined in (A).



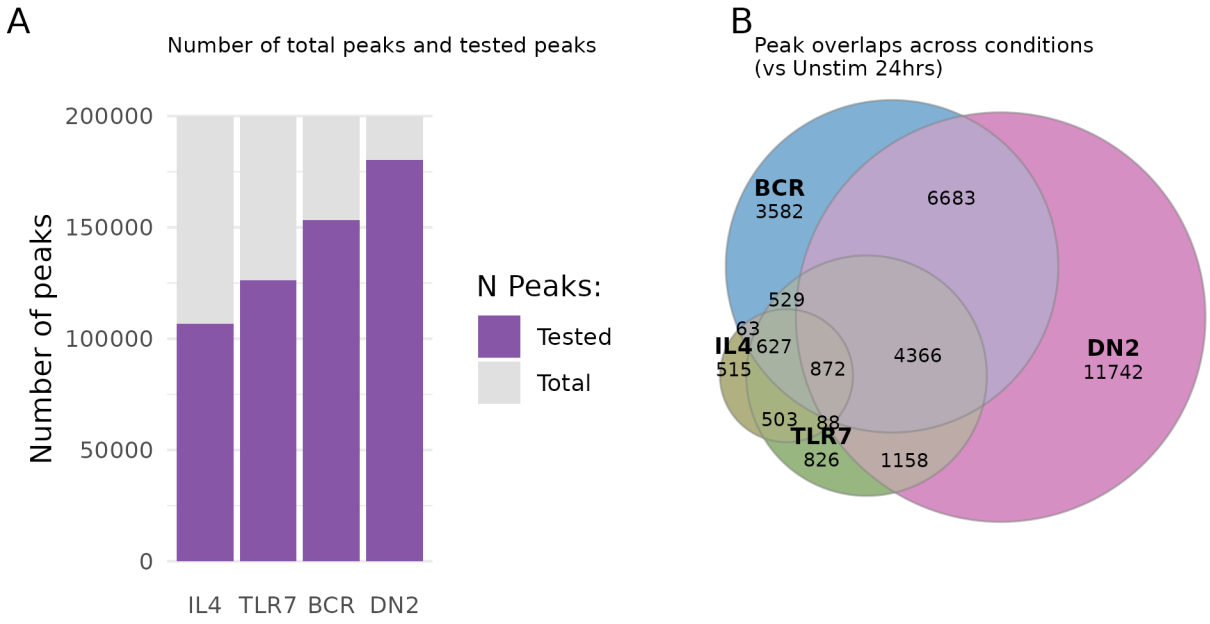
Supplementary Figure 7. Activity of bulk-derived gene modules in single-cell subsets. WGCNA modules derived from bulk RNA-seq (N = 6) for each activation condition were scored in the single-cell data (N = 5, Seurat AddModuleScore). Module scores indicate activity of gene modules (x-axis) derived from each activation condition (panels) across the 13 identified single-cell B cell states (y-axis).



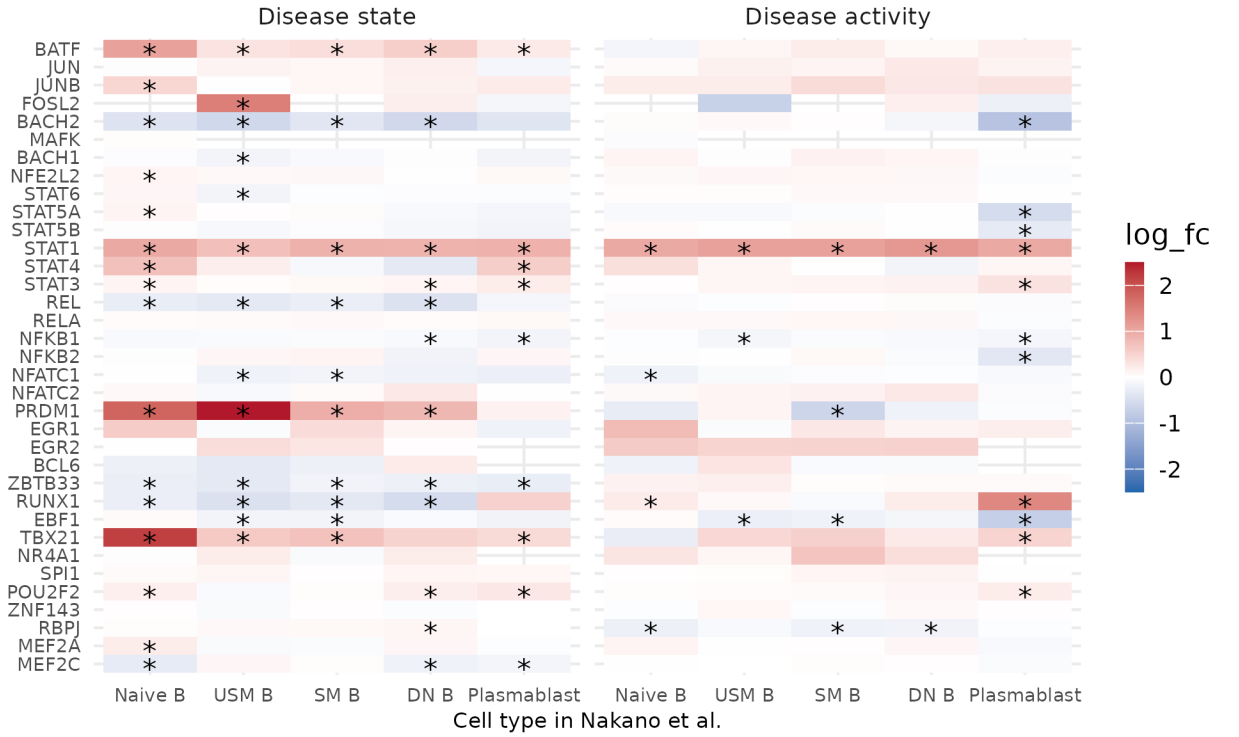
Supplementary Figure 8. Inter-subject variability of IMD-associated gene expression across cell states. Donor-level breakdown of the immune-mediated disease (IMD)-associated gene expression shown in Figure 3f. Each panel is a single-cell cluster; within a panel, each column is a donor (N = 5). Dot size, percentage of expressing cells; dot fill, average scaled expression from Seurat DotPlot function.



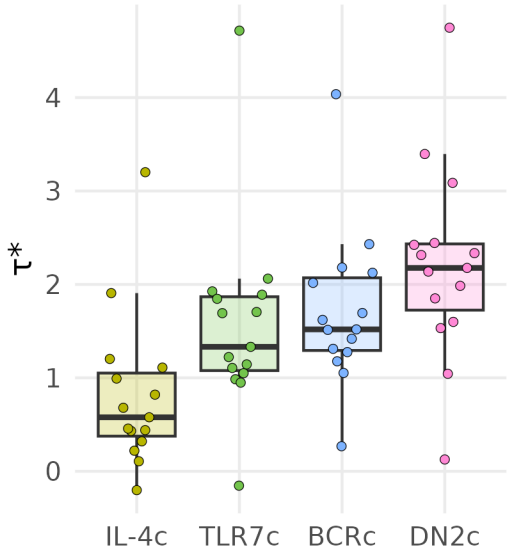
Supplementary Figure 9. ATAC-seq tracks at representative loci across B cell activation conditions. Bulk ATAC-seq accessibility signal at six representative loci across conditions (N = 3). Asterisks indicate regions differentially accessible relative to the unstimulated 24-hour condition ("Unstim_24") at 1% FDR (DESeq2).



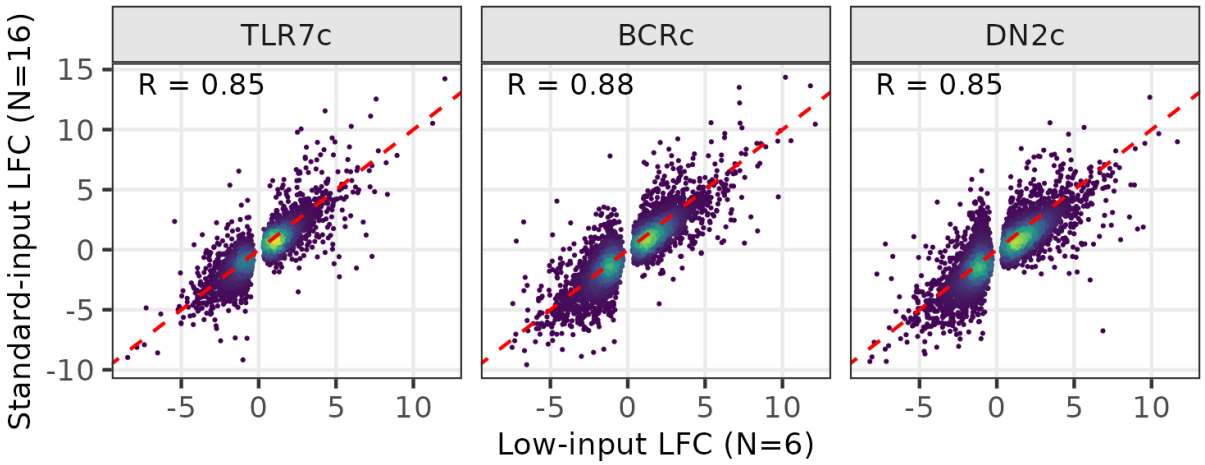
Supplementary Figure 10. ATAC-seq peak statistics and overlap. (A) Total number of consensus peaks (gray) and tested peaks (purple) for each activation conditions versus the unstimulated control at 24 hours; "tested peaks" passed DESeq2 independent filtering. **(B)** Euler diagram illustrating the intersection of peaks with significantly increased accessibility in activation conditions relative to the unstimulated control (N = 3, FDR < 0.01).



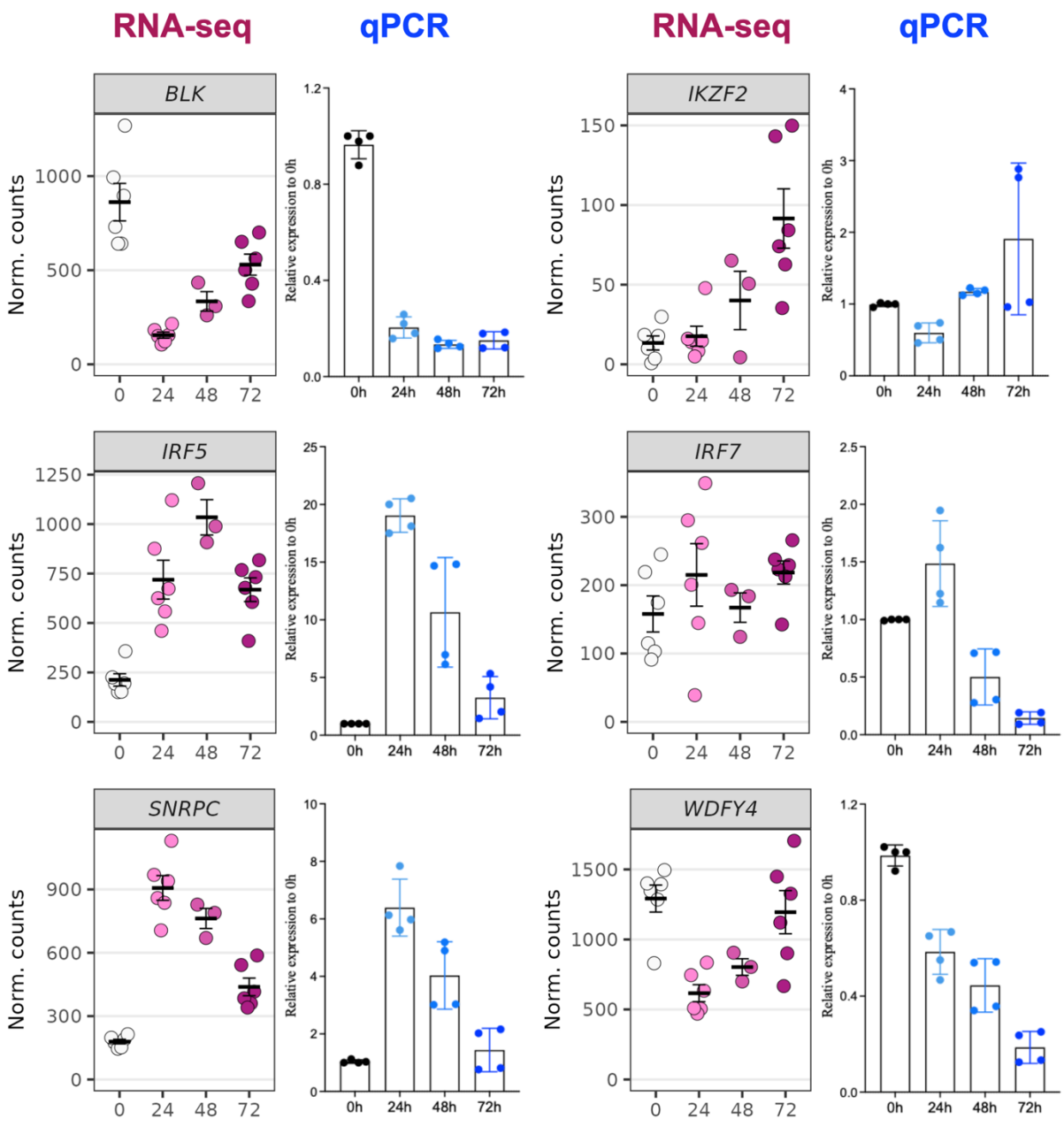
Supplementary Figure 11. Status of transcription factor genes with activation-associated motif enrichment across B cell disease signatures. Genes encoding transcription factors whose DNA-binding motifs are enriched in regions gaining accessibility upon B cell activation (Figure 4c) were examined in B cell gene signatures from Nakano et al. (2022). Heatmap: transcription factor genes (y-axis) and B cell subsets (x-axis), colored by log fold change (logFC). For the disease-state signature (left), logFC reflects patients versus controls; for the disease-activity signature (right), logFC reflects high versus low disease activity.



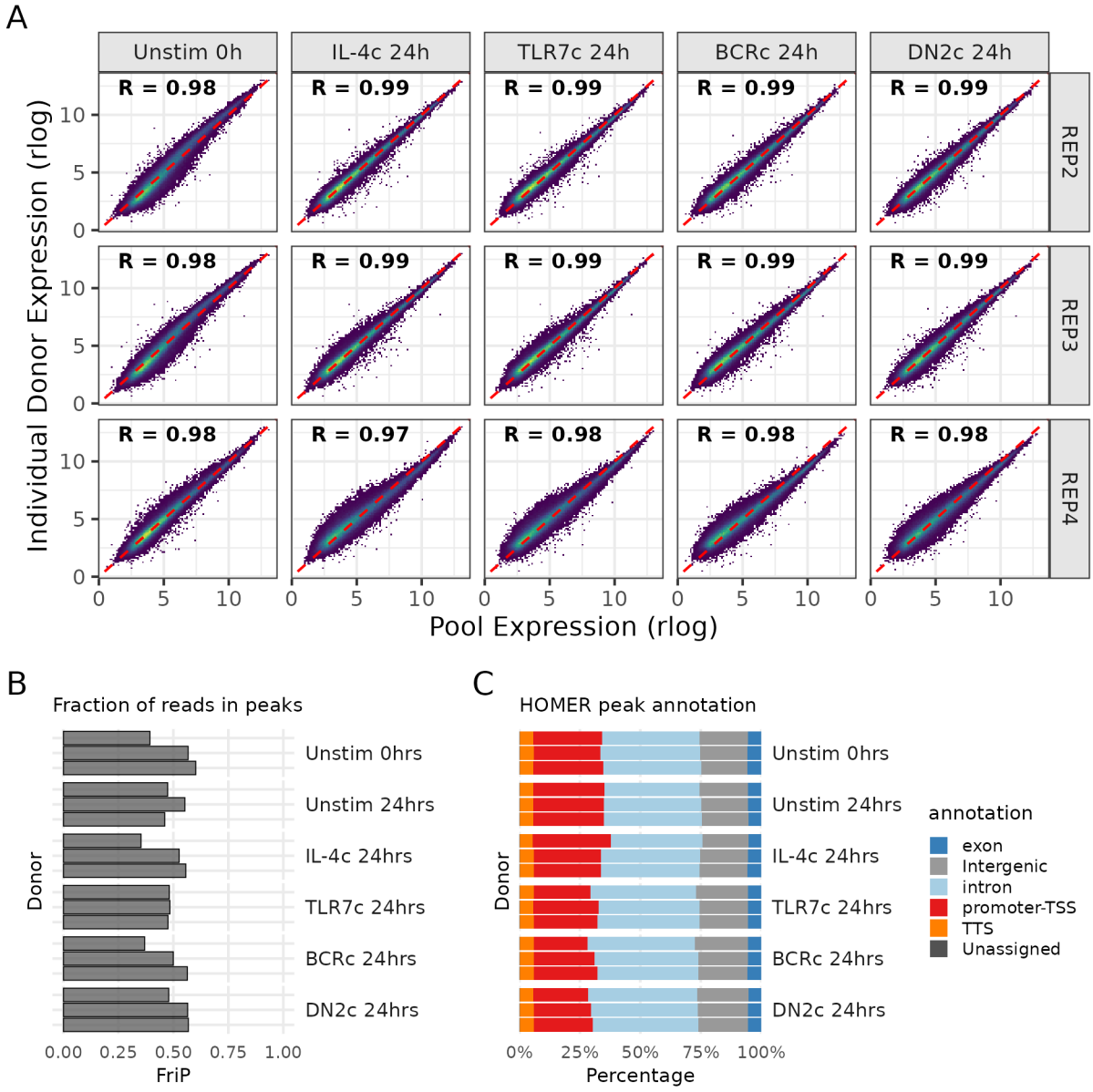
Supplementary Figure 12. Distribution of LDSC Tau* coefficients across 15 immune-mediated diseases. Stratified LD-score regression (S-LDSC) heritability enrichment of immune-mediated disease GWAS within chromatin regions that gain accessibility upon B cell activation, for the simulation conditions IL-4c, TLR7c, BCRc, and DN2c at 24 h (N = 3, Figure 4d). Tau* is the standardized per-SNP effect size of the annotation on trait heritability, jointly estimated with the baseline model and scaled to be comparable across traits. The 15 immune-mediated diseases analyzed here were allergy/eczema, asthma, adult-onset asthma, childhood-onset asthma, celiac disease, Crohn's disease, covid, inflammatory bowel disease, multiple sclerosis, primary biliary cholangitis, psoriasis, rheumatoid arthritis, systemic lupus erythematosus, type 1 diabetes, and ulcerative colitis. For traits with more than one GWAS dataset, the plotted Tau* is the average across studies. The GWAS datasets are listed in Supplementary Data 3, sheet "ATAC heritability enrichment".



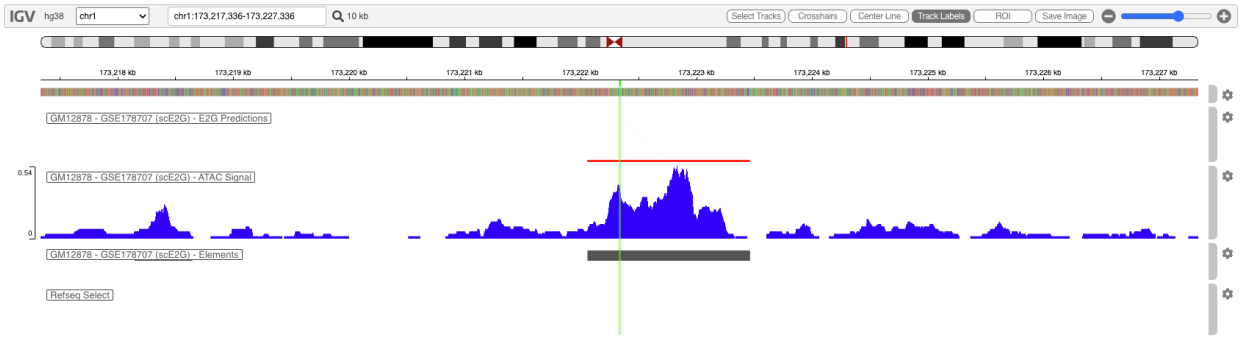
Supplementary Figure 13. Comparison of gene expression fold changes between low-input and standard-input RNA-seq. Log fold changes (LFC) between 0 and 24 h for low-input RNA-seq (x-axis) versus standard-input RNA-seq (y-axis) across conditions (panels). Genes significant at 5% FDR in the low-input are shown with their corresponding LFCs in the standard-input dataset. Point density is indicated by the color scale (yellow/green: high; purple: low). R: Pearson correlation coefficient. $P < 2.2 \times 10^{-16}$ for all panels.



Supplementary Figure 14. Comparison of gene expression dynamics measured by bulk RNA-seq in total B cells and qPCR in naïve B cells across the DN2c time course. Expression of six representative genes measured by bulk RNA-seq and quantitative PCR (qPCR) across the 0–72 h DN2c time course. RNA-seq was performed on total B cells (N = 6, DESeq2-normalized counts); qPCR was performed on naïve B cells from a different set of donors (N = 4, relative expression normalized to the 0-hour time point). Note that *WDFY4* has >60 exons and qPCR captures only a small portion of the gene, which can bias quantification if isoform usage changes over time. Each point is an individual donor; the horizontal lines (RNA-seq) or bars (qPCR) indicate the mean; error bars show the SEM. Source data for qPCR are provided as a Source Data file.



Supplementary Figure 15. Quality control and reproducibility of ATAC-seq libraries. (A) Reproducibility of chromatin accessibility: signal in individual libraries (y-axis, rows) versus an independent 3-donor pool (x-axis), stratified by condition (columns). Point density is indicated by the color scale (yellow/green: high; purple: low); the red dashed line is the identity line (y = x). R: Pearson correlation coefficient. $P < 2.2 \times 10^{-16}$ for all panels. **(B)** Fraction of reads in peaks (FRiP) per sample. **(C)** Genomic annotation of peaks per sample, determined by HOMER.



Supplementary Figure 16. rs2205960 overlaps an accessible chromatin region in the GM12878 lymphoblastoid cell line. Chromatin accessibility (ATAC-seq) at the *TNFSF4* locus in GM12878, an Epstein–Barr virus–transformed immortalized B cell line (LCL), from ENCODE data accessed through the Enhancer2Gene (E2G) portal (https://e2g.stanford.edu/). The green vertical line marks the position of rs2205960.