nature communications

Article | Article in Press

https://doi.org/10.1038/s41467-026-78656-w

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

Received: 4 March 2026

Accepted: 30 September 2026

Published online: 23 October 2026

Check for updates

Vitor R. C. Aguiar[1,2,18], Marcella E. Franco[1,2,3,18], 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[1,17], 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] ✉

B cells are central for immunity and immune-mediated diseases (IMD), but how distinct signaling pathways drive diverse activation responses over time is not comprehensively understood. Here, we obtain peripheral blood B cells from 27 female healthy donors for in vitro stimulation, profiling the response at multiple time points using RNA-seq, single-cell RNA-seq with surface protein markers (CITE-seq), and ATAC-seq. Using diverse stimuli including activation of the B cell receptor (BCR), Toll-like receptor 7 (TLR7), TLR9, CD40, and differentiation into double negative 2 (DN2) $IgD^{-}CD27^{-}CD11c^{+}CD21^{-}$ B cells, our analyses uncover activation-dependent B cell states, stimulation-specific transcription factors and regulatory regions, the dynamic response of IMD-associated gene expression, and specific open chromatin regions associated with IMD genetic risk. Furthermore, we also identify a lupus risk variant in an enhancer regulating *TNFSF4* expression. The analysis data can be accessed via an interactive browser: https://mgalab.shinyapps.io/bcellactivation. Our findings underscore how profiling the full breadth of B cell activation pathways may be important for understanding B cell biology in health and disease.

B lymphocytes are essential players in the immune system through their ability to produce antibodies, present antigens, and release cytokines. While these functions are critical for protecting the host from infections and maintaining immune homeostasis, dysregulated B cell activity can drive the development and progression of immune-mediated diseases (IMD). B cells are highly dynamic, transitioning through various activation and differentiation states in response to diverse external stimuli. Signals from the B cell receptor (BCR), co-stimulatory molecules, cytokines, and pattern recognition receptors converge to define the activation trajectory of B cells. This capacity to reach diverse activation states is critical for immune defense, but when dysregulated, it can also drive autoimmunity or contribute to the development of allergies and cancer[1–3].

A full list of affiliations appears at the end of the paper. ✉e-mail: mgutierr@broadinstitute.org

Understanding how B cells transition into pathogenic states is critical for discerning their role in disease progression and developing targeted therapeutic interventions. One key example of a likely pathogenic B cell state is the double negative 2 (DN2) subset, characterized by the absence of CD27 and IgD expression and the presence of CD11c expression (with some groups also defining them as $CD21^-$ $CD11c^+$ B cells)[4,5]. DN2 B cells are considered a pre-plasmablast state that might emerge through an extrafollicular pathway[4], an alternative maturation pathway independent of germinal center dynamics[6]. These cells are expanded in the peripheral blood of patients with systemic lupus erythematosus (SLE), rheumatoid arthritis (RA), multiple sclerosis (MS), autoimmune anemia, systemic sclerosis (SSc), Sjogren's disease, Crohn's disease (CD), and severe SARS-CoV-2 infection[4,7–9]. DN2 B cells have also been identified within tumor-infiltrating B cells in over 15 cancer types[10]. In mice, $CD11c^+$ B cells accumulate with age (a.k.a. age-associated B cells, ABCs) and can cause autoimmunity[9,11,12]. It has been shown that this likely pathogenic subset can be generated in vitro via stimulation of the BCR and Toll-like receptor 7, alongside key cytokines (IFN-γ, IL-21, IL-2 and BAFF)[4,13].

However, while specific drivers of states like DN2 have been identified, we still lack a comprehensive molecular understanding of how B cells respond to different stimuli over time and of the broader B cell activation landscape. A significant barrier to progress has been the relatively low frequency of B cells in blood and technical challenges of in vitro cultivation. Therefore, most existing functional genomics efforts have focused on cells in a steady state or under a narrow range of stimuli or time points[14,15], leaving the vast majority of the B cell activation space underexplored. It is increasingly recognized that the regulatory elements governing B cell function are highly context-dependent; many genes and pathways only become active under particular states of cellular activation[14,15]. Without a systematic map of these states, we cannot fully grasp how B cells are wired to respond to their environment, or how that transcriptional and epigenetic wiring is fundamentally altered in disease.

This gap in knowledge has direct implications for interpreting the genetic basis of IMDs. While genome-wide association studies (GWAS) have identified hundreds of variants associated with IMDs, most reside in non-coding regions and likely exert their effects by modulating gene regulation[16]. Evidence suggests that risk variants for SLE, MS, SSc, primary biliary cholangitis (PBC), psoriasis, and CD, and RA are highly enriched in B cell-specific regulatory elements[14,16,17]. However, mapping these variants to target genes is often unsuccessful because the relevant regulatory effects may be "hidden" in specific activation states that have not yet been profiled[18,19].

A systematic investigation of immune activation pathways is therefore essential not only to understand B cell biology but to reveal disease-relevant activation conditions and uncover the mechanisms by which genetic risk manifests. In this study, we address these challenges by employing a multi-omics approach—integrating bulk RNA-seq, single-cell RNA-seq with surface protein profiling (CITE-seq), and ATAC-seq—to dissect chromatin remodeling, gene expression, and alternative splicing in response to various activation pathways. By targeting key BCR, TLR and cytokine signaling pathways, we systematically profile B cells to define distinct activation states. This multi-omics resource provides a detailed map of the dynamic changes occurring during activation, offering new opportunities to investigate B cell regulation and providing a framework to understand how disease-associated variants influence B cell biology.

Results

Multi-omics profiling of B cell activation

We recruited 27 unrelated healthy donors from the Mass General Brigham Biobank (MGBB) and the Boston Children's Hospital Biorepository of Adult Healthy Controls, and enriched for $CD19^+$ cells from peripheral blood (Supplementary Fig. 1a, b). We stimulated these B cells in six activation conditions and two control conditions, and we profiled samples at different time points with bulk RNA-seq, bulk ATAC-seq, single-cell RNA-seq, and CITE-seq (Fig. 1).

We used the following stimulation conditions in order to target different activation pathways important in B cells: (1) "CD40c": CD40 ligand (CD40L) to trigger CD40 signaling, which in vivo is mediated by interaction with T cells and possibly other cell types, along with IL-4 to improve B cell survival[20]; (2) "TLR9c": CpG oligodeoxynucleotides to stimulate TLR9, a receptor primarily responsible for detecting bacterial DNA, mimicking bacterial infection[21], combined with IL-4[22]; (3) "TLR7c": resiquimod (R848) to activate TLR7, which recognizes RNA ligands of viral origin, simulating viral infection, along with IL-4[21]; (4) "BCRc": anti-IgG/IgM antibodies to activate the B cell receptor (BCR) signaling pathway, together with IL-4[14,22]; (5) "BCR/TLR7c": dual BCR and TLR7 stimulation with anti-IgG/IgM and R848, supplemented with IL-4; (6) "DN2c": a DN2 B cell-inducing cocktail that includes dual BCR and TLR7 stimulation with anti-IgG/IgM and R848, along with additional cytokines (IL-2, IL-21, BAFF, and IFN-γ). Together these DN2c stimulants have been shown to differentiate naïve B cells into the "double negative 2" subset (IgD^- $CD27^-$ $CD11c^+$) (Supplementary Fig. 1e–h), which is transcriptomically and functionally similar to the DN2 B cells found in SLE patients[4,13]. Here we do not study DN2 B cells themselves, but we study the activation states that are induced by the stimulation cocktail that can differentiate naïve B cells into DN2 B cells. IL-4 was not added to "DN2c". To ensure effective B cell activation, we verified that each of these stimulation conditions induced upregulation of the activation marker CD69 by flow cytometry (Supplementary Fig. 1c, d).

Additionally, we used two control conditions: "Unstim" which includes unstimulated B cells at 0 h as well as cells in media only for 4 and 24 h (we did not include later time points due to excessive cell death if B cells are not stimulated); and "IL-4c", which includes media and IL-4 (Fig. 1).

Dynamic gene expression programs in activated B cells

Using low-input RNA-seq, we profiled the in vitro activated B cells at multiple time points (0, 4, 24, 48, 72 h) from 6 adult healthy individuals (Fig. 1). Since different activation cocktails can cause variability in cell viability, especially at the later time points (Supplementary Fig. 1i), the profiled B cells at each time point and activation condition were pre-sorted for live cells (7-AAD^-). Principal component analysis (PCA) on the top 2000 most variable genes shows that samples group by their activation cell state (Fig. 2a). PC1 separates samples to a great extent by the time points, with the 4-h time point forming one branch, the 24-h time point forming a second branch, and the 48 and 72-h time points clustering closer together in a third branch. PC2 strongly separates samples by the in vitro conditions with control conditions at the top, followed by CD40c, TLR7c, and TLR9c activation conditions, which are separated from the bottom conditions that include BCR stimulation (BCRc, BCR/TLR7c, DN2c) (Fig. 2a, inset). In line with the PCA, pairwise differential expression analysis shows higher numbers of differentially expressed genes for 4 versus 24 h than for 48 versus 72 h, as well as higher number of differentially expressed genes with respect to control conditions in the three BCR stimulation conditions than in the non-BCR stimulation conditions (Fig. 2b). These results suggest that in studies investigating B cell activation response of genes of interest, the time point evaluated matters and that stimulating the BCR itself induces a larger transcriptomic response than stimulating other receptors alone.

The RNA-seq profiling reveals stimulus-specific dynamic responses for genes following B cell stimulation (Fig. 2c and Supplementary Fig. 14). For example, *AICDA*, which encodes the enzyme activation-induced cytidine deaminase involved in class switch recombination and somatic hypermutation, was strongly upregulated in late time points of conditions that include BCR stimulation. *BANK1*, which

encodes the B cell scaffold protein with ankyrin repeats 1 that transduces BCR signals, showed sustained downregulation in conditions that include BCR stimulation, whereas expression quickly recovered after 4 h in the other conditions. *FCER2* (which encodes CD23, an IgE receptor) was repressed only in DN2c. *IRF5*, which codes for an interferon regulatory transcription factor with multiple functions, showed anticorrelated expression dynamics between conditions that include BCR stimulation and those that do not. *TBX21* (which encodes the transcription factor T-bet) was specifically activated in DN2c and not in any other condition. Finally, *XBP1*, which encodes X-box-binding Protein 1, a marker of plasma cell differentiation, showed early upregulation at 4 h in all conditions but exhibited further strong upregulation at 72 h in BCRc, whereas in DN2c it was downregulated.

Next, to understand the broad patterns of gene dynamics, we used weighted gene co-expression network analysis (WGCNA) to leverage the correlation of gene expression to group genes into transcriptional programs ("modules")[23]. Here, we employed gene expression data from unstimulated B cells as the 0-h baseline and transformed gene expression values using variance-stabilizing transformation (VST) to understand expression dynamics relative to time independently of total gene expression levels. When applied to the DN2c time course, we identified eight gene modules (Fig. 2d). To understand the biological processes enriched in each module, we used Gene Ontology (GO) enrichment analysis (Fig. 2e and Supplementary Data 2). This analysis unraveled the functional dynamics of B cell activation. For example, genes involved in B cell activation (modules 4, 5, and 8) are initially downregulated with a rise in gene expression in late time points. Genes in module 4, which has *BTK* (necessary in BCR signaling) as one of its main hub genes, recover higher expression levels already at 24 h, whereas genes in module 5, with *BLK* (also involved in BCR signaling) among the main hub genes, display a slower rise in expression levels. Module 8, with an even slower rise in expression levels, had no significant pathway enrichment, but its main hub gene is *FCRL1*, a promoter of B cell activation[24]. At the same time as these B cell activation genes are downregulated, genes involved in RNA processing and ribosome biogenesis (module 1) go up and peak at 4 h. At 24 h, cellular respiration genes (module 3) are upregulated, followed by a strong induction of mitosis-related genes (module 7) at 48 and 72 h. Overall, these enrichment patterns suggest that once B cells are activated with DN2c, they transiently de-prioritize transcription of

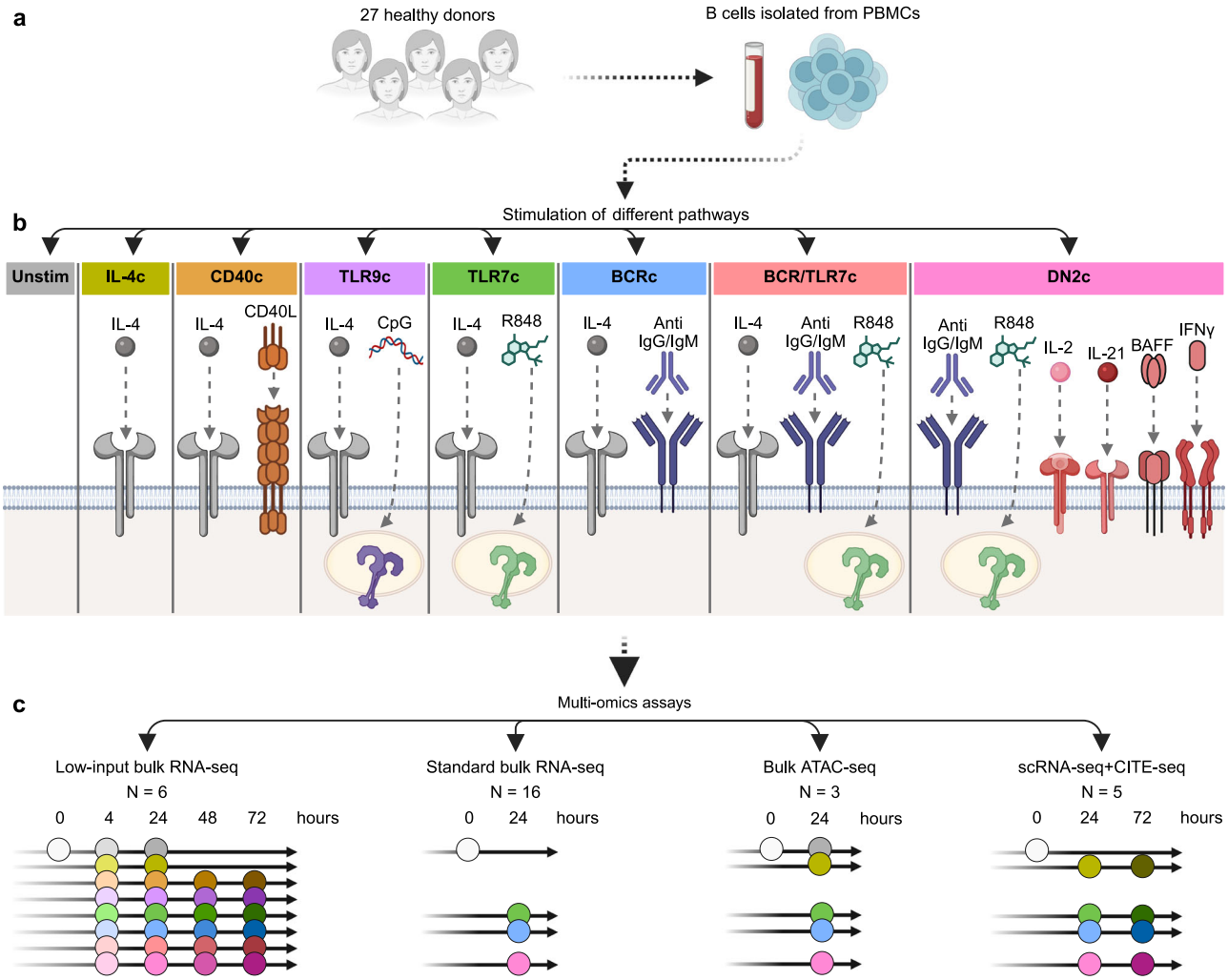


Fig. 1 | Experimental design and overview of B cell activation resource. a To characterize how distinct signaling pathways drive human B cell activation over time, we isolated peripheral blood B cells from 27 healthy donors and stimulated them in vitro under conditions engaging different activation pathways. **b** Activation conditions used (Unstimulated control, IL-4c, CD40c, TLR9c, TLR7c, BCRc, BCR/TLR7c, and DN2c). **c** We profiled the response to stimuli across a time course with multiple genomic assays (low-input RNA-seq, standard-input RNA-seq, single-cell RNA-seq with cell-surface protein quantification (CITE-seq), and bulk ATAC-seq). "N" indicates the number of donors used for each assay; the activation or control conditions used are color-coded as in (**b**). The three donors for which there is ATAC-seq data were profiled for standard bulk RNA-seq as well. "0 hours" indicates freshly isolated B cells not placed in media; gray conditions of 4 and 24 h indicate B cells cultured in media only ("Unstim"). Created in BioRender. Rezende da Costa Aguiar, V. (2026) [https://BioRender.com/ogog9hu].

https://doi.org/10.1038/s41467-026-78656-w

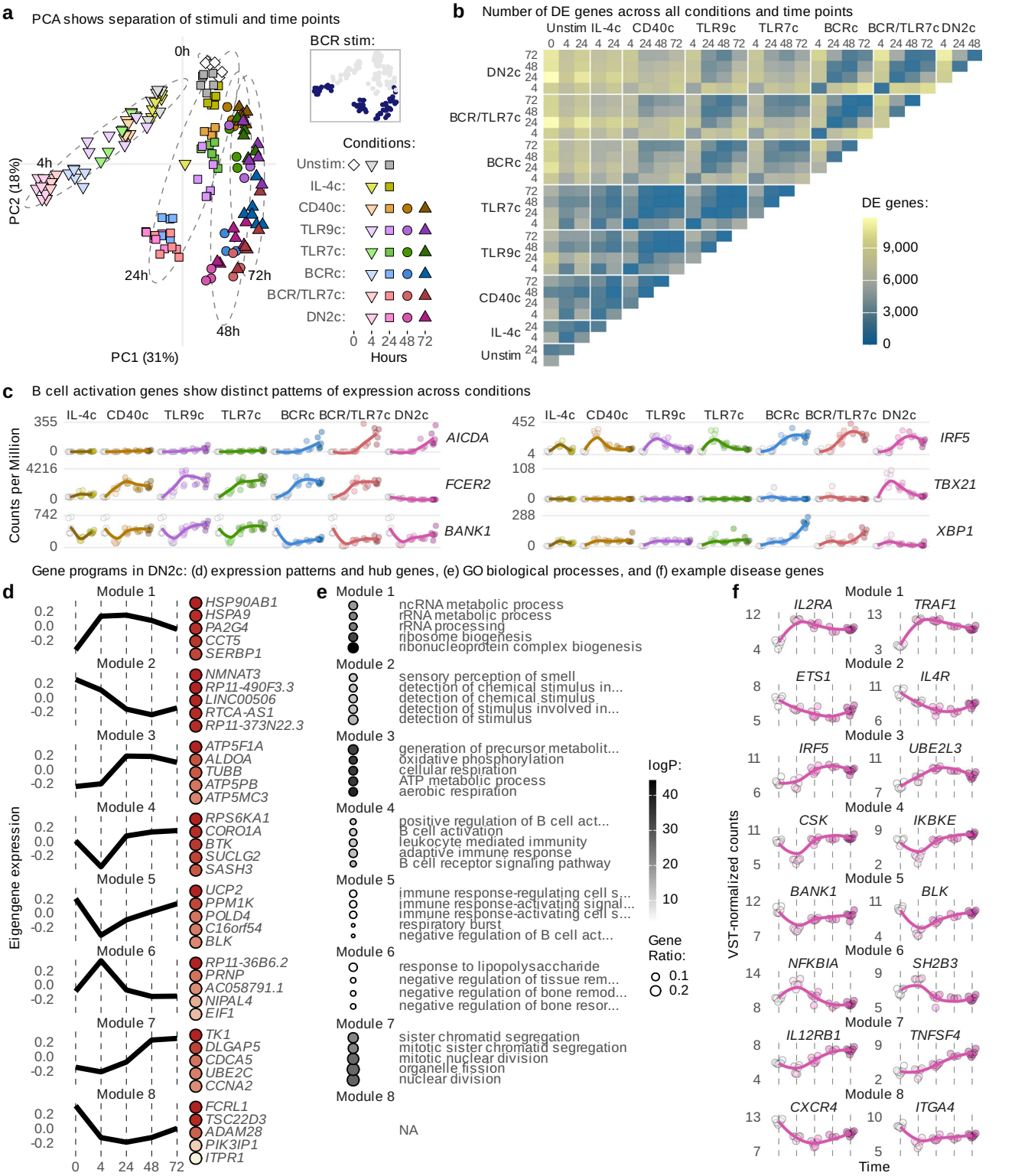


Fig. 2 | Pathway-dependent gene expression and transcriptional dynamics in activated B cells over time. Peripheral-blood B cells from six healthy donors, stimulated in vitro and profiled by low-input RNA-seq over 0–72 h. **a** Principal component analysis (PCA) with the top 2000 most variable genes; position reflects transcriptome similarity, separating conditions and time points. Inset: the same plot highlighting BCR-stimulated conditions. **b** Differentially expressed (DE) gene counts between condition/time-point pairs (edgeR; Benjamini-Hochberg adjusted $P < 0.05$). **c** Expression (counts per million) of selected genes over time; points, donors; lines, per-condition loess trends. **d** Weighted gene co-expression network analysis (WGCNA) modules in DN2c: module eigengene (expression summary) versus time (hours). Right, main hub genes per module, colored by intra-module connectivity (kIM). **e** Gene Ontology (GO) biological processes enriched per module (**d**) by over-representation analysis (clusterProfiler); fill, $-\log_{10} P$; size, gene ratio. **f** Expression of selected immune-mediated disease-associated genes from modules in (**d**).

B cell activation genes and, instead, prioritize transcription of translational machinery genes, which probably aids in subsequent production of energy which is then used for B cell proliferation.

We performed the same module identification and pathway enrichment analysis for the other activation time courses (Supplementary Fig. 2) and evaluated the degree of module sharing among activation conditions. From the modules identified in each activation time course, we calculated the WGCNA statistic $Z_{summary}$ to estimate the degree of preservation with each of the other activation time courses (Supplementary Fig. 3)[25]. We observed that the three activation conditions that stimulate the BCR (DN2c, BCRc, BCR/TLR7c) have higher module preservation among them, and less preservation with the non-BCR activation conditions (CD40c, TLR7c, TLR9c). A similar pattern is observed in the reciprocal direction, with CD40c, TLR7c, and TLR9c having higher module preservation among them, and lower with the three BCR-stimulating conditions. For example, DN2c module 7 (Fig. 2d), which is the nuclear division-enriched module that presents late upregulation, has high preservation with BCRc and BCR/TLR7c ($Z > 20$), and moderate to low preservation with TLR9c, TLR7c, and CD40c ($Z < 12$). Despite the high preservation among conditions that include BCR stimulation, this "proliferation program" is most prominent in DN2c; genes overexpressed in DN2c compared to BCRc at 72 h reveal further enrichment for biological processes related to DNA replication (Supplementary Data 4). In contrast, CD40c module 7 (Supplementary Fig. 2), which presents late downregulation of ribosome biogenesis pathway genes, has high preservation with TLR9c and TLR7c, but lower preservation with the three BCR-activating conditions. Notably, several modules were highly conserved among all 6 activation conditions. For example, DN2c module 4 (Fig. 2d), showing quick down-regulation followed by slow recovery of genes enriched in B cell activation and B cell signaling pathways, shows Z values > 20 with all other time courses. In contrast, TLR7c module 8 which presents a "zig-zag" expression pattern and no enriched pathways (Supplementary Fig. 2), has the lowest preservation with all other time courses. Overall, these results suggest there are some "universal" response programs in B cell activation, some that are dependent on whether the BCR was stimulated or not, and some that may be unique to a particular activation condition.

Next, we turned back to our main module analysis on the DN2c time course and assessed the dynamic transcriptional programs to which IMD-associated genes belong (Fig. 2f). We defined IMD-associated genes as the gene at each genome-wide significant GWAS risk locus prioritized by the Open Targets Genetics Locus-to-Gene (L2G) model[26]. Module 1 includes, for example, *IL2RA*, which encodes the alpha chain of the IL-2 receptor complex (also known as CD25) and is a central immune modulator associated with several IMDs[27–36], and *TRAF1*, whose expression drives risk of inflammatory arthritis in children and adults[34,37]. *ETS1* and *IL4R* are members of module 2. *ETS1*, an important regulator of lymphocyte differentiation, is associated with multiple IMDs[28,33,34,38–41]. *IL4R*, involved in IgE class switching and secretion from B cells among other functions[42], is associated with severe asthma and is a therapeutic target for asthma[29,43]. As example genes in module 3, *IRF5* is an interferon response gene implicated in autoimmune diseases such as RA, SLE, Sjögren's, SSc[34,40,44–46], while *UBE2L3* is involved in ubiquitination to target proteins for degradation and is also associated with several IMDs[32,34,40,46–48]. In module 4, *CSK* encodes a Src kinase that regulates B cell activation and is associated with SLE and SSc[46,49], and *IKBKE*, which is associated with SLE and encodes the kinase IKKε[50]. In module 5, we have *BANK1* and *BLK*. *BANK1* is associated with SLE[40,46]; while *BLK* is associated with RA, SLE, and SSc[34,40,45,46]. In module 6, *SH2B3* encodes an adaptor protein involved in various signaling activities by growth factors and cytokine receptors, and is implicated in dozens of human traits including autoimmune diseases[31,33–36,46–48], and *NFKBIA*, a negative regulator of NFκB that is associated with asthma and psoriasis[29,51]. In module 7, *IL12RB1* encodes a subunit of the IL-12 receptor that is associated with SSc[49], whereas *TNFSF4* encodes OX40L on the surface of B cells and is associated with asthma, eczema, RA, and SLE[27,29,46,52]. In module 8, *CXCR4* is associated with MS[33], and *ITGA4*, which encodes the integrin subunit alpha four involved in cell motility and signaling, and is associated with inflammatory bowel disease (IBD)[47]. Notably, within a disease, candidate causal genes belong to different gene modules. For example, given the modules identified in DN2c, five candidate SLE risk genes are tightly associated (kME > 0.85) with module 4, four genes with module 5, three genes with module 6, two genes with module 3, and one each with modules 1 and 7 (Supplementary Fig. 4). Overall, this wide dispersion of IMD-associated genes across gene modules underscores the complexity of IMDs, showing that candidate causal genes may play key roles at different time points of B cell response to activation.

Finally, to further assess the clinical relevance of these activation states, we intersected our gene expression data with transcriptional signatures derived from sorted B cell subsets of SLE patients[53]. We examined the overlap with "disease-state" signatures (reflecting disease establishment by comparing inactive SLE versus healthy controls) and "disease-activity" signatures (reflecting disease exacerbation by comparing high versus low disease activity) across naïve, memory, and double negative B cells, as well as plasmablasts[53]. We found that 22–78% of genes upregulated in the SLE disease-state signature—and 14–71% of genes in the SLE disease-activity signature—are also upregulated during in vitro B cell activation (Supplementary Fig. 5). This enrichment was most pronounced in conditions involving BCR stimulation at later time points (24, 48, and 72 h). These results suggest that in vitro B cell activation, particularly via BCR engagement, successfully recapitulates key transcriptional programs relevant for both the establishment and exacerbation of SLE.

Single-cell profiling shows pathway-dependent B cell states

We performed single-cell RNA-seq coupled with surface protein expression profiling (CITE-seq) on total B cells extracted from five genotyped healthy participants. We assayed the following conditions: unstimulated cells at day 0, and IL-4c, TLR7c, BCRc, and DN2c at 24 and 72 h (Fig. 1). We performed this experiment in three batches—one batch including two donors, and two batches on the same three donors—pooling cells from multiple conditions and donors into one run at each batch. Before pooling, cells from different conditions were tagged with specific hashtag oligonucleotides. After pooling and before cell encapsulation, live B cells were sorted ($CD19^+$ 7-AAD^-). After sequencing, demultiplexing, and quality control to remove empty droplets, doublets, and dying cells, we obtained ~37,700 cells.

Single-cell RNA-seq revealed stimulus-dependent B cell states (Fig. 3). By looking at both RNA expression and surface protein expression, we observed that BCR engagement in BCRc and DN2c leads to decreased expression of IgD/IgM on the surface of cells in comparison with the Day 0, IL-4c, and TLR7c conditions ($p_{adj} < 0.05$, Fig. 3c and Supplementary Data 5). Concordantly, at 24 h, cells in these conditions also upregulated CD69 protein expression in comparison to IL-4c and TLR7c ($p_{adj} < 0.05$), indicating high B cell activation after BCR stimulation (Fig. 3c and Supplementary Data 5). Next, we quantified discrete B cell subpopulations based on expression thresholds for established markers (Supplementary Fig. 6). We found that BCRc and DN2c contained high frequencies of IgD^- $CD27^-$ cells, and were enriched for proliferating *MKI67*-high cells and $CD11c^+$ B cells (Fig. 3e). The BCRc condition contained the highest levels of IgD^- $CD27^+$ B cells and of cells highly expressing *XBP1*, a marker of plasma cell differentiation (Fig. 3e).

Transcriptome-based clustering identified 13 distinct states (Fig. 3b, d). Naïve-like B cells expressing *TCL1A*, *FOS*, and *DUSP1* were mostly found among unstimulated cells (cluster 12). Upon stimulation, B cells changed to different cell states. For example, by 72 h, BCRc-stimulated cells had distributed largely into a *CD27*$^+$ "memory-like"

https://doi.org/10.1038/s41467-026-78656-w

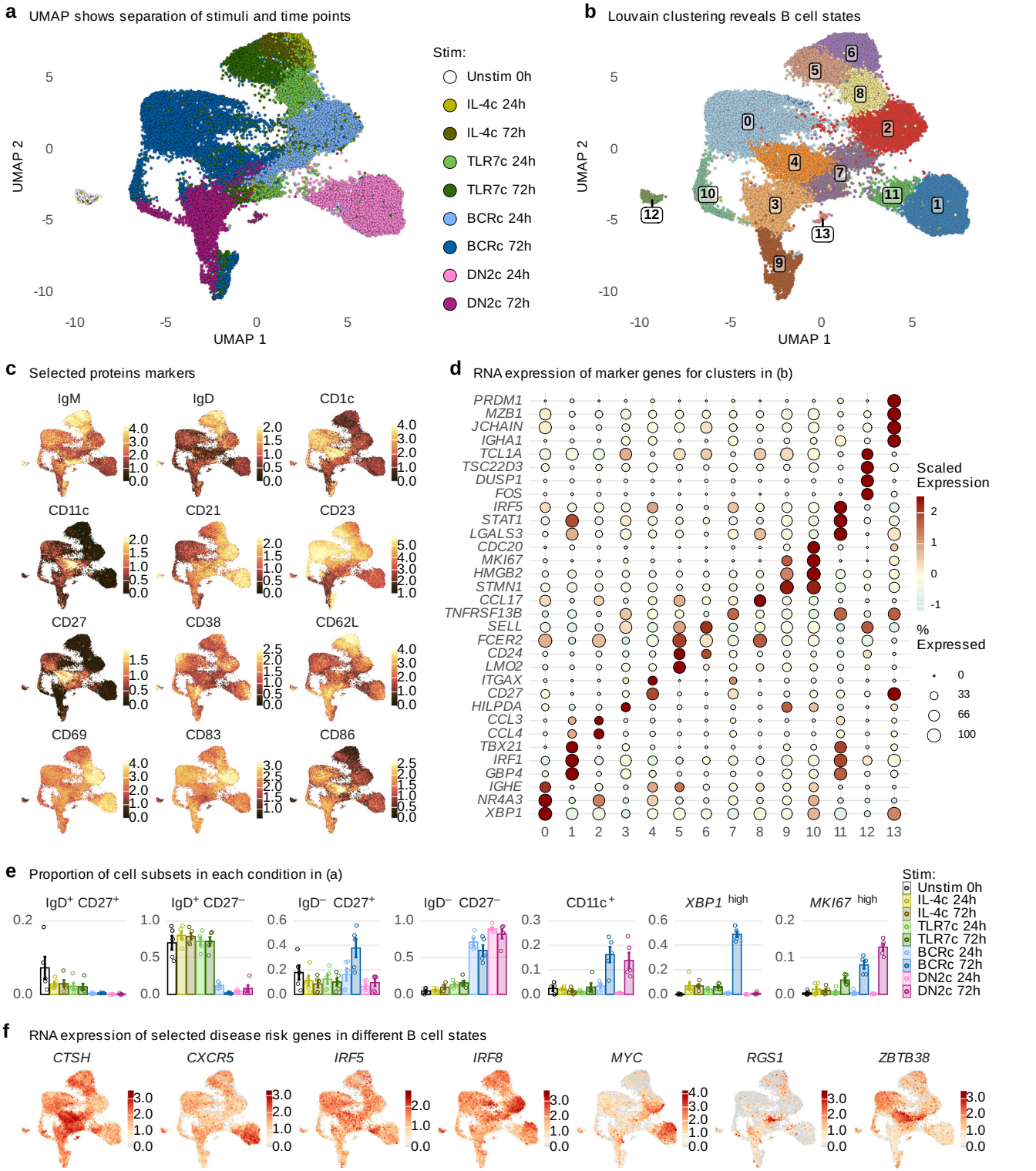


Fig. 3 | Single-cell RNA expression coupled with quantification of 137 surface protein markers (CITE-seq) reveals pathway-dependent B cell fates. B cells from five healthy donors were stimulated and profiled by droplet-based single-cell RNA-seq with oligo-conjugated antibodies (CITE-seq). **a** Uniform Manifold Approximation and Projection (UMAP) computed from gene expression principal components after batch integration with Harmony. **b** UMAP as in (**a**) colored by transcriptome-based clusters (Louvain clustering). **c** UMAP as in (**a**) colored by surface-protein expression of selected surface markers. **d** RNA expression of selected cluster marker genes identified by differential expression across clusters. **e** Proportion of B cell subsets, defined by gating on marker protein or gene expression (genes in italic), within each condition and time point. Points represent individual donors ($N = 5$); bars indicate the mean and error bars the SEM. **f** UMAP as in (**a**), colored by RNA expression of selected IMD-associated genes.

state (cluster 4; mean 18% [range: 7–30%]), a *XBP1*$^{+}$ *IGHE*$^{+}$ "early plasmablast-like" state (cluster 0; mean 67% [range: 52–78%]), and *MKI67*$^{+}$ "proliferative" states (clusters 9 and 10; mean 9% [range: 6–15%]).

To investigate the dynamics of our bulk-derived gene modules across single-cell states, we quantified module activity scores at the single-cell level. This analysis revealed marked heterogeneity in module expression across cell states (Supplementary Fig. 7). For example, following DN2c stimulation, the bulk derived module 7 ("proliferation program", Fig. 2d, e) showed high expression specifically within single-cell clusters 9 and 10 (proliferating B cells, Fig. 3b, d), and not in cluster 3 where most of the other cells from the DN2c 72-h timepoint group. This single-cell integration demonstrates that the broad temporal trajectories observed in bulk can be driven by specific cell states.

Finally, we looked at the expression patterns of IMD-associated genes across different B cell states (Fig. 3f; see Supplementary Fig. 8 for individual-level data by cluster). For example, *CTSH*, which encodes for Cathepsin H and is associated with MS and type 1 diabetes (T1D)[33,36], is significantly enriched within the CD27^{+} cluster 4 that we observed at 72 h after BCRc stimulation (5% FDR, Supplementary Data 5). *CXCR5*, which encodes a B cell homing factor and is located in a risk locus for multiple IMDs[29,32,33], is significantly enriched within the DN2c population at 24 h (5% FDR, Supplementary Data 5). Transcription factor gene *IRF5*, associated with many autoimmune diseases and one of the strongest non-MHC associations with SLE[27,34,40,44–46], is expressed mainly in cluster 11 (5% FDR, Fig. 3f and Supplementary Data 5), a state of DN2c with high expression of *STAT1* and *LGALS3* (Fig. 3d). *IRF8*, another interferon response transcription factor gene associated with IMDs such as IBD, PBC, SLE, and SSc[46,47,54–56], exhibits heterogeneous expression patterns, with strong, significant enrichment within the BCRc condition at 24 h (5% FDR, Supplementary Data 5). *RGS1*, which is associated with multiple autoimmune diseases[33,48,57,58], is significantly upregulated within the CD27^{+} cluster 4 (5% FDR, Supplementary Data 5). Finally, *ZBTB38*, associated with asthma[29], increases expression in BCRc with time (5% FDR, Supplementary Data 5), consistent with the bulk RNA-seq data (5% FDR). Overall, these examples illustrate how expression levels of disease-associated genes can be highly context-dependent.

B cell activation induces epigenomic changes relevant for IMDs

Different stimulation conditions can alter B cell states not only by modifying gene expression but also by remodeling chromatin. To investigate how chromatin accessibility changes with activation, we performed ATAC-seq on three healthy donors for conditions TLR7c, BCRc, and DN2c, at 24 h, and used as controls unstimulated cells at 0 h, and cells in IL-4c or in a media-only condition for 24 h.

We observed significant differences in chromatin accessibility across conditions. We performed PCA on the top 5000 most variable peaks. PC1 captures a gradient of activation strength, with control conditions on the left, followed by TLR7c, then BCRc, and finally DN2c at the other extreme (Fig. 4a). We performed differential accessibility analysis across all pairwise combinations of conditions (Fig. 4b, Supplementary Fig. 9). The number of differentially accessible peaks between activation conditions and unstimulated cells goes in line with the observed PC1 gradient, with DN2c presenting the highest number of differentially accessible regions compared to unstimulated cells at 24 h, followed by BCRc and then TLR7c. These results are concordant with what we observed in the RNA-seq data, with DN2c being the strongest stimulation, inducing the largest number of changes.

To identify which transcription factors might be involved in the chromatin remodeling of B cells after activation, we selected peaks that are significantly more accessible at 1% FDR in an activation condition compared to the unstimulated condition at 24 h (second column in Fig. 4b). We further filtered for peaks with a large effect size ($\log_2$ fold change > 1) and employed HOMER to identify transcription factor (TF) motifs enriched in open chromatin regions at each activation condition (Supplementary Data 3)[59]. These motifs were then further filtered to belong to TFs that appear expressed in the low input RNA-seq data at 4 or 24 h post-activation (Fig. 4c). Overall, activated B cells demonstrated enrichment of multiple motifs, with DN2c exhibiting the highest number of enriched motifs and the greatest fold enrichment. Motifs of TFs that harbor basic leucine zipper (bZIP) domains showed the strongest enrichments, with the most significant TF motif being the one of BATF. BATF has previously been described as a pioneer transcription factor involved in chromatin remodeling and class switch recombination in lymphocytes. The DN2c condition also exhibited significant enrichment of the IRF family of TFs, as well as *PRDM1* (BLIMP-1) and T-bet.

To further investigate the relevance of these enriched TF motifs to disease, we identified those corresponding to transcription factors differentially expressed in the Nakano et al. SLE "disease-state" and "disease-activity" B cell signatures[53]. The TF motifs enriched in activation-dependent open chromatin regions correspond to >30 TFs that are in the SLE "disease-state" or "disease-activity" signatures (Supplementary Fig. 11). For example, *BATF* and *STAT1* are upregulated in naïve B cells, unswitched memory B cells, switched-memory B cells, double negative B cells, and plasmablasts of SLE patients compared to healthy controls. *STAT1* is further upregulated in all B cell subsets of SLE patients with high disease activity compared with those with low disease activity, together with *STAT3* and *POU2F2* (OCT2) in plasmablasts, and *RUNX1* in naïve B cells and plasmablasts.

Finally, we used linkage disequilibrium score regression for specifically expressed genes (LDSC-SEG) to integrate our chromatin data with GWAS data to test whether chromatin regions that become accessible following stimulation are enriched for genetic risk of IMD. This is a method that has been widely used to find the specific cell types or cell states at which disease risk variants likely affect gene regulation, hence pointing to disease-relevant cell states that likely mediate genetic susceptibility (e.g., activated T cells for RA, rhinovirus-infected airway epithelial cells for asthma)[14,60,61]. We utilized this method by providing annotations composed of the open chromatin regions that are significantly more accessible in each of the activation conditions at 24 h versus the 24 h non-stimulated control (1% FDR), i.e., IL-4c vs. non-stim, TLR7c vs. non-stim, BCRc vs. non-stim, and DN2c vs. non-stim. We then provided summary statistics of 15 immune-mediated diseases (some of which had multiple independent GWAS performed), and 6 "negative control" diseases or traits for which we do not expect an immune-mediated mechanism, hence we do not expect an enrichment of genetic risk in open chromatin regions specific for activated B cells. Using this method, we observed significant heritability enrichment (1% FDR) in activation-dependent peaks for multiple IMD traits, including RA, SLE, T1D, MS, PBC, IBD and its subtype CD, and celiac disease (Fig. 4d and Supplementary Data 3). For most of these diseases, the heritability enrichment was higher for the open chromatin regions that become more accessible with the DN2c (Supplementary Fig. 12). Notably, we did not find significant enrichment in "negative control" traits for which we would not expect activated B cells to be relevant, such as type 2 diabetes (T2D), schizophrenia (SCZ), major depressive disorder (MDD), LDL and HDL cholesterol, and height (Fig. 4d). We included a COVID-19 infection GWAS given that it has been reported that patients with severe COVID-19 present expansion of DN2 B cells[62], but we did not find a significant heritability enrichment in open chromatin regions specific for activated B cells for this trait. Overall, these findings underscore the relevance of studying activated B cells to understand genetic associations with IMDs and suggest that some GWAS risk loci with yet unexplained mechanisms can be uncovered by studying their regulatory effects in activated B cells.

https://doi.org/10.1038/s41467-026-78656-w

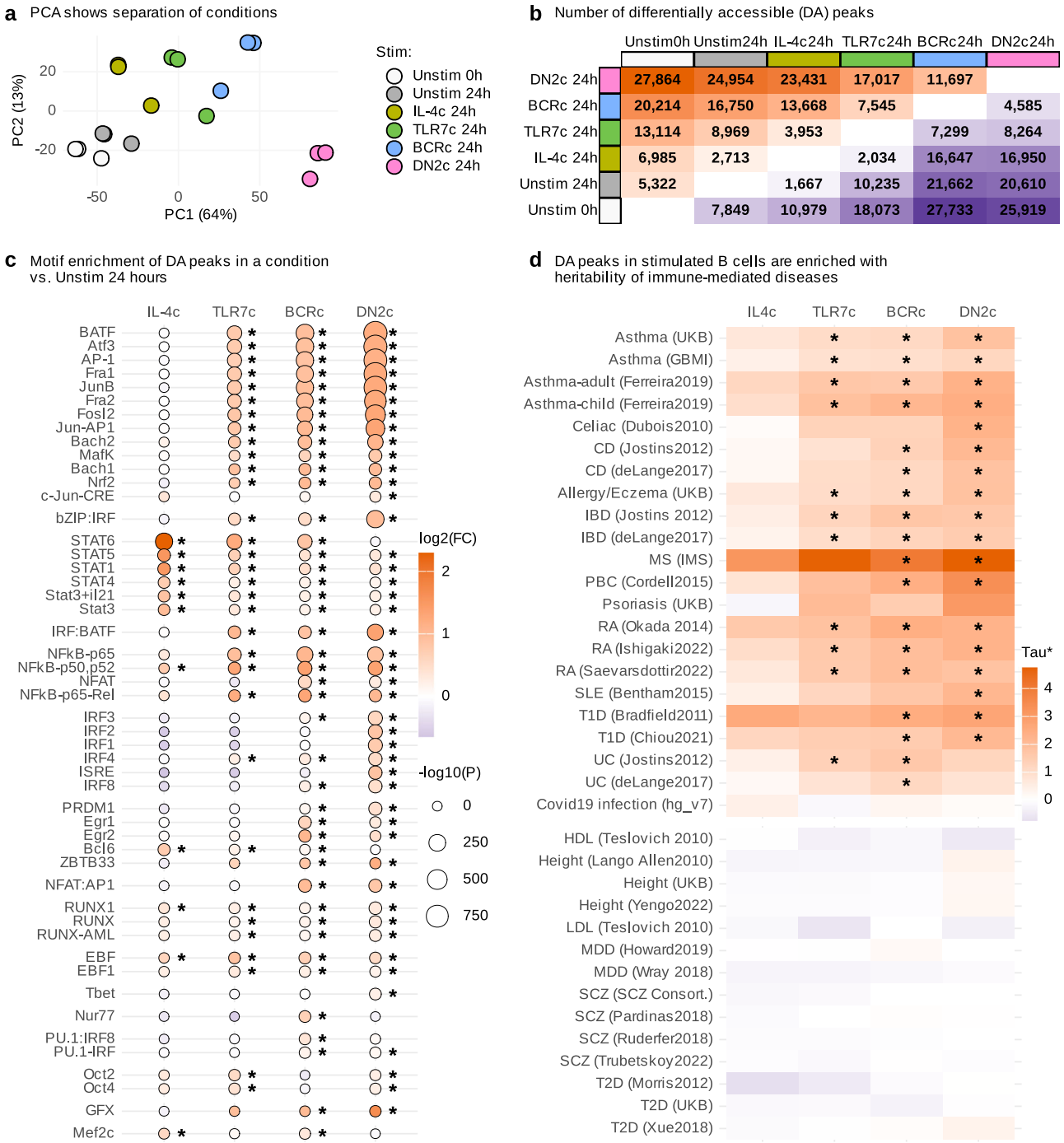


Fig. 4 | Chromatin profiling in activated B cells using bulk ATAC-seq uncovers disease-relevant cell states. B cells from three healthy donors were stimulated and profiled by bulk ATAC-seq at 0 and 24 h. **a** Principal component analysis on the 5000 most variable peaks, grouping samples by overall accessibility. **b** Pairwise counts of differentially accessible peaks (DESeq2, 1% FDR) between conditions; orange indicates higher accessibility in the row versus column conditions; purple indicates lower accessibility in the column versus row conditions. **c** HOMER transcription factor motif enrichment (binomial test, one-sided) in regions gaining accessibility at 24 h post-stimulation relative to unstimulated controls. Dot fill, log2 fold change; dot size, −log10 P; asterisks denote significance after Benjamini-Hochberg correction (FDR < 0.01). **d** Stratified LD-score regression applied to specifically expressed/accessible annotations (LDSC-SEG) estimating heritability enrichment of immune-mediated diseases (top) and control traits (bottom) within these stimulation-induced regions. Tau*, standardized effect size; asterisks denote FDR < 0.01.

B-cell activation-specific chromatin region overlapping IMD risk variant regulates *TNFSF4*

Motivated by the IMD heritability enrichment in stimulated B cells, we sought to investigate whether open chromatin regions of activated B cells could provide insights into potential regulatory mechanisms in IMD risk loci. To do this, we fine-mapped SLE GWAS data and intersected it with our ATAC-seq data to find putative causal variants falling in open chromatin regions that are specific to activated B cells. As proof of concept for this approach, we decided to focus on the chr1q25.1 risk locus, where a single candidate causal variant (rs2205960) was found in independent cross-ancestry fine-mapping studies leveraging linkage disequilibrium differences across

https://doi.org/10.1038/s41467-026-78656-w

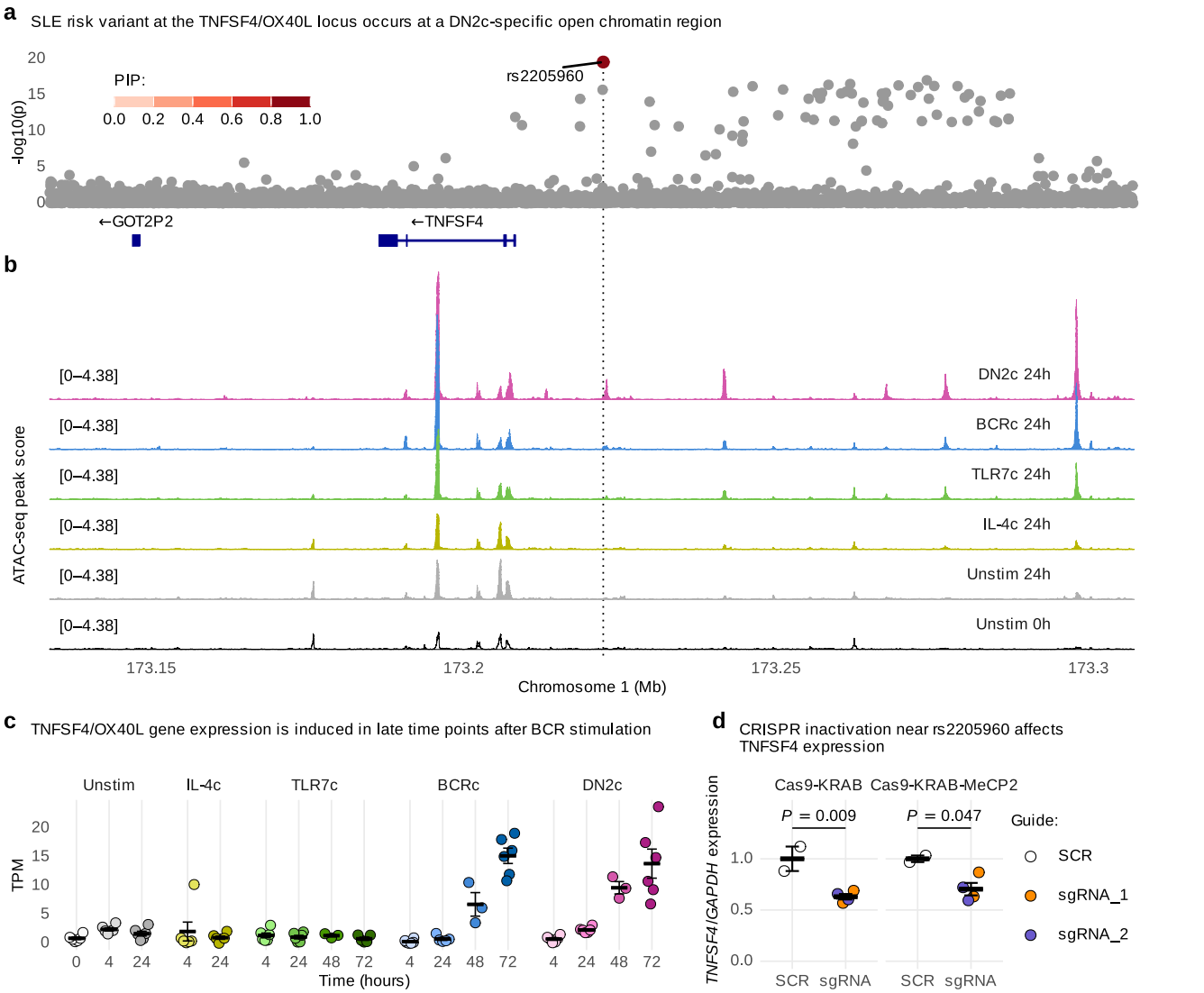


Fig. 5 | An SLE risk variant overlapping a DN2c-specific open chromatin region lies in an enhancer that regulates TNFSF4. a Regional association from the Langefeld et al. SLE GWAS at the TNFSF4 locus; fine-mapped variant (rs2205960) is highlighted and colored by posterior inclusion probability (PIP) from SuSiE fine-mapping. **b** ATAC-seq accessibility tracks across conditions at the genomic region in (**a**). **c** TNFSF4 expression (low-input RNA-seq) in the same conditions shown in (**b**); points are donors, horizontal lines the mean, error bars the SEM. TPM transcripts per million. **d** CRISPRi targeting the region containing the putative causal variant reduces TNFSF4 RNA expression in GM12878 cells, using two repressor systems (dCas9-KRAB and dCas9-KRAB-MeCP2). Each points shows an experimental unit replicate (mean of two qPCR technical replicates; $n = 2$ per guide), horizontal lines indicate the mean, error bars the SEM; expression is relative to GAPDH and standardized to the scrambled control (SCR). For each repressor system, GM12878 cells were split into 6 independent experimental units, 2 with the scramble control sgRNA (SCR), and 2 for each sgRNA targeting the regulatory element. The experimental units from both sgRNAs were compared to SCR by a two-tailed unpaired Student's t-test on ΔCt values. Source data for (**d**) are provided as a Source Data file.

populations[58,63]. Additionally, we performed fine-mapping with in-sample linkage disequilibrium estimation in the Langefeld et al. GWAS data[40]. Concordant with previous findings, rs2205960 was the single variant included in the main credible set with a posterior probability greater than 0.99 (Fig. 5a).

This intergenic variant is located 15 kb upstream of *TNFSF4*, within a region where, in our data, chromatin is open exclusively in the DN2c condition (Fig. 5b). *TNFSF4* is expressed at late time points in conditions that include BCR engagement (BCRc and DN2c) and is more expressed in DN2c than in BCRc (Fig. 5c). To investigate potential regulatory targets of rs2205960, we utilized the Enhancer2Gene (E2G) portal, a comprehensive resource integrating predictive models, chromatin state measurements, 3D contacts, and large-scale genetic perturbations to map enhancer-gene interactions across more than 350 cell types and tissues[64]. We found 7 enhancer-to-gene predictions for rs2205960, all pointing to *TNFSF4* as the target gene. The top 3 (with scores >0.5) are for lymphoblastoid cell lines (LCLs, Epstein-Barr Virus-transformed immortalized B cells), one for GM06990, based on the ENCODE-rE2G model, and two for GM12878, based on the ENCODE-rE2G and the scE2G models (Supplementary Fig. 16). The bottom 4 predictions with scores <0.5 are for activated T cells. Given the scarcity of primary B cell activation datasets, these data in LCLs corroborate our observation that this locus acts as a regulatory element within the activated B cell lineage. To assess whether the open chromatin region overlapping rs2205960 indeed functions as a regulatory element for *TNFSF4* expression, we used CRISPR inactivation (CRISPRi) in the GM12878 cell line. First, we generated stable LCLs that constitutively express dCas9-KRAB or dCas9-KRAB-MECP2 systems, which have repressor domains that can epigenetically suppress the target region of interest. We designed two single guide RNAs to bind

near rs2205960 and we nucleofected them separately in each of the CRISPRi system cell lines. Inhibition of the region that contains the variant reduced *TNFSF4* expression with both single guide RNAs and both repressor systems (Fig. 5d), consistent with this region acting as an enhancer that regulates *TNFSF4*.

Pervasive stimulus-dependent splicing changes during B cell activation

Cell states are shaped not only by the epigenome and the level of expression of genes, but also by what particular isoforms are expressed by each gene. To quantify alternative splicing in a robust way, high-depth RNA-seq libraries that cover the full transcript are needed. Low input RNA-seq yields full transcript information but, starting from 1000 cells per sample, it yields low complexity libraries that present shortcomings for isoform quantification. Droplet-based single-cell RNA-seq only quantifies one end (5′ or 3′) of each gene, making it difficult to ascertain which middle exons are included or excluded. Hence, we performed standard RNA-seq with deep sequencing (at least 40 M pairs of 150 bp reads per sample) in 16 healthy donors. We utilized 500,000 cells per sample to capture a wide dynamic range of expressed isoforms. Given the larger input material, we performed full transcriptome profiling on only four conditions per donor: unstimulated B cells (the 0 h time point), as well as TLR7c, BCRc, and DN2c for 24 h (Fig. 6a). Gene expression changes between activation conditions and unstimulated B cells replicated well with the ones observed for the low input RNA-seq data (Supplementary Fig. 13).

Next, we quantified RNA splicing with the LeafCutter pipeline[65], which uses an annotation-free algorithm to measure intron excision rates and performs differential splicing analysis with a Dirichlet-multinomial model. At 5% FDR, we identified 5115, 5760 and 5668 genes with differential splicing between unstimulated B cells and 24 h-activated B cells with TLR7c, BCRc, and DN2c, respectively (Fig. 6b, c). Between activation conditions, we found 501–3399 genes with significant differential splicing at 5% FDR, with the largest differences found between TLR7c and the two conditions that stimulate the BCR (BCRc and DN2c, Fig. 6b). We performed gene set enrichment analysis (from Gene Ontology) on the differentially spliced genes between each of the activation conditions and non-stimulated cells (Fig. 6d). For all three activation conditions "mRNA splice site selection" was within the top 2 gene sets enriched, with other gene sets related to RNA splicing also significantly enriched. This has also been observed in activated T cells and PBMCs, where genes that code for proteins involved in alternative splicing and/or RNA processing, such as *CELF2*, *HNRNPD*, and *SFRS2*, present differential splicing upon immune activation[66]. Notably, differentially spliced genes for TLR7c and BCRc are enriched in germinal center formation pathways, but this gene set is not significant in DN2c. As a proof-of-concept to demonstrate how stimulation-dependent splicing in B cells can capture trait-relevant events, we highlighted two examples of previously characterized colocalizations between trait genetic associations and splicing quantitative trait loci (sQTL)[67]. At the *BCL2A1* gene, we examined a splicing event linked to a dynamic splicing-QTL during B cell maturation that colocalized with lymphocyte count association, and found that the same junction is significantly upregulated following DN2c stimulation (Fig. 6e). Similarly, a *STAT6* splicing junction—whose splicing-QTLs in both $CD16^+$ myeloid and NK cells colocalize with the genetic association signal for asthma susceptibility—is also induced in DN2c-activated B cells (Fig. 6f). Overall, our analysis of alternative splicing shows that isoform usage drastically changes in B cell activation, and which specific activation pathways are induced can also influence alternative splicing of genes. It remains to be determined which of these transcript isoform changes lead to differences in protein function.

Discussion

B cells play crucial roles in our immune system, such as producing antibodies and modulating other immune cell types. Their dysfunction is implicated in many IMDs. However, despite their importance in health and disease, insights into how healthy B cells respond to different stimuli remain limited. In this study, we generated multi-omics data to comprehensively profile gene expression and chromatin remodeling during B cell activation through multiple receptor pathways.

Using bulk RNA-seq, we profiled how B cells transcriptionally respond to different activation conditions over time. We identified gene programs (or modules) that illuminate how B cells respond to stimulatory signals, including transcriptional changes associated with B cell activation, energy generation, ribosome biogenesis, protein translation, and cell division. Each of these modules exhibits distinct temporal dynamics. While these modules are conserved across activation conditions, BCR stimulation is crucial for the activation of a "proliferation program" at 48 and 72 h. This program is observed in all three conditions that include BCR stimulation, and it is most prominent in DN2c. For example, the gene module itself is larger in the DN2c condition. Further, gene ontology analysis of genes overexpressed in DN2c compared to BCRc at 72 h reveals significant enrichment for biological processes related to DNA replication, chromosome segregation, and nuclear division (Supplementary Data 4). In conditions lacking BCR stimulation (CD40c, TLR7c, TLR9c), the proliferation program is not detected. Additionally, in non-BCR stimulated conditions, a module of genes coding for ribosomal proteins, strongly enriched for "cytoplasmic translation", decreases in expression over time. Notably, this downregulation does not occur upon BCR stimulation, indicating that BCR engagement might sustain protein synthesis for longer time periods (Supplementary Fig. 2).

To investigate B cell activation at the single-cell level, we performed single-cell RNA-seq. We found that B cells give rise to different cell states depending on the activation condition. By applying clustering and differential gene expression analysis, we identified IgD^+ $CD27^-$ "naïve-like" B cells, IgD^- $CD27^+$ "memory-like" B cells, *XBP1*-expressing B cells, and *MKI67*-expressing proliferating B cells, and how the frequency of these subsets varies depending on the experimental condition. After BCR stimulation in BCRc and DN2c, IgM/IgD protein expression is markedly reduced by 24 h, and by 72 h, a proliferating $MKI67^+$ subset emerges, with this subset being more abundant in DN2c. These results are concordant with the observations of the proliferation gene module identified in the bulk RNA-seq analysis. Furthermore, the majority of CD27-high cells at 72 h are from BCRc, whereas DN2c appears to suppress this cell state.

Next, we performed ATAC-seq to study the chromatin remodeling at 24 h post-stimulation. Our data suggest a gradient of activation strength across conditions: IL-4c, TLR7c, BCRc, and DN2c, with DN2c exhibiting the most wide-spread changes. Transcription factor motif analysis of activation-dependent open chromatin regions revealed strong enrichment for bZIP-containing TFs, such as BATF and ATF3, in all conditions except IL-4c, with enrichment increasing toward DN2c. On the other hand, binding sites of the transcriptional repressor BCL6 show the opposite trend, with no significant enrichment in DN2c. Notably, DN2c exhibits specific or enhanced motif enrichment for IRF family TFs (IRF1, IRF2, IRF3, IRF4, IRF8), BLIMP-1 (*PRDM1*), and T-bet *(TBX21)*. Most of these TFs are in risk loci for IMDs. For example, *IRF8* is a likely target gene for risk variants for SLE, UC, PBC, MS, SSc, and RA[33,34,46,47,49,55,56,68]. *PRDM1* is a likely target gene for variants associated with IBD, UC, CD, and psoriasis[47,54,55,69]. *TBX21* is in risk loci for autoimmune thyroid disease, asthma, childhood-onset asthma, and eczema[29,31,70]. These findings suggest that studying human B cells activated with the DN2c can be of high relevance to understand the role of IMD-associated TFs in B cell function and differentiation.

https://doi.org/10.1038/s41467-026-78656-w

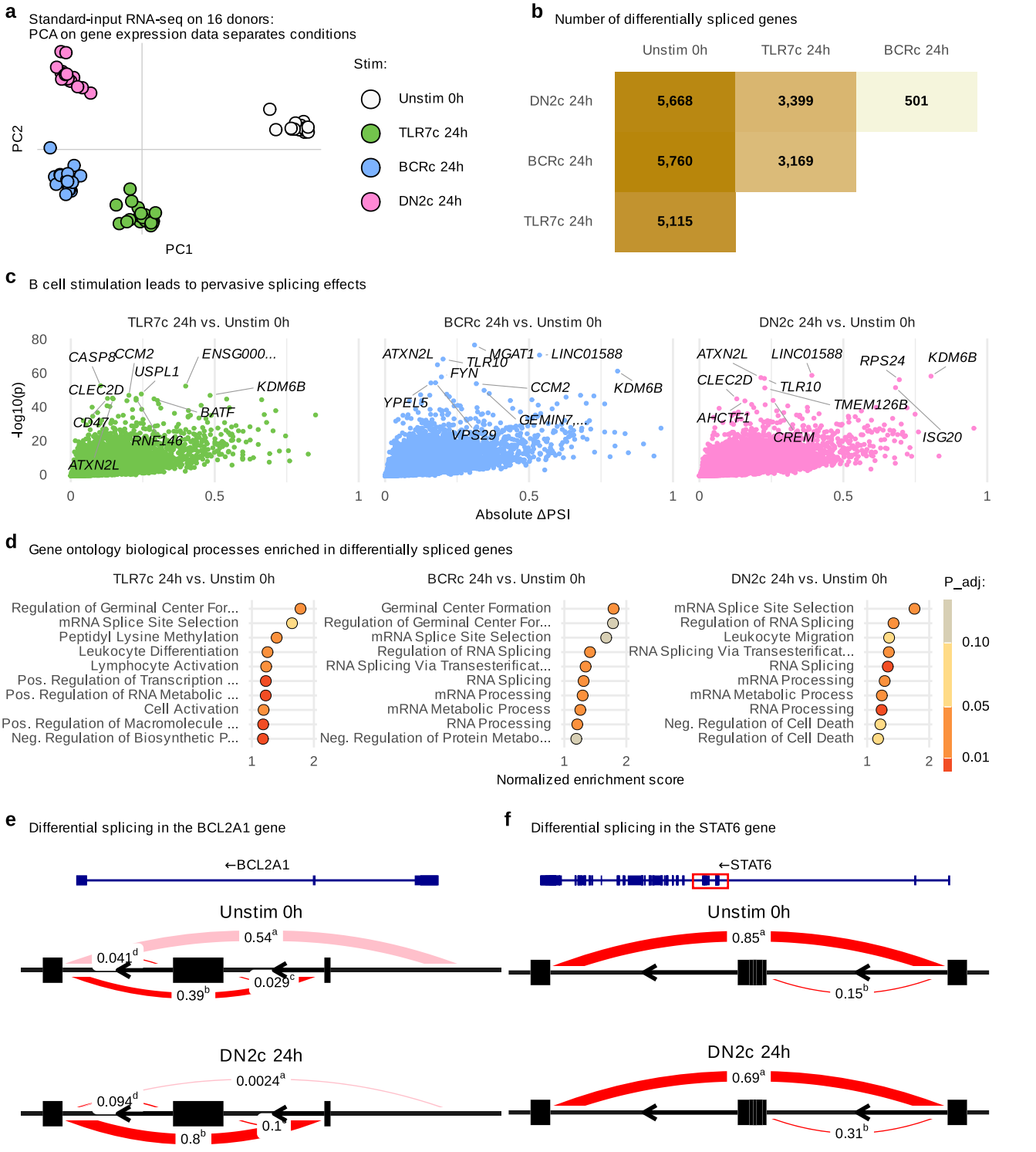


Fig. 6 | Pervasive alternative splicing changes during B cell activation. Standard-input bulk RNA-seq from 16 donors was used to quantify differential intron usage (LeafCutter). **a** Principal component analysis on gene expression data from 16 donors. **b** Number of differentially spliced genes (LeafCutter, 5% FDR) between every pair of conditions. **c** Splicing effect size, ΔPSI (percent spliced in) between each activation condition and Unstim 0 h, against −log10 P; each point is a gene, represented by its most significant intron cluster (per-cluster likelihood-ratio test, Dirichlet-multinomial model). **d** Gene Ontology biological processes enriched among differentially spliced genes in (**c**). Normalized enrichment scores and *P* values with Benjamini-Hochberg adjustment from pre-ranked gene set enrichment analysis (fgsea), with genes ranked by -log10 P of differential splicing (one-sided, scoreType = "pos"). **e** Differential splicing event at the BCL2A1 gene between DN2c and Unstim 0 h; numbers indicate the average proportion of junction usage and line widths are scaled to it. **f** As in (**e**), for the STAT6 gene.

By integrating our chromatin data with publicly available GWAS data, we found that activation-dependent open chromatin regions are enriched for immune-mediated disease heritability. Although this enrichment does not necessarily imply causality of risk variants in B cells, it suggests that studying risk variants in activated rather than resting B cells could provide valuable insights into IMD pathogenesis. As a proof of concept, we examined the SLE risk locus chr1q25.1[40]. By performing fine-mapping, we identified the intergenic SNP rs2205960 as the candidate causal variant in the locus. This SNP is located in a genomic region with accessible chromatin specifically in the DN2c condition. In a B cell line (LCL), we show that this open chromatin region is an enhancer that regulates *TNFSF4*. *TNFSF4* encodes OX40L, a tumor necrosis factor (TNF) superfamily member expressed on the surface of antigen-presenting cells that binds OX40 (*TNFRSF4*) on the surface of T cells, triggering NF-kB activation. The *TNFSF4* locus is associated with other immune-mediated diseases apart from SLE, such as RA, asthma, and eczema[27,29,40,46,52]. OX40L–CAR-Tregs have been developed to deplete OX40L-expressing cells and control alloreactivity and autoimmunity[71]. Our data shows that *TNFSF4* is lowly expressed in resting B cells, but it is strongly upregulated upon BCR stimulation (in BCRc and DN2c), thus suggesting that its function is more relevant after activation. It is important to note that it remains to be determined if the candidate causal variant itself regulates *TNFSF4* during B cell activation with DN2c and how this could affect B cell function. Public eQTL datasets indicate rs2205960 is marginally associated with *TNFSF4* expression in induced pluripotent stem cells and monocytes ($P > 0.0015$)[72,73], but these associations are far from the significance threshold required for calling an eQTL. *TNFSF4* has significant eQTL variants in T cells, thyroid, and blood, but none of the leading variants in these tissues are in linkage disequilibrium with rs2205960 ($r^2 < 0.08$)[74–76]. It is possible that rs2205960 is not the causal variant in the locus and/or that it affects *TNFSF4* in other non-B-cell cell states. We believe that even if rs2205960 is in a functional enhancer in LCLs (immortalized B cells) the variant itself is not associated with *TNFSF4* expression in LCLs because its regulatory effects may be very context-specific within B cell states—indeed, in primary B cells subjected to different stimuli this region is only open in DN2c-activated B cells. Manku et al. previously identified rs2205960 as a likely causal variant overlapping a binding site for NF-κBp65 (*RELA*)[63]. Notably, we observed that *RELA* is significantly upregulated 4 h post-DN2c activation, preceding the chromatin opening at rs2205960 (24 h) and the subsequent peak of *TNFSF4* expression (48–72 h). This temporal sequence suggests a model where activation-induced NF-κB signaling opens the rs2205960 enhancer to drive *TNFSF4* expression. Ultimately, this locus exemplifies the necessity of selecting appropriate activation states to resolve the regulatory mechanisms underlying IMD risk.

Lastly, we explored splicing changes during B cell activation. Splicing has gained attention in cancer research, leading to the development of therapeutic splicing modulators[77]. Moreover, recent studies have shown that a proportion of the IMD risk variants (17–25%) might exert their effects by altering RNA splicing[78,79]. For example, Mu et al. reported that among sixteen SLE GWAS loci that colocalized with different types of QTLs, ten colocalized exclusively with splicing QTLs[78]. We propose that investigating splicing across diverse B activation states may help elucidate uncharacterized IMD associations. Here we provide a comprehensive catalog of splicing changes occurring during B cell activation and in a stimulus-dependent manner. This resource can aid in the identification of splicing-mediated regulatory mechanisms and support future studies that analyze how these events may get altered in disease.

Despite the broad range of stimuli and high depth of molecular profiling in our study, a number of limitations should be considered. First, the study was performed using in vitro stimulation of B cells from healthy adult female donors. While this design facilitates controlled perturbation and avoids disease-related confounders, it may not fully capture the in vivo complexity of B cell activation in disease contexts or across diverse populations of different age groups, sex, and genetic ancestries. Second, although we incorporated a wide spectrum of stimuli relevant to key immunological pathways, additional microenvironmental cues that exist in tissue niches (such as particular cytokine combinations and interactions with other cell types) are not captured in this system. Furthermore, it is hard to know whether the concentration of our stimulants is lower or higher than that encountered by B cells in the human body. Third, due to the low abundance of B cells in peripheral blood and our aim to maximize the number of activation conditions profiled, we stimulated total B cell populations rather than isolating specific subsets. As a result, the observed transcriptional and epigenomic responses represent averaged signals across heterogeneous B cell populations, which may obscure subset-specific dynamics (such as for naïve, memory, or transitional B cells), including differences in cell viability upon stimulation. However, our qPCR validation for six disease-associated genes show that starting with naïve B cells we can expect similar expression level dynamics in response to DN2c-activation compared to our low input RNA-seq data starting from total B cells (Supplementary Fig. 14). Fourth, while we linked open chromatin regions to disease-associated variants, further experimental validation will be needed to confirm the causal variants and their precise gene targets.

Our work is complementary to previous studies of B cell multi-omics. For example, Guo et al. conducted bulk RNA-sequencing, proteomics and metabolomics analyses on B cells cultured for 24 h in control and activation conditions (10 in total), many of which are similar to those used in our study[15]. This overlap allows for direct comparisons while also enabling us to expand upon their findings using our additional time points, ATAC-seq and single-cell RNA-seq data. For instance, the authors found that *BCAT1* is highly upregulated following combined simulation of BCR and TLR9, leading to mTOR activation and promoting B cell growth. In our dataset, *BCAT1* exhibits stronger activation in the DN2c condition. Our multiple time points reveal that *BCAT1* is associated with module 1 (Fig. 2), a gene module characterized by early activation at 4 h.

In summary, our multi-omics approach provides a comprehensive profile of B cell activation in response to different stimuli. This publicly available and easy-to-access resource will facilitate future studies investigating B cell function and its implications in immune-mediated diseases.

Methods

Samples

We recruited 27 adult female donors aged 20–60 years from the Mass General Brigham Biobank (MGBB) through the Joint Biology Consortium Recruitment Core, and the Boston Children's Hospital Biorepository of Adult Healthy Controls. All donors were recruited as part of lupus research studies. Participants recruited through the MGBB Biobank received US$100 in gift cards as compensation. Participants recruited through the BCH Biorepository received US$10 in gift cards. All use of human material was approved by the Mass General Brigham Institutional Review Board (protocol 2019P003709) and the Boston Children's Hospital Institutional Review Board (protocols IRB-P00039644 and IRB-P00038916). All recruited volunteers provided written informed consent.

Reagents and Materials

A list of all antibodies, commercial reagents, kits, and their corresponding catalog numbers is provided in Supplementary Data 1.

B cell processing

Blood samples were collected from healthy donors using EDTA-coated tubes. Peripheral blood mononuclear cells (PBMC) were subsequently isolated by Ficoll-Paque (GE Healthcare) density gradient

centrifugation. B cells were further isolated from PBMCs using a Total B Cell Isolation Kit (Stemcell Technologies), and their purity was confirmed through flow cytometry (FACS) analysis, which had a median of 99.45% (Supplementary Fig. 1a, b). Cells were stimulated with the following conditions for the different experiments: (1) **IL-4c:** 20 ng/ml IL-4; (2) **CD40c:** 20 ng/ml IL-4 + 50 ng/ml sCD40L; (3) **TLR9c:** 20 ng/ml IL-4 + 2.5 µg/ml CpG; (4) **TLR7c:** 20 ng/ml IL-4 + 1 µM R848, (5) **BCRc:** 20 ng/ml IL-4 + 10 µg/ml anti-IgG/IgM, (6) **BCR/TRL7c:** 20 ng/ml IL-4 + 10 µg/ml anti-IgG/IgM + 1 µM R848; (7) **DN2c:** 10 ng/ml BAFF + 10 ng/ml IL-21 + 2.12 ng/ml IL-2 + 20 ng/ml IFN-γ + 1 µg/ml R848 + 5 µg/ml anti-IgG/IgM.

Flow cytometry

Samples were analyzed using a BD LSRFortessa X-20 Cell Analyzer. Samples were acquired on the BD LSRFortessa X-20 Cell Analyzer equipped with 355 nm, 405 nm, 488 nm, 561 nm, and 640 nm lasers using FACSDiva software. For flow cytometry staining, fluorochrome-conjugated antibodies targeting the markers indicated in the corresponding figures and throughout the manuscript were used according to standard extracellular and intracellular staining protocols. Compensation and gating analyses were performed using FlowJo v10. Antibody information, including target marker, fluorochrome, clone, dilution, vendor, and catalog number, is provided in Supplementary Data 1.

Low-input bulk RNA sequencing

At the designated time points, 1000 live B cells per sample were sorted directly into the lysis buffer (Supplementary Fig. 1j). RNA-seq for total B cells from six participants was performed at the Broad Institute using the SmartSeq2 library preparation and 75 bp paired-end reads, including two control conditions and six activation conditions (Fig. 1). We trimmed adapter sequences from the reads and performed QC with Trim Galore and FastQC[80,81]. Samples with less than 2 million sequenced reads were removed. We aligned reads to the Gencode v38 "primary assembly" reference transcriptome with Salmon (v1.5.1)[82].

We performed PCA in R (v4.1.2) with the top 2000 most variable genes after VST normalization with DESeq2 (v1.34)[83]. In order to test for differential gene expression across conditions and time points we used edgeR v3.36[84], applying default filtering of minimum expression levels. We used a significance threshold of $p < 0.05$ on the Benjamini & Hochberg adjusted p-values. As shown in Supplementary Fig. 13, gene expression changes between unstimulated B cells at 0 h and activation conditions replicate well with the ones found for standard-input RNA-seq using higher complexity libraries and a larger number of donors (Pearson correlation ≥ 0.85). We used WGCNA v1.72 to leverage correlation of gene expression to identify gene modules[23]. The input gene set was restricted to genes exhibiting significant temporal changes, as modeled by a cubic spline approach in edgeR. Raw counts were imported to DESeq2 using tximport. Genes with low expression were filtered (requiring ≥ 4 normalized counts in at least 3 samples) and VST transformation was applied to the count matrix. We constructed a signed co-expression network. A soft-thresholding power was selected for each condition using the pickSoftThreshold function. The adjacency matrix was calculated using Pearson correlation and transformed into a Topological Overlap Matrix (TOM). Gene modules were identified with a detect cut height of 0.99 and minimum module size of 60 genes, keeping genes with minimum module connectivity of 0.8. Highly correlated modules were merged using a cut height of 0.2. To understand the biological processes enriched in each module, we performed gene ontology (GO) analysis with the clusterProfiler package (v4.2.2)[85], after selecting genes with kME ≥ 0.9 in any module and the top 500 genes in each module by kME value.

Disease-associated gene definition

At each GWAS risk locus passing the genome-wide significance threshold, we selected the prioritized gene according to the Open Targets Genetics Locus-to-Gene (L2G) algorithm[26]. This machine learning framework integrates fine-mapping credible sets with functional genomics data (e.g., chromatin interaction, canonical QTLs) to prioritize causal targets. This algorithm has limitations and will not always identify the true causal gene. It also does not provide any information of whether the risk haplotype affects gene regulation specifically in activated B cells, for which expression quantitative trait loci studies are not available.

Standard-input bulk RNA sequencing

We performed RNA-seq on samples from 16 participants, at resting state and three stimulation conditions at 24 h (TLR7c, BCRc, DN2c), targeting 40 million reads. Following culture, 500,000 cells per sample were resuspended in the lysis buffer, and RNA was extracted using the AllPrep DNA/RNA Mini or Micro Kit (Qiagen), depending on the cell number. We performed QC with Trim Galore and FastQC[80,81]. We aligned reads to the reference genome GRCh38 with STAR (v2.7.9a) in two-pass mode[86], accepting up to 4% of mismatches. In order to identify possible sample swaps we used the MBV method to evaluate the concordance of the RNA-sequencing reads with the genotypes from MGB Biobank[87].

For splicing analyses, we mapped the reads as described above, additionally adding the flag "--outSAMstrandField intronMotif" for STAR. We used regtools v0.5.1 to extract junctions[88], and performed intron clustering and differential splicing analysis between pairs of conditions with LeafCutter v0.2.9[65]. PCA via the leafviz visualization tool revealed an outlier that had a low proportion of mapped reads to the genome which was removed. Differential splicing analysis was performed with default parameters. We performed enrichment analysis on differentially spliced genes with the fgsea R package v1.20.0[89], ranking genes by their most extreme p value across intron clusters, and testing enrichment in Gene Ontology biological process gene sets (MSigDB C5, v2022.1.Hs). Fgsea was run in one-sided mode (scoreType = "pos"). For each gene set, fgsea computes an enrichment score that reflects how strongly the set's genes are concentrated toward the top of the ranked list. Enrichment scores were normalized for gene-set size. P values were obtained by comparing each observed enrichment score to a null distribution built from randomly sampled gene sets of the same size, and adjusted for multiple testing using the Benjamini-Hochberg method. We used the leafviz R package for visualization of splicing events (https://github.com/jackhump/leafviz).

Single-cell RNA sequencing

B cells from five healthy donors were activated using a reverse time course to accommodate simultaneous processing and staining. Cells were cultured and activated for 72 h, 24 h, or left unstimulated (0 h). At the 0 h time point, all cells were collected and stained with Hash Tag Oligonucleotides (HTO) and Antibody-Derived Tags (ADT) using TotalSeq-C antibodies (BioLegend). The samples were then pooled, and live (7-AAD$^-$) CD19$^+$ cells were sorted for sequencing (Supplementary Fig. 1j). Droplet encapsulation and sequencing were performed at the Center for Cellular Profiling, Brigham and Women's Hospital, using the droplet-based 10X Genomics Chromium scRNA-seq platform. Sequencing was performed in two batches. Raw sequencing reads were processed into count matrices using Cell Ranger v6.1.1.

We used Seurat v4 for quality control and data analysis[90–93]. We removed droplets with less than 500 genes and more than 10% of mitochondrial reads. We applied the "LogNormalize" method to gene counts, and centered log ratio (CLR) transformation to ADT and HTO counts. We performed HTO-based demultiplexing with demuxmix v1.6.0 using the regression-based mixture model incorporating the number of detected genes per cell as a covariate[94]. We performed

genotype-based demultiplexing (using MGB Biobank genome-wide SNP genotypes) with demuxlet including biallelic SNPs with a Phred quality score ≥ 20 and imputation $R^2 \geq 0.9$[95]. We removed cells marked as doublets by HTO-based demultiplexing and/or genotype-based demultiplexing. For PCA, we used centered and scaled data after regressing for total UMIs and percentage of mitochondrial reads and we chose the top 2000 variable genes. We integrated data from different libraries and donors with Harmony v1.2.0 using the defaults of theta = 2 and all PCA dimensions[96]. We performed clustering with the Louvain clustering algorithm using the 30 harmonized PCs. To annotate clusters, we identified the top markers per cluster with the *FindAllMarkers* function in Seurat using a logFC threshold of 0.5 and minimum percentage of expressing cells of 33%. We quantified the proportion of B cell subsets based on gene or protein markers of interest by "gating" cells using IgD, CD27, CD11c, *MKI67* and *XBP1*. Thresholds for marker positivity were determined by evaluating expression distributions across clusters (Supplementary Fig. 6).

To evaluate condition- and cluster-specific changes in gene expression and surface protein abundance while accounting for inter-individual variability, we performed pseudobulk differential analysis. Single-cell RNA and antibody-derived tag (ADT) counts were aggregated to the donor level within each experimental condition or annotated B cell cluster. Differential testing was conducted using the generalized linear model (GLM) framework implemented in the edgeR R package (version 3.36.0)[84]. Count matrices were normalized using the trimmed mean of M-values (TMM) method. Statistical significance was assessed using quasi-likelihood (QL) F-tests, and nominal p-values were adjusted for multiple comparisons using the Benjamini-Hochberg false discovery rate (FDR) method (Supplementary Data 5).

Bulk ATAC-seq

We isolated 100,000 nuclei from B cells per sample and delivered them to the Center for Functional Cancer Epigenetics, Dana-Farber Cancer Institute, for ATAC-seq library preparation and sequencing. The conditions included here were: unstimulated cells straight from PBMCs (0 h), unstimulated cells in media for 24 h, cells treated with IL-4c for 24 h, cells treated with TLR7c for 24 h, cells treated with BCRc for 24 h, and cells treated with DN2c for 24 h. Sequencing data was processed with the Nextflow ATAC-seq pipeline v2.0[97,98]. Briefly, this pipeline first uses FastQC to verify sequencing quality, adapter contamination, and GC content. Then, it uses Trim Galore to trim adapters and remove low-quality reads. Reads are aligned with BWA-MEM to the GRCh38 reference genome (Gencode v42 annotation), PCR duplicates are removed with Picard, and only high-quality alignments are retained. Reads aligning to black-listed regions (ENCODE hg38 blacklist; "hg38-blacklist.v3.bed") are filtered out. Peaks are called with MACS2 in individual samples, and then overlapping peaks across samples are collapsed to form the "consensus" peaks. To assess library quality, the Fraction of Reads in Peaks (FRiP) is calculated for each sample. Additionally, peaks are annotated to genomic features (e.g., promoters, exons, introns, intergenic regions) using HOMER (Supplementary Fig. 15). Quantification of open chromatin in each sample is performed counting reads overlapping the consensus peaks. Counts were then normalized with the rlog method for PCA and with the median of ratios method for differential accessibility analysis with DESeq2[83]. In addition to the samples described above coming from 3 individual donors and shown in our main analyses, ATAC-seq library preparation and sequencing was also performed for nuclei pooled from 3 independent donors (from cells coming from the same activation conditions). Supplementary Fig. 15 shows scatter plots comparing the normalized counts of each of the three individual donors versus the pool of the three independent donors, for each of the activation or unstimulated conditions. Given replication was high for all samples (Pearson correlation coefficient ≥ 0.97) no samples were removed from our analyses. Differential accessibility analysis was performed with DESeq2 v1.34 accounting for donor effects[83]. Differentially accessible peaks at 1% FDR in an activation condition at 24 h with respect to the unstimulated condition at 24 h were selected for downstream analyses. We used HOMER v4.11 for motif enrichment analysis using all the consensus peaks as the background set, a 200 bp window centered on the peak, masking repeat-masked regions, and restricting to vertebrate transcription factors only[59]. For heritability enrichment analysis, we used the linkage disequilibrium score regression (LDSC) method, with pre-processed GWAS summary statistics, linkage disequilibrium values, and SNP lists made available by the authors[60,99]. Briefly, binary annotations were created by identifying 1000 Genomes EUR SNPs located within accessible chromatin regions, extending peaks by a 500 bp window. LD scores were calculated within a 1 centimorgan (cM) window. The analysis was restricted to the intersection of the reference panel and the HapMap3 SNP list, excluding the MHC region. To quantify the contribution of accessible chromatin regions to trait heritability, we first estimate the total SNP heritability and annotation-specific standard deviations with S-LDSC with the baseline model. To evaluate specific enrichment of our B cell activation-specific peak sets, we performed LDSC-Seg analysis. To enable comparison across traits and annotations, we calculated the standardized regression coefficient Tau*[100].

CRISPR validation experiments

GM12878 lymphoblastoid cell line (LCL), an ENCODE Tier 1 EBV-transformed human B cell line, was obtained from the Coriell Institute. Cells were cultured in RPMI 1640 medium supplemented with 10% fetal bovine serum, 1% penicillin-streptomycin, and 1% L-glutamine, at a density of 1×10^5 cells/mL at 37 °C in a humidified incubator with 5% CO_2.

To generate GM12878 cell lines stably expressing the CRISPR interference (CRISPRi) system, two separate cell lines were established using either the lentiCRISPRi(v1)-Blast (Addgene #170067, expressing dCas9-KRAB) or lentiCRISPRi(v2)-Blast transfer plasmid. Lentiviral particles for each construct were produced independently by transfecting HEK293T cells[101]. Briefly, HEK293T cells were seeded at 3×10^5 cells in 6-well plates in DMEM supplemented with 10% FCS. At ~70% confluency (~24 h post-seeding), cells were co-transfected using TransIT-LT1 transfection reagent (Mirus Bio) and Opti-MEM (Gibco) with a packaging mix containing 150 ng pMD2.G (VSV-G; Addgene #12259), 400 ng psPAX2 (Addgene #12260), and 500 ng of the respective transfer plasmid. Culture medium was replaced with fresh complete RPMI supplemented with 10% FCS 24 h post-transfection. Viral supernatants were harvested at 48 h and 72 h post-transfection and passed through a 0.45 μm syringe filter. For the initial transduction (48 h harvest), filtered viral supernatant was applied to target GM12878 cells in the presence of 4 μg/mL polybrene, followed by spinoculation at 300 × g for 2 h at 32 °C. At 72 h post-transfection, a second round of filtered viral supernatant was added directly to the cells. Transduced GM12878 cells were selected 24 h after the second viral addition using puromycin (3 μg/mL) for 10 days. Two single guide RNAs (sgRNAs) targeting the regulatory region containing the rs2205960 variant were designed using the CRISPRpick tool (Broad Institute)[102], selecting guides with the highest predicted on-target efficacy and minimal off-target potential. All sgRNA sequences, including a non-targeting scrambled control, are listed in Supplementary Data 1.

Nucleofection was performed using the SE Cell Line 4D-Nucleofector X Kit (Lonza) following the manufacturer's instructions. Briefly, 300,000–500,000 GM12878 cells were resuspended in SE solution and nucleofected with 6 μL of 30 μM sgRNA using program DN-100. Immediately after nucleofection, cells were transferred into pre-warmed R10 medium and cultured for 72 h to allow gene silencing.

Total RNA was extracted using a micro-RNA isolation kit (Qiagen), followed by cDNA synthesis using AffinityScript cDNA Synthesis Kit (Agilent). Quantitative PCR (qPCR) was performed using qPCR Brilliant

SYBR Master Mix (Agilent) on a QuantStudio 3 Real-Time PCR System (Applied Biosystems). *TNFSF4* transcript levels were measured using the primers listed in Supplementary Data 1. Data are presented as relative gene expression normalized to the housekeeping gene *GAPDH*[103], using the ΔΔCt method. For each repressor system, each sgRNA and the scrambled control was assayed in two independent silencing experiments. The two sgRNAs, as independent perturbations of the same regulatory element, were pooled and compared to the scrambled control by two-sample t-test.

GWAS fine-mapping

We performed Bayesian statistical fine-mapping of the Langefeld et al. SLE GWAS data[40]. We selected bi-allelic SNPs that were genotyped in the immunochip in a window of 500 kb at each genome-wide significant locus, and applied SuSiE with in-sample LD to estimate variant posterior probabilities and credible sets with at least 90% coverage[104].

Statistics and reproducibility

This study generated a multi-omics resource of primary human B cell activation. Peripheral blood B cells were isolated from 27 healthy donors and stimulated in vitro under multiple activation conditions across a time course, then profiled by low-input RNA-seq ($N = 6$), standard-input RNA-seq ($N = 16$), single-cell RNA-seq with surface-protein quantification (CITE-seq; $N = 5$), and bulk ATAC-seq ($N = 3$). CRISPRi validation was performed in the GM12878 lymphoblastoid cell line.

No statistical method was used to determine sample sizes, which were determined by donor and sample availability and are consistent with those typical for genomic profiling studies of this kind.

Data were excluded only according to established quality-control criteria: low-input RNA-seq samples with fewer than 2 million reads were removed; one standard-input RNA-seq sample was removed as an outlier with a low proportion of mapped reads; and in the single-cell data, droplets with fewer than 500 genes or more than 10% mitochondrial reads, and doublets, were removed.

No formal randomization procedure was used. The assignment of donors to assays was based on cell availability. Sample collection, experiments, and computational analyses were performed by different investigators, and both the experimental and analytical work were carried out without access to donor identity.

Reproducibility was supported by multiple donors per assay and by concordance across independent measurements. All antibodies, reagents, and kits used in this study, with their catalog numbers, are listed in Supplementary Data 1. Full details of statistical tests and software are provided, and the analysis code, processed data, and raw sequencing data are available as described in the relevant Methods sections.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

The raw sequencing data generated in this study are available in dbGaP under accession code phs004221.v1.p1 [https://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study_id=phs004221.v1.p1]. The data include raw RNA-seq, CITE-seq, and ATAC-seq sequence reads (FASTQ/BAM) from 27 donors. Because these are individual-level human genomic sequencing data that represent potentially identifying information, and in accordance with the study protocols and the written informed consent provided by participants, they are under controlled access. Access is overseen by the NIAMS Data Access Committee and can be requested through the dbGaP Authorized Access portal, requiring endorsement by an institutional Signing Official and a signed Data Use Certification; use is restricted to health/medical/biomedical research (HMB). Requests are typically reviewed within a few weeks. Approved access is granted for a period of one (1) year, with the option to renew access. All processed data, including count matrices and differential expression/accessibility summary statistics, have been deposited in Zenodo [https://zenodo.org/uploads/15595470]. We made processed data publicly available via an interactive website that can be easily queried for visualization of gene expression, single cell profiles, chromatin accessibility, and splicing events [https://mgalab.shinyapps.io/bcellactivation/]. Publicly available data used in this study include: Reference genome and transcript annotations, GENCODE v38 [https://www.gencodegenes.org/human/release_38.html], and GENCODE v42 [https://www.gencodegenes.org/human/release_42.html]; ENCODE blacklist regions (hg38) [https://www.encodeproject.org/files/ENCFF356LFX/]; Open Targets Platform (Locus-to-Gene predictions) [https://platform.opentargets.org]; "disease-state" and "disease-activity" B cell signatures (Supplementary Table S2 of Nakano et al.)[53]; LDSC reference files including pre-processed GWAS summary statistics, LD scores, and SNP lists, and baseline model (Zenodo) [https://zenodo.org/records/10515792]; GWAS summary statistics from Langefeld et al. 2017, GWAS Catalog [https://www.ebi.ac.uk/gwas/studies/GCST007400]; Enhancer-to-Gene (E2G) predictions, E2G Browser [https://e2g.stanford.edu/]; MSigDB C5 (GO) gene sets, v2022.1.Hs [https://www.gsea-msigdb.org/gsea/msigdb]. All data are included in the Supplementary Information or available from the authors, as are unique reagents used in this Article. The raw numbers for charts and graphs are available in the Source Data file whenever possible. Source data are provided with this paper.

Code availability

The code used for analyses is available on GitHub [https://github.com/gutierrez-arcelus-lab/bcellactivation] and archived on Zenodo at [https://doi.org/10.5281/zenodo.22664586][105]. The source code for the Shiny app is available on GitHub [https://github.com/gutierrez-arcelus-lab/shinybcells]. Most software used in this project was installed and configured by Biogrids[106].

References

1. Rawlings, D. J., Metzler, G., Wray-Dutra, M. & Jackson, S. W. Altered B cell signalling in autoimmunity. *Nat. Rev. Immunol.* **17**, 421–436 (2017).
2. Satitsuksanoa, P. et al. B cells: the many facets of B cells in allergic diseases. *J. Allergy Clin. Immunol.* **152**, 567–581 (2023).
3. Laumont, C. M., Banville, A. C., Gilardi, M., Hollern, D. P. & Nelson, B. H. Tumour-infiltrating B cells: immunological mechanisms, clinical impact and therapeutic opportunities. *Nat. Rev. Cancer* **22**, 414–430 (2022).
4. Jenks, S. A. et al. Distinct effector B cells induced by unregulated toll-like receptor 7 contribute to pathogenic responses in systemic lupus erythematosus. *Immunity* **49**, 725–739.e6 (2018).
5. Vinuesa, C. G., Shen, N. & Ware, T. Genetics of SLE: mechanistic insights from monogenic disease and disease-associated variants. *Nat. Rev. Nephrol.* **19**, 558–572 (2023).
6. Elsner, R. A. & Shlomchik, M. J. Germinal center and extrafollicular B cell responses in vaccination, immunity, and autoimmunity. *Immunity* **53**, 1136–1150 (2020).
7. Wing, E. et al. Double-negative-2 B cells are the major synovial plasma cell precursor in rheumatoid arthritis. *Front. Immunol.* **14**, 1241474 (2023).
8. Castleman, M. J. et al. Autoantibodies elicited with SARS-CoV-2 infection are linked to alterations in double negative B cells. *Front. Immunol.* **13**, 988125 (2022).
9. Mouat, I. C., Goldberg, E. & Horwitz, M. S. Age-associated B cells in autoimmune diseases. *Cell. Mol. Life Sci.* **79**, 402 (2022).
10. Yang, Y. et al. Pan-cancer single-cell dissection reveals phenotypically distinct B cell subtypes. *Cell* **187**, 4790–4811.e22 (2024).

11. Rubtsov, A. V. et al. Toll-like receptor 7 (TLR7)–driven accumulation of a novel CD11c+ B-cell population is important for the development of autoimmunity. *Blood* **118**, 1305–1315 (2011).
12. Nickerson, K. M. et al. Age-associated B cells are heterogeneous and dynamic drivers of autoimmunity in mice. *J. Exp. Med.* **220**, e20221346 (2023).
13. Zumaquero, E. et al. IFNγ induces epigenetic programming of human T-bethi B cells and promotes TLR7/8 and IL-21 induced differentiation. *eLife* **8**, e41641 (2019).
14. Calderon, D. et al. Landscape of stimulation-responsive chromatin across diverse human immune cells. *Nat. Genet.* **51**, 1494–1505 (2019).
15. Guo, R. et al. Multiomic analysis reveals a key BCAT1 role in mTOR activation by B cell receptor and TLR9. *J. Clin. Investig.* **135**, e186258 (2025).
16. Farh, K. K.-H. et al. Genetic and epigenetic fine mapping of causal autoimmune disease variants. *Nature* **518**, 337–343 (2015).
17. Soskic, B. et al. Chromatin activity at GWAS loci identifies T cell states driving complex immune diseases. *Nat. Genet.* **51**, 1486–1493 (2019).
18. Umans, B. D., Battle, A. & Gilad, Y. Where are the disease-associated eQTLs? *Trends Genet.* **37**, 109–124 (2021).
19. Mostafavi, H., Spence, J. P., Naqvi, S. & Pritchard, J. K. Systematic differences in discovery of genetic effects on gene expression and complex traits. *Nat. Genet.* **55**, 1866–1875 (2023).
20. Possamaï, D., Pagé, G., Panès, R., Gagnon, É & Lapointe, R. CD40L-stimulated B lymphocytes are polarized toward APC functions after exposure to IL-4 and IL-21. *J. Immunol.* **207**, 77–89 (2021).
21. Malaspina, A. et al. CpG oligonucleotides enhance proliferative and effector responses of B cells in HIV-infected individuals. *J. Immunol.* **181**, 1199–1206 (2008).
22. Van Belle, K. et al. Comparative in vitro immune stimulation analysis of primary human B cells and B cell lines. *J. Immunol. Res.* **2016**, 1–9 (2016).
23. Langfelder, P. & Horvath, S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinform.* **9**, 559 (2008).
24. Mamidi, M. K. et al. FCRL1 immunoregulation in B cell development and malignancy. *Front. Immunol.* **14**, 1251127 (2023).
25. Langfelder, P., Luo, R., Oldham, M. C. & Horvath, S. Is my network module preserved and reproducible? *PLoS Comput. Biol.* **7**, e1001057 (2011).
26. Mountjoy, E. et al. An open approach to systematically prioritize causal variants and genes at all published human GWAS trait-associated loci. *Nat. Genet.* **53**, 1527–1533 (2021).
27. Ishigaki, K. et al. Multi-ancestry genome-wide association analyses identify novel genetic mechanisms in rheumatoid arthritis. *Nat. Genet.* **54**, 1640–1651 (2022).
28. Ferreira, M. A. et al. Shared genetic origin of asthma, hay fever and eczema elucidates allergic disease biology. *Nat. Genet.* **49**, 1752–1757 (2017).
29. Ferreira, M. A. R. et al. Genetic architectures of childhood- and adult-onset asthma are partly distinct. *Am. J. Hum. Genet.* **104**, 665–684 (2019).
30. López-Isac, E. et al. Combined genetic analysis of juvenile idiopathic arthritis clinical subtypes identifies novel risk loci, target genes and key regulatory mechanisms. *Ann. Rheum. Dis.* **80**, 321–328 (2021).
31. Saevarsdottir, S. et al. FLT3 stop mutation increases FLT3 ligand level and risk of autoimmune thyroid disease. *Nature* **584**, 619–623 (2020).
32. Jostins, L. et al. Host–microbe interactions have shaped the genetic architecture of inflammatory bowel disease. *Nature* **491**, 119–124 (2012).
33. International Multiple Sclerosis Genetics Consortium et al. Multiple sclerosis genomic map implicates peripheral immune cells and microglia in susceptibility. *Science* **365**, eaav7188 (2019).
34. Okada, Y. et al. Genetics of rheumatoid arthritis contributes to biology and drug discovery. *Nature* **506**, 376–381 (2014).
35. Ji, S.-G. et al. Genome-wide association study of primary sclerosing cholangitis identifies new risk loci and quantifies the genetic relationship with inflammatory bowel disease. *Nat. Genet.* **49**, 269–273 (2017).
36. Onengut-Gumuscu, S. et al. Fine mapping of type 1 diabetes susceptibility loci and evidence for colocalization of causal variants with lymphoid gene enhancers. *Nat. Genet.* **47**, 381–386 (2015).
37. Wang, Q. et al. Identification of a regulatory pathway governing TRAF1 via an arthritis-associated non-coding variant. *Cell Genom.* **3**, 100420 (2023).
38. the EArly Genetics and Lifecourse Epidemiology (EAGLE) Eczema Consortium. Multi-ancestry genome-wide association study of 21,000 cases and 95,000 controls identifies new risk loci for atopic dermatitis. *Nat. Genet.* **47**, 1449–1456 (2015).
39. Dubois, P. C. A. et al. Multiple common variants for celiac disease influencing immune gene expression. *Nat. Genet.* **42**, 295–302 (2010).
40. Langefeld, C. D. et al. Transancestral mapping and genetic load in systemic lupus erythematosus. *Nat. Commun.* **8**, 16021 (2017).
41. Mahajan, A. et al. Fine-mapping type 2 diabetes loci to single-variant resolution using high-density imputation and islet-specific epigenome maps. *Nat. Genet.* **50**, 1505–1513 (2018).
42. Yanagihara, Y. et al. Functional significance of IL-4 receptor on B cells in IL-4-induced human IgE production. *J. Allergy Clin. Immunol.* **96**, 1145–1151 (1995).
43. Steinke, J. W. Anti–interleukin-4 therapy. *Immunol. Allergy Clin. North Am.* **24**, 599–614 (2004).
44. Taylor, K. E. et al. Genome-wide association analysis reveals genetic heterogeneity of Sjögren's syndrome according to ancestry. *Arthritis Rheumatol.* **69**, 1294–1305 (2017).
45. Márquez, A. et al. Meta-analysis of Immunochip data of four autoimmune diseases reveals novel single-disease and cross-phenotype associations. *Genome Med.* **10**, 97 (2018).
46. Bentham, J. 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).
47. De Lange, K. M. et al. Genome-wide association study implicates immune activation of multiple integrin genes in inflammatory bowel disease. *Nat. Genet.* **49**, 256–261 (2017).
48. Trynka, G. et al. Dense genotyping identifies and localizes multiple common and rare variant association signals in celiac disease. *Nat. Genet.* **43**, 1193–1201 (2011).
49. López-Isac, E. et al. GWAS for systemic sclerosis identifies multiple risk loci and highlights fibrotic and vasculopathy pathways. *Nat. Commun.* **10**, 4955 (2019).
50. Wang, Q. et al. High-throughput identification of functional regulatory SNPs in systemic lupus erythematosus. *Nat. Commun.* **15**, 6804 (2024).
51. Tsoi, L. C. et al. Large scale meta-analysis characterizes genetic architecture for common psoriasis associated variants. *Nat. Commun.* **8**, 15382 (2017).
52. Pasanen, A. et al. Identifying atopic dermatitis risk loci in 1,094,060 individuals with subanalysis of disease severity and onset. *J. Investig. Dermatol.* **144**, 2417–2425 (2024).
53. Nakano, M. et al. Distinct transcriptome architectures underlying lupus establishment and exacerbation. *Cell* **185**, 3375–3389.e21 (2022).

54. Liu, J. Z. et al. Association analyses identify 38 susceptibility loci for inflammatory bowel disease and highlight shared genetic risk across populations. *Nat. Genet.* **47**, 979–986 (2015).
55. Liu, J. Z. et al. Dense fine-mapping study identifies new susceptibility loci for primary biliary cirrhosis. *Nat. Genet.* **44**, 1137–1141 (2012).
56. Gorlova, O. et al. Identification of novel genetic markers associated with clinical phenotypes of systemic sclerosis through a genome-wide association strategy. *PLoS Genet.* **7**, e1002178 (2011).
57. Chiou, J. et al. Interpreting type 1 diabetes risk with genetics and single-cell epigenomics. *Nature* **594**, 398–402 (2021).
58. Wang, Y.-F. et al. Identification of 38 novel loci for systemic lupus erythematosus and genetic heterogeneity between ancestral groups. *Nat. Commun.* **12**, 772 (2021).
59. Heinz, S. et al. Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol. Cell* **38**, 576–589 (2010).
60. Finucane, H. K. et al. Heritability enrichment of specifically expressed genes identifies disease-relevant tissues and cell types. *Nat. Genet.* **50**, 621–629 (2018).
61. Djeddi, S. et al. Rhinovirus infection of airway epithelial cells uncovers the non-ciliated subset as a likely driver of genetic risk to childhood-onset asthma. *Cell Genom.* **4**, 100636 (2024).
62. Woodruff, M. C. et al. Extrafollicular B cell responses correlate with neutralizing antibodies and morbidity in COVID-19. *Nat. Immunol.* **21**, 1506–1516 (2020).
63. Manku, H. et al. Trans-ancestral studies fine map the SLE-susceptibility locus TNFSF4. *PLoS Genet.* **9**, e1003554 (2013).
64. Gschwind, A. R. et al. An encyclopedia of human enhancer–gene regulatory interactions. *Nature* https://doi.org/10.1038/s41586-026-10781-4. (2026).
65. Li, Y. I. et al. Annotation-free quantification of RNA splicing using LeafCutter. *Nat. Genet.* **50**, 151–158 (2018).
66. Martinez, N. M. & Lynch, K. W. Control of alternative splicing in immune responses: many regulators, many predictions, much still to learn. *Immunol. Rev.* **253**, 216–236 (2013).
67. Tian, C. et al. Single-cell RNA sequencing of peripheral blood links cell-type-specific regulation of splicing to autoimmune and inflammatory diseases. *Nat. Genet.* **56**, 2739–2752 (2024).
68. Cordell, H. J. et al. An international genome-wide meta-analysis of primary biliary cholangitis: Novel risk loci and candidate drugs. *J. Hepatol.* **75**, 572–581 (2021).
69. Barrett, J. C. et al. Genome-wide association defines more than 30 distinct susceptibility loci for Crohn's disease. *Nat. Genet.* **40**, 955–962 (2008).
70. Grosche, S. et al. Rare variant analysis in eczema identifies exonic variants in DUSP1, NOTCH4 and SLC9A4. *Nat. Commun.* **12**, 6618 (2021).
71. Rui, X. et al. Human OX40L-CAR-Tregs target activated antigen-presenting cells and control T cell alloreactivity. *Sci. Transl. Med.* **16**, eadj9331 (2024).
72. Kilpinen, H. et al. Common genetic variation drives molecular heterogeneity in human iPSCs. *Nature* **546**, 370–375 (2017).
73. Schmiedel, B. J. et al. Impact of genetic polymorphisms on human immune cell gene expression. *Cell* **175**, 1701–1715.e16 (2018).
74. GTEx Consortium The GTEx Consortium atlas of genetic regulatory effects across human tissues. *Science* **369**, 1318–1330 (2020).
75. Lepik, K. et al. C-reactive protein upregulates the whole blood expression of CD59 - an integrative analysis. *PLOS Comput. Biol.* **13**, e1005766 (2017).
76. Chen, L. et al. Genetic drivers of epigenetic and transcriptional variation in human immune cells. *Cell* **167**, 1398–1414.e24 (2016).
77. Rogalska, M. E., Vivori, C. & Valcárcel, J. Regulation of pre-mRNA splicing: roles in physiology and disease, and therapeutic prospects. *Nat. Rev. Genet.* **24**, 251–269 (2023).
78. Mu, Z. et al. The impact of cell type and context-dependent regulatory variants on human immune traits. *Genome Biol.* **22**, 122 (2021).
79. El Garwany, O. et al. Splicing QTL mapping in stimulated macrophages associates low-usage splice junctions with immune-mediated disease risk. *Nat. Commun.* **16**, 7205 (2025).
80. Krueger, F. et al. FelixKrueger/TrimGalore: v0.6.10 - add default decompression path. Preprint at https://doi.org/10.5281/ZENODO.7598955 (2023).
81. Andrews, S. A quality control tool for high throughput sequence data. Babraham Bioinformatics.
82. Patro, R., Duggal, G., Love, M. I., Irizarry, R. A. & Kingsford, C. Salmon provides fast and bias-aware quantification of transcript expression. *Nat. Methods* **14**, 417–419 (2017).
83. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).
84. Robinson, M. D., McCarthy, D. J. & Smyth, G. K. edgeR: a bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**, 139–140 (2010).
85. Wu, T. et al. clusterProfiler 4.0: a universal enrichment tool for interpreting omics data. *Innov. Camb. Mass* **2**, 100141 (2021).
86. Dobin, A. et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).
87. Fort, A. et al. MBV: a method to solve sample mislabeling and detect technical bias in large combined genotype and sequencing assay datasets. *Bioinformatics* **33**, 1895–1897 (2017).
88. Cotto, K. C. et al. Integrated analysis of genomic and transcriptomic data for the discovery of splice-associated variants in cancer. *Nat. Commun.* **14**, 1589 (2023).
89. Korotkevich, G. et al. Fast gene set enrichment analysis. Preprint at https://doi.org/10.1101/060012 (2016).
90. Hao, Y. et al. Integrated analysis of multimodal single-cell data. *Cell* **184**, 3573–3587.e29 (2021).
91. Stuart, T. et al. Comprehensive integration of single-cell data. *Cell* **177**, 1888–1902.e21 (2019).
92. Butler, A., Hoffman, P., Smibert, P., Papalexi, E. & Satija, R. Integrating single-cell transcriptomic data across different conditions, technologies, and species. *Nat. Biotechnol.* **36**, 411–420 (2018).
93. Satija, R., Farrell, J. A., Gennert, D., Schier, A. F. & Regev, A. Spatial reconstruction of single-cell gene expression data. *Nat. Biotechnol.* **33**, 495–502 (2015).
94. Klein, H.-U. demuxmix: demultiplexing oligonucleotide-barcoded single-cell RNA sequencing data with regression mixture models. Preprint at https://doi.org/10.1101/2023.01.27.525961 (2023).
95. Kang, H. M. et al. Multiplexed droplet single-cell RNA-sequencing using natural genetic variation. *Nat. Biotechnol.* **36**, 89–94 (2018).
96. Korsunsky, I. et al. Fast, sensitive and accurate integration of single-cell data with Harmony. *Nat. Methods* **16**, 1289–1296 (2019).
97. Ewels, P. A. et al. The nf-core framework for community-curated bioinformatics pipelines. *Nat. Biotechnol.* **38**, 276–278 (2020).
98. Harshil, Patel et al. nf-core/atacseq: [2.1.2] - 2022-08-07. Zenodo https://doi.org/10.5281/ZENODO.2634132 (2023).
99. Finucane, H. K. et al. Partitioning heritability by functional annotation using genome-wide association summary statistics. *Nat. Genet.* **47**, 1228–1235 (2015).
100. Gazal, S. et al. Linkage disequilibrium-dependent architecture of human complex traits shows action of negative selection. *Nat. Genet.* **49**, 1421–1427 (2017).
101. Wang, L. W. et al. Modulating gene expression in Epstein-Barr Virus (EBV)-positive B cell lines with CRISPRa and CRISPRi. *Curr. Protoc. Mol. Biol.* **121**, 1–31 (2018). 31.1313.18.

102. Sanson, K. R. et al. Optimized libraries for CRISPR-Cas9 genetic screens with multiple modalities. *Nat. Commun.* **9**, 5416 (2018).
103. R.Walker, M. et al. Induction of FoxP3 and acquisition of T regulatory activity by stimulated human CD4+CD25− T cells. *J. Clin. Investig.* **112**, 1437–1443 (2003).
104. Wang, G., Sarkar, A., Carbonetto, P. & Stephens, M. A simple new approach to variable selection in regression, with application to genetic fine mapping. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **82**, 1273–1300 (2020).
105. R. C. Aguiar, V. gutierrez-arcelus-lab/bcellactivation: v1.0.0 — B cell activation multi-omics analysis. Preprint at https://doi.org/10.5281/ZENODO.22664586 (2026).
106. Morin, A. et al. Collaboration gets the most out of software. *eLife* **2**, e01456 (2013).

Acknowledgements

We would like to thank Dr. Duane Wesemann and his laboratory (Brigham & Women's Hospital) for helpful discussions on B cell activation, Stephen Gazal for guidance on application of LDSC-SEG, the Gutierrez-Arcelus and Nigrovic laboratories for helpful comments.

Author contributions

M.G.-A. conceived the study, secured funding, and supervised the work. V.R.C.A., M.E.F., and M.G.-A. refined research goals, designed the experiments, and drafted the manuscript. M.E.F. performed the optimization of experiments and the multiomics resource experiments. V.R.C.A. performed data analysis and visualization for the pilot experiments and the multiomics resource. N.A.A. and M.Co. performed the CRISPR and qPCR validation experiments. D.F.-S., M.Ch., Q.X., N.H., and V.P. supported data analyses and/or data visualization. C.D.L., H.A., and M.M. supported the GWAS fine-mapping analyses. Y.L. and B.E.G. supported CRISPR tool development. M.T., B.W., and A.W. performed pilot experiments that supported the development of methods for the multiomics resource. R.C.-D., R.L.-B., R.D., Q.W., M.B., P.L., L.H., M.H.C., A.D.S., and E.M. provided support in experimental methodology and/or B cell expertise. S.J., K.S.C., E.Z., and I.S. provided key support in DN2c B cell stimulation experiments and interpretation of results. P.C., M.G., and H.L. supported ATAC-seq experiments. Z.Z., J.C., and K.W. supported CITE-seq experiments. J.A.S. coordinated donor recruitment. P.A.N. provided resources, served as a scientific advisor, and provided critical review of the research direction and results. All authors reviewed and edited the manuscript.

Funding

V.R.C.A. was supported by a Joint Biology Consortium microgrant (Parent grant NIH/NIAMS 2P30 AR070253). Maria Gutierrez-Arcelus was supported by NIH/NIAID 2P30AR070253 and U19AI095219, the Lupus Research Alliance Lupus Innovation Award, the Lupus Research Alliance Empowering Research Career Development Award, the Vic Braden Family Arthritis National Research Foundation fellowship, the Gilead Sciences Rheumatology Research Scholars Award. This work was supported by two Boston Children's Hospital Office of Faculty Development/Basic & Clinical Translational Research Executive Committees Faculty Career Development Fellowships (to V.R.C.A. and M.G.A.). N.H. was supported by an MD fellowship from the Boehringer Ingelheim Fonds. P.A.N. was supported by the Lupus Research Alliance, NIH/NIAMS 2P30AR070253, 2R01AR065538, R01AR073201, and R01AR075906. J.A.S. is supported by the National Institute of Arthritis and Musculoskeletal and Skin Diseases (R01 AR080659, R01 AR077607, P30 AR070253, and P30 AR072577), the National Heart, Lung, and Blood Institute (R01 HL155522), the Rheumatology Research Foundation, the Arthritis Foundation, the R. Bruce and Joan M. Micke Research Scholar Fund, and the Llura Gund Award funded by the Gordon and Llura Gund Foundation. The project described was supported by Clinical Translational Science Award 1UL1TR002541-01 to Harvard University and Brigham and Women's Hospital from the National Center for Research Resources. B.E.G. was supported by a Broad Institute Founder's grant. Y.L. was supported by a Lymphoma Research Foundation award. C.D. Langefeld was supported by P01 AR49084, NIAMS, Program Project in the Genetics of SLE (UAB-SLE2), and P01 AI083194, NIH, Genomics of Lupus (SLEGEN P01). This work was supported by NIH/NIAID 1R01AI198376, and the Cell Discovery Network, a collaborative initiative funded by the Manton Foundation and the Warren Alpert Foundation at Boston Children's Hospital.

Competing interests

L.H. receives salary support from CARRA and investigator-initiated grant support from BMS. L.H. has received consulting fees from Sobi, speaking fees from Pfizer, and provided legal consulting services for Phillips, Parker, Orberson & Arnett. M.H.C. serves as a consultant for and holds equity in Attovia Therapeutics. J.A.S. has received research support from 10x Genomics, Boehringer Ingelheim, Bristol Myers Squibb, Johnson & Johnson, and Sonoma Biotherapeutics unrelated to this work. J.A.S. has performed consultancy for AbbVie, Anaptys, AstraZeneca, Boehringer Ingelheim, Bristol Myers Squibb, GSK, Immunovant, Invivyd, Johnson & Johnson, Novartis, Pfizer, Sun, and UCB unrelated to this work. The funders had no role in the decision to publish or preparation of this manuscript. The content is solely the responsibility of the authors and does not necessarily represent the official views of Harvard University, its affiliated academic health care centers, or the National Institutes of Health. The remaining authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41467-026-78656-w.

Correspondence and requests for materials should be addressed to Maria Gutierrez-Arcelus.

Peer review information *Nature Communications* thanks the anonymous reviewer(s) for their contribution to the peer review of this work. A peer review file is available.

Reprints and permissions information is available at http://www.nature.com/reprints

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

© The Author(s) 2026

This Article in Press is shared early to give you faster access to new research. It is citable and carries a permanent DOI. The final edited version will replace it automatically.

[1]Division of Immunology, Boston Children's Hospital, Harvard Medical School, Boston, MA, USA. [2]Program in Medical and Population Genetics, Broad Institute of MIT and Harvard, Cambridge, MA, USA. [3]Max Delbrück Center for Molecular Medicine in the Helmholtz Association, Berlin Institute for Medical System Biology, Berlin, Germany. [4]Biotechnology Department, Faculty of Science, Cairo University, Giza, Egypt. [5]Division of Immunology, Institute of Infectious Diseases and Molecular Medicine, Faculty of Health Sciences, University of Cape Town, Cape Town, South Africa. [6]Division of Rheumatology, Department of Medicine V, Heidelberg University Hospital, Heidelberg, Germany. [7]Division of Infectious Diseases, Department of Medicine, Brigham and Women's Hospital, Harvard Medical School, Boston, MA, USA. [8]Department of Medicine, Division of Rheumatology, Lowance Center for Human Immunology, Emory University, Atlanta, GA, USA. [9]Department of Microbiology, University of Alabama at Birmingham, Birmingham, AL, USA. [10]Center for Cellular Profiling, Brigham and Women's Hospital, Harvard Medical School, Boston, MA, USA. [11]Center for Functional Cancer Epigenetics, Dana-Farber Cancer Institute, Harvard Medical School, Boston, MA, USA. [12]Department of Biostatistics and Data Science, Wake Forest School of Medicine, Winston-Salem, NC, USA. [13]Center for Precision Medicine, Wake Forest School of Medicine, Winston-Salem, NC, USA. [14]Division of Rheumatology, Inflammation, and Immunity, Brigham and Women's Hospital, Harvard Medical School, Boston, MA, USA. [15]Berlin Institute of Health (BIH) Translational Research Unit, Berlin, Germany. [16]Charité – Universitätsmedizin Berlin, Berlin, Germany. [17]Present address: Institute of Regenerative Medicine and Biotherapy, INSERM, University of Montpellier, Montpellier, France. [18]These authors contributed equally: Vitor R. C. Aguiar, Marcella E. Franco. e-mail: mgutierr@broadinstitute.org