PRDM1 shapes germinal center B-cell clonal diversity by gating chromatin accessibility during light-to-dark zone transition

Godhev Kumar Manakkat Vijay[1,7], Bala Ramaswami[2,7], Steven Gierlack[1,7], Nicholas A. Pease[1,7], Peter Gerges[1], Dianyu Chen[3,4], Kairavee Thakkar[5,6], Luis Mena Hernandez[1], Swapnil Keshari[1], Heping Xu[3,4], Nathan Salomonis[5], Jishnu Das[1], David M. Rothstein[2#] and Harinder Singh[1#]

[1]Center for Systems Immunology and Department of Immunology, University of Pittsburgh, Pittsburgh, PA, USA; [2]Thomas E. Starzl Transplantation Institute, Department of Surgery, University of Pittsburgh, Pittsburgh, PA, USA; [3]Key Laboratory of Growth Regulation and Translational Research of Zhejiang Province, School of Life Sciences, Westlake University, Hangzhou 310024, Zhejiang, China; [4]Center for Infectious Disease Research, Westlake Laboratory of Life Sciences and Biomedicine, Hangzhou 310024, Zhejiang, China; [5]Division of Bioinformatics, Cincinnati Children's Hospital and Medical Center, Cincinnati, Ohio, USA; [6]Department of Pharmacology and Physiology, University of Cincinnati, College of Medicine, Cincinnati, Ohio, USA; [7]These authors contributed equally

[#]Corresponding authors: rothsteind@upmc.edu; harinder@pitt.edu

Abstract

Germinal-center (GC) B-cell responses are defined by many positive regulators of affinity maturation, but few components that restrain clonal dominance, notably *Nr4a1*, are known. We reveal an unsuspected role for PRDM1 (BLIMP1), a plasma-cell determinant, as a feedback regulator of affinity maturation. Single cell RNA-seq and BCR-seq showed that B-cell–specific *Prdm1* loss drives an exaggerated GC reaction with larger clones, increased somatic hypermutation, and greater clonal dominance, independent of *Nr4a1*. Single cell chromatin profiling with base-resolution modelling indicated that PRDM1 represses expression of BCR-signaling genes and gates chromatin accessibility at ISRE, EICE, NF-κB, and POU (Oct) motifs. In the absence of PRDM1, enhanced engagement of signaling-inducible transcription factors promotes G1–S transition during light-zone (LZ) selection and fuels dark-zone (DZ) expansion. Thus, PRDM1 attenuates BCR signaling and constrains the LZ→DZ transition, fine-tuning clonal competition thereby maintaining repertoire diversity. The chromatin-encoded checkpoint could be leveraged to modulate vaccine responses.

Introduction

Extensive studies have delineated signaling systems and transcription factors (TFs) that positively regulate germinal center (GC) B cell responses[1-11]. In the LZ, T follicular helper (T_{fh})-derived CD40L and ICOS along with B cell receptor (BCR) signals license B cell clones to pass through the G1-S checkpoint; the cells then re-enter the DZ for clonal expansion and AID-mediated hypermutation. Among TFs that promote GC initiation, selection, and output, BCL6 programs GC identity. MYC licenses positively selected LZ cells for re-entry and proliferation in the DZ, whereas FOXO1 controls the program underlying rapid cell divisions. PU.1 and SPIB via interactions with IRF8 help maintain GC identity and selection, acting on ETS–IRF composite elements (EICE)[12,13]. BATF supports robust GC responses and class switching[14]. The POU TF module comprised of OCT1/POU2F1, OCT2/POU2F2, and their co-activator OBF1/POU2AF1 maintains BCR-dependent transcription required for GC expansion[15]. Canonical and alternative NF-κB components c-Rel and p52/RelB also promote GC maintenance, a p52–ETS1 complex has been recently shown to induce OCT1 and OBF1, thereby representing a feedforward regulatory loop[16,17].

In contrast, only one signaling-induced negative regulator of the GC response that restricts clonal dominance has been identified, notably *Nr4a1* (NUR77)[18,19]. We posited that PRDM1 (BLIMP1) which promotes exit of GC B cells into the plasma cell (PC) pathway, could also act as a signaling-induced transient feedback gate to restrain the GC B cell response, thereby reducing clonal dominance. In keeping with this hypothesis, low and heterogenous expression of PRDM1 has been reported in GC B cells and its loss has been shown to result in larger GCs[20,21].

Using single-cell (sc) transcriptome, V(D)J and chromatin profiling coupled with single nucleotide-resolution accessibility modeling of regulatory DNA sequences, we show that *Prdm1*-deficient B cells mount an exaggerated GC reaction characterized by larger clone sizes and enhanced affinity maturation culminating in greater clonal dominance. This phenotype is not attributable merely to impaired exit of GC B cells into the PC pathway. Rather, integrated genomic analyses indicate that PRDM1 constrains expression of genes encoding components of the BCR-signaling cascade. Loss of this feedback augments

chromatin engagement of signaling-inducible TFs at EICE, NF-κB and POU (OCT) motifs, promoting the G1–S transition during LZ selection and fueling DZ expansion.

Results

PRDM1 deficient B cells undergo more robust GC response

To test the hypothesized function of PRDM1 in GC B cells, we utilized *Prdm1*$^{fl/fl}$*Cd19*$^{cre/+}$ (*Prdm1* cKO) and *Prdm1*$^{+/+}$*Cd19*$^{cre/+}$ (WT) mice immunized with NP-CGG alum and LPS (Fig. 1a). Loss of PRDM1 resulted in increased frequencies and numbers of NP^{+} GC B cells at 7- and 14-days post-immunization (d.p.i.) (Fig. 1b-d and Extended data Fig. 1a). Notably, the increase in absolute numbers of LZ and DZ GCBCs in the *Prdm1* cKO mice was accompanied with a decreased relative frequency of LZ GCBCs (Fig. 1e-g and Extended Data. Fig. 1b,c). These results were consistent with the hypothesis that PRDM1 constrains the GC response, during the LZ to DZ transition.

scRNA-seq reveals PRDM1 constrains LZ to DZ transition

To analyze the impact of PRDM1 perturbation on transcriptional programming during the LZ to DZ transition, we performed scRNA-seq on NP-specific B cells[7] at 14 d.p.i. (Fig. 1a). Analyses of the WT scRNA-seq datasets using the ICGS2 computational pipeline[22] identified five distinct cell clusters that could be distinguished based on biologically informative, marker genes (Fig. 2a and Supplementary Table 1). Within the two LZ clusters, LZ-1 manifested the highest expression of B cell activation markers, including *Gpr183*, *Cd55* and *Cd38*. In contrast, LZ-2 showed elevated expression of genes involved in antigen presentation and B cell receptor (BCR) signaling, including *Cd83*, *H2-DMa*, *H2-Oa* and *Nfkbia*. The two DZ clusters, DZ-1 and DZ-2, resolved cells within the G1→S and G2→M phases of the cell cycle, respectively. DZ-1 was characterized by high expression of DNA replication genes (*Mcm5*, *Mcm7*, *Pcna)*, whereas DZ-2 manifested higher expression of mitotic components and their regulators, including *Cenpf*, *Ccnb1* and *Cdk1* (Fig. 2a and Supplementary Table 2). Notably, the analysis identified a distinctive GCBC cluster, herein termed transition zone (TZ), which exhibited elevated expression of ribosomal biogenesis and translation genes, including *Npm1*, *Nop58*, *Fbl* and *Eif5a* (Fig. 2a,b and Supplementary Table 2). Consistent with this transcriptional module, TZ cells showed the highest expression of *Myc* target genes,

highlighting their importance as key intermediates that integrate activation and proliferation cues during the transitioning of LZ cells into the DZ[23,24] (Fig. 2c). Gene signature analyses using curated GC B cell gene sets[7] substantiated the conclusion that the TZ cluster represents an intermediate state, as the cells displayed the highest *Myc* score and also an increase in G1/S associated genes compared with LZ cells (Extended Data Fig. 2a-c). We note that a subset of TZ cells also expressed increased levels of transcripts that have previously been associated with cells transitioning from the LZ to the DZ, termed gray zone (GZ) GCBC[25] (Fig. 2d, Extended Data Fig. 2d,e and Supplementary Table 3). To substantiate the developmental trajectory inferred by the annotations of the ICGS2 clusters based on their marker genes, we performed RNA velocity using scVelo[26], which leveraged the dynamics of spliced and unspliced transcripts. This analysis reinforced the inference that TZ cells arise from LZ-2 and give rise to DZ-1 cells (Fig. 2e).

Next, cellHarmony was used to align the *Prdm1* cKO GC B cells with their WT counterparts[27]. Consistent with the phenotypic analysis (Fig. 1), the relative frequency of *Prdm1* cKO DZ GC B cells was increased, in relation to their LZ counterparts (Extended Data Fig. 2f). Furthermore, the scRNAseq analysis suggested that the inversion of LZ and DZ frequencies may occur within TZ cells. To explore the molecular mechanisms by which PRDM1 impacts GC state transitions, we performed differential gene expression analyses. In keeping with the dominant molecular actions of PRDM1 as a transcriptional repressor we observed a larger number of upregulated compared with downregulated differentially expressed genes (DEGs), 749 and 260, respectively (Supplementary Table 4). The largest number of upregulated DEGs were observed in the LZ-2 and TZ clusters (Extended Data Fig. 2g). In the LZ-2 cluster, *Prdm1* cKO GC B cells exhibited enrichment of genes involved in BCR signaling and translation (Fig. 2f and Supplementary Table 5). In contrast, genes associated with the G1/S transition, DNA replication and cell cycle progression were upregulated in the TZ cluster (Fig. 2g and Supplementary Table 5). Notably, both clusters showed increased expression of genes known to be regulated by the TFs MYC, SPI1 (PU.1) and IRF8 (Fig. 2h,i and Supplementary Table 5). Consistent with the above analyses, the G1/S signature scores were increased in TZ and DZ-1 clusters (Fig. 2j). Taken together these transcriptional alterations agreed with the hypothesized function of PRDM1 as a negative feedback

regulator, the absence of which in the LZ, results in enhanced BCR signaling and precocious programming of *Myc*-mediated protein translation and the G1→S transition.

Expression of *Prdm1* in GC B cells was previously reported as low and sparse[7,28]. To determine whether *Prdm1* expression is in accord with its proposed function in constraining the LZ to DZ transition, we quantified the expression of *Prdm1* across various WT GC clusters in comparison with *Nr4a1*. In keeping with its induction during B cell activation as a consequence of BCR signaling[18], *Nr4a1* transcripts were detected with the highest frequency (~30%) and median values, in the LZ-1 cluster (Fig. 2k and Extended data Fig. 2h). We note that *Nr4a1* expression was not impacted in *Prdm1* cKO GC B cells (Supplementary Table 4). Though *Prdm1* transcripts were only detected in a small fraction of WT GC B cells (~1%), like *Nr4a1,* they manifested highest median values in the LZ-1 cells (Fig. 2l and Extended data Fig. 2i). Importantly, *Prdm1* and *Nr4a1* transcripts were strongly correlated in cells that displayed their coincident expression (Fig. 2m). Thus, despite sparse detection of *Prdm1* transcripts in LZ cells, its expression pattern was consistent with its induction in BCR activated LZ cells and in restraining the LZ → DZ transition not dependent on controlling *Nr4a1* expression .

PRDM1 attenuates affinity maturation and clonal dominance

The above results implied that affinity maturation and clonal dominance in the GC may be accelerated with the loss of PRDM1. To directly test this possibility, we leveraged scBCR-seq data generated in parallel with the scRNA-seq. To analyze NP-specific GC B cell clones, we focused on those harboring V_H186.2 heavy chain gene segment (IGHV1-72*01) rearrangements which dominate the NP-specific GC B cell response at its peak[7,29] (Supplementary Table 6). The W33L amino acid substitution has been shown to increase NP-binding affinity of the BCR containing the IGHV1-72*01 gene segment by ten-fold and represents the dominant mutation after positive selection during a peak GC response[7]. In keeping with the increased numbers and relative frequencies of NP-specific DZ GC B cells, we noted larger clone sizes and increased SHM within *Prdm1* cKO GC B cells during the initiation of the GC B cell response (7 d.p.i.). However, enhanced affinity maturation was not observable at this early stage (Fig. 3a-c). As was the case at 7 d.p.i., the clonal sizes of the

Prdm1 cKO GC B cells were significantly larger (14 d.p.i.) than their WT counterparts (Fig. 3d) suggesting that mutant cells in the DZ are expanded not only due to more efficient LZ →DZ transition but possibly due to faster cell cycle times. In keeping with these observations, the mutation frequencies of the IGHV1-72*01 gene segment were also increased in *Prdm1* cKO GC B cells at 14 d.p.i. (Fig. 3e). Importantly at the peak of the GC response, there was a significant increase in the frequency of W33L mutation bearing clones in the absence of PRDM1 (Fig. 3f).

To analyze clonal dominance, we enumerated all clonotypes based on their paired IgH and IgL sequences in all WT or *Prdm1* cKO NP-specific B cells. We then constructed Lorenz curves that display cumulative fraction of cells versus cumulative fraction of clonotypes, after ranking clonotypes from smallest to largest. In such an analysis a repertoire with every clonotype being equally represented, would generate a perfectly uniform distribution represented by a diagonal (Fig. 3g). As expected for non-uniform Ig repertoires, both the WT and *Prdm1* cKO distributions were displaced from the diagonal. Importantly the *Prdm1* cKO distribution showed greater skewing from the diagonal indicating that a smaller subset of clonotypes were accounting for a large fraction of cells. Next we used the Gini coefficient to quantify this skewing based on area defined by the observed Lorenz curve and the diagonal. Values of the Gini coefficient can range from 0 (uniform repertoire) to 1 (maximal clonal dominance with all cells being of a single clonotype). The Gini coefficient for *Prdm1* cKO NP-specific Ig repertoire was significantly higher than that of WT cells (Fig. 3h) thereby providing evidence for clonal dominance. As a complementary Ig repertoire diversity metric, we computed Shannon entropy scores for the two distributions. In contrast with the Gini index, Shannon entropy increases with number of clonotypes and uniformity in their distributions. As expected Shannon entropy values were lower for the *Prdm1* cKO repertoires (Fig. 3i). Together, these data show that, in the absence of PRDM1, affinity maturation and clonal dominance are enhanced at the peak of the GC response.

PRDM1 gates chromatin accessibility at ISRE, EICE, NF-κB, and Oct motifs

To analyze the molecular consequences of PRDM1 action on chromatin accessibility and to infer its target genes in GC states, we performed scMultiome-seq, jointly profiling gene

expression and chromatin accessibility using NP^+B220^+ splenocytes isolated from WT and *Prdm1* cKO mice (14 d.p.i.). MIRA-based joint topic modeling[30], which involves representation of each cell according to weighted mixtures of gene expression and chromatin accessibility revealed six discrete clusters, C1–C6 (Fig. 4a). Gene expression and chromatin accessibility topics that were specific for clusters C1-C6 are displayed (Fig. 4b,c) and their associated genes and chromatin coordinates are provided (Supplementary Table 7). Notably, a subset of C4 cells displayed the highest *Myc* gene signature score (Fig. 4d). Support vector machine (SVM) classification was used to analyze the correspondence between the six WT multiome clusters with the five reference scRNA-seq clusters, the latter designated as "states" hereafter (Fig. 4e-g and Extended Fig. 3a,b). Multiome clusters C1 and C2 were enriched for the LZ-1 state, C3 for LZ-2, C4 for TZ, C5 for DZ-1, and C6 for DZ-2. As was the case for scRNA-seq, multiome clusters derived from *Prdm1* cKO cells showed an overall correspondence with their WT counterparts, suggesting preservation of global GC state identities despite the perturbation. SVM classification was next used to infer alterations in relative frequencies of GC states because of the perturbation. The *Prdm1* cKO C3 cells manifested a reduction in relative frequency of LZ-2 classified cells with an increase in DZ-1 classified cells (Fig. 4h). Similarly, *Prdm1* cKO C4 cells displayed a reduction in relative frequency of TZ classified cells with an increase in DZ-1 cells (Fig. 4i). Thus, PRDM1 KO LZ and TZ cells exhibited DZ-like features. Overall, these results were consistent with the scRNA-seq analysis, collectively demonstrating that PRDM1 loss enhances LZ→TZ and TZ→DZ state transitions. Consistent with this interpretation, analysis of alterations in chromatin accessibility revealed a substantial increase in upregulated differentially accessible regions (upDARs) in the C3 and C4 *Prdm1* cKO clusters compared with C1 and C2 (Fig. 4j). Many of these upDARs were associated with G1→S genes whose expression was increased in C3 and C4 *Prdm1* cKO cells (see below).

To analyze TF motif contributions to chromatin structure in WT GCBCs and their perturbations in the *Prdm1* cKO counterparts, at base-pair resolution, we utilized ChromBPNet[31]. The deep learning model predicts TF motif activity, while accounting for Tn5 transposase cutting bias. This approach enabled the systematic identification of TF motifs underlying chromatin accessibility changes during GCBC state transitions. The scMultiome

dataset was used to train cluster-specific ChromBPNet models on pseudobulk ATAC-seq profiles derived from a given GCBC cluster. Each of these cluster-specific models was then used to generate corresponding base-pair resolution contribution score (CS) predictions across the union set of ATAC-seq peaks. These cluster-specific contribution score tracks revealed short DNA sequences, referred to as "seqlets", that were predicted to be important in determining the accessibility profiles of their corresponding ATAC-seq peaks. Strikingly, many of the ChromBPNet seqlets corresponded to known TF motifs thereby implicating action of cognate TFs in binding to such sequences and controlling chromatin accessibility (Supplementary Table 8). Single nucleotide CS predictions for selected motifs, aggregated across all instances of their occurrence within accessible chromatin regions, are shown for each GCBC cluster (Fig. 5a). The analysis inferred dynamic action of multiple signaling induced TFs in regulating chromatin accessibility across GC B cell states with an increase in NFKB, MYC and POU activity preceding that of AP-1 (BATF).

We next measured the enrichment of TF motif family seqlets in the upregulated DARs of *Prdm1* cKO GC B cells (Supplementary Table 9). The upregulated C3 and C4 DARs were enriched for seqlets which matched EICE and ISRE motifs known to be bound by PU.1/SPIB:IRF4/IRF8 and IRF1/IRF2, respectively as well as PRDM1, (Fig. 5b,c). PRDM1 has been shown to antagonize the forementioned activators binding to EICE and ISRE motifs by recognition of an overlapping DNA sequence[32,33]. Additionally, we also observed enrichment for PAX, NFKB and POU motif seqlets in these regions, which predicted increased activities of their cognate TFs in the absence of PRDM1. These findings suggested a chromatin gating function of PRDM1 during the G1→S transition during which it antagonizes the action of constitutively expressed as well as signaling induced TFs in GC B cells.

PRDM1 represses expression of BCR-signaling genes in GC B cells

To distinguish the primary and secondary effects of loss of PRDM1 on chromatin accessibility and gene expression, we sought to identify its direct transcriptional targets by integrating DEGs with ChromBPNet seqlet predictions and upDARs (Fig. 6a). A shared set of DEGs from the multiome-seq and scRNA-seq (Supplementary Table 10) enabled the identification of presumptive PRDM1-repressed target genes. These genes (n=380) were upregulated in the

absence of PRDM1 and were associated with one or more ISRE, EICE or PRDM1 motif seqlets within an OCR that was ± 200 kb of their transcription start site (TSS) (Fig. 6b,c and Extended Data Fig. 4a-d). Within this target gene set we identified those that were associated with one or more upDARs (n=154) that in turn contained a PRDM1-related motif seqlet (n=23) (Supplementary Table 11). Gene set enrichment analysis of these PRDM1 targets revealed significant overrepresentation of pathways involved in BCR signaling e.g., *Syk, Lyn, Grb2, Itpr2* and G1→S progression e.g., *Ccnd3, Mcm3, Rfc4, Pold4* (Fig. 6d,e and Supplementary Table 11). These results suggest that PRDM1 functions to constrain the LZ to DZ transition by directly repressing expression of genes involved in BCR signaling and cell cycle progression.

Discussion

Our analysis has revealed an unsuspected function of PRDM1 as a feedback attenuator of the GC B cell response. We propose that transient and sub threshold levels of PRDM1, that are induced by BCR and CD40 signaling, feedback to inhibit signal transduction through these pathways and in so doing lower clonal dominance that can accompany affinity maturation. ChromBPNet analysis of chromatin accessibility in *Prdm1* cKO GC B cells delineated candidate upregulated target genes within BCR-signaling pathways e.g., *Syk, Lyn, Grb2, Itpr2* whose regulatory regions harbor ISRE/EICE/PRDM1 seqlets contributing to open chromatin. We propose that PRDM1 restrains transcription at these elements in part by opposing IRF and ETS family activators e.g., PU.1/SPIB:IRF8[32,33]. Enhanced BCR signaling via modest yet coordinated increases in expression of multiple genes encoding signaling components would be expected to elevate signaling-induced TFs. In line with such a signaling-amplification model, upDARs in *Prdm1* cKO LZ cells were enriched for NF-κB and POU (Oct) seqlets, consistent with increased activity of the cognate TFs. Recent work indicates that the NF-κB p52–ETS1 complex promotes GC formation by inducing POU2F1 (OCT1) and POU2AF1 (OBF1)[17], factors that are required to sustain BCR signaling and GC maintenance[15]. We identified *Pou2af1* as a presumptive PRDM1 target gene, it displayed increased expression in *Prdm1*-deficient GC B cells, suggesting that PRDM1 restrains POU activity in part via direct repression of *Pou2af1*. Notably, *Prdm1* expression is reduced in p52–ETS1-deficient GC B cells, consistent with a model in which BCR/CD40-driven NF-

κB activity transiently induces PRDM1 as well as *Pou2af1*. Taken together, the p52–ETS→POU (positive) and p52–ETS→PRDM1 (negative) gene regulatory interactions form an incoherent module that tunes GC amplitude and diversity, reminiscent of the previously proposed IRF4–BCL6–PRDM1 incoherent feed-forward architecture[34]. How these two regulatory modules interface with *Nr4a1* remains to be more fully explored. *Nr4a1* is an immediate early gene that is induced by BCR signaling[18] wheras *Prdm1* appears to manifest slower signaling dependent induction[21]. Notably, we did not observe diminished expression of *Nr4a1* in *Prdm1*-deficient GC B cells indicating that PRDM1 action in tuning GC amplitude and Ig repertoire diversification is not mediated through controlling the expression of *Nr4a1*.

Strong T_{fh}–B-cell interactions promote PC differentiation via *Prdm1* induction, whereas weaker interactions favor recycling[28]; our data extend this paradigm by showing that PRDM1 also restricts GC expansion by gating the LZ→DZ transition. While this manuscript was in preparation, related work using bone-marrow chimeras, independently demonstrated a B-cell–intrinsic role for PRDM1 in limiting GC expansion[35]. Importantly the two studies confirm and complement each other's findings. The reported ~2-fold increase in MYC protein and S-phase GC fractions aligns with our observation of precocious G1–S programming and elevated MYC target signatures in *Prdm1* cKO cells; the absence of increased *Myc* transcripts in our datasets suggests post-transcriptional or post-translational mechanisms of MYC elevation. Our multiome analyses point to PRDM1-dependent chromatin gating as a direct means to tune expression of BCR-signaling genes; modest, coordinated upregulation of components such as *Syk*, *Lyn*, and *Grb2* in the absence of PRDM1 likely accounts for increased NF-κB and POU chromatin action, promoting G1–S entry and DZ proliferation. We propose that transient, NF-κB-linked PRDM1 induction in the LZ feeds back to dampen signaling components and constrain NF-κB/POU activity, thereby limiting LZ→DZ flux. In the absence of PRDM1, positively selected clones undergo premature G1–S commitment and increased DZ proliferation. Thus, PRDM1 functions as a molecular brake on GC progression that limits clonal dominance and shapes repertoire diversity. Therapeutic targeting of the PRDM1 chromatin-encoded GC checkpoint may prove useful in modulating the specificity and breadth of vaccine responses.

References:

1. Victora, G.D. & Nussenzweig, M.C. Germinal Centers. *Annual Review of Immunology* **40**, 413-442 (2022).

2. Papin, A., Cesarman, E. & Melnick, A. 3D chromosomal architecture in germinal center B cells and its alterations in lymphomagenesis. *Current Opinion in Genetics & Development* **74**, 101915 (2022).

3. Foy, T.M. et al. gp39-CD40 interactions are essential for germinal center formation and the development of B cell memory. *The Journal of experimental medicine* **180**, 157-163 (1994).

4. Dent, A.L., Shaffer, A.L., Yu, X., Allman, D. & Staudt, L.M. Control of Inflammation, Cytokine Expression, and Germinal Center Formation by BCL-6. *Science (New York, N.Y.)* **276**, 589-592 (1997).

5. Ersching, J. et al. Germinal Center Selection and Affinity Maturation Require Dynamic Regulation of mTORC1 Kinase. *Immunity* **46**, 1045-1058.e1046 (2017).

6. Luo, W., Weisel, F. & Shlomchik, M.J. B cell receptor and CD40 signaling are rewired for synergistic induction of the c-Myc transcription factor in germinal center B cells. *Immunity* **48**, 313-326. e315 (2018).

7. Chen, D. et al. Coupled analysis of transcriptome and BCR mutations reveals role of OXPHOS in affinity maturation. *Nature Immunology* **22**, 904-913 (2021).

8. Luo, W. et al. IL-21R signal reprogramming cooperates with CD40 and BCR signals to select and differentiate germinal center B cells. *Science Immunology* **8**, eadd1823 (2023).

9. Liu, D. et al. T–B-cell entanglement and ICOSL-driven feed-forward regulation of germinal centre reaction. *Nature* **517**, 214-218 (2015).

10. Chen, S.T., Oliveira, T.Y., Gazumyan, A., Cipolla, M. & Nussenzweig, M.C. B cell receptor signaling in germinal centers prolongs survival and primes B cells for selection. *Immunity* **56**, 547-561.e547 (2023).

11. Shlomchik, M.J. & Weisel, F. Germinal center selection and the development of memory B and plasma cells. *Immunological Reviews* **247**, 52-63 (2012).

12. Xu, H. et al. Regulation of bifurcating B cell trajectories by mutual antagonism between transcription factors IRF4 and IRF8. *Nat Immunol* **16**, 1274-1281 (2015).

13. Wang, H. et al. Transcription factors IRF8 and PU. 1 are required for follicular B cell development and BCL6-driven germinal center responses. *Proceedings of the National Academy of Sciences* **116**, 9511-9520 (2019).

14. Ise, W. et al. The transcription factor BATF controls the global regulators of class-switch recombination in both B cells and T cells. *Nature immunology* **12**, 536-543 (2011).

15. Song, S. et al. OBF1 and Oct factors control the germinal center transcriptional program. *Blood* **137**, 2920-2934 (2021).

16. De Silva, N.S. et al. Transcription factors of the alternative NF-κB pathway are required for germinal center B-cell development. *Proceedings of the National Academy of Sciences* **113**, 9063-9068 (2016).

17. Morgan, D. et al. The transcription complex p52–ETS1 is essential for germinal center formation. *Nature Immunology* (2025).

18. Brooks, J.F. et al. Negative feedback by NUR77/Nr4a1 restrains B cell clonal dominance during early T-dependent immune responses. *Cell reports* **36** (2021).

19. Tan, C. et al. NR4A nuclear receptors restrain B cell responses to antigen when second signals are absent or limiting. *Nature Immunology* **21**, 1267-1279 (2020).

20. Radtke, D. & Bannard, O. Expression of the plasma cell transcriptional regulator Blimp-1 by dark zone germinal center B cells during periods of proliferation. *Frontiers in immunology* **9**, 3106 (2019).

21. Martins, G. & Calame, K. Regulation and Functions of Blimp-1 in T and B Lymphocytes. *Annual Review of Immunology* **26**, 133-169 (2008).

22. Venkatasubramanian, M., Chetal, K., Schnell, D.J., Atluri, G. & Salomonis, N. Resolving single-cell heterogeneity from hundreds of thousands of cells through sequential hybrid clustering and NMF. *Bioinformatics* **36**, 3773-3780 (2020).

23. Calado, D.P. et al. The cell-cycle regulator c-Myc is essential for the formation and maintenance of germinal centers. *Nature Immunology* **13**, 1092-1100 (2012).

24. Dominguez-Sola, D. et al. The proto-oncogene MYC is required for selection in the germinal center and cyclic reentry. *Nature Immunology* **13**, 1083-1091 (2012).

25. Kennedy, D.E. et al. Novel specialized cell state and spatial compartments within the germinal center. *Nature Immunology* **21**, 660-670 (2020).

26. La Manno, G. et al. RNA velocity of single cells. *Nature* **560**, 494-498 (2018).

27. DePasquale, E.A.K. et al. cellHarmony: cell-level matching and holistic comparison of single-cell transcriptomes. *Nucleic Acids Research* **47**, e138-e138 (2019).

28. Ise, W. et al. T Follicular Helper Cell-Germinal Center B Cell Interaction Strength Regulates Entry into Plasma Cell or Recycling Germinal Center Cell Fate. *Immunity* **48**, 702-715.e704 (2018).

29. Manakkat Vijay, G.K. et al. Temporal dynamics and genomic programming of plasma cell fates. *Nature Immunology* **25**, 1097-1109 (2024).

30. Lynch, A.W. et al. MIRA: joint regulatory modeling of multimodal expression and chromatin accessibility in single cells. *Nature methods* **19**, 1097-1108 (2022).

31. Pampari, A. et al. ChromBPNet: bias factorized, base-resolution deep learning models of chromatin accessibility reveal cis-regulatory sequence syntax, transcription factor footprints and regulatory variants. *BioRxiv*, 2024.2012. 2025.630221 (2025).

32. Minnich, M. et al. Multifunctional role of the transcription factor Blimp-1 in coordinating plasma cell differentiation. *Nat Immunol* **17**, 331-343 (2016).

33. Kuo, T.C. & Calame, K.L. B lymphocyte-induced maturation protein (Blimp)-1, IFN regulatory factor (IRF)-1, and IRF-2 can bind to the same regulatory sites. *Journal of immunology (Baltimore, Md. : 1950)* **173**, 5556-5563 (2004).

34. Sciammas, R. et al. An incoherent regulatory network architecture that orchestrates B cell diversification in response to antigen signaling. *Molecular Systems Biology* **7**, 495-495 (2011).

35. Conter, L. et al. BLIMP1 controls GC B cell expansion and exit through regulating cell cycle progression and key transcription factors BCL6 and IRF4. *Cell reports* **44** (2025).

Fig. 1

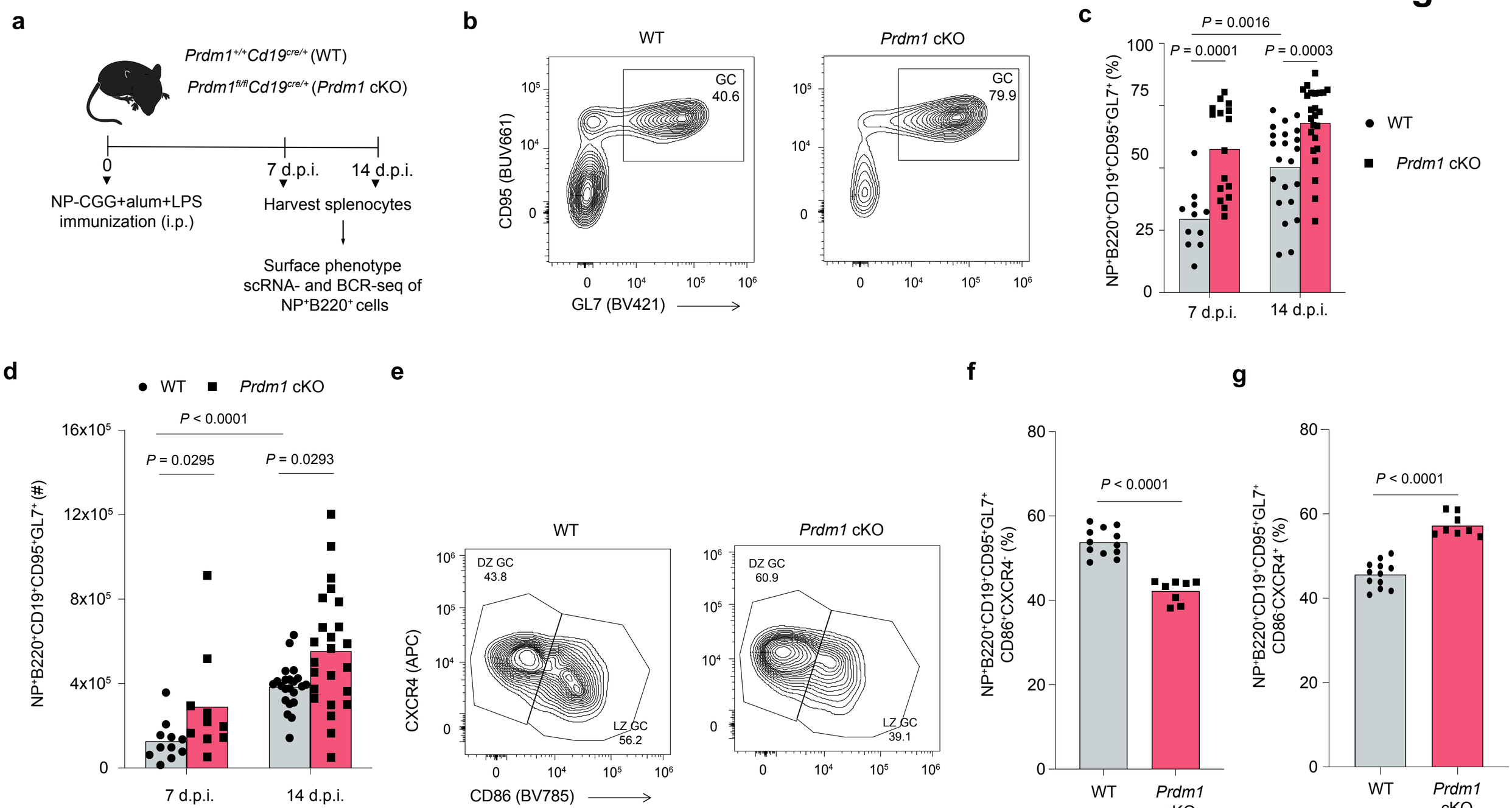
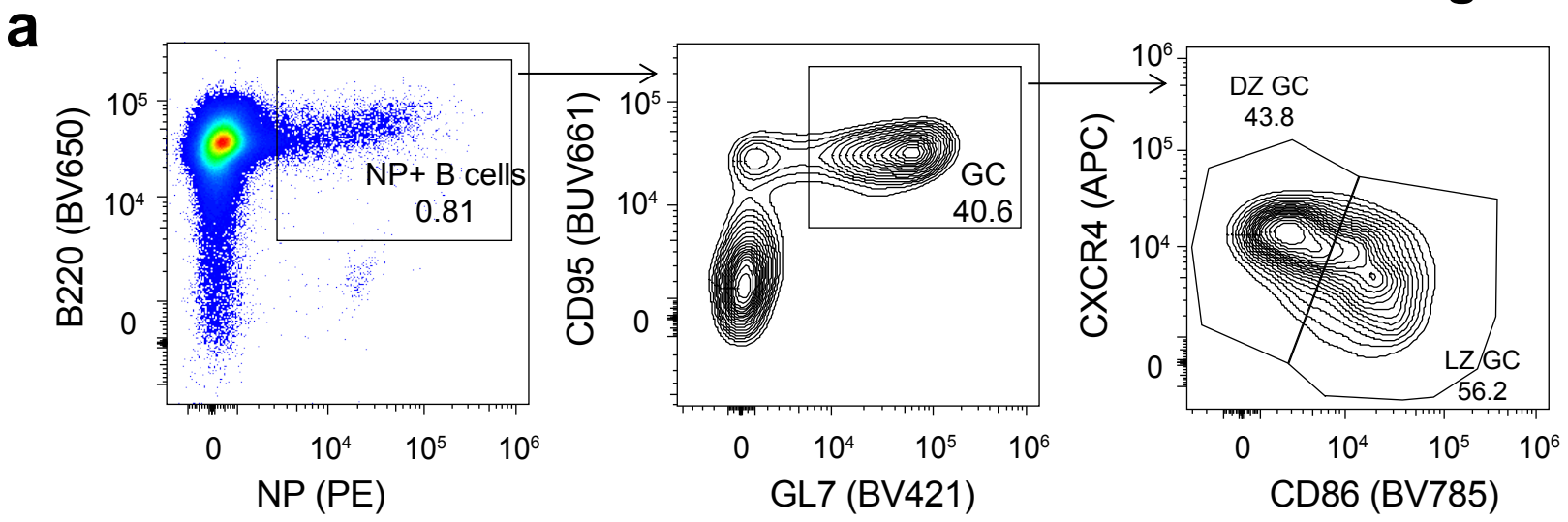
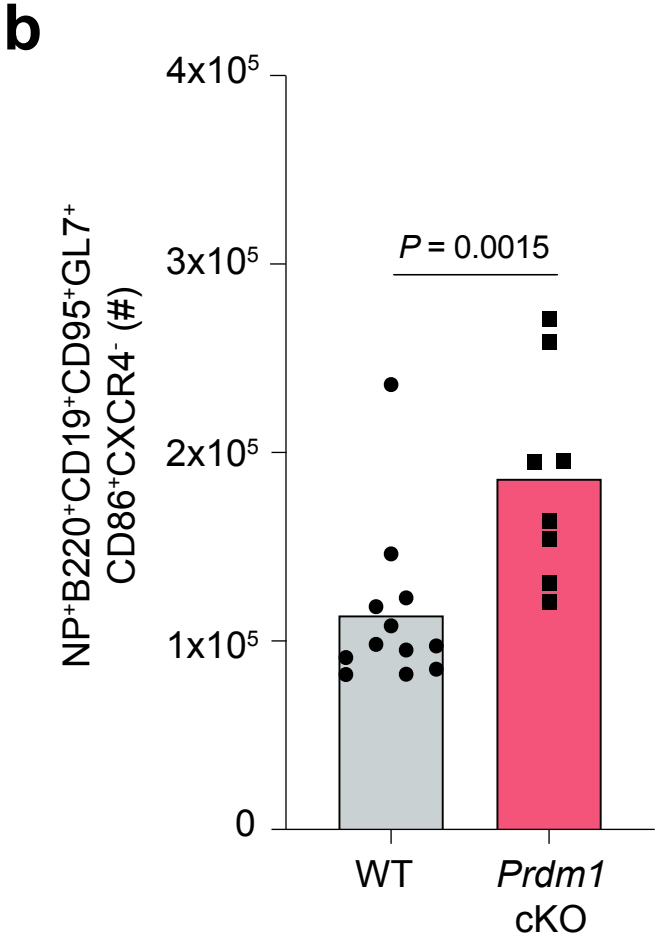
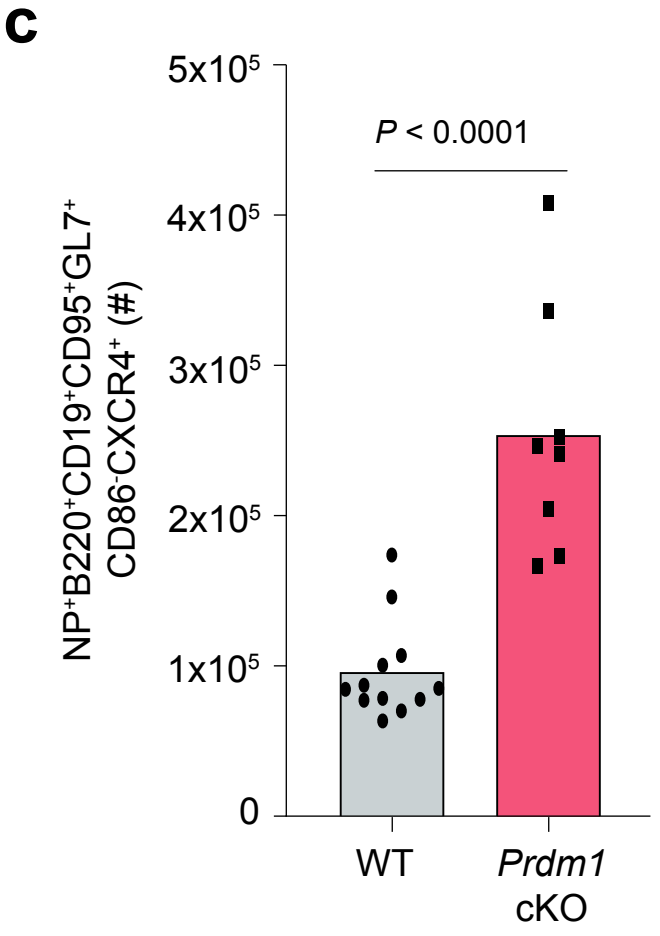

a
Prdm1+/+Cd19cre/+ (WT)
Prdm1fl/flCd19cre/+ (Prdm1 cKO)
0
7 d.p.i.
14 d.p.i.
NP-CGG+alum+LPS immunization (i.p.)
Harvest splenocytes
Surface phenotype scRNA- and BCR-seq of NP+B220+ cells
b
WT
Prdm1 cKO
GC 40.6
GC 79.9
CD95 (BUV661)
GL7 (BV421)
c
P = 0.0016
P = 0.0001
P = 0.0003
NP+B220+CD19+CD95+GL7+ (%)
7 d.p.i.
14 d.p.i.
WT
Prdm1 cKO
d
WT
Prdm1 cKO
P < 0.0001
P = 0.0295
P = 0.0293
NP+B220+CD19+CD95+GL7+ (#)
7 d.p.i.
14 d.p.i.
e
WT
Prdm1 cKO
DZ GC 43.8
LZ GC 56.2
DZ GC 60.9
LZ GC 39.1
CXCR4 (APC)
CD86 (BV785)
f
P < 0.0001
NP+B220+CD19+CD95+GL7+ CD86-CXCR4- (%)
WT
Prdm1 cKO
g
P < 0.0001
NP+B220+CD19+CD95+GL7+ CD86-CXCR4+ (%)
WT
Prdm1 cKO

Fig. 1. PRDM1 loss results in GC expansion and positively skews DZ/LZ ratio. a, Experimental design for analyzing frequencies, single cell transcriptional states and BCR sequences of NP-specific B cells elicited by immunization of WT and *Prdm1* cKO mice. **b,** Representative flow plots showing the frequency of NP^+ GC B cells in WT and *Prdm1* cKO mice (14 d.p.i.). **c,** Frequencies of NP^+ GC B cells in WT and *Prdm1* cKO mice (7 and 14 d.p.i.). **d,** Total number of NP-specific GC B cells in WT and *Prdm1* cKO mice (7 and 14 d.p.i.). **e,** Representative flow plots showing the frequencies of LZ and DZ GC B cells in WT and *Prdm1* cKO mice. **f,g,** Frequencies of NP-specific LZ and DZ GC B cells in WT and *Prdm1* cKO mice (14 d.p.i.) respectively. Each symbol (**c,d,f,g**) represents an individual mouse. Data in panels **b-d** are from three different experiments and **e-g** are from two independent experiments. Statistical significance was tested by Mann-Whitney test (**c,d,f,g**).

Extended Data Fig. 1







Extended Data Fig. 1. PRDM1 loss results in GC expansion and positively skews DZ/LZ ratio. a, Gating strategy for analyzing NP-specific GC B cells in a WT mouse (14 d.p.i.). **b,c** Total number of NP-specific LZ and DZ GC B cells in WT and *Prdm1* cKO mice (14 d.p.i.) respectively. Data in panels **b,c** are from two independent experiments. Statistical significance was tested by Mann-Whitney test (**b,c**).

Fig. 2

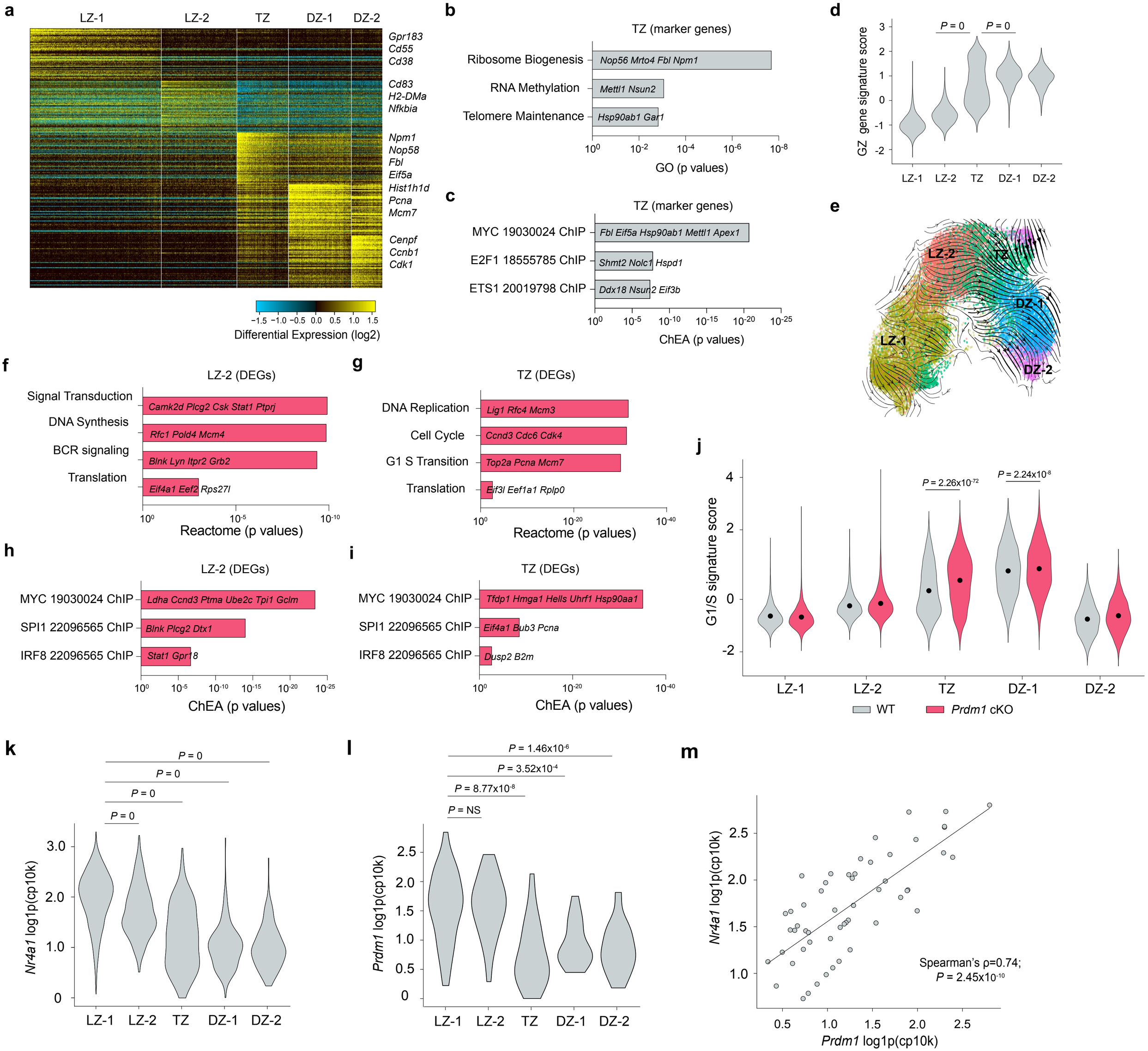
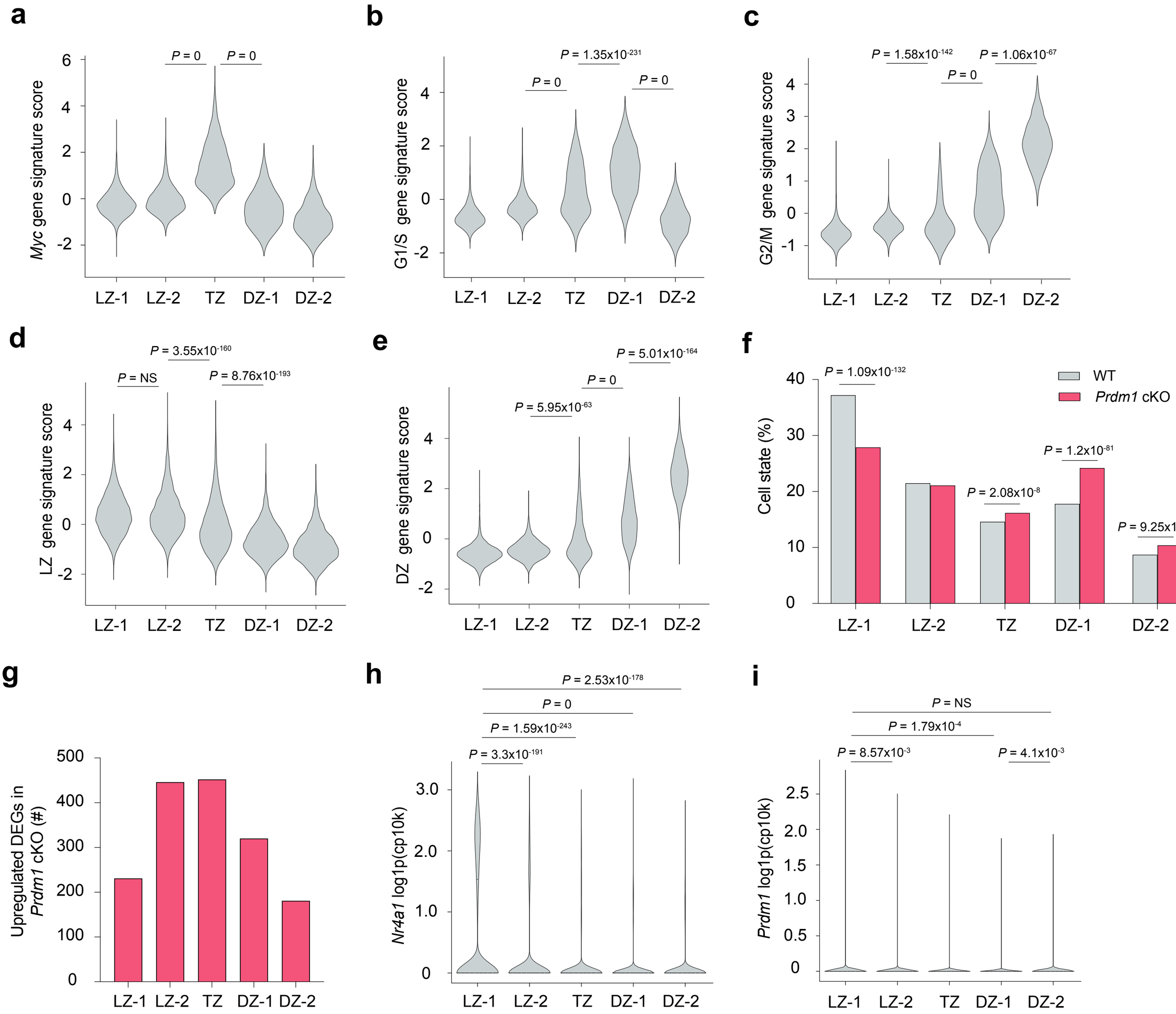

a
LZ-1
LZ-2
TZ
DZ-1
DZ-2
Gpr183
Cd55
Cd38
Cd83
H2-DMa
Nfkbia
Npm1
Nop58
Fbl
Eif5a
Hist1h1d
Pcna
Mcm7
Cenpf
Ccnb1
Cdk1
−1.5 −1.0 −0.5 0.0 0.5 1.0 1.5
Differential Expression (log2)
b
TZ (marker genes)
Ribosome Biogenesis
Nop56 Mrto4 Fbl Npm1
RNA Methylation
Mettl1 Nsun2
Telomere Maintenance
Hsp90ab1 Gar1
GO (p values)
c
TZ (marker genes)
MYC 19030024 ChIP
Fbl Eif5a Hsp90ab1 Mettl1 Apex1
E2F1 18555785 ChIP
Shmt2 Nolc1 Hspd1
ETS1 20019798 ChIP
Ddx18 Nsun2 Eif3b
ChEA (p values)
d
GZ gene signature score
P = 0
P = 0
LZ-1
LZ-2
TZ
DZ-1
DZ-2
e
LZ-2
TZ
DZ-1
LZ-1
DZ-2
f
LZ-2 (DEGs)
Signal Transduction
Camk2d Plcg2 Csk Stat1 Ptprj
DNA Synthesis
Rfc1 Pold4 Mcm4
BCR signaling
Blnk Lyn Itpr2 Grb2
Translation
Eif4a1 Eef2 Rps27l
Reactome (p values)
g
TZ (DEGs)
DNA Replication
Lig1 Rfc4 Mcm3
Cell Cycle
Ccnd3 Cdc6 Cdk4
G1 S Transition
Top2a Pcna Mcm7
Translation
Eif3l Eef1a1 Rplp0
Reactome (p values)
h
LZ-2 (DEGs)
MYC 19030024 ChIP
Ldha Ccnd3 Ptma Ube2c Tpi1 Gclm
SPI1 22096565 ChIP
Blnk Plcg2 Dtx1
IRF8 22096565 ChIP
Stat1 Gpr18
ChEA (p values)
i
TZ (DEGs)
MYC 19030024 ChIP
Tfdp1 Hmga1 Hells Uhrf1 Hsp90aa1
SPI1 22096565 ChIP
Eif4a1 Bub3 Pcna
IRF8 22096565 ChIP
Dusp2 B2m
ChEA (p values)
j
G1/S signature score
P = 2.26x10-72
P = 2.24x10-8
LZ-1
LZ-2
TZ
DZ-1
DZ-2
WT
Prdm1 cKO
k
Nr4a1 log1p(cp10k)
P = 0
P = 0
P = 0
P = 0
LZ-1
LZ-2
TZ
DZ-1
DZ-2
l
Prdm1 log1p(cp10k)
P = 1.46x10-6
P = 3.52x10-4
P = 8.77x10-8
P = NS
LZ-1
LZ-2
TZ
DZ-1
DZ-2
m
Nr4a1 log1p(cp10k)
Spearman's ρ=0.74;
P = 2.45x10-10
Prdm1 log1p(cp10k)

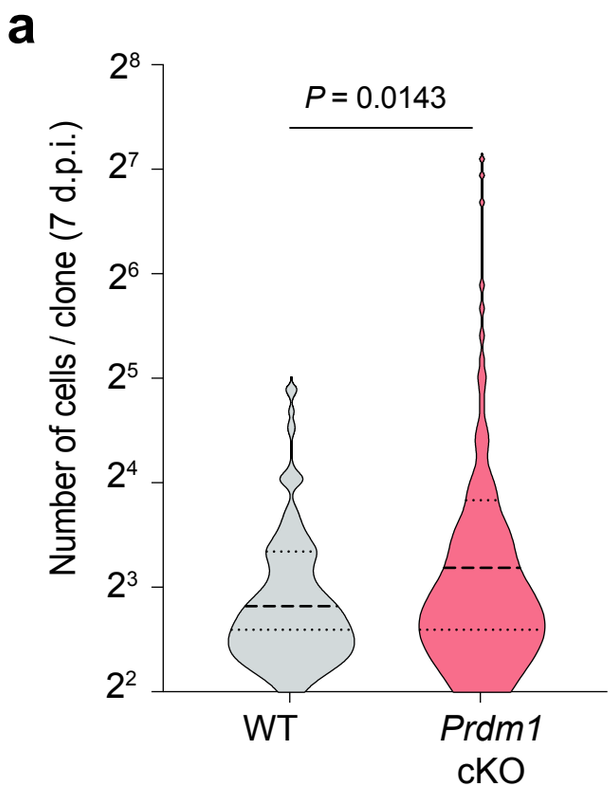
Fig. 2. PRDM1 constrains the GC B cell response by antagonizing the LZ to DZ transition. a, scRNA-seq analysis of splenic $NP^{+}B220^{+}$ cells isolated at 14 d.p.i. from WT mice (n=3). Heatmap was generated using cluster-specific marker genes delineated by the MarkerFinder algorithm in AltAnalyze **(Methods).** Columns in heatmap represent cells (n=28,809); rows represent MarkerFinder genes (n=300). Select MarkerFinder genes for each cell cluster are displayed on right of Heatmap. Cell clusters were generated using ICGS2 within AltAnalyze and annotated with their inferred genomic states. LZ, TZ and DZ refer to light zone, transitional zone and dark zone GC B cell states, respectively. **b,c,** Bar plots displaying key GO pathways **(b)** and TF target genes from the ChEA datasets **(c)** associated with the top 60 genes in TZ state. **d,** Violin plot displaying the GZ signature scores for the indicated cell clusters. **e,** Developmental trajectory of GC B cells constructed by scVelo on 2-D UMAP space **(Methods). f-i,** Bar plots displaying key Reactome pathways and TF target genes from the ChEA datasets associated with the upregulated DEGs (*Prdm1* cKO versus WT) in LZ-2 and TZ states. **j,** Plot displaying the G1/S signature scores across indicated cell clusters in WT and *Prdm1* cKO mice (14 d.p.i., n=3) **(Methods)**. **k,l,** Violin plots displaying log1p(cp10k)-normalized expression of *Nr4a1* **(k)** and *Prdm1* **(l)** in non-zero expressing cells across the indicated cell clusters. **m,** Scatter plot displaying *Prdm1* versus *Nr4a1* expression in non-zero expressing cells. Spearman's correlation and two-tailed p values are indicated. Each dot (**e**) represents an individual cell. Selected genes within pathways are displayed (**b, c, f-i**). Data in panels **a-l** are from three different experiments. Statistical significance was tested by Fisher's Exact test and Benjamini-Hochberg method for correction of multiple-hypotheses testing (**b, c, f-i**), Mann-Whitney test (**j**) and Kruskal-Wallis with Dunn's multiple comparison test (**k, l**).

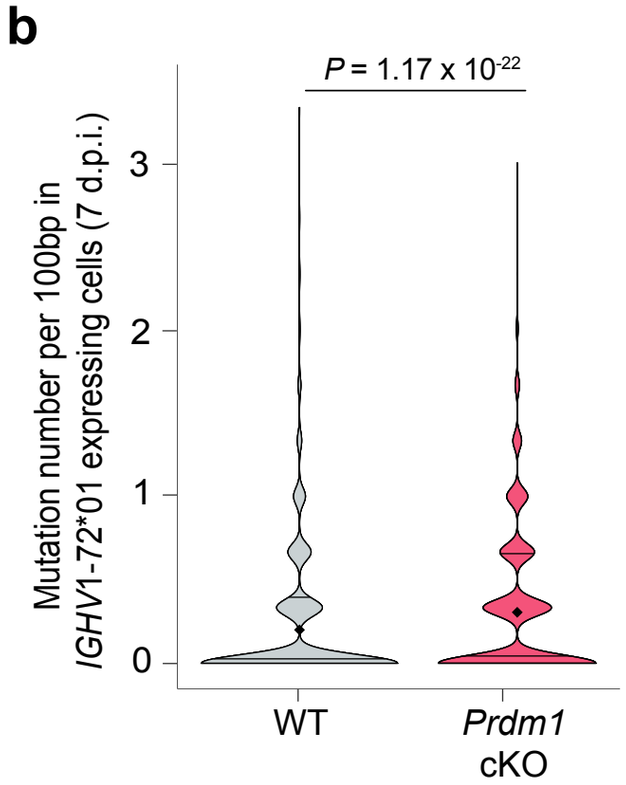
Extended Data Fig. 2

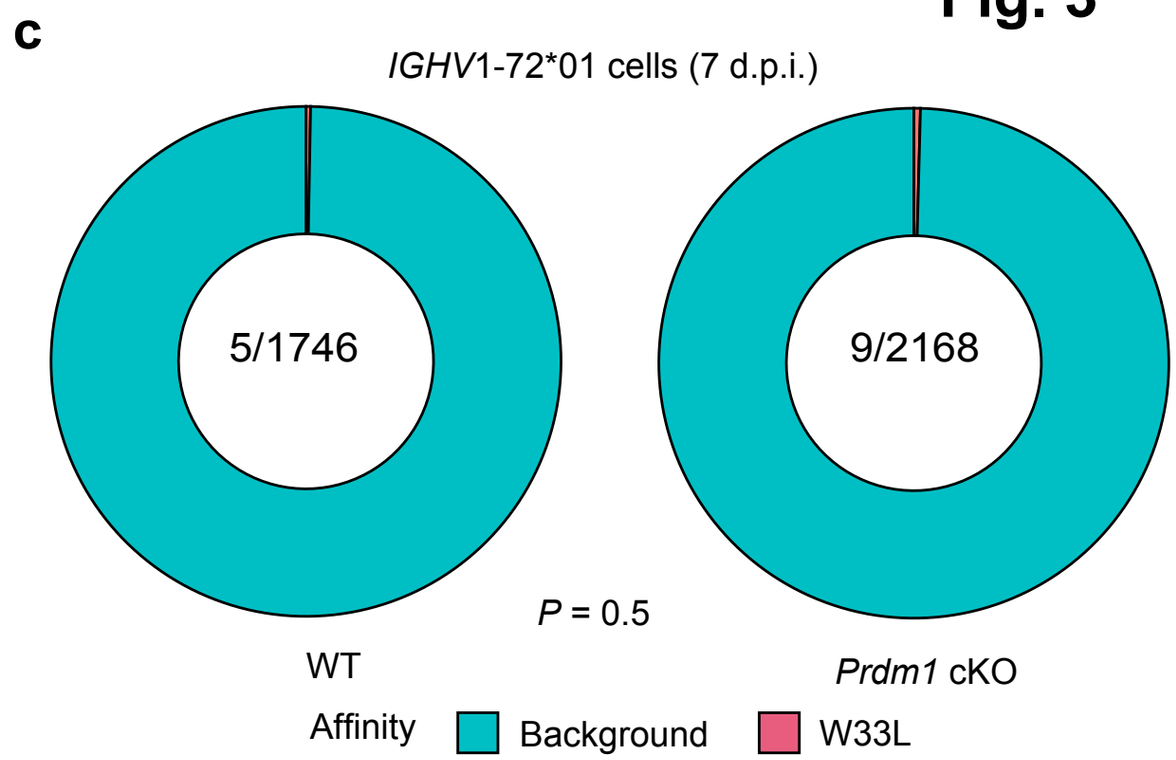


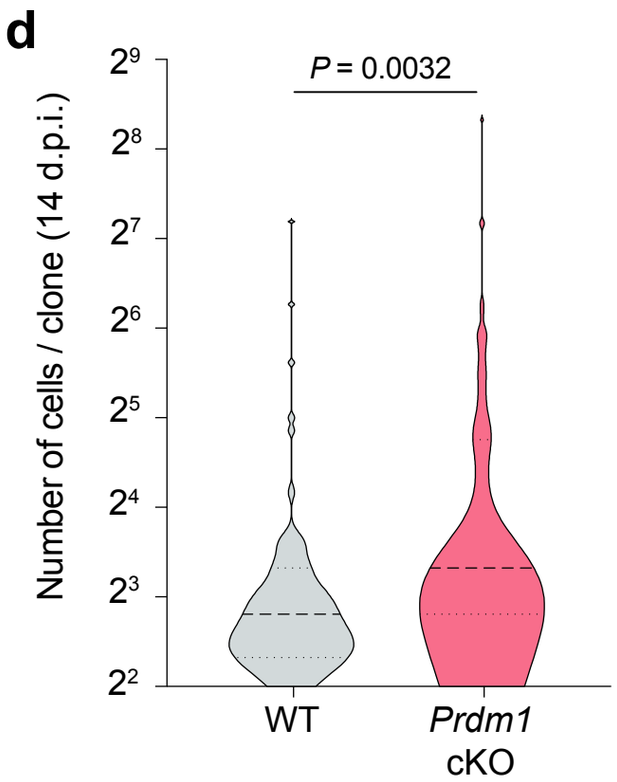
Extended Data Fig. 2. PRDM1 constrains the GC B cell response by antagonizing the LZ to DZ transition. a-e, Violin plots displaying the *Myc* target gene, G1/S, G2M, LZ, and DZ signature scores for the indicated cell clusters. **f,** Bar plot displaying the frequencies of indicated cell clusters in WT and *Prdm1* cKO NP-specific GC B cells by clustering of scRNA-seq data using ICGS2 referenced clusters and cellHarmony. **g,** Plot displaying the number of upregulated DEGs across indicated cell clusters in *Prdm1* cKO compared to WT. **h,i** Violin plots displaying log1p(cp10k)-normalized expression of *Nr4a1* **(h)** and *Prdm1* **(i)** in all cells across the indicated cell clusters. Data in panels **a-i** are from three independent experiments. Statistical significance was tested by Kruskal-Wallis with Dunn's multiple comparison test (**a-e, h-i**) and Fisher's Exact test (**f**).

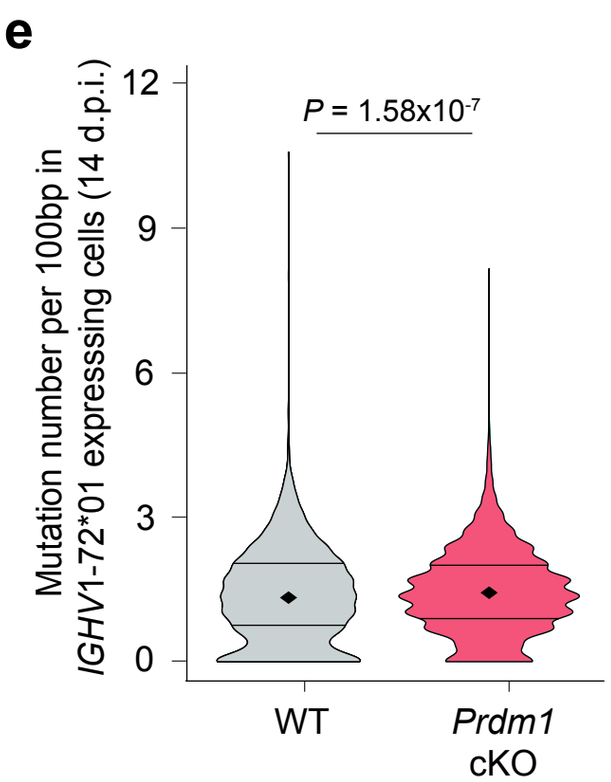
Fig. 3

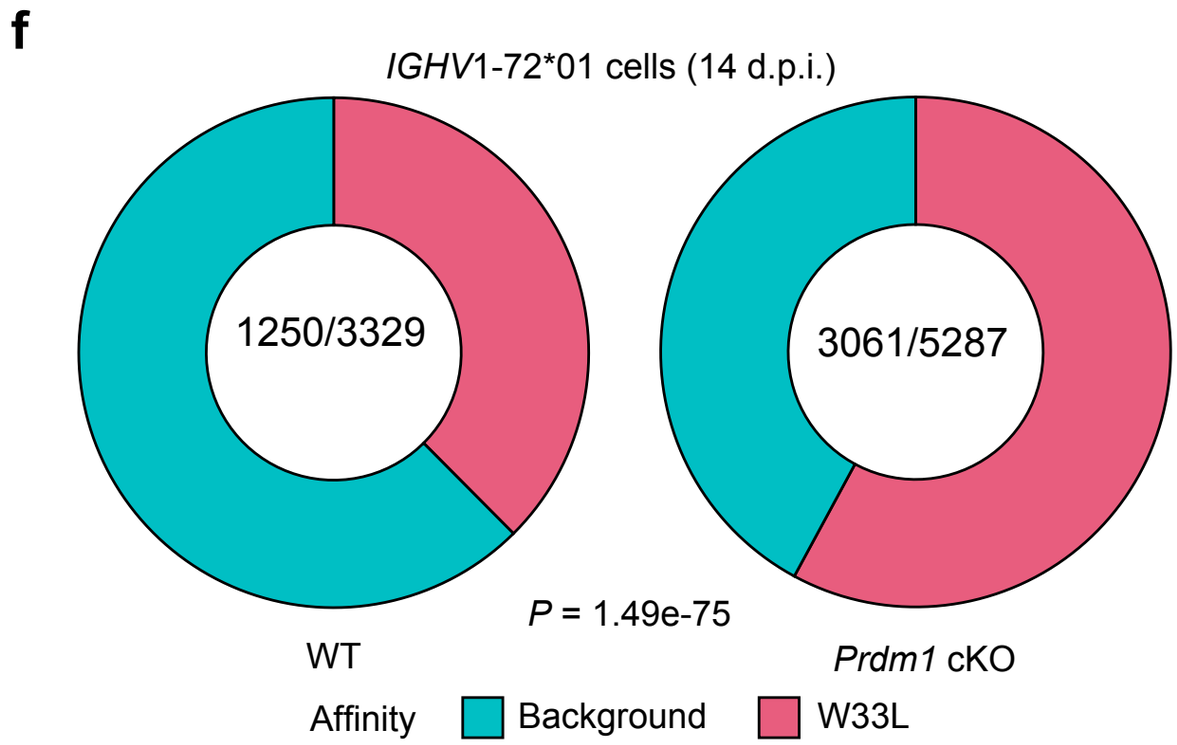

a
P = 0.0143
Number of cells / clone (7 d.p.i.)
WT
Prdm1 cKO

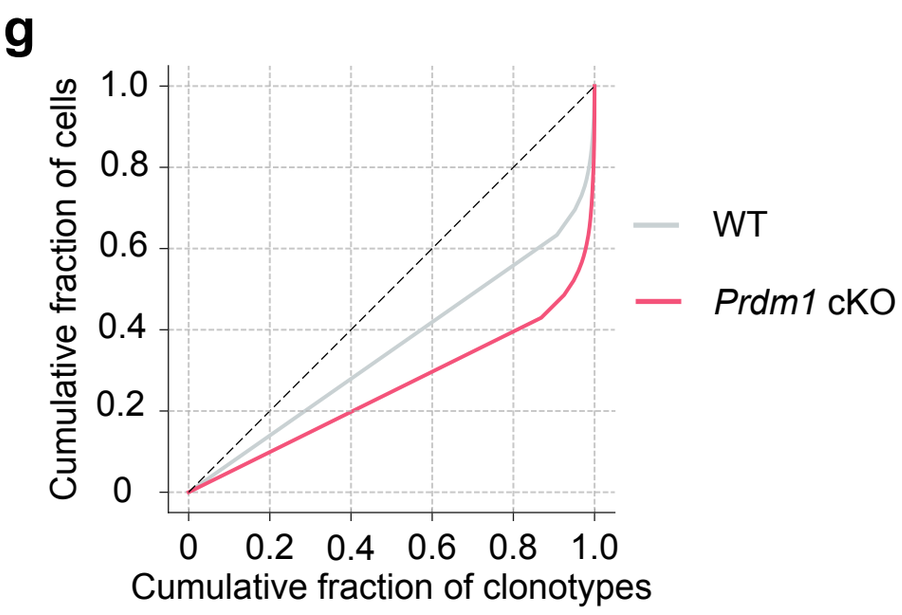

b
P = 1.17 x 10-22
Mutation number per 100bp in IGHV1-72*01 expressing cells (7 d.p.i.)
WT
Prdm1 cKO

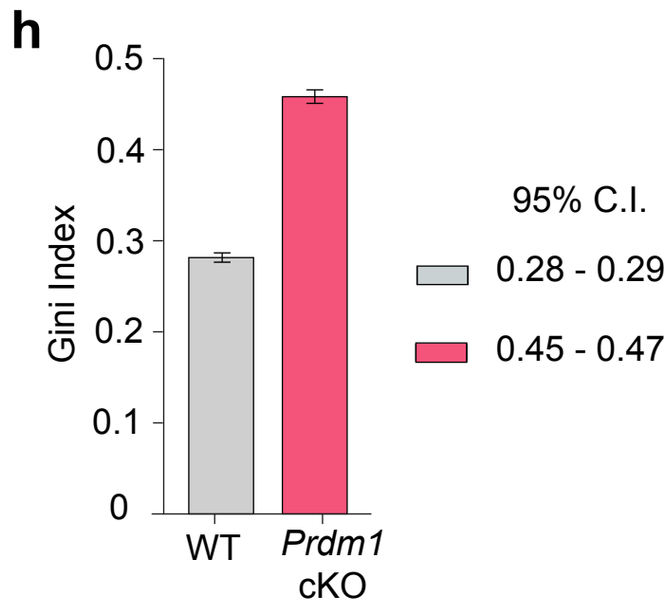

c
IGHV1-72*01 cells (7 d.p.i.)
5/1746
9/2168
P = 0.5
WT
Prdm1 cKO
Affinity
Background
W33L


d
P = 0.0032
Number of cells / clone (14 d.p.i.)
WT
Prdm1 cKO


e
P = 1.58x10-7
Mutation number per 100bp in IGHV1-72*01 expresssing cells (14 d.p.i.)
WT
Prdm1 cKO


f
IGHV1-72*01 cells (14 d.p.i.)
1250/3329
3061/5287
P = 1.49e-75
WT
Prdm1 cKO
Affinity
Background
W33L


g
Cumulative fraction of cells
Cumulative fraction of clonotypes
WT
Prdm1 cKO


h
Gini Index
95% C.I.
0.28 - 0.29
0.45 - 0.47
WT
Prdm1 cKO

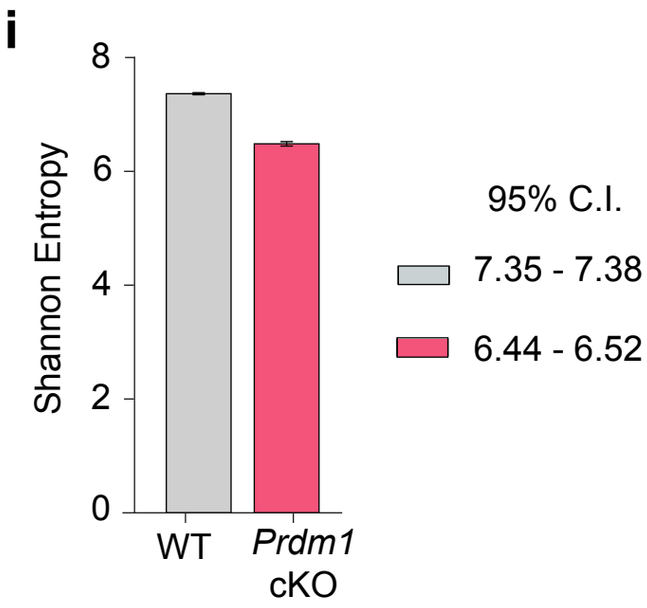

i
Shannon Entropy
95% C.I.
7.35 - 7.38
6.44 - 6.52
WT
Prdm1 cKO

Fig. 3. PRDM1 constrains affinity maturation and clonal dominance. a, scBCR-seq analysis for the numbers of cells in distinct *IGHV1-72*01* clones in WT and *Prdm1* cKO mice (7 d.p.i.). **b,** scBCR-seq analysis for the mutation frequencies within *IGHV1-72*01* gene segments of GC B cells in WT and *Prdm1* cKO mice (7 d.p.i.). **c,** Pie-charts displaying the proportion of *IGHV1-72*01* gene segments harboring high affinity W33L mutations in splenic NP^+B220^+ cells of indicated genotypes (7 d.p.i.). **d,** scBCR-seq analysis for the numbers of cells in distinct *IGHV1-72*01* clones in WT and *Prdm1* cKO mice (14 d.p.i.). **e,** scBCR-seq analysis for the mutation frequencies within *IGHV1-72*01* gene segments of GC B cells in WT and *Prdm1* cKO mice (14 d.p.i.). **f,** Pie-charts displaying the proportion of *IGHV1-72*01* gene segments harboring high affinity W33L mutations in splenic NP^+B220^+ cells of indicated genotypes (14 d.p.i.). **g-i,** Clonal diversity of BCR sequences in WT and *Prdm1* cKO GC B cells. **g,** Lorenz plots displaying the clonal distribution, skewed curves reflect clonal diversity. **h,i,** Bar plots displaying the Gini Index values **(h)** and the Shannon Entropy scores **(i)** for the BCR repertoires, respectively. For comparison across samples, data was rarefied to a common paired-cell depth per sample and computed with 95% confidence intervals. Data in panels **d-i** are from three different experiments. Statistical significance was tested by Mann-Whitney test (**a,b,d,e**) and Chi square test (**c,f**).

Fig. 4

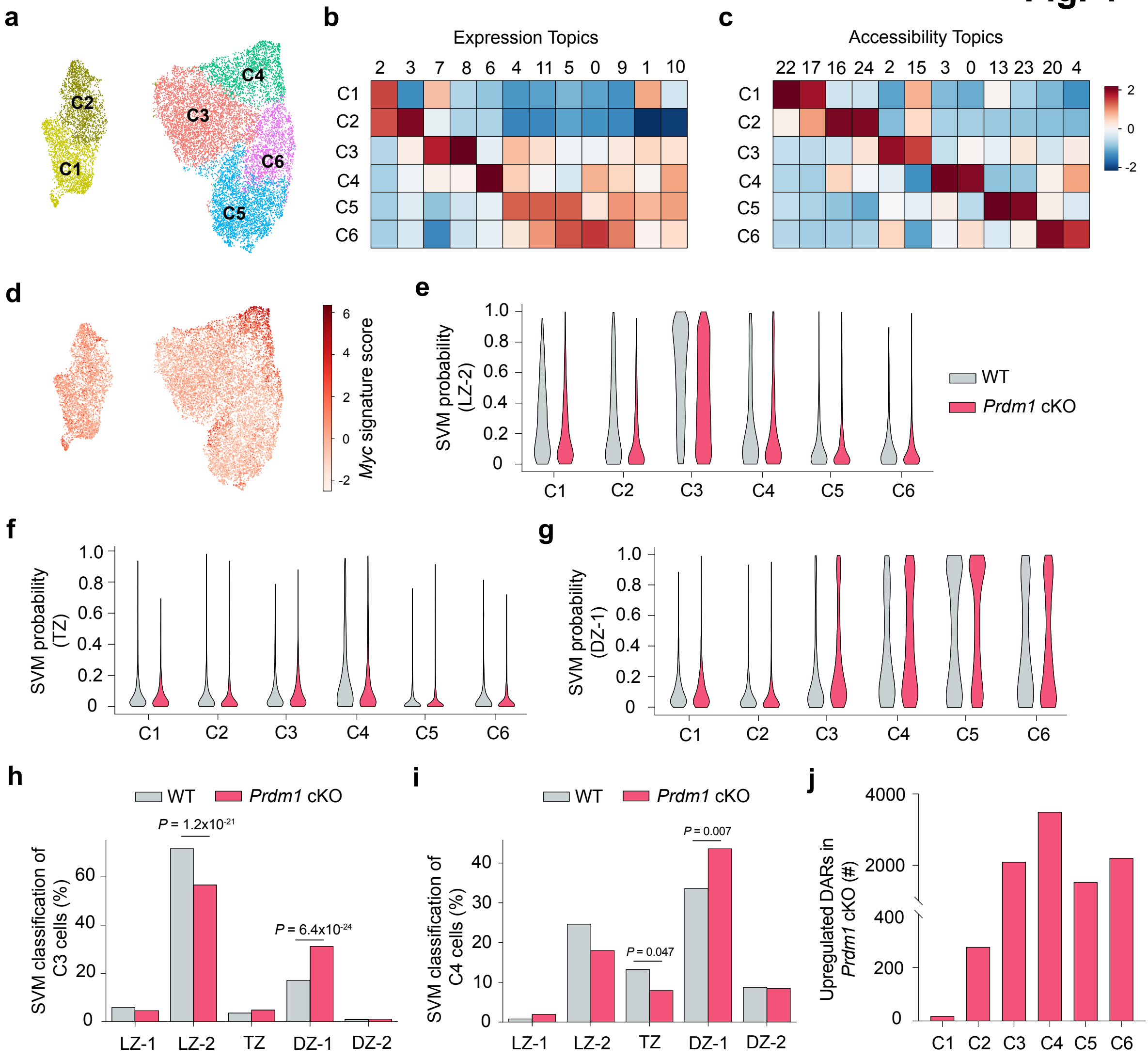
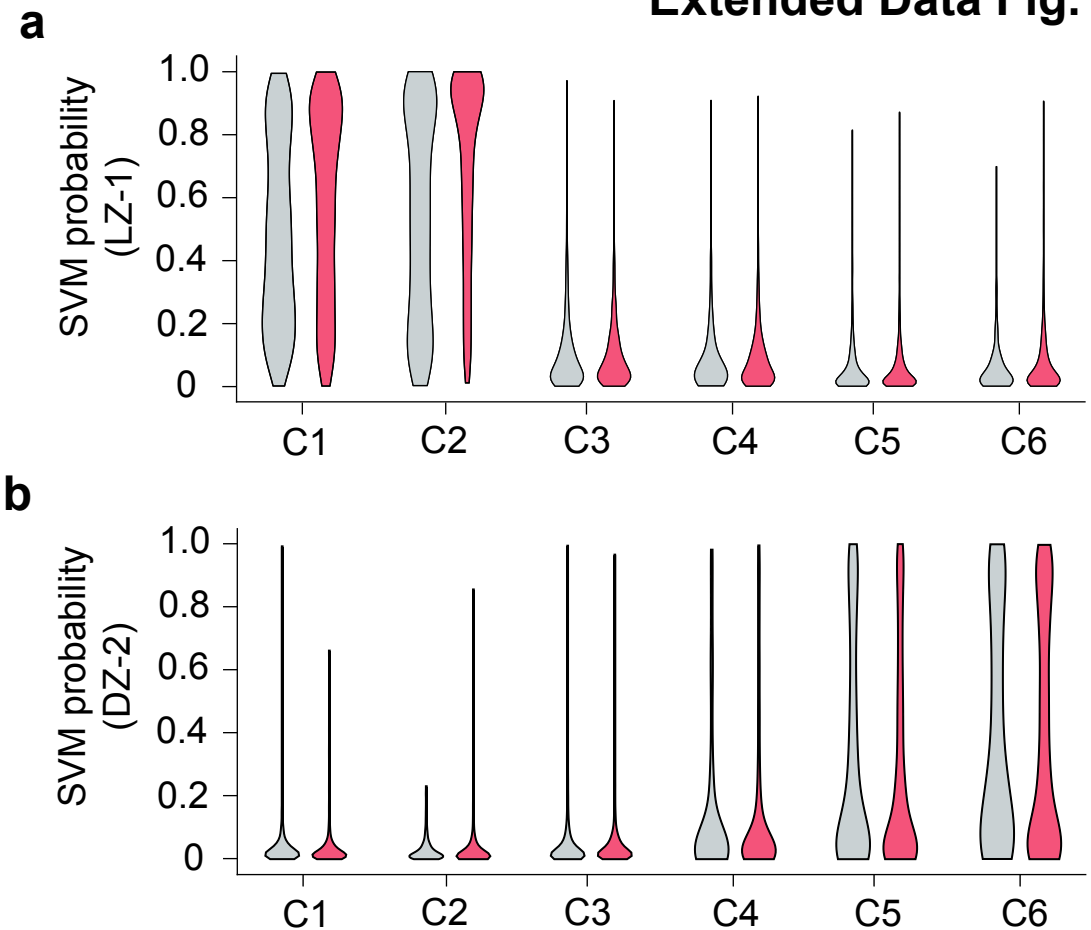

a
C1
C2
C3
C4
C5
C6
b
Expression Topics
2 3 7 8 6 4 11 5 0 9 1 10
c
Accessibility Topics
22 17 16 24 2 15 3 0 13 23 20 4
2
0
-2
d
6
4
2
0
-2
Myc signature score
e
SVM probability (LZ-2)
WT
Prdm1 cKO
f
SVM probability (TZ)
g
SVM probability (DZ-1)
h
WT
Prdm1 cKO
P = 1.2x10-21
P = 6.4x10-24
SVM classification of C3 cells (%)
LZ-1 LZ-2 TZ DZ-1 DZ-2
i
WT
Prdm1 cKO
P = 0.007
P = 0.047
SVM classification of C4 cells (%)
LZ-1 LZ-2 TZ DZ-1 DZ-2
j
Upregulated DARs in Prdm1 cKO (#)
4000
2000
400
200
0
C1 C2 C3 C4 C5 C6

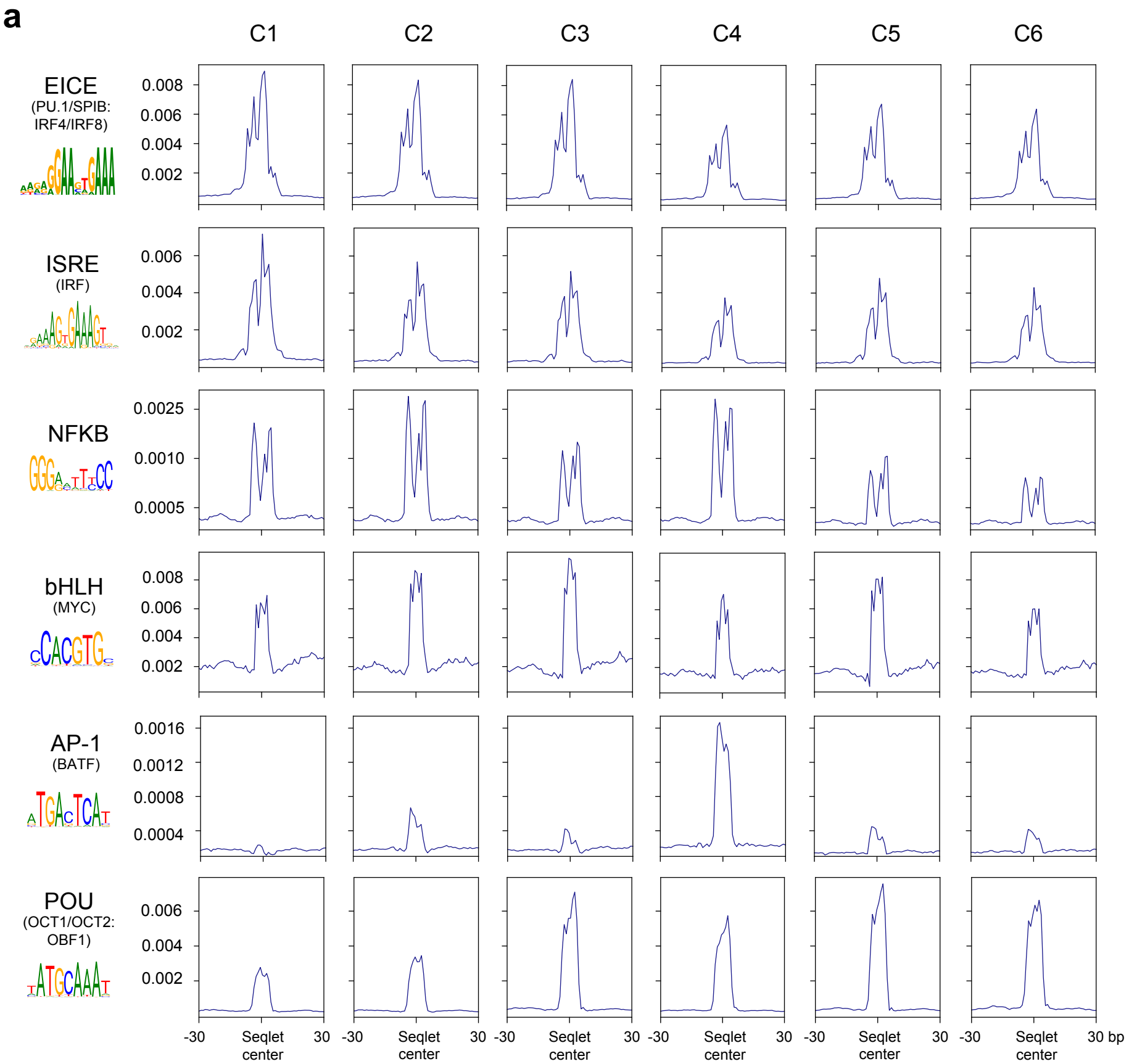
Fig. 4. Multiome joint topic modelling reveals transcription and chromatin features underlying enhanced LZ-DZ transition in absence of PRDM1. a, NP^+B220^+ cells isolated at 14 d.p.i. from WT and *Prdm1* cKO mice were profiled by scMultiome-seq (n=15,155). UMAP plot displaying 6 clusters (C1-6) was generated using MIRA and joint topic modelling of both RNA and ATAC features (WT=4,764 and *Prdm1* cKO=10,391). **b,c** Heatmaps showing relative prominence (z-score of average weight) of each gene expression **(b)** or accessibility **(c)** topic within designated cell clusters **(4a)**. **d,** UMAP displaying the *Myc* gene signature score across the multiome-seq clusters. **e-g,** Violin plots displaying the distribution of indicated ICGS2-delineated transcriptional states (y-axis) across the six scMultiome-seq defined clusters (x-axis) in WT and *Prdm1* cKO GC B cells. Probabilities were determined by SVM classification. **h,i** Bar plots displaying the frequency of WT and *Prdm1* cKO GC B cells within the C3 **(h)** and C4 **(i)** clusters that were assigned to indicated ICGS2-transcriptional states, as determined by SVM classification. **j,** Bar plot displaying the number of upregulated DARs for indicated scMultiome-seq clusters in *Prdm1* cKO compared with WT counterparts (14 d.p.i.). Statistical significance was tested by Fisher's Exact test 2x2 contingency table (**h,i**).

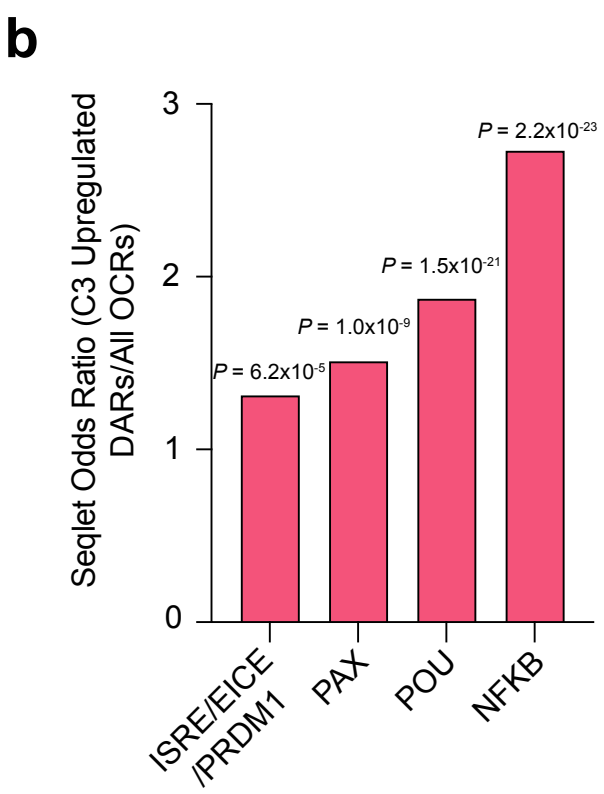
Extended Data Fig. 3



Extended Data Fig. 3. Multiome joint topic modelling reveals transcription and chromatin features underlying enhanced LZ-DZ transition in absence of PRDM1. a,b Violin plots displaying the distribution of indicated ICGS2-delineated transcriptional states (y-axis) across the six scMultiome-seq defined clusters (x-axis) in *Prdm1* cKO GC B cells. Probabilities were determined by SVM classification.

Fig. 5


a
C1
C2
C3
C4
C5
C6
EICE
(PU.1/SPIB:
IRF4/IRF8)
ISRE
(IRF)
NFKB
bHLH
(MYC)
AP-1
(BATF)
POU
(OCT1/OCT2:
OBF1)
0.008
0.006
0.004
0.002
0.0025
0.0010
0.0005
0.0016
0.0012
0.0008
0.0004
-30
Seqlet
center
30
30 bp


b
Seqlet Odds Ratio (C3 Upregulated
DARs/All OCRs)
3
2
1
0
P = 6.2x10-5
P = 1.0x10-9
P = 1.5x10-21
P = 2.2x10-23
ISRE/EICE
/PRDM1
PAX
POU
NFKB

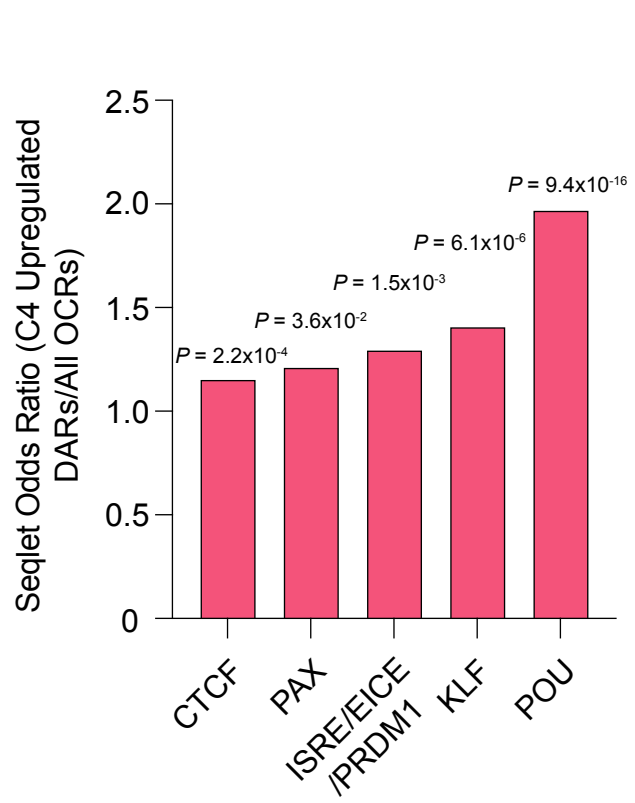

c
Seqlet Odds Ratio (C4 Upregulated
DARs/All OCRs)
2.5
2.0
1.5
1.0
0.5
0
P = 2.2x10-4
P = 3.6x10-2
P = 1.5x10-3
P = 6.1x10-6
P = 9.4x10-16
CTCF
PAX
ISRE/EICE
/PRDM1
KLF
POU

Fig. 5. Loss of PRDM1 enhances chromatin accessibility at PAX, POU and NFKB actionable regions during LZ to DZ transition. a, Aggregated ChromBPNet-derived CS signals at selected TF motifs for indicated GC B cell clusters within all peaks. Motif instances were identified by HOMER. Position weight matrix logos are displayed. **b,c** Bar plots displaying the relative increase in indicated seqlets predicted by ChromBPNet within the upregulated DARs in the C3 **(b)** and C4 **(c)** cluster of *Prdm1* cKO GC B cells. Statistical significance was tested by Fisher's Exact test and Benjamini-Hochberg method for correction of multiple-hypotheses testing (**b,c**).

Fig. 6

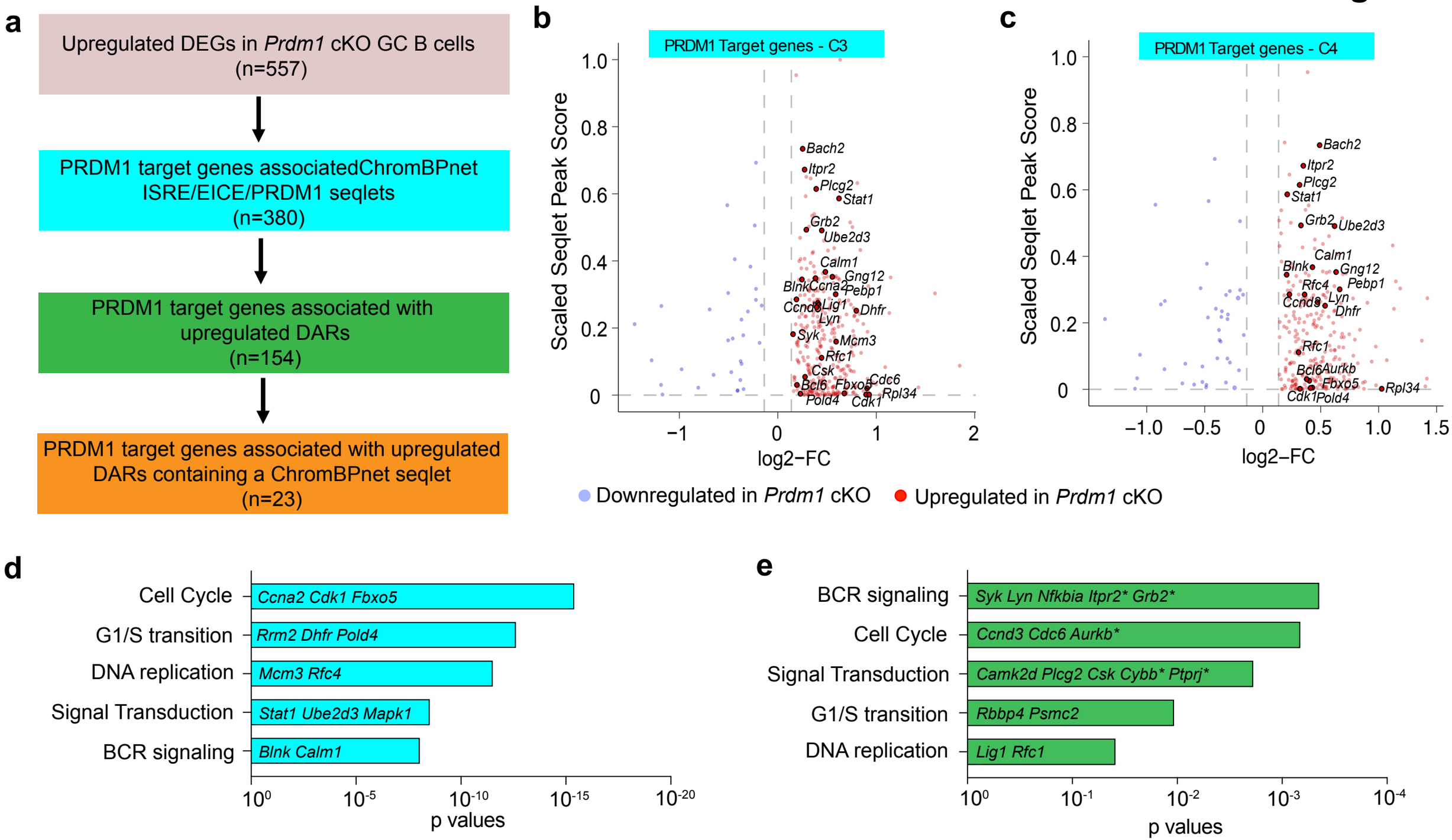
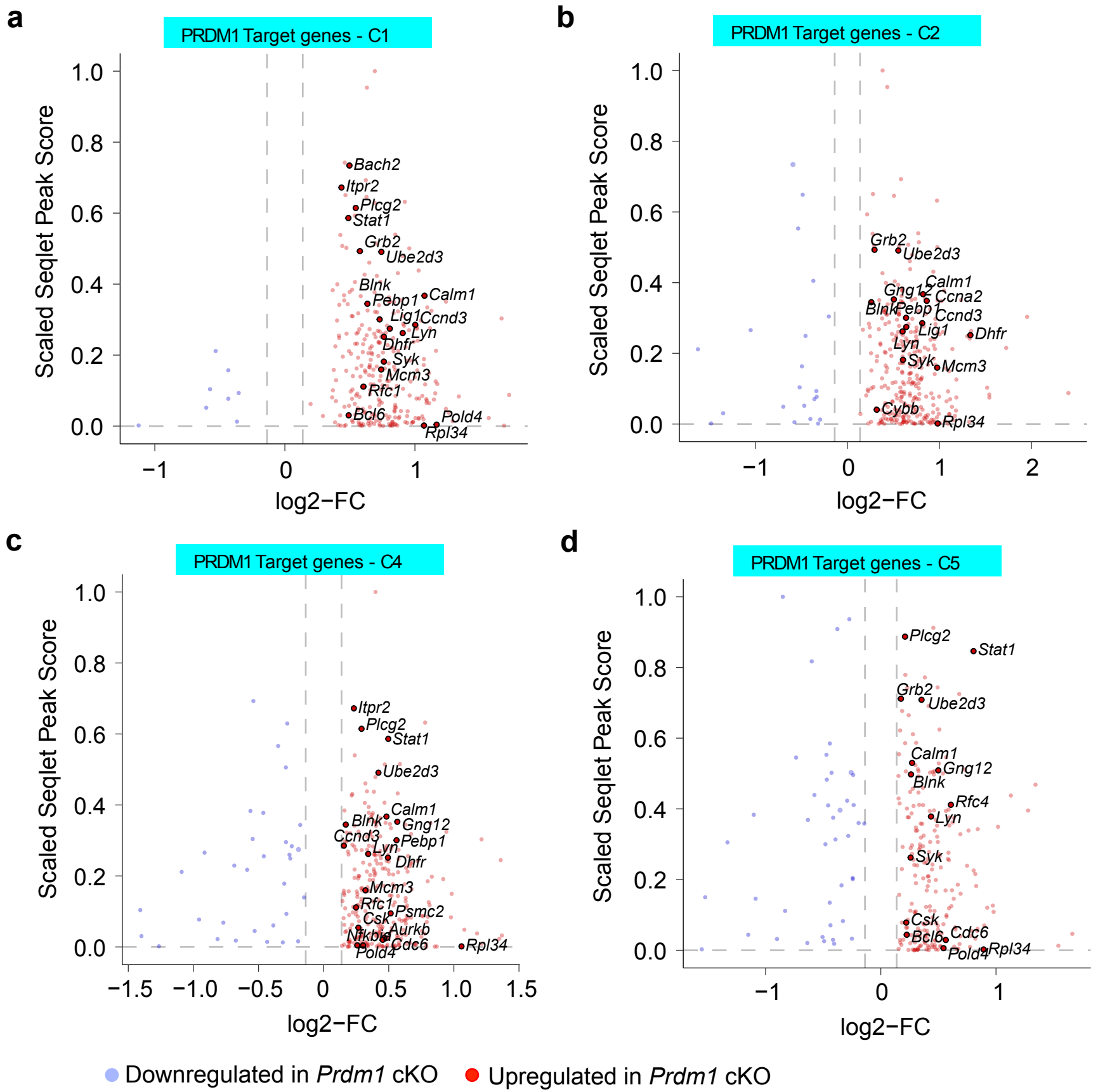

a
Upregulated DEGs in Prdm1 cKO GC B cells (n=557)
PRDM1 target genes associatedChromBPnet ISRE/EICE/PRDM1 seqlets (n=380)
PRDM1 target genes associated with upregulated DARs (n=154)
PRDM1 target genes associated with upregulated DARs containing a ChromBPnet seqlet (n=23)
b
PRDM1 Target genes - C3
Scaled Seqlet Peak Score
log2−FC
Bach2
Itpr2
Plcg2
Stat1
Grb2
Ube2d3
Calm1
Gng12
Blnk
Ccna2
Pebp1
Ccnd3
Lig1
Dhfr
Lyn
Syk
Mcm3
Rfc1
Csk
Bcl6
Fbxo5
Cdc6
Pold4
Cdk1
Rpl34
c
PRDM1 Target genes - C4
Scaled Seqlet Peak Score
log2−FC
Bach2
Itpr2
Plcg2
Stat1
Grb2
Ube2d3
Calm1
Blnk
Gng12
Rfc4
Pebp1
Ccnd3
Lyn
Dhfr
Rfc1
Bcl6
Aurkb
Fbxo5
Rpl34
Cdk1
Pold4
Downregulated in Prdm1 cKO
Upregulated in Prdm1 cKO
d
Cell Cycle
Ccna2 Cdk1 Fbxo5
G1/S transition
Rrm2 Dhfr Pold4
DNA replication
Mcm3 Rfc4
Signal Transduction
Stat1 Ube2d3 Mapk1
BCR signaling
Blnk Calm1
p values
e
BCR signaling
Syk Lyn Nfkbia Itpr2* Grb2*
Cell Cycle
Ccnd3 Cdc6 Aurkb*
Signal Transduction
Camk2d Plcg2 Csk Cybb* Ptprj*
G1/S transition
Rbbp4 Psmc2
DNA replication
Lig1 Rfc1
p values

Fig. 6. ChromBPNet modeling predicts PRDM1 mediated repression of BCR signaling genes. **a,** Schematic rationale for identification of PRDM1 target genes within upregulated DEGs in *Prdm1* cKO GC B cell clusters. Shared DEGS in scRNA-seq and multiome datasets (n=557) were used for downstream analysis. Color codes indicate target genes based on specified criteria. PRDM1 target genes were associated (TSS +/- 200kb) with one or more ISRE, EICE and PRDM1 motif seqlets based on ChromBPNet models. These were annotated for association with one or more upDARs and ISRE, EICE and/or PRDM1 motif seqlets located within them. Numbers in parentheses denote the union of upregulated genes across the six clusters at each filtering step. **b,c,** Volcano plots displaying the PRDM1 target genes for the indicated clusters. Selected PRDM1 target genes are displayed. **d,e,** Bar plots displaying key Reactome pathways associated with the PRDM1 target genes identified by ChromBPNet seqlets **(d)** and ChromBPNet seqlets associated with one or more DARs **(e)**. Selected genes from the pathway enrichment analysis are highlighted. Asterisks indicate PRDM1 target genes containing ISRE, EICE and/or PRDM1 ChromBPNet seqlets within an associated upregulated DAR. Statistical significance was tested by Fisher's Exact test and Benjamini-Hochberg method for correction of multiple-hypotheses testing (**d,e**).

Extended Data Fig. 4



Extended Data Fig. 4. ChromBPNet modeling predicts PRDM1 mediated repression of BCR signaling genes. **a-d,** Volcano plots displaying the PRDM1 target genes for the indicated clusters as in Fig. **4b,c**. Selected PRDM1 target genes are displayed.

Methods

Mice. *Prdm1*$^{fl/fl}$ (008100) and *Cd19*$^{cre/cre}$ (006785) mice were purchased from Jackson laboratory and were intercrossed to generate *Prdm1*$^{fl/fl}$*Cd19*$^{cre/+}$ (*Prdm1* cKO) and *Prdm1*$^{+/+}$ *Cd19*$^{cre/+}$ (WT) littermates. Mice were housed under specific pathogen-free conditions (20–26 °C; 30–70% relative humidity; 12-h light/dark cycle). All procedures complied with University of Pittsburgh IACUC protocols 21059197 and 24054803.

Immunization. Eight to ten week old *Prdm1* cKO and WT mice were immunized intraperitoneally with 50 μg NP(23)-CGG (Biosearch Technologies) mixed with 50% (v/v) imject alum (Thermo Fisher Scientific) and 1 μg LPS (Sigma).

Flow cytometry. Spleens were dissociated to single-cell suspensions in PBS (Ca^{2+}/Mg^{2+}-free) + 2% FBS + 2 mM EDTA (FACS buffer). Erythrocytes were lysed in ACK lysis buffer (Thermo Fisher) for 5 min at RT, quenched in FACS buffer, and washed. After a brief PBS wash to remove serum proteins, cells were stained with Zombie NIR (BioLegend) and anti-CD16/32 (2.4G2, 25 μg/mL) for 10 min on ice, followed by fluorochrome-conjugated antibodies (listed in Supplementary Table 12) for 20 min on ice. Standard doublet exclusion (FSC-A/H, SSC-A/H), FMO/compensation controls, and NP binding controls (unconjugated CGG; NP-negative samples) were used. Cells were fixed in BD Cytofix (#554655) for 20 min at 4 °C, washed, and resuspended in PBS. Data were acquired on a Cytek Aurora and analyzed in FlowJo (Tree Star). GC B cells were gated as B220^{+} GL-7^{+} Fas^{+}, with NP^{+} subsets as indicated.

Isolation of NP^{+}B220^{+} cells. For sorting, splenocytes (7 and 14 d.p.i.) were stained with eFluor780 live/dead, B220 (RA3-6B2; PE-Dazzle 594), and NP–PE (NP(30)–PE) for 20 min at 4 °C and sorted as NP^{+}B220^{+} on a FACSAria II (70 μm nozzle; 4 °C). Post-sort purity were above 95%.

scRNA-seq and BCR-seq. Sorted NP^{+}B220^{+} cells were suspended in 5% FBS and loaded on a Chromium Controller (10x Genomics; Single Cell 5′ v2). Libraries were prepared as per the manufacturer's protocol as described previously[1]. From the cDNA, one aliquot was used for gene-expression library prep; a second was amplified for VDJ using the BCR kit. Libraries were sequenced on an Illumina NovaSeq 6000 with manufacturer recommended read

structures for 5′ GEX and VDJ. Target depths were ~30k reads/cell (GEX) and ~5k reads/cell (VDJ).

scRNA-seq data processing. FASTQs were processed with Cell Ranger v6.1.2 using the mm10 reference. UMI count matrices from WT 14 d.p.i. NP^+B220^+ cells (HDF5) were analyzed with AltAnalyze ICGS2 (default parameters) to delineate clusters and marker genes[2]. Initial cell state labels were assigned by ICGS2/GO-Elite enrichment and refined by literature markers. Non-B-lineage contaminants were removed. All scRNA-seq data are in GEO: GSE305877.

scRNA-Seq annotation and integration. WT clusters served as reference states for supervised alignment of *Prdm1* cKO cells using cellHarmony[3] (centroid-based label transfer; Pearson $r \geq 0.7$ on $\log_2$-scaled counts). Fisher's exact test was performed to assess differential cell type frequency.

RNA velocity analyses. Velocyto v0.17.17[4] was used to generate loom files with spliced/unspliced counts for LZ, TZ, and DZ clusters. Pre-processing and dynamical modeling used scVelo v0.3.3[5]. Scanpy v1.11.4[6] with default settings was used to generate velocity streams over the UMAP manifold.

Gene Expression Analyses. Expression values for *Prdm1* and *Nr4a1* were normalized to counts-per-10k and log1p-transformed "expressing" cells were defined from the counts layer as >0. Analyses were stratified by GC states (LZ-1, LZ-2, TZ, DZ-1, DZ-2) and visualized with violin plots.

Differential gene expression analyses. For scRNA-seq datasets, DEG analysis was performed using cellHarmony between the reference (WT) and the query (*Prdm1* cKO) datasets. Genes with an absolute fold ≥ 1.1 ($\log_2$ fold change ≥ 0.13) and empirical Bayes t-test p-value < 0.05 (Benjamini Hochberg corrected) were considered to be differentially expressed. For multiome-seq data, differential gene expression analysis was performed using Scanpy (v1.10.3) between reference (WT) and query (*Prdm1* cKO) cells for each cluster. Genes with a Benjamini-Hochberg correct pooled-variance Student's t-test P value < 0.05 and absolute $\log_2$ fold change ≥ 0.13 were considered to be differentially expressed.

BCR analysis of 5'-end scRNA-seq. VDJ FASTQs were processed with Cell Ranger VDJ v6.1.2 with GRCm38 mouse reference. BCR and GEX barcodes were linked to couple clonotypes with transcriptomes. NP-specific clones were defined as cells bearing IGHV1-72*01 (VH186.2) rearrangements; cells lacking paired GEX–VDJ were excluded. Clonotypes were defined from the top-UMI productive IGH and IGK/IGL chains using CDR3 amino-acid sequence + V + J genes (alleles collapsed), as described[1,7]. SHM rates and W33L calls were as described[7].

BCR Clonotype Analysis. Within each sample, all clonotypes were enumerated based on their paired IgH and IgL sequences and then cells were grouped by clonotype. Lorenz curves plotted cumulative cells versus cumulative clonotypes. To equalize depth for estimation of Gini coefficients and Shannon entropy scores, clone sizes were rarefied to the minimum paired-cell count across samples (subsampling without replacement). For each replicate, the Gini coefficient[8] and Shannon entropy scores[9] were computed from clone counts and summarized per-sample rarefied means with 95% CIs (2.5–97.5th percentiles of replicate distributions). Per-genotype values are the mean of experiment-level rarefied Gini indices and Shannon entropy scores, weighted equally across experiments.

scMultiome-seq. Sorted NP^+B220^+ splenocytes were washed twice with PBS containing 0.04% BSA and lysed in 100 µL of chilled lysis buffer [10 mM Tris-HCl, 10 mM NaCl, 3 mM $MgCl_2$, 0.1% Tween-20, 0.1% IGEPAL, 0.01% digitonin, 1% BSA, 1 mM DTT, 1 U/µL RNase inhibitor (Protector, Sigma)]. Cells were mixed by pipetting and incubated on ice for 3 min. Lysis was quenched with 1 mL of chilled wash buffer (same composition as lysis buffer excluding digitonin) and mixed gently. Nuclei were washed three times with wash buffer and resuspended in diluted nuclei buffer (10x Genomics) supplemented with 1 mM DTT and 1 U/µL RNase inhibitor. Nuclei were counted, resuspended in chilled PBS (750–1,000 nuclei/µL), and aliquoted (180 µL per sample). Labeling reactions were quenched with 2% BSA in PBS, followed by centrifugation at 500 × g for 4 min at 4 °C. After pooling and washing, nuclei were resuspended in diluted nuclei buffer and processed according to the Chromium Next GEM Single Cell Multiome ATAC + Gene Expression User Guide (10x Genomics) through the pre-amplification PCR step. Ten microliters of the pre-amplified product were fractionated using 0.6× SPRI bead cleanup to separate cDNA (gene expression) and gDNA

(ATAC) products. Final libraries were prepared according to the manufacturer's protocol and sequenced on a NovaSeq 6000 (Illumina) obtaining a depth of 30,000 reads for the mRNA and ATAC libraries.

scMultiome data analysis. FASTQ files were processed with Cell Ranger ARC (v2.0.2) using the GRCm38 (2011) reference genome for alignment. Outputs from individual GEM wells were aggregated using the 'aggr' function. Cell and gene filtering were performed in Scanpy (1.11.4), retaining cells with >2000 detected genes and <10% mitochondrial reads, and genes expressed in at least 20 cells. Cell cycle effects were regressed out using the 'regress_out' function in SCANPY with G1/S and G2/M gene sets. MIRA v2.1.1 was used for joint topic modeling of expression and accessibility. A joint embedding integrated topic weights, followed by UMAP and Leiden clustering. Peaks were called with MACS2 v3.0.3 (see below).

SVM based cell state predictions. A multiclass SVM (scikit-learn v1.3.2) was trained on standardized gene-expression features from the scRNA-seq reference states (ICGS2 labels) and applied to the Multiome dataset to infer LZ-1, LZ-2, TZ, DZ-1, DZ-2 probabilities per cell. The maximum-probability class was assigned as the label; probability distributions were retained for statistics[10]. Receiver operating characteristic curves (using Youden's J statistic calculated per Leiden cluster) defined high-confidence probability cutoffs for each state.

MIRA Topic Modeling. MIRA (v2.1.1)[11] topic modeling was applied to both RNA expression and ATAC accessibility data, inferring topic weights for each cell. Weights were averaged per cluster to build a cluster × topic matrix, then each topic was z-score normalized across clusters. The topics showing the highest variance between clusters were selected as cluster specific topics.

ChromBPNet Models and Analyses. Cluster/genotype-specific pseudo-bulk BAMs were generated from single-cell multiome dataset by using Sinto's 'filterbarcodes' function and used to call peaks with MACS2 (v2.2.7.1). These peak sets were merged and used to train cluster/genotype-specific ChromBPNet (v0.1.7) models[12]. A Tn5 bias model was trained by running ChromBPNet 'bias pipeline' with nonpeak sequences from inaccessible chromatin

regions. Bias-corrected BigWig files were generated by running ChromBPNet 'pred_bw' and base-pair resolution contribution score predictions were generated by running ChromBPNet 'contribs_bw' with the top accessibility topics (MIRA) per cluster, and cluster specific upregulated DAR peak sets. The resulting profile scores AnnData objects were used as input into TFModisco 'motifs' function using sampling of 10,000 seqlets and the 'report' function using the Jaspar 2024 Core Vertebrates (non-redundant) PFM collection. Seqlets matching motifs in the same Jaspar motif cluster family were merged for quantification of TF motif family seqlet instances sampled from either the union set of peaks or cluster specific upregulated DARs.

Differential Chromatin Accessibility Analyses. scATAC data (multiome) were normalized and log transformed using Scanpy (v1.10.3)[6]. Peaks were annotated with genomic feature information. Differential chromatin accessibility was calculated using Muon v0.1.6 between *Prdm1* cKO and WT cells within each Leiden cluster. Peaks with a false discovery rate < 0.05 and absolute $\log_2$ fold change ≥ 0.13 were classified as differentially accessible regions (DARs)[13].

PRDM1 Candidate Target Gene Analysis. HOMER annotatePeaks.pl was used to identify ATAC peaks harboring PRDM1/ISRE/EICE motifs. For each motif instance, the dot product between the ChromBPNet contribution-score vector and the motif's position-weight matrix was computed; instances with positive mean dot product were classified as seqlet hits. Genomic coordinates of seqlet hits were intersected with ATAC peaks linked to cluster DEG lists. Two additional notations were added to candidate PRDM1-repressed genes: (i) overlap with ≥1 up-DAR with the gene body; and (ii) co-localization of a PRDM1-related seqlet and an up-DAR.

Statistical analysis. For the biological datasets, we used Prism version 10 (GraphPad) for differential testing. Pairwise statistical tests were performed using a two-tailed Student's t-test or Mann-Whitney test, whereas multicondition comparisons were performed using a one-way ANOVA with Tukey's multiple comparison test or Kruskal Wallis with Dunn's multiple comparison test based on the distribution, as indicated. Difference in proportions were calculated using Fisher exact test and Benjamini-Hochberg method for correction for multiple hypotheses testing. Two-sample Kolmogorov–Smirnov tests were used to compare

the state-probability distributions between *Prdm1* cKO and WT cells within each Leiden cluster (significance at $p < 0.05$). Aggregate motif counts in upDARs versus background peaks from multiome-seq were tabulated per cluster. Fisher's exact tests (SciPy v1.9) with Benjamini–Hochberg correction were used for identifying enriched motifs. Co-expression in double-positive cells was assessed with Spearman correlations (95% CIs via Fisher z and a cluster-blocked permutation p-value). Co-detection enrichment used Fisher's exact on 2×2 tables, and detection prevalence by state was summarized with Wilson 95% CIs plus global chi-square/trend tests and pairwise Fisher (odds ratios).

The R code defining the clones and all other codes are provided at (https://github.com/pitt-csi).

References:

1. Manakkat Vijay, G.K. et al. Temporal dynamics and genomic programming of plasma cell fates. *Nature Immunology* **25**, 1097-1109 (2024).

2. Venkatasubramanian, M., Chetal, K., Schnell, D.J., Atluri, G. & Salomonis, N. Resolving single-cell heterogeneity from hundreds of thousands of cells through sequential hybrid clustering and NMF. *Bioinformatics* **36**, 3773-3780 (2020).

3. DePasquale, E.A.K. et al. cellHarmony: cell-level matching and holistic comparison of single-cell transcriptomes. *Nucleic Acids Research* **47**, e138-e138 (2019).

4. La Manno, G. et al. RNA velocity of single cells. *Nature* **560**, 494-498 (2018).

5. Bergen, V., Lange, M., Peidli, S., Wolf, F.A. & Theis, F.J. Generalizing RNA velocity to transient cell states through dynamical modeling. *Nature biotechnology* **38**, 1408-1414 (2020).

6. Wolf, F.A., Angerer, P. & Theis, F.J. SCANPY: large-scale single-cell gene expression data analysis. *Genome Biology* **19**, 15 (2018).

7. Chen, D. et al. Coupled analysis of transcriptome and BCR mutations reveals role of OXPHOS in affinity maturation. *Nature Immunology* **22**, 904-913 (2021).

8. Dannebaum, R. et al. Highly multiplexed immune repertoire sequencing links multiple lymphocyte classes with severity of response to COVID-19. *EClinicalMedicine* **48** (2022).

9. Lai, L. et al. Composition and diversity analysis of the B-cell receptor immunoglobulin heavy chain complementarity-determining region 3 repertoire in patients with acute rejection after kidney transplantation using high-throughput sequencing. *Experimental and Therapeutic Medicine* **17**, 2206-2220 (2019).

10. Huang, Y. & Zhang, P. Evaluation of machine learning approaches for cell-type identification from single-cell transcriptomics data. *Briefings in bioinformatics* **22**, bbab035 (2021).

11. Lynch, A.W. et al. MIRA: joint regulatory modeling of multimodal expression and chromatin accessibility in single cells. *Nature methods* **19**, 1097-1108 (2022).

12. Pampari, A. et al. ChromBPNet: bias factorized, base-resolution deep learning models of chromatin accessibility reveal cis-regulatory sequence syntax, transcription factor footprints and regulatory variants. *BioRxiv*, 2024.2012. 2025.630221 (2025).

13. Bredikhin, D., Kats, I. & Stegle, O. MUON: multimodal omics analysis framework. *Genome biology* **23**, 42 (2022).

Acknowledgments

We thank the expert personnel within Division of Laboratory Animal Resources, the Flow Cytometry Core, the Single Cell Core and UPMC Genome Center at the University of Pittsburgh for their invaluable assistance. We particularly wish to acknowledge the Lafyatis lab for help with the scRNA-seq and BCR-seq library preparations and Stuart Hay for setting up the viewer for scRNA-seq datasets. This research was supported by the UPMC ITTC fund and NIH grants (RO1 AI145064; H.S.) and (RO1 AI114587; D.M.R.).

Author contributions

H.S. and D.M.R. conceived and supervised the study. G.K.M.V. and B.R. performed all experiments. G.K.M.V., S.G., N.P., and P.H.G. performed the computational analysis. D.C., K.T., and S.K. helped with select computational analyses. N.P. and L.M.H. helped with the experiments. H.X., N.S., J.D and D.M.R. provided critical inputs during the study. G.K.M.V. and H.S. wrote the manuscript with input from other authors.

Competing interests

The authors declare no competing interests.