Gene regulatory network determinants of rapid recall in human memory CD4+ T cells

Alexander Katko[1]^, Svetlana Korinfskaya[2]^, Anthony T. Bejjani[1], Seyifunmi M. Owoeye[1], Zi F. Yang[1], Akshata Rudrapatna[1], Joseph A. Wayman[1], Michael Kotliar[2], Leah C. Kottyan[2,3,4], Artem Barski[2,3,4]*, Emily R. Miraldi[1,4,5]*

[1]Division of Immunobiology, Cincinnati Children's Hospital Medical Center, Cincinnati, OH 45229
[2]Division of Allergy and Immunology, Cincinnati Children's Hospital Medical Center, Cincinnati, OH 45229
[3]Division of Human Genetics, Cincinnati Children's Hospital Medical Center, Cincinnati, OH 45229
[4]Department of Pediatrics, University of Cincinnati College of Medicine, Cincinnati, OH 45267
[5]Division of Biomedical Informatics, Cincinnati Children's Hospital Medical Center, Cincinnati, OH 45229

^co-first
*co-corresponding

Abstract

Rapid recall is the hallmark of memory T cells. While naïve T cells require days to mount an effector response to a new threat, antigen-experienced memory T cells can produce cytokines within hours of the repeat encounter. The establishment of memory and control of rapid recall across lifespan is poorly understood, yet the mechanisms are fundamental to pathogen defense and immunological diseases. Epigenetic poising was proposed as a likely mechanism. Indeed, compared to naïve, memory cells exhibit enhanced chromatin accessibility proximal to rapid recall genes, but the transcription factors (**TFs**) that establish, maintain and utilize these putative regulatory elements are unknown. Here, we leverage single-nuclei (**sn**)multiome-seq (simultaneous snRNA-seq and snATAC-seq) to (1) characterize the dynamic activation responses of naïve and memory CD4+ T cell subsets and (2) reconstruct the underlying gene regulatory networks, at genome scale. Our analysis uncovered thousands of genes and putative regulatory elements with rapid-recall dynamics, shared and unique across the memory populations. A core of memory-associated TFs (MAF, PRDM1, RUNX2, RBPJ and KLF6) is predicted to orchestrate rapid recall and maintain accessible chromatin at rapid-recall gene enhancers in resting memory cells. We integrate GWAS to link our T cell populations and their dynamic chromatin landscapes to human phenotypes, nominating CD4+ T cell populations, rapid recall responses and gene regulatory mechanisms that might mediate genetic risk to autoimmune and inflammatory diseases.

Introduction

Memory CD4+ T cells are critical for protective immunity and tolerance. Upon initial antigen encounter, naïve CD4+ T cells undergo clonal expansion and, over several days, differentiate into cytokine-producing T helper (**Th**) cells. Depending on the cytokine and microenvironmental cues coincident with initial antigen encounter, the differentiating naïve cells acquire diverse immune capabilities: defense against intracellular pathogens (so-called Type 1 Th cells (**Th1**))[1], extracellular microbes (**Th17**)[2] and large extracellular parasites (**Th2**)[3,4]), antibody class switching and affinity maturation (T follicular helper cells (**Tfh**))[5–7] and tolerance (induced regulatory T cells (**iTreg**))[8]. The powerful defense capabilities of effector cells can cause collateral damage to the host. Most effector cells die following pathogen clearance, but some survive as memory cells. Memory cells possess a "rapid recall" ability to secrete cytokine and initiate clonal expansion within hours of re-exposure to antigen[9]. Rapid recall enables potent pathogen defense and is important for successful vaccination[10]. The phenomenon is remarkable because memory cells otherwise persist in a polar-opposite state (near quiescence) across decades and even lifespan[11], their powerful effector and proliferation capabilities latent in the absence of antigen. Dysregulated rapid recall responses pose serious health threats: failure to mount contributes to chronic or life-threatening infections, while inappropriate, excessive responses drive allergic and autoimmune conditions.

Despite its importance, the mechanisms underlying rapid recall and its long-term maintenance are poorly understood. Both naïve and memory CD4 T cells can be activated by antigen-presenting cells (**APCs**). T cell receptor (**TCR**) stimulation by class II major histocompatibility (**MHCII**)-antigen complex and co-stimulation (e.g., via B7-CD28 axis) are sufficient for activation of both cell types. Memory cells have a higher barrier to activation than naïve cells[12], are less dependent on co-stimulation[13] and some do not even express CD28 co-receptor. While differences in ZAP70, LAT and SLP76 have been described[14–16], the expression and activation kinetics of TCR signaling proteins are generally similar between naïve and memory T cells[17], and these signals converge to activate common downstream TFs, including AP-1[18], NFAT[19] and NFkB[20,21]. We and others thus hypothesized that the rapid recall ability of memory cells is regulated epigenetically, via local chromatin state that influences transcription factor (**TF**) binding at key loci[22–25]. Initial evidence for this hypothesis was provided by a study, which identified differential DNA methylation at cytokine gene promoters between naïve and memory cells, and epigenetic inheritance by daughter cells in the absence of TCR activation[26].

Naïve T cell activation and effector cell differentiation dramatically alters the epigenome. Early single-gene and more recent 'omics studies report alterations in DNA methylation, deposition of lineage-specific histone marks at regulatory elements, chromatin opening at key cytokine genes and changes in 3D chromatin structure. In resting memory T cells, histone modifications at regulatory elements of cytokine genes correlated with their rapid recall upon activation. These epigenetic modifications are thought to speed binding of activation-induced TFs to their regulatory elements, enabling faster gene expression kinetics in memory versus naïve T cells. Multiple TFs regulate expression of the so-called "rapid-recall genes". These TFs include (1) TCR-inducible factors such as AP-1[18,27], NFATs[19] and NFkB[28], (2) cytokine-induced STATs[29,30] and Th subset-specific regulators such as TBET[31], GATA3[31], RORC[32] or FOXP3[33,34], and (3) non-inducible factors from ETS[35] and RUNX[36,37] families, posited to maintain the epigenetic state in resting memory T cells. Given the large number of potential regulators, a comprehensive approach is needed to elucidate the gene regulatory networks (**GRNs**) orchestrating rapid recall.

GRNs describe the control of gene expression by TFs. Genome-scale inference of GRNs from perturbation time-course gene expression and chromatin state data (e.g., using RNA-seq,

ChIP-seq, ATAC-seq) provides a high-throughput means to infer thousands of regulatory interactions. CD4+ T cell biology has been an inspiring frontier for development of novel genome-scale GRN inference methods[38–43]. However, these prior works mainly focused on naïve cell activation/differentiation and not rapid recall specifically. Similarly, genomic characterizations (e.g., RNA-seq, ChIP-seq) of the rapid recall process, involving side-by-side comparison of TCR activation dynamics between memory and naïve cells, exist, but have not yet been integrated into a comprehensive GRN model. Their utilization for this purpose, however, is suboptimal, as these existing studies characterize epigenome at the population level, obscuring heterogeneity, an important GRN signal, in the human CD4+ T cell compartment.

Multi-modal, single-cell (**sc**)-genomics technologies provide a novel and necessary strategy for GRN modeling of rapid recall. Memory T cells are composed of populations of central (**TCM**) and effector (**TEM**) memory cells, distinguished by their homing, proliferation capacity and ability to produce cytokines. It has been notoriously difficult to resolve resting human TEM populations based on sc-transcriptomics alone. Thus, multi-modal, sc-technologies are needed.

Here, we apply single-nuclei (**sn**)-multiome-seq technology (parallel transcriptome and chromatin accessibility) to a T-cell activation time course of naïve and memory human CD4+ T cells, to elucidate the molecular drivers of rapid recall at genome scale. Multi-modal analyses (1) enabled resolution of the CD4+ T cell subsets across the time course, including resting TEM populations (Th1, Th2, Th17, Treg, cytotoxic and MHC class II-expressing populations) and (2) identified thousands of “rapid recall” genes and accessible chromatin regions exhibiting faster activation dynamics in memory versus naïve cells. Furthermore, hundreds of genes and thousands of accessibility regions distinguished resting memory and naïve populations; these molecular correlates represent potential drivers of rapid recall programming and maintenance at rest. We built genome-scale GRNs to identify TF regulators of gene expression and chromatin state across the CD4+ T cell populations, at rest and in response to T-cell activation. GRN analysis nominated a set of core TFs, RUNX2, PRDM1 (BLIMP1), MAF, RBPJ and KLF6, responsible for both resting maintenance and early activation dynamics of rapid recall genes. Additionally, we shed light on TCM functional roles following activation, identifying potential for accelerated proliferation and acquisition of effector phenotypes. Integration of GWAS risk variants with our sc-resolved maps of chromatin accessibility implicated memory Th subsets and their rapid recall ability in genetic risk mediation of inflammatory and autoimmune diseases. Furthermore, we integrate genetic risk variants into the GRN, to nominate disease-relevant regulatory mechanisms in CD4 T cells.

Results

Resolution of naïve, TCM and TEM populations across a TCR activation time course for GRN inference

We designed T-cell activation time course experiments to unravel, at genome scale, the gene regulatory networks (**GRNs**) governing the rapid recall phenomenon in human memory CD4+ T cells (**Fig. 1A-C**). Given CD4 T cell heterogeneity, we selected the single-nuclei (**sn**) multiome-seq technology, to simultaneously characterize the chromatin landscape (accessibility by snATAC-seq) and gene expression (snRNA-seq) at single-cell resolution. CD4 T cells from the blood of 4 healthy donors were prepared by negative magnetic selection. Within the CD4 pool, the T effector memory ($CD27^-$, **TEM**) population is low abundance but includes cells capable of high production of lineage-specific cytokines. Thus, we increased the fraction of TEM cells by partially depleting naive and T central memory (**TCM**) using anti-CD27 magnetic beads

(**Methods**). Cells were then activated with anti-CD3/28 beads for early (2.5h), peak memory (5h) and late (15h) activation timepoints or remained resting and subjected to 10x snMultiome analysis (16 experiments: 4 donors by 4 timepoints). To control for batch effects, resting cells from a fifth control donor were "spiked" into each experiment immediately before cell lysis; this strategy improved the quality of inferences in downstream analyses (**Fig. 1A, S1A-H, Methods**).

The multiome-seq analyses yielded 65,862 high-quality CD4 T cell nuclei, which composed 15 clusters (**Fig. 1B, Methods**), robustly recovered from each donor replicate (**Fig. S1I-L**). A cluster of "stressed" cells, mainly from 15h, lacked evidence of TCR activation, expressed stress genes and were removed from subsequent analyses (**Methods**). Next, we aggregated gene expression or accessibility signal from individual cells to generate pseudobulk profiles per CD4 T cell subset (e.g., Th2), time point and donor (**Methods**). Importantly, pseudobulk analyses mitigate data sparsity issues, leveraging biological variation across replicates to more accurately estimate false discovery rate (**FDR**) and assess reproducibility[44,45].

We used principal component analysis (**PCA**) to identify the major sources of variation in the gene expression and chromatin accessibility data. Activation was the dominant source of variation, principal component (**PC**) 1, in both data types (**Fig. S2A-F, Methods**). However, PCA also revealed data-modality differences. While PC1 was composed of genes peaking at 2.5-5h for transcriptome, the major trend (PC1) in the accessibility data corresponded to peaks monotonically gaining or losing accessibility over the time course. For gene expression, distinct temporal patterns dominated the gene expression data, corresponding to PCs 1-3, totaling 44% of the data variance. As expected, PC1 had strong contributions from the activation-dependent genes such as *IRF4* and *CD69.* For transcriptome, PC4 was the first PC to distinguish cell types: effectors (Th1, Th2, Th17, MHCII, CTL) from non-effectors (naïve, TCM, Treg). In contrast, for chromatin accessibility, the second most important trend, PC2, separated effectors from the non-effectors, while PC3 and PC4 captured a mix of cell-type and temporal variation. Hierarchical clustering complemented PCA. While transcriptomes formed four main clusters corresponding to timepoints (resting, 2.5h, 5h and 15h) **(Fig. S2G)**, clustering of accessible chromatin profiles was driven by both cell type and time: all resting populations clustered together, while, for activation timepoints, effectors and non-effectors clustered separately **(Fig. S2H).** These results underscore the complementary information provided by snRNA- and snATAC-seq.

Although transcriptome alone was sufficient to define activated naïve, TCM and TEM, chromatin accessibility was key to resolving (1) resting naïve versus TCM and (2) resting Th1, Th2 and Th17 populations (**Fig. 1D-E**). In the resting state, TCM and naïve cells share common signature genes (**Fig. 1D**), but resting TCM were distinguished from naive based on accessible chromatin proximal to *IL2* and *CXCR5* (e.g., **Fig. 1E, Fig. S2I**).

In addition to Th1, Th2 and Th17, we identified two small TEM populations, "MHCII", "CTL". Both populations expressed activation markers *IRF4* and *CD44* by 2.5h and lacked expression of homing TCM co-stimulatory molecule *CD27* (**Fig. 1C, S2J-K**). The "MHCII" population was distinguished by elevated expression of MHC Class II and antigen presentation (*CD74*, *CIITA*, *PSD4*) genes, at baseline and upon activation (**Fig. 1C**, **S2L**).

The CTL populations expressed markers of cytotoxicity (*EOMES*, *GNLY*, *PRF1*), and activation further induced *GNLY* and *PRF1* expression. CTL also shared genes in common with Th1 and Th17 cells **(Fig. 1D).** Similar to resting Th1 and Th17 cells, *TBX21*, *IFNG*, and *RORC* were only detected in activated CTL. Upon activation, CTL induced *RORC* expression as robustly as Th17 cells, but without concomitant *IL17A/F*.

Finally, the "TCM/TEM" population was uniquely detected at activation timepoints, with cell numbers increasing from 2.5 to 15h **(Fig. S1K-L).** The TCM/TEM cells robustly expressed activation markers (early induction of *IRF4* and *CD44*, **Fig. S2I-J**), but with mixed central and effector memory features: expression of TCM markers *SELL* and *LEF1* and the cytokine *IL21*

(Fig. 1C). In addition, TCM/TEM shared accessible chromatin regions with TEM populations at several effector gene loci (e.g., CCR family members). Globally, by both transcriptome and accessible chromatin, TCM/TEM clustered more closely with TCM than TEM, suggesting that TCM/TEM are a TCM population that gains TEM features upon TCR activation (**Fig. S2A-H**).

These clusters – and their associated transcriptomes and accessible chromatin profiles, resolved by donor and timepoint – served as input to the *Inferelator* algorithm for GRN inference (**Fig. 1C**)[38,40]. From the chromatin accessibility we derived putative regulatory elements, using, for the first time, state-of-the-art neural network models *maxATAC* for TF binding site (**TFBS**) prediction [46] and Trac-loop data[47] to associate predicted TFBS with potential target genes. The resulting prior of ATAC-derived TF-gene interactions was used to estimate protein TF activities and guide gene expression modeling (GRN inference) **(Methods)**[40,48]**.** The resulting GRN describes 151,155 regulatory (TF-gene) interactions between 761 TFs and 15,116 genes **(Table S1.** Analyses of the GRN identified TFs involved in memory maintenance, orchestration of activation responses in memory and naïve cells and subset-specific gene expression programs (described below).

Thousands of rapid-recall genes exhibit faster, more extreme up- and down-regulation in activated memory cells

Having resolved the major CD4 T cell populations, we investigated their transcriptional responses to T-cell activation. In total, we identified 10,042 dynamic, activation-dependent genes: 2086 unique to memory populations, 21 unique to naive cells and 7935 common to both ($|\log_2(FC)| > 0.58$, FDR=10% for at least one timepoint relative to t=0 in at least one population, (**Table S2**). We defined rapid recall (**RR**) genes as those with faster, more extreme dynamics in memory versus naive cells (**Methods**). This analysis identified 5831 RR genes (**Table S2**). The RR genes clustered according to dynamics (e.g., timepoint of peak expression, increasing or decreasing) and whether the trends were observed in the TEM populations (e.g., Th1, Th2, Th17, MHCII and/or CTL), the TCM population or both groups (**Fig. 2A-B**, **Table S2**). We functionally annotated the RR gene clusters and queried the GRN for predicted TF regulators (**Fig. 2C, S3A-E**, **Methods**, **Table S3**).

TEM and TCM shared several upregulated gene clusters with maximal expression at early, middle, early and middle ("early+middle") or late timepoints (332, 432, 715 and 696 genes, respectively). As swift cytokine production is a defining feature of rapid recall, early, "early+middle" and middle RR genes are enriched in pathways essential for cytokine production, including ribosome biogenesis, translation, protein folding and protein localization. For the early and middle RR genes, the GRN identified known (IRF4[49], BATF[50], NFATC1[51,52] and NFAT5[53]) and novel (KLF6, POU2F2, HIVEP3) regulators in differentiation, cytokine response and immune effector function processes. **Fig. S3D**, **Table S3**). The GRN newly identified TF regulators of protein synthesis and transport pathways with "early+middle" cluster dynamics (**Fig. 2D**). For example, FOXK2 is a predicted regulator of ribosome biogenesis, protein folding and nuclear protein import, targeting ribosome assembly genes (*URB1*, *DDX46*), protein folding chaperones (*CCT5*, *CCT7*, *VAPA*), nucleoporins (*NUP93*, *NUP188*) and importins (*IPO11*, *IPO7*). The gene targets of IRF1, IRF4 and ETS2 are also enriched for ribosome biogenesis. MYB, STAT3, BATF, PRDM1 (BLIMP1) and MNT are predicted to regulate protein localization via induction of genes involved in intracellular transport (*COP1*, *XPOT*) and protein exocytosis (*CDC42*, *SCFD2*). Other TFs, including XBP1, MYC and E2F4, play roles in stress response, protein folding, biosynthetic processes (nucleotide, amino acid) and cell-cycle. Across the shared upregulated RR gene clusters with early and middle dynamics, many predicted regulators are themselves early (*NFAT5*, *NFATC1*, *BACH2*), early+middle (*CDC5L*, *GLIS3*) or middle (*XBP1*, *TBX21*, *IRF4*) rapid recall

genes (**Fig S3F-G**).

”Late” genes, maximal at 15 hours, include a second wave of genes to support massive protein synthesis (e.g., enriched in protein transport and folding, ER stress, cytokine production) and a new wave of genes involved in cell proliferation (e.g., enriched in DNA replication, cell cycle) Many regulators (e.g., VDR, NFE2L1, PLAGL1, ATF5) of late-upregulated genes are themselves late-upregulated genes (**Fig. S3F-G**). Indeed, we identified a sequential regulatory cascade, in which TFs belonging to the early and middle RR clusters upregulate six of the ten late-induced late controller TFs.

Two gene clusters were specific to T effector memory: “TEM up” (260 genes) and “TEM down” (493 genes). TEM-up genes are dramatically up-regulated (on average 3.4-fold) in TEM relative to naive and TCM across all timepoints; they display mixed dynamics, with most (87%) upregulated following activation. Functionally, they are enriched for signaling pathways associated with canonical effector functions. The cluster includes cytokines (*IL2*, *IL22*), chemokine receptors (*CCR2*, *CCR5*, *CCR6*), TFs (*MAF*, *RORA*, *RUNX2*, *MYBL1*) and costimulatory molecules (*LIGHT*, *CD58*, *CD2*). A core of mutually co-regulating TFs regulates these genes: RORA, RORC, MAF, RUNX2, MYBL1, KLF6, MYB, PRDM1 and GATA3 (**Fig 2E**). These TFs exhibit high expression and GRN-predicted TF activity (**TFA**) in resting TEM (**Fig. S3F-G**, **Methods**). Upon activation, their TFAs are maintained (RORA, RORC, MAF, PRDM1) or decreased (RUNX2, KLF6, MYBL1). The TEM-down genes are more lowly expressed in resting TEM and rapidly repressed further by 2h. In TCM and naïve, these genes are also downregulated, but to a lesser extent and more gradually. TEM-down genes include the naïve-associated TF *LEF1* and TCR signal transduction genes (*ZAP70*, *TEC*, *MALT1*).

Three rapid recall gene clusters exhibit higher amplitude downregulation in memory versus naïve populations. In total, RR downregulated genes (2818) are nearly as frequent as upregulated genes (3003). The ”down late recovery” cluster (793 genes) is comprised of genes repressed at the 2.5- and 5-hour timepoint, with return to or slightly exceeding baseline at 15 hours. Naïve cells exhibit similar dynamics but at lower amplitude. These genes are enriched for DNA repair, DNA replication, cell-cycle processes and nuclear division. They include DNA polymerases (*POLD3*, *POLE2*, *POLA2*), cyclin dependent kinases (*CDK2*), double-stranded break repair enzymes (*BRCA1*, *BRCA2*) and three of the six origin recognition complex genes (*ORC1*, *ORC3*, *ORC5*).

The “down weak recovery” cluster shows similar dynamics, but these repressed genes only recover to ~50% of their baseline expression by 15 hours. They are involved in processes important for cell survival and homeostasis in the absence of antigen (e.g., endocytosis, macroautophagy). At 15 hours, these genes may regain expression to support a prolonged immune response by, for example, recycling organelles damaged during the activation. The gene cluster also contains *TOX*, which is involved in[54–56] and promotes *IL10* production thereby attenuating pro-inflammatory responses[57].

Finally, “down” genes monotonically decrease, with faster downregulation in memory cells. Some genes (*PTPRC*, *FYB1*, *PIK3CA*, *TRAT1*, *TNFRSF21*) are directly involved in TCR signaling. Others are connected to pathways that may interfere with activation; protein ubiquitination (*TRIM5*, *BIRC2*, *PRKN*), autophagy *(TOM1, SESN3*, *SIRT1*) and heterochromatin formation (*MECP2*, *PPM1D*). The GRN predicts that several TFs with activation-dependent activity (e.g., NFAT, IRF4, RUNX2, THAP1 and MAF) not only activate RR genes but also repress RR “down” genes, suggesting dual activities. Other “down” gene regulators (ETV3, MAFG and NFE2L3), are predicted to be primarily repressors. A second mechanism for sustained gene repression is inhibition of gene activators. The TFs TCF7, KLF2 and ELF1 are both activators of and members of the “down” cluster, with known and predicted regulatory interactions among themselves[58] (**Fig. S4A**). Following activation, PRDM1, BACH2 and STAT4 are predicted to

repress *TCF7*, *KLF2* and *ELF1*. In support, *Tcf7* is a target of Prdm1 in murine effector CD8+ T cells[59].

While many genes exhibited RR dynamics across multiple memory populations, nearly half (45%, 2653 genes) were unique to a single memory subset (**Fig. S3A-C**). We thus sought to characterize the unique activation responses across the memory populations. Although individual genes were induced in the expected populations (e.g., *IL17A/F* in Th17 cells, *GNL* in CTL and antigen presentation genes in MHCII TEM), we detected very few pathway enrichments in the subset-specific rapid recall gene sets, indicating that the transcriptional responses, at a pathway level, are largely shared across TEM subsets.

In addition to the rapid recall genes, there were 4694 genes with similar activation dynamics in naive and memory cells (**Fig. S4A, Table S2**). These genes formed five clusters: early (816), middle (767), late (1067), down (759) and down late recovery (1285). Genes upregulated at early and middle timepoints were enriched for mRNA processing, splicing, translation, protein folding and other pathways supporting large-scale protein production. Their predicted regulators are IRFs, STATs, XBP1, KLFs and E2F4. Late upregulated genes were enriched for DNA repair and oxidative phosphorylation; predicted regulators included NR1H2, EGR3 and ATFs. Downregulated genes were involved in protein catabolism ("down") and DNA repair and autophagy ("down recovery"). Notably, many RR regulators (PRDM1, MAF, RUNX2, FOXK2, KLF6, etc.) were not predicted to regulate genes with memory-naïve shared activation dynamics.

Activated TCM diverge into highly proliferative and TEM-like cell states

Compared to TEM, TCM cells are more restricted to lymph nodes, exhibit an increased proliferative capacity and reduced production of effector cytokines[60,61]. It is hypothesized that TCM serve as a reservoir for effector cells. Upon activation, murine TCM can differentiate into effector cells and migrate to peripheral tissues to orchestrate immune responses locally[62,63]. Another study reported that *in vitro* activation of sorted human CD4+ TCM caused the TCM to acquire Tfh cytokine signatures[64].

Our study provided an opportunity to explore human TCM function and regulatory mechanisms in the context of activation, relative to TEM. The TCM population is composed of three subpopulations: TCM, TCM/TEM (**Fig. 1B, C**), and "proliferating TCM". The proliferating TCM population is a transcriptional state, apparent only through clustering of cell transcriptomes, independently of the accessible chromatin data (**Fig. 3A**). Indeed, we previously reported that proliferation genes are poised (i.e. have the chromatin environment similar to expressed genes) in resting T cells[65]. Within our time course, the proliferating TCM state is primarily detected at the 15-hour timepoint (95%) and accessed by both TCM and TCM/TEM cells (**Fig. 3A**).

The proliferating TCM exhibit increased expression of DNA repair genes (*CHEK1*, *DNA2*) and DNA replication genes (*MCM10*, *POLQ* and *HELLS*) (**Fig. S4C**), suggesting enhanced proliferative potential. Relative to other memory populations, the induced RR genes of TCM at 15h are enriched for cell-proliferation genes: cell cycle, DNA replication, "G2/M transition" and others (**Fig. 3B, Table S4**). The GSEA results are supported by single-cell cell-cycle phase predictions (**Fig. S4D**). Proliferating TCM had the largest proportion of cells predicted to be in S, G2 or M phases, relative to all other cell-types at 15h ($P_{adj} < 0.1$, pairwise t-test).

We next visualized the dynamics of rapid recall genes involved in cell cycle, identifying three major trends (**Fig. 3C**)**.** The first gene cluster peaks at 2.5h in memory cells. The cluster includes *CDK4* and other regulators of early cell cycle progression: *MYB, MCMBP*, *CDC37* and *CDC123.* The second gene cluster increases expression over the course of activation; it includes genes involved in G1 and S phase (*CDK6*, *CCND2*, *CINP*, *PES1*) and G2/M transition and mitosis

(*CDK7, LMNA, TUBB,* CENP family). The third cluster consists of late RR genes involved in DNA replication and S phase progression (*CDK2*, *CDC7*, *CHECK1*, *HELLS*, MCM and BRCA families). The proliferating TCM subset exhibits the highest expression of genes from this cluster, suggesting increased competency to start cell division.

Increased MYC activity correlates with the proliferative potential of TCM (**Fig. 3D**). While there is no difference in *MYC* expression at 2.5 or 5 hours between TCM and the other populations (**Fig. S4E**), the GRN-predicted TFAs of both MYC and its binding partner ZBTB17 (MIZ1) are elevated in TCM, indicating post-transcriptional regulation of MYC in memory cells (**Fig. 3D**). TCM and TCM/TEM at 2.5h have the highest ZBTB17 activity, relative to all other TEM populations ($P_{adj} < 0.1$, t-test). Another potential mechanism is the activation-dependent downregulation of *GSK3B*, as GSK3B promotes MYC degradation (**Fig. 3D**)[66]. Consistent with elevated MYC activity, TCM have the highest expression of several MYC-activated targets, including *CCND2* (Cyclin D), *CDK4* and *CDK6*, which enable faster cell-cycle progression (**Fig. 3D**).

Cytokine-stimulated human TCM can be polarized into a "pre-TEM" state, suggesting that TCM might replenish TEM in inflammatory conditions[64,67]. Here, our data suggests that T cell activation causes a subset of TCM cells to take on effector features, leading to the TCM/TEM population (**Fig. 1B, 3A**). The assertion that the TCM/TEM are derived from TCM is based on their overall similarity in both transcriptome and accessibility to TCM (**Fig. S2A-H)**. Relative to TCM, the TCM/TEM have increased expression of cytokine genes, including *TNF*, *IFNG*, *IL5*, *IL9,* and *IL21* (**Fig 3E**, **Fig. S4G**), and increased accessibility in these loci (**Fig. 3F**).

To nominate TFs contributing to the putative TCM transition to TEM, we created a core regulatory network for the TCM/TEM population (**Fig. 3G-H**, **Methods**). Several TFs core to TEM gene regulation (**Fig. 2G**), including MAF, KLF6, PRDM1 and RORA, were predicted to drive gene expression differences between TCM/TEM and TCM, suggesting that these TFs may also contribute to TCM acquisition of effector phenotypes. However, there were also important differences. The core TEM factor RUNX2 was notably absent, as *RUNX2* or its predicted activity was not increased in TCM/TEM–relative to TCM (**Fig. 3H**). Additionally, our TCM/TEM network highlighted a role for TET2, whose transcript was increased in TCM/TEM, relative to both TEM and TCM populations (**Fig. S4F**). TET2's DNA demethylation activity is crucial to reshaping the chromatin landscape during T helper differentiation[68,69]. Thus, TET2 may likewise be required to remodel chromatin and generate effector potential in TCM/TEM. Our GRN predicts that TET positively regulates *CXCR5*. In support, murine Tet2 is required for efficient demethylation of the *Cxcr5* locus in $CD4^+$ T cells[69]. Although CXCR5 is most famously a Tfh receptor[70], *CXCR5* was also expressed by a subset of circulating TCM and interpreted as a "pre-TEM" marker[64,71]. Collectively, our findings suggest that activated TCM have greater proliferative potential than TEM and may serve as a reservoir for TEM cells. GRN analysis identified potential regulators of both processes: elevated MYC activity correlates with the TCM proliferation potential, while TET2 and a subset of TEM-core TFs are predicted to alter chromatin and transcriptome, leading to effector potential acquisition.

Most activation-responsive accessible chromatin regions exhibit rapid recall dynamics

Having characterized transcriptional responses to activation, we next investigated the underlying epigenetic landscape. Of the 324,074 accessible chromatin regions ("peaks") identified in our study, 157,355 were dynamic (83,317 unique to memory, 5,452 unique to naive and 68,586 dynamic in both). Mirroring rapid recall genes, 46,134 "rapid recall peaks" were defined based on faster, higher-amplitude activation-induced opening or closing in memory versus naive cells (**Methods**). Unlike RR genes, which compose most dynamic genes (58%), RR peaks are a smaller fraction of all dynamic peaks (29%). In addition, while RR genes comprised nearly equal

numbers of up- and downregulated genes, RR peaks primarily gained accessibility during activation (78%).

The RR peaks clustered into seven dynamic patterns (**Fig.4A-B, Table S5**). Accessibility in the "early open" cluster (4517 peaks) peaked at 2.5h and diminished slightly at subsequent timepoints. The "middle open" cluster (8363 peaks) monotonically gained accessibility throughout the time course. The "middle open TEM" cluster (6441 peaks) were elevated in resting TEM and gained accessibility across the time course (**Fig. 4B**). There were two "late-opening" peak clusters, maximally accessibility at 15h, distinguished by the magnitude of accessibility increase between 0 and 15h, which was nearly 4-fold higher for the "late open strong" cluster (median $\log_2$(FC) = 3.2, 10,578 peaks) relative to the "late open weak" cluster (median $\log_2$(FC) = 1.3, 6097 peaks). Finally, activation rapidly diminished accessibility in the "rapid closed" (4651 peaks) and "rapid closed TEM" (5486 peaks) clusters. These peak clusters are exemplified by intronic peaks of several RR genes: *CD58* (early open), *VIM* (middle open), *MAF* (middle open TEM) and *LEF1* (rapid close) (**Fig. 4C**).

We also identified 39,622 activation-dependent peaks with similar dynamics between memory and naive cells, which we termed "shared activation-dependent peaks" or "act peaks" (**Fig. S5A**). These peaks clustered into five patterns. The "act open" cluster (8,136 peaks) exhibited early opening dynamics (i.e., maximal gain by 2.5h). The "act open TCM+naïve" cluster (7,927 peaks) similarly exhibited early opening dynamics, but the accessibility gain at 2.5h was at least two-folder higher in TCM and naïve populations relative to TEM. Across the CD4 populations, the "act late open" cluster (8,733 peaks) steadily gained accessibility at each timepoint, achieving maximal accessibility at 15h. The "act close" cluster (8,691 peaks) closed in response to activation, with the greatest loss in accessibility between 0 and 2.5h. Finally, the "act close naïve" cluster (6,134 peaks) was composed of peaks with higher baseline accessibility that monotonically decreased upon activation in naive cells. Although few of the "act close naïve" peaks were dynamic in TCM, in TEM, accessibility was rapidly diminished by the 2.5h time point and then gradually increased to achieve accessibility on par with naïve cells at 15h.

We linked RR peaks (putative enhancers) with their potential gene targets. 53% of the RR peaks were promoter-proximal (+/-5kb of a gene TSS). Relative to all accessible chromatin regions in our study, RR peaks were more frequently gene-distal (OR=2.13, P<0.001) and more likely to form promoter-enhancer interactions, as detected from Trac-loop of resting and activated CD4 T cells (OR=1.25, P<0.001)[47]. Similarly, activated peaks were also enriched for gene-distal regions (OR = 2.08, P<0.001) and Trac-loop enhancers (OR = 1.15, p < 0.001).

We linked 88% of RR peaks and 90% of activation-shared peaks to potential target genes using Trac-loop interactions (21% and 23%, respectively) and a linear proximity rule (±10kb gene body) for the remaining peaks (67% for both). In general, RR peak sets were most enriched for the RR gene sets with corresponding dynamics (**Fig. S5B**). For example, peaks with early opening dynamics were most strongly enriched proximally to genes with "early", "early+middle", and "middle open TEM" RR dynamics. The "late open strong" peaks were enriched proximally to the "late up", "down late recovery" and "down weak recovery" RR genes, suggesting a role for these regions not only in promotion of rapid-up genes but also recovery of gene expression at 15h. Shared activation-dependent peaks were also strongly enriched proximal to both RR and activation-shared genes. For example, "act open" and "act open TCM+naïve" peaks sets were enriched proximally to early, middle and down RR genes as well activation-shared early up-regulated genes (**Fig. S5B**).

To identify TF drivers of activation-dependent chromatin remodeling (**Fig. 4A, Fig. S4A**), we built a novel multivariate modeling framework. While traditional TF motif enrichment analyses provided some clues about TF regulators of the accessibility patterns, motif enrichment analyses do not leverage the important relationship between TF activities and chromatin accessibility

signal. Post-hoc assessment of this relationship is not trivial, when considering the multivariate, context-specific interactions of TFs across the 122 pseudobulk ATAC profiles (e.g., as shown in **Fig. 4A**). Thus, we explicitly modeled chromatin accessibility as a function of GRN-derived protein TF activities, using regularized regression to incorporate TFBS predictions (**Methods**). We visualized the predicted TF drivers of chromatin accessibility across peak sets with RR and naïve-memory shared activation dynamics (**Fig. 4D**). We compared the relative activities of these TFs between memory and naive cells, to identify factors driving TCR activation chromatin dynamics in both contexts. Several TFs, including PRDM1, RARG, TBX21, MAF, RORC and RUNX2, are more active in memory cells and specifically regulate accessible chromatin with RR dynamics, but not peaks with naïve-shared dynamics. In contrast, other memory-associated TFs (MYBL1, ETS2, BACH1, BACH2, HLF, RBPJ) regulate the shared activation-dependent peaks, suggesting different regulatory mechanisms by which memory and naive cells control accessibility of the same peaks. Many AP-1 factors (e.g., JUN, JUND, FOS, FOSL1) had similar activity in naïve and memory cells and were predicted to regulate both RR and activation peaks. In contrast, NFKB-associated TFs (RELA, NFKB2) and FOSB regulated shared activation but not RR peaks. Several naive-associated TFs were predicted to repress chromatin accessibility changes in RR peaks: EGR1 and NR2C2 repress opening of early rapid recall peaks, while LEF1 and ZNF460 are predicted to maintain the rapidly closed TEM peaks open in naïve cells.

In summary, hundreds of thousands of accessible chromatin regions exhibit activation-dependent dynamics and are linked to genes with corresponding dynamics. GRN modeling prioritized specific TFs, unique or common to naïve and memory populations, responsible for orchestrating a diverse RR and naïve-memory shared dynamic patterns.

Identification of TFs governing memory maintenance at rest

We previously hypothesized that the rapid recall ability of memory T cells is encoded in their resting-state epigenome[23]. Other studies suggested that members of RUNX and ETS TF families are responsible for maintaining memory-defining open chromatin elements at rest[72]. To uncover memory persistence mechanisms from our high-resolution multiome data, we identified accessibility chromatin regions differential between resting naive and memory populations and searched for candidate controllers: TFs (1) with binding site enrichment and (2) differentially expressed between resting memory and naive populations (**Fig. 5A-B**, **Tables S5, S6**). Our analysis yielded 32,661 candidate regulatory elements, potentially encoding T cell memory. The majority exhibited increased accessibly in memory populations (23,055 peaks, 71%), which were resolved into four clusters based on memory subset-specific patterns. The Th1/Th17/CTL cluster (4383 peaks) had the greatest accessibility in CTL followed by Th17 and Th17 populations, while the Th2/MHCII cluster (2904 peaks) was most accessible in Th2 and MHCII cells. These clusters were enriched for binding sites of their associated master regulators (TBX21, RORC, EOMES or GATA3) as well as regulators of the TEM activation response (PRDM1, MAF, RUNX2), nominated by GRN (**Fig. 2E**) and accessibility network (**Fig. 4D**) analyses. These TFs were also supported by elevated transcript in the corresponding memory populations. The other memory-associated clusters were distinguished by relatively low (13,860 peaks) or high (2036 peaks) accessibility in Treg. By motif enrichment and transcript expression, BATF, MAF and RUNX2 are implicated in the resting maintenance of both clusters. Finally, two clusters exhibited increased accessibility in naive cells, were moderately accessibility in TCM and distinguished by moderate or low Treg accessibility ("naive+Treg" and "naive" clusters, respectively). Motif enrichment and gene expression criteria nominated several TFs, including four LEF1, ETS1, TCF7 and EGR2, as regulators of the naïve-associated clusters.

We next explored how regions of resting memory-associated chromatin might underlie

rapid recall dynamics of peaks and genes. The phenomenon is exemplified by the *IFNG* locus, in which a rapid recall peak, linked to the *IFNG* by trac-loop, exhibited low-level but significantly increased accessibility in Th1 relative to naïve cells prior to activation (green box, **Fig. 5C**). Indeed, many of the rapid recall peaks were differentially accessible in resting memory versus naive cells. Apart from the "late weak open" cluster, all other RR peaks with opening dynamics were significantly enriched for peaks open in resting memory versus naive cells (**Fig. 5D**), supporting that pre-existing chromatin accessibility in resting memory cells contributes to rapid recall. Likewise, the RR peak cluster with closing dynamics (Close TEM) was enriched for peaks with less accessibility in resting memory versus naïve cells, suggesting epigenetic mechanisms for faster, RR closing of accessible chromatin regions. In contrast, none of the activation-dependent peak sets with shared memory-naïve dynamics were enriched for resting memory-associated accessible chromatin (**Fig. 5D**).

Prior studies found little evidence that rapid recall was driven by enhanced expression or activity of canonical TCR signaling proteins[17]. Here, we reexamine this possibility, taking advantage of sc-resolved memory subset transcriptomes, to discover gene expression differences, between memory and naïve populations at rest, potentially contributing to rapid recall. In total, there were 1907 genes differentially expressed between naïve and at least one memory population (1028 upregulated in memory, 869 upregulated in naive) (**Fig. 5E, Table S2**). Using the trac-loop dataset, resting memory-up peaks were enriched for enhancer links to resting memory-up genes, while resting naive-up peaks were enriched for resting naive-up genes (one-sided Fisher's exact tests, $P < 0.01$). GSEA revealed a common theme of cell adhesion, activation and migration among the upregulated memory genes. The "Treg" memory-associated gene cluster (highest resting expression in Treg, 161 genes) was enriched for immune regulation pathways and included Treg-associated genes like *FOXP3*, *CTLA4* and *IKZF2*. Naïve-upregulated genes included TCR signaling molecules (*LAT*, *LCK*, *PLCL1*) and naive-associated (*FOXO1*, *LEF1*, *TCF7*) and other TFs (*ETS1*, *EGR1*).

Using the GRN, we mapped TFs (**Fig. 5B**) to genes that might contribute to rapid recall **(Fig. 5F, Methods).** The TFs KLF6, MAF, RBPJ, RUNX2, NR3C1, SMAD3, PRDM1, RORA and MYBL1 were core regulators of RR-relevant resting memory genes. TF regulation of co-stimulatory molecules *CD58*, *CD2*, *TNFRSF18*, *CD63*, *TNFSF14* and *CD70* at rest may contribute to RR. For example, CD58 and its binding partner CD2 increase T-cell responsiveness to activation and downstream cytokine production through PLCγ1 dependent Ca^{2+} signaling[73]; thus, baseline increase in *CD58* and *CD2* expression may drive rapid-recall responses *in vivo*. Collectively, our analyses nominate accessible chromatin regions, transcripts, TFs and GRNs, uniquely utilized by resting memory cells to enable their rapid recall ability.

Memory-associated TFs are implicated in early differentiation of naïve cells to effectors

Limiting to naïve cells, we used pseudotemporal ordering to characterize the naive compartment's heterogeneity across resting and activated timepoints. The pseudotemporal ordering correlated with the temporal dynamics of our design, excepting some unstimulated cells (low *IRF4*, *CD44* expression) from all timepoints, concentrated early in the trajectory (**Fig. 6A, B**). Next, we visualized module scores summarizing accessibility in the putative "memory-encoding" peak sets (**Fig. 5A**) on a per-cell basis (**Fig. 6C-E**). A subset of naive cells at late pseudotime, mostly from 15h, gain accessibility in regions open in resting memory and lose accessibility in regions defining resting naïve cells. These so-called "precocious" naive cells acquire memory-associated chromatin features faster than other activated naïve cells at 15h. This trend extends beyond the resting memory-specific regions to include accessibility changes in the RR and shared activation peak sets. Relative to other naïve activated cells at 15h, the precocious naive gained (or lost)

more accessibility in peak sets opening (or closing) in activated cells (**Fig. 6C, D, S5A-B**).

We explored potential TF regulators of the precocious naive cells (**Fig. 6E, F**). Like resting and activated TEM, naive precocious cells exhibited increased *KLF6* expression (**Fig. 6E, F**). Additionally, naive precocious displayed increased expression of *RUNX3* and *TET2,* both of which are characteristic of effector T cells[74,75]. Genome-wide, hundreds of genes have significant differences in either expression (667) or proximal accessibility (397) between naïve precocious and other 15h naive activated cells (**Fig. 6G**, **Table S7**). Some effector genes (*MAF*, *TET2* and *RBPJ*) were already elevated in naïve precocious, while other genes (*IL2*, *IFNG* and *TBX21*) had a significant increase in proximal accessibility but not expression, suggesting transcriptional priming.

Thus, the putative memory-encoding peaks: (1) exhibit accessibility in resting memory but not naive cells and (2) gain accessibility in activated naive cells by 15h. As a result, the "memory-encoding" peaks appear functionally important to both rapid recall (memory cells) and the differentiation of activated naive cells into effectors. Interestingly, both processes implicate shared TFs (*KLF6*, *MAF*, *RBPJ*) as well as unique TFs (e.g., *RUNX2* unique to memory and *TET2* and *RUNX3* unique to precocious naive cells).

Memory-associated accessible chromatin regions and subset-defining GRNs are implicated in mediating genetic risk for allergic, autoimmune and inflammatory diseases

CD4 T cells and their activation states mediate genetic risk for autoimmune and inflammatory diseases[76–80]. Genetic risk variants for these and other complex diseases are predominantly noncoding, enriched in regulatory elements and, if causal, thought to alter TF binding and gene expression[81]. To examine whether accessible chromatin regions defined by our study (e.g., rapid recall-associated) might mediate genetic disease risk, we utilized the Regulatory Element Locus Intersection (**RELI**) algorithm[77] to systematically assess the colocalization of GWAS-identified sets of genetic risk variants with accessible chromatin regions (**Table S8**, **Methods**). Of the 387 disease risk variant sets (phenotypes) tested, 44 significantly colocalized with at least one of our 18 accessible chromatin maps (FDR = 10%, **Fig. 7A**). Accessible chromatin regions distinguishing resting memory from naive populations were most significantly enriched for genetic risk variants, followed by accessible chromatin regions that opened upon activation, both RR and memory-naive shared.

Clustering of disease-peak set enrichments yielded four phenotype groups **(Fig. 7A).** The "Th2/MHCII disease" cluster had the strongest enrichments in peaks defined by high accessibility in resting Th2 and MHCII cells. Nearly all diseases in this group were atopic, allergic or respiratory, underscoring the well-established role of type 2 immunity in these conditions. The "Treg disease" cluster, which included the autoimmune and inflammatory conditions (e.g., hypothyroidism, rheumatoid arthritis, type 1 diabetes), were most enriched in resting memory-associated peaks that were also accessible in Treg (other resting memory-associated peak sets were elevated in other memory populations but not Treg). Primary biliary cholangitis (**PBC**) was in the Treg-driven disease group; in support, PBC has been associated with an altered Th17/Treg balance[82,83]. A third cluster involved less significant enrichments, lacked coherent patterns and was not analyzed further. The fourth cluster, "TEM/TCM shared diseases", encompassed both allergic diseases, such as hay fever and asthma, and inflammatory diseases, including ulcerative colitis, Crohn's disease and celiac disease. Although diseases in this cluster most significantly overlapped resting memory "TEM+TCM shared" peak sets, they were also enriched for RR peak sets (e.g., "middle open", "late open strong") suggesting that RR dynamics might be impacted in these diseases. A subset also overlapped peaks with naive-memory shared opening dynamics ("act open").

To explore potential CD4 T subset-specific gene regulatory mechanisms for these

diseases, we linked risk variants co-localized with accessible chromatin to genes, using promoter-enhancer interactions from Trac-loop[47] and micro-C data[84] to recover distal interactions and linear proximity (+/-1kb of TSS) for gene-proximal interactions. Using this method, 60% of overlapped risk loci were linked to putative target genes. Within an intron of *FGFR1OP,* the PBC-associated risk variant rs12529876 linked to the *CCR6* promoter 131kb downstream (**Fig 7B**). In patients with PBC, the frequency of CCR6^{+} immune cells in the hepatic portal tracts strongly correlated with the severity of inflammation[85]. Our analysis positions rs125892876 within a memory-associated accessible chromatin region, exhibiting elevated accessibility in resting memory populations (belongs to the "TEM+TCM Shared" cluster) and RR dynamics ("Late open strong"). In support that this putative enhancer drives *CCR6* expression, its accessibility correlates with *CCR6* expression across CD4+ T populations, over the activation time course (rho = 0.65, $P < 0.01$, **Fig. 7C**). Chemokine receptor CCR6 and its ligand CCL20 play important roles in several other autoinflammatory diseases, including inflammatory bowel disease, rheumatoid arthritis and psoriatic arthritis[86–88]. CCR6 may drive these diseases by enhancing Th17 migration to tissue lesions[89]. In our data, both *CCR6* expression and the accessibility of its putative enhancer are highest in Th17 and the Th17-polarized CTL populations.

After linking risk loci to genes, we used the GRN to predict underlying regulatory mechanisms, focusing on risk variants for the first disease cluster (asthma, atopic respiratory) that co-localized with resting memory Th2/MHCII-defining accessible chromatin regions (**Fig. 7A** green box, **7D-E**). Of the thirteen risk-variant-linked genes, eight displayed Th2- or MHCII-specific expression at rest or upon activation (**Fig. 7F**), and we queried the GRN to identify TF regulators with elevated Th2- and/or MHCII-specific activities at rest (**Fig. 7G**). The genes elevated in resting MHCII and/or Th2 populations *(KIF1B*, *TH2LCRR*, *HLA-DQB1, AHI1, IQGAP1)* also exhibited MHCII and/or Th2-specific dynamics upon activation, suggesting potential enhancer activity for the risk-variant linked putative regulatory regions. *RAD50*, *IL10* and *IL13* were not elevated in resting Th2 or MHCII but exhibited Th2-specific RR dynamics, suggesting a role for the risk-variant linked enhancers in Th2 poising of these genes. The GRN predicted known and novel regulators of these risk-variant-linked genes. For example, GATA3 is well known while KLF11 is a novel *IL13* regulator. In contrast, all regulators of MHCII-cell-associated risk genes are novel, given the novelty of this population (relative to well-studied Th2).

As a second example, we explored risk variant mechanisms for the second disease cluster (autoimmune and inflammatory conditions), enriched in resting memory Treg-defining accessible chromatin regions (**Fig. 7A** blue box, **S6E-F**). Of the seven risk-linked genes, six displayed Treg-specific expression. *IL2RA, CTLA4*, *ZNF292* and *ZC3H12D* exhibited both Treg-specific expression at rest and upon activation (**Fig. S6G-H**), suggesting Treg-specific activity for the linked risk variant-containing regulatory elements. FOXP3, PRDM1, XBP1 and ATF4 were predicted to maintain baseline expression and RR poising of these risk-associated Treg genes.

Our analysis of GWAS variants connected disease-associated polymorphisms to diverse and dynamic accessible chromatin regions defining T cell states. Within CD4+ T cell populations and contexts, the GRN enabled identification of gene regulatory mechanisms potentially contributing to genetic risk and pathogenesis of inflammatory and autoimmune diseases.

Discussion

Memory CD4 T cells protect against repeat infections, via rapid cytokine production and clonal expansion upon antigen re-exposure. While many molecular components (receptors, TFs and epigenetic features) have been reported, the regulatory networks governing memory responses and rapid recall were poorly understood. Bulk 'omics studies highlighted the role of the epigenome

in regulating memory CD4 T-cell response[22–25] but were unable to resolve regulatory elements active in heterogeneous subsets. In this study, sn-multiome-seq provided high-resolution, multi-modal analysis of CD4 T cell populations, especially the TEM compartment, which was resolved into five populations: Th1, Th17, Th2, MHCII and CTL. Paired transcriptome-accessibility data were critical to our resolution of resting Th subsets, as well as heterogeneity in TCM and naïve cells. In our previous studies[23,27], memory T cell populations possessed positive chromatin marks proximal to rapid recall genes (incl. subset-specific genes, *IFNG*, *IL4*, *IL17A)*. With multiome-seq resolution, we confirmed that these resting chromatin accessibility regions were indeed subset-specific.

Multiome-powered GRN inference revealed novel regulatory circuits governing rapid recall responses in memory CD4 T cells. In complex mammalian settings, integration of TF binding site (**TFBS**) predictions from chromatin accessibility critically improves GRN inference from gene expression data[40,90] . Here, in a first-time application to GRN inference, we utilized maxATAC, a cutting-edge deep neural network[46,] to generate genome-wide "*in silico* ChIP-seq", from snATAC-seq, for hundreds of TFs in our sc-resolved CD4 T populations – a scale and resolution infeasible with existing TFBS measurement technologies. These TFBS predictions, along with experimentally measured promoter-enhancer interactions[47], enhanced genome-scale GRN inference from the multiome-seq data. From the GRN, we predicted distinct and overlapping combinations of TFs orchestrating transcriptional and chromatin accessibility activation responses, enabling the precise control of dynamics (e.g., early versus late) and subset-specific responses.

CD4 T cells underwent extensive transcriptional reprogramming following TCR activation, particularly during the early-to-middle activation phase (2.5-5h). In this window, cells upregulated effector molecules and their protein synthesis machinery. GRN analysis predicted known (MYC, XBP1, IRF4) and novel (FOXK2) regulators of these processes. The wave of regulators between 5 and 15 hours (e.g., VDR, NFE2L1, PLAGL1, ATF5) controlled a distinct set of biological pathways, facilitating cellular homeostasis through induction of ER stress response and proliferation programs.

Our analysis implicated TFs (MYBL1, RORA, PRDM1, KLF6, RUNX2, RBPJ, MAF, NR3C1, SMAD3) in memory maintenance at rest. As an example, we predict that SMAD3 promotes memory cell survival through *IL7R* regulation while actively suppressing RR pathways prior to activation, suggesting a checkpoint role preventing premature T cell responses. These memory-encoding factors orchestrate diverse programs targeting survival, cell-cycle and signal transduction genes, including enhanced expression of costimulatory molecules other than *CD28*. Specifically, upregulation of *CD58* and its binding partner *CD2* might contribute to stronger activation responses and RR kinetics of memory T cells *in vivo*. Excepting RUNX2, these TFs maintain high activity post-activation, contributing to enhanced cytokine production, in addition to memory maintenance.

Our study uncovered heterogeneity among activated TCM cells. We identified activated "TCM" and a second population "TCM/TEM" that exhibited mixed features of TCM and TEM. Genome-wide, TCM/TEM transcriptomes and chromatin accessibility profiles were more similar to TCM than TEM, but, relative to TCM, the TCM/TEM exhibit increased proximal accessibility and/or expression of effector genes, including *IL21, TNF* and *IFNG*. Many of the TFs composing the resting memory core (KLF6, MAF, PRDM1, SMAD3 and RORA but not RUNX2 or RORC) were predicted regulators of the gene signatures distinguishing TEM/TCM from TCM, along with ZEB2 and DNA demethylation enzyme TET2. Indeed, extensive DNA demethylation was previously implicated in TCM-to-TEM transitions[91,92]. At 15h, TCM RR genes were more enriched for cell-cycle genes than the TEM populations, and we discovered a transcriptional state, accessed by both TCM and TEM/TCM populations, with elevated S-phase gene expression.

Together, these findings support an increased proliferative potential and role for TCM in repopulation of circulating and tissue-homing effector memory cells[61].

Members of the resting memory core were also predicted regulators of the activated "precocious" naive subpopulation (~10% of activated naïve), which exhibited accelerated activation kinetics by chromatin accessibility and transcriptome. However, there were also differences between the predicted regulators of resting memory and precocious naïve. Mirroring TCM/TEM, *TET2* was elevated in precocious naïve cells, presumably for a similar purpose: DNA demethylation at effector loci. There were also differences in RUNX family utilization. Our data suggests that RUNX3 contributes to the initial stages of naïve-cell differentiation to effectors, while RUNX2 uniquely contributes to resting memory maintenance, through its known roles in chromatin structure maintenance and regulation of cytokine production[72,93–97].

The resting memory core TFs may be important mediators of genetic risk for immune-mediated disease, as accessible chromatin maps defining the resting memory subsets (Th2/MHCII, Treg, pan-TEM/TCM) were most frequently and significantly associated with genetic risk variants for immune-mediated diseases. In addition, we nominated early (2-5h) activation-dependent accessible chromatin dynamics, both RR and naïve-memory shared, as important contexts for genetic-risk mediation by CD4 T cells. Focusing on atopic and asthma variant colocalizations with the resting Th2/MHCII-associated chromatin, we used experimentally derived promoter-enhancer interactions and queried the GRN to identify regulators of the risk-variant target genes. Indeed, resting memory core TFs RUNX2 and KLF6 appeared in the resulting subnetwork.

In conclusion, we describe a genome-scale GRN governing rapid recall and early activation timepoints in human naïve and memory T cell subsets. This sn-multiome-inferred model provides a foundation for future studies aimed at deciphering the transcriptional networks governing T cell responses in health and disease. In this manuscript, we highlight but a subset of regulatory mechanisms from the GRN and rich, underlying data. To maximize the resource, the GRNs and sn-multiome data are interactively available, along with open-source codebases to reproduce results here-in, from: https://github.com/MiraldiLab/RapidRecall.

Methods

Cell processing

Five leukocyte depletion filters containing cells from de-identified male donors were obtained from the University of Cincinnati Hoxworth Blood Center, Cincinnati, Ohio. Filters were back-flushed with PBS, and PBMCs were prepared by Ficoll centrifugation. Human CD4 T cells were isolated from PBMCs using CD4 isolation magnetic beads (STEM CELL #19052). To increase the fraction of TEM cells, naive and TCM cells were partially depleted using anti-CD27-biotin coated streptavidin magnetic beads (Invitrogen Dynabeads Biotin Binder 11047; CD27 antibody Biolegend 302804). Cells were cryopreserved in 50% CTS OpTmizer medium (A1048501), 40% human serum (Sigma H4522), 10% DMSO (ThermoFisher) and stored in liquid nitrogen. Cells were prepared for 16 multiome-seq experiments (4 donors x 4 time points). Thawed, purified cells were plated in CTS OpTmizer with 1% Penicillin-Streptomycin (Gibco 15140122) and incubated at 37°C for the activation duration time: 15h/5h/2.5h. Activation was performed using Dynabeads Human T-Activator CD3/CD28 (Gibco 11131D). Resting cells were incubated in the same medium and temperature for the same time as activated, but without activation beads added. To prevent cell-free DNA contamination originating from dead cells, DNase I (1:100) was added directly into the medium (NEB M0303S) during the last 30min of incubation. After incubation, cells were

washed with PBS + 0.1% BSA twice. For each multiome experiment, cells from the spike-in control donor (resting timepoint) were mixed with cells from the experimental donor at the time point of interest at a ratio of 8% (ctrl) to 92% (each test donor), immediately preceding the nuclei prep. Nuclei were prepared according to the modified 10X Genomics protocol: 100ul of cold lysis buffer (10mM Tris-HCl (pH 7.4), 10mM NaCl, 3mM $MgCl_2$, 1% BSA, 0.025% Tween-20, 0.025% Nonidet P40, 0.00125% Digitonin, 1mM DTT, 1 U/μl RNase inhibitor (Roche)) was added to cell pellets. Samples were incubated for 40s on ice, washed with 1mL of cold wash buffer (10mM Tris-HCl (pH 7.4), 10mM NaCl, 3mM $MgCl_2$, 1% BSA, 0.025% Tween-20, 1mM DTT, 1 U/μl RNase inhibitor (Roche)) and filtered with Flowmi cell strainer (Bel-Art H13680-0040).

scRNA and scATAC libraries preparation

The Chromium Next GEM Single Cell Multiome ATAC + Gene Expression assay was performed according to the manufacturer's instructions (10x Genomics) with the Chromium Next GEM Single Cell Multiome ATAC + Gene Expression Reagent Bundle (PN-1000283) kit. Briefly, nuclei suspended in the multiome buffer were combined with a transposition mix and incubated for 1h. The transposition reaction was combined with the kit's master mix and loaded together with partitioning oil and gel beads into the Chromium Next GEM Chip J Single Cell Kit (PN-1000234) to generate a gel bead-in-emulsion (**GEM**). The GEMs were incubated to produce 10x-barcoded DNA from the transposed fragments (for ATAC) and 10x-barcoded, full-length cDNA from poly-adenylated mRNA. Next, the GEMs were broken, the barcoded DNA and cDNA were purified, and the purified products were subjected to a pre-amplification PCR step. The pre-amplified material served as starting material for ATAC and RNA library construction per 10x protocol. ATAC libraries and GEX libraries were pooled separately. The pools were sequenced on the NovaSeq 6000 sequencer with SP or S1 flow cells, using the following sequencing parameters: R1: 50 cycles, i7: 8 cycles, i5: 24 cycles, R2: 49 cycles (for ATAC) or R1: 28 cycles, i7: 10 cycles, i5: 10 cycles, R2: 90 cycles (for RNA).

Whole genome sequencing and variant calling

DNA was isolated using a PureLink Genomic DNA Kit (Thermo Fisher). Whole-genome sequencing was performed using DNBseq next-generation sequencing technology. Libraries were sequenced on an Illumina NovaSeq to generate 150-base paired-end reads. Sequencing reads were aligned, and variants were identified using the Genome Analysis Toolkit (**GATK**) Unified Genotyper following GATK Best Practices[98–100].

scRNA-seq and scATAC-seq quality control

Initial alignment, filtering and UMI counting were performed using cellranger-arc-2.0.2 and the 10x genome GRCh38 reference genome within Scientific Data Analysis Platform (https://scidap.com)[101]. Then, using Seurat v5.1.0, the following quality thresholds were used to select quality cells: detected genes > 400, transcript unique molecular identifiers (**UMI**) > 750, percent mitochondrial transcripts < 20%, ATAC fragments > 2000, fragments in blacklist < 5%, nucleosome signal < 2.5, TSS enrichment > 2[102]. Initial clustering revealed small populations of non-CD4+ T cells, including Natural Killer, CD8+ T and myeloid cells based on either *CD8A* or lack of *CD3E* expression. These cells were removed from the analysis. We then used several criteria to filter likely doublet nuclei. We removed nuclei (1) predicted to be doublets based on both RNA (scDblFinder v1.15.2) and ATAC (ArchR v1.0.2) modalities and (2) those with unexpectedly large RNA (>10k UMI) and ATAC (>50k fragment) libraries[103,104]. For scDblFinder,

a merged SingleCellExperiment object was used for input (additional params: Cluster = TRUE, samples = "Dataset"). For ArchR, the 16 individual fragment files were used as input, as well as each samples expected number of doublets. The expected number of doublets was calculated by multiplying the average doublet rate reported by the Cell Ranger User Guide[105] by the number of cells in the sample. Across the 16 multiome-seq experiments, our QC pipeline yielded 65,862 quality, CD4+ T cell nuclei.

Sample integration and clustering

Prior to sample integration, each sample's Cell Ranger fragment file was modified to include both transposase cut sites[38,46]. Cut sites were filtered if they overlapped ENCODE GRCh38 blacklist regions. Peak calling was performed by MACS2 V2.1.4[106,107] on initial clustering of each Seurat dataset individually (MACS2 parameters: --shift -25 --extsize 50 --tsize 50 --pvalue 1E-8 --keep-dup all)[38]. A reference peak set was generated by taking the union of peaks called across all samples and merging overlapping peaks. A (peaks x cells) counts matrix was then generated using *FeatureMatrix* from Signac v1.13.0. This counts matrix replaced the original counts data generated from Cell Ranger.

Next, the Seurat v5.1.0 pipeline was applied to the RNA and ATAC data in parallel. For RNA, the data was log-normalized, centered and scaled by z-scoring, and then dimensionality reduced to the first 30 PCs. For ATAC, data was first normalized using run term frequency inverse document frequency (the *RunTFIDF* function) followed by latent semantic indexing (**LSI**) dimension reduction (the *RunSVD* function) using components 2-30. Component 1 was removed due to high correlation with library size (fragments per nucleus). Peaks with ≥1 cut site in 20 or more cells were included in the dimension reduction. After dimensional reductions for both RNA and ATAC were calculated, each data modality was then separately batch-corrected using Harmony v3.8[108]. The batch variable provided to Harmony was dataset (1-16), with either the PCA or LSI reduction as input. The output of Harmony is a batch-corrected data dimension reduction (harmony_rna, harmony_atac). The two data modalities were integrated using Seurat's weighted nearest neighbor (**WNN**)[102] method using the Harmony dimension reductions as input. Cells were clustered using the smart local moving (**SLM**) algorithm. Other clustering algorithms (Louvain and Leiden) gave comparable results. After clustering, another round of peak calling was performed with MACS2 (see above parameters), with cells aggregated per cluster-timepoint. These peak sets were merged (union) to generate a penultimate reference peak set. The above ATAC-seq steps (normalizing, dimension reduction, harmony integration) and combination with transcriptome (wnn, clustering) were repeated to achieve the final clustering solution: 15 clusters and 65,862 cells. Cluster identities were determined based on gene markers and accessibility near effector genes and genome-scale analyses, as described in **Results** (**Fig. 1C-D**, **S2**). A cluster of "stressed" cells (3.1% of total) bore markers of cell stress (*TXNIP*, *PDCD4*), lacked activation markers (*CD44, IRF4*) and was excluded from downstream analyses.

Identification of "Donor0" spike-in control cells

To distinguish the spike-in control ("Donor0") from experimental donor cells in each multiome-seq experiment, we used genetic variant calls, available from our WGS of the spike-in control (Donor0) or inferred from the RNA and ATAC multiome modalities for the four experimental donors, using the standard GATK pipeline (v4.2.4.1)[109]. First, we identified high-confidence experimental donor cells from each sample using Souporcell v2.5[110]. Souporcell was run independently for the RNA and ATAC data, using Cell Ranger BAM files as input. To improve Souporcell speed and

performance, we provided a single nucleotide variant (**SNV**) reference of common variants occurring in at least 5% of individuals from the 1000 Genomes Project[111]; additional Souporcell parameters included: --skip_remap True, --ignore True, -k 2, --no_umi True (ATAC only). Second, to further improve genotype-based mapping of cells to experimental donors (lacking WGS), donor-cell predictions supported by both RNA and ATAC analyses were used to infer genotypes for the experimental donors, serving as valuable input to a second round of Souporcell analysis. For variant calling in these experimental donors, the Cell Ranger BAM files from each sample were filtered based on predicted experimental cell barcodes (supported by RNA and ATAC) and variants were called using the standard Freebayes v1.3.7 pipeline[112] applied to the RNA sequencing data. Finally, another round of Souporcell was performed on each sample, this time using the full gene expression BAM file and each donor's predicted VCF files as input.

Pseudobulk batch correction, data normalization and differential analyses

We aggregated gene expression or accessibility signal from individual cells to generate pseudobulk profiles per CD4+ T cell subset, time point and donor. Importantly, the spike-in control donor cells provided a pseudobulk sample condition ("Donor0", resting cells) that remained constant across all 16 multiome-seq experiments (donor-timepoints). This batch-control design is required to distinguish biological from technical variation across experiments, as standard batch correction methods (ComBat, DESeq2, EdgeR)[113–115] require at least one common control sample across experiments. However, the use of spike-in donor control cells is not common practice in sc-genomics study designs. Inclusion of a spike-in donor (8% of nuclei per experiment) (1) slightly reduces sequencing data available for the experimental donor-condition of interest (i.e., incurs expense) and (2) the processing of the spike-in cells introduces additional steps and therefore risk to sensitive nuclei prep protocols, which may not be feasible or acceptable for, e.g., challenging tissue samples or rare patient samples.

We therefore evaluated the impact of our spike-in study design on inference quality, both visually and quantitatively. As described above, single-cell Harmony batch correction was critical to resolving nuclei across experiments into cell populations (**Fig. S1A-C**). The pseudobulk gene expression counts resulting from Harmony batch-corrected cell integration and WNN clusters served as input to three pseudobulk batch-correction approaches: (1) no pseudobulk batch correction, (2) spike-in informed batch correction, where each of the 16 multiome-seq experiments was treated as a batch with a common pseudobulk (spike-in) sample, (3) batch correction ignoring the spike-in control pseudobulk (for we supplied an incorrect design in which the four timepoint experiments per donor were treated as a single batch ("batch = donor"), resulting in 4 batches that preserved timepoint information). Option (3) is commonplace in the absence of spike-in control cells, but problematic for several reasons: (a) Biological variability due to donor differences is modeled as technical and subtracted off as a batch effect. (b) Despite targeting similar numbers of nuclei for input to the 10x instrument, the number of quality nuclei and their per-nucleus library sizes are often variable across experiments. Lumping four experiments into a single batch, based on donor, makes it difficult to control for this important technical covariate (among others). For batch correction, we used Combat-Seq from sva v3.46.0, using "option 2" spike-in informed parameters (batch=dataset, group=celltype-timepoint) or "option 3" parameters (batch=donor, group=celltype-timepoint).

For data normalization and differential analysis, we used DESeq2 (v.1.38.3) and the variance stabilizing transformation (**VST**). We also leveraged the single-cell nature of our data for differential gene expression analysis, filtering genes that met our DESeq2 statistical cutoffs (FDR=10%, $\log_2|FC|>0.58$) but were detected in fewer than 5% of nuclei within one condition of a comparison (e.g., 2.5h activation versus resting in Th2 cells). As previously[38], this additional sc-

based differential criterion improved biological signal, removing low-quality differential genes supported by pseudobulk analysis alone. For differential analysis of the accessible chromatin data, we performed no filtering based on % nuclei with peak signal, because peak signal per nucleus is limited by DNA copy number and extreme signal from a small number of cells cannot occur. This rationale was supported by visual inspection of ATAC signal quantification (e.g., **Fig. 4A, 5A, S5A**). For differential analysis of ATAC signal, FDR=10%, $\log_2|FC|>0.58$ were also applied.

To visually compare the three pseudobulk batch-correction strategies, we identified the top 1000 most batch-dependent dynamic genes resulting from the "no batch correction" strategy for visualization. First, from the "no batch correction" analysis, dynamic genes were identified as those differential for at least one activation timepoint relative to t=0, for at least one cell type. Next, we identified those dynamic genes with the greatest batch effects, using the following strategy: (1) We calculated the coefficient of variation (**CV**) for each gene at each celltype_timepoint across the four donors, using VST counts from the "no batch correction" strategy. (2) For each gene, the CVs were averaged, and the 1000 genes with the highest average CV were visualized in the heatmap (**Fig. S1D-F**). As expected, performing no pseudobulk batch correction resulted in the largest batch effects. We quantitatively assessed the three pseudobulk batch-correction strategies by comparing the number of differential genes resulting from each strategy. In addition, we considered that the incorrect "batch = donor" assumption is only likely for experiments lacking spike-in control cells, and, in that case, a "batch = donor" design would benefit from additional experimental donor cells (i.e., a slight signal gain, as the experimental donors would contribute 100% (as opposed to 92%) of cells in the experiment). For this reason, we added a simulation condition to our comparisons, randomly replacing spike-in donor cells with experimental donor cells for each experiment. (Similar to bootstrapping, our up-sampling simulation underestimates gene variance, providing a challenging comparator for the spike-in strategy.) For all comparisons, the spike-in batch control strategy resulted in the most differentially expressed genes. The relative increase in DEGs was dependent on the strength of the biological signal: While the spike-in strategy increased activation-dependent DEGs by 8-12%, depending on cell population timepoint (**Fig. S1G**), the spike-in strategy increased DEGs between difficult-to-distinguish memory populations at rest by 20-30% (**Fig. S1H**). Not only do these results demonstrate the benefit of the spike-in design in our study, they provide support for this batch-control approach in future sc-genomics studies.

For the scATAC-seq data, the same processing steps were followed, with the exception that we did not run combat-seq to control for batch effects, as visual inspection of the data (e.g., analogously to **Fig. S1D** analysis) did not identify gross batch effects.

Pseudobulk samples that were too sparse (library size <150k for RNA, <400k for ATAC) were excluded from downstream analyses, with the following exception: at the 2.5h timepoint, no donor met the RNA or ATAC cutoff for the Th2 and Th17 populations. Given the importance of these populations at this time point, we took the following approach: Following batch correction with ComBat-seq (RNA) and proceeding DESeq2 normalization, RNA counts were combined for donors "1 and 2" and donors "3 and 4", resulting in two pseudobulk gene expression and chromatin accessibility estimates at 2.5h for Th17 and Th2 cells.

For visualization of pseudobulk chromatin accessibility tracks (e.g., **Fig. 1E**), data were library size normalized following the method used in Signac[116].

Dendrograms

Raw pseudobulk RNA and ATAC counts matrices were downsampled to 150,000 reads and 400,000 fragments, respectively, to account for differences in library sizes across populations.

Following the batch correction and normalization strategies describe above, downsampled counts matrices were batch-corrected (RNA only) and VST-normalized prior to z-scoring (both). Features with a coefficient of variation of < 0.01 for expression and < 0.001 for accessibility were filtered out, resulting 12,225 genes and 23,136 ATAC peaks. Hierarchical clustering was performed using *ComplexHeatmap* (v 2.14.0) with Ward's method and Euclidian distance.

Identification of rapid-recall genes and accessible chromatin regions

Rapid recall genes or peaks were defined as features with faster, more extreme activation dynamics (up or down) in memory populations than in naive cells. Therefore, a feature is rapid recall if:

1. It is dynamic in at least one memory population (differential at 2.5h, 5h or 15h relative to t=0h) AND
2. at the above dynamic timepoint(s) in the corresponding memory population(s), the feature is differential relative to naïve cells.
3. The sign of the $\log_2$(FC) estimates for (1) and (2) are the same (thereby ensuring that at least one memory population has faster, more extreme dynamics than naïve).

Differential analysis criteria are as described above.

GSEA

We conducted gene set enrichment analysis (**GSEA**) using the Gene Ontology (**GO**) Biological Processes pathways obtained from the Molecular Signatures Database (**mSigDB**, accessed April 8, 2024). Statistical significance was assessed using Fisher's exact test, with the analysis restricted to pathways containing a minimum of 5 genes. The background gene set comprised all genes present in the above pathways. Pathway enrichments were considered statistically significant at a (**FDR**) threshold of 10%. Multiple testing correction was applied using the Benjamini-Hochberg procedure.

Motif scanning and enrichment

We performed motif scanning using CIS-BP motifs (v2.0.0) on all peaks in our reference set (described above) using FIMO from Meme Suite version 5.4.1 (raw P-value threshold = 0.00001, first-order Markov model background)[117]. Enrichment of TF binding sites in peaks sets compared to background (all peaks) was calculated using a Fisher exact test (FDR = 10%) and Benjamini-Hochberg procedure to correct for multiple hypothesis testing.

Cell cycle scoring

To score cell-cycle phase for each cell, we downloaded genes associated with each cell phase from tinyatlas (update from 9/13/17)[118]. We then used Seurat's *CellCycleScoring* function to assign scores at the single-cell level, using the phase with the highest score to classify cells.

GRN model construction

Selection of Target Genes for GRN. We selected 15,169 target genes for GRN inference, based on differential expression (FDR=10%, $\log_2|FC|>0.58$) within CD4+ T cell subsets across timepoints, as well as across subsets at a particular timepoint.

Selection of Potential Regulators. We intersected a curated list of 1639 human TFs[119], with our list of target genes (above), yielding an initial set of 761 regulators. *RORC* did not meet our definition of target gene but was added as a potential regulator, given its known importance to CD4+ T cell biology[120]. *CTCF* was a target gene, but we removed it from consideration as a potential regulator[121], given its mode of regulation as an insulator. Thus, we considered a total of 761 potential regulators for GRN inference.

maxATAC prediction of TFBS. We used maxATAC (version 1.0.5)[46] to generate genome-wide *in silico* TF binding site (**TFBS**) predictions for the 110 human TFs that met our potential regulator criteria and had maxATAC models available: ARID3A, ARNT, ATF1, ATF2, ATF3, ATF7, BACH1, BHLHE40, CEBPB, CREB1, CREM, CUX1, E4F1, EGR1, ELF1, ELF4, ELK1, ESRRA, ETS1, ETV6, FOS, FOSL1, FOSL2, FOXK2, FOXP1, GABPA, GATA3, GATAD2B, HSF1, IKZF1, IRF3, JUN, JUNB, JUND, KMT2A, LEF1, MAFF, MAFK, MAX, MAZ, MBD2, MEF2A, MNT, MXI1, MYB, MYC, NFATC3, NFE2L2, NFIA, NFIC, NFXL1, NFYA, NFYB, NKRF, NR2C1, NR2C2, NRF1, PBX3, PKNOX1, RELA, REST, RFX1, RFX5, RUNX1, SETDB1, SKIL, SMAD5, SP1, SREBF1, SREBF2, SRF, STAT5A, TBP, TCF7, TCF7L2, TCF12, USF2, YBX1, YY1, ZBED1, ZBTB7A, ZBTB11, ZBTB33, ZBTB40, ZEB1, ZFX, ZHX2, ZKSCAN1, ZNF24, ZNF143, ZNF207, ZNF217, ZNF274, ZNF282, ZNF384, ZNF407, ZNF569, ZNF592, ZSCAN29, ZZZ3. We generated maxATAC predictions for each of the fourteen CD4 T populations in **Fig. 1B**, as they all met the recommended minimum library size (5M fragments) for maxATAC prediction from snATAC-seq data. We followed default data processing of ATAC signal for input to the pre-trained maxATAC models, to generate a total of 1540 genome-wide "*in silico* TF ChIP-seq" tracks (110 TFs x 14 CD4 populations in resting or activated contexts). To arrive at a final set of TFBS, we thresholded maxATAC scores according to the value that maximized the average validation F1-score reported per TF model[46].

Trac-loop association of putative regulatory regions (ATAC peaks) with target genes. Human CD4+ T cell Trac Loop data was downloaded from GSE87254[47]. We combined unique Trac regions from both resting and activated cells into one list. Genomic regions were lifted from hg19 to hg38 using UCSC's Lift Genome Annotation tool with a minimum ratio of remapped bases of 0.95 (6 of 38812 regions could not be mapped to Hg38). To infer enhancer-promoter interactions in our data, we overlapped our reference ATAC peak set with Trac regions directly connected to a gene promoter.

Prior network of TF-gene interactions inferred from ATAC data. The initial prior was based off the reference peak set previously described. Putative TF binding sites were predicted using maxATAC as described above. A TF-gene interaction was included in the prior if a predicted binding site for that TF was present in an ATAC peak within 10kb of the gene. In addition, interactions were included if the motif was present in a TRAC loop enhancer for the gene, even if the enhancer was >10kb from the gene of interest. Finally, for TFs with no maxATAC models available, TF binding site prediction was based on TF motif scanning (see above).

GRN inference. Gene expression is modeled as a multivariate linear combination of transcription factor activities (**TFAs**):

$$x_{ij} = \Sigma_{k \in TFs} b_{ik} a_{kj} \quad \textbf{(Equation 1)},$$

where x_{ij} corresponds to the expression of gene i in pseudobulk sample biological replicate j (e.g., "resting Th1, Donor 2"), a_{kj} is the activity of TF k in condition j, and b_{ik} describes the regulatory effect of TF k on gene i [38,40,122]). While gene expression is directly measured in our

multiome-seq study design, protein TF activities (a_{kj}) are not. Based on previous GRN benchmarking in CD4 T cells (bulk RNA-seq and ATAC-seq[40] and scRNA-seq and scATAC-seq[38]) we used two complementary TFA estimation strategies:

(1) Prior-based TFA assumes that protein TFAs can be estimated based on the expression of a TF's target genes, using partial prior knowledge of TF-gene interactions (e.g., as inferred from ATAC-seq and Trac-loop above). TFA is calculated from the normalized gene expression matrix $X \in \mathbb{R}^{|\text{genes}|\times|\text{samples}|}$ and the prior network matrix $P \in \mathbb{R}^{|\text{genes}|\times|TFs|}$using the least-squares solution of the equation[48]:

$$X = P \cdot \text{A} \quad \textbf{(Equation 2)},$$

(2) TF mRNA serves as an estimate of protein TFA; this is a reasonable assumption for some TFs, whose major mode of regulation is transcriptional (as opposed to post-transcriptional).

For each TFA estimation strategy separately, TF-gene regulatory interactions, b_{ik}, were identified via optimization of the following objective function[40]:

$$\text{B} = \text{argmin}|X - \text{BA}|_2^2 + \Lambda|\text{B}|_1 \quad \textbf{(Equation 3)},$$

where $X \in \mathbb{R}^{|\text{genes}|\times|\text{samples}|}$ is the gene expression matrix, $A \in \mathbb{R}^{|\text{TFs}|\times|\text{samples}|}$ is the TF activity matrix, $B \in \mathbb{R}^{|\text{genes}|\times|TFs|}$ is the matrix of TF-gene interaction coefficients, and $\Lambda \in \mathbb{R}^{|\text{genes}|\times|TFs|}$ is the matrix of penalties associated with each interaction. The penalty matrix incorporates prior information via a decreased penalty (Λ_{ik}) for interactions with prior support ($\Lambda_{ik} = 0.5$ with prior support, $\Lambda_{ik} = 1$ for no prior support). While the adaptive LASSO penalty encourages selection of TF-gene interactions supported by the prior, novel interactions, strongly supported by the gene expression model, can be inferred. In this way the two sources of error in the prior (false positive and false negative interactions) can be corrected through the inference procedure. The ability to infer novel interactions is especially critical for TFs lacking a maxATAC or TF motif model (e.g., TET2) and therefore ATAC-derived prior information. The LASSO regularization parameter λ using StARS[123], with 100 subsamples and a subsample fraction of 0.63. Confidence scores for TF-gene interactions were defined as the sum of (1) subsamples in which the edge was nonzero and (2), to break ties between equally stable edges, the correlation between the TF's activity and its predicted target gene's TFA[40]. We evaluated the GRN's predictive performance (predictive R^2 on left-out test gene expression data) as a function of model complexity (average number of regulators per gene)[40]; this analysis identified an average of 10 regulators per target gene as optimal for GRNs derived using both prior-based TFA and TF mRNA. Confidences from these two networks were rank-combined using the maximum[40], then the top 151,140 interactions were selected and reranked and signed, using partial correlation between genes and their TF regulators to determine interaction sign.

GRN analyses

Final transcription factor activity estimates were calculated by substituting the signed, ranked confidences from the final GRN for the prior matrix in **Equation 3**.

Functional annotation of TFs using the GRN. (e.g., **Fig S3D**) To determine whether a TF activated (or repressed) gene pathways, we tested for significant overlap between its positive (or negative) target genes and gene pathways, using GSEA (described above).

GRN subnetworks encoding TCM/TEM (e.g., **Fig. 3G**). We selected TFs that were differentially expressed between TCM/TEM and TCM in the same direction at all 3 activation timepoints (FDR = 10%). Target genes for this network included genes upregulated in TCM/TEM

compared to TCM in at least one activation timepoint.

GRN subnetworks encoding memory at rest. (e.g., **Fig. 5F**) TFs predicted to encode resting memory were prioritized based on several criteria:

(1) The TF's motif is enriched in a peak cluster with increased accessibility in memory versus naïve cells (e.g., **Fig. 5A** clusters Th1/Th17/CTL, Th2/MHCII)

(2) The TF is differentially expressed between at least one memory population and naïve cells at rest

(3) The TF has at least two predicted upregulated target genes (positive edge) that show differential expression between naïve cells and at least one memory population at rest, where the expression is higher in the memory population.

Nine TFs met these criteria: PRDM1, MYBL1, RORA, KLF6, RUNX2, NR3C1, SMAD3, MAF and RBPJ. Target genes were upregulated in memory cells compared to naive at rest and involved in the biological pathways enriched in memory upregulated genes (**Fig. 5E**).

Disease GRN subnetworks. We constructed GRNs centered on the accessible chromatin peak set most significantly overlapping the "Th2/MHCII" or "Treg" phenotype variant sets (green and blue boxes in **Fig .7A**), to arrive at a Th2/MHCII (**Fig. 7D**) and Treg (**Fig. S6A**) disease GRNs. We first collected all risk variants with an overlapping peak from the peak-cluster of interest (Th2/MHCII or Shared Treg). These peaks were linked to genes using the method described in *Prior network of TF-gene interactions inferred from ATAC data*, with the following exceptions (1) we included promoter-enhancer interactions from micro-C of CD4 T cells (in addition to Trac-loop) [47,84] and (2) the proximity rule was based on a more stringent criterion +/- 1kb promoter (**Table S9**). We combined both the resting and activated micro-C data and included loops detected from both algorithms (HiCCUPS and mustache)[124,125]. Any SNP-linked target genes detected in < 5% cells of the relevant cell-type (Th2/MHCII or Treg) were excluded from the analysis. We then visualized the expression dynamics of the target genes as well as the activities of all predicted TF regulators from the GRN. Target genes and their TF regulators were retained in the GRN if their expression and TFA were elevated in the cell-type of interest at rest or their activation dynamics were specific to the cell-type of interest.

Modeling chromatin accessibility as a function of TF activities

We built a novel multivariate model to infer TF regulation of activation-dependent chromatin accessibility (**Fig. 4A, Fig. S5A**). TF motif enrichment analyses nominate potential TF regulators of accessibility patterns, but do not directly leverage important relationships between TF activities and chromatin accessibility signals. Post-hoc assessment of this relationship is non-trivial, when considering the multivariate, context-specific interactions of TFs across the 122 pseudobulk ATAC profiles (e.g., as shown in **Fig. 4A**). Thus, we explicitly modeled chromatin accessibility as a function of GRN-derived protein TF activities, using regularized regression to incorporate the TFBS predictions.

Similar to GRN inference, we modeled chromatin accessibility as a multivariate linear function of TF activities, constructing a model for each dynamic accessibility pattern in our dataset:

$$y_{ij} = \Sigma_{k \in TFs} c_{ik} a_{kj} \quad \textbf{(Equation 4)},$$

where $y_{ij} \in \mathbb{R}^{|\text{peak set}| \times |\text{sample}|}$ corresponds to pattern i in pseudobulk sample biological replicate j (e.g., "resting Th1, Donor 2"), $a_{kj} \in R^{|\text{TF}| \times |\text{sample}|}$ is the activity of TF k in condition j, as estimated from the final GRN (described above), and $c_{ik} \in \mathbb{R}^{|\text{TF}| \times |\text{peak set}|}$ describes the regulatory effect of

TF k on accessibility pattern i. In total, we constructed models for 12 dynamic accessibility patterns, pertaining to the median peak accessibility per cluster (7 RR clusters from **Fig. 4A**, and 5 naïve and memory-shared activation-dependent clusters from **Fig. S5A**).

The prior information was derived from motif-based TFBS prediction in each peak cluster (parameters and motif database as described above). We first counted the fraction of peaks per cluster with at least one binding site and calculated enrichment relative to the background of all reference peaks (one-sided Fisher exact test). For each cluster independently, each TF is ranked according to p-value, yielding adaptive LASSO penalty weighting factors $\gamma_{ik} \in [0,1]$ in which the most enriched TF k receives the smallest weight and therefore smallest penalty:

$$\text{penalty} = \gamma_{ik} \cdot \lambda_{ik} \qquad \textbf{(Equation 5)}$$

The resulting $\Lambda_{ik} \in \mathbb{R}^{|\text{peak set}| \times |\text{TF}|}$ penalty matrix is used to guide network inference using an adaptive elastic net objective function:

$$\text{C} = \text{argmin}|Y - \text{CA}|_2^2 + \Lambda\left(\alpha|\text{C}|_1 + \frac{1-\alpha}{2}|\text{C}|_2\right) \qquad \textbf{(Equation 6)},$$

where $Y \in \mathbb{R}^{|\text{peak set}| \times |\text{samples}|}$ is the accessibility matrix, $A \in \mathbb{R}^{|\text{TFs}| \times |\text{samples}|}$ is the activity matrix, $\Lambda \in \mathbb{R}^{|\text{genes}| \times |TFs|}$ is the matrix of penalties associated with each interaction and $\alpha \in [0,1]$ is parameter that controls the relative importance of the ridge (L2) versus LASSO (L1) penalty. A small L2 penalty stabilizes the optimization, and we set $\alpha = 0.99$ to prioritize the L1 penalty and model sparsity[126]. Regularized regression was performed with GLMNet v0.7.0 in julia. λ_{ik} selection and network interaction confidences were estimated using StARS, as described for GRN inference.

Visualization of Accessibility Networks. (e.g., **Fig. 4D**) We visualized any regulatory interaction between TF and peak cluster observed in 75 out of 100 subsamples.

GWAS trait genetic risk variants and chromatin accessibility co-localization analysis

Mirroring the approach in[80], we utilized a curated set of 538 human diseases and phenotypes that had ≥20 independent risk loci from GWAS in European ancestry[127]. We removed the 136 phenotypes associated with lab-collected values, leaving 387 for testing (**Table S8**). We applied the Regulatory Element Locus Intersection (**RELI**) algorithm[77] to identify CD4 T-cell context-specific chromatin accessibility regions that significantly co-localized with genetic risk variant sets [80]. Bonferroni correction was applied to account for multiple hypothesis testing. Disease phenotype variant and chromatin accessibility sets were reported as significantly co-localized, if the enrichment achieved $P_{adj} < 0.1$ and was supported by peak-variant overlaps in at least three independent risk loci. In **Fig. 7A**, phenotypes were clustered using k-means and a distance metric based on Pearson correlation.

Acknowledgements

This project was supported by National Institute of Health (**NIH**) funding: R01AI153442 (AB, ERM), U01AI150748 (ERM, LCK), R01AI173314 (ERM, LCK), P30AR070549 (LCK). We thank Yuriy Baglaenko, David Hildeman, Matthew Weirauch and Claire Chougnet for helpful feedback on the study. We thank the CCRF Single Cell Genomics Facility (RRID:SCR_022653) for assistance optimizing the multiome-seq assay for this study. Some of our images were created with BioRender.com (**Fig. 1A**).

Author contributions

AB and ERM conceived of the study. AB and ERM designed the experiments, which were executed and optimized by SK. AK and ERM designed the computational analyses, which were executed by AK with support from ATB, SMO, ZFY, AR, JAW and MK. AK, SF, AB and ERM analyzed the experimental data and modeling results. AK, SF, AB and ERM wrote the manuscript. All authors provided feedback on the manuscript.

Declaration of interests

A.B. is a co-founder of Datirium, LLC, the developer of SciDAP.com. M.K. and A.B. developed CWL-Airflow, which is licensed to Datirium, LLC.

Data availability

Raw and processed multiome-seq data and WGS have been deposited in GEO under accession number GSE282266 (Multiome) and GSE287942 (WGS).

References

1. Mosmann, T.R., Cherwinski, H., Bond, M.W., Giedlin, M.A., and Coffman, R.L. (1986). Two types of murine helper T cell clone. I. Definition according to profiles of lymphokine activities and secreted proteins. The Journal of Immunology *136*. https://doi.org/10.4049/jimmunol.136.7.2348.
2. Ivanov, I.I., Atarashi, K., Manel, N., Brodie, E.L., Shima, T., Karaoz, U., Wei, D., Goldfarb, K.C., Santee, C.A., Lynch, S. V, et al. (2009). Induction of intestinal Th17 cells by segmented filamentous bacteria. Cell *139*, 485–498. https://doi.org/10.1016/j.cell.2009.09.033.
3. Van De Veerdonk, F.L., and Netea, M.G. (2010). T-cell subsets and antifungal host defenses. Preprint, https://doi.org/10.1007/s12281-010-0034-6 https://doi.org/10.1007/s12281-010-0034-6.
4. Sallusto, F. (2016). Heterogeneity of Human CD4(+) T Cells Against Microbes. Annu Rev Immunol *34*, 317–334. https://doi.org/10.1146/ANNUREV-IMMUNOL-032414-112056.
5. Breitfeld, D., Ohl, L., Kremmer, E., Ellwart, J., Sallusto, F., Lipp, M., and Förster, R. (2000). Follicular B helper T cells express CXC chemokine receptor 5, localize to B cell follicles, and support immunoglobulin production. J Exp Med *192*, 1545–1552. https://doi.org/10.1084/jem.192.11.1545.
6. Schaerli, P., Willimann, K., Lang, A.B., Lipp, M., Loetscher, P., and Moser, B. (2000). CXC chemokine receptor 5 expression defines follicular homing T cells with B cell helper function. J Exp Med *192*, 1553–1562. https://doi.org/10.1084/jem.192.11.1553.
7. Johnston, R.J., Poholek, A.C., DiToro, D., Yusuf, I., Eto, D., Barnett, B., Dent, A.L., Craft, J., and Crotty, S. (2009). Bcl6 and Blimp-1 are reciprocal and antagonistic regulators of T follicular helper cell differentiation. Science *325*, 1006–1010. https://doi.org/10.1126/science.1175870.
8. Kanamori, M., Nakatsukasa, H., Okada, M., Lu, Q., and Yoshimura, A. (2016). Induced Regulatory T Cells: Their Development, Stability, and Applications. Preprint, https://doi.org/10.1016/j.it.2016.08.012 https://doi.org/10.1016/j.it.2016.08.012.
9. Sprent, J., and Surh, C.D. (2002). T cell memory. Annu Rev Immunol *20*, 551–579. https://doi.org/10.1146/annurev.immunol.20.100101.151926.
10. Kaech, S.M., Wherry, E.J., and Ahmed, R. (2002). Effector and memory T-cell differentiation: Implications for vaccine development. Preprint, https://doi.org/10.1038/nri778 https://doi.org/10.1038/nri778.
11. Hammarlund, E., Lewis, M.W., Hansen, S.G., Strelow, L.I., Nelson, J.A., Sexton, G.J., Hanifin, J.M., and Slifka, M.K. (2003). Duration of antiviral immunity after smallpox vaccination. Nat Med *9*. https://doi.org/10.1038/nm917.
12. Mehlhop-Williams, E.R., and Bevan, M.J. (2014). Memory CD8+ T cells exhibit increased antigen threshold requirements for recall proliferation. Journal of Experimental Medicine *211*. https://doi.org/10.1084/jem.20131271.
13. Croft, M., Bradley, L.M., and Swain, S.L. (1994). Naive versus memory CD4 T cell response to antigen. Memory cells are less dependent on accessory cell costimulation and can respond to many antigen-presenting cell types including resting B cells. The Journal of Immunology *152*. https://doi.org/10.4049/jimmunol.152.6.2675.
14. Farber, D.L., Acuto, O., and Bottomly, K. (1997). Differential T cell receptor-mediated signaling in naive and memory CD4 T cells. Eur J Immunol *27*. https://doi.org/10.1002/eji.1830270838.
15. Hussain, S.F., Anderson, C.F., and Farber, D.L. (2002). Differential SLP-76 Expression and TCR-Mediated Signaling in Effector and Memory CD4 T Cells. The Journal of

Immunology *168*. https://doi.org/10.4049/jimmunol.168.4.1557.
16. Kersh, E.N., Kaech, S.M., Onami, T.M., Moran, M., Wherry, E.J., Miceli, M.C., and Ahmed, R. (2003). TCR signal transduction in antigen-specific memory CD8 T cells. J Immunol *170*, 5455–5463. https://doi.org/10.4049/JIMMUNOL.170.11.5455.
17. Lai, W., Yu, M., Huang, M.-N., Okoye, F., Keegan, A.D., and Farber, D.L. (2011). Transcriptional Control of Rapid Recall by Memory CD4 T Cells. The Journal of Immunology. https://doi.org/10.4049/jimmunol.1002742.
18. Macián, F., López-Rodríguez, C., and Rao, A. (2001). Partners in transcription: NFAT and AP-1. Preprint, https://doi.org/10.1038/sj.onc.1204386 https://doi.org/10.1038/sj.onc.1204386.
19. Rao, A. (1994). NF-ATp: a transcription factor required for the co-ordinate induction of several cytokine genes. Immunol Today *15*. https://doi.org/10.1016/0167-5699(94)90007-8.
20. Teijaro, J.R., Turner, D., Pham, Q., Wherry, E.J., Lefrançois, L., and Farber, D.L. (2011). Cutting Edge: Tissue-Retentive Lung Memory CD4 T Cells Mediate Optimal Protection to Respiratory Virus Infection. The Journal of Immunology *187*. https://doi.org/10.4049/jimmunol.1102243.
21. Oeckinghaus, A., and Ghosh, S. (2009). The NF-kappaB family of transcription factors and its regulation. Preprint, https://doi.org/10.1101/cshperspect.a000034 https://doi.org/10.1101/cshperspect.a000034.
22. Cuddapah, S., Barski, A., and Zhao, K. (2010). Epigenomics of T cell activation, differentiation, and memory. Curr Opin Immunol *22*, 341–347. https://doi.org/10.1016/j.coi.2010.02.007.
23. Barski, A., Cuddapah, S., Kartashov, A. V, Liu, C., Imamichi, H., Yang, W., Peng, W., Lane, H.C., and Zhao, K. (2017). Rapid Recall Ability of Memory T cells is Encoded in their Epigenome. Sci Rep *7*, 39785. https://doi.org/10.1038/srep39785.
24. Avni, O., Lee, D., Macian, F., Szabo, S.J., Glimcher, L.H., and Rao, A. (2002). Th cell differentiation is accompanied by dynamic changes in histone acetylation of cytokine genes. Nat Immunol *3*. https://doi.org/10.1038/ni808.
25. Fields, P.E., Kim, S.T., and Flavell, R.A. (2002). Cutting Edge: Changes in Histone Acetylation at the IL-4 and IFN-γ Loci Accompany Th1/Th2 Differentiation . The Journal of Immunology *169*. https://doi.org/10.4049/jimmunol.169.2.647.
26. Fitzpatrick, D.R., Shirley, K.M., and Kelso, A. (1999). Cutting Edge: Stable Epigenetic Inheritance of Regional IFN-γ Promoter Demethylation in CD44highCD8+ T Lymphocytes. The Journal of Immunology *162*. https://doi.org/10.4049/jimmunol.162.9.5053.
27. Yukawa, M., Jagannathan, S., Vallabh, S., Kartashov, A. V., Chen, X., Weirauch, M.T., and Barski, A. (2020). AP-1 activity induced by co-stimulation is required for chromatin opening during T cell activation. Journal of Experimental Medicine *217*. https://doi.org/10.1084/JEM.20182009.
28. Sen, R., and Baltimore, D. (1986). Multiple nuclear factors interact with the immunoglobulin enhancer sequences. Cell *46*, 705–716. https://doi.org/10.1016/0092-8674(86)90346-6.
29. Darnell, J.E., Kerr, Ian M., and Stark, G.R. (1994). Jak-STAT Pathways and Transcriptional Activation in Response to IFNs and Other Extracellular Signaling Proteins. Science (1979) *264*. https://doi.org/10.1126/science.8197455.
30. Vahedi, G., Takahashi, H., Nakayamada, S., Sun, H.W., Sartorelli, V., Kanno, Y., and OShea, J.J. (2012). STATs shape the active enhancer landscape of T cell populations. Cell *151*. https://doi.org/10.1016/j.cell.2012.09.044.

31. Joulin, V., Bories, D., Eléouet, J.F., Labastie, M.C., Chrétien, S., Mattéi, M.G., and Roméo, P.H. (1991). A T-cell specific TCR delta DNA binding protein is a member of the human GATA family. EMBO J *10*. https://doi.org/10.1002/j.1460-2075.1991.tb07706.x.
32. Hirose, T., Smith, R.J., and Jetten, A.M. (1994). ROR-γ: The Third Member of ROR/RZR Orphan Receptor Subfamily That Is Highly Expressed in Skeletal Muscle. Biochem Biophys Res Commun *205*. https://doi.org/10.1006/bbrc.1994.2902.
33. Zhou, L., Lopes, J.E., Chong, M.M.W., Ivanov, I.I., Min, R., Victora, G.D., Shen, Y., Du, J., Rubtsov, Y.P., Rudensky, A.Y., et al. (2008). TGF-B-induced Foxp3 inhibits TH17 cell differentiation by antagonizing RORγt function. Nature *453*, 236–240. https://doi.org/10.1038/nature06878.
34. Fontenot, J.D., Gavin, M.A., and Rudensky, A.Y. (2003). Foxp3 programs the development and function of CD4+CD25+ regulatory T cells. Nat Immunol *4*, 330–336. https://doi.org/10.1038/ni904.
35. WASYLYK, B., HAHN, S.L., and GIOVANE, A. (1993). The Ets family of transcription factors. Eur J Biochem *211*. https://doi.org/10.1111/j.1432-1033.1993.tb19864.x.
36. Egawa, T., Tillman, R.E., Naoe, Y., Taniuchi, I., and Littman, D.R. (2007). The role of the Runx transcription factors in thymocyte differentiation and in homeostasis of naive T cells. Journal of Experimental Medicine *204*, 1945–1957. https://doi.org/10.1084/jem.20070133.
37. Gergen, J.P., and Butler, B.A. (1988). Isolation of the Drosophila segmentation gene runt and analysis of its expression during embryogenesis. Genes Dev *2*. https://doi.org/10.1101/gad.2.9.1179.
38. Wayman, J.A., Thomas, A., Bejjani, A., Katko, A., Almanan, M., Godarova, A., Korinfskaya, S., Cazares, T.A., Yukawa, M., Kottyan, L.C., et al. (2023). An atlas of gene regulatory networks for T memory cells in youth and old age. bioRxiv.
39. Henriksson, J., Chen, X., Gomes, T., Ullah, U., Meyer, K.B., Miragaia, R., Duddy, G., Pramanik, J., Yusa, K., Lahesmaa, R., et al. (2019). Genome-wide CRISPR Screens in T Helper Cells Reveal Pervasive Crosstalk between Activation and Differentiation. Cell *176*. https://doi.org/10.1016/j.cell.2018.11.044.
40. Miraldi, E.R., Pokrovskii, M., Watters, A., Castro, D.M., De Veaux, N., Hall, J.A., Lee, J.-Y., Ciofani, M., Madar, A., Carriero, N., et al. (2019). Leveraging chromatin accessibility for transcriptional regulatory network inference in T Helper 17 Cells. Genome Res *29*, 449–463. https://doi.org/10.1101/gr.238253.118.
41. Gustafsson, M., Gawel, D.R., Alfredsson, L., Baranzini, S., Bjorkander, J., Blomgran, R., Hellberg, S., Eklund, D., Ernerudh, J., Kockum, I., et al. (2015). A validated gene regulatory network and GWAS identifies early regulators of T cell-associated diseases. Sci Transl Med *7*. https://doi.org/10.1126/scitranslmed.aad2722.
42. Yosef, N., Shalek, A.K., Gaublomme, J.T., Jin, H., Lee, Y., Awasthi, A., Wu, C., Karwacz, K., Xiao, S., Jorgolli, M., et al. (2013). Dynamic regulatory network controlling TH17 cell differentiation. Nature *496*, 461–468. https://doi.org/10.1038/nature11981.
43. Ciofani, M., Madar, A., Galan, C., Sellars, M., MacE, K., Pauli, F., Agarwal, A., Huang, W., Parkurst, C.N., Muratet, M., et al. (2012). A validated regulatory network for Th17 cell specification. Cell *151*. https://doi.org/10.1016/j.cell.2012.09.016.
44. Squair, J.W., Gautier, M., Kathe, C., Anderson, M.A., James, N.D., Hutson, T.H., Hudelle, R., Qaiser, T., Matson, K.J.E., Barraud, Q., et al. (2021). Confronting false discoveries in single-cell differential expression. Nat Commun *12*. https://doi.org/10.1038/s41467-021-25960-2.
45. Heumos, L., Schaar, A.C., Lance, C., Litinetskaya, A., Drost, F., Zappia, L., Lücken, M.D., Strobl, D.C., Henao, J., Curion, F., et al. (2023). Best practices for single-cell analysis

across modalities. Nat Rev Genet *24*. https://doi.org/10.1038/s41576-023-00586-w.
46. Cazares, T.A., Rizvi, F.W., Iyer, B., Chen, X., Kotliar, M., Bejjani, A.T., Wayman, J.A., Donmez, O., Wronowski, B., Parameswaran, S., et al. (2023). maxATAC: Genome-scale transcription-factor binding prediction from ATAC-seq with deep neural networks. PLoS Comput Biol *19*, e1010863. https://doi.org/10.1371/journal.pcbi.1010863.
47. Lai, B., Tang, Q., Jin, W., Hu, G., Wangsa, D., Cui, K., Stanton, B.Z., Ren, G., Ding, Y., Zhao, M., et al. (2018). Trac-looping measures genome structure and chromatin accessibility. Nat Methods *15*, 741–747. https://doi.org/10.1038/s41592-018-0107-y.
48. Arrieta-Ortiz, M.L., Hafemeister, C., Bate, A.R., Chu, T., Greenfield, A., Shuster, B., Barry, S.N., Gallitto, M., Liu, B., Kacmarczyk, T., et al. (2015). An experimentally supported model of the Bacillus subtilis global transcriptional regulatory network . Mol Syst Biol *11*. https://doi.org/10.15252/msb.20156236.
49. Glasmacher, E., Agrawal, S., Chang, A.B., Murphy, T.L., Zeng, W., Lugt, B. Vander, Khan, A.A., Ciofani, M., Spooner, C.J., Rutz, S., et al. (2012). A genomic regulatory element that directs assembly and function of immune-specific AP-1 - IRF complexes. Science (1979) *338*. https://doi.org/10.1126/science.1228309.
50. Sahoo, A., Alekseev, A., Tanaka, K., Obertas, L., Lerman, B., Haymaker, C., Clise-Dwyer, K., McMurray, J.S., and Nurieva, R. (2015). Batf is important for IL-4 expression in T follicular helper cells. Nat Commun *6*, 7997. https://doi.org/10.1038/ncomms8997.
51. Clipstone, N.A., and Crabtree, G.R. (1992). Identification of calcineurin as a key signalling enzyme in T-lymphocyte activation. Nature *357*. https://doi.org/10.1038/357695a0.
52. Jain, J., Loh, C., and Rao, A. (1995). Transcriptional regulation of the IL-2 gene. Curr Opin Immunol *7*, 333–342. https://doi.org/10.1016/0952-7915(95)80107-3.
53. Alberdi, M., Iglesias, M., Tejedor, S., Merino, R., López-Rodríguez, C., and Aramburu, J. (2017). Context-dependent regulation of Th17-associated genes and IFNγ expression by the transcription factor NFAT5. Immunol Cell Biol *95*. https://doi.org/10.1038/icb.2016.69.
54. Khan, O., Giles, J.R., McDonald, S., Manne, S., Ngiow, S.F., Patel, K.P., Werner, M.T., Huang, A.C., Alexander, K.A., Wu, J.E., et al. (2019). TOX transcriptionally and epigenetically programs CD8+ T cell exhaustion. Nature *571*, 211–218. https://doi.org/10.1038/s41586-019-1325-x.
55. Alfei, F., Kanev, K., Hofmann, M., Wu, M., Ghoneim, H.E., Roelli, P., Utzschneider, D.T., von Hoesslin, M., Cullen, J.G., Fan, Y., et al. (2019). TOX reinforces the phenotype and longevity of exhausted T cells in chronic viral infection. Nature *571*, 265–269. https://doi.org/10.1038/s41586-019-1326-9.
56. Scott, A.C., Dündar, F., Zumbo, P., Chandran, S.S., Klebanoff, C.A., Shakiba, M., Trivedi, P., Menocal, L., Appleby, H., Camara, S., et al. (2019). TOX is a critical regulator of tumour-specific T cell differentiation. Nature *571*, 270–274. https://doi.org/10.1038/s41586-019-1324-y.
57. Canaria, D.A., Rodriguez, J.A., Wang, L., Yeo, F.J., Yan, B., Wang, M., Campbell, C., Kazemian, M., and Olson, M.R. (2023). Tox induces T cell IL-10 production in a BATF-dependent manner. Front Immunol *14*. https://doi.org/10.3389/fimmu.2023.1275423.
58. Utzschneider, D.T., Delpoux, A., Wieland, D., Huang, X., Lai, C.Y., Hofmann, M., Thimme, R., and Hedrick, S.M. (2018). Active Maintenance of T Cell Memory in Acute and Chronic Viral Infection Depends on Continuous Expression of FOXO1. Cell Rep *22*. https://doi.org/10.1016/j.celrep.2018.03.020.
59. Rutishauser, R.L., Martins, G.A., Kalachikov, S., Chandele, A., Parish, I.A., Meffre, E., Jacob, J., Calame, K., and Kaech, S.M. (2009). Blimp-1 promotes terminal differentiation of virus-specific CD8 T cells and represses the acquisition of central memory T cell properties. Immunity *31*.

60. Lam, N., Lee, Y., and Farber, D.L. (2024). A guide to adaptive immune memory. Nat Rev Immunol *24*, 810–829. https://doi.org/10.1038/s41577-024-01040-6.
61. Sallusto, F., Lenig, D., Förster, R., Lipp, M., and Lanzavecchia, A. (1999). Two subsets of memory T lymphocytes with distinct homing potentials and effector functions. Nature *401*, 708–712. https://doi.org/10.1038/44385.
62. Raphael, I., Joern, R.R., and Forsthuber, T.G. (2020). Memory CD4+ T Cells in Immunity and Autoimmune Diseases. Cells *9*. https://doi.org/10.3390/CELLS9030531.
63. Wherry, E.J., Teichgräber, V., Becker, T.C., Masopust, D., Kaech, S.M., Antia, R., von Andrian, U.H., and Ahmed, R. (2003). Lineage relationship and protective immunity of memory CD8 T cell subsets. Nat Immunol *4*, 225–234. https://doi.org/10.1038/ni889.
64. Rivino, L., Messi, M., Jarrossay, D., Lanzavecchia, A., Sallusto, F., and Geginat, J. (2004). Chemokine receptor expression identifies pre-T helper (Th)1, pre-Th2, and nonpolarized cells among human CD4+ central memory T cells. Journal of Experimental Medicine *200*. https://doi.org/10.1084/jem.20040774.
65. Barski, A., Jothi, R., Cuddapah, S., Cui, K., Roh, T.-Y., Schones, D.E., and Zhao, K. (2009). Chromatin poises miRNA- and protein-coding genes for expression. Genome Res *19*, 1742–1751. https://doi.org/10.1101/gr.090951.109.
66. Kazi, A., Xiang, S., Yang, H., Delitto, D., Trevino, J., Jiang, R.H.Y., Ayaz, M., Lawrence, H.R., Kennedy, P., and Sebti, S.M. (2018). GSK3 suppression upregulates β-catenin and c-Myc to abrogate KRas-dependent tumors. Nat Commun *9*. https://doi.org/10.1038/s41467-018-07644-6.
67. Geginat, J., Sallusto, F., and Lanzavecchia, A. (2001). Cytokine-driven proliferation and differentiation of human naive, central memory, and effector memory CD4+ T cells. Journal of Experimental Medicine *194*. https://doi.org/10.1084/jem.194.12.1711.
68. Li, J., Li, L., Sun, X., Deng, T., Huang, G., Li, X., Xie, Z., and Zhou, Z. (2021). Role of Tet2 in Regulating Adaptive and Innate Immunity. Front Cell Dev Biol *9*. https://doi.org/10.3389/FCELL.2021.665897.
69. Baessler, A., Fuchs, B., Perkins, B., Richens, A.W., Novis, C.L., Harrison-Chau, M., Sircy, L.M., Thiede, K.A., and Hale, J.S. (2023). Tet2 deletion in CD4+ T cells disrupts Th1 lineage commitment in memory cells and enhances T follicular helper cell recall responses to viral rechallenge. Proc Natl Acad Sci U S A *120*. https://doi.org/10.1073/PNAS.2218324120.
70. Schaerli, P., Willimann, K., Lang, A.B., Lipp, M., Loetscher, P., and Moser, B. (2000). CXC chemokine receptor 5 expression defines follicular homing T cells with B cell helper function. Journal of Experimental Medicine *192*. https://doi.org/10.1084/jem.192.11.1553.
71. Chevalier, N., Jarrossay, D., Ho, E., Avery, D.T., Ma, C.S., Yu, D., Sallusto, F., Tangye, S.G., and Mackay, C.R. (2011). CXCR5 Expressing Human Central Memory CD4 T Cells and Their Relevance for Humoral Immune Responses. The Journal of Immunology *186*. https://doi.org/10.4049/jimmunol.1002828.
72. Bevington, S.L., Cauchy, P., and Cockerill, P.N. (2017). Chromatin priming elements establish immunological memory in T cells without activating transcription. BioEssays. https://doi.org/10.1002/bies.201600184.
73. Liu, S.Q.J., and Golan, D.E. (1999). T-cell stimulation through the T-cell receptor/CD3 complex regulates CD2 lateral mobility by a calcium/calmodulin-dependent mechanism. Biophys J *76*. https://doi.org/10.1016/S0006-3495(99)77327-3.
74. Ichiyama, K., Chen, T., Wang, X., Yan, X., Kim, B.-S., Tanaka, S., Ndiaye-Lobry, D., Deng, Y., Zou, Y., Zheng, P., et al. (2015). The methylcytosine dioxygenase Tet2 promotes DNA demethylation and activation of cytokine gene expression in T cells. Immunity *42*, 613–626. https://doi.org/10.1016/j.immuni.2015.03.005.

75. Cruz-Guilloty, F., Pipkin, M.E., Djuretic, I.M., Levanon, D., Lotem, J., Lichtenheld, M.G., Groner, Y., and Rao, A. (2009). Runx3 and T-box proteins cooperate to establish the transcriptional program of effector CTLs. J Exp Med *206*, 51. https://doi.org/10.1084/JEM.20081242.
76. Calderon, D., Nguyen, M.L.T., Mezger, A., Kathiria, A., Müller, F., Nguyen, V., Lescano, N., Wu, B., Trombetta, J., Ribado, J. V., et al. (2019). Landscape of stimulation-responsive chromatin across diverse human immune cells. Nat Genet *51*. https://doi.org/10.1038/s41588-019-0505-9.
77. Harley, J.B., Chen, X., Pujato, M., Miller, D., Maddox, A., Forney, C., Magnusen, A.F., Lynch, A., Chetal, K., Yukawa, M., et al. (2018). Transcription factors operate across disease loci, with EBNA2 implicated in autoimmunity. Nat Genet *50*, 699–707. https://doi.org/10.1038/s41588-018-0102-3.
78. Soskic, B., Cano-Gamez, E., Smyth, D.J., Ambridge, K., Ke, Z., Matte, J.C., Bossini-Castillo, L., Kaplanis, J., Ramirez-Navarro, L., Lorenc, A., et al. (2022). Immune disease risk variants regulate gene expression dynamics during CD4+ T cell activation. Nat Genet *54*. https://doi.org/10.1038/s41588-022-01066-3.
79. Daga, K., Agarwal, S., Moti, Z., Lee, M.B.K., Din, M., Wood, D., Modat, M., and Booth, T.C. (2024). Machine Learning Algorithms to Predict the Risk of Rupture of Intracranial Aneurysms: a Systematic Review. Clin Neuroradiol. https://doi.org/10.1007/s00062-024-01474-4.
80. Wayman, J.A., Yang, Z., Angerman, E., Bonkowski, E., Jurickova, I., Chen, X., Bejjani, A.T., Parks, L., Parameswaran, S., Miethke, A.G., et al. (2024). Accessible chromatin maps of inflammatory bowel disease intestine nominate cell-type mediators of genetic disease risk. bioRxiv.
81. Farh, K.K.-H., Marson, A., Zhu, J., Kleinewietfeld, M., Housley, W.J., Beik, S., Shoresh, N., Whitton, H., Ryan, R.J.H., Shishkin, A.A., et al. (2015). Genetic and epigenetic fine mapping of causal autoimmune disease variants. Nature *518*, 337–343. https://doi.org/10.1038/nature13835.
82. Wang, R., Li, B., Huang, B., Li, Y., Liu, Q., Lyu, Z., Chen, R., Qian, Q., Liang, X., Pu, X., et al. (2024). Gut Microbiota-Derived Butyrate Induces Epigenetic and Metabolic Reprogramming in Myeloid-Derived Suppressor Cells to Alleviate Primary Biliary Cholangitis. Gastroenterology *167*, 733-749.e3. https://doi.org/10.1053/j.gastro.2024.05.014.
83. Kudira, R., Yang, Z.F., Osuji, I., Karns, R., Bariya, P., Pfuhler, L., Mullen, M., Taylor, A., Ji, H., Lages, C.S., et al. (2024). Cholestasis alters polarization and suppressor function of hepatic regulatory T cells. bioRxiv. https://doi.org/10.1101/2024.05.17.594680.
84. Oguchi, A., Suzuki, A., Komatsu, S., Yoshitomi, H., Bhagat, S., Son, R., Bonnal, R.J.P., Kojima, S., Koido, M., Takeuchi, K., et al. (2024). An atlas of transcribed enhancers across helper T cell diversity for decoding human diseases. Science *385*, eadd8394. https://doi.org/10.1126/science.add8394.
85. Zhang, M., Qin, Z., Huang, Y., Tian, W., Li, Y., Wang, C., Zhao, W., Dai, Y., Shi, X., Gershwin, M.E., et al. (2024). Association of CCR6 functional polymorphisms with Primary Biliary Cholangitis. J Transl Autoimmun *8*, 100234. https://doi.org/10.1016/j.jtauto.2024.100234.
86. Varona, R., Cadenas, V., Flores, J., Martínez-A, C., and Márquez, G. (2003). CCR6 has a non-redundant role in the development of inflammatory bowel disease. Eur J Immunol *33*, 2937–2946. https://doi.org/10.1002/eji.200324347.
87. Roghani, S.A., Lotfi, R., Soroush, M.G., Khorasanizadeh, A., Feizollahi, P., Assar, S., Soufivand, P., Pournazari, M., Mohammadi Kish, Z., and Taghadosi, M. (2023).

Increased gene expression of CCR6 and RORγt in peripheral blood cells of rheumatoid arthritis patients and their correlation with anti-cyclic citrullinated peptide and disease activity. Immun Inflamm Dis *11*, e1112. https://doi.org/10.1002/iid3.1112.
88. Shi, Z., Garcia-Melchor, E., Wu, X., Getschman, A.E., Nguyen, M., Rowland, D.J., Wilson, M., Sunzini, F., Akbar, M., Huynh, M., et al. (2021). Targeting the CCR6/CCL20 Axis in Entheseal and Cutaneous Inflammation. Arthritis Rheumatol *73*, 2271–2281. https://doi.org/10.1002/art.41882.
89. Wang, C., Kang, S.G., Lee, J., Sun, Z., and Kim, C.H. (2009). The roles of CCR6 in migration of Th17 cells and regulation of effector T-cell balance in the gut. Mucosal Immunol *2*, 173–183. https://doi.org/10.1038/mi.2008.84.
90. Duren, Z., Chen, X., Jiang, R., Wang, Y., and Wong, W.H. (2017). Modeling gene regulation from paired expression and chromatin accessibility data. Proc Natl Acad Sci U S A *114*. https://doi.org/10.1073/pnas.1704553114.
91. Abdelsamed, H.A., Moustaki, A., Fan, Y., Dogra, P., Ghoneim, H.E., Zebley, C.C., Triplett, B.M., Sekaly, R.-P., and Youngblood, B. (2017). Human memory CD8 T cell effector potential is epigenetically preserved during in vivo homeostasis. J Exp Med *214*, 1593–1606. https://doi.org/10.1084/jem.20161760.
92. Durek, P., Nordström, K., Gasparoni, G., Salhab, A., Kressler, C., de Almeida, M., Bassler, K., Ulas, T., Schmidt, F., Xiong, J., et al. (2016). Epigenomic Profiling of Human CD4+ T Cells Supports a Linear Differentiation Model and Highlights Molecular Regulators of Memory Development. Immunity *45*, 1148–1161. https://doi.org/10.1016/J.IMMUNI.2016.10.022.
93. Bakshi, R., Hassan, M.Q., Pratap, J., Lian, J.B., Montecino, M.A., Van Wijnen, A.J., Stein, J.L., Imbalzano, A.N., and Stein, G.S. (2010). The human SWI/SNF complex associates with RUNX1 to control transcription of hematopoietic target genes. J Cell Physiol *225*, 569–576. https://doi.org/10.1002/jcp.22240.
94. Bevington, S.L., Cauchy, P., Piper, J., Bertrand, E., Lalli, N., Jarvis, R.C., Gilding, L.N., Ott, S., Bonifer, C., and Cockerill, P.N. (2016). Inducible chromatin priming is associated with the establishment of immunological memory in T cells. EMBO J *35*, 515–535. https://doi.org/10.15252/embj.201592534.
95. Guo, H., and Friedman, A.D. (2011). Phosphorylation of RUNX1 by cyclin-dependent kinase reduces direct interaction with HDAC1 and HDAC3. Journal of Biological Chemistry *286*, 208–215. https://doi.org/10.1074/jbc.M110.149013.
96. Lichtinger, M., Ingram, R., Hannah, R., Müller, D., Clarke, D., Assi, S.A., Lie-A-Ling, M., Noailles, L., Vijayabaskar, M.S., Wu, M., et al. (2012). RUNX1 reshapes the epigenetic landscape at the onset of haematopoiesis. EMBO Journal *31*, 4318–4333. https://doi.org/10.1038/emboj.2012.275.
97. Zhang, F., Meng, G., and Strober, W. (2008). Interactions among the transcription factors Runx1, RORgammat and Foxp3 regulate the differentiation of interleukin 17-producing T cells. Nat Immunol *9*, 1297–1306. https://doi.org/10.1038/ni.1663.
98. McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M., et al. (2010). The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res *20*, 1297–1303. https://doi.org/10.1101/gr.107524.110.
99. DePristo, M.A., Banks, E., Poplin, R., Garimella, K. V, Maguire, J.R., Hartl, C., Philippakis, A.A., del Angel, G., Rivas, M.A., Hanna, M., et al. (2011). A framework for variation discovery and genotyping using next-generation DNA sequencing data. Nat Genet *43*, 491–498. https://doi.org/10.1038/ng.806.
100. Van der Auwera, G.A., Carneiro, M.O., Hartl, C., Poplin, R., del Angel, G., Levy-

Moonshine, A., Jordan, T., Shakir, K., Roazen, D., Thibault, J., et al. (2013). From FastQ Data to High-Confidence Variant Calls: The Genome Analysis Toolkit Best Practices Pipeline. Curr Protoc Bioinformatics *43*. https://doi.org/10.1002/0471250953.bi1110s43.

101. Kotliar, M., Kartashov, A., and Barski, A. (2025). Accelerating Single-Cell Sequencing Data Analysis with SciDAP: A User-Friendly Approach. Methods Mol Biol *2880*, 255–292. https://doi.org/10.1007/978-1-0716-4276-4_13.
102. Hao, Y., Hao, S., Andersen-Nissen, E., Mauck, W.M., Zheng, S., Butler, A., Lee, M.J., Wilk, A.J., Darby, C., Zager, M., et al. (2021). Integrated analysis of multimodal single-cell data. Cell *184*. https://doi.org/10.1016/j.cell.2021.04.048.
103. Granja, J.M., Corces, M.R., Pierce, S.E., Bagdatli, S.T., Choudhry, H., Chang, H.Y., and Greenleaf, W.J. (2021). ArchR is a scalable software package for integrative single-cell chromatin accessibility analysis. Nat Genet *53*, 403–411. https://doi.org/10.1038/s41588-021-00790-6.
104. Germain, P.-L., Lun, A., Garcia Meixide, C., Macnair, W., and Robinson, M.D. (2021). Doublet identification in single-cell sequencing data using scDblFinder. F1000Res *10*, 979. https://doi.org/10.12688/f1000research.73600.2.
105. CellRanger Tutorial https://www.10xgenomics.com/support/software/cell-ranger/latest/tutorials.
106. Gaspar, J.M. (2018). Improved peak-calling with MACS2. bioRxiv.
107. Zhang, Y., Liu, T., Meyer, C.A., Eeckhoute, J., Johnson, D.S., Bernstein, B.E., Nusbaum, C., Myers, R.M., Brown, M., Li, W., et al. (2008). Model-based analysis of ChIP-Seq (MACS). Genome Biol *9*, R137. https://doi.org/10.1186/gb-2008-9-9-r137.
108. Korsunsky, I., Millard, N., Fan, J., Slowikowski, K., Zhang, F., Wei, K., Baglaenko, Y., Brenner, M., Loh, P. ru, and Raychaudhuri, S. (2019). Fast, sensitive and accurate integration of single-cell data with Harmony. Nat Methods *16*. https://doi.org/10.1038/s41592-019-0619-0.
109. Betschart, R.O., Thiéry, A., Aguilera-Garcia, D., Zoche, M., Moch, H., Twerenbold, R., Zeller, T., Blankenberg, S., and Ziegler, A. (2022). Comparison of calling pipelines for whole genome sequencing: an empirical study demonstrating the importance of mapping and alignment. Sci Rep *12*. https://doi.org/10.1038/s41598-022-26181-3.
110. Heaton, H., Talman, A.M., Knights, A., Imaz, M., Gaffney, D.J., Durbin, R., Hemberg, M., and Lawniczak, M.K.N. (2020). Souporcell: robust clustering of single-cell RNA-seq data by genotype without reference genotypes. Nat Methods *17*. https://doi.org/10.1038/s41592-020-0820-1.
111. 1000 Genomes Project Consortium, Abecasis, G.R., Altshuler, D., Auton, A., Brooks, L.D., Durbin, R.M., Gibbs, R.A., Hurles, M.E., and McVean, G.A. (2010). A map of human genome variation from population-scale sequencing. Nature *467*, 1061–1073. https://doi.org/10.1038/nature09534.
112. Garrison, E., and Marth, G. (2012). Haplotype-based variant detection from short-read sequencing -- Free bayes -- Variant Calling -- Longranger. arXiv preprint arXiv:1207.3907.
113. Zhang, Y., Parmigiani, G., and Johnson, W.E. (2020). ComBat-seq: Batch effect adjustment for RNA-seq count data. NAR Genom Bioinform *2*. https://doi.org/10.1093/nargab/lqaa078.
114. Johnson, W.E., Li, C., and Rabinovic, A. (2007). Adjusting batch effects in microarray expression data using empirical Bayes methods. Biostatistics *8*. https://doi.org/10.1093/biostatistics/kxj037.
115. Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol *15*.

https://doi.org/10.1186/s13059-014-0550-8.
116. Stuart, T., Srivastava, A., Madad, S., Lareau, C.A., and Satija, R. (2021). Single-cell chromatin state analysis with Signac. Nat Methods *18*, 1333–1341. https://doi.org/10.1038/s41592-021-01282-5.
117. Grant, C.E., Bailey, T.L., and Noble, W.S. (2011). FIMO: scanning for occurrences of a given motif. Bioinformatics *27*, 1017–1018. https://doi.org/10.1093/bioinformatics/btr064.
118. Kirchner, R., and HBC (2018). Tinyatlas.
119. Lambert, S.A., Jolma, A., Campitelli, L.F., Das, P.K., Yin, Y., Albu, M., Chen, X., Taipale, J., Hughes, T.R., and Weirauch, M.T. (2018). The Human Transcription Factors. Cell *172*, 650–665. https://doi.org/10.1016/j.cell.2018.01.029.
120. Ivanov, I.I., McKenzie, B.S., Zhou, L., Tadokoro, C.E., Lepelley, A., Lafaille, J.J., Cua, D.J., and Littman, D.R. (2006). The orphan nuclear receptor RORgammat directs the differentiation program of proinflammatory IL-17+ T helper cells. Cell *126*, 1121–1133. https://doi.org/10.1016/j.cell.2006.07.035.
121. Bell, A.C., West, A.G., and Felsenfeld, G. (1999). The protein CTCF is required for the enhancer blocking activity of vertebrate insulators. Cell *98*, 387–396. https://doi.org/10.1016/s0092-8674(00)81967-4.
122. Bonneau, R., Reiss, D.J., Shannon, P., Facciotti, M., Hood, L., Baliga, N.S., and Thorsson, V. (2006). The Inferelator: an algorithm for learning parsimonious regulatory networks from systems-biology data sets de novo. Genome Biol *7*, R36. https://doi.org/10.1186/gb-2006-7-5-r36.
123. Liu, H., Roeder, K., and Wasserman, L. (2010). Stability Approach to Regularization Selection (StARS) for High Dimensional Graphical Models. Adv Neural Inf Process Syst *24*, 1432–1440.
124. Wingett, S., Ewels, P., Furlan-Magaril, M., Nagano, T., Schoenfelder, S., Fraser, P., and Andrews, S. (2015). HiCUP: pipeline for mapping and processing Hi-C data. F1000Res *4*, 1310. https://doi.org/10.12688/f1000research.7334.1.
125. Roayaei Ardakany, A., Gezer, H.T., Lonardi, S., and Ay, F. (2020). Mustache: multi-scale detection of chromatin loops from Hi-C and Micro-C maps using scale-space representation. Genome Biol *21*, 256. https://doi.org/10.1186/s13059-020-02167-0.
126. Friedman, J., Hastie, T., and Tibshirani, R. (2010). Regularization Paths for Generalized Linear Models via Coordinate Descent. J Stat Softw *33*, 1–22.
127. Cerezo, M., Sollis, E., Ji, Y., Lewis, E., Abid, A., Bircan, K.O., Hall, P., Hayhurst, J., John, S., Mosaku, A., et al. (2025). The NHGRI-EBI GWAS Catalog: standards for reusability, sustainability and diversity. Nucleic Acids Res *53*, D998–D1005. https://doi.org/10.1093/nar/gkae1070.

Figure Legends

Fig. 1. Resolution of naïve, TCM and TEM populations across resting and TCR-activation timepoints for GRN inference. (A) Experimental design: CD4+ T cells were activated using anti-CD3/CD28 beads and collected at four time points: resting, 2.5h, 5h, and 15h, for a total of 16 multiome-seq experiments (4 timepoints x 4 donors). To control for batch effects, cells from a fifth, "Ctrl", donor (t=0) were spiked into each experiment immediately before lysis. (**B)** UMAP visualization, integrating transcriptome and chromatin accessibility from the resulting 65,862 high-quality nuclei. "**Act**" refers to activated cells from all activated timepoints (2.5-15h), combined. **(C)** Overview of the *Inferelator* for GRN inference from RNA-seq and ATAC-seq data. TF binding site (**TFBS**) prediction from ATAC-seq was accomplished using deep neural network models (**maxATAC**), when available, or TF motif scanning. Putative TFBS were linked to target genes using both promoter-enhancer interactions derived from Trac-loop data and linear proximity. The resulting "prior" of ATAC-derived TF-gene interactions enables estimation of protein TF activities and guides gene expression modeling (GRN inference). **(D)** Expression of select marker genes (transcripts per million (**TPM**), $\log_2$-normalized relative to the mean expression of each gene). **(E)** Normalized chromatin accessibility (see **Methods**) at select marker genes. (Panel A was created with biorender.com.)

Fig. 2. Rapid-recall transcriptional responses and their predicted TF regulators. (A) RR genes exhibited faster, higher-magnitude activation dynamics in memory compared to naive populations (n = 5831 genes, **Methods**). Donor, cell-type and timepoint-resolved pseudobulk gene expression values (VST counts) were z-scored and clustered. **(B)** Median gene expression profile per cluster, as a function of timepoint, for naïve, TCM and TEM populations; "TEM" represents the combined expression of Th1, Th2 and Th17 populations. **(C)** Select GO Biological Pathway enrichments per cluster ($P_{adj} < 0.1$). (No enrichments were identified for the TEM Down cluster.) GRN subnetworks for select gene pathways enriched in the **(D)** "Early + Middle" and **(E)** "TEM Up" RR clusters. TFs are depicted as grey boxes and target genes as circles, colored according to their associated GO term. Target genes are limited to those (1) belonging to the given cluster and GO pathway and (2) regulated by at least one predicted TF regulator of the cluster (**Methods**). Solid arrows represent interactions supported by both gene expression modeling and the ATAC-derived "prior" network, while dashed arrows depict interactions supported by gene expression modeling alone.

Fig. 3. Characterization of the activated TCM compartment identifies proliferative and TEM potentials. (A) The proliferating TCM ("Prolif. TCM") are a transcriptional state, identified via clustering of the transcriptome data alone, (RNA UMAP, upper left; light gray dots correspond to naïve or resting populations). UMAP visualization of activated memory cells by transcriptome alone (left lower panels) or accessibility alone (lower right panels) reveal that the proliferating TCM are a transcriptional state accessed by both TCM and TCM/TEM populations; upper panels show all activated memory populations, while lower panels highlight proliferating TCM, with other populations grayed out. **(B)** GSEA of each memory population's 15-hour RR genes. The TCM

population has the strongest enrichments in cell-cycle associated pathways. (Note TCM/TEM are excluded from this analysis, as they were not observed in resting conditions and rapid-recall genes are defined relative to the resting timepoint.) **(C)** The gene expression dynamics of rapid-recall genes associated with cell-cycle. "P." stands for "TCM-proliferating". Line plots represent the average gene expression per population and cluster, as a function of time. **(D)** Literature-based regulatory network of MYC supported by our data and the GRN (see text). The TF activities of MYC and its binding partner ZBTB17 are increased in TCM relative to other populations. Consistent with the TFA prediction, MYC's activating target gene *CCND2* is increased, and its repressed target *CDKN1B* is decreased in activated TCM. *GSK3B*, an inhibitor of MYC, is rapidly downregulated in TCM. **(E)** TCM/TEM (yellow) have increased expression of several cytokines (*IFNG*, *IL21*, *TNF*) compared to TCM (blue); Th1 expression (green) is shown for reference. **(F)** Normalized accessibility signal at selected regions differential between activated TCM and TCM/TEM (2.5h, 5h, 15h combined); $*P_{adj} < 0.1$, FC > 1.5. **(G)** Core GRN for the TCM/TEM population, for select TFs and genes upregulated in TCM/TEM relative to TCM across the activation timepoints. Node color represents the fold change between TCM/TEM and TCM, averaged across the activation timepoints. **(H)** Fold changes of select TFs between TCM/TEM and TCM; $*P_{adj} < 0.1$. The maximum VST counts for TCM/TEM and TCM (any activation timepoint) are displayed.

Fig. 4. Chromatin accessibility rapid-recall dynamics and predicted TF drivers. (A) 46,134 accessible chromatin regions exhibit rapid recall dynamics (i.e., faster, higher-amplitude response to activation in memory than naïve cells, **Methods**). **(B)** The average accessibility signal per cluster, for naïve, TCM, and TEM populations. (TEM combines Th1, Th2 and Th17 signal.) **(C)** Normalized ATAC signal at peaks with early open (*CD58*), middle open (*VIM*), middle open TEM (*MAF*), and rapid closed TEM (*LEF1*) dynamics. **(D)** A network model describing the TF regulation of activation-dependent chromatin accessibility (**Methods**). The network includes both RR peak clusters (**Fig. 4A**) and peaks with similar activation dynamics in memory and naïve populations (**Fig. S5A**). TF node colors indicate the relative difference in TFA between memory and naïve populations, at each of the four timepoints. Red indicates average higher activity in memory cells, while blue indicates higher activity in naive.

Fig. 5. **TF regulation of memory-associated accessible chromatin regions and genes at rest. (A)** Clustering of chromatin accessibility regions differential between naive and at least one memory population at rest (n = 32,219 peaks). **(B)** Putative TF regulators per accessibility cluster in (A), where TFs (1) are differentially expressed between resting naive and memory populations, with a fold-change that positively correlates with the cluster's accessibility signal, and (2) have motif enrichment in the corresponding peak cluster. **(C)** Normalized ATAC signal at the *IFNG* locus for naïve, Th1, Th2 and Th17 cells. The blue box highlights a RR peak at the *IFNG* promoter, with accessibility in Th1 and, to a lesser extent, Th17 cells. Green and gold boxes highlight accessibility in *IFNG* distal regulatory regions (supported by Trac-loop); these peaks show opposite trends: RR opening (green box) and RR closing (gold box). Zoom-in boxes show accessibility in resting Th1 but not naive populations. Violin plots display *IFNG* expression. **(D)**

For each RR or memory-naïve shared activation peak sets (**Fig. 4A, S5A**), the number of peaks that were also differentially in resting memory and naïve cells (A) is indicated. Red (or blue) asterisk indicates significant enrichment of "memory-up" (or "naïve-up") peak sets by Fisher exact test, $P_{adj} < 0.1$. **(E)** 1907 genes were differentially expressed between naïve and at least one memory population at rest. **(F)** GRN subnetwork of "resting memory maintenance", highlighting TFs and target genes in resting memory cells predicted to contribute to rapid recall. Target genes are colored by function. Red arrows are activating interactions, blue arrows are inhibitory.

Fig. 6. Enhanced activation kinetics in a subset of naive activated cells. WNN UMAP of all naïve CD4 T cells colored by **(A)** predicted pseudotime, **(B)** experimental timepoint, and **(C, D)** average accessibility of peaks from either the naïve-associated or memory-associated peaks at rest (from **Fig. 5A**, **Methods**). Scores > 0 indicate above-average peak-set accessibility, relative to all other CD4 T cells. A subset of late pseudotime "precocious" naïve cells (blue box) shows positive "Memory Up" scores. **(E)** Per-cell "Memory Up" and "Naïve Up" accessibility scores and expression ($\log_2$(TPM)) of select TFs across naïve-cell pseudotime. **(F)** Density plots of TF expression in naïve T cells. **(G)** Co-clustering of gene expression and summed proximal accessibility for 849 genes with differential expression or proximal accessibility between precocious naïve (defined as "Memory Up" module score > 0.025) and other activated naïve cells at 15h ($P_{adj} < 0.1$).

Fig. 7. Integrated GWAS-GRN analyses nominate subset- and context-specific disease regulatory mechanisms in CD4 T cells. **(A)** Significant colocalizations of GWAS phenotype genetic risk variant sets and accessible chromatin regions (resting memory- and naïve-associated, RR and those with naïve-memory shared activation dynamics); $^*P_{adj} < 0.1$ (**Methods**). The clustered phenotypes are annotated based on the peak sets with the most enrichments (e.g., "Treg"). Green and blue boxes indicate risk variant sets selected for downstream GRN analysis. The size of each peak set is indicated. **(B)** The PSC risk variant (rs12529876), within the *FGFR1OP* locus, is linked to the *CCR6* promoter via track-looping 131kb away. The variant overlaps a RR peak ("middle TEM" dynamics) that is also a resting memory-associated peak ("TEM+TCM shared" cluster). **(C)** The accessibility of the distal *CCR6* enhancer in (B) correlates with *CCR6* expression ($R = 0.65$, $P < 0.01$). **(D)** GRNs centered on target genes linked (via trac-loop, micro-C or promoter proximity) to genetic risk variants in the "Th2/MHCII cluster" (green box in A). Only genes detected in >5% of Th2 or MHCII cells were included. TFs were included if they exhibited Th2 or MHCII-specific activity at rest or post-activation. Outlined nodes represent Th2 of MHCII RR genes. (**E**) Peak accessibility at risk variants (**F**) target gene expression and (**G**) TF activity of regulators. For panels E-G, the left heatmap represents the resting state (t=0h), while the right heatmap shows activated dynamics. Font colors indicate Th2-specific (purple) and MHCII-specific (blue) peak accessibility (E), gene expression (F) and TFA (G).

Fig. 1

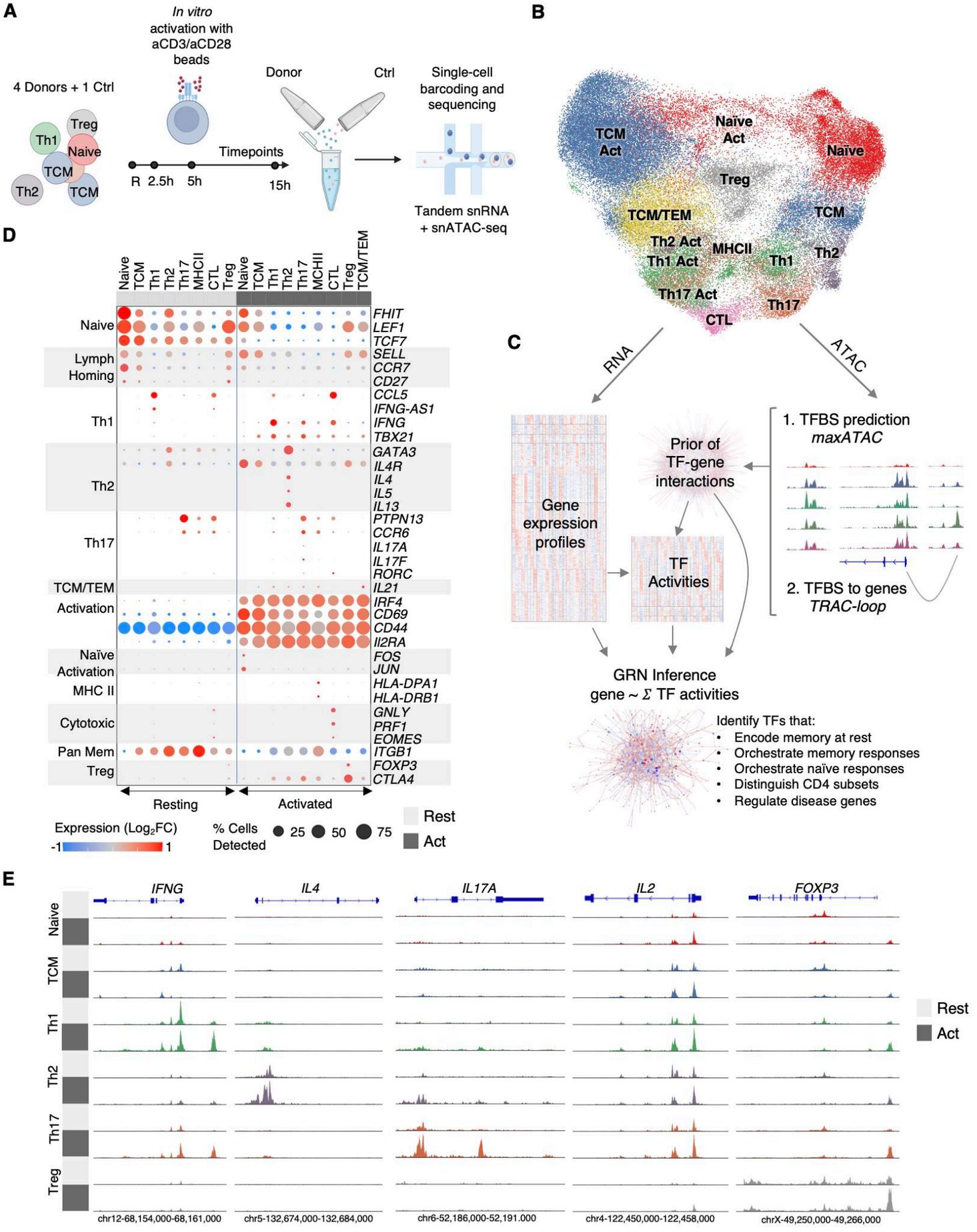

A
In vitro activation with aCD3/aCD28 beads
4 Donors + 1 Ctrl
Th1
Treg
Naive
TCM
Th2
TCM
Timepoints
R 2.5h 5h 15h
Donor
Ctrl
Single-cell barcoding and sequencing
Tandem snRNA + snATAC-seq
B
TCM Act
Naïve Act
Naïve
Treg
TCM/TEM
TCM
Th2 Act
Th1 Act
MHCII
Th1
Th2
Th17 Act
Th17
CTL
C
RNA
ATAC
Gene expression profiles
Prior of TF-gene interactions
TF Activities
1. TFBS prediction maxATAC
2. TFBS to genes TRAC-loop
GRN Inference
gene ~ Σ TF activities
Identify TFs that:
• Encode memory at rest
• Orchestrate memory responses
• Orchestrate naïve responses
• Distinguish CD4 subsets
• Regulate disease genes
D
Naive TCM Th1 Th2 Th17 MHCII CTL Treg Naive TCM Th1 Th2 Th17 MCHII CTL Treg TCM/TEM
Naive
Lymph Homing
Th1
Th2
Th17
TCM/TEM
Activation
Naïve Activation
MHC II
Cytotoxic
Pan Mem
Treg
FHIT
LEF1
TCF7
SELL
CCR7
CD27
CCL5
IFNG-AS1
IFNG
TBX21
GATA3
IL4R
IL4
IL5
IL13
PTPN13
CCR6
IL17A
IL17F
RORC
IL21
IRF4
CD69
CD44
Il2RA
FOS
JUN
HLA-DPA1
HLA-DRB1
GNLY
PRF1
EOMES
ITGB1
FOXP3
CTLA4
Resting
Activated
Expression (Log2FC)
-1
1
% Cells Detected
25
50
75
Rest
Act
E
IFNG
IL4
IL17A
IL2
FOXP3
Naive
TCM
Th1
Th2
Th17
Treg
chr12-68,154,000-68,161,000
chr5-132,674,000-132,684,000
chr6-52,186,000-52,191.000
chr4-122,450,000-122,458,000
chrX-49,250,000-49,266,000
Rest
Act

Fig. 2

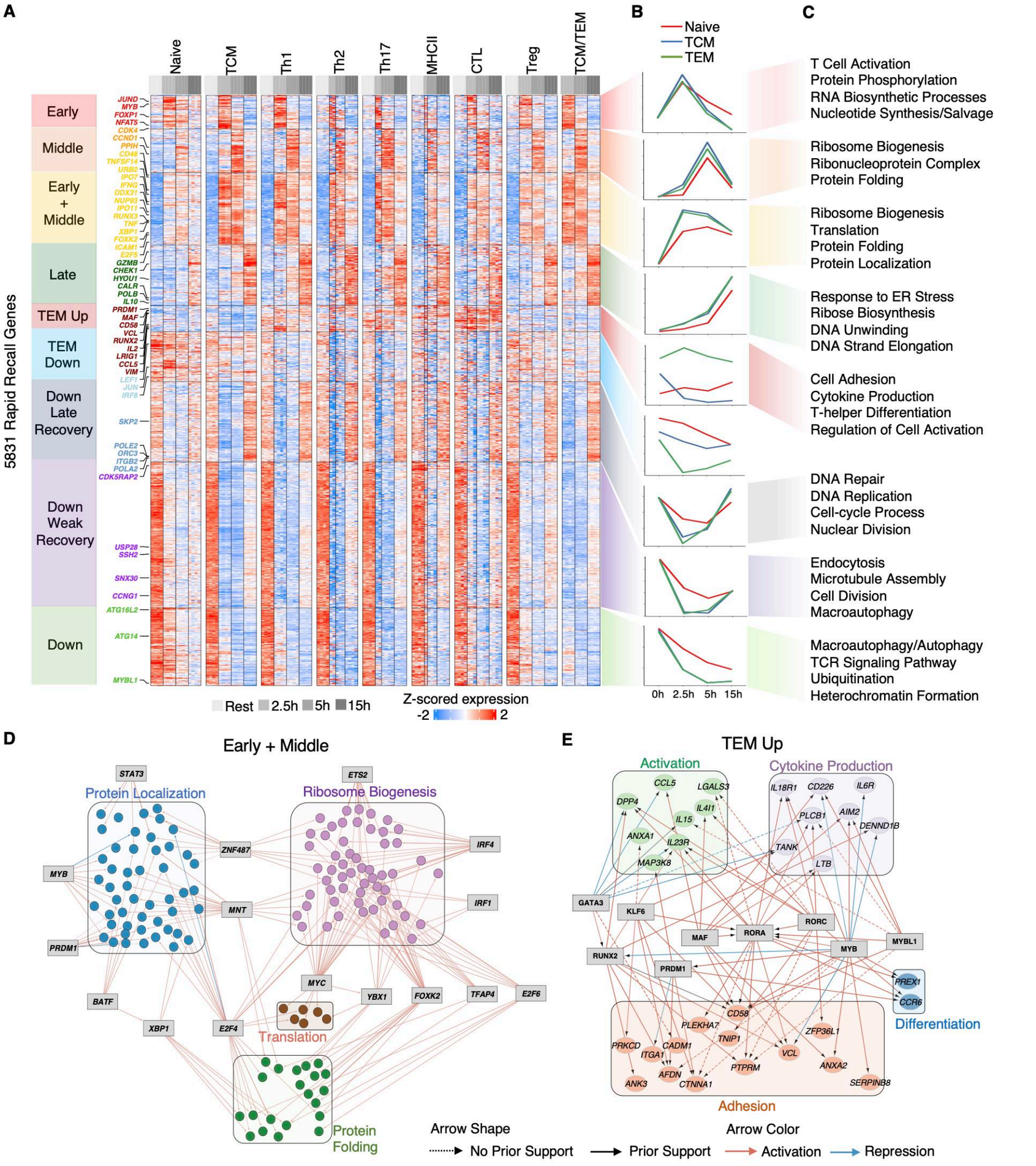

A
Naive
TCM
Th1
Th2
Th17
MHCII
CTL
Treg
TCM/TEM
5831 Rapid Recall Genes
Early
Middle
Early + Middle
Late
TEM Up
TEM Down
Down Late Recovery
Down Weak Recovery
Down
JUND
MYB
FOXP1
NFAT5
CDK4
CCND1
PPIH
CD48
TNFSF14
URB2
IPO7
IFNG
DDX31
NUP93
IPO11
RUNX3
TNF
XBP1
FOXK2
ICAM1
E2F5
GZMB
CHEK1
HYOU1
CALR
POLB
IL10
PRDM1
MAF
CD58
VCL
RUNX2
IL2
LRIG1
CCL5
VIM
LEF1
JUN
IRF8
SKP2
POLE2
ORC3
ITGB2
POLA2
CDK5RAP2
USP28
SSH2
SNX30
CCNG1
ATG16L2
ATG14
MYBL1
Rest
2.5h
5h
15h
Z-scored expression
-2
2
B
Naive
TCM
TEM
0h
2.5h
5h
15h
C
T Cell Activation
Protein Phosphorylation
RNA Biosynthetic Processes
Nucleotide Synthesis/Salvage
Ribosome Biogenesis
Ribonucleoprotein Complex
Protein Folding
Ribosome Biogenesis
Translation
Protein Folding
Protein Localization
Response to ER Stress
Ribose Biosynthesis
DNA Unwinding
DNA Strand Elongation
Cell Adhesion
Cytokine Production
T-helper Differentiation
Regulation of Cell Activation
DNA Repair
DNA Replication
Cell-cycle Process
Nuclear Division
Endocytosis
Microtubule Assembly
Cell Division
Macroautophagy
Macroautophagy/Autophagy
TCR Signaling Pathway
Ubiquitination
Heterochromatin Formation
D
Early + Middle
Protein Localization
Ribosome Biogenesis
STAT3
ETS2
ZNF487
IRF4
MYB
MNT
IRF1
PRDM1
MYC
YBX1
FOXK2
TFAP4
E2F6
BATF
XBP1
E2F4
Translation
Protein Folding
E
TEM Up
Activation
Cytokine Production
CCL5
LGALS3
DPP4
IL4I1
IL15
ANXA1
IL23R
MAP3K8
IL18R1
CD226
IL6R
PLCB1
AIM2
DENND1B
TANK
LTB
GATA3
KLF6
RORC
MAF
RORA
MYBL1
MYB
RUNX2
PRDM1
PREX1
CCR6
Differentiation
CD58
PLEKHA7
TNIP1
ZFP36L1
PRKCD
CADM1
ITGA1
VCL
PTPRM
ANXA2
AFDN
ANK3
CTNNA1
SERPINB8
Adhesion
Arrow Shape
No Prior Support
Prior Support
Arrow Color
Activation
Repression

Fig. 3

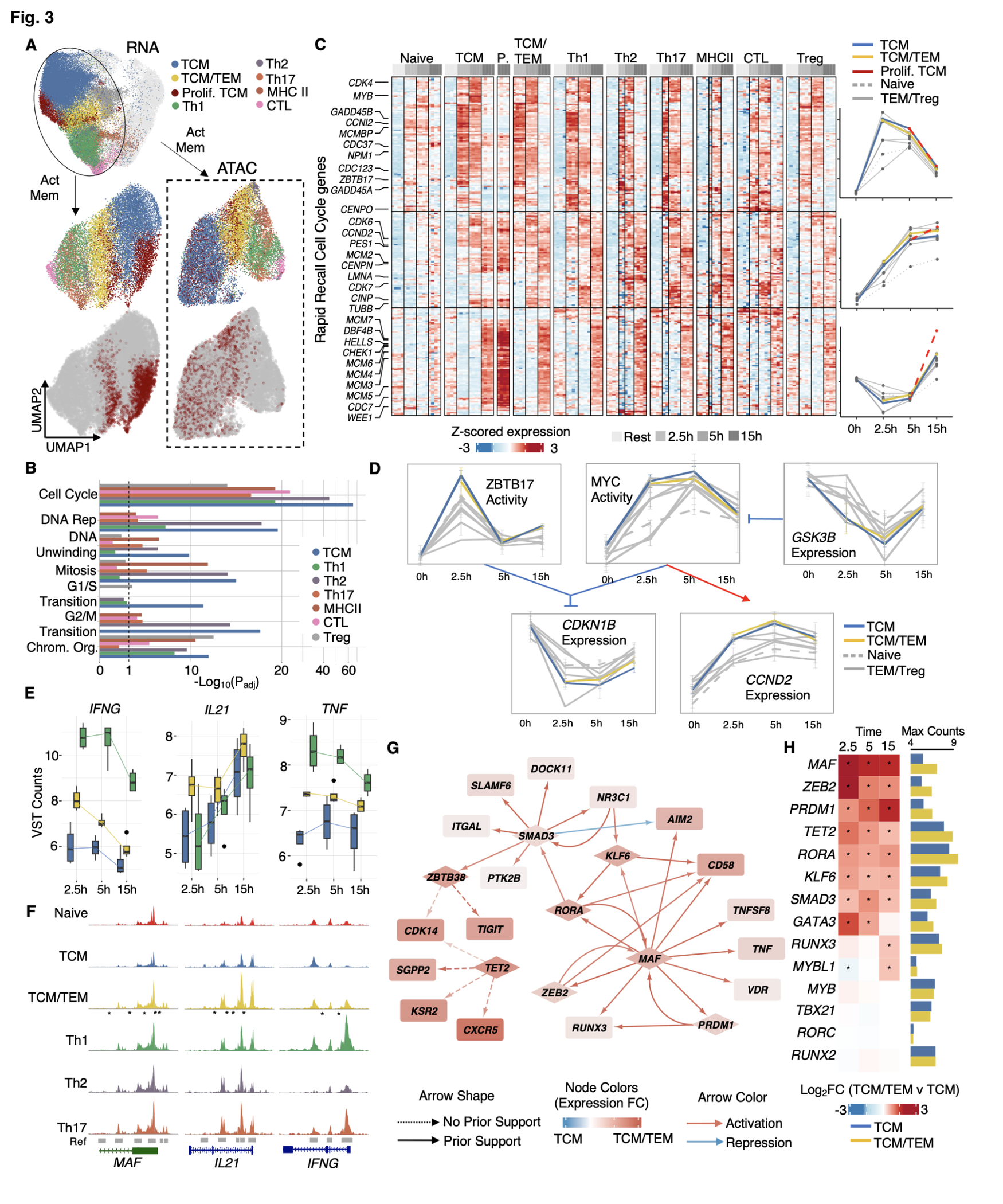

A
RNA
TCM
TCM/TEM
Prolif. TCM
Th1
Th2
Th17
MHC II
CTL
Act
Mem
ATAC
UMAP2
UMAP1
B
Cell Cycle
DNA Rep
DNA Unwinding
Mitosis
G1/S Transition
G2/M Transition
Chrom. Org.
-Log10(Padj)
C
Naive
TCM
TCM/ TEM
Th1
Th2
Th17
MHCII
CTL
Treg
Rapid Recall Cell Cycle genes
CDK4
MYB
GADD45B
CCNI2
MCMBP
CDC37
NPM1
CDC123
ZBTB17
GADD45A
CENPO
CDK6
CCND2
PES1
MCM2
CENPN
LMNA
CDK7
CINP
TUBB
MCM7
DBF4B
HELLS
CHEK1
MCM6
MCM4
MCM3
MCM5
CDC7
WEE1
Z-scored expression
Rest
2.5h
5h
15h
TCM/TEM
Prolif. TCM
Naive
TEM/Treg
D
ZBTB17 Activity
MYC Activity
GSK3B Expression
CDKN1B Expression
CCND2 Expression
E
IFNG
IL21
TNF
VST Counts
F
Naive
TCM
TCM/TEM
Th1
Th2
Th17
Ref
MAF
IL21
IFNG
G
DOCK11
SLAMF6
NR3C1
ITGAL
SMAD3
AIM2
KLF6
CD58
ZBTB38
PTK2B
RORA
TNFSF8
CDK14
TIGIT
MAF
TNF
SGPP2
TET2
ZEB2
VDR
KSR2
CXCR5
RUNX3
PRDM1
Arrow Shape
No Prior Support
Prior Support
Node Colors (Expression FC)
TCM
TCM/TEM
Arrow Color
Activation
Repression
H
Time
Max Counts
MAF
ZEB2
PRDM1
TET2
RORA
KLF6
SMAD3
GATA3
RUNX3
MYBL1
MYB
TBX21
RORC
RUNX2
Log2FC (TCM/TEM v TCM)
TCM
TCM/TEM

Fig. 4

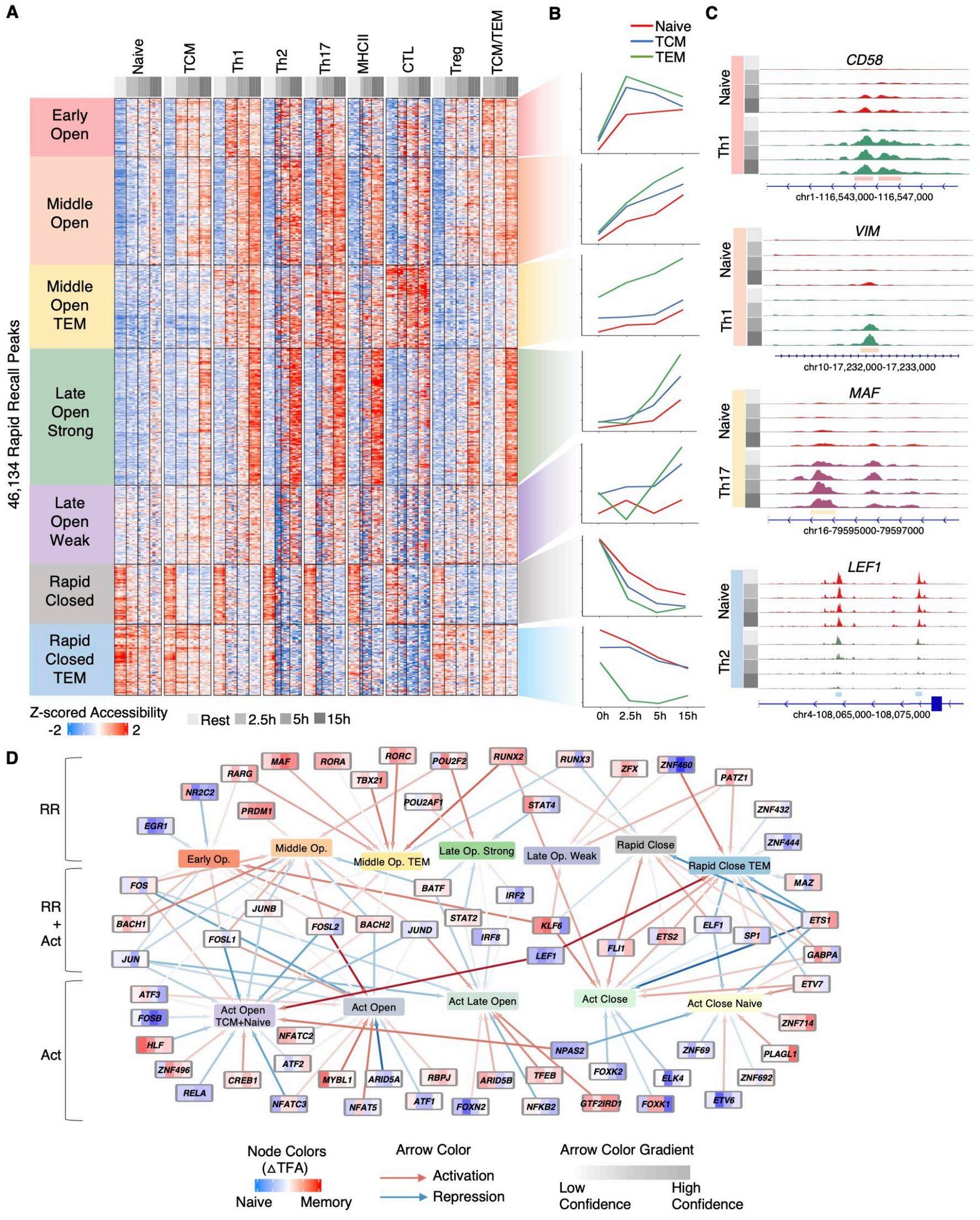

A
Naive
TCM
Th1
Th2
Th17
MHCII
CTL
Treg
TCM/TEM
Early Open
Middle Open
Middle Open TEM
Late Open Strong
Late Open Weak
Rapid Closed
Rapid Closed TEM
46,134 Rapid Recall Peaks
Z-scored Accessibility
-2
2
Rest
2.5h
5h
15h
B
Naive
TCM
TEM
0h
2.5h
5h
15h
C
CD58
Naive
Th1
chr1-116,543,000-116,547,000
VIM
Naive
Th1
chr10-17,232,000-17,233,000
MAF
Naive
Th17
chr16-79595000-79597000
LEF1
Naive
Th2
chr4-108,065,000-108,075,000
D
RR
RR + Act
Act
Early Op.
Middle Op.
Middle Op. TEM
Late Op. Strong
Late Op. Weak
Rapid Close
Rapid Close TEM
Act Open TCM+Naive
Act Open
Act Late Open
Act Close
Act Close Naive
Node Colors (ΔTFA)
Naive
Memory
Arrow Color
Activation
Repression
Arrow Color Gradient
Low Confidence
High Confidence

Fig. 5

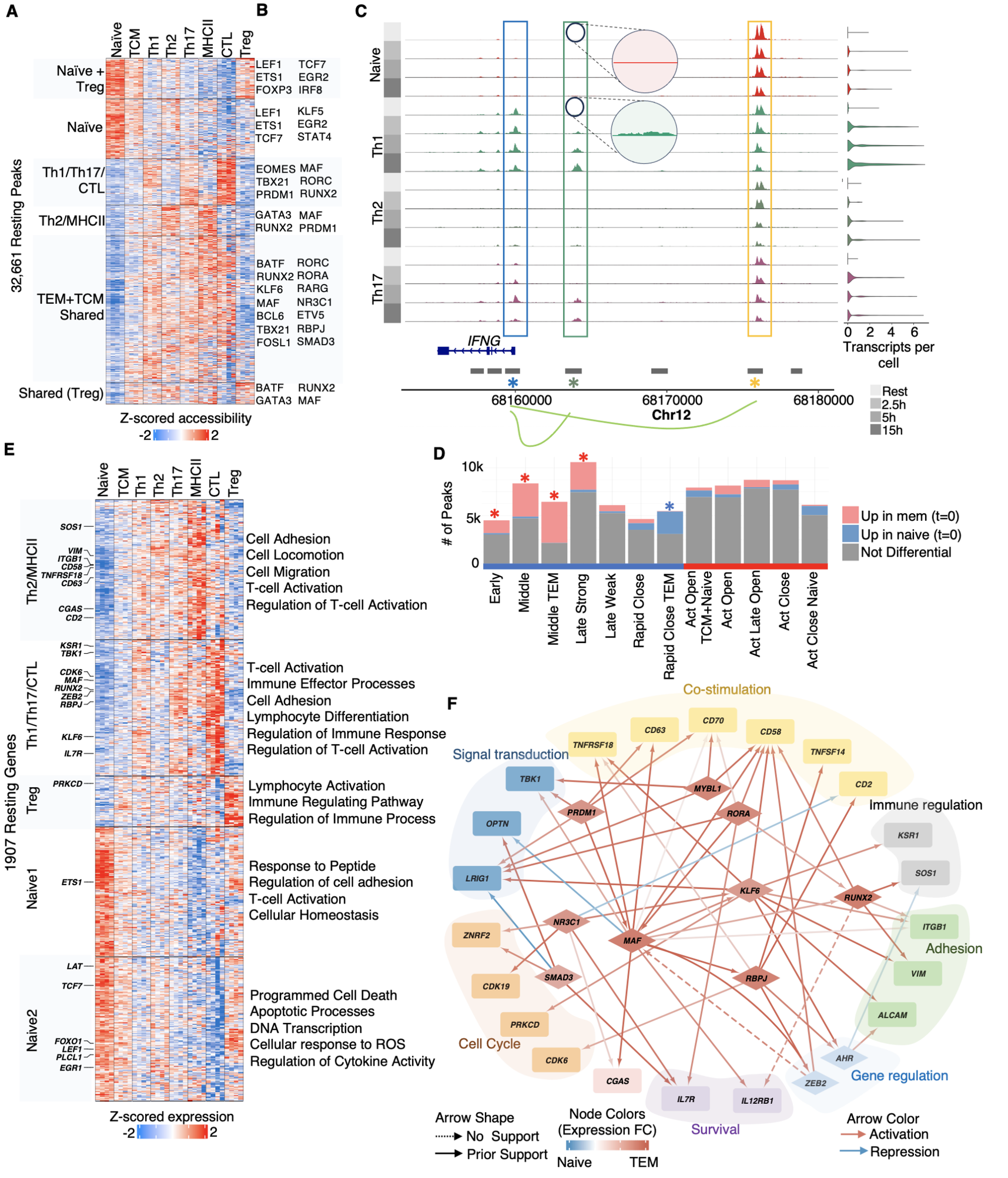

A
B
Naïve
TCM
Th1
Th2
Th17
MHCII
CTL
Treg
Naïve + Treg
LEF1 TCF7
ETS1 EGR2
FOXP3 IRF8
Naïve
LEF1 KLF5
ETS1 EGR2
TCF7 STAT4
Th1/Th17/ CTL
EOMES MAF
TBX21 RORC
PRDM1 RUNX2
Th2/MHCII
GATA3 MAF
RUNX2 PRDM1
TEM+TCM Shared
BATF RORC
RUNX2 RORA
KLF6 RARG
MAF NR3C1
BCL6 ETV5
TBX21 RBPJ
FOSL1 SMAD3
Shared (Treg)
BATF RUNX2
GATA3 MAF
32,661 Resting Peaks
Z-scored accessibility
-2
2
C
Naive
Th1
Th2
Th17
IFNG
68160000
68170000
68180000
Chr12
0 2 4 6
Transcripts per cell
Rest
2.5h
5h
15h
D
10k
5k
0
# of Peaks
Early
Middle
Middle TEM
Late Strong
Late Weak
Rapid Close
Rapid Close TEM
Act Open TCM+Naive
Act Open
Act Late Open
Act Close
Act Close Naive
Up in mem (t=0)
Up in naive (t=0)
Not Differential
E
Naive
TCM
Th1
Th2
Th17
MHCII
CTL
Treg
Th2/MHCII
SOS1
VIM
ITGB1
CD58
TNFRSF18
CD63
CGAS
CD2
Cell Adhesion
Cell Locomotion
Cell Migration
T-cell Activation
Regulation of T-cell Activation
Th1/Th17/CTL
KSR1
TBK1
CDK6
MAF
RUNX2
ZEB2
RBPJ
KLF6
IL7R
T-cell Activation
Immune Effector Processes
Cell Adhesion
Lymphocyte Differentiation
Regulation of Immune Response
Regulation of T-cell Activation
Treg
PRKCD
Lymphocyte Activation
Immune Regulating Pathway
Regulation of Immune Process
Naive1
ETS1
Response to Peptide
Regulation of cell adhesion
T-cell Activation
Cellular Homeostasis
Naive2
LAT
TCF7
FOXO1
LEF1
PLCL1
EGR1
Programmed Cell Death
Apoptotic Processes
DNA Transcription
Cellular response to ROS
Regulation of Cytokine Activity
1907 Resting Genes
Z-scored expression
-2
2
F
Co-stimulation
CD70
CD63
CD58
TNFRSF18
TNFSF14
CD2
Signal transduction
TBK1
MYBL1
PRDM1
RORA
OPTN
LRIG1
Immune regulation
KSR1
SOS1
KLF6
RUNX2
NR3C1
ZNRF2
MAF
ITGB1
Adhesion
VIM
SMAD3
RBPJ
CDK19
ALCAM
PRKCD
Cell Cycle
CDK6
AHR
CGAS
ZEB2
Gene regulation
IL7R
IL12RB1
Survival
Arrow Shape
No Support
Prior Support
Node Colors (Expression FC)
Naive
TEM
Arrow Color
Activation
Repression

Fig. 6

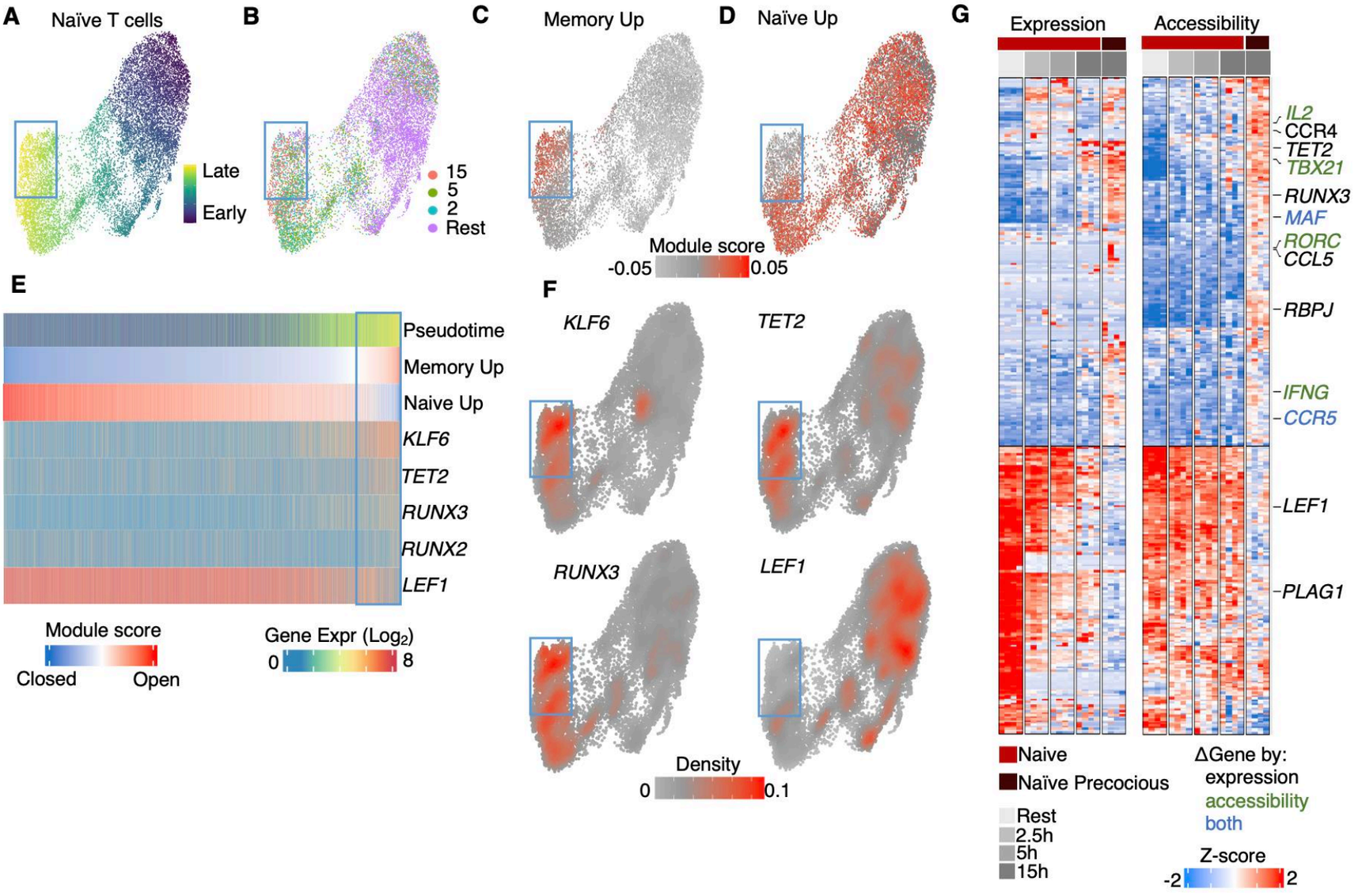
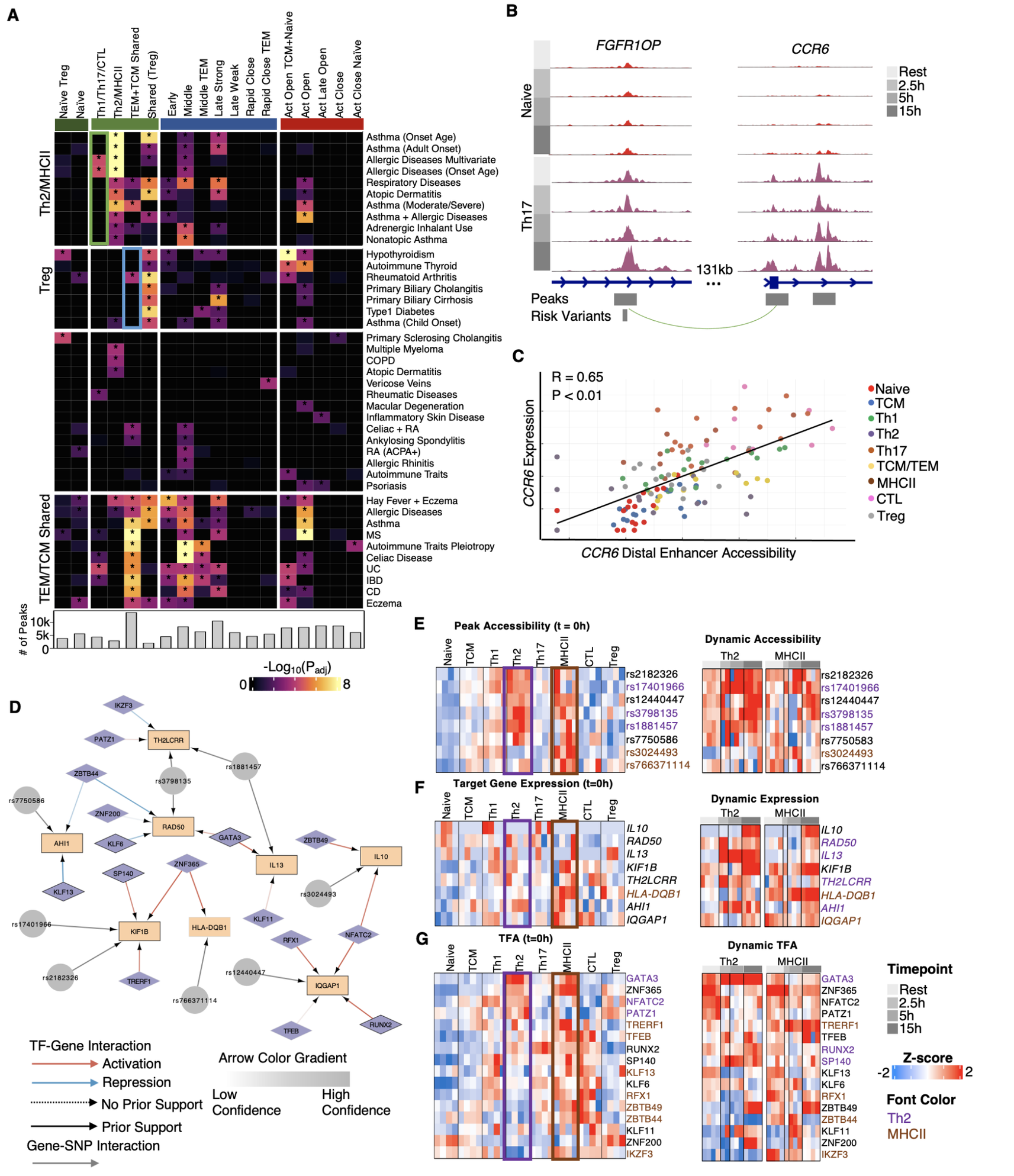

A
Naïve T cells
Late
Early
B
15
5
2
Rest
C
Memory Up
D
Naïve Up
Module score
-0.05
0.05
E
Pseudotime
Memory Up
Naive Up
KLF6
TET2
RUNX3
RUNX2
LEF1
Module score
Closed
Open
Gene Expr (Log2)
0
8
F
KLF6
TET2
RUNX3
LEF1
Density
0
0.1
G
Expression
Accessibility
IL2
CCR4
TET2
TBX21
RUNX3
MAF
RORC
CCL5
RBPJ
IFNG
CCR5
LEF1
PLAG1
Naive
Naïve Precocious
Rest
2.5h
5h
15h
ΔGene by:
expression
accessibility
both
Z-score
-2
2

Fig. 7


A
Naive Treg
Naïve
Th1/Th17/CTL
Th2/MHCII
TEM+TCM Shared
Shared (Treg)
Early
Middle
Middle TEM
Late Strong
Late Weak
Rapid Close
Rapid Close TEM
Act Open TCM+Naive
Act Open
Act Late Open
Act Close
Act Close Naïve
Th2/MHCII
Treg
TEM/TCM Shared
Asthma (Onset Age)
Asthma (Adult Onset)
Allergic Diseases Multivariate
Allergic Diseases (Onset Age)
Respiratory Diseases
Atopic Dermatitis
Asthma (Moderate/Severe)
Asthma + Allergic Diseases
Adrenergic Inhalant Use
Nonatopic Asthma
Hypothyroidism
Autoimmune Thyroid
Rheumatoid Arthritis
Primary Biliary Cholangitis
Primary Biliary Cirrhosis
Type1 Diabetes
Asthma (Child Onset)
Primary Sclerosing Cholangitis
Multiple Myeloma
COPD
Atopic Dermatitis
Vericose Veins
Rheumatic Diseases
Macular Degeneration
Inflammatory Skin Disease
Celiac + RA
Ankylosing Spondylitis
RA (ACPA+)
Allergic Rhinitis
Autoimmune Traits
Psoriasis
Hay Fever + Eczema
Allergic Diseases
Asthma
MS
Autoimmune Traits Pleiotropy
Celiac Disease
UC
IBD
CD
Eczema
# of Peaks
10k
5k
0
-Log10(Padj)
0
8
B
FGFR1OP
CCR6
Naive
Th17
Rest
2.5h
5h
15h
131kb
Peaks
Risk Variants
C
R = 0.65
P < 0.01
CCR6 Expression
CCR6 Distal Enhancer Accessibility
Naive
TCM
Th1
Th2
Th17
TCM/TEM
MHCII
CTL
Treg
D
IKZF3
PATZ1
TH2LCRR
ZBTB44
rs3798135
rs1881457
rs7750586
ZNF200
RAD50
AHI1
KLF6
GATA3
ZBTB49
IL10
IL13
SP140
ZNF365
KLF13
rs3024493
rs17401966
KIF1B
HLA-DQB1
KLF11
RFX1
NFATC2
rs2182326
TRERF1
rs12440447
IQGAP1
rs766371114
TFEB
RUNX2
TF-Gene Interaction
Activation
Repression
No Prior Support
Prior Support
Gene-SNP Interaction
Arrow Color Gradient
Low Confidence
High Confidence
E
Peak Accessibility (t = 0h)
Naive
TCM
Th1
Th2
Th17
MHCII
CTL
Treg
rs2182326
rs17401966
rs12440447
rs3798135
rs1881457
rs7750586
rs3024493
rs766371114
Dynamic Accessibility
Th2
MHCII
rs2182326
rs17401966
rs12440447
rs3798135
rs1881457
rs7750583
rs3024493
rs766371114
F
Target Gene Expression (t=0h)
IL10
RAD50
IL13
KIF1B
TH2LCRR
HLA-DQB1
AHI1
IQGAP1
Dynamic Expression
G
TFA (t=0h)
GATA3
ZNF365
NFATC2
PATZ1
TRERF1
TFEB
RUNX2
SP140
KLF13
KLF6
RFX1
ZBTB49
ZBTB44
KLF11
ZNF200
IKZF3
Dynamic TFA
Timepoint
Rest
2.5h
5h
15h
Z-score
-2
2
Font Color
Th2
MHCII

Supplemental Figure Legends

Fig. S1. Evaluation of the spike-in donor batch-correction strategy and multiome-seq data characteristics (QC and compositions), related to Fig. 1. UMAP of **(A)** unharmonized and **(B,C)** harmonized data colored by either donor or cell population. Harmony integration reduces the effect of donor on clustering, thereby improving aggregation of nuclei by cell population / timepoint for downstream pseudobulk analyses (see **Methods**). Three batch-correction strategies for pseudobulk expression data were evaluated visually (**D-F**) and quantitatively (**G-H**), see **Methods**. (**D-F**) The 1000 activation-dynamic genes with greatest batch effects are displayed for z-scored VST counts resulting from the three pseudobulk batch-correction strategies: **(D)** no batch correction, **(E)** incorrect "batch = donor" design and **(F)** "batch = dataset" correction enabled by the spike-in strategy**.** Black box highlights a gene cluster where batch effects are especially apparent in the "no batch" strategy (D) but absent from the other two strategies (E-F). **(G)** Total number of dynamic genes identified by differential DEG analysis for each major T-cell population, using "batch = dataset" (red), batch = donor (light blue) or no batch correction (dark blue). An additional counts matrix was simulated by up-sampling an additional 8% of experimental cells and applying the batch = donor correction (beige) (**Methods**). **(H)** Number of differential genes identified between Th1, Th2 and Th17 cell populations at rest; methods as described for (G). Spike-in correction with batch = dataset resulted in the largest number of differential genes for both comparison types (G,H), providing strong support for using a spike-in control in the experimental design. **(I)** UMI counts and **(J)** fragment counts for each major cell population are relatively consistent. **(K)** Number of cells in each major cell population stacked by timepoint. **(L)** Cell number for each major cell class (Naïve, Stressed, TCM, TEM, TCM/TEM, Treg). TEM is composed of Th1, Th2, Th17, MHCII and CTL.

Fig. S2. Assessment of variation in the gene expression and chromatin accessibility data, related to Fig. 1. (A-B) Principal component analysis (**PCA**) of pseudobulk gene expression profiles. Datapoint color and shape show cell-type and timepoint, respectively. PC1-PC3 correlate with timepoint, while PC4 describes cell-type variation. 95% confidence ellipses are drawn for each of the 4 timepoints. (**C-D**) Loadings plots of selected transcripts contributing to the PCs. **(E-F)** PCA of pseudobulk chromatin accessibility profiles. PC1 describes timepoint variation, while PC2 describes cell-type variation. PC3-4 has a mix of timepoint and cell-type contributions. Hierarchical clustering of the pseudobulk (**G**) expression or (**H**) accessibility profiles. Only features with a coefficient of variation > 0.01 were included. Data were z-scored and clustered using Ward's method and Euclidean distance. The RNA profiles clustered by timepoint, while clustering of the ATAC profiles reflected a mix of timepoint and cell-type variation. **(I-J)** Normalized expression of activation-inducible *IRF4* and *CD44* in each population. **(K)** Time-dependent expression dotplot of MHCII-associated genes across the TEM populations. Color represents average log expression across the four donors and size represents percent of cells with at least one transcript detected. MHCII cluster has the highest expression of antigen presentation genes among all T cells.

Fig. S3. Characterization of T-cell activation transcriptional responses, related to Fig. 2. (A) Pseudobulk gene expression heatmap of the 1076 RR genes that are uniquely increased in only one memory population (e.g., "subset-specific"; FDR = 10%, $\log_2$(FC) > 0.58). The compositions of memory subset-specific- versus non-specific RR genes that are **(B)** up-regulated (1076 out of 2372 genes) or **(C)** down-regulated (1577 out of 3773 genes). Most RR genes (>50%) are defined based on more than one population, and TCM have the greatest fraction of subset-specific RR genes. Compositions are influenced by both biological factors (functional similarity of subsets) and technical, as we had less power to detect differential genes in lower-abundance populations (e.g., CTL and MHCII), **(D)** Pathway enrichments of each TF's predicted activated target genes, using GO Molecular Pathways. Enriched pathways shown reached significance for at least one TF (Fisher's exact test, FDR = 10%). Pathway enrichment analysis of each TF's repressed target genes yielded few significant enrichments and are therefore not shown. **(E)** Pathway enrichment of each RR gene cluster, using GO terms. Any pathway achieving significance in at least one cluster is included (Fisher's exact test, FDR = 10%). Z-scored **(F)** TF activity (**TFA**) and **(G)** gene expression, resolved by donors, cell types and timepoints, for select TFs (**Methods**). TFs mentioned in the text are highlighted in blue.

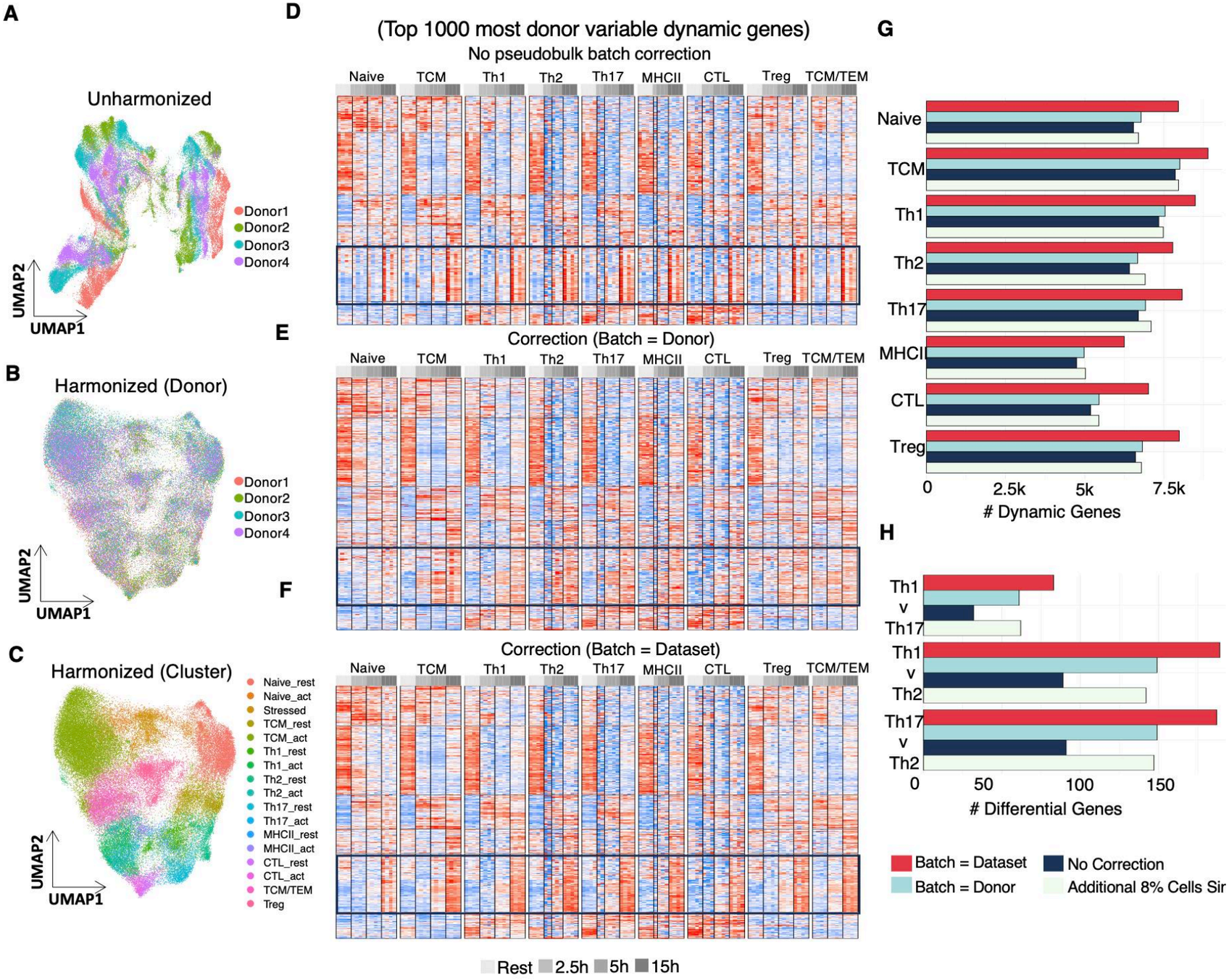
Fig. S4. Characterization of shared T-cell activation transcriptional responses, proliferating TCM and TCM/TEM, related to Fig. 2 and Fig. 3. (A) K-means clustering of pseudobulk gene expression profiles for all activated, non-RR genes. Gene set enrichment analysis (Fisher's exact test) with GO terms was performed for each cluster, and the top pathways, achieving (FDR = 10%) are listed. Median gene expression for the major T-cell populations is shown to the right of each cluster. TEM is the average of Th1, Th2 and Th17. **(B)** Prediction of regulatory interactions driving the "Rapid Down" gene cluster's dynamics. **(C)** Select "Proliferating TCM" marker genes, differentially increased in the proliferating TCM relative to each of the other T-cell populations at 15 hours (FDR = 10%, Log_2FC > 0.58). Dot size represents the percentage of cells with transcript detected, and color is average Log_2(TPM). **(D)** Percentage of each T-cell population predicted to be proliferating (S, G2/M phase combined). Error bars are standard error (n = 4 donors) per cell-type timepoint. Relative to any other population at any other timepoint, proliferating TCM have a significantly higher proportion of cells in S or G2/M phase (FDR = 10%). **(E)** Normalized expression of *MYC* across T-cell populations. **(F)** Normalized expression of *TET2* across T-cell populations. TCM/TEM has significantly higher expression of *TET2* when compared to all other cell types except Th2 (FDR = 10%, Log_2FC > 0.58). **(G)** Gene expression of select cytokines and TFs. TCM/TEM (yellow box) have higher expression when compared to TCM (blue box).

Fig. S5. Conserved activation-responsive chromatin accessibility and the enrichment of activation-responsive chromatin in activation-responsive gene sets, related to Fig. 4 and Fig. 6. (A) K-means clustering of accessible chromatin regions with shared activation-dynamics across naïve and memory populations. Median peak accessibility for the major T-cell populations are shown to the right of each cluster. TEM is the average of Th1, Th2 and Th17. **(B)** Gene targets for RR (**Fig. 4A)** and memory-naïve shared activation-dependent (A) peak clusters were

determined using Trac-looping (when available) or proximity (**Methods**). Fisher exact test was used to identify significant overlap of peak set gene targets and RR (**Fig. 2A**) and naïve-memory shared activation-dependent (**Fig. S4A**) gene clusters. Dot color represents log-transformed adjusted p-values, while the size indicates the percentage of peaks within each peak set linked to genes within each RR gene cluster. **(C)** Cohen's d effect sizes for module score differences between activated naive at 15h and precocious naive populations for each activation-dependent peak sets (**Figs. 4A, S5A**). Effect sizes were calculated using donor-averaged module scores (n=4 donors per cell type); Cohen's d is the standardized difference between group means divided by the pooled standard deviation. $^{*}P_{adj} < 0.1$, $^{**}P_{adj} < 0.01$, $^{***}P_{adj} < 0.001$.

Fig. S6. GRN subnetwork for resting Treg chromatin regions co-localized with variants from the "Treg disease" cluster, related to Fig. 7. (A) GRNs centered on target genes linked (via trac-loop, micro-C or promoter proximity) to genetic risk variants in the "Treg cluster" (blue box in **Fig. 7A**). Only genes expressed in at least 5% of Treg were included. TF regulators were included if they exhibited Treg-specific activity at rest or post-activation. Outlined nodes represent rapid recall genes for Treg. Red arrows indicate activating interactions, while blue arrows denote inhibitory interactions. Solid arrows represent interactions supported by both gene expression modeling and the ATAC-derived "prior" network, while dashed arrows depict interactions supported by gene expression modeling alone. Arrow color gradient reflects confidence levels, from low to high. (**B**) Peak accessibility (**C**) target gene expression and (**D**) transcription factor activity for TF regulators. For panels B-D, the left heatmap represents the resting state (t=0h), while the right heatmap shows activated dynamics.

Fig. S1 – related to Fig. 1



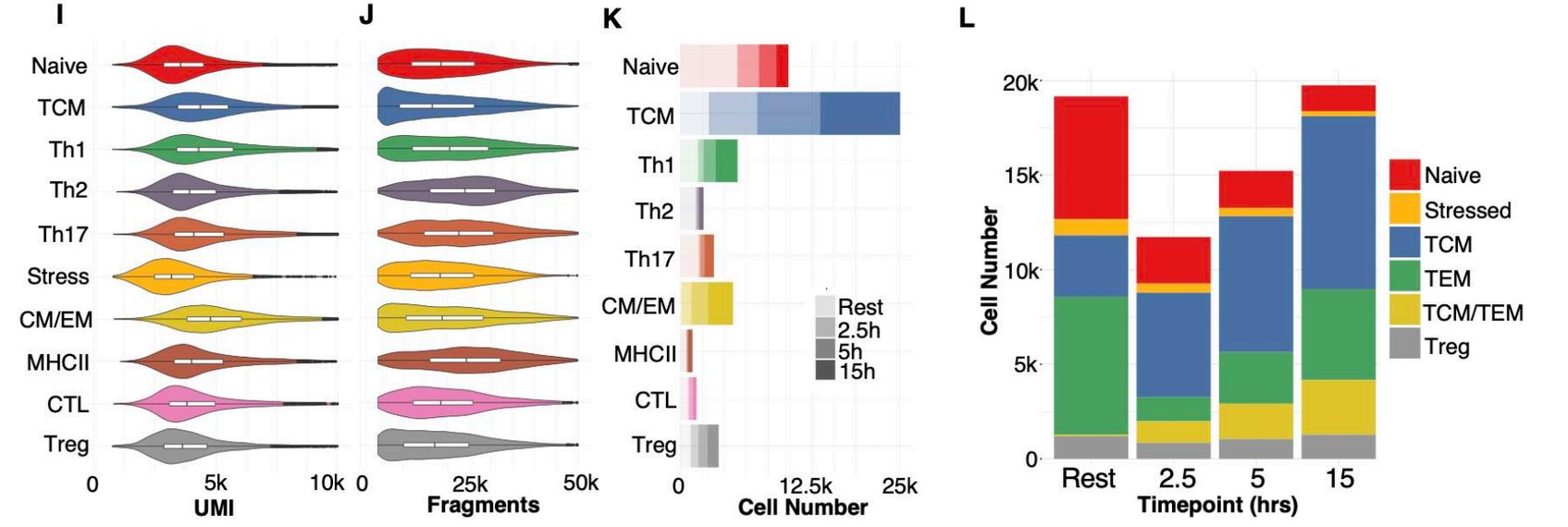


Fig. S2 – related to Fig. 1

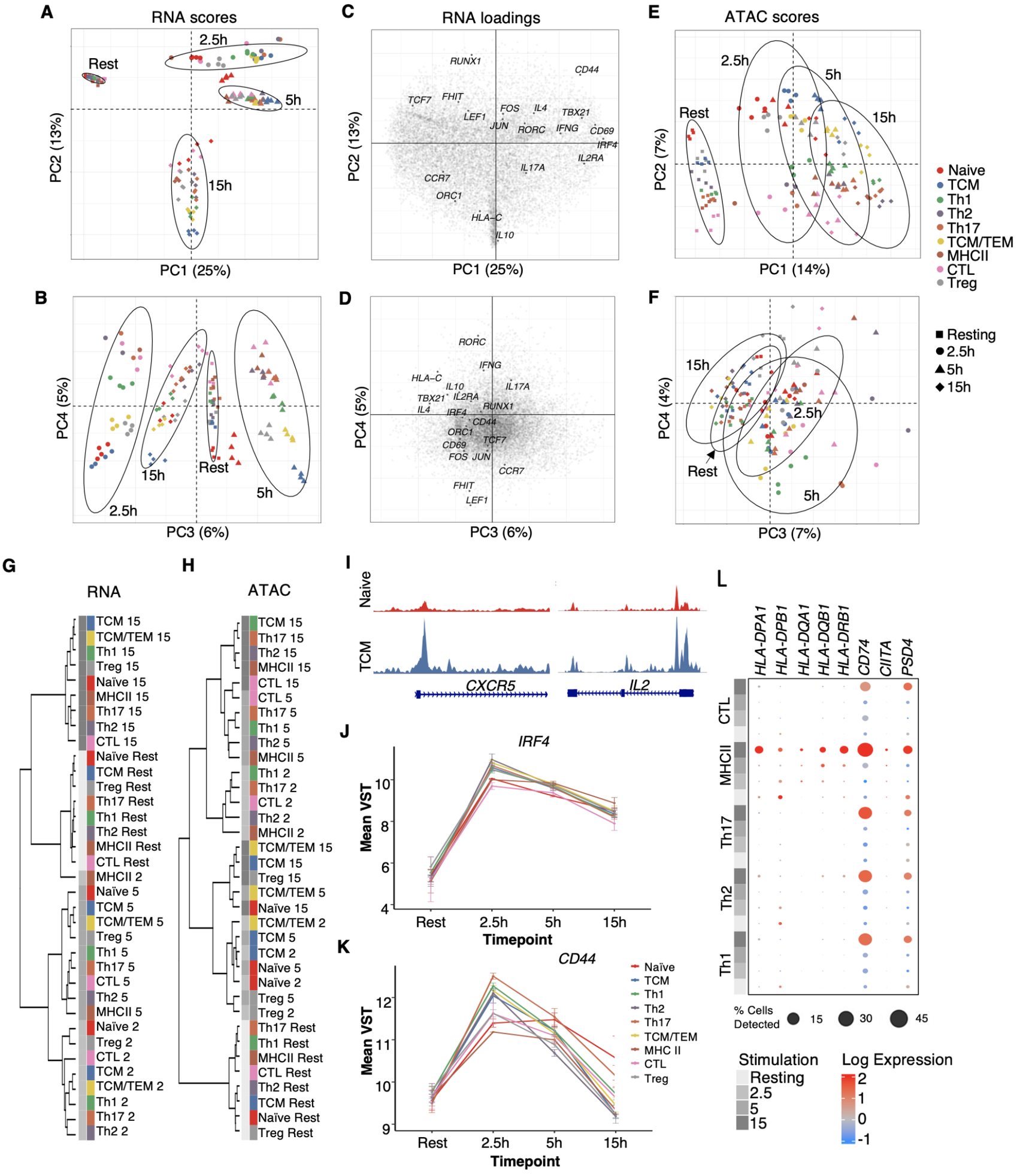


Fig. S3 – related to Fig. 2

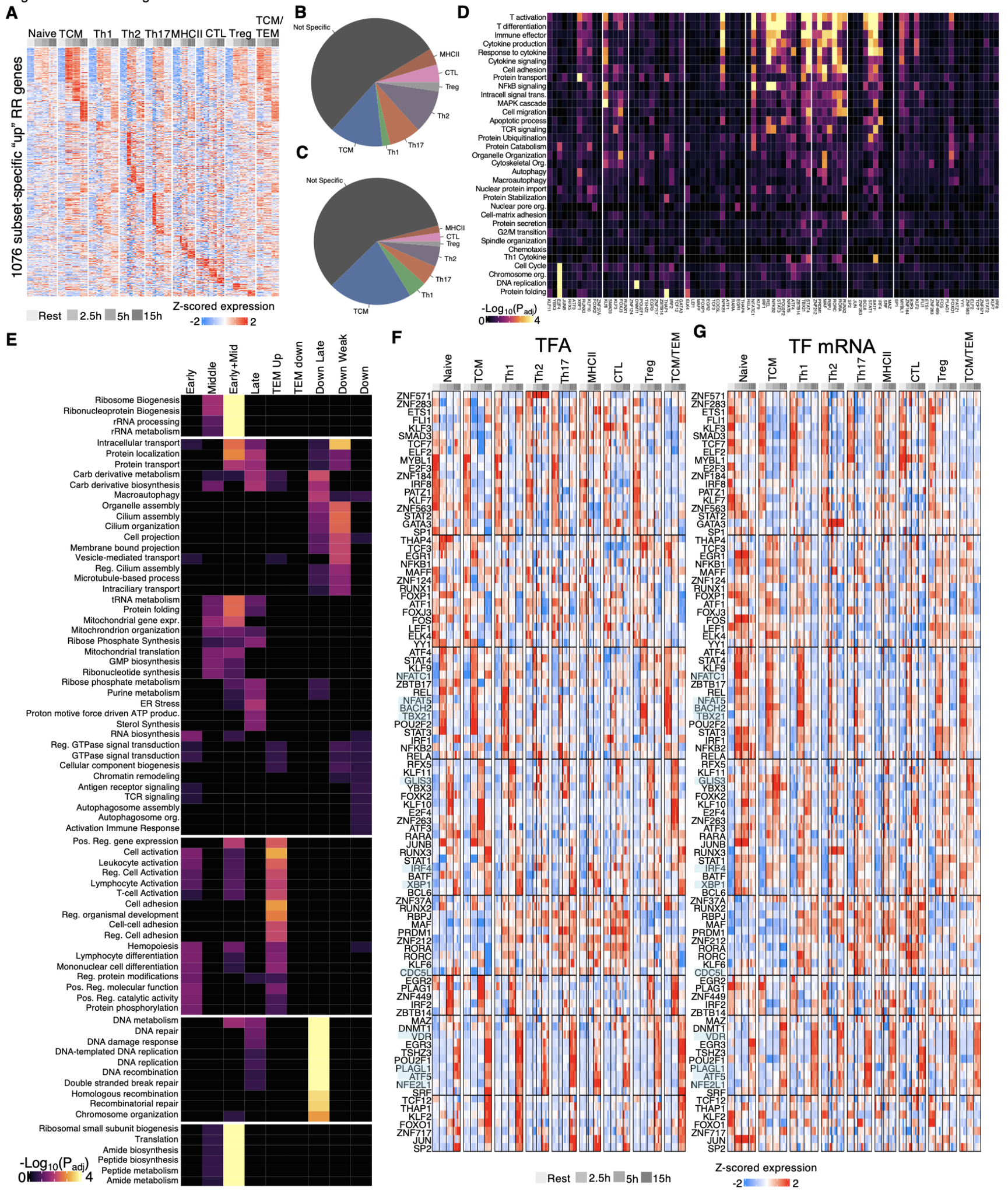


Fig. S4 – related to Fig. 2 and 3

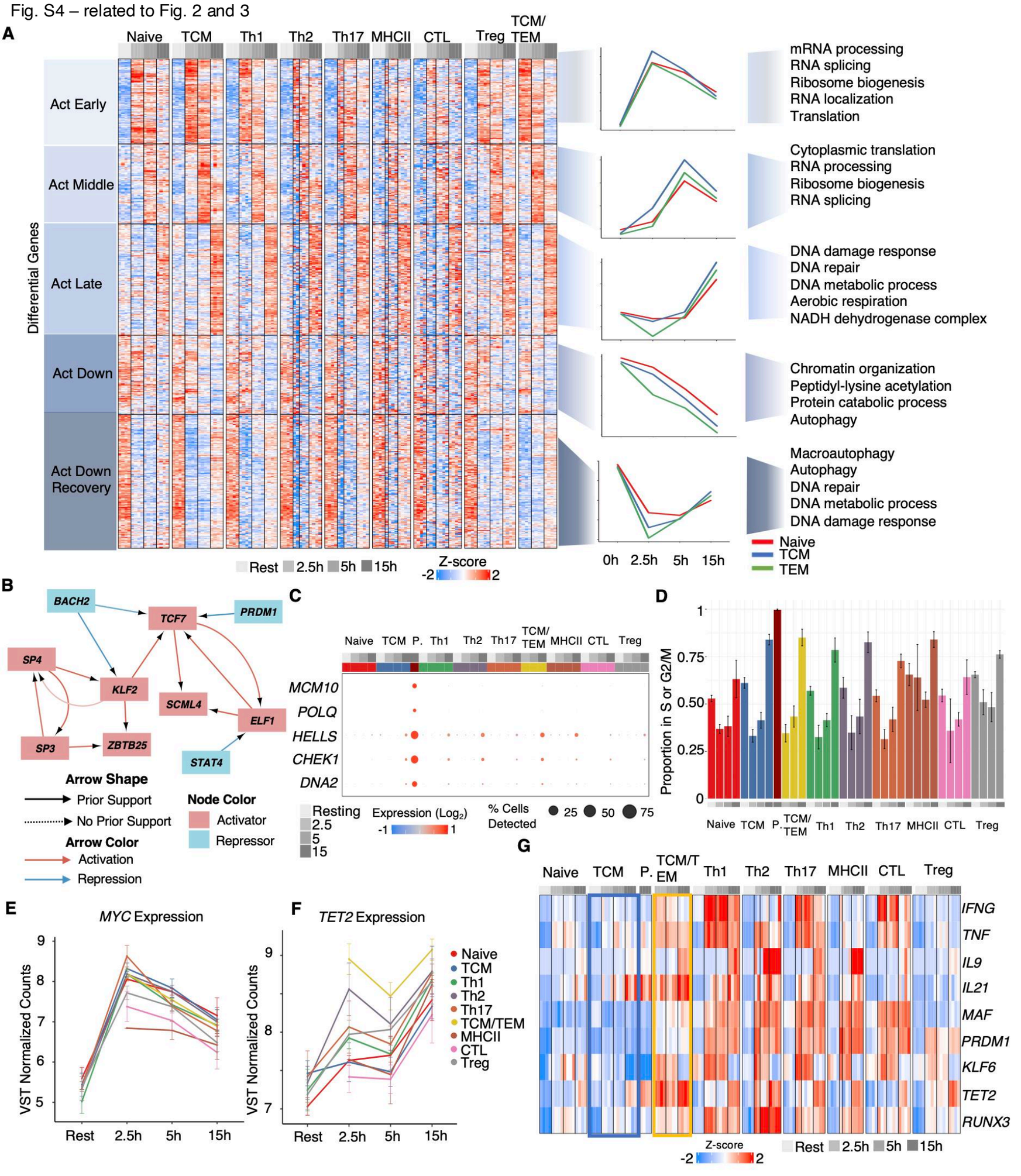


Fig. S5 – related to Fig. 4 and 6

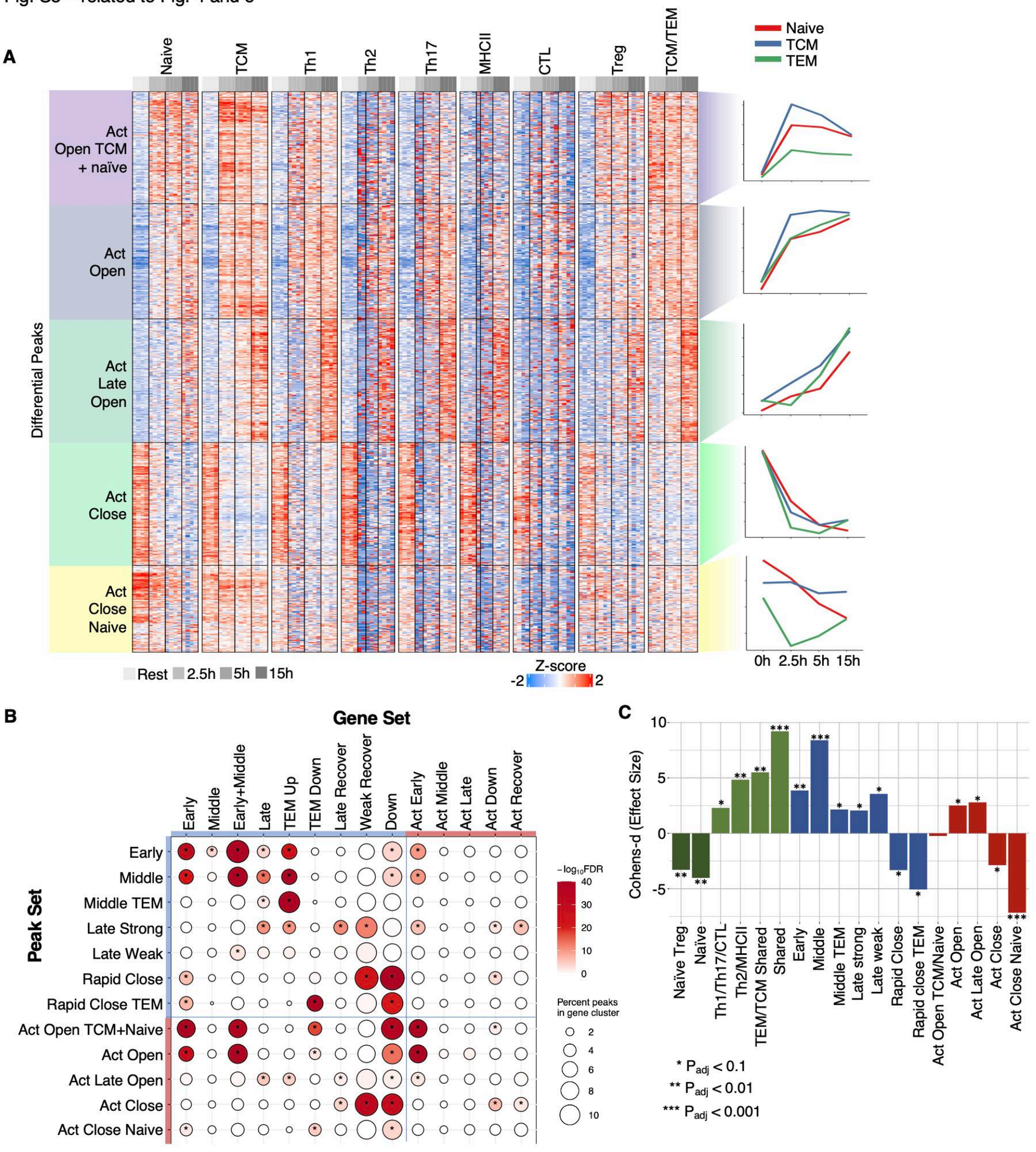
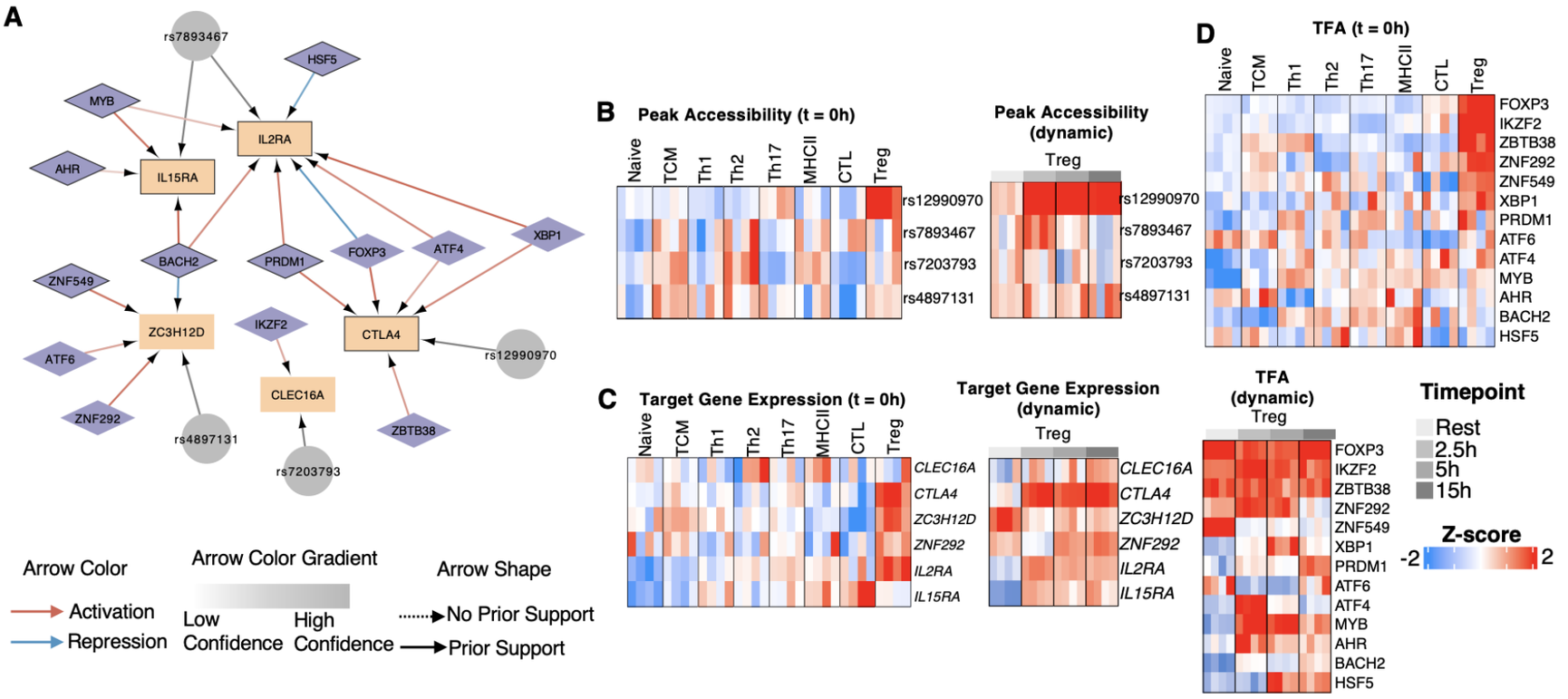


Fig. S6 – related to Fig. 7



Supplemental Table Legends

Table S1. Gene regulatory network and functional enrichment of TF target genes. Predicted TF-gene interactions from the gene regulatory network are presented, along with confidence values such as stability scores and the correlation between Transcription Factor Activity (**TFA**) and target gene expression. Additionally, TF functions are annotated using Gene Set Enrichment Analysis (**GSEA**) of each TF's predicted up-regulated target genes. (**A**) "GRN", (blue) The list of TF regulator - target gene interactions identified by the GRN; (**B**) "TFFunctionsPositive" (green) GO Enrichment Analysis Results for targets of each TF regulator with $P_{adj} \leq 0.1$ terms shown (**Fig. S3D**).

Table S2. Differential gene expression and identification of rapid recall, dynamic, and resting genes. Differential gene expression results used to identify rapid recall, dynamic, and resting (differential between memory and naïve at baseline) genes. (**A**) "RapidRecallGenes" (blue) Genes categorized as rapid recall (both up-and down-regulated during activation) for any cell type (**Fig. 2A**). (**B**) "ActGenes" (green) Genes that are differential between Resting and Activated (2.5h, 5h, 15h) timepoints (**Fig. S4A**). (**C**) "RestingGenes" (yellow) Genes that are differential between naive and memory populations for the timepoint 0h (**Fig. 5A**).

Table S3. Rapid recall and dynamic gene cluster enrichments. We performed GSEA on each rapid call (**Fig. 2A**) and dynamic (**Fig. S4A)** gene cluster to identify functional roles. Similarly, we preformed enrichment of predicted TF up-regulated target genes in each cluster to find likely TF regulators. (**A**) "RR_TFEnrichment" (blue). The list of TFs that are predicted to regulate genes in each RR cluster in **Fig. 2A**. (**B**) "RR_PathwayEnrichment" (green) GO analysis results for each cluster in **Fig. 2A**. (**C**) "Act_TFEnrichment" (yellow) provides the list of TFs that are predicted to regulate genes in each activation dynamic cluster in **Fig. S4A**. (**D**) "Act_PathwatEnrichment" (purple). GO analysis results of each cluster from **Fig. S4A**. $P_{adj} \leq 0.1$ terms are shown.

Table S4. Enrichment of late (15h) rapid recall genes in cell-cycle related GO pathways. GSEA results for each memory population's 15h rapid recall genes for cell-cycle related pathways in the GO database. "TCM/TEM" was not included because this cell population does not have a baseline (t=0) comparator.

Table S5. Differential peak accessibility and identification of rapid recall, dynamic, and resting peaks. Differential peak accessibility results used to identify rapid recall, dynamic, and resting (differential between memory and naïve at baseline) chromatin regions. We also include the cluster identity of each peak from **Fig. 4A**, **Fig. S5A**, and **Fig. 5A**. (**A**) "RapidRecallPeaks" (blue) Peaks categorized as rapid recall (both up-and down-regulated during activation) for any cell type (**Fig. 4A**). (**B**) "ActPeaks"(green) Peaks that are differential between Resting and Activated (2.5h, 5h, 15h) timepoints (**Fig. S5A**). (**C**) "RestingPeaks" (yellow) Peaks that are differential between naive and memory populations for the timepoint 0h (**Fig. 5E**).

Table S6. TF motif enrichment in peak clusters. We calculated the enrichment of each TF's binding motifs from the CIS-BP database in the rapid recall (**Fig. 4A**), dynamic (**Fig. S5A**) and resting (**Fig. 5A**) peak sets.

Table S7. Differential gene expression and gene accessibility between naïve (t=15h) and naïve precocious. DESeq2 results for the comparison between activated naïve (t=15h) and the

naïve precocious population. This table includes results for gene expression (**A**, blue) and chromatin accessibility (**B**, green) (total accessibility across promoter and gene body).

Table S8. Regulatory Element Locus Intersection (RELI) enrichments. Enrichment of peak-locus intersections for each phenotype across the rapid recall, dynamic, and resting peak sets. The table includes number of overlaps and statistics associated with the enrichment for each condition.

Table S9. Peak to gene linkages from proximity, Trac, and micro-C. We linked each peak to a gene based on TSS proximity (+/- 1kb) (**A**), Trac-looping (**B**), and micro-C (**C**). This table includes each peak-gene association separated by linkage method. Trac-loop data are for resting and 16 hrs activated total human CD4 T cells[35]. Micro-C data are for resting and 24 hrs activated total human CD4 T cells[58].