Supplementary Methods

Patient details and treatment

Patients either had myeloablative conditioning (cyclophosphamide and fractinonated total body irradiation) or reduced intensity conditioning (fludarabine and melphalan) (Supplementary Table 1). All allogenic HSCT patients had prophylactic cyclosporin A and high-dose acyclovir (day -5 to day 28 or discharge) with subsequent valganciclovir (to day 100). Patients with high CMV DNAemia (all NIR patients except p47) were treated with ganciclovir (14 days) but only two patients p4 and p13 were on treatment (just finished) when the blood sample was taken.

JARID2 and HDAC6 inhibition

JARID2 and HDAC6 inhibition studies were performed on CD8+ T cells isolated from 40 ml blood from healthy donors using the RosetteSep™ Human $CD8^+$ T Cell Enrichment Cocktail (StemCell Technology, Cambridge, UK). For JARID2 inhibition, cells were cultured with JIB-04 (46 nM or 1150 nM; Selleckchem, Houston, TX) for 4 h. For HDAC6 inhibition, cells were cultured with CAY10603 (20 nM, 500 nM, or 2500 nM; Selleckchem) for 2 h then stimulated with phorbol myristate acetate (PMA; 10 ng/mL, Sigma-Aldrich, St. Louis, MI) and ionomycin (1 nM, Sigma-Aldrich) for 2 h.

Immunofluorescence microscopy

Immunofluorescence microscopy was performed on cytospins prepared from FACS sorted, CMV-specific T cells. The cells were permeabilised with 1% Triton X-100 for 20 min and probed with mouse anti-JARID2 (ab93288, Abcam, Cambridge, UK), goat anti-HDAC6 (ab196351, Abcam), rabbit anti-KDM6B (ab169197, Abcam), mouse anti-PI3KCG, mouse anti-NFATC1 (sc-7294, Santa Cruz Biotechnology, Dallas, TX), or rabbit

anti-KLF16 (NMP2-30580 Novus Biologicals, Littleton, CO), followed by anti-mouse Alexa-Fluor-568 (A10037, Life Technologies), anti-goat Alexa-Fluor-633 (A20182, Life Technologies), or anti-rabbit Alexa-Fluor-488 (A21206, Life Technologies). Coverslips were mounted (ProLong Diamond Antifade Reagent, Life Technologies) and imaged by confocal laser scanning microscopy (Leica DMI8, x100 oil immersion lens). Total cytoplasmic (PI3KGC) or nuclear (all others) fluorescence intensity was measured (ImageJ, NIH, Bethesda, USA) in a minimum of 20 cells. The 20 plus cells were randomly selected on the DAPI only images to ensure integrity of the cells before quantification of the protein staining.

Global chromatin accessibility

FAIRE was carried out as previously described[1]. Briefly, cells were cross-linked with 1% formaldehyde and lysed. Chromatin-bound DNA was sheared by sonication, phenol-chloroform extracted, and purified with the Zymo-SpinTM I kit (Zymo Research, Irvine, CA). FAIRE DNA libraries and a mixed total input were prepared with the NEBNext® Ultra DNA Library Prep Kit for Illumina® (New England BioLabs, Ipswich, MA). FAIRE libraries were single-end 75bp sequenced on an Illumina NextSeq at the Ramaciotti Centre for Genomics. Samples from 4 IR (p01, p8, p15, and p37) and 4 NIR (p04, p14, p25, and p47) recipients were separately pooled.

Flow cytometry

To detect or isolate CMV-specific or naïve T cells, PBMCs were incubated with APC- or PE-conjugated MHC class I multimers specific for the HLA A*01:01-restricted epitope VTEHDTLLY, HLA A*02:01-restricted epitope NLVPMVATV, HLA B*07:02-restricted

epitope ELKRKMIYM (Immudex, Copenhagen, Denmark). Cells were incubated with PerCP-Cy5.5-conjugated anti-CD8 (eBioscience; Thermo Fisher Scientific, Waltham, MA), Alexa-Fluor-700- or Pacific Blue-conjugated anti-CD4 (BD Biosciences; Thermo Fisher Scientific), live/dead near IR (Thermo Fisher Scientific), PE-Cy7 CCR7 (BD Bioscience), and CD45RA-FITC (BD Biosciences). To assess MHC-multimer frequencies, cells were acquired using a BD LSR Fortessa and data analysed in FlowJo (TreeStar). To isolate CMV-specific and naïve T cells (CD45RA+ CCR7+), cells were sorted using a FACS Aria-IIIu cell sorter (BD Biosciences).

Gene expression analysis

Gene expression was measured by single-end 75bp sequencing on a NextSeq500 at the Ramaciotti Centre for Genomics, University of New South Wales, Australia. RNA was isolated using the RNAeasy micro kit (Qiagen, CA). First-strand cDNA was synthesized from 1 μg total RNA using the SuperScript™ III First-Strand Synthesis System (Invitrogen, Thermo Fisher Scientific), and qRT-PCR was carried out with gene-specific TaqMan probes on the Applied Biosystems ViiA 7 Real-Time PCR System (Life Technologies). RNA libraries were prepared from DNase-treated RNA using the SMARTer Total RNA-seq pico input (Clontech Laboratories; Mountain View, CA) and single-end 75bp sequenced on a NextSeq500 at the Ramaciotti Centre for Genomics, University of New South Wales, Australia.

Bioinformatics analysis

Sequences were adapter stripped (Trimmomatic-0.33), cleaned of ribosomal sequences (TagDust, RNAseq), and mapped to the Hg19 genome with Bowtie2 (2.2.5) (FAIRE-seq)[2] or Hisat2 (RNA-seq)[3]. RNA-seq gene expression was measured using HOMER

(pc3, -count exon)[4] and variance stabilization and differential expression were performed in DeSeq2[5] with a false discovery rate (FDR) of <0.2. The less stringent FDR was chosen because considerable variation is expected in these human patient samples and the signatures were to be used for further comparison to epigenomic data which benefits from reduced false negative rates. DeSeq2 values incorporate the zero-centered normal prior and thus shrinkage is greater for the log2-fold change estimates from genes with low counts and high dispersion. Genes on ChrY or with low counts were discarded. Although it is worth noting that the number of reads mapped to ChrY from p14 (female donor to male patient), was as low as that from p01 (female donor to female patient), supporting the assertion that the profiled cells are derived from the transplanted donor cells.

For FAIRE-seq, duplicate (via Picard), multi-mapping reads were removed, and enriched regions were determined against a mixed total input sample using the broad option in MACS2[6] and filtered on q-values <0.01 and ≥3 fold enrichment. Specific peaks were those only detected in either IR or NIR and whose counts were at least 1.5-fold higher in one sample than the other. HOMER was used to create bedGraph files, annotate peaks, and find Gene Ontology (GO) and motif enrichment. CD8+ naïve and memory-annotated chromatin states were condensed from RoadMap data[7]. CLOVER ($p<0.05$)[8] was used to find motif enrichment using the JASPAR 2016 position weight matrices and background sequences with regions standardised to 500bp and matching GC content from HOMER. Genes were classified as epigenomic regulators according to EpiFactors[9] and transcription factors according to their 'Protein Function' in the Protein Atlas and grouped by InterPro domain (similar to previous publications[10]).

Enriched regions in publicly available ChIP-seq data were obtained from CISTROME[11], with regions with the same factor and cell type concatenated to produce 1640 ChIP sets. Bedtools[12] was used to calculate the overlap with our region sets (after standardising to 500bp and liftOver to Hg38). ChIP sets that (1) overlapped less than 15% of the IR or NIR regions, or (2) had similar (<1.2-fold difference) occurrence in IR and NIR regions were discarded before creating the occurrence matrix. A score of 1 was given when an overlap was detected and 0 if not, and factor clustering was performed with Pearson distances and average linkage. JARID2 (iPSC, GSM1180134)[13] and HDAC6 (activated T cells, GSM393957)[14] ChIP-seq tracks were obtained via CISTROME and converted to Hg19 using CrossMap.

Enrichment in the ranked genelist from the RNA-seq data (normalised by library size in DeSeq2, unlogged) was calculated by geneset enrichment analysis (GSEA; Signal2Noise, weighted, nominal p-values <0.05, FDR <0.35)[15]. Rank metrics and normalised enrichment scores (NES) for our genesets in specific GEO datasets were also calculated by GSEA (Signal2Noise, weighted).

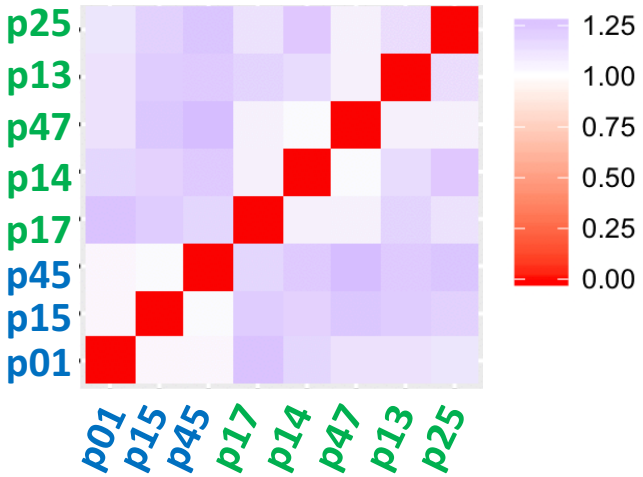
Ggplot2, cluster, and factoextra were used for visualization, similarity calculations, PCA and other clustering. String[16] was used to create functional association networks that were visualised in Cytoscape[17] with molecular function classified primarily by 'Protein Function' in the Protein Atlas[18].

References

1. Simon JM, Giresi PG, Davis IJ, Lieb JD. A detailed protocol for formaldehyde-assisted isolation of regulatory elements (FAIRE). *Curr Protoc Mol Biol*. 2013;Chapter 21:Unit21 26.
2. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. *Nat Methods*. 2012;9(4):357-359.
3. Kim D, Langmead B, Salzberg SL. HISAT: a fast spliced aligner with low memory requirements. *Nat Meth*. 2015;12(4):357-360.
4. Heinz S, Benner C, Spann N, et al. Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol Cell*. 2010;38(4):576-589
5. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology*. 2014;15(12):550.
6. Zhang Y, Liu T, Meyer CA, et al. Model-based analysis of ChIP-Seq (MACS). *Genome Biol*. 2008;9(9):R137.
7. Bernstein BE, Stamatoyannopoulos JA, Costello JF, et al. The NIH Roadmap Epigenomics Mapping Consortium. *Nat Biotechnol*. 2010;28(10):1045-1048.
8. Frith MC, Fu Y, Yu L, Chen JF, Hansen U, Weng Z. Detection of functional DNA motifs via statistical over-representation. *Nucleic Acids Res*. 2004;32(4):1372-1381.
9. Medvedeva YA, Lennartsson A, Ehsani R, et al. EpiFactors: a comprehensive database of human epigenetic factors and complexes. *Database (Oxford)*. 2015;2015:bav067.
10. Jolma A, Yan J, Whitington T, et al. DNA-binding specificities of human transcription factors. *Cell*. 2013;152(1-2):327-339.
11. Liu T, Ortiz JA, Taing L, et al. Cistrome: an integrative platform for transcriptional regulation studies. *Genome Biology*. 2011;12(8):R83.

12. Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics*. 2010;26.
13. Kaneko S, Bonasio R, Saldana-Meyer R, et al. Interactions between JARID2 and noncoding RNAs regulate PRC2 recruitment to chromatin. *Mol Cell*. 2014;53(2):290-300.
14. Wang Z, Zang C, Cui K, et al. Genome-wide mapping of HATs and HDACs reveals distinct functions in active and inactive genes. *Cell*. 2009;138(5):1019-1031.
15. Subramanian A, Tamayo P, Mootha VK, et al. Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Proceedings of the National Academy of Sciences*. 2005;102(43):15545-15550.
16. Szklarczyk D, Morris JH, Cook H, et al. The STRING database in 2017: quality-controlled protein-protein association networks, made broadly accessible. *Nucleic Acids Res*. 2017;45(D1):D362-d368.
17. Shannon P, Markiel A, Ozier O, et al. Cytoscape: A Software Environment for Integrated Models of Biomolecular Interaction Networks. *Genome Research*. 2003;13(11):2498-2504.
18. Uhlen M, Oksvold P, Fagerberg L, et al. Towards a knowledge-based Human Protein Atlas. *Nat Biotechnol*. 2010;28(12):1248-1250

A


p25
p13
p47
p14
p17
p45
p15
p01
1.25
1.00
0.75
0.50
0.25
0.00
p01
p15
p45
p17
p14
p47
p13
p25

B

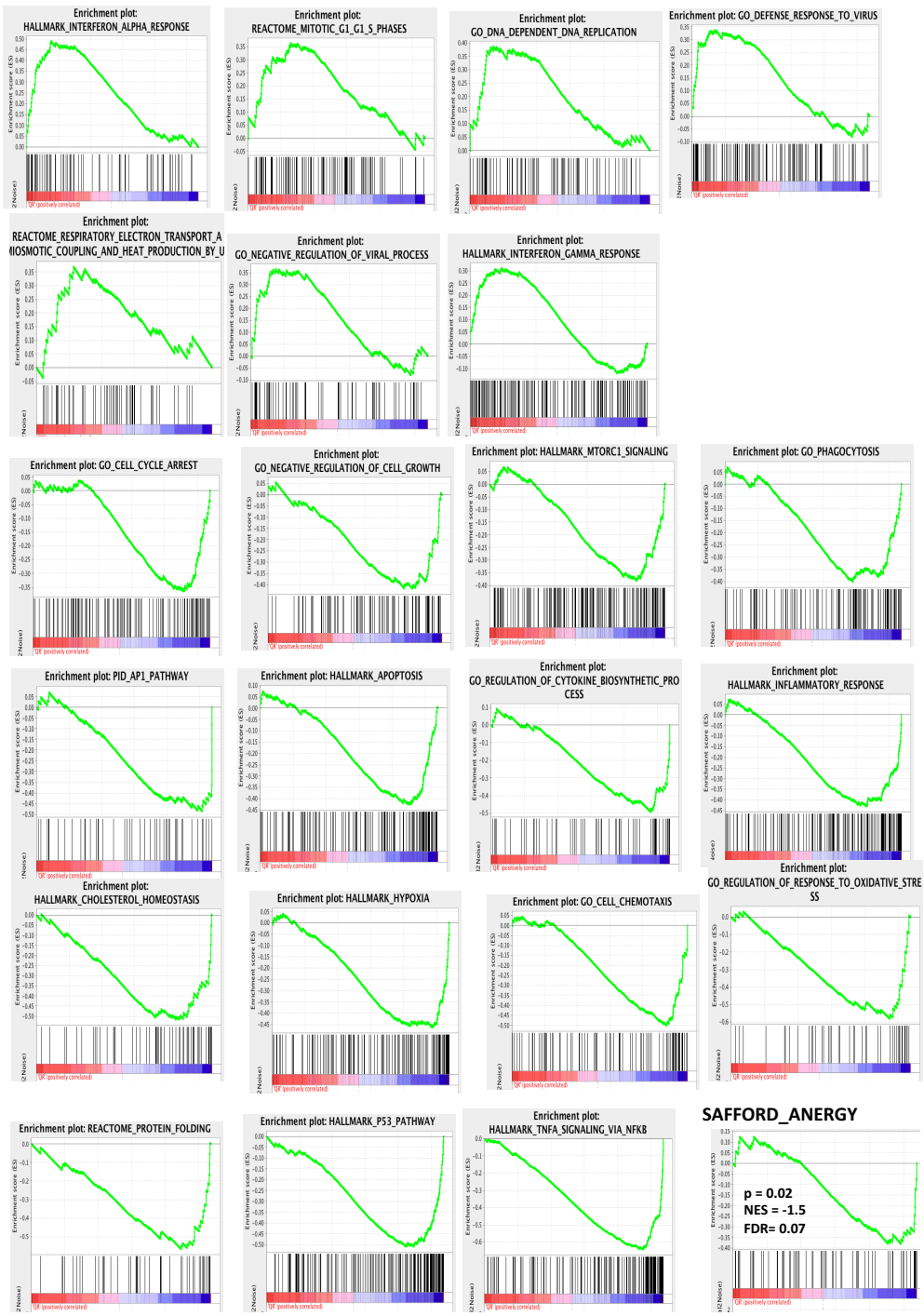

Enrichment plot: HALLMARK_INTERFERON_ALPHA_RESPONSE
Enrichment plot: REACTOME_MITOTIC_G1_G1_S_PHASES
Enrichment plot: GO_DNA_DEPENDENT_DNA_REPLICATION
Enrichment plot: GO_DEFENSE_RESPONSE_TO_VIRUS
Enrichment plot: REACTOME_RESPIRATORY_ELECTRON_TRANSPORT_A
IOSMOTIC_COUPLING_AND_HEAT_PRODUCTION_BY_U
Enrichment plot: GO_NEGATIVE_REGULATION_OF_VIRAL_PROCESS
Enrichment plot: HALLMARK_INTERFERON_GAMMA_RESPONSE
Enrichment plot: GO_CELL_CYCLE_ARREST
Enrichment plot: GO_NEGATIVE_REGULATION_OF_CELL_GROWTH
Enrichment plot: HALLMARK_MTORC1_SIGNALING
Enrichment plot: GO_PHAGOCYTOSIS
Enrichment plot: PID_AP1_PATHWAY
Enrichment plot: HALLMARK_APOPTOSIS
Enrichment plot: GO_REGULATION_OF_CYTOKINE_BIOSYNTHETIC_PRO
CESS
Enrichment plot: HALLMARK_INFLAMMATORY_RESPONSE
Enrichment plot: HALLMARK_CHOLESTEROL_HOMEOSTASIS
Enrichment plot: HALLMARK_HYPOXIA
Enrichment plot: GO_CELL_CHEMOTAXIS
Enrichment plot: GO_REGULATION_OF_RESPONSE_TO_OXIDATIVE_STRE
SS
Enrichment plot: REACTOME_PROTEIN_FOLDING
Enrichment plot: HALLMARK_P53_PATHWAY
Enrichment plot: HALLMARK_TNFA_SIGNALING_VIA_NFKB
SAFFORD_ANERGY
p = 0.02
NES = -1.5
FDR= 0.07

Figure S1

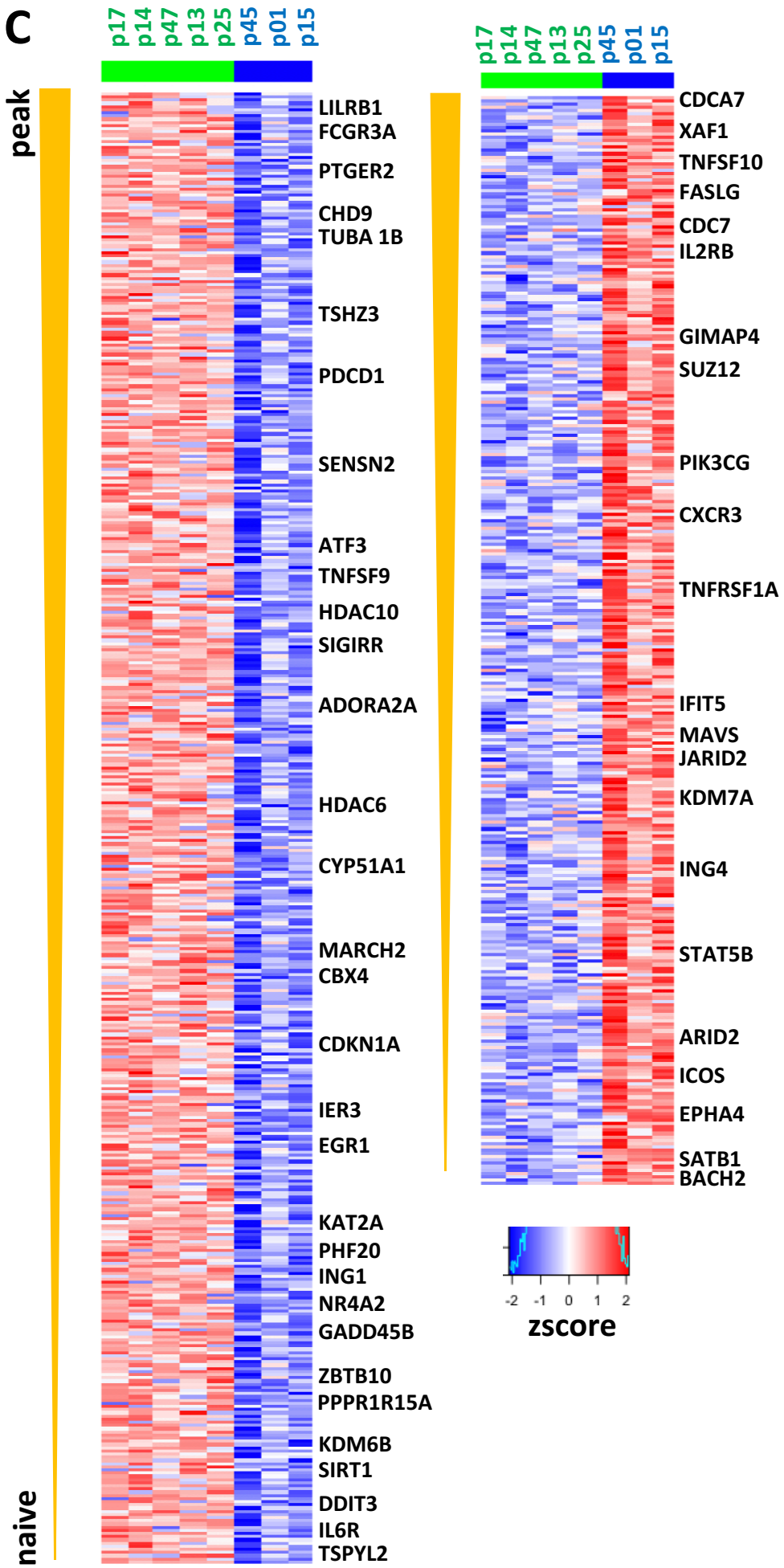

C
p17
p14
p47
p13
p25
p45
p01
p15
peak
naive
LILRB1
FCGR3A
PTGER2
CHD9
TUBA 1B
TSHZ3
PDCD1
SENSN2
ATF3
TNFSF9
HDAC10
SIGIRR
ADORA2A
HDAC6
CYP51A1
MARCH2
CBX4
CDKN1A
IER3
EGR1
KAT2A
PHF20
ING1
NR4A2
GADD45B
ZBTB10
PPPR1R15A
KDM6B
SIRT1
DDIT3
IL6R
TSPYL2
CDCA7
XAF1
TNFSF10
FASLG
CDC7
IL2RB
GIMAP4
SUZ12
PIK3CG
CXCR3
TNFRSF1A
IFIT5
MAVS
JARID2
KDM7A
ING4
STAT5B
ARID2
ICOS
EPHA4
SATB1
BACH2
-2
-1
0
1
2
zscore

Figure S1

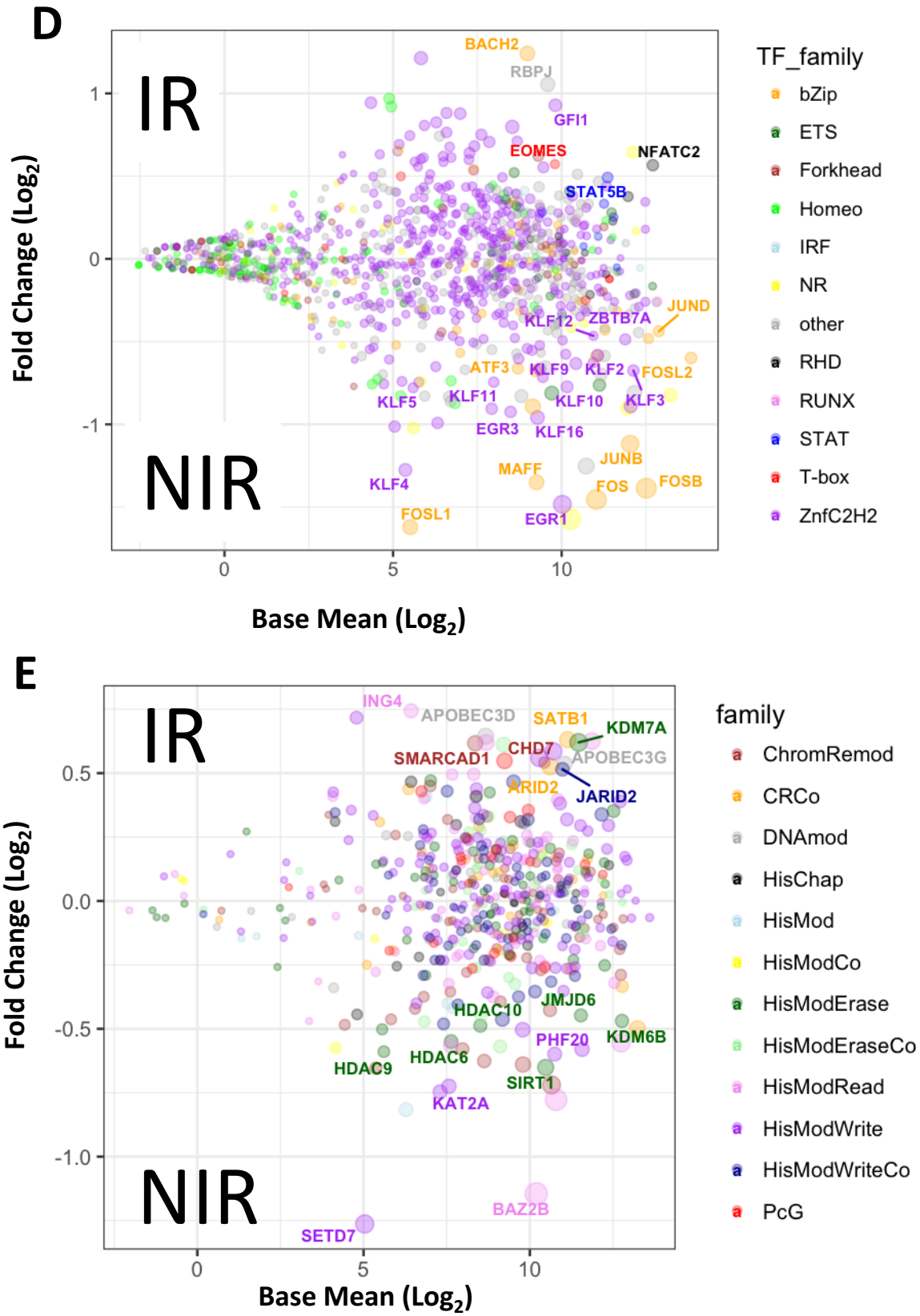
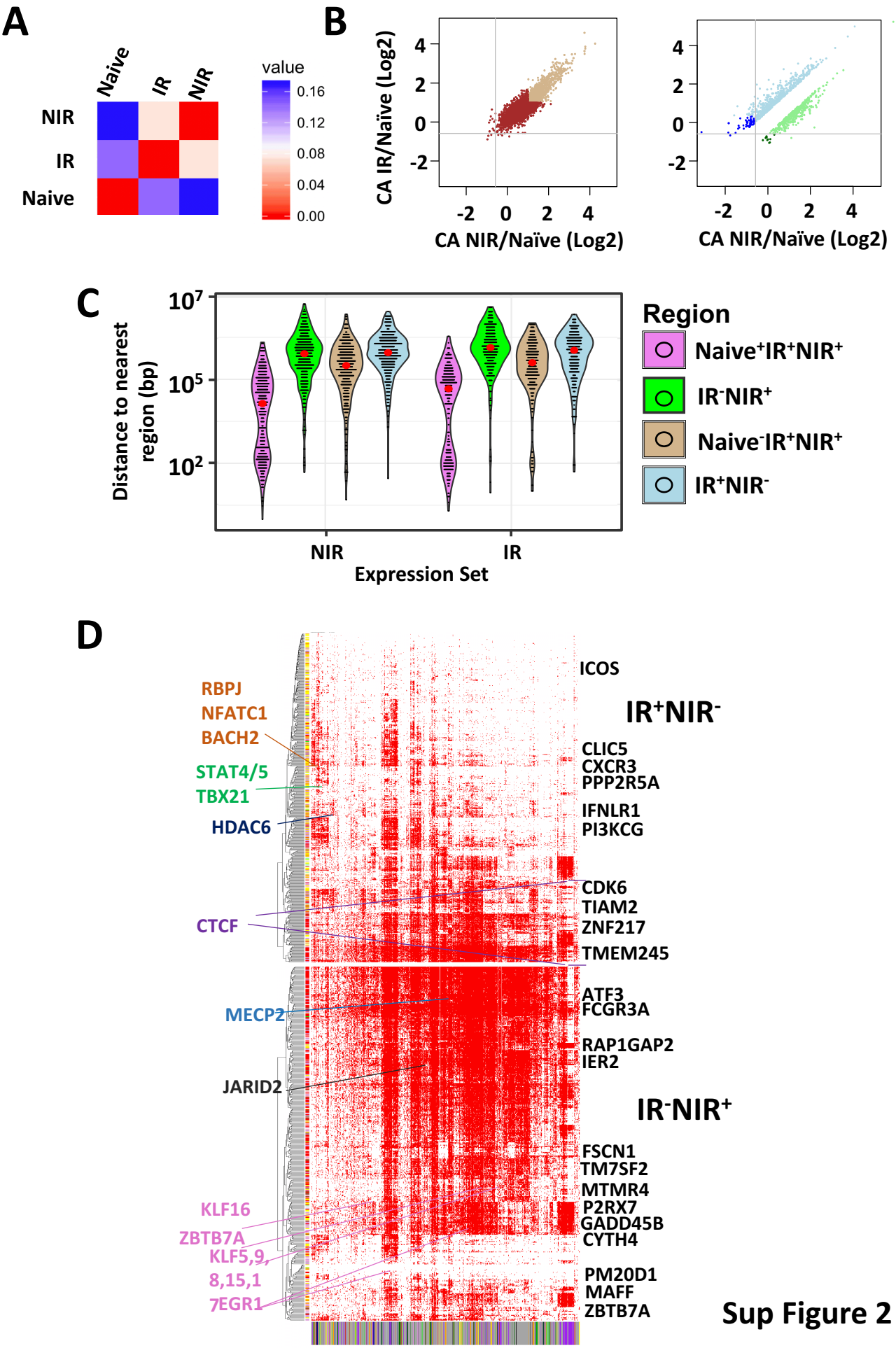

D
IR
NIR
BACH2
RBPJ
GFI1
EOMES
NFATC2
STAT5B
JUND
KLF12
ZBTB7A
ATF3
KLF9
KLF2
FOSL2
KLF5
KLF11
KLF10
KLF3
EGR3
KLF16
JUNB
MAFF
FOS
FOSB
KLF4
FOSL1
EGR1
Fold Change (Log2)
Base Mean (Log2)
TF_family
bZip
ETS
Forkhead
Homeo
IRF
NR
other
RHD
RUNX
STAT
T-box
ZnfC2H2
E
IR
NIR
ING4
APOBEC3D
SATB1
KDM7A
SMARCAD1
CHD7
APOBEC3G
ARID2
JARID2
HDAC10
JMJD6
PHF20
KDM6B
HDAC9
HDAC6
SIRT1
KAT2A
BAZ2B
SETD7
Fold Change (Log2)
Base Mean (Log2)
family
ChromRemod
CRCo
DNAmod
HisChap
HisMod
HisModCo
HisModErase
HisModEraseCo
HisModRead
HisModWrite
HisModWriteCo
PcG

Figure S1

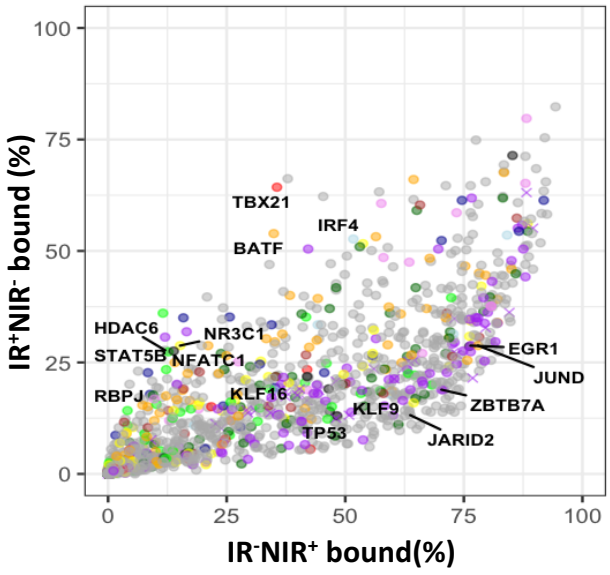
Supplementary Figure 1: Identification and characterisation of gene signatures associated with reactive and non-reactive CMV+ lymphocytes.

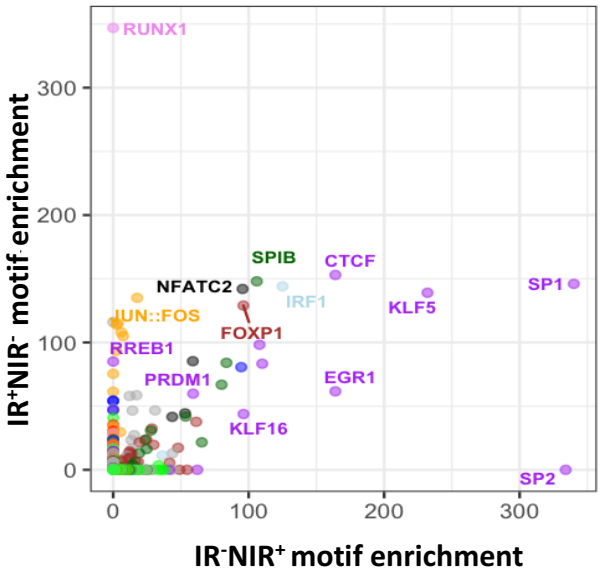
(A) Similarities in transcriptome profiles of CMV-specific memory CD8+ lymphocytes from patients undergoing HSCT were assessed using Pearson correlations. Non-reactive (NIR) patients in green and red and reactive (IR) patients in blue. (B) GSEA and ranking of genes from more in NIR to more in IR reveals enrichment of genesets at either ends of the scale. (C) Genes differentially expressed in lymphocytes from IR and NIR recipients using false discovery rate (FDR) 0.2 cut-off and ranked according to their response in peak CMV infection (GSE12589). Z-score scaled. Expression assessed by rank metric (weighted Signal2Noise). Normalised enrichment score shown. * p-value < 0.05. (D-E) Expression of transcription factors (D) and chromatin modulators (E) in reactive and non-reactive cells. Sized by significance, where larger = more significant. Values incorporate zero-centred normal priors and thus shrinkage is greater for the log2-fold change estimates from genes with low counts and high dispersion.


A
Naive
IR
NIR
NIR
IR
Naive
value
0.16
0.12
0.08
0.04
0.00
B
CA IR/Naïve (Log2)
CA NIR/Naïve (Log2)
CA NIR/Naïve (Log2)
C
Distance to nearest region (bp)
Region
Naive+IR+NIR+
IR-NIR+
Naive-IR+NIR+
IR+NIR-
NIR
IR
Expression Set
D
RBPJ
NFATC1
BACH2
STAT4/5
TBX21
HDAC6
CTCF
MECP2
JARID2
KLF16
ZBTB7A
KLF5,9,
8,15,1
7EGR1
ICOS
IR+NIR-
CLIC5
CXCR3
PPP2R5A
IFNLR1
PI3KCG
CDK6
TIAM2
ZNF217
TMEM245
ATF3
FCGR3A
RAP1GAP2
IER2
IR-NIR+
FSCN1
TM7SF2
MTMR4
P2RX7
GADD45B
CYTH4
PM20D1
MAFF
ZBTB7A

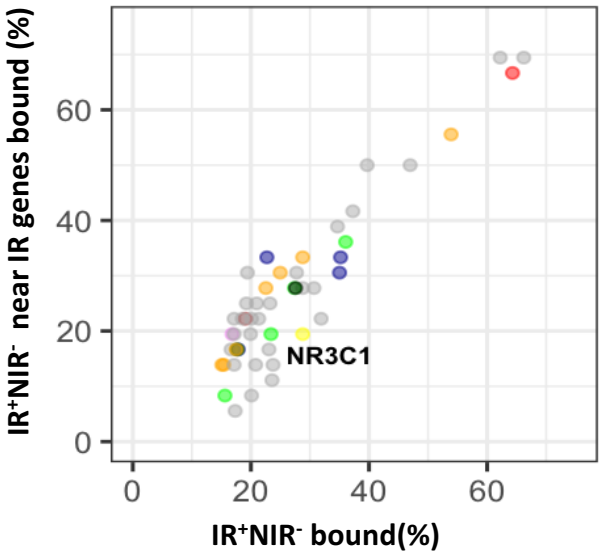
Sup Figure 2

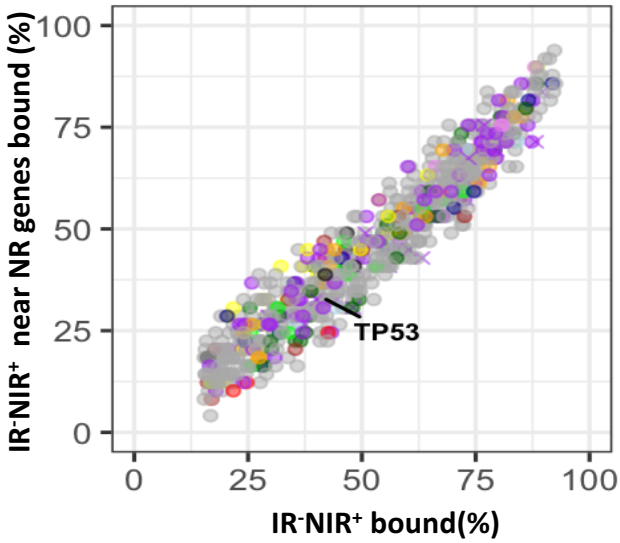
E



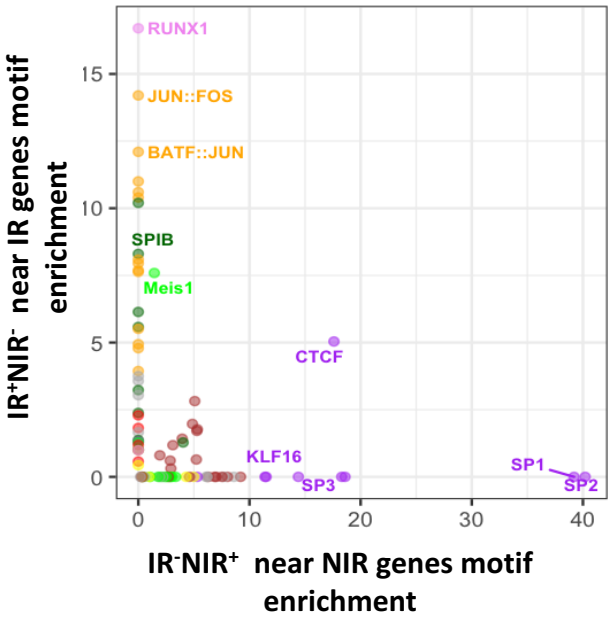


F





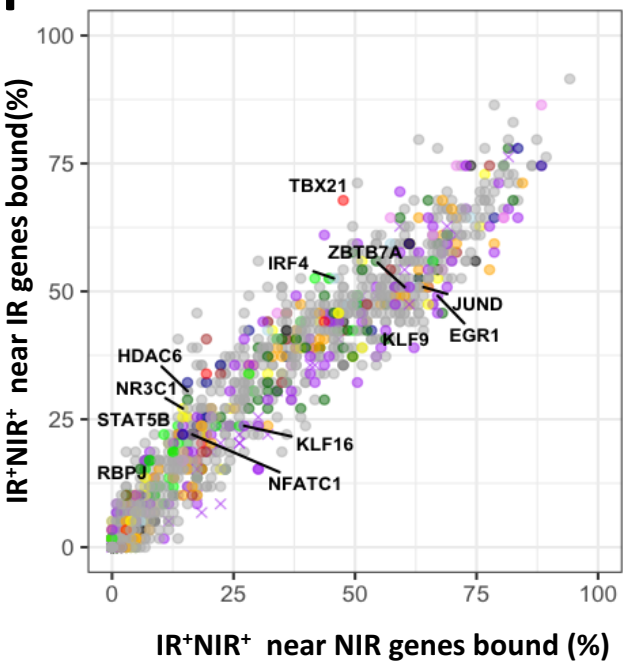
G

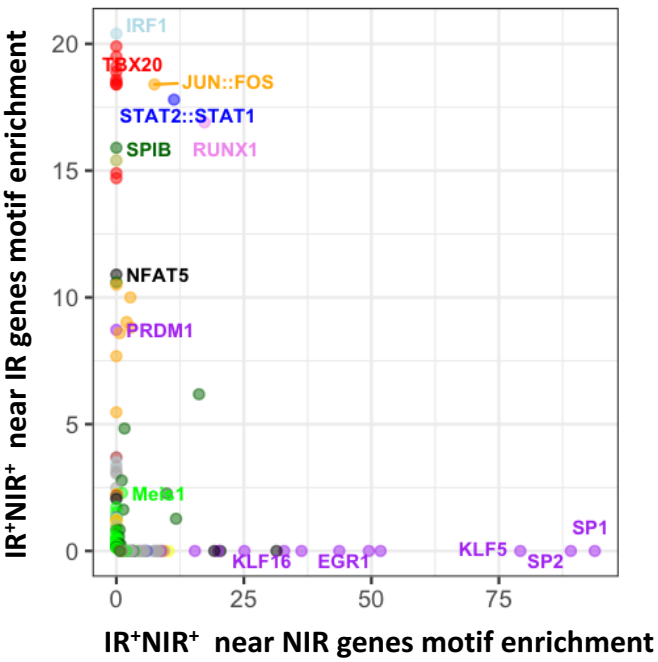


TF_Family
- bZip
- ETS
- Forkhead
- Homeo
- NR
- other
- T-box
- RUNX
- ZnfC2H2

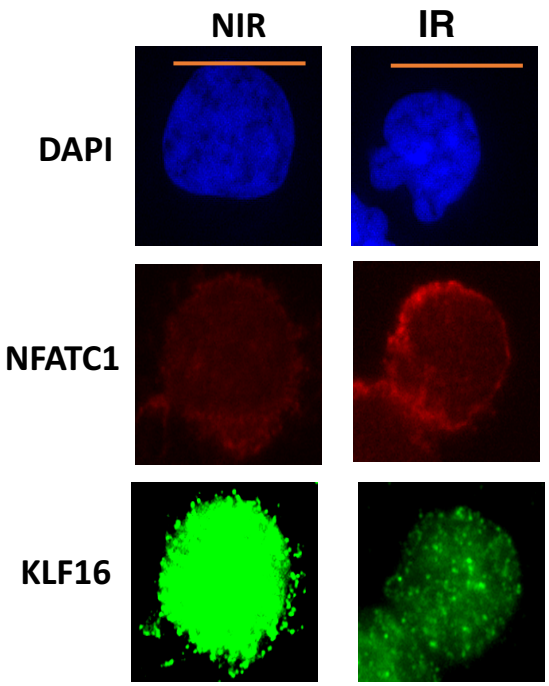
Sup Figure 2

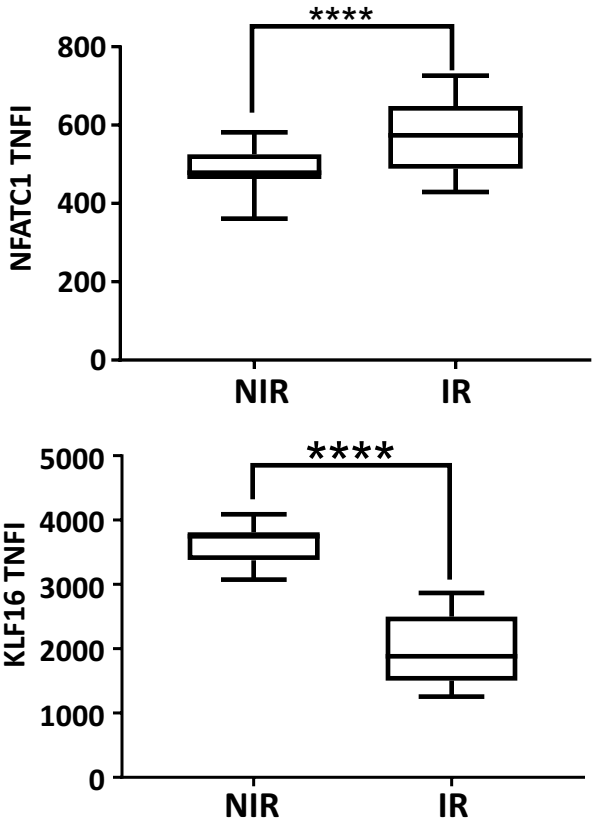
H





I

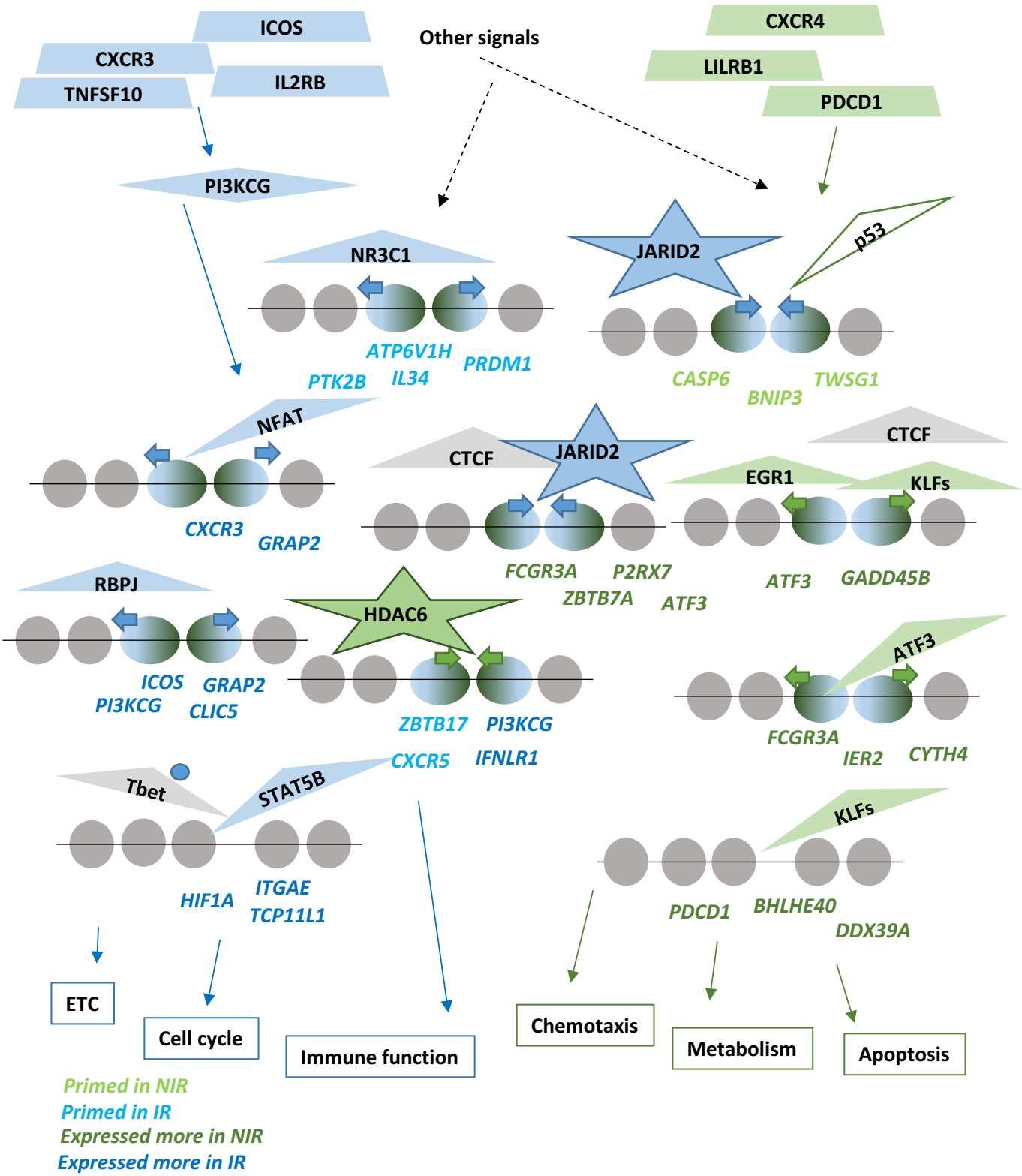




Sup Figure 2

Supplementary Figure 2: Identification and characterisation of chromatin accessibility in reactive and non-reactive CMV+ lymphocytes.

(A) Similarities between the accessibility profiles of CMV-specific memory CD8+ T cells from patients undergoing HSCT and normal naïve CD8+ were assessed using Pearson correlations (B) The groups of regions with differential chromatin accessibility (CA) can be further subgrouped by their CA in naïve cells (C) The distances from the transcriptional start site of genes with higher expression in cells from NIR or IR recipients, to the nearest accessible region. Median shown in red. (D) Factors from ChIP-seq experiments in CISTROME were clustered by whether they bound (red) or not (white) in the $IR^{+}NIR^{-}$ or $IR^{-}NIR^{+}$ accessible regions. Several regions near differentially expressed genes are named. Chromatin state in memory cells CD8+ is shown on left (see **Figure 4F**), colours below show motif family. (E) Transcription factor and chromatin modifier binding in the $IR^{+}NIR^{-}$ and $IR^{-}NIR^{+}$ accessible regions as assessed by the proportion bound in publicly available ChIP-seq data from CISTROME and enrichment of JASPAR binding motifs. Motif enrichment measured by Clover Enrichment Score, against regions with similar GC content. (F) The proportion $IR^{+}NIR^{-}$ and $IR^{-}NIR^{+}$ accessible regions that are within 50 kb of differentially expressed genes that are bound by DNA-binding proteins from the CISTROME ChIP-seq datasets. (G) Enrichment of motifs in $IR^{+}NIR^{-}$ and $IR^{-}NIR^{+}$ accessible regions that are within 50 kb of differentially expressed genes. (H) Factor binding and enrichment of motifs in $naïve^{-}$ $IR^{+}NIR^{+}$ regions that are within 50 kb of differentially expressed genes. (I) Protein levels and localisation of NFATC1 and KLF16 in reactive and non-reactive lymphocytes. Total nuclear (TNFI) or cytoplasmic (TCFI) florescence intensity were measured in a minimum of n = 20 cells for 2 different donors. DAPI shows nucleus. **** $p < 0.0001$ Scale bar is 5μm. (E-H) X is CTCF binding. (G) All significant ($p < 0.05$) motifs plotted. Enrichment measured by Clover Enrichment Score.


ICOS
CXCR3
IL2RB
TNFSF10
Other signals
CXCR4
LILRB1
PDCD1
PI3KCG
NR3C1
ATP6V1H
PRDM1
PTK2B
IL34
JARID2
p53
CASP6
TWSG1
BNIP3
NFAT
CXCR3
GRAP2
CTCF
JARID2
FCGR3A
P2RX7
ZBTB7A
ATF3
CTCF
EGR1
KLFs
ATF3
GADD45B
RBPJ
ICOS
GRAP2
PI3KCG
CLIC5
HDAC6
ZBTB17
PI3KCG
ATF3
FCGR3A
IER2
CYTH4
CXCR5
IFNLR1
Tbet
STAT5B
HIF1A
ITGAE
TCP11L1
KLFs
PDCD1
BHLHE40
DDX39A
ETC
Cell cycle
Immune function
Chemotaxis
Metabolism
Apoptosis
Primed in NIR
Primed in IR
Expressed more in NIR
Expressed more in IR

Sup Figure 3

Supplementary Figure 3: Schematic of the regulation of accessibility in immune reactive (IR) and non-immune reactive (NIR) CMV lymphocytes.

Gene regulation in memory lymphocytes derived from hematopoietic stem cell transplants and responding to CMV infection with different efficiencies. Receptors (trapezoids), kinases (diamonds), transcription factors (triangles), and chromatin remodellers (stars) that are more highly expressed in reactive (blue) or non-reactive (green) cells give rise to chromatin with differential accessibility. In some cases, this contributes to differences in gene expression (dark italics), while in others it may prime genes (light italics) for expression upon receipt of additional signals.