



A microRNA expression and regulatory element activity atlas of the mouse immune system

Samuel A. Rose^{1,2}, Aleksandra Wroblewska^{1,2}, Maxime Dhainaut¹, Hideyuki Yoshida³, Jonathan M. Shaffer⁴, Anela Bektesevic^{1,2}, Benjamin Ben-Zvi^{1,2}, Andrew Rhoads⁵, Edy Y. Kim^{6,7}, Bingfei Yu⁸, Yonit Lavin¹, Miriam Merad^{1,9}, Jason D. Buenrostro^{10,11}, Brian D. Brown^{1,2}✉ and the Immunological Genome Consortium*

To better define the control of immune system regulation, we generated an atlas of microRNA (miRNA) expression from 63 mouse immune cell populations and connected these signatures with assay for transposase-accessible chromatin using sequencing (ATAC-seq), chromatin immunoprecipitation followed by sequencing (ChIP-seq) and nascent RNA profiles to establish a map of miRNA promoter and enhancer usage in immune cells. miRNA complexity was relatively low, with >90% of the miRNA compartment of each population comprising <75 miRNAs; however, each cell type had a unique miRNA signature. Integration of miRNA expression with chromatin accessibility revealed putative regulatory elements for differentially expressed miRNAs, including miR-21a, miR-146a and miR-223. The integrated maps suggest that many miRNAs utilize multiple promoters to reach high abundance and identified dominant and divergent miRNA regulatory elements between lineages and during development that may be used by clustered miRNAs, such as miR-99a/let-7c/miR-125b, to achieve distinct expression. These studies, with web-accessible data, help delineate the cis-regulatory elements controlling miRNA signatures of the immune system.

miRNAs are a large class of noncoding RNAs that post-transcriptionally regulate gene expression¹ and are critical for proper immune cell development and function^{2–4}. Identification of the miRNA expression profile of each immune cell is an important goal towards understanding the regulatory network of the immune system. Molecular profiling has been carried out to determine the miRNA profiles of different immune cells^{5–13}, which has aided in the identification of biologically relevant miRNAs. However, most of these studies tended to focus primarily on closely related hematopoietic populations, such as B and T lymphocytes or macrophages^{5,6,9–11,13}, and this limits possibilities for comparative analysis across the immune system.

Even less well characterized is the promoter and enhancer landscape that controls the miRNA expression states of immune cells. Certain studies have identified some of the *cis*-regulatory elements controlling the expression of specific immune-functioning miRNAs, such as miR-223 (refs. ^{14,15}), miR-146a¹⁶ and miR-181 (ref. ¹⁷). While these and other studies have provided insights into the regulation of particular miRNAs in specific cellular contexts, global analyses of miRNA *cis*-regulation have been limited. miRNA promoter and enhancer sites have been mapped by inference, using histone modifications and other molecular data, or by using empirical sequencing-based methods in cells with limited *Drosha* activity^{18–23}. However, these studies were performed mostly in cell lines or bulk tissues and generally lacked paired measurements of mature miRNA abundance and regulatory element activity, precluding opportunities

to link specific *cis*-elements to miRNA expression. Thus, we currently lack an understanding of the *cis*-regulatory elements controlling miRNA expression and their activity patterns in the immune system.

Here we built a resource of miRNA expression profiles from 63 primary mouse immune cell populations spanning hematopoietic and stromal lineages, and connected these profiles to open chromatin regions (OCRs). miRNA complexity in each cell type was relatively low, with >90% of the miRNA compartment of each population comprised of <75 miRNAs; however, each cell type and cell state had a unique miRNA signature. An integrated analysis of miRNA expression with chromatin accessibility across immune cells established the active miRNA promoter and enhancer landscape of the immune system. This revealed divergent and convergent miRNA-associated *cis*-regulatory elements used between immune populations, and provided evidence of dominant or additive effects of different promoters for the same miRNA. The use of multiple promoters supports a means to reach the high abundance required for miRNAs to achieve suppressive activity. In addition to public release, these data have been integrated into a web-based browser (<http://shiny.immgen.org/mirna-browser/>) enabling queries of miRNA abundance estimates alone or with respect to predicted binding sites on individual transcripts.

Results

A miRNA expression atlas of 63 immune populations. To obtain an atlas of miRNA signatures across the immune system we isolated

¹Precision Immunology Institute, Icahn School of Medicine at Mount Sinai, New York, NY, USA. ²Department of Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, New York, NY, USA. ³YCI Laboratory for Immunological Transcriptomics, RIKEN Center for Integrative Medical Sciences, Kanagawa, Japan. ⁴Qiagen, Frederick, MA, USA. ⁵Department of Pathology, Harvard Medical School, Boston, MA, USA. ⁶Division of Pulmonary and Critical Care Medicine, Brigham and Women's Hospital, Boston, MA, USA. ⁷Harvard Medical School, Boston, MA, USA. ⁸Division of Biology, University of California San Diego, La Jolla, CA, USA. ⁹Department of Oncological Sciences, Icahn School of Medicine at Mount Sinai, New York, NY, USA. ¹⁰Broad Institute of MIT and Harvard, Cambridge, MA, USA. ¹¹Department of Stem Cell and Regenerative Biology, Harvard University, Cambridge, MA, USA.

*A list of authors and their affiliations appears at the end of the paper. ✉e-mail: brian.brown@mssm.edu

Table 1 | Cell types profiled

Cell type	Abbreviation	Organ	Markers	Gender
Peritoneal macrophages	MF.PC	Peritoneal cavity	ICAM2 ⁺ F4/80 ⁺	Male
Peritoneal macrophages + poly(I:C)	MF.pIC.PC	Peritoneal cavity	ICAM2 ⁺ F4/80 ⁺	Male
Peritoneal macrophages	MF.fem.PC	Peritoneal cavity	ICAM2 ⁺ F4/80 ⁺	Female
Microglia	MF.microglia.CNS	Brain	CD45 ⁺ F4/80 ⁺ CD11b ⁺	Male
Lung macrophages	MF.Alv.Lu	Lung	F4/80 ⁺ CD11c ⁺ SiglecF ⁺	Male
Lung macrophages + poly(I:C)	MF.pIC.Alv.Lu	Lung	F4/80 ⁺ CD11c ⁺ SiglecF ⁺	Male
Spleen macrophages	MF.RP.Sp	Spleen	F4/80 ⁺ CD11b ⁺ CD11c ^{inter}	Male
LI macrophages	MF.LI	Large intestine	CD11b ⁺ CD64 ⁺ CD11c ^{inter}	Male
SI macrophages	MF.SI	Small intestine	CD11b ⁺ CD64 ⁺ CD11c ^{inter}	Male
Ly6C ^{high} monocytes	Mo.6C ⁺ II ⁻ .Bl	Blood	B220 ⁻ CD3 ⁻ Ly6G ⁻ CD115 ⁺ CD11b ⁺ Ly6C ^{hi}	Male
Ly6C ^{low} monocytes	Mo.6C ⁻ II ⁻ .Bl	Blood	B220 ⁻ CD3 ⁻ Ly6G ⁻ CD115 ⁺ CD11b ⁺ Ly6C ^{lo}	Male
Ly6C ^{high} monocytes	Mo.6C ⁺ II ⁻ .BM	Bone marrow	B220 ⁻ CD3 ⁻ Ly6G ⁻ CD117 ⁻ CD115 ⁺ CD11b ⁺ Ly6C ^{hi}	Male
Dendritic cells	DC.Sp	Spleen	TCRb ⁻ B220 ⁻ CD11c ⁺ MHCII ^{high} Fit3 ⁺	Male
Neutrophils	GN.Sp	Spleen	TCRb ⁻ B220 ⁻ CD11c ⁻ Gr1 ⁺	Male
Neutrophils	GN.fem.Sp	Spleen	TCRb ⁻ B220 ⁻ CD11c ⁻ Gr1 ⁺	Female
Naive peritoneal mast cells	MC.PC	Peritoneal cavity	FceR1α ⁺ CD117 ⁺	Male
IgE-sensitized peritoneal mast cells	MC.IgE.PC	Peritoneal cavity	CD45 ⁺ CD11b ⁻ CD11c ⁻ CD19 ⁻ CD4 ⁻ CD8 ⁻ FceR1α ⁺ CD117 ⁺	Male
IgE-activated peritoneal mast cells	MC.IgEX.PC	Peritoneal cavity	CD45 ⁺ CD11b ⁻ CD11c ⁻ CD19 ⁻ CD4 ⁻ CD8 ⁻ FceR1α ⁺ CD117 ⁺	Male
CD19 ⁺ B cells	B.Sp	Spleen	CD19 ⁺ IgM ⁺	Male
CD19 ⁺ B cells	B.fem.Sp	Spleen	CD19 ⁺ IgM ⁺	Female
CD19 ⁺ CD93 ⁺ CD5 ⁺ B cells	B.93 ⁺ .Sp	Spleen	Ly6C ⁻ Ter119 ⁻ Gr1 ⁻ CD11c ⁻ CD19 ⁺ CD23 ⁻ AA4 ⁺ CD5 ⁺	Male
CD19 ⁺ CD93 ⁻ CD5 ⁺ B cells	B.93 ⁻ .Sp	Spleen	Ly6C ⁻ Ter119 ⁻ Gr1 ⁻ CD11c ⁻ CD19 ⁺ CD23 ⁻ AA4 ⁻ CD5 ⁺	Male
Follicular B cells	B.Fo.Sp	Spleen	CD5 ⁻	Male
Activated follicular B cells	B.Fo.algM.Sp	Spleen	Ly6C ⁻ Ter119 ⁻ Gr1 ⁻ CD11c ⁻ CD19 ⁺ CD23 ⁺ AA4 ⁻ CD5 ⁻	Male
B1a B cells	B1b.PC	Peritoneal cavity	CD19 ⁺ CD5 ⁺ CD43 ⁺	Male
Activated B1a B cells	B1b.algM.PC	Peritoneal cavity	CD19 ⁺ CD5 ⁺ CD43 ⁺	Male
Thymic CD4 ⁺ CD8 ⁺ CD3 ^{lo} DP	T.DP.Th	Thymus	CD4 ⁺ CD8 ⁺ CD3 ^{lo}	Male
Naive CD8 T cells	T8.Nve.Sp	Spleen	TCRb ⁺ CD4 ⁻ CD8 ⁺	Male
CD8 T cells CD44 ⁻ OT-I	T8.OT1.Sp	Spleen	B220 ⁻ CD4 ⁻ MHCII ⁻ NK1.1 ⁻ Gr-1 ⁻ CD8 ⁺ CD44 ⁻ CD45.1 ⁺	Male
CD8 T cells KLRG1 ⁻ CD127 ⁺ + LisOVA, day8	T8.MP.LisOVA.d8.Sp	Spleen	B220 ⁻ CD4 ⁻ MHCII ⁻ NK1.1 ⁻ Gr-1 ⁻ CD8 ⁺ CD45.1 ⁺ KLRG1 ⁻ CD127 ⁺	Male
CD8 T cells KLRG1 ⁺ CD127 ⁻ + LisOVA, day8	T8.TE.LisOVA.d8.Sp	Spleen	B220 ⁻ CD4 ⁻ MHCII ⁻ NK1.1 ⁻ Gr-1 ⁻ CD8 ⁺ CD45.1 ⁺ KLRG1 ⁺ CD127 ⁻	Male
CD8 memory T cells KLRG1 ⁺ 127 ⁺ + LisOVA, day 60	T8.mem.LisOVA.d60.Sp	Spleen	B220 ⁻ CD4 ⁻ MHCII ⁻ NK1.1 ⁻ Gr-1 ⁻ CD8 ⁺ CD45.1 ⁺ KLRG1 ⁺ CD127 ⁺	Male
Naive CD4 T cells	T4.Nve.Sp	Spleen	TCRb ⁺ CD4 ⁺ CD8 ⁻	Male
Activated T4	T4.Vb8 ⁺ .SEB.Sp	Spleen	TCRb ⁺ CD4 ⁺ CD8 ⁻ Vb8 ⁺	Male
Activated T4	T8.Vb8 ⁻ .SEB.Sp	Spleen	TCRb ⁺ CD4 ⁺ CD8 ⁻ Vb8 ⁻	Male
Regulatory T cells	Treg.4.25 ^{hi} .Sp	Spleen	TCRb ⁺ CD4 ⁺ CD25 ⁺	Male
Regulatory T cells	Treg.4.FP3 ⁺ .Co	Large intestine	CD4 ⁺ CD25 ⁺ FoxP3 ⁺	Male
Naive Tgd	Tgd.Sp	Spleen	CD8 ⁻ CD19 ⁻ Gr-1 ⁻ Ter119 ⁻ TCRb ⁻ TCRgd ⁺	Male
Tgd CD27 ⁺	Tgd.27 ⁺ .LN	Lymph nodes	TCRgd ⁺ CD27 ⁺	Male
Tgd CCR6 ⁺	Tgd.CCR6 ⁺ .LN	Lymph nodes	TCRgd ⁺ CCR6 ⁺	Male
Tgd CD27 ⁺ , activated	Tgd.27 ⁺ .aCD3.LN	Lymph nodes	TCRgd ⁺ CD27 ⁺	Male
Tgd CCR6 ⁺ , activated	Tgd.CCR6 ⁺ .aCD3.LN	Lymph nodes	TCRgd ⁺ CCR6 ⁺	Male
NKT cells	NKT.Sp	Spleen	CD8 ⁻ CD19 ⁻ Gr-1 ⁻ Ter119 ⁻ NK1.1 ^{inter} TCRb ^{inter} : replicates 1–3 CD1d-tetramer ⁺ (loaded with αGalCer): replicates 4–5	Male
NKT cells 50-μg LPS, 3h	NKT.Sp.LPS.h3.50	Spleen	CD1d-tetramer ⁺ (loaded with αGalCer)	Male

Continued

Table 1 | Cell types profiled (continued)

Cell type	Abbreviation	Organ	Markers	Gender
NKT cells 165-µg LPS, 3 h	NKT.Sp.LPS.h3.165	Spleen	CD1d-tetramer ⁺ (loaded with αGalCer)	Male
NKT cells 50-µg LPS, 1 day	NKT.Sp.LPS.d1.50	Spleen	CD1d-tetramer ⁺ (loaded with αGalCer)	Male
NKT cells 165-µg LPS, 1 day	NKT.Sp.LPS.d1.165	Spleen	CD1d-tetramer ⁺ (loaded with αGalCer)	Male
NKT cells 50-µg LPS, 3 day	NKT.Sp.LPS.d3.50	Spleen	CD1d-tetramer ⁺ (loaded with αGalCer)	Male
NKT cells 165-µg LPS, 3 day	NKT.Sp.LPS.d3.50	Spleen	CD1d-tetramer ⁺ (loaded with αGalCer)	Male
Spleen NK	NK.Sp	Spleen	CD8 ⁻ CD19 ⁻ Gr-1 ⁻ Ter119 ⁻ NK1.1 ⁺	Male
Spleen NK, infected (MCMV)	NK.MCMV.Sp	Spleen	CD8 ⁻ CD19 ⁻ Gr-1 ⁻ Ter119 ⁻ NK1.1 ⁺	Male
DX5 ⁺ NK cells, steady state	NK.49b ⁺ .Lv	Liver	NK1.1 ⁺ CD49b ⁺	Male
DX5 ⁺ NK cells, infected (MCMV)	NK.49b ⁺ .MCMV.Lv	Liver	NK1.1 ⁺ CD49b ⁺	Male
DX5 ⁻ NK cells, steady state	NK.49b ⁻ .Lv	Liver	NK1.1 ⁺ CD49b ⁻	Male
DX5 ⁻ NK cells, infected (MCMV)	NK.49b ⁻ .MCMV.Lv	Liver	NK1.1 ⁺ CD49b ⁻	Male
Fibroblastic reticular cells	FRC.SLN	Lymph nodes	CD45 ⁺ gp38 ⁺ CD31 ⁻	Male
Blood endothelial cells	BEC.SLN	Lymph nodes	CD45 ⁺ gp38 ⁻ CD31 ⁺	Male
Lymphatic endothelial cells	LEC.SLN	Lymph nodes	CD45 ⁺ gp38 ⁺ CD31 ⁺	Male
Double-negative	DN.SLN	Lymph nodes	CD45 ⁺ gp38 ⁻ CD31 ⁻	Male
FRC, LCMV activated	FRC.LCMV.SLN	Lymph nodes	CD45 ⁺ gp38 ⁺ CD31 ⁻	Male
BEC, LCMV activated	BEC.LCMV.SLN	Lymph nodes	CD45 ⁺ gp38 ⁻ CD31 ⁺	Male
LEC, LCMV activated	LEC.LCMV.SLN	Lymph nodes	CD45 ⁺ gp38 ⁺ CD31 ⁺	Male
DN, LCMV activated	DN.LCMV.SLN	Lymph nodes	CD45 ⁺ gp38 ⁻ CD31 ⁻	Male

63 populations of immune and stromal cells (Table 1). This included cells in steady state, as well as the same populations isolated after injection of a pathogen, pathogen-associated molecule or other stimuli. All cells were isolated according to a standard ImmGen protocol and double-flow sorted for improved purity.

miRNA profiling was performed by quantitative PCR (qPCR) array to measure the expression of 923 different miRNAs. qPCR was chosen because it was reported by the miRQC study to have the highest sensitivity and specificity with the lowest input requirement for miRNA profiling, including comparisons with sequencing²⁴. Although qPCR cannot be used for discovery of new miRNAs, sequencing-based approaches are prone to biases that affect absolute miRNA quantification^{25,26}. Because all conserved mammalian miRNAs have probably been discovered¹, we opted to focus on quantification of known miRNAs to expand the cellular catalog and to integrate these results with orthogonal datasets. Importantly, the 923 miRNAs analyzed represent the vast majority of confidently annotated and conserved miRNAs.

Data were normalized by creating a scaling factor across samples based on the arithmetic mean of all miRNA assays. Threshold cycle (Ct) values of expression were converted to linear units (arbitrary units, AU) by calibration to the limit of reliable detection as determined by a standard curve of ten different synthetic miRNAs, and by inter-replicate variability (Extended Data Fig. 1a,b and Supplementary Table 1). Importantly, standard curves showed no clear amplification or detection differences, indicating little bias and supporting the reliability of the platform. Finally, we performed an *in vivo* sensor assay, as previously described²⁷, to measure suppressive activity on a target transcript in relation to miRNA expression levels (Extended Data Fig. 1c). This indicated that miRNAs expressed >32 AU were capable of suppression *in vivo*, and that our abundance measurements were correlated to suppressive activity for multiple miRNAs across cell types (Extended Data Fig. 1d,e). The data generated here represent the broadest available compendium of standardized miRNA expression values for mouse immune cells.

The miRNA compartment of immune cells is low complexity. Of 923 mature miRNAs assayed, we detected 442 in at least one cell type. On average, each cell type expressed ~185 unique miRNAs at an expression level within the reliable range of detection (>1 AU; Fig. 1a), consistent with prior studies using both sequencing and qPCR^{5,6}. Considering a more stringent high-abundance expression cutoff corresponding to detectable suppressive activity in the *in vivo* suppressor assay (>32 AU), the average miRNA compartment of each immune population comprised ~50 unique miRNAs (Fig. 1a). It is worth noting that there may be miRNAs reliably detected at lower abundance (1–32 AU) that are highly expressed in an uncharacterized subpopulation but, even for a subpopulation present at 10%, miRNAs detected <4 AU would still be below or close to threshold. Cumulative proportional abundance distributions indicated that >75% of the total miRNA molecules within a cell type contained <25 mature species of miRNA, with >50% of mature miRNAs within most immune cell types explained by the five most highly expressed miRNAs (Fig. 1b and Extended Data Fig. 1f). These data suggest that the active miRNA compartment of immune cells is smaller and less complex than may be generally appreciated.

Unique miRNA signatures in immune and stromal populations.

Despite the low complexity of abundant miRNAs within immune cells, differentiating features were present across all cell types and states profiled. miRNA coexpression profiles segregated primarily with respect to developmental lineage (Fig. 1c). A small number of miRNAs was enriched in the three main lymphoid lineages relative to each other, in addition to miR-125a-5p being present in natural killer (NK) and T cells but not in B cells (Fig. 2a). Notably, myeloid cells showed substantial variation with respect to both tissue and cell type. Mast cells clustered distinctly from all other cell types while macrophages had differences across tissues, of which the most distinct were lung macrophages and microglia. Amongst the various tissue-resident macrophages we identified 11 miRNAs specifically enriched in lung macrophages, 11 in microglia, five in intestinal macrophages and two in peritoneal cavity (PC) macrophages

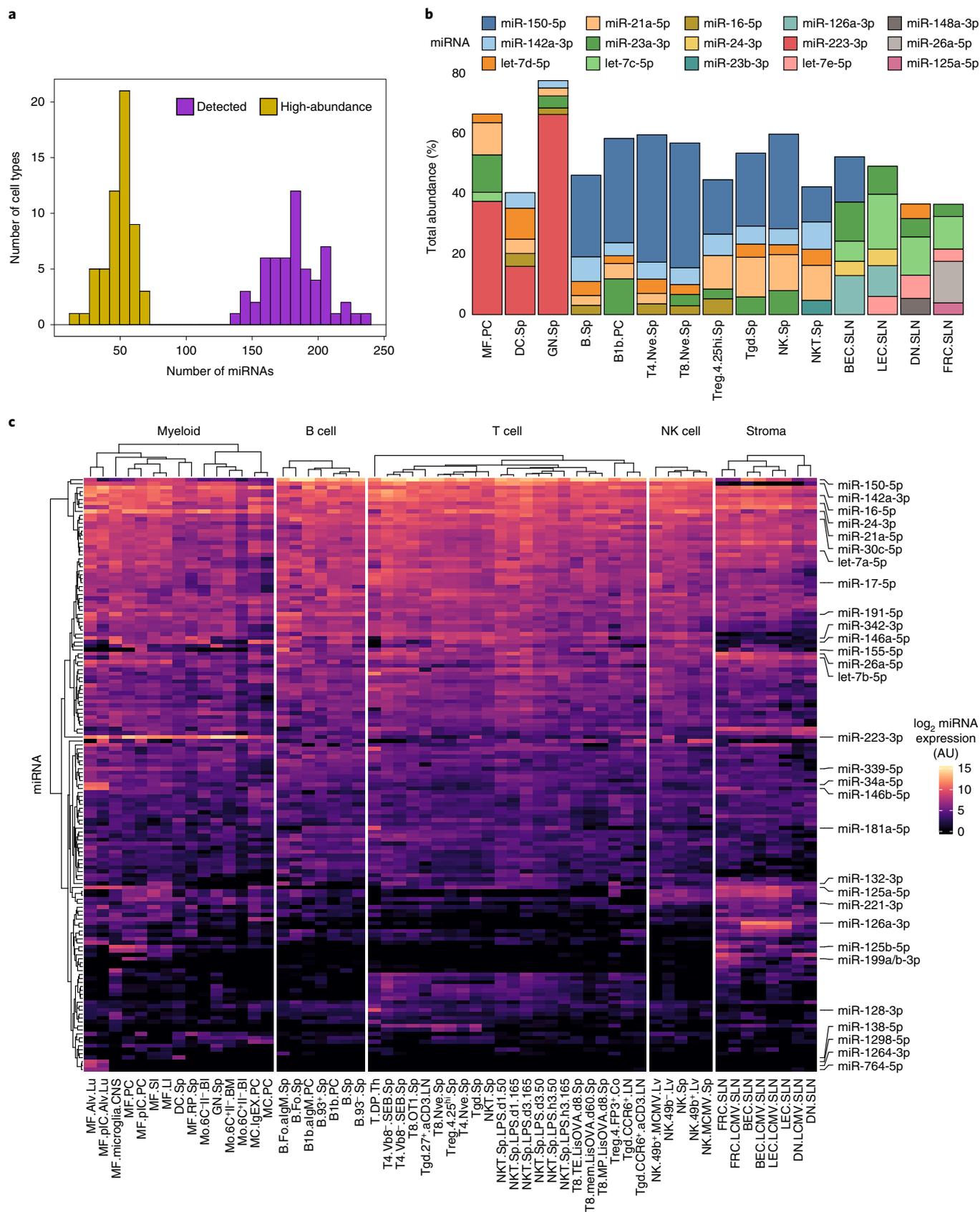


Fig. 1 | The miRNome of immune cells. a, The number of miRNAs passing detection (>1 AU) and high-abundance (>32 AU) cutoffs in all cell types profiled. To pass either threshold, a given miRNA's abundance had to be over the threshold in all replicates from a given cell type ($n=1-5$). **b**, Percentage of total miRNA abundance of the five most highly expressed miRNAs within the indicated cell populations. **c**, Absolute expression (\log_2 AU) of all miRNAs expressed above the high-abundance threshold across the cell compendium (150 in total). Expression values <0 \log_2 AU were set to 0. miRNAs are hierarchically clustered by Euclidean distance and complete linkage; immune populations are clustered by Pearson correlation distance and complete linkage.

(Fig. 2b). Some signatures reflected tissue context, such as expression of the inflammation-induced miRNA miR-155-5p³ in intestinal macrophages. Seven miRNAs (miR-1264-3/5p, miR-1298-5p, miR-764-5p, miR-448-3p and miR-1912-3/5p) expressed exclusively by lung macrophages that map to the intronic sequence of a single gene, *Htr2c*. Interestingly, while the expression of *Htr2c* is relatively modest within the coding transcriptome of lung macrophages (Extended Data Fig. 2b), the *Htr2c* miRNA cluster was amongst the most highly expressed miRNAs in these cells, implying that the dominant role of the locus may be miRNA output.

miRNA profiles of mature, steady-state T cell subsets were ~90% correlated (Extended Data Fig. 2a), although a small set of miRNAs distinguished each of the different populations (Fig. 2c). Of these, several were enriched in one subset of T cells, including miR-223, which was specific for $\gamma\delta$ T cells, and the miR-15/16 cluster found downregulated in iNKT cells. Tissue of residence had variegated effects across lymphoid cells, where colonic regulatory T cells (Tregs) showed large differences from splenic Tregs but NK cells from the liver and B cells from the PC were generally similar to their splenic counterparts (Fig. 1c). Tregs from the large intestine had changes in miRNAs linked to regulation of T cell activation and differentiation, such as the miR-23~27~24 clusters²⁸ and miR-342 (ref. ²⁹) (Extended Data Fig. 2c), possibly reflecting tissue-specific regulation of these processes or T cell activation levels within the gut.

Twenty-two miRNAs showed consistent changes in three or more perturbation conditions. miR-155 and miR-147 were upregulated in a broad set of stimuli consistent with expectations³⁰, the miR-132/212 cluster was most induced in both B cell populations treated with anti-IgM, several miR-17 family members were upregulated in liver NK cells after mouse cytomegalovirus (MCMV) infection and NKT cells treated with LPS, and miR-7 was notably induced in several stromal populations following lymphocytic choriomeningitis virus (LCMV) infection (Fig. 2d and Extended Data Fig. 2d). Additionally, miR-423-3p and miR-342-3p appeared to be downregulated across many perturbations, a trend previously observed for miR-342-3p in T cells³¹.

Given the breadth of our study, we also sought miRNAs with highly restricted expression. To identify cell-specific miRNAs, the tissue specificity index Tau (τ) was calculated for all expressed miRNAs in steady state. We identified 31 miRNAs having high τ scores (>0.92), with $>45\%$ of their abundance confined to a single cell type (Fig. 2e). Several of these may represent tissue-dependent expression patterns where our dataset had only one representative cell type. For example, miR-9-5p was highly restricted to microglia and is reported to be expressed by other cells in the brain³². However, other miRNAs were found to be predominately expressed in specific cell subsets from a repeatedly sampled tissue, such as miR-138-5p in dendritic cells (DCs) from the spleen and miR-214-3p in fibroblastic reticular cells (FRCs) from the lymph node.

These data identify unique miRNA modules between immune cell types which, in some cell types, include the expression of highly specific miRNAs.

The miRNA promoter/enhancer landscape of the immune system. As noted, there has been relatively little examination of the genome-wide regulatory landscape of miRNAs in the immune system. To address this gap, we integrated miRNA expression profiles with chromatin accessibility data from a parallel study including assay for a transposase-accessible chromatin using sequencing (ATAC-seq) profiles of 90 immune cell populations³³ (<http://rstats.immgen.org/Chromatin/chromatin.html>). To map promoter usage, we overlaid the ATAC-seq data onto well-established miRNA transcriptional start site (TSS)/promoter annotations. The annotations were based on seven separate sources that used inference- and sequencing-based methods^{18–22}, as well as overlapping host transcripts from GenCode³⁴. To ensure promoter function in immune cells, we used available biochemical data (Methods) in populations similar to those in our ATAC-seq analysis, including histone mark chromatin immunoprecipitation sequencing (ChIP-seq), RNA polymerase II (POLR2A) and global run-on sequencing (GRO-seq) to restrict mapping to annotated promoters enriched with H3K4me3 marks, in addition to filtering based on POLR2A activity and nascent RNA. This identified 1,315 miRNA TSS/promoter OCRs connected to 323 miRNAs detected in our immune-cell profiling (Extended Data Fig. 3a,b). While not all annotated promoters may function as a promoter in every cell type, we did find that in similar populations there were nascent transcripts spanning from promoter/TSS to miRNA at the vast majority of annotated miRNA promoters found open (1,398/1,653 promoter-to-miRNA connections), supporting their function as promoters in immune cells. Annotated promoter OCRs within 2 kb of each other were further grouped into distinct promoter regions for subsequent analyses.

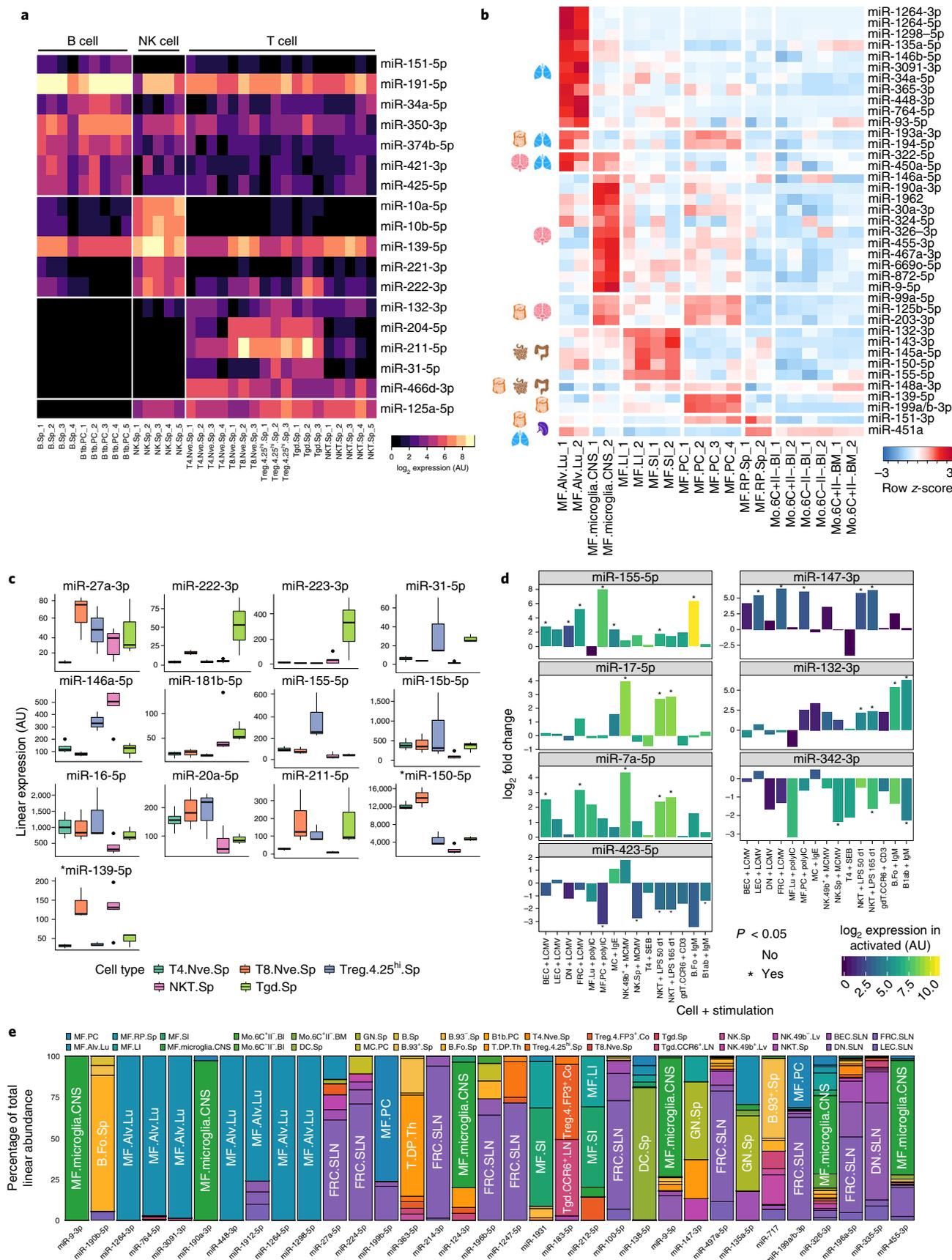
To identify distal *cis*-regulatory elements (dCREs), such as enhancers, that control miRNA expression we used an approach similar to the ImmGen *cis*-Atlas³³. In 22 samples with miRNA expression and ATAC-seq data (Extended Data Fig. 3c), the accessibility at OCRs within 100 kb of annotated miRNA promoter regions was tested for correlation with miRNA expression. This yielded 395 significant associations (adjusted $P < 0.1$) for 140/344 miRNAs, the majority of which (331/395) were positive correlations and thus represent potential enhancer elements of the miRNA (Fig. 3a). Most of the associated dCREs tended to be in close proximity to a miRNA TSS/promoter, with $>50\%$ of dCRE associations occurring within 15 kb of the promoter (Extended Data Fig. 3d). Over 95% of dCREs were located in a topologically associated domain (TAD) with the miRNA promoter, based on TAD boundaries as determined by Johanson et al.³⁵ using *in situ* Hi-C in four cell types also analyzed in our study (Extended Data Fig. 3e). Some miRNAs had a large number (over ten) of significantly associated open dCREs (Extended Data Fig. 3f), potentially representing coordinately activated domains.

Chromatin accessibility often serves as a proxy for CRE activity yet other molecular features, including H3K27ac and transcriptional activity, are more accurate predictors of CRE function. Therefore we used public H3K27ac and H3K4me1 ChIP-seq as

Fig. 2 | Population and cell type miRNA signatures. **a**, Lymphoid-lineage-enriched miRNAs. Heatmap presents miRNAs enriched in the indicated lineage as compared to other lymphoid lineages (limma two-sided adjusted $P < 0.05$ and \log_2 fold change (FC) > 1.5). Only one of 13 T cell-enriched miR-466-669 cluster members is displayed. **b**, Tissue-resident macrophage-enriched miRNAs. Heatmap presents miRNAs enriched in one tissue macrophage population versus at least three out of four other tissue macrophages (limma two-sided adjusted $P < 0.1$ and \log_2 FC > 3). **c**, miRNAs differentially expressed between different subsets of T cells. miRNAs were selected based on limma two-sided test, unadjusted $P < 0.1$ and \log_2 FC > 1.25 , as compared to at least three out of the other four populations. * indicates miRNA just below the threshold. For all boxplots in this manuscript, upper and lower border of boxes represent the 25th and 75th percentile, respectively, and the middle line is the median. Whiskers extend $1.5 \times$ the interquartile range, and points represent values beyond the whiskers. $n = 3$ for Treg, gdT and CD8; $n = 4$ for CD4; $n = 5$ for NKT. **d**, \log_2 FC of all perturbation comparisons for seven selected, broadly reactive miRNAs. Bars are shaded by mean \log_2 expression in the perturbed samples and are marked with an asterisk if the miRNA was differentially expressed in the perturbation (limma two-sided test unadjusted $P < 0.05$, \log_2 FC > 1 and expression > 4 AU in perturbed or > 32 AU in steady-state population). $n = 2$ for all steady-state and stimulated populations except steady-state MF.PC ($n = 4$) and NK.Sp, NKT and B1ab ($n = 5$). **e**, Cell-type-restricted miRNAs. Graph presents miRNAs with $\tau > 0.92$ and 45% of their total linear abundance in the atlas restricted to a single-cell population.

well as GRO/precision nuclear run-on sequencing (PRO)-seq data (Methods) from similar cell populations to further characterize dCREs (Extended Data Fig. 3c,g), focusing on those whose

accessibility positively associated with miRNA expression (putative enhancers). We found that 278/279 dCREs were marked by H3K4me1, 273/279 by H3K27ac and 182/279 were transcribed



in at least one immune cell type. We observed a strong correlation between accessibility and H3K27ac levels at putative enhancers (Pearson $P=0.63$; Extended Data Fig. 3h), a high proportion of activity signals (H3K27ac and nascent transcripts) at accessible dCREs when the linked miRNA was at high abundance (Extended Data Fig. 3i) and, for 174/331 putative enhancer–miRNA associations, there was a significantly positive correlation (adjusted $P<0.1$) between dCRE H3K27ac levels and expression of the associated miRNA. Taken together, accessibility correlation, histone mark signatures and nascent RNA support the notion that positively associated dCREs are probable enhancers and that activity at these dCREs helps control expression of the associated miRNA.

To further corroborate enhancer activity, we selected 12 dCREs at the miR-142 and miR-21a loci for functional analysis (Extended Data Fig. 4a,b). Each of the elements was cloned into a luciferase reporter and transfected into RAW 264.7 macrophages, where miR-142 and miR-21a are both highly expressed. Eight of 12 dCRE resulted in increased luciferase activity compared to controls, indicative of enhancer activity (Fig. 3b). Although several dCREs did not increase luciferase levels, this does not exclude enhancer function because enhancer activity assays are limited by the lack of native context for the element. Thus, functional analysis of selected dCREs supports the likelihood that these elements serve as enhancers.

Our integrated miRNA atlas provided a means to understand *cis*-regulatory regions influencing miRNA abundance. For example, ten OCRs correlated with miR-223-3p abundance along with the proximal promoter previously described¹⁴ (Extended Data Fig. 4c). The promoter was accessible in myeloid cells, where miR-223 is highly expressed, and also in B cells, where miR-223 expression is modest. However, B cells lacked accessibility at an element ~1.5 kb upstream of the proximal promoter, which was universally open in myeloid cells. Moreover microglia, which express lower levels of miR-223 than other tissue macrophages (Fig. 1c), lacked accessibility at most associated dCREs in contrast to other myeloid cells. This was also reflected by lower H3K4me1 and H3K27ac levels at the miR-223-associated dCRE in microglia (Extended Data Fig. 4d). These features suggest a basal level of transcription driven by the miR-223 core promoter, modulated by the use of dCRE with varying degrees of specificity in the myeloid lineage.

miR-146a, a negative regulator of nuclear factor- κ B signaling¹⁶, also exhibited differential regulatory element usage. It was generally more highly expressed in lymphoid cells at steady state (Fig. 1c), and our analysis found a putative enhancer region supported by histone mark and nascent RNA signatures in T cells (Fig. 3c and Extended Data Fig. 4e). To validate the putative enhancer we used CRISPR to excise the predicted element, as well as the annotated miR-146a promoter. We generated ribonuclear protein complexes (RNPs) of Cas9 with single-guide RNAs targeting flanking regions of the elements and electroporated them into primary mouse T cells. This resulted in decreased expression of miR-146a in cells where either the promoter or putative enhancer was targeted compared to cells where an unrelated gene (*B2m*) was targeted (Fig. 3d). These experiments support the role of the identified dCRE in enhancing expression of miR-146a in T cells.

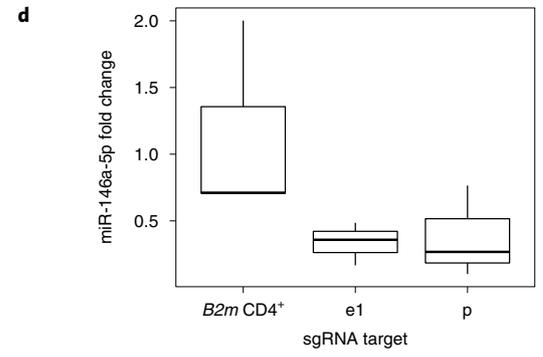
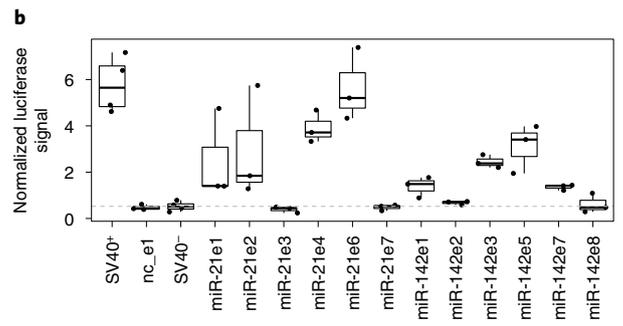
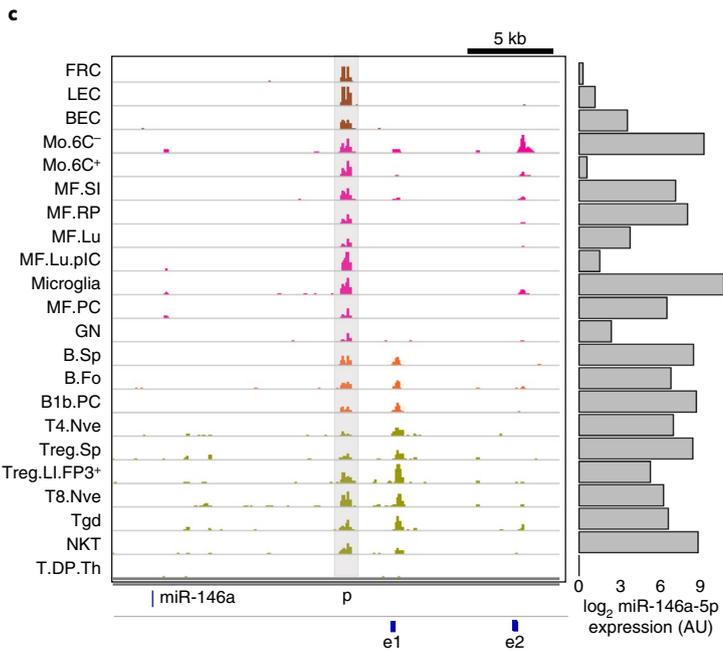
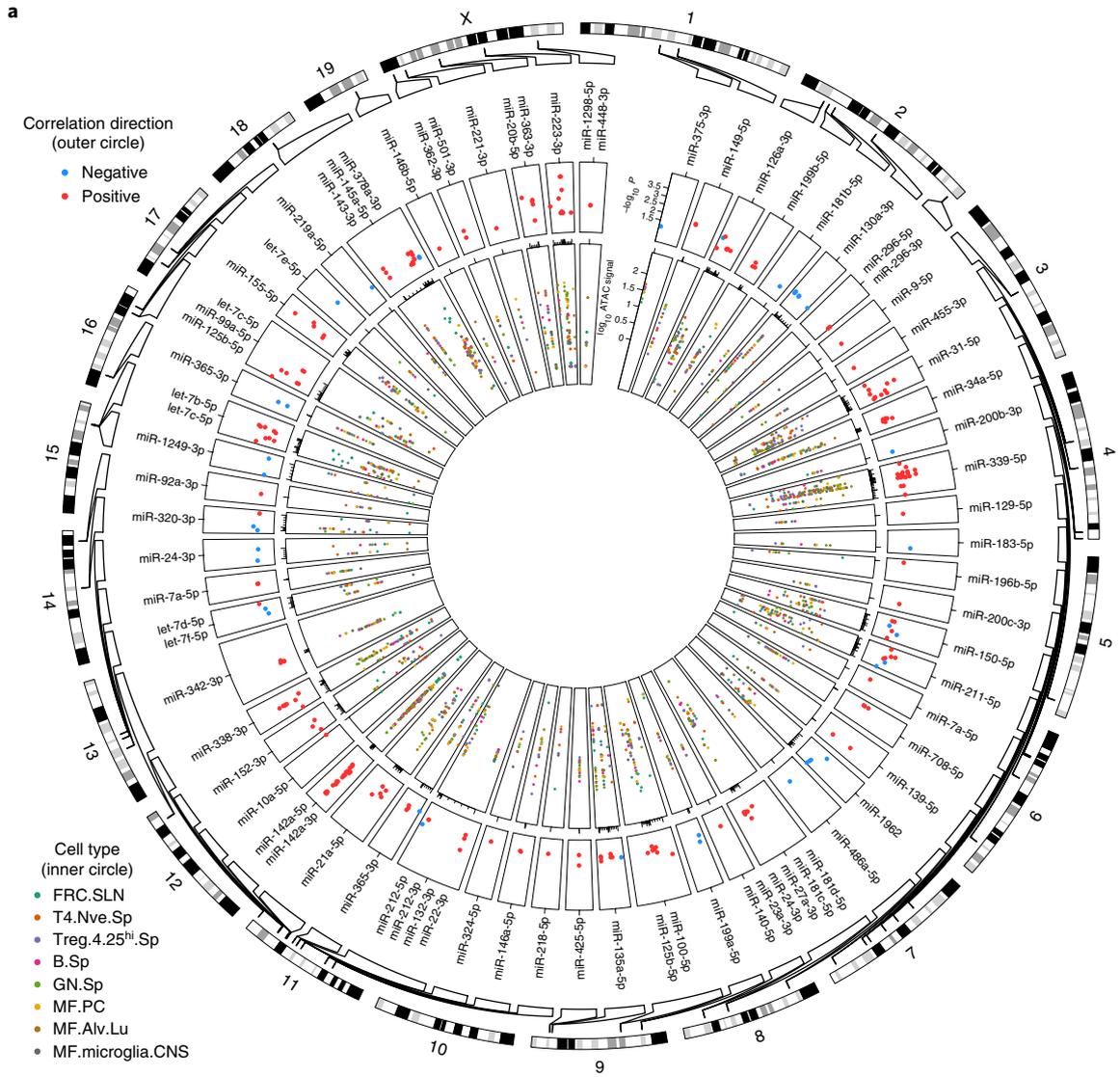
Multi-promoter usage associated with increased miRNA abundance. The miRNA promoter/enhancer landscape indicated that most expressed miRNAs had a single annotated promoter region, but that 28% (53/191) of conserved miRNAs had multiple annotated promoter regions at a single locus (Fig. 4a). Some miRNAs had up to five promoter regions, or as many as eight when accounting for duplicated miRNA loci (Extended Data Fig. 5a), and miRNAs with more than one promoter tended to be more conserved overall. Of miRNAs with multiple promoters, 35 were within protein-coding genes, long intergenic noncoding RNAs or processed transcripts while 18 were intergenic. Though it is known that many miRNAs have multiple annotated promoters^{20,36,37}, we are not aware of a global analysis examining how chromatin accessibility at these promoters relates to miRNA abundance. To assess this, we looked at miRNA expression in relation to the number of accessible annotated promoter regions across the genome in each cell type. Strikingly, there was a strong trend of increasing average miRNA abundance with increasing number of open promoter regions (Fig. 4b; $P<2\times 10^{-16}$, adjusted $R^2=0.21$). At many individual loci we also found similar associations, with 25/49 multipromoter miRNAs tested showing significant ($P<0.1$) relationships between open promoter region count and miRNA expression. This included miRNAs with multiple promoters at the same locus (miR-99a), duplicated miRNAs with individual promoters (miR-19b) and duplicated miRNAs with multiple promoters at each locus (miR-29b) (Fig. 4c–e).

With further restriction of our analysis to populations where we had H3K4me3 ChIP-seq and nascent RNA detection of promoter activity, in addition to ATAC-seq, the same trend of additive promoter effects on miRNA expression was observed (Extended Data Fig. 5b,c), and it also confirmed multiple active promoter regions at specific loci, such as the miR-29a/b-1 cluster (Extended Data Fig. 5d). These stringent promoter definitions provide a conservative estimate that, in a single cell type, a fraction of miRNAs use more than one promoter to control expression levels.

While multiple promoter regions may contribute to the overall abundance of a miRNA, they probably do so with different efficiencies. To study this, we built a stepwise regression model for each expressed miRNA with multiple promoters—focusing on nonduplicated miRNAs to simplify analyses to dynamics at individual loci—relating promoter accessibility to miRNA abundance. In 19/27 miRNAs tested there was at least one promoter that explained a significant ($P<0.1$) amount of variability in miRNA abundance. In 15/17 miRNAs with positive expression–accessibility relationships, proximal promoters accounted for more of the variation in miRNA abundance levels compared with the most distally annotated promoters (Extended Data Fig. 6). Therefore, at miRNA loci with multiple active promoters, proximal promoters typically appeared to have a greater influence in modulation of miRNA abundance.

One example was miR-21a, which has two promoters: one associated with its host gene, *Vmp1* and another in an intron of *Vmp1*, which is more proximal to miR-21a^{37,38}. Shared and divergent regulatory patterns have been observed previously at these two promoters in human cell lines³⁸. Our regression model found a stronger connection between miR-21a expression and accessibility at the

Fig. 3 | miRNA distal element discovery by OCR and expression associations. **a**, OCR accessibility to miRNA expression associations across 22 immune and stromal populations: 395 significant associations were found (adjusted $P<0.1$). Only associations for conserved miRNAs were plotted. Outer circle represents the $-\log_{10}$ adjusted P value of the Pearson correlation between an OCR and miRNA. Inner circle represents \log_{10} OCR activities in the indicated samples. **b**, Normalized luciferase signal in reporter-construct-transfected RAW 264.7 macrophages across *cis*-elements selected through associations in **a**. SV40⁺ represents the positive control and nc_e1 is a negative control enhancer region; SV40⁻ measures basal promoter (p) activity with no regulatory element added. The dashed line indicates mean SV40⁻ signal. $n=3$ –4 transfections. **c**, ATAC-seq signal at the miR-146a locus across 22 immune cell populations with both ATAC-seq and miRNA measurements. Lane height is standardized across all. e1 corresponds to enhancers targeted by flanking sgRNAs in **d**. **d**, miR-146a-5p expression measured by qPCR in activated CD4⁺ T cells following sgRNA–Cas9 RNP electroporation targeting enhancer 1 (e1) and the promoter of miR-146a or *B2m* as control. miR-146a-5p expression was measured at day 5 after treatment, and values were normalized to 5S rRNA. Shown are fold-change values relative to mean expression in control cells (*B2m* sgRNAs). $n=3$ mice, two independent experiments.



intronic promoter than the distal host gene promoter (Fig. 4f and Extended Data Fig. 6a,b), consistent with previous characterizations in human cells^{37,38}. In addition, putative enhancer elements were also closer to the intronic promoter. To assess the contribution of either promoter region to miR-21a abundance, we electroporated activated CD4⁺ T cells with CRISPR–Cas9 RNPs targeting flanking regions of either the proximal or distal promoter. We observed a decrease in miR-21a abundance under conditions targeting the proximal promoter, whereas editing of the distal (host gene) promoter had little effect (Fig. 4g). Interestingly, editing the proximal promoter decreased T cell numbers (Fig. 4h), consistent with the role of miR-21a in suppression of apoptosis during T cell expansion³⁹. Although we did not observe a decrease in miR-21a when the host gene promoter was targeted, we cannot exclude the possibility that this was due to either poor CRISPR efficiency, negative selection of edited cells or that this promoter has a greater role in other cell types. Our results suggest that variation in miR-21a expression across immune cells is primarily driven by the proximal intronic promoter rather than the host gene promoter.

Duplicated miRNA loci show consistent and divergent activity. Duplication of miRNA genes leads to identical mature miRNAs arising from distinct loci¹. This creates challenges in understanding miRNA regulation, because contributions from each locus cannot be determined from miRNA expression alone. We sought to resolve production sites for duplicated miRNAs using our promoter activity map. This revealed different origins of production for the same miRNA in different cell types. For example, miR-199a/b-3p is expressed predominantly in PC macrophages and fibroblasts and has three potential loci for production. OCR analysis suggested that PC macrophages have only one active locus (miR-199b) whereas fibroblasts show activity at all three loci in the production of miR-199a/b-3p (Fig. 5a,b), an observation confirmed by H3K4me3 and nascent RNA analysis (Extended Data Fig. 7a).

These findings prompted us to examine promoter behavior between all duplicated miRNA loci. We calculated the Manhattan distance of binarized open/closed values for each promoter at one locus compared to all promoter regions at the other locus across all ATAC-seq samples in the *cis*-regulatory atlas. This revealed instances of consistent and divergent activity (Fig. 5a,c and Extended Data Fig. 7b). miR-16 and miR-30c, two duplicated miRNAs that were highly and ubiquitously expressed, had both of their loci open across all cells. In contrast, miR-128 had one locus constitutively open (miR-128-1) while the other (miR-128-2) had promoters whose activity was largely restricted to developing lymphocytes and progenitors (Fig. 5a,d and Extended Data Fig. 8a). This suggests that the miR-128-2 locus is responsible for generation of the high expression levels observed during lymphocyte development (Extended Data Fig. 8b), which is supported by nascent RNA signatures in proB cells (Extended Data Fig. 8c). Let-7a, miR-19b/miR-92a (of clusters

miR-17~92 and miR-106a~363), miR-218 and miR-26a had a similar dichotomy at their two loci, highlighting the situational nature of how many duplicated loci are used. Notably, several of these had significant associations between the number of promoter sites used and average miRNA expression, implying that parallel control of both loci is important for setting of expression levels.

Differential promoter usage between miR-125b paralogs. miR-125b is a critical regulator of hematopoietic stem cell maintenance and differentiation and macrophage activation, and has been implicated in various hematopoietic malignancies⁴⁰. miR-125b is situated across two paralogous miRNA clusters: miR-100/let-7a-2/miR-125b-1 on chromosome 9 and miR-99a/let-7c-1/miR-125b-2 on chromosome 16. There were three annotated promoter regions associated with miR-125b-1 and four with miR-125b-2. The two loci had similar promoter structures, with several promoters able to produce polycistronic host transcripts for all three cluster members⁴¹, along with a separate promoter corresponding to transcripts specific for miR-125b but not the other cluster members. Promoters for the polycistronic transcripts that were active at either locus behaved similarly, open early in hematopoietic development and extinguished following differentiation into mature cell types (Fig. 5a,e,f). Interestingly, miR-125b-specific promoters did not behave in the same way (Extended Data Fig. 8d): the miR-125b-1-specific promoter was open only in fibroblasts while the miR-125b-2-specific promoter had an accessibility pattern similar to the developmentally regulated distal promoters. Activity at this promoter was supported by nascent RNA in thioglycolate-elicited macrophages, suggesting an apparent use in myeloid cells under certain conditions (Extended Data Fig. 8e). The juxtaposition of miR-125b loci highlights lineage- and developmental-stage-specific use of certain promoters, and suggests that miR-125b-2 utilizes a promoter for independent control from its other cluster members across hematopoiesis and certain tissue macrophages.

Discussion

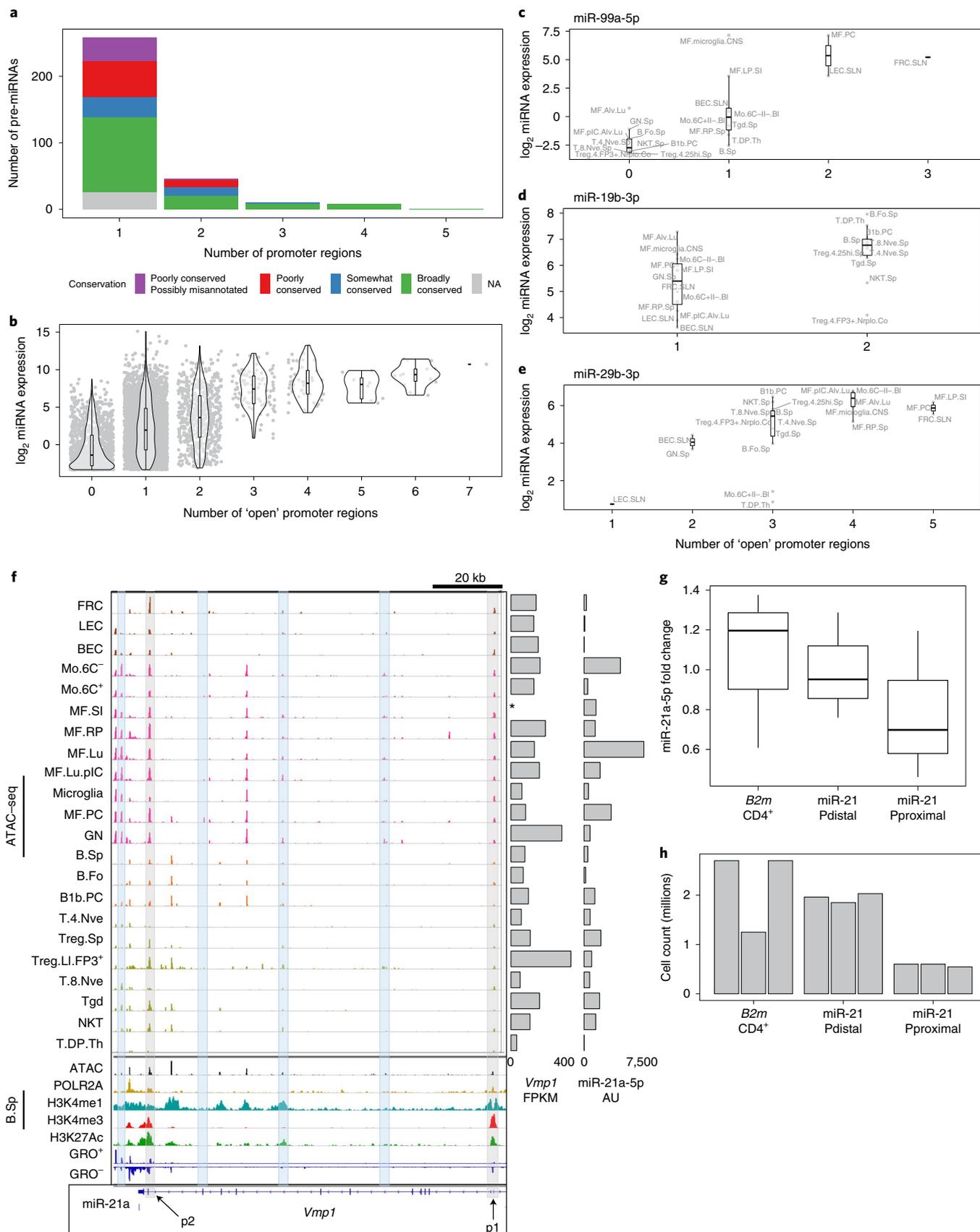
Here we present an atlas of miRNA epigenetic and expression profiles across the mouse immune system. By pairing OCR and expression measurements, we established a wide-ranging map of putative CREs controlling miRNA expression within different immune populations and uncovered themes of miRNA promoter and enhancer use. This sheds light on miRNA *cis*-regulatory control in immune cells and provides a map of miRNA promoter and enhancer element usage across the immune system.

miRNA activity is highly dependent on concentration^{27,42,43}, and thus it is important that miRNA signatures incorporate abundance thresholds. We used several metrics, including an in vivo miRNA sensor assay, to empirically derive a threshold of likely activity. There were ~50 miRNAs surpassing this cutoff in each cell type while ~185 miRNAs were consistently detected above background. Meanwhile, a

Fig. 4 | Additive effects of multiple miRNA promoters. **a**, Number of promoter regions at individual loci for expressed miRNAs, colored by TargetScan conservation categories. **b–e**, miRNA expression as compared to its number of open promoter regions. Open is defined as Benjamini–Hochberg-adjusted MACS2 $P < 0.05$; annotated promoter OCRs within 2 kb of each other were merged and summarized to the minimum MACS2 P value and maximum ATAC signal. **b**, Global comparison across all 22 cell types with paired miRNA and ATAC-seq measurements in 236 detected and confidently annotated miRNAs with promoter annotations. **c–e**, Individual miRNA associations for miR-99a-5p (**c**), miR-19b-3p (**d**) and miR-29b-3p (**e**) ($n = 22$ populations). **f**, Chromatin accessibility at the miR-21a locus in all overlapping populations between miRNA and ATAC-seq profiling, with host gene (*Vmp1*) and miRNA expression at the side. Asterisk indicates missing RNA-seq data. Shaded gray bars indicate promoter sites labeled as p1 for the host gene promoter and p2 for the intronic promoter; shaded blue bars indicate OCRs correlated with miR-21a expression. Overlaid at the bottom are POLR2A, H3K4me1, H3K4me3 and H3K27Ac ChIP-seq, and GRO-seq plots from public data on splenic B cells to highlight promoter and enhancer structures. Read density track heights are standardized across the 22 ATAC-seq samples in the top portion and normalized within each lane in the bottom portion. **g**, miR-21a-5p expression measured by qPCR on day 4 after sgRNA–Cas9 RNP electroporation and activation. Effects of miR-21a promoter deletion were measured in CD4⁺ T cells; fold changes are relative to mean B2m CD4⁺ expression ($n = 3$ mice, three experiments). **h**, Total T cell counts at day 4 after electroporation/activation under conditions targeting miRNA *cis*-regulatory elements or an off-target control (*B2m*) ($n = 3$ mice, three experiments). NA, not applicable. FPKM, fragments per kilobase exon per million mapped reads.

majority of overall abundance was generally explained by five or fewer very highly expressed miRNAs. In populations overlapping with previously studied cell types, our signatures are consistent⁵⁻¹³

but provide a more defined thresholding relevant to interpretation of gene signatures. To aid in prioritization of miRNAs regulating a given target, we also developed an analysis tool incorporating



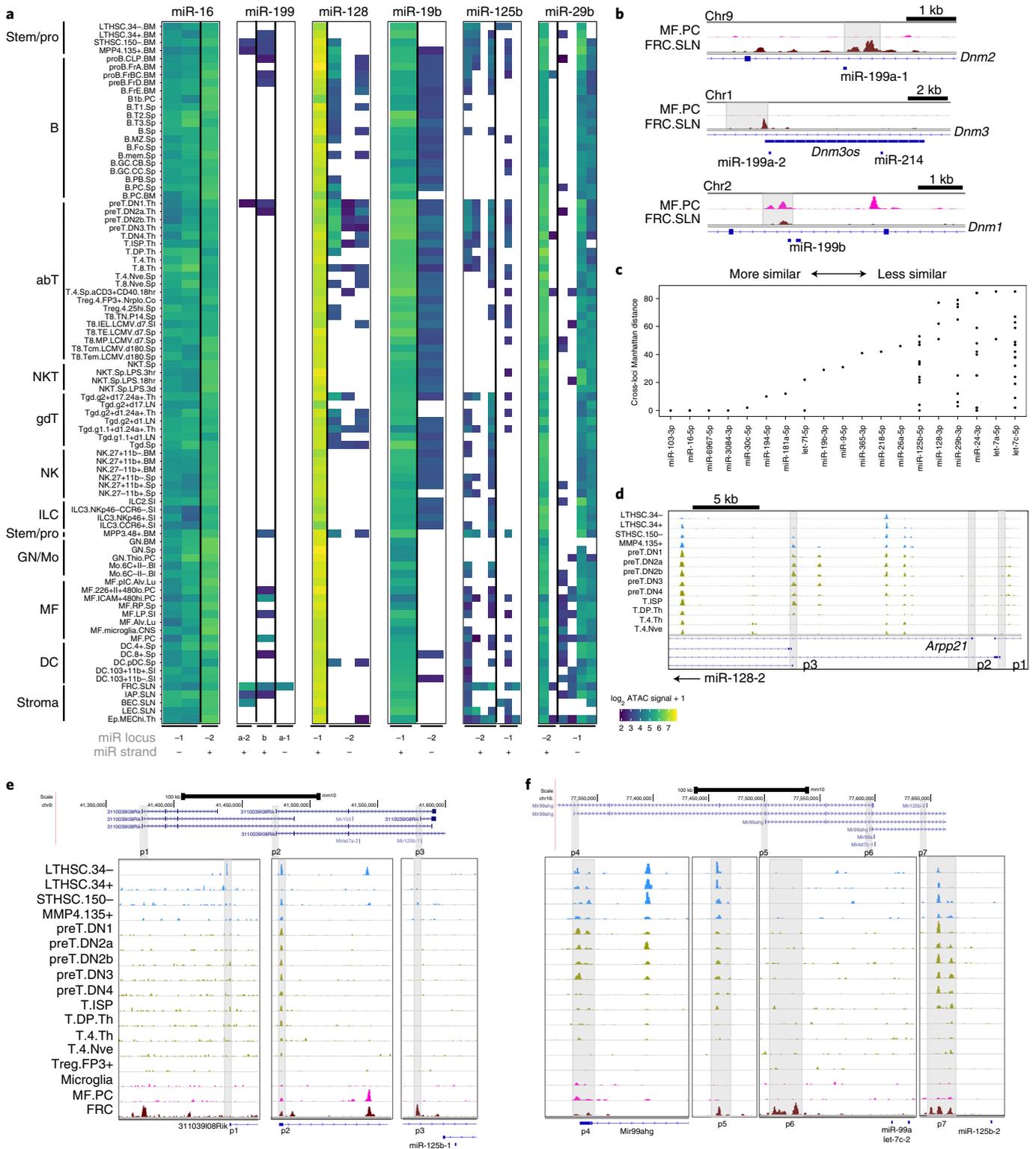


Fig. 5 | Chromatin accessibility patterns at multi-promoter and duplicated miRNAs. **a**, Promoter region accessibility patterns for six select duplicated miRNAs. Colorless boxes represent OCRs not detected above background. For each locus, if the miRNA is on the positive strand, the promoters are ordered from furthest to nearest from left to right, with the opposite true for miRNAs on the negative strand. abT, T cells with an $\alpha\beta$ TCR; ILC, innate lymphoid cell; GN/Mo, neutrophils and monocytes; MF, macrophages. **b**, IGV plot of miR-199a/b loci in PC macrophages and FRCs. Promoter regions highlighted in gray. **c**, Similarity of promoter region activity across duplicated miRNA loci. Manhattan distance shown for each promoter region at one locus as compared to all promoter regions at the other locus across all 87 *cis*-atlas samples. **d**, *Cis*-element accessibility at miR-128-2 locus over T cell development. Promoter regions are indicated by shaded gray bars. **e, f**, Promoter accessibility at miR-125b loci over T cell development and in microglia, PC macrophages and fibroblasts. RefSeq transcript tracks for the miR-125b-1 locus (left) and miR-125b-2 locus (right) are shown at the top, with zoomed-in views of ATAC-seq read density at annotated miRNA promoter locations in selected immune cell types from T cell development, microglia, PC macrophages and FRC stromal cells below. Promoters highlighted as in **d**.

our miRNA expression database with miRNA target prediction (Transcript-View: <http://shiny.immgen.org/mirna-browser/>). This enhances existing target prediction resources by incorporation of abundance information to determine probable functional miRNA to target messenger RNA relationships in different cell types.

How proper miRNA concentration is achieved is a central question of miRNA biology. Many mechanisms have been described as to how this process is regulated, including post-transcriptional regulation of miRNA processing and miRNA degradation by complementary target RNAs¹. Transcription is another major mechanism of miRNA control, but surprisingly little is known about the genomic elements orchestrating this process for many miRNAs. Through analysis of paired OCR accessibility and miRNA expression measurements, we were able to map the miRNA promoter and dCRE activity landscape of immune cells. While chromatin accessibility is not always a perfect measure of activity at promoters or dCRE, it is a useful proxy⁴⁴. Importantly, we used additional measures, such as H3K4me3 and H3K27ac histone marks and nascent RNA, to support *cis*-regulatory activity found by ATAC-seq. However, some associated OCRs may not be functioning as promoters or enhancers but could be structural elements, or, in the case of negatively associated elements, potential repressors. Association of some elements may also result from confounding variation as seen at a certain rate in other studies⁴⁵. However, the many associations between chromatin accessibility, histone mark enrichment and miRNA abundance add confidence to our miRNA *cis*-regulatory map.

Our analysis rediscovered known biology at the miR-223 locus¹⁵ while predicting additional regulatory elements that could provide more granular control over miR-223 expression within the myeloid lineage. Interestingly, other highly expressed and/or lineage-specific miRNAs, such as miR-21a and miR-142, had many dCREs associated with their expression as well, with five and 12 associated dCREs, respectively, resembling super-enhancer control regions for these miRNAs. This may indicate a need for coordination through many elements to achieve high expression, redundancy of enhancers to ensure proper expression or an ability to precisely titrate miRNA levels through combinatorial use of dCREs. The plausibility of these mechanisms is supported by our reporter assay, showing that selected DNA sequences from these regions can independently promote transcription. In addition, our analysis identified putative regulatory regions for many immune-relevant miRNAs, including miR-146a, which was validated by CRISPR-Cas9 editing in primary mouse T cells.

Our study indicates that cells have multiple miRNA promoters concurrently active at certain miRNA loci. Other studies have reported the presence of several promoters for a given miRNA^{20,36,37}, but these studies generally lacked parallel promoter and expression measurements across a range of cells to assess the relationship between the two. Proximal promoters and sometimes certain loci (for example, miR-128-2) appeared to have predominant effects on miRNA expression levels, extending to functionally dominant regulatory effects as observed on T cell expansion after editing of the miR-21a proximal promoter. The use of proximal promoters to upregulate miRNA expression may be an orientational consequence of evolution, supporting a model proposed by Ozsolak et al. in which miRNAs originating from longer host transcripts evolved closer promoters to increase miRNA production efficiency³⁷. In addition to reduction of transcriptional distance, these promoters are probably associated with regulatory elements and TADs more suited to miRNA production than their host transcript, especially for miRNAs whose distal promoters are also for protein-coding genes. Characterization of these predominant promoters and their associated elements, along with cognate transcription factors, will provide insight into the mechanisms controlling production of these miRNAs.

While activity at some individual promoters best accounted for variance in miRNA levels across immune cells, we also observed

that a higher number of open promoter regions was associated with higher miRNA expression. This suggests a regulatory logic in which aggregation of pri-miRNA from several transcriptional initiation regions increases miRNA concentration, a phenomenon we found generally supported by nascent-RNA-inferred transcription units in relevant immune cell types. We and others have suggested that high miRNA abundance is probably achieved by the stability of miRNAs, which can have a half-life of >48 h (refs. ^{46,47}). This assumption partly derived from the fact that abundance of a miRNA could be much higher than its host gene, sometimes by more than an order of magnitude. The finding that some highly expressed miRNAs probably have multiple active promoters, and proximal promoters additional to the host gene promoter, supports a mechanism in which parallel transcription also enables miRNAs to reach high cellular concentrations. Alternatively, additional accessible promoters may switch to serve as enhancers for the same miRNA.

Our analysis revealed cell-type-specific usage of many promoter elements, including many duplicated loci with divergent promoter activity. In addition to driving miRNA production with varying efficiency, the use of alternative promoters allows for a different set of parameters to govern miRNA expression under different conditions. It has been demonstrated that various miRNA TSS sites are associated with different enhancer elements that can form alternative responses to stimuli, such as tumor necrosis factor- α (ref. ³⁶). Additionally, the use of different TSS sites can also transcribe different host genes and pri-miRNAs which, in some cases, may impose additional regulation during processing through pri-miRNA tertiary structures⁴⁸. Differential promoter usage could serve to balance associated miRNAs with differing functions. For instance, studies have proposed that miR-99/100 and let-7 (miR-125b cluster members) may oppose the pro-proliferative role of miR-125b in HSCs by negative regulation of transcripts promoting proliferation⁴¹. The promoter separating miR-125b from polycistronic control could enable a shift in the balance between these families of miRNA to change the proliferative status of the cell under the right conditions.

The studies described here provide a deeper understanding of the miRNome of immune cells and provide a resource for studying the regulatory landscape of this important layer of post-transcriptional control of immune cell biology.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at <https://doi.org/10.1038/s41590-021-00944-y>.

Received: 8 January 2020; Accepted: 26 April 2021;

Published online: 7 June 2021

References

1. Bartel, D. P. Metazoan microRNAs. *Cell* **173**, 20–51 (2018).
2. Mehta, A. & Baltimore, D. MicroRNAs as regulatory elements in immune system logic. *Nat. Rev. Immunol.* **16**, 279–294 (2016).
3. O'Connell, R. M., Rao, D. S., Chaudhuri, A. A. & Baltimore, D. Physiological and pathological roles for microRNAs in the immune system. *Nat. Rev. Immunol.* **10**, 111–122 (2010).
4. Montagner, S., Dehó, L. & Monticelli, S. MicroRNAs in hematopoietic development. *BMC Immunol.* **15**, 14 (2014).
5. Kuchen, S. et al. Regulation of microRNA expression and abundance during lymphopoiesis. *Immunity* **32**, 828–839 (2010).
6. Mildner, A. et al. Mononuclear phagocyte miRNome analysis identifies miR-142 as critical regulator of murine dendritic cell homeostasis. *Blood* **121**, 1016–1027 (2013).
7. Landgraf, P. et al. A mammalian microRNA expression atlas based on small RNA library sequencing. *Cell* **129**, 1401–1414 (2007).
8. Monticelli, S. et al. MicroRNA profiling of the murine hematopoietic system. *Genome Biol.* **6**, R71 (2005).

9. Basso, K. et al. Identification of the human mature B cell miRNome. *Immunity* **30**, 744–752 (2009).
10. Wu, H. et al. miRNA profiling of naïve, effector and memory CD8 T cells. *PLoS ONE* **2**, e1020 (2007).
11. Butovsky, O. et al. Identification of a unique TGF- β -dependent molecular and functional signature in microglia. *Nat. Neurosci.* **17**, 131–143 (2014).
12. Agudo, J. et al. The miR-126-VEGFR2 axis controls the innate response to pathogen-associated nucleic acids. *Nat. Immunol.* **15**, 54–62 (2013).
13. Fehniger, T. A. et al. Next-generation sequencing identifies the natural killer cell microRNA transcriptome. *Genome Res.* **20**, 1590–1604 (2010).
14. Fukao, T. An evolutionarily conserved mechanism for microRNA-223 expression revealed by microRNA gene profiling. *Cell* **129**, 617–631 (2007).
15. Fazi, F. et al. A minicircuitry comprised of microRNA-223 and transcription factors NFI-A and C/EBP α regulates human granulopoiesis. *Cell* **123**, 819–831 (2005).
16. Taganov, K. D., Boldin, M. P., Chang, K. J. & Baltimore, D. NF- κ B-dependent induction of microRNA miR-146, an inhibitor targeted to signaling proteins of innate immune responses. *Proc. Natl Acad. Sci. USA* **103**, 12481–12486 (2006).
17. Ye, Z. et al. Regulation of miR-181a expression in T cell aging. *Nat. Commun.* **9**, 3060 (2018).
18. Kirigin, F. F. et al. Dynamic microRNA gene transcription and processing during T cell development. *J. Immunol.* **188**, 3257–3267 (2012).
19. Georgakilas, G. et al. microTSS: accurate microRNA transcription start site identification reveals a significant number of divergent pri-miRNAs. *Nat. Commun.* **5**, 5700 (2014).
20. Chang, T. C., Pertea, M., Lee, S., Salzberg, S. L. & Mendell, J. T. Genome-wide annotation of microRNA primary transcript structures reveals novel regulatory mechanisms. *Genome Res.* **25**, 1401–1409 (2015).
21. Marson, A. Connecting microRNA genes to the core transcriptional regulatory circuitry of embryonic stem cells. *Cell* **134**, 521–533 (2008).
22. de Rie, D. An integrated expression atlas of miRNAs and their promoters in human and mouse. *Nat. Biotechnol.* **35**, 872–878 (2017).
23. Suzuki, H. I., Young, R. A. & Sharp, P. A. Super-enhancer-mediated RNA processing revealed by integrative microRNA network analysis. *Cell* **168**, 1000–1014 (2017).
24. Mestdagh, P. et al. Evaluation of quantitative miRNA expression platforms in the microRNA quality control (mirQC) study. *Nat. Methods* **11**, 809–815 (2014).
25. Jayaprakash, A. D., Jabado, O., Brown, B. D. & Sachidanandam, R. Identification and remediation of biases in the activity of RNA ligases in small-RNA deep sequencing. *Nucleic Acids Res.* **39**, e141 (2011).
26. Giraldez, M. D. et al. Comprehensive multi-center assessment of small RNA-seq methods for quantitative miRNA profiling. *Nat. Biotechnol.* **36**, 746–757 (2018).
27. Brown, B. D. et al. Endogenous microRNA can be broadly exploited to regulate transgene expression according to tissue, lineage and differentiation state. *Nat. Biotechnol.* **25**, 1457–1467 (2007).
28. Cho, S. et al. miR-23 approximately 27 approximately 24 clusters control effector T cell differentiation and function. *J. Exp. Med.* **213**, 235–249 (2016).
29. Trifari, S. et al. MicroRNA-directed program of cytotoxic CD8⁺ T-cell differentiation. *Proc. Natl Acad. Sci. USA* **110**, 18608–18613 (2013).
30. O'Connell, R. M., Rao, D. S. & Baltimore, D. microRNA regulation of inflammatory responses. *Annu. Rev. Immunol.* **30**, 295–312 (2012).
31. Rodríguez-Galán, A., Fernández-Messina, L. & Sánchez-Madrid, F. Control of immunoregulatory molecules by miRNAs in T cell activation. *Front. Immunol.* **9**, 2148 (2018).
32. He, M. et al. Cell-type-based analysis of microRNA profiles in the mouse brain. *Neuron* **73**, 35–48 (2012).
33. Yoshida, H. et al. The cis-regulatory atlas of the mouse immune system. *Cell* **176**, 897–912 (2019).
34. Harrow, J. et al. GENCODE: the reference human genome annotation for the ENCODE project. *Genome Res.* **22**, 1760–1774 (2012).
35. Johanson, T. M. et al. Transcription-factor-mediated supervision of global genome architecture maintains B cell identity. *Nat. Immunol.* **19**, 1257–1264 (2018).
36. Bouvy-Liivrand, M. et al. Analysis of primary microRNA loci from nascent transcriptomes reveals regulatory domains governed by chromatin architecture. *Nucleic Acids Res.* **45**, 12054 (2017).
37. Ozsolak, F. et al. Chromatin structure analyses identify miRNA promoters. *Genes Dev.* **22**, 3172–3183 (2008).
38. Ribas, J. et al. A novel source for miR-21 expression through the alternative polyadenylation of VMP1 gene transcripts. *Nucleic Acids Res.* **40**, 6821–6833 (2012).
39. Ruan, Q. et al. MicroRNA-21 regulates T-cell apoptosis by directly targeting the tumor suppressor gene *Tipe2*. *Cell Death Dis.* **5**, e1095 (2014).
40. O'Connell, R. M. et al. MicroRNAs enriched in hematopoietic stem cells differentially regulate long-term hematopoietic output. *Proc. Natl Acad. Sci. USA* **107**, 14235–14240 (2010).
41. Emmrich, S. et al. miR-99a/100–125b tricistrons regulate hematopoietic stem and progenitor cell homeostasis by shifting the balance between TGF β and Wnt signaling. *Genes Dev.* **28**, 858–874 (2014).
42. Mullokandov, G. et al. High-throughput assessment of microRNA activity and function using microRNA sensor and decoy libraries. *Nat. Methods* **9**, 840–846 (2012).
43. Bosson, A. D., Zamudio, J. R. & Sharp, P. A. Endogenous miRNA and target concentrations determine susceptibility to potential ceRNA competition. *Mol. Cell* **56**, 347–359 (2014).
44. Buenrostro, J. D., Giresi, P. G., Zaba, L. C., Chang, H. Y. & Greenleaf, W. J. Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nat. Methods* **10**, 1213–1218 (2013).
45. Gasperini, M., Tome, J. M. & Shendure, J. Towards a comprehensive catalogue of validated and target-linked human enhancers. *Nat. Rev. Genet.* **21**, 292–310 (2020).
46. Baccarini, A. et al. Kinetic analysis reveals the fate of a microRNA following target regulation in mammalian cells. *Curr. Biol.* **21**, 369–376 (2011).
47. Kingston, E. R. & Bartel, D. P. Global analyses of the dynamics of mammalian microRNA metabolism. *Genome Res.* **29**, 1777–1790 (2019).
48. Treiber, T., Treiber, N. & Meister, G. Regulation of microRNA biogenesis and its crosstalk with other cellular pathways. *Nat. Rev. Mol. Cell Biol.* **20**, 5–20 (2019).

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

© The Author(s), under exclusive licence to Springer Nature America, Inc. 2021

the Immunological Genome Consortium

Oscar Aguilar¹², Rhys Allan¹³, Janice Arakawa-Hoyt¹², Jilian Astarita¹⁴, K. Frank Austen¹⁵, Nora Barrett¹⁵, Alev Baysoy¹⁶, Christophe Benoist¹⁶, Brian D. Brown^{1,2}, Matthew Buechler¹⁴, Jason Buenrostro¹⁷, Maria Acebes Casanova¹, Kyunghye Choi¹⁸, Kaitavjeet Chowdhary¹⁶, Marco Colonna¹⁸, Ty Crowl⁸, Tianda Deng⁸, Jigar V. Desai¹⁹, Fiona Desland¹, Maxime Dhainaut¹, Jiarui Ding²⁰, Claudia Dominguez¹⁴, Daniel Dwyer¹⁵, Michela Frascoli²¹, Shani Gal-Oz²², Ananda Goldrath⁸, Ricardo Grieshaber-Bouyer¹⁵, Baosen Jia¹, Tim Johanson¹³, Stefan Jordan¹, Joonsoo Kang²¹, Varun Kapoor¹⁴, Ephraim Kenigsberg¹, Edy Y. Kim^{6,7}, Joel Kim¹, Ki wook Kim¹⁸, Evgeny Kiner¹⁶, Mitchell Kronenberg²³, Lewis Lanier¹², Catherine Laplace¹⁶, Caleb Lareau¹⁷, Yonit Lavin¹, Andrew Leader¹, Jisu Lee²¹, Assaf Magen¹, Barbara Maier¹, Alexandra Maslova²⁴, Diane Mathis¹⁶, Adelle McFarland¹⁸, Miriam Merad^{1,9}, Etienne Meunier²⁴, Paul Monach¹⁵, Sara Mostafavi²⁴, Soren Muller¹⁴, Christoph Muus²⁰, Hadas Ner-Gaon²², Quyhn Nguyen⁸,

Peter A. Nigrovic¹⁵, Kouta Niizuma¹², German Novakovsky²⁴, Stephen Nutt¹³, Kayla Omilusik⁸, Adriana Ortiz-Lopez¹⁶, Mallory Paynich²³, Vincent Peng¹⁸, Marc Potempa¹⁶, Rachana Pradhan¹⁴, Sara Quon⁸, Ricardo Ramirez¹⁶, Deepshika Ramanan¹⁶, Gwendalyn Randolph¹⁸, Aviv Regev²⁰, Andrew Rhoads⁵, Samuel A. Rose^{1,2}, Kumba Seddu¹⁶, Tal Shay²², Avishai Shemesh¹⁶, Justin Shyer¹⁴, Christopher Smilie²⁰, Nick Spidale²¹, Ayshwarya Subramanian²⁰, Katelyn Sylvia²¹, Julie Tellier¹³, Shannon Turley¹⁴, Brinda Vijaykumar¹⁶, Amy Wagers¹⁷, Chendi Wang²⁴, Peter L. Wang¹⁸, Aleksandra Wroblewska^{1,2}, Liang Yang¹⁶, Aldrin Yim¹⁸, Hideyuki Yoshida³ and Bingfei Yu⁸

¹²Department of Microbiology & Immunology, University of California, San Francisco, CA, USA. ¹³The Walter and Eliza Hall Institute of Medical Research, Parkville, Victoria, Australia. ¹⁴Department of Cancer Immunology, Genentech, South San Francisco, CA, USA. ¹⁵Division of Rheumatology, Immunology and Allergy, Brigham and Women's Hospital, Boston, MA, USA. ¹⁶Department of Immunology, Harvard Medical School, Boston, MA, USA. ¹⁷Department of Stem Cell and Regenerative Biology, Harvard University, Cambridge, MA, USA. ¹⁸Department of Pathology and Immunology, Washington University School of Medicine, St. Louis, MO, USA. ¹⁹National Institute of Allergy and Infectious Diseases, National Institutes of Health, Bethesda, MD, USA. ²⁰Broad Institute of Massachusetts Institute of Technology and Harvard, Cambridge, MA, USA. ²¹Department of Pathology, University of Massachusetts Medical School, Worcester, MA, USA. ²²Department of Life Sciences, Ben-Gurion University of the Negev, Be'er Sheva, Israel. ²³La Jolla Institute for Immunology, La Jolla, CA, USA. ²⁴Department of Statistics, University of British Columbia, Vancouver, British Columbia, Canada.

Methods

Mice. All cells used for miRNA analysis were obtained from male 6–8-week-old C57BL/6J mice from the Jackson Laboratory, used 6–8 days from reception. For the miRNA sensor assay, CD45.1 C57BL/6J (B6.SJL-*Ptprca*^u *Peprc*^u/BoyCr1) male mice were used as donors. Mice were housed in specific-pathogen-free facilities. The use of all mice for these studies was in accordance with institutional guidelines, with review and approval by the Institutional Animal Care and Use Committee at the Icahn School of Medicine at Mount Sinai.

Cell isolation, sorting and culture. Cells were purified according to the standardized ImmGen standard operations protocol using the lineage markers highlighted in Table 1. Peritoneal cell suspensions were obtained by lavage of the peritoneal cavity with 10 ml of cold PBS. Single-cell suspensions were obtained by mechanical disruption of tissues followed by erythrocyte lysis using ACK buffer (Lonza). Where necessary, tissues were digested at 37°C for 15 min in glucose-containing medium. Cells were sorted using a BD FACSAria IIu cell sorter at the Icahn School of Medicine at Mount Sinai (neutrophils and dendritic cells from the spleen; macrophages from the peritoneal cavity, lung, small and large intestine, spleen and brain; monocytes from the blood and bone marrow; naive and activated CD4⁺ T cells, naive CD8⁺ T cells, regulatory Tregs from the spleen; naive splenic NK and invariant natural killer T cells (iNKT cells); Brigham and Women's Human Immunology Flow Core (mast cells from the peritoneal cavity, naive and activated NKT cells from the spleen); Harvard Medical School (double-positive T cells from the thymus, Tregs from the colon); UCSD (naive and activated CD8⁺ T cells from the spleen); Fox Chase Cancer Center (B cells from the peritoneal cavity and the spleen); University of Massachusetts Medical School ($\gamma\delta$ T cells from the lymph nodes); UCSF (NK cells from the spleen and the liver); and Genentech (stromal cells). The majority of populations profiled were performed in duplicate to maximize breadth, with the exception of B1b.PC, NK.Sp and NKT.Sp ($n = 5$); MF.PC, GN.Sp, B.Sp and T4.Nve.Sp ($n = 4$); DC.Sp, Treg.4.25hi.Sp, T8.Nve.Sp and Tgd.Sp ($n = 3$); and MC.IgE.PC, Tgd.27+.LN, NK.49b-.MCMV.Lv and female populations, with one replicate, that were retained for completeness; however no analyses were done with single-replicate populations. RAW 264.7 macrophages were cultured in DMEM supplemented with 10% fetal bovine serum, penicillin and streptomycin.

RNA isolation and miRNA profiling. RNA was isolated using the miRNeasy Micro Kit (Qiagen), and miRNA was profiled using the miScript miRNA PCR array. Out of 940 miRNAs measured, 17 assays were removed from the analysis due to improper annotation in miRbase v.22. miRNAs were profiled on three separate qPCR plates and in four processing batches. A standard curve was generated using ten synthetic miRNAs: mmu-miR-126a-5p, mmu-miR-140-5p, mmu-miR-142a-3p, mmu-miR-147-3p, mmu-miR-150-5p, mmu-miR-155-5p, mmu-miR-199a-5p, mmu-miR-210-3p, mmu-miR-223-3p and mmu-miR-451a (Supplementary Table 1).

miRNA sensors. Bidirectional lentiviral vectors were used for miRNA sensors, as previously described¹. Bone marrow cells were isolated from CD45.1 C57BL/6J mice, and lineage-positive cells (CD5⁺, CD45R⁺ (B220), CD11b⁺, Gr-1⁺ (Ly6G/C), 7-4⁺ and Ter119⁺) were depleted using the Lineage Cell Depletion Kit (Miltenyi). Remaining cells were transduced overnight with miR-24-3p sensor, miR-652-3p sensor or a control vector in StemSpan SFEM expansion medium (StemCell Technologies) supplemented with 50 ng ml⁻¹ stem cell factor (SCF), 10 ng ml⁻¹ IL-3, 10 ng ml⁻¹ Flt3L, 20 ng ml⁻¹ IL-6 (PeproTech) and 5 μ g ml⁻¹ polybrene (Millipore). CD45.2 recipient mice were irradiated with two doses of 600 rad with a 4-h interval. Transduced lineage-negative cells were collected, washed, counted and transplanted into lethally irradiated recipient mice. After 8 weeks, mice were sacrificed and blood, spleen and cells from the peritoneal cavity were collected and analyzed for green fluorescent protein (GFP) and mutant nerve growth factor receptor (NGFR) expression. The following antibodies were used: PE-Cy7-conjugated anti-mouse CD19, APC-conjugated anti-TCR β , PerCP-Cy5.5-conjugated anti-CD4, AlexaFluor700-conjugated anti-Ly6G, eFluor 450-conjugated anti-CD45.1, APC-eFluor 780-conjugated anti-CD45.2 (Thermo Fisher Scientific), BV510-conjugated anti-CD8a (BioLegend) and PE-conjugated anti-human CD271 (NGFR) (BD Bioscience). Data were acquired on a BD FACS Fortessa and analyzed with FlowJo software (TreeStar). Percentage of suppression was calculated as described previously⁴⁹ (Supplementary Table 2).

Luciferase assay. We used a pGL3-control vector (Promega) with its SV40 enhancer deleted (pGL3-no SV40, a gift from the Wysocka Laboratory⁵⁰), for our luciferase reporter assay experiments. The pGL3-control vector with the SV40 enhancer intact was used as a positive control, and the pGL3-no SV40 with no insert or with a predicted inactive region (nc_e1) was used as negative control. Genomic sequence in putative enhancer regions at the miR-21a and miR-142 loci, either directly correlated with miRNA expression or linked by a shared H3K27ac domain, was PCR amplified from RAW 264.7 macrophage DNA and cloned into the pGL3-no SV40 vector with T4 DNA ligase (NEB) after NheI and XhoI digestion. For transfection, a procedure previously outlined by Cheung et al. was used⁵¹. Briefly, RAW 264.7 macrophages were seeded at 200,000 cells per 24-well plate and left for 4 h in the incubator at 37°C. After 4 h the cells were cotransfected

with 0.5 μ g of reporter construct along with Renilla luciferase vector pRL-SV40 (Promega) at a 1:20 dilution using Lipofectamine 3000 (ThermoFisher) and a 1:4 ratio of plasmid to transfection reagent. After 4 h transfection medium was removed and replaced with fresh DMEM. Luminescence was measured on a SpectraMax M5 (Molecular Devices) 24 h after removal of transfection medium using the Dual-Luciferase Reporter Assay System (Promega), with two technical replicates per biological replicate and at least three biological replicates per condition ($n = 3$ –4 transfections). Normalized luciferase signal was calculated by averaging technical replicates, subtracting background signal measured for each luciferase in untreated wells and then dividing firefly luciferase signal by the Renilla signal for the same sample. Primers used to clone cCREs and luciferase data can be found in Supplementary Table 3.

RNP electroporation. T cell RNP electroporation experiments were performed following an optimized protocol from Nüssing et al.⁵². RNPs were generated by incubation of 0.6 μ l of Cas9 protein at 10 μ g μ l⁻¹ (Alt-R S.p. Cas9 Nuclease V3, IDT) with 1 μ l of each sgRNA (ThermoFisher; Supplementary Table 4) at 0.3 nmol μ l⁻¹ and water up to 5 μ l at room temperature for 10 min. CD4⁺ or CD8⁺ T cells were isolated from splenocytes in single-cell suspension using negative selection by EasySep mouse T cell isolation kit (Stemcell Technologies). RNPs were electroporated into 2 million isolated T cells per condition in a Lonza 4D-Nucleofector X using the P4 primary cell nucleofector kit (Lonza) with the CMI37 setting. Following electroporation, T cells were activated for 3 days as described by Wroblewska et al.⁵³. Following activation, T cells were cultured for 24–48 h with IL-2, counted and subsequently resuspended in Qiazol (Qiagen) followed by RNA extraction using a miRNeasy Micro kit (Qiagen) with the small RNA enrichment protocol. Small RNA was reverse transcribed using the miRCURY LNA RT kit (Qiagen) with 20 ng of input, diluted 1:6 in molecular-grade water and then qPCR was performed with miRCURY LNA miRNA PCR assay (Qiagen) for either miR-146a-5p with a reference assay of 5S ribosomal RNA ($n = 3$ mice) or miR-21a-5p with a reference assay of miR-142-3p ($n = 3$ mice). Resulting Ct values were normalized using the delta-Ct method, and fold changes were calculated relative to an off-target control sgRNA targeting the *B2m* gene (Supplementary Table 4).

Bioinformatics procedures. Normalization and batch correction. Critical threshold values were obtained from qPCR profiling with an upper limit of 40. Across-sample normalization was performed for each qPCR plate separately as follows: (1) the arithmetic mean of all assays for each sample was calculated; (2) the median of sample means was divided by the arithmetic mean for each sample to create a scaling factor; and (3) Ct values for each sample were multiplied by their scaling factor. Following normalization, individual assay outliers were identified either by visual inspection or based on a z-score >4, relative to all other samples (Supplementary Table 5). Next, normalization was redone with outlier assays set as missing. Missing values were then imputed using the mice package in R⁵⁴ with default settings and ten iterations.

Batch correction was performed only for a subset of miRNAs (223 in total; Supplementary Table 5) that were assessed as having a significant association with batch, because batch correction of all miRNAs would have removed signal from cell-type-specific miRNAs seen in only a limited number of batches. Batch-associated miRNAs were selected including all cell types that had at least one replicate in two different batches. Variation due to cell type was removed using the removeBatchEffect function from the limma package⁵⁵, then an *F*-test in limma was performed with respect to batch for each miRNA on the resulting values. miRNAs were selected for batch correction if they had a Bonferroni adjusted $P < 0.01$ and mean Ct < 35 across all samples involved. After miRNA selection, batch correction was performed using ComBat⁵⁶ only for selected miRNAs.

Following normalization and batch correction, Ct values were converted to arbitrary units (AU) for easier interpretation. We determined a limit of reliable detection at Ct = 28, which was assessed through the standard curves of ten miRNAs and inter-replicate variability at different Ct thresholds. Our AU scale was designed so that a Ct value of 28 corresponded to a value of 1 AU. Ct values were converted to AU with the formula: $2^{(28 - \text{exp}[i,j])}$, where exp is the Ct value matrix for all samples and miRNAs, *i* is a given miRNA and *j* is a given cell type. A pseudocount of 0.1 was added to all assays following conversion. Expressed miRNAs were denoted as those that were detected at >1 AU in all replicates of at least one multireplicate cell type, giving 442 miRNAs expressed in the immune system from our panel. For all miRNA-family-based expression values, the linear AU values of all miRNAs within a given family, based on TargetScan v.7 (ref. ⁵⁷) annotations, were summed.

Cross-reactivity analysis. Quantitative PCR assays can sometimes be subject to cross-reactivity with other RNA molecules due to similarity between target sequences. To identify situations where this might arise, we calculated the Levenshtein edit distance between all pairs of miRNAs using the BioStrings package in R⁵⁸. Next, using mean log₂ AU population expression values, the Pearson correlation was calculated for all pairs of miRNAs with an edit distance <6, removing miRNA pairs in the same cluster. Potential cross-reactive pairs were flagged if they had a correlation slope >0.7, leaving 29 miRNA pairs with potential

overlap in signal (Supplementary Table 6). While it was difficult to definitively prove an overlap in signal, instances were noted where directionality of the signal probably fell towards one miRNA or the other based on public data or chromatin accessibility in the region. In certain instances, these miRNAs were removed or modified in display figures.

Cell specificity indices. To measure the specificity of expression we used the Tau statistic, calculated for each miRNA using the formula: $\tau = \frac{\sum_{i=1}^N (1-x_i)}{N-1}$ (ref. 59), where N is the number of cell types and x_i the expression value for a given cell type divided by the maximum expression value for that miRNA across samples. Mean \log_2 AU values for steady-state cell types were used as expression values. Only miRNAs expressed at >4 AU in all replicates of at least one cell type were used. For Tau calculation, \log_2 AU values <0 were set to 0. Selection of cell-specific miRNAs was performed by taking miRNAs with $\tau > 0.92$ and $>45\%$ of their total linear abundance in a single cell type. Specificity scores can be found in Supplementary Table 7.

Differential expression and clustering. Differential expression (DE) was performed under several circumstances that required adjustment of cutoffs, depending on how many samples were involved in the test and the aim of the analysis. Although there were several conditions, \log_2 AU values and the two-sided statistical framework from the limma package in R³⁵ were always used to assess significance and fold change. Cutoffs and thresholds for each particular test are listed in their respective figure legends. A strategy for selection of miRNAs specific to certain groups of cells among many was used in the analysis of tissue macrophages, lymphoid lineage cells, mature T cells and perturbed populations. This method used two-condition limma DE tests between every combination of two cell groups, selected against each other. Following the DE tests for all two-group comparisons, miRNAs were assessed for significance and consistency of fold change between one cell type and all other cell groups (lymphoid), all but one other cell group (tissue macrophage, mature T cell) or three or more perturbation conditions.

The samples used in most comparisons are delineated in the heatmaps or boxplots showing the results of DE, and all replicates for a given population were always used. Female samples were excluded from all analysis. DE statistics from all tests are included in Supplementary Table 8.

Hierarchical clustering of all miRNA profiling samples was performed using mean \log_2 AU abundance values. Only miRNAs reaching high abundance (>32 AU) in at least one cell type were kept for input, leaving 150 miRNAs. Expression values <0 \log_2 AU were set to 0. One minus the Pearson correlation between cell types was used as the distance measure, and clustering was done using hclust in R with complete agglomeration for cell types. Euclidean distance and complete agglomeration were used to cluster miRNAs.

miRNA promoter annotation. Annotation aggregation was used to determine TSS/promoter sites for as many miRNAs as possible. We aggregated annotations from seven different sources and linked these to OCRs identified in the ImmGen *cis*-Atlas³³. The term promoter is used as an encompassing term to define these annotations, because some sources predicted a specific TSS site and some a less specific promoter region. Sources included three empirical sequencing-based annotations from Chang et al.²⁰, De Rie et al.²² (denoted Fantom5) and Georgakilas et al.¹⁹ (denoted drshaKO_mESC), and three inference-based annotations from Georgakilas et al.¹⁹ (denoted microTSS), Marson et al.²¹ and Kirigin et al.²¹. Additionally, we used start sites of annotated transcripts in Gencode³⁴ release M20 that had a transcript support level of 1–3 or 'NA' and fully overlapped with a pre-miRNA. Cell types in which biochemical mappings were done are listed in Supplementary Table 9. In addition, only Chang et al. transcripts that fully overlapped with the pre-miRNA were used. For Marson et al. predictions, only those with a score >0 were used. miR-451a required a specific literature annotation⁶⁰, because it was not represented in the other studies. miRbase v.22 was used for all coordinates and annotations of miRNA sequences, with coordinates in all annotation files converted to mm10 by the UCSC batch conversion tool if necessary (<https://genome.ucsc.edu/cgi-bin/hgLiftOver>). Several entries within these files were modified after visual inspection for proper strandedness, orientation, naming and corroboration by other data such as publicly available nascent RNA sequencing^{61–68} (BioProject accession nos. PRJNA436058, PRJNA437913, PRJNA259674, PRJNA211259, PRJNA360121, PRJNA248435, PRJNA338193 and PRJNA482330), POLR2A ChIP-seq^{69,70} (BioProject accession nos. PRJNA302909 and PRJNA129939), cap analysis gene expression (CAGE)-seq peaks (http://fantom.gsc.riken.jp/5/datafiles/latest/extra/CAGE_peaks/) and conservation scores (<http://hgdownload.cse.ucsc.edu/goldenpath/mm10/phastCons60way>). The resulting aggregated annotation set provided us with 15,093 miRNA promoter annotations in the mouse genome.

Annotations collected from these sources were then mapped to active genomic regions in the immune system by intersecting them with OCR coordinates as defined in the ImmGen *cis*-Atlas study (BioProject accession no. PRJNA392905). For our analysis we chose to include the blacklisted OCRs (<https://sites.google.com/site/anshulkundaje/projects/blacklists>) and regions showing homology to chrM removed previously, because some of these regions overlap with miRNA promoter sites (that is, the miR-17~92 cluster). After filtering 518,844 total OCRs

for sites with at least five edge counts in at least two libraries, 505,802 OCRs remained for use in our analysis, all of which were 250 base pairs (bp) in width. Signal intensities and OCR *P* values were calculated in the same manner as in a previous study³³. Overlaps between OCRs and TSS/promoter annotations were found using BEDTools v.2.26.0 (ref. 71), with the 'closest' command. The Chang et al. and GenCode annotations were retained as annotated promoters if an OCR was within 100 bp upstream or 25 bp downstream of the promoter prediction. Fantom5, Georgakilas et al. and Kirigin et al. overlaps were retained if the linked OCR was <50 bp distant, and Marson et al. annotated promoters were retained only in the case of a direct overlap. These cutoffs were determined after manual curation of each annotation source.

Filtering of putative TSS/promoter sites was performed as follows. For filtering and QC H3K4me3 and POLR2A ChIP-seq data were downloaded for all 13 adult mouse bulk tissues from the ENCODE consortium⁷², in addition to H3K4me3 data from Lara-Astiaso et al. (indexing-first ChIP (iChIP))⁷³ (BioProject accession no. PRJNA255796), which is comprised of 16 broad hematopoietic cell types. H3K4me3 peaks for the iChIP samples were called using the same methodology as the original paper, through the Homer v.4.8 software suite⁷⁴. H3K4me3 peaks for ENCODE tissues were downloaded using the data portal (<https://www.encodeproject.org/>) in the bed broadPeak format, while POLR2A ChIP-seq data were downloaded in the bed narrowPeak format using optimal idr thresholded peaks. Putative promoters were then filtered out if they did not overlap with an H3K4me3 peak from either ENCODE bulk tissues or iChIP populations. Also, if a promoter OCR was predicted by Chang et al. or Marson et al. alone, these were retained only if they overlapped a POLR2A ChIP-seq peak from the ENCODE samples, because these predictions had higher rates of false positives. The resulting filtered set of annotations had, overall in mice, 1,315 promoter OCRs for 323/390 expressed pre-miRNAs, 381/442 expressed mature miRNAs and 1,417/1,965 mature miRNAs (Supplementary Table 9). Pre-miRNAs detected by expression but lacking annotation were generally from either duplicated miRNAs active at only one locus, lowly expressed or were specifically expressed in populations profiled for miRNA and not ATAC-seq. Additional miRNAs lacking annotation were, in some cases, identified as potentially cross-reactive with other miRNAs in the qPCR assay and were flagged in the final quality control (Supplementary Table 6). An additional possibility is that a miRNA is exogenously derived in the cell types we already have. Promoter/TSS regions were also manually inspected for concurrence with publicly downloaded GRO-seq data, and regions were flagged or pruned if it appeared they were suspect in producing miRNA-spanning transcripts, or if it was clear that the peak did not intersect the true TSS site in representative immune populations. This was prioritized to maintain promoter/TSS annotations as close to the TSS and immediate proximal promoter as possible. Peaks annotated as promoter/TSS sites remained defined as such throughout the study.

Chromatin accessibility and miRNA expression joint analysis. To associate miRNA expression variation with changes in chromatin accessibility globally, a correlation analysis was carried out using the 22 overlapping samples for which there was ATAC-seq and miRNA expression profiling (Extended Data Fig. 3c). Here, only OCRs open above background (open: Benjamini–Hochberg adjusted MACS2 $P < 0.05$ throughout study) and with activity >10 in at least one of the overlapping samples were used. Additionally, OCRs denoted as miRNA TSS/promoter sites or TSS sites in the *cis*-atlas study (UCSC mRNA and long noncoding RNA annotations) were removed. To identify candidate regulatory regions, Pearson correlation was calculated between the \log_2 ATAC-seq activity at each separate OCR within 100 kb of an annotated miRNA promoter and that miRNA's \log_2 AU expression value. Correlation *P* values were adjusted for multiple testing by Benjamini–Hochberg⁷⁵ correction. Significant associations were determined by an adjusted *P* value cutoff of 0.1, providing 395 significant associations for 140/344 miRNAs tested detected in the overlap samples and with promoter annotations (Fig. 3 and Supplementary Table 10). Analysis of whether predicted dCREs fell within the same TAD was done using TAD boundaries as defined by Johanson et al.³⁵. Each dCRE was checked for concurrence with the miRNA promoter TAD in splenic neutrophils, B cells, CD4⁺ T cells and mouse embryonic fibroblasts (MEFs). Discrepancies between promoter and dCRE TADs in any of the cell types are reported in Supplementary Table 10.

Promoter regression analyses. A stepwise linear regression framework was used to determine the contributions of multiple distinct promoters at a single locus for the same miRNA on its expression (Extended Data Fig. 6). To select for distinct promoter regions, annotated promoter OCRs were merged using the BEDTools merge command, taking OCRs within 2 kb of each other and summarizing the ATAC-seq signal and MACS2 *P* values per cell type to the maximum and minimum in the merged OCRs, respectively. Next, conserved and expressed miRNAs with multiple merged promoter regions at the same locus were selected for testing, 27 in total. Linear regression models were built in a stepwise manner for each multipromoter miRNA, using the stepAIC function with direction set to 'both', from the MASS package in R³⁶, with chromatin accessibility at merged promoter regions as the predictor of miRNA expression (Supplementary Table 11). Analysis to examine miRNA expression in relation to the number of open promoter regions

was done in a similar fashion (Fig. 4b–e). Linear models were built, at both the global (all miRNAs at once) and individual miRNA level, using the `lm` function in R, with the integer count of open promoter regions as the predictor and `log2` AU miRNA expression as the response (Supplementary Table 11). An unadjusted *P* value cutoff of 0.1 was used to determine significance of regression terms for individual miRNA models. The additional filters for this analysis using nascent RNA and H3K4me3 are explained in the following two sections.

Histone mark ChIP-seq analysis in aligned populations. Histone mark ChIP-seq datasets from Lavin et al.⁷⁷ (BioProject accession no. PRJNA267535), iChIP⁷³ and ENCODE consortium data from MEFs⁷² were downloaded and mapped to the mm10 reference genome using Bowtie2 (ref.⁷⁸) v.2.3.4.3 with the settings ‘-very-sensitive-local-no-mixed-no-discordant’. Tag directories were created from the resulting bam files using Homer, allowing for only uniquely mapped reads and a maximum of three identical reads per position. ChIP-seq signal was then quantified by measuring the number of tags per 10 million mapped tags in 2-kb regions centered on ATAC-seq peaks identified in the *cis*-atlas, excluding those in ENCODE blacklisted regions not in miRNA loci. The resulting signals were then quantile normalized separately for each histone mark (Supplementary Data 1), and a threshold of 15 or more on the normalized scale was used to delineate the presence of a mark. Pearson correlations between H3K27ac signal and miRNA expression or ATAC-seq signal were calculated using log values, and can be found in Supplementary Table 10.

Nascent RNA analysis. Sequencing reads from GRO/PRO-seq datasets^{51–68} were trimmed and pruned via Cutadapt v.2.10 (ref.⁷⁹), with settings ‘-a ‘A{100}’ -q 10 -m 25. Subsequently, Bowtie2 with settings ‘-local -N 0 filtered-out reads mapping to ribosomal RNA (https://support.illumina.com/sequencing/sequencing_software/igenome.html), non-miRNA blacklisted regions or small Cajal-body-specific RNA/small nucleolar RNA loci (RFAM v.14.2 (ref.⁸⁰)). The remaining reads were mapped to the mm10 genome by Bowtie2 with settings ‘-local -N 1. Tag directories and bigwig files were then created from the resulting bam files using default settings and merging replicate samples. De novo transcript calling was performed in Homer with the `findPeaks` function and the option ‘-style groseq flag and -uniqmap. Similar to Bouvy-Liivrand et al.³⁶, several `minBodySize`, `minReadDepth`, `bodyFold` and `tssFold` settings were used to account for depth and quality differences across nascent RNA experiments. There were four settings used to call transcripts: ‘-minBodySize 1000 -minReadDepth 10, -minBodySize 4000 -minReadDepth 5 -tssFold 3 -bodyFold 3, -minBodySize 900 -minReadDepth auto and -minBodySize 800 -minReadDepth auto -tssFold 3 -bodyFold 2.5. For each cell type, all transcripts called were assigned transcript clusters (connected de novo transcripts) using BEDTools cluster with settings ‘-s -d 750. A miRNA-spanning nascent transcript was defined as one that initiated within 2 kb of an annotated promoter region and was part of a transcript cluster that spanned the pre-miRNA. dCRE transcripts were counted if a called nascent transcript initiated within 100 bp of the element. Called transcripts for each cell type can be found in Supplementary Data 2.

Duplicated loci distance. To calculate distance values across duplicated miRNA loci, merged promoter values were used as generated above. Each promoter was coded as 1 or 0 depending on the open or closed status in a given cell. The Manhattan distance, which equates to the number of discrepancies between open/closed status, was then calculated between each pairing of promoter regions at different loci for the same mature miRNA using all 87 populations from the ImmGen *cis*-Atlas (Fig. 5c).

Visualization. All ATAC-seq and other read pile-up data were visualized in Integrative Genomics Viewer (IGV)⁸¹. All tracks were always normalized to the highest-value peak among each data type unless otherwise specified. Numbers on the left-hand side indicate track height data range.

Data modifications. The name miR-199a-3p (accession no. MIMAT0000230) was changed to miR-199a/b-3p because it has the same mature sequence as miR-199b-3p (accession no. MIMAT0004667).

Statistics and reproducibility. All statistical procedures and calculations were performed in the R language for statistical computing using the packages and methods described above. miRNA profiling sample sizes were chosen to maximize breadth of populations as stated above, with no statistical method used to determine sample size. Sample quality was assessed by clustering, Ct value comparisons as in Extended Data Fig. 1b and overall level of nonmissing qPCR assay amplification. Populations with only a single replicate were excluded from the overall analysis. The investigators were not blinded during experiments and outcome assessment. All boxplots show the upper and lower borders of boxes representing the 25th and 75th percentile, respectively, and the middle line is the median. Whiskers extend 1.5 × interquartile range and points represent values beyond the whiskers.

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

Data availability

Data that support the findings of this study are available on the ImmGen website (www.immgen.org), and raw and processed miRNA expression data, including sample metadata, are available at the GEO (accession no. GSE144081). Additionally, tables with ATAC-seq signal, *P* values and peak locations with previously blacklisted peaks included are provided (Supplementary Data 3). Sequence Read Archive ID and other sample information for downloaded ChIP-seq and nascent RNA datasets can be found in Supplementary Table 12. Processed histone mark and nascent RNA data are available in Supplementary Data 1 and 2, respectively. Source data for all figures in this manuscript are provided. External mRNA-seq and ATAC-seq data were downloaded through the ImmGen website. miRNA promoter annotations were downloaded from the supplements of their respective studies cited in the text or the GENCODE database (<https://www.genecodegenes.org/mouse/>). TAD boundary data were downloaded from Johanson et al.³⁵. miRNA conservation and other information was downloaded from TargetScan v.7 (www.targetscan.org). ENCODE blacklist regions for mm10 were downloaded from <https://sites.google.com/site/anshulkundaje/projects/blacklists>. rRNA sequence and sca/snoRNA loci were retrieved from iGenomes (https://support.illumina.com/sequencing/sequencing_software/igenome.html) and RFAM v.14.2 (<https://rfam.xfam.org/>), respectively. CAGE peaks from the FANTOM5 consortium were downloaded from their website (http://fantom.gsc.riken.jp/5/datafiles/latest/extra/CAGE_peaks/). phastCons conservation scores were downloaded from <http://hgdownload.cse.ucsc.edu/goldenpath/mm10/phastCons60way>. Source data are provided with this paper.

Code availability

Custom code used in analysis will be made available upon request. Code for normalization and batch correction of qPCR data is available at <https://github.com/srose89/ImmGen-miRNA>.

References

- Baccarini, A. & Brown, B. D. Monitoring microRNA activity and validating microRNA targets by reporter-based approaches. *Methods Mol. Biol.* **667**, 215–233 (2010).
- Prescott, S. L. et al. Enhancer divergence and *cis*-regulatory evolution in the human and chimp neural crest. *Cell* **163**, 68–83 (2015).
- Cheung, S. T., Shakibakho, S., So, E. Y. & Mui, A. L. F. Transfecting RAW264.7 cells with a luciferase reporter gene. *J. Vis. Exp.* **100**, 52807 (2015).
- Nüssing, S. et al. Efficient CRISPR/Cas9 gene editing in uncultured naive mouse T cells for in vivo studies. *J. Immunol.* **204**, 2308–2315 (2020).
- Wroblewska, A. et al. Protein barcodes enable high-dimensional single-cell CRISPR screens. *Cell* **175**, 1141–1155 (2018).
- van Buuren, S. & Groothuis-Oudshoorn, K. Mice: multivariate imputation by chained equations in R. *J. Stat. Softw.* <http://hdl.handle.net/10.18637/jss.v045.i03> (2011).
- Ritchie, M. E. et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* **43**, e47 (2015).
- Johnson, W. E., Li, C. & Rabinovic, A. Adjusting batch effects in microarray expression data using empirical Bayes methods. *Biostatistics* **8**, 118–127 (2007).
- Agarwal, V., Bell, G. W., Nam, J. W. & Bartel, D. P. Predicting effective microRNA target sites in mammalian mRNAs. *Elife* **4**, e05005 (2015).
- Pagès, H., Aboyoun, P., Gentleman, R. & DebRoy, S. Biostrings: efficient manipulation of biological strings. R package v2.46.0 (2017); <https://bioconductor.org/packages/Biostrings>
- Yanai, I. et al. Genome-wide midrange transcription profiles reveal expression level relationships in human tissue specification. *Bioinformatics* **21**, 650–659 (2005).
- Dore, L. C. et al. A GATA-1-regulated microRNA locus essential for erythropoiesis. *Proc. Natl Acad. Sci. USA* **105**, 3333–3338 (2008).
- Bönelt, P. et al. Precocious expression of Blimp1 in B cells causes autoimmune disease with increased self-reactive plasma cells. *EMBO J.* **38**, e100010 (2019).
- Danko, C. G. et al. Dynamic evolution of regulatory element ensembles in primate CD4⁺T cells. *Nat. Ecol. Evol.* **2**, 537–548 (2018).
- Hah, N. et al. Inflammation-sensitive super enhancers form domains of coordinately regulated enhancer RNAs. *Proc. Natl Acad. Sci. USA* **112**, E297–E302 (2015).
- Kaikkonen, M. U. et al. Remodeling of the enhancer landscape during macrophage activation is coupled to enhancer transcription. *Mol. Cell* **51**, 310–325 (2013).
- Nair, S. J. Phase separation of ligand-activated enhancers licenses cooperative chromosomal enhancer assembly. *Nat. Struct. Mol. Biol.* **26**, 193–203 (2019).
- Nelson, V. L. PPAR γ is a nexus controlling alternative activation of macrophages via glutamine metabolism. *Genes Dev.* **32**, 1035–1044 (2018).
- Wei, C. Repression of the central splicing regulator RBFox2 is functionally linked to pressure overload-induced heart failure. *Cell Rep.* **10**, 1521–1533 (2015).

68. Zhu, Y. Comprehensive characterization of neutrophil genome topology. *Genes Dev.* **31**, 141–153 (2017).
69. Mostafavi, S. Parsing the interferon transcriptional network and its disease associations. *Cell* **164**, 564–578 (2016).
70. Escoubet-Lozach, L. et al. Mechanisms establishing TLR4-responsive activation states of inflammatory response genes. *PLoS Genet.* **7**, e1002401 (2011).
71. Quinlan, A. R. & Hall, I. M. The BEDTools manual. (2010); <https://github.com/arq5x/bedtools2>
72. The ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**, 57–74 (2012).
73. Lara-Astiaso, D. et al. Chromatin state dynamics during blood formation. *Science* **345**, 943–949 (2014).
74. Heinz, S. et al. Simple combinations of lineage-determining transcription factors prime *cis*-regulatory elements required for macrophage and B cell identities. *Mol. Cell* **38**, 576–589 (2010).
75. Benjamini, Y. & Hochberg, Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc. Series B Stat. Methodol.* **57**, 289–300 (1995).
76. Venables, W. N. & Ripley, B. D. *Modern Applied Statistics with S-Plus* (Springer, 2002).
77. Lavin, Y. et al. Tissue-resident macrophage enhancer landscapes are shaped by the local microenvironment. *Cell* **159**, 1312–1326 (2014).
78. Langmead, B. & Salzberg, S. Bowtie2. *Nat. Methods* **9**, 357–359 (2012).
79. Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet J.* **17**, 10–12 (2011).
80. Kalvari, I. et al. Rfam 14: expanded coverage of metagenomic, viral and microRNA families. *Nucleic Acids Res.* **49**, D192–D200 (2021).
81. Robinson, J. T. et al. Integrative genomics viewer. *Nat. Biotechnol.* **29**, 24–26 (2011).

Acknowledgements

We thank all members of the ImmGen Consortium for their comments and advice throughout this work and for their critical feedback on the manuscript, especially

C. Benoist (Harvard), L. Lanier (UCSF), S. Nutt (WEHI), P. Monach (Boston U), S. Turley (Genetech) and D. Hasson (Mount Sinai). B.D.B. was supported by the NIH (nos. R01AI113221 and R01AT011326), the Cancer Research Institute and the Alliance for Cancer Gene Therapy. M.M. was supported by the NIH (nos. R01CA257195 and R01CA254104), and S.A.R. by no. T32AI007605. J.D.B. acknowledges support from the NIH Director's New Innovator Award (no. DP2HL151353). This work was also supported by the NIH (no. R24AI072073) to the ImmGen Consortium.

Author contributions

S.A.R. designed and performed experiments, analyzed data and wrote the manuscript. A.W. and M.D. performed experiments and edited the manuscript. H.Y. analyzed data and edited the manuscript. B.B.-Z. and A.B. provided technical assistance. J.M.S. carried out qPCR profiling. A.R., E.Y.K., B.Y. and Y.L. provided samples. M.M. designed and supervised the research, analyzed data and edited the manuscript. J.D.B. analyzed data and edited the manuscript. B.D.B. designed and supervised the research, analyzed data and wrote the manuscript.

Competing interests

J.M.S. is an employee and stockholder of Qiagen Sciences. J.D.B. holds patents related to ATAC-seq and is on the scientific advisory boards of Camp4, Seqwell, and Celsee. The remaining authors declare no competing interests.

Additional information

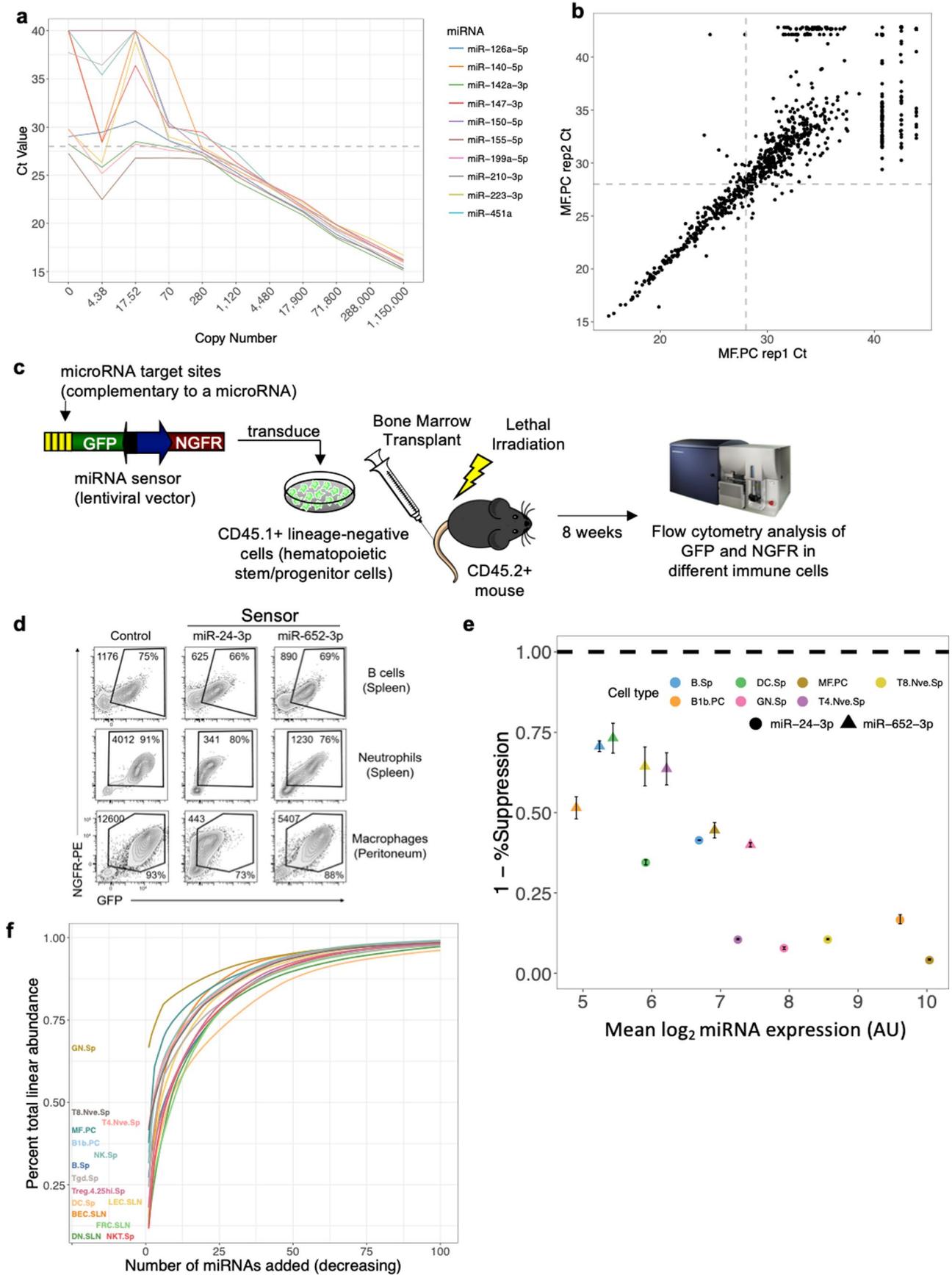
Extended data is available for this paper at <https://doi.org/10.1038/s41590-021-00944-y>.

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41590-021-00944-y>.

Correspondence and requests for materials should be addressed to B.D.B.

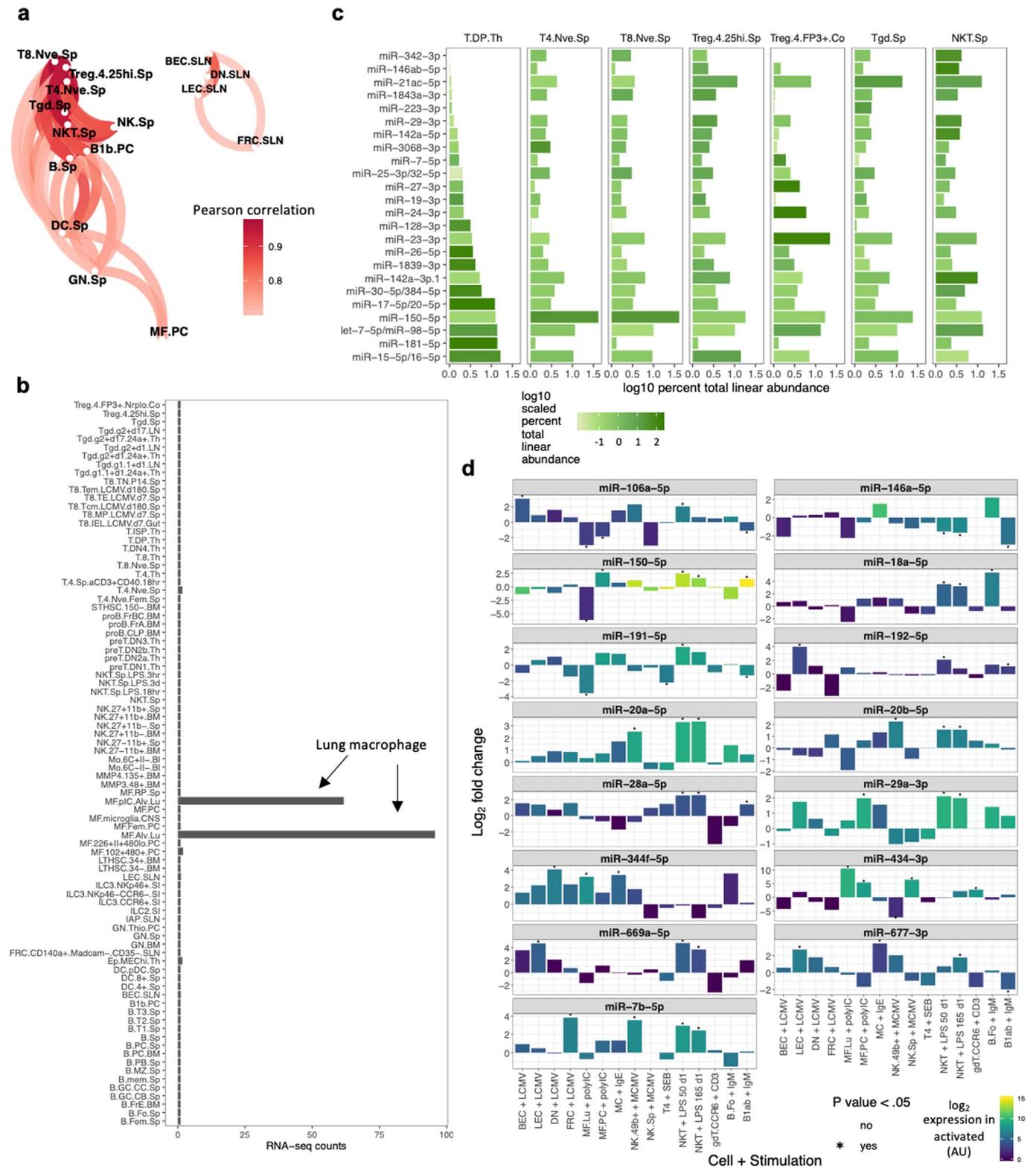
Peer review information *Nature Immunology* thanks Musa Mhlanga and Massimiliano Pagani for their contribution to the peer review of this work. Jamie D. K. Wilson and Laurie Dempsey were the primary editors on this article and managed its editorial process and peer review in collaboration with the rest of the editorial team.

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

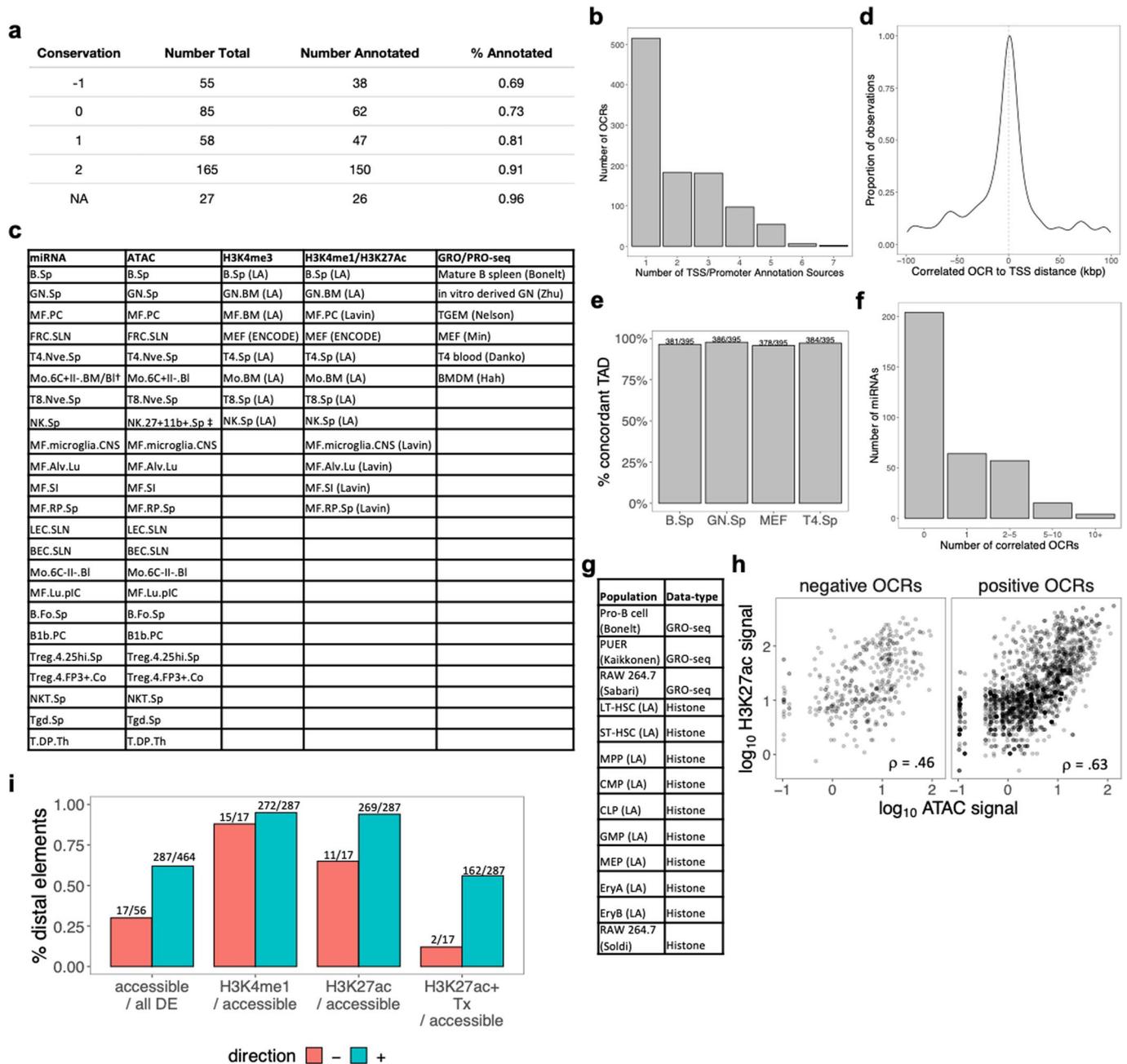


Extended Data Fig. 1 | See next page for caption.

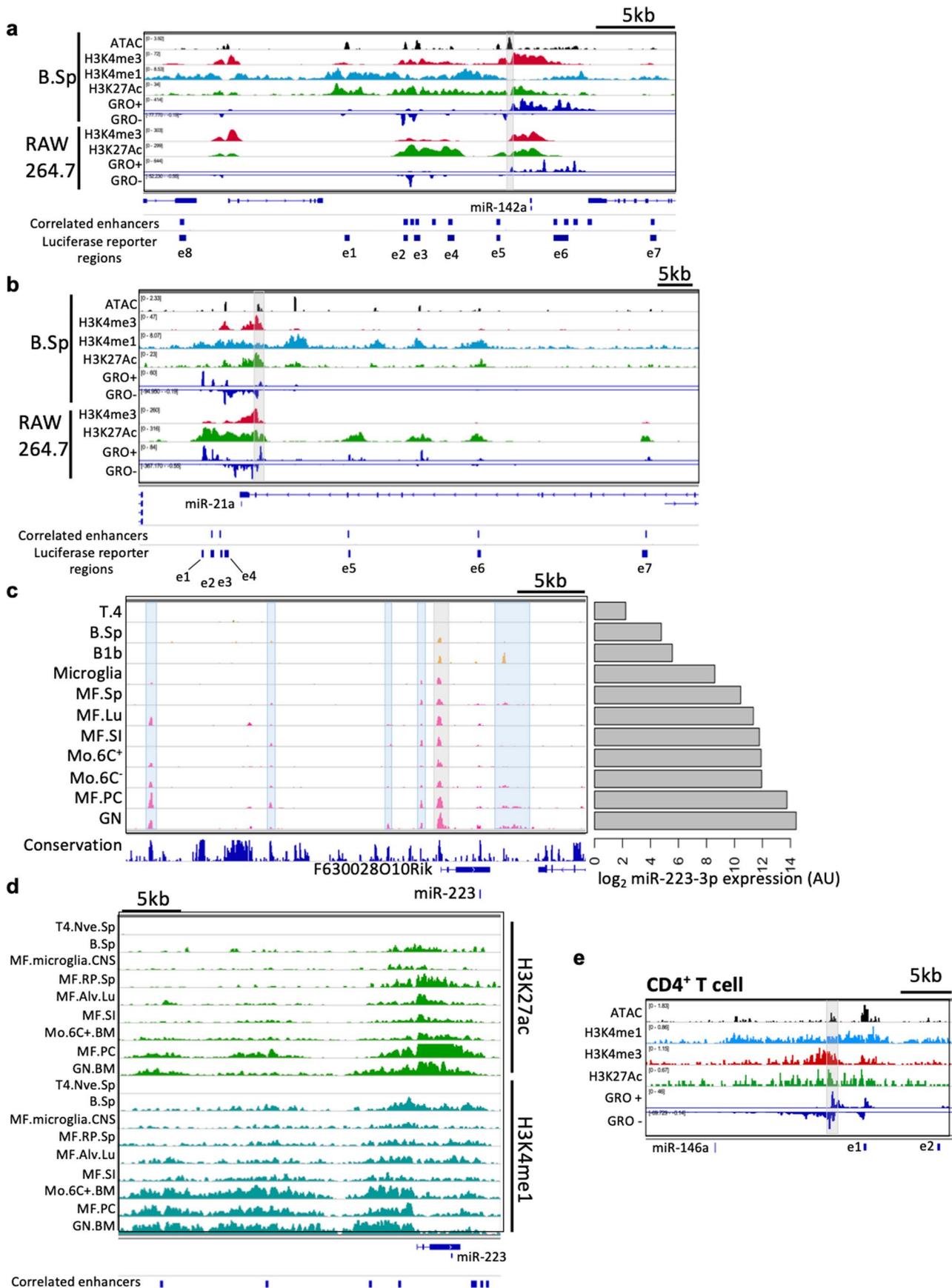
Extended Data Fig. 1 | Determination of miRNA abundance thresholds. **a**, qPCR standard curve generated for 10 different miRNAs using synthetic RNA mimics. Plotted are copies of synthetic mature miRNA species input into qPCR reaction against the corresponding Ct value. Horizontal grey dashed line marks a Ct value of 28. **b**, Representative scatterplot of Ct values for two PC macrophage replicates against each other. **c**, Schematic diagram of the in vivo miRNA sensor assay. Lineage negative cells were isolated from CD45.1⁺ C57BL/6 mice, transduced with lentiviral-based sensors for miR-24-3p or miR-652-3p, or a control vector, and transplanted in to lethally irradiated CD45.2⁺ mice. After 8 weeks, cells from the spleen and peritoneal cavity were collected, stained for immune cell markers, and NGFR, and analyzed by flow cytometry. **d**, Representative flow plots showing logarithmic fluorescence intensities for NGFR and GFP from mice that received a sensor for the indicated microRNA. Values are the percent of NGFR⁺ GFP⁺ cells in the total population. Accompanying integer values for each plot represent MFI of GFP of all NGFR⁺ cells within the plot. **e**, Correlation between miRNA sensor suppression and miRNA expression in 7 cell types from the 11-cell set. Percent suppression was calculated as one minus the target miRNA GFP/NGFR median fluorescence intensity divided by the average of the same ratio for all 3 non-targeting control replicates in a given cell type. Error bars represent the standard error of the mean (miR-24-3p n = 2 mice; miR-652-3p n = 4 mice). **f**, Cumulative percent of total linear abundance within a given cell type compared to the number of miRNAs added in decreasing order of expression.



Extended Data Fig. 2 | miRNA abundance patterns in T cells and after perturbation. **a**, Pearson correlation of the 11-cell immune subset and lymph node stroma cells based on miRNA expression signatures. Expression data was filtered on miRNAs that are high-abundance (>32 AU) in at least one cell type. Only correlations between samples >0.7 are plotted. **b**, *Htr2c* read counts in ImmGen cis-Atlas samples. **c**, Percent linear abundance of the 15 most highly expressed miRNA and miRNA families in T cell subsets and DP thymocytes cells. Bars are shaded by Z-score value of the miRNA family across populations. **d**, miRNAs changing consistently in 3 or more perturbation conditions not highlighted in Fig. 2d. (limma two-sided unadjusted $p < 0.05$, \log_2 FC > 1, and expression >4AU in perturbed or >32AU in steady-state population; $n = 2$ for all activated and stimulated populations except: NK.Sp = 5, NKT = 5, B1ab = 5).

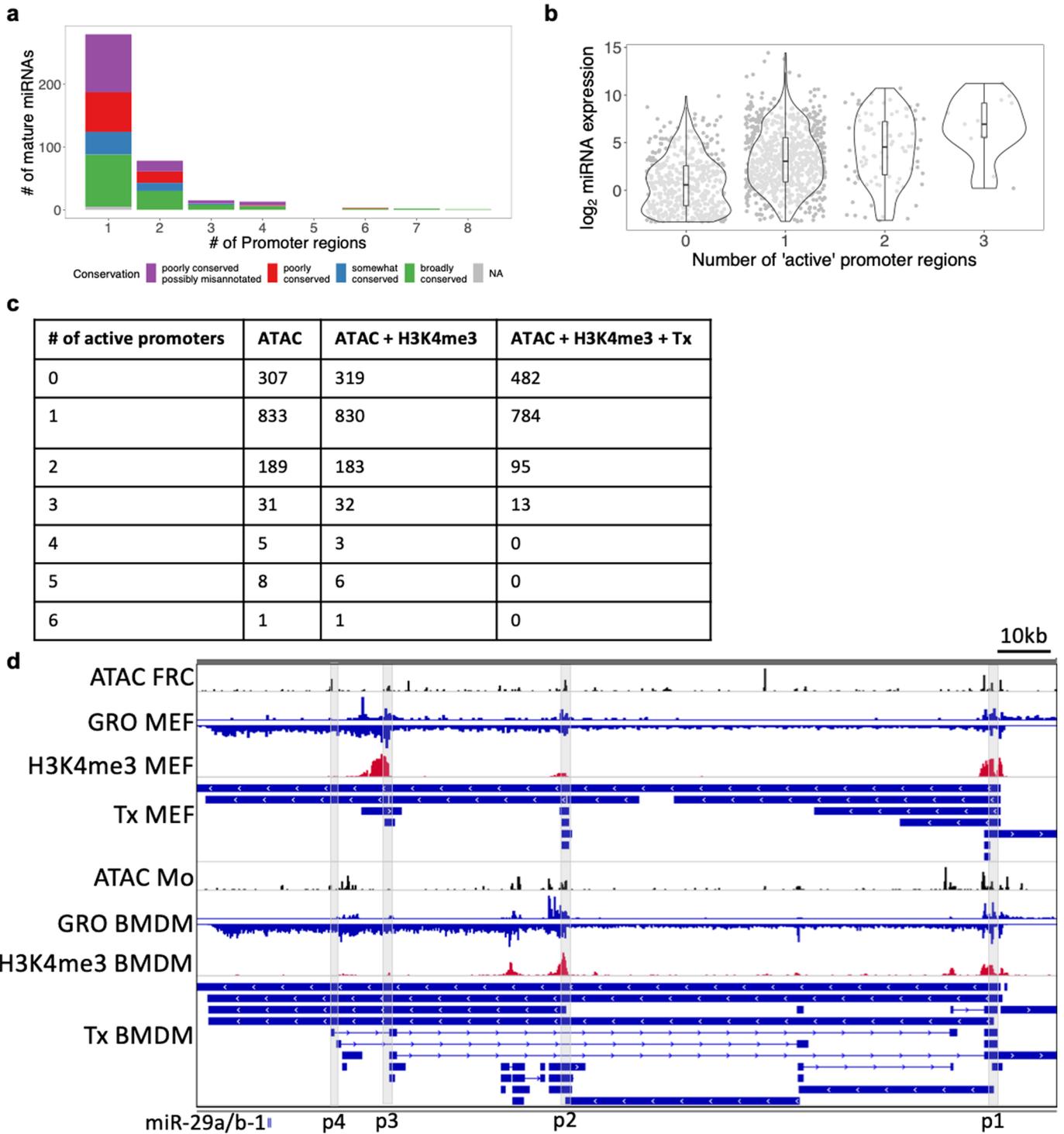


Extended Data Fig. 3 | Characterization of miRNA associate cis-elements. **a**, Table displaying the number of pre-miRNAs (having an expressed mature isoform) with promoter annotations after aggregation, broken down by TargetScan V7 conservation category. **b**, The number of annotation sources from compiled studies annotating a particular OCR as a promoter region. **c**, Table displaying the alignment of cell types from externally downloaded datasets with ImmGen miRNA and ATAC-seq populations for integrative analysis. † = BM monocyte miRNA profile only used for H3K27ac signal to miRNA expression correlation. ‡ = NK cells were not part of 22 overlapping cell types used for miRNA to ATAC-seq correlations. **d**, Distance from significantly correlated OCRs to an annotated miRNA promoter/TSS. **e**, Percent of correlated OCRs within the same TAD as the promoter for the same miRNA according to TAD definitions in 4 listed cell types from Johanson *et al.* **f**, Frequency of miRNAs with different numbers of significantly correlated OCRs. **g**, Unaligned additional datasets incorporated in promoter and enhancer analyses. **h**, \log_{10} ATAC-seq signal compared with \log_{10} H3K27ac signal at associated distal elements across the 6 fully aligned populations in **c**. **i**, Fraction of associated DEs in either direction of effect above or below high-abundance miRNAs in the 6 fully aligned populations meeting various molecular criteria of active enhancer elements. Bars from left to right represent the number of accessible putative DEs of total possible for expressed miRNAs, the number of accessible putative DEs marked with H3K4me1 or H3K27ac, and the number of accessible putative DEs marked with H3K27ac and with nascent RNA transcripts detected.

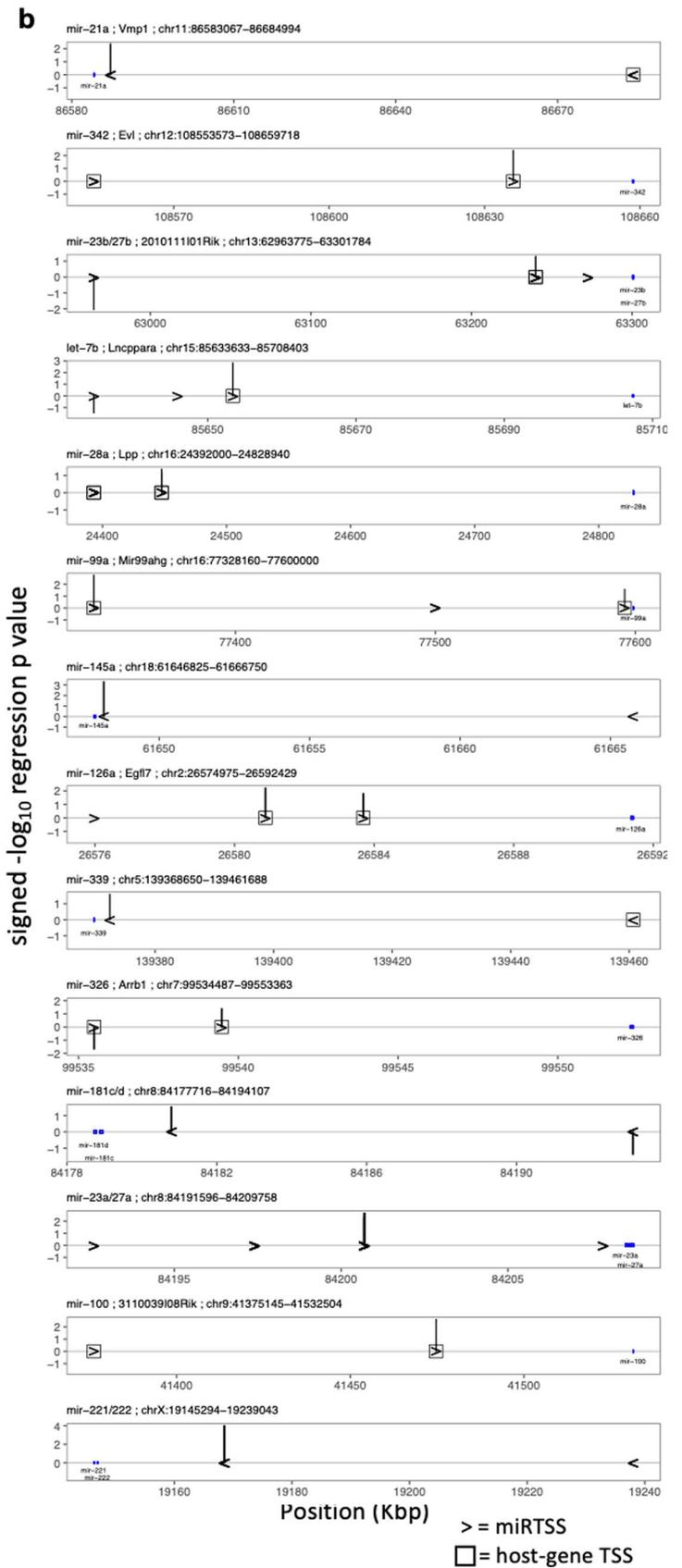
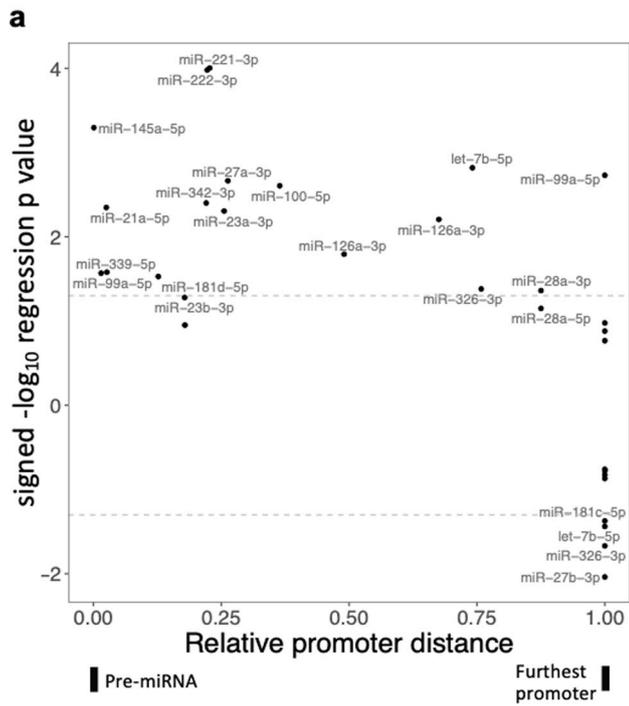


Extended Data Fig. 4 | See next page for caption.

Extended Data Fig. 4 | Histone mark and nascent RNA visualization at select miRNA loci. **a,b**, IGV plot of layered available molecular information at miR-142 (a) and miR-21a (b) loci in splenic B cells and RAW 264.7 macrophages. Correlated DEs from Fig. 3a and regions selected for luciferase reporter assays displayed in Fig. 3b are labeled. Lanes are normalized individually. Promoter regions are shaded in gray for all panels. **c**, Representative read pile-up tracks of ATAC-seq signal, all normalized to same scale, showing differential *cis*-element accessibility at the miR-223 locus in select cells. Peak highlighted in gray is the pri-miR-223 promoter region and peaks highlighted in light blue are correlated elements with miR-223 expression from analysis in Fig. 3a. **d**, H3K27ac and H3K4me1 read pile-ups in aligned histone mark populations at the miR-223 locus. Tracks are normalized by histone mark. **e**, IGV plots of histone mark and nascent RNA signatures at associated distal elements in CD4⁺ T cells at the miR-146a locus. e1 corresponds to the enhancer site targeted by flanking sgRNAs in Fig. 3d. Lanes are normalized individually.



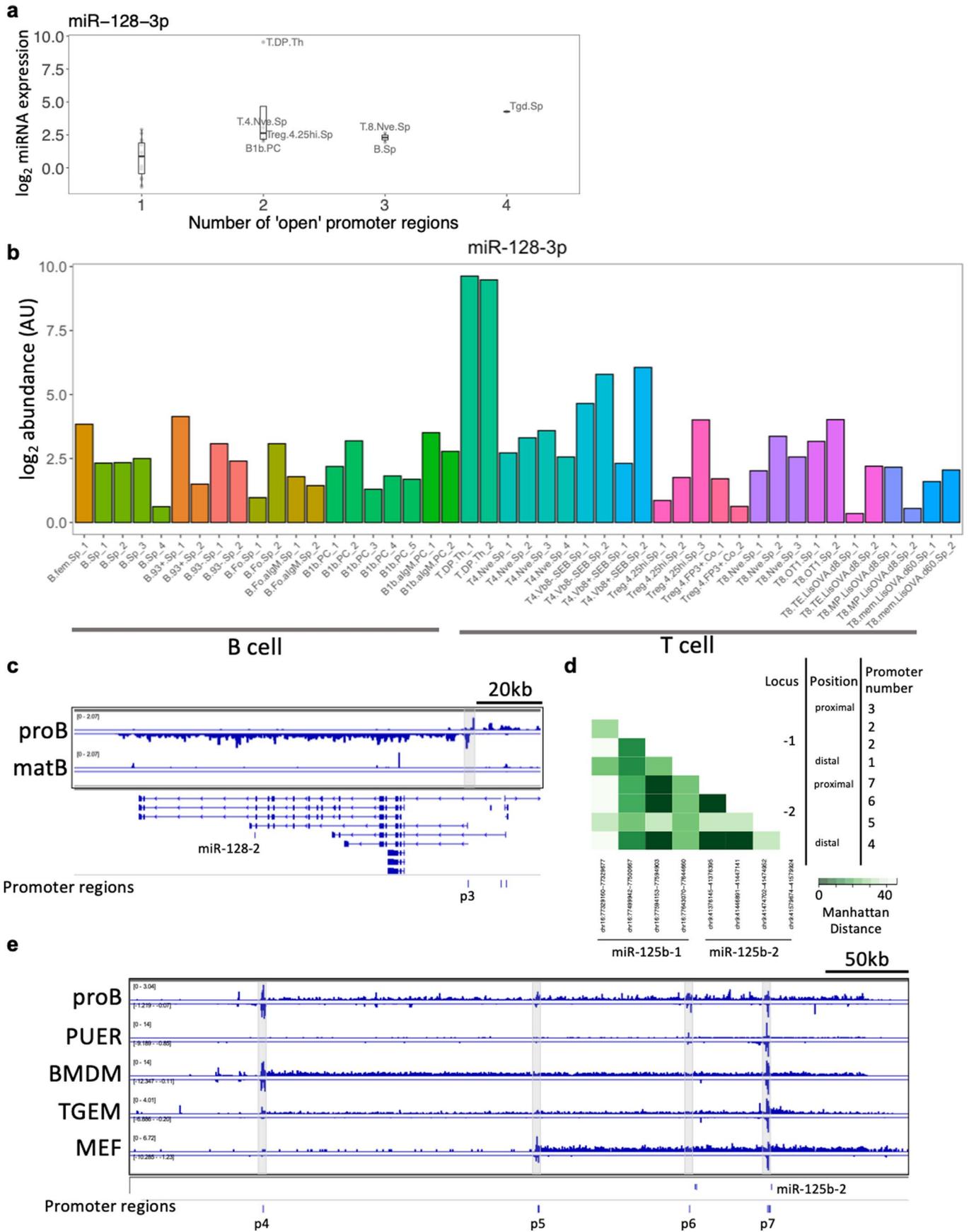
Extended Data Fig. 5 | Histone marks and nascent RNA support promoter additivity. **a**, Fraction of mature miRNAs (includes duplicated) with a given number of promoter regions, colored by TargetScan conservation categories. **b**, miRNA \log_2 expression compared to its number of 'active' promoter regions across the genome using the 6 aligned populations with chromatin mark, nascent RNA, accessibility, and miRNA expression measurements. Active promoters are defined as accessible by ATAC-seq, presence of H3K4me3, and a nascent transcript detected initiating from the promoter region and spanning the miRNA. ($n = 6$ populations) **(c)** Aligned dataset read pile-ups and *de novo* nascent transcript calls at the miR-29a/b-1 locus in BMDMs and MEFs illustrating multiple promoter use. All tracks are normalized independently. **d**, Number of active promoters for each expressed miRNA across the 6 aligned populations with or without histone mark and nascent RNA criteria.



Extended Data Fig. 6 | See next page for caption.

Extended Data Fig. 6 | Step-wise regression at multi-promoter loci. **a**, Individual promoter region associations to miRNA expression at multi-promoter loci. Each point represents the strength and direction of association from a promoter region accessibility to expression step-wise regression at a multi-promoter miRNA loci, plotted against the miRNA to promoter region distance relative to the most distal promoter. Gray dotted lines indicate p value of 0.05. Only associations with $p < 0.1$ are labeled with text. **b**, Stepwise regression associations for each multi-promoter miRNA with a significant association. Each locus is labeled with the miRNA, coordinates, and host-gene if available. Arrows indicate the most distal position for each promoter region in the locus. Boxes indicate an annotated host-transcript isoform TSS. Height of bars over promoter regions represents the signed $-\log_{10}$ p value in the stepwise regression using promoter accessibility as a predictor of miRNA expression.

Extended Data Fig. 7 | Promoter accessibility in multi-copy miRNAs. a, ATAC-seq, GRO-seq, and H3K4me3 ChIP-seq data at miR-199 loci in FRC.SLN cells and MEFs. **b,** Heatmap of OCR accessibility at TSS/promoter regions for 14 select duplicated miRNAs with promoter annotations at both loci. Clear boxes represent OCRs not detected above background. For each locus, if the miRNA is on the positive strand the promoters are ordered from furthest to closest going left to right. The opposite is true for miRNAs on the negative strand.



Extended Data Fig. 8 | See next page for caption.

Extended Data Fig. 8 | miR-128 and miR-125b promoter activities. **a**, Number of open merged promoter regions compared to \log_2 AU miRNA expression for miR-128-3p across 22 overlapping miRNA and ATAC-seq samples. ($n = 22$ populations) **(b)** ImmGen miRNA Browser view of miR-128-3p expression across B and T cells. **c**, GRO-seq read pile-ups in pro-B cells and mature B cells at the miR-128-2 locus. Active promoter in progenitor cells highlighted in gray. **d**, Heatmap of pairwise Manhattan distance values between promoter regions of miR-125b-1 and miR-125b-2. Promoter numbers correspond to Fig. 3e,f. **e**, GRO-seq read pile-ups normalized within each row across selected cell types at the miR-125b-2 locus. Promoter regions highlighted in gray.

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see [Authors & Referees](#) and the [Editorial Policy Checklist](#).

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a Confirmed

- The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement
- A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
- The statistical test(s) used AND whether they are one- or two-sided
Only common tests should be described solely by name; describe more complex techniques in the Methods section.
- A description of all covariates tested
- A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
- A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
- For null hypothesis testing, the test statistic (e.g. F , t , r) with confidence intervals, effect sizes, degrees of freedom and P value noted
Give P values as exact values whenever suitable.
- For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
- For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
- Estimates of effect sizes (e.g. Cohen's d , Pearson's r), indicating how they were calculated

Our web collection on [statistics for biologists](#) contains articles on many of the points above.

Software and code

Policy information about [availability of computer code](#)

Data collection

qPCR data was collected by Qiagen and processed using custom code in the R computing language as outlined in the methods. Code used for normalization and batch correction of miRNA qPCR data can be found at <https://github.com/srose89/ImmGen-miRNA>.

Data analysis

Flow cytometry data was analyzed in FlowJo v6 or Cytobank. Statistical analyses and processing of qPCR data were performed in the R (version 3.6.1) computing language using custom code or packages including: limma (v3.34.9), mice (v2.46.0), MASS (v7.3-51.4), BEDTools (v2.25.0), and Biostrings (v2.52.0). Alignment of public nascent RNA and ChIP-seq data was done using Bowtie2 (v2.3.4.3) and nascent RNA data was also pre-processed with Cutadapt (v2.10). Signal quantification and peak calling for public ChIP-seq data, as well as de novo transcript definition for nascent RNA data, was done using Homer (v4.8). Aligned read pile-ups were visualized with IGV (v2.4.1). The UCSC batch conversion tool (<https://genome.ucsc.edu/cgi-bin/hgLiftOver>) was used for genome coordinate lift over to mm10.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research [guidelines for submitting code & software](#) for further information.

Data

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

External mRNA-seq data was downloaded through the ImmGen website. ATAC-seq data was collected through the ImmGen website. ChIP-seq and GRO/PRO-seq datasets were downloaded through GEO from studies cited in the text. miRNA promoter annotations were downloaded from supplement of their respective studies cited in the text or the GENCODE database (<https://www.genencodegenes.org/mouse/>). TAD boundary data was downloaded from Johanson et al as cited in the text. Conservation and other information was downloaded from TargetScan V7 (www.targetscan.org). Problematic sequence regions on the ENCODE blacklist were downloaded from <https://sites.google.com/site/anshulkundaje/projects/blacklists>. rRNA sequence and sca/snoRNA loci were retrieved from igenomes (<http://>

igenomes.illumina.com.s3-website-us-east-1.amazonaws.com) and RFAM v14.2, respectively. CAGE peaks from the FANTOM5 consortium were downloaded from their website (http://fantom.gsc.riken.jp/5/datafiles/latest/extra/CAGE_peaks/). phastCons conservation scores were downloaded from <http://hgdownload.cse.ucsc.edu/goldenpath/mm10/phastCons60way>.

Data that support the findings of this study are available on the ImmGen website (www.immgen.org), and raw and processed miRNA expression data including sample metadata are available at GEO accession GSE144081. Additionally, tables with ATAC-seq signal, p-values, and peak locations with previously blacklisted peaks included are provided (Supplementary Data 3). SRA ID and other sample information for downloaded ChIP-seq and nascent RNA datasets can be found in Supplementary Table 12. Processed histone mark and nascent RNA data is available in Supplementary Data 1 and 2, respectively. Source data for all figures in this manuscript are provided.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

Life sciences Behavioural & social sciences Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	Sample sizes for qPCR profiling were determined based on a maximization of breadth of cell types with adequate reproducibility. The sample sizes were sufficient to determine strong changes because of the reproducibility of biological replicates demonstrated in both high and low replicate cell types. Explicit sample sizes for qPCR profiling or molecular experiments are listed in the methods section. No sample size calculation was done, but based on prior ImmGen experience and reliable clustering of replicate samples across batches this sample size was deemed sufficient given the resources available. Sample sizes for miRNA sensor, RNP electroporation, and Luciferase experiments were determined by a minimization of the number of mice or cells required while still able to assess miRNA repression and promoter/enhancer effects through reproducible and consistent measurements across batches as seen in the data. No explicit sample size calculations were made prior to these experiments.
Data exclusions	Female samples and single replicate samples were excluded from many analyses because of a lack of reproducibility for these samples. Individual qPCR assay outlier measurements were identified either by visual inspection or based on a Z-score greater than 4 relative to all other samples. These measurements were set to missing and imputed using the mice package in R with default settings and 10 iterations.
Replication	miRNA sensor assays were used to validate abundance measurements by showing that miRNA abundance was highly correlated with suppression of a GFP sensor in vivo (n = 2 mice miR-24-3p; n = 4 mice miR-652-3p). Histone mark ChIP-seq (H3K4me1/3, H3K27ac) and nascent RNA sequencing (GRO/PRO-seq) were used to support predicted miRNA cis-regulatory elements. Luciferase (n = 3-4 transfections) and CRISPR mediated deletion experiments (n = 3 mice for each; 3 independent experiments for miR-21, 2 independent experiments for miR-146a) were also used to support the putative regulatory elements. Signal of each luciferase was the average of two technical replicates. Trends observed overall were consistent across experiments.
Randomization	Samples or miRNAs were split into groups based on literature annotations or sometimes conservation, respectively. Normalization and batch correction were done to control for array intensity differences, the batch correction was able to be done for a subset of miRNAs because of the same cell type being present in multiple batches of qPCR profiling.
Blinding	Blinding was not relevant to this study because data collection was focused on sorted populations of cells which required supervised analysis for linking immune cells across data modalities and analyzing differences between immune populations by lineage, cell type, or perturbation. Unbiased clustering was performed to ensure the groupings used were reflected in the miRNA expression data previous to analysis.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
<input type="checkbox"/>	<input checked="" type="checkbox"/> Antibodies
<input type="checkbox"/>	<input checked="" type="checkbox"/> Eukaryotic cell lines
<input checked="" type="checkbox"/>	<input type="checkbox"/> Palaeontology
<input type="checkbox"/>	<input checked="" type="checkbox"/> Animals and other organisms
<input checked="" type="checkbox"/>	<input type="checkbox"/> Human research participants
<input checked="" type="checkbox"/>	<input type="checkbox"/> Clinical data

Methods

n/a	Involved in the study
<input checked="" type="checkbox"/>	<input type="checkbox"/> ChIP-seq
<input type="checkbox"/>	<input checked="" type="checkbox"/> Flow cytometry
<input checked="" type="checkbox"/>	<input type="checkbox"/> MRI-based neuroimaging

Antibodies

Antibodies used	-PE-Cy7-conjugated anti-mouse CD19 (clone eBio1D3, 1:400), APC-conjugated anti-TCRb (clone H57-597, 1:200), PerCP-Cy5.5-conjugated anti-CD4 (clone RM4-5, 1:400), AlexaFluor 700-conjugated anti-Ly6G (clone RB6-8C5, 1:400), eFluor 450-conjugated anti-CD45.1 (clone A20, 1:200), APC-eFluor 780-conjugated anti-CD45.2 (clone 104, 1:200) (Thermo Fisher Scientific), BV510-conjugated anti-CD8a (clone 53-6.7, BioLegend, 1:400), and PE-conjugated anti-human CD271 (NGFR) (clone C40-1457, BD Bioscience, 1:200)
Validation	The antibodies used for sorting were validated part of the ImmGen SOP (see https://www.immgen.org/ImmGenProtocols.html for details).

Eukaryotic cell lines

Policy information about [cell lines](#)

Cell line source(s)	RAW 264.7 cells were a kind gift from the Brody Laboratory at the Icahn School of Medicine at Mount Sinai.
Authentication	No authentication of cell lines was performed.
Mycoplasma contamination	Raw macropahges tested negative for mycoplasma contamination before incubation in an all mycoplasma negative incubator.
Commonly misidentified lines (See ICLAC register)	No commonly misidentified cell lines were used.

Animals and other organisms

Policy information about [studies involving animals](#); [ARRIVE guidelines](#) recommended for reporting animal research

Laboratory animals	All cells used for miRNA analysis were obtained from male 6- to 8-week-old C57BL/6J mice from the Jackson Laboratory, used 6 to 8 days from reception. For miRNA sensor, 8-week-old CD45.1 C57BL/6J (B6.SJL-Ptprca Pepcb/BoyCrl) male mice were used as donors. Mice were housed in specific pathogen-free facilities.
Wild animals	This study did not involve wild animals.
Field-collected samples	This study did not involve samples collected from the field.
Ethics oversight	The use of all mice for these studies was in accordance with institutional guidelines with review and approval by the Institutional Animal Care and Use Committee of the Icahn School of Medicine at Mount Sinai.

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Flow Cytometry

Plots

Confirm that:

- The axis labels state the marker and fluorochrome used (e.g. CD4-FITC).
- The axis scales are clearly visible. Include numbers along axes only for bottom left plot of group (a 'group' is an analysis of identical markers).
- All plots are contour plots with outliers or pseudocolor plots.
- A numerical value for number of cells or percentage (with statistics) is provided.

Methodology

Sample preparation	Cells were purified according to the standardized ImmGen standard operations protocol using the lineage markers highlighted in Table 1. Peritoneal cell suspensions were obtained by lavage of the peritoneal cavity with 10 ml cold PBS. Single-cell suspensions were obtained by mechanical disruption of tissues followed by erythrocyte lysis using ACK buffer (Lonza). Where necessary, tissues were digested at 37°C for 15 minutes, in glucose-containing medium.
Instrument	BD FACSAria IIu cell sorter was used to sort miRNA qPCR samples at MSSM, and the BD FACS Fortessa was used for miRNA sensor analysis.
Software	BD FACSDiva was used to collect data and FlowJo v6 (TreeStar) was used for analysis.
Cell population abundance	Double sorted to achieve 100 percent purity according to ImmGen SOP.
Gating strategy	Gating strategy for populations is available on the ImmGen website (https://www.immgen.org/Databrowser19/DatabrowserPage.html) by clicking on respective populations within the gene skyline browsers (i.e. http://rstats.immgen.org/)

Skyline/resources/Sorting_PDFs_ATAC/T_4_Nve_Sp.pdf). Representative sort reports for select myeloid populations are provided in the supplementary information. Additionally, cells were gated for live/dead in Fig. S1d,e.

Tick this box to confirm that a figure exemplifying the gating strategy is provided in the Supplementary Information.