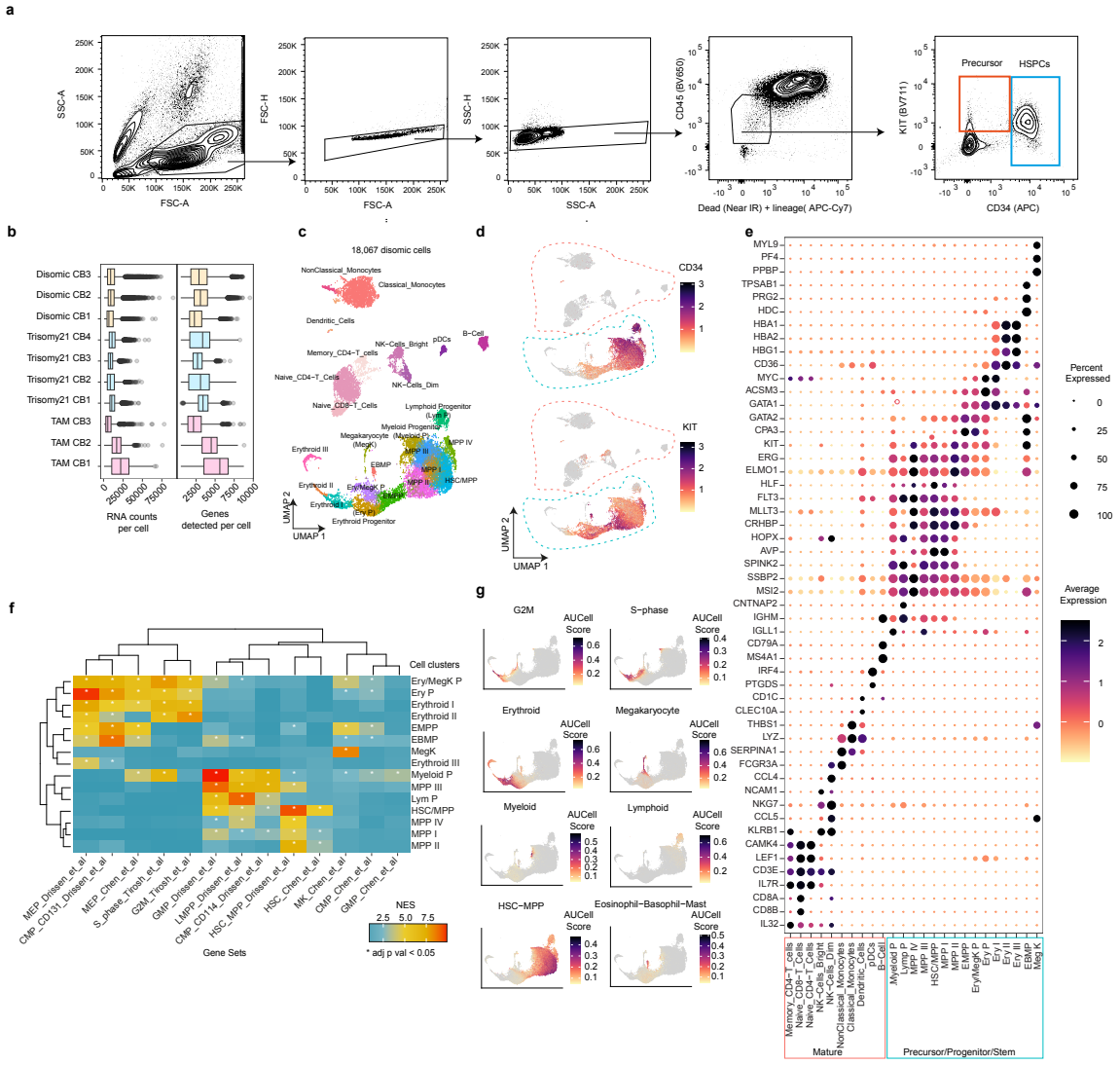
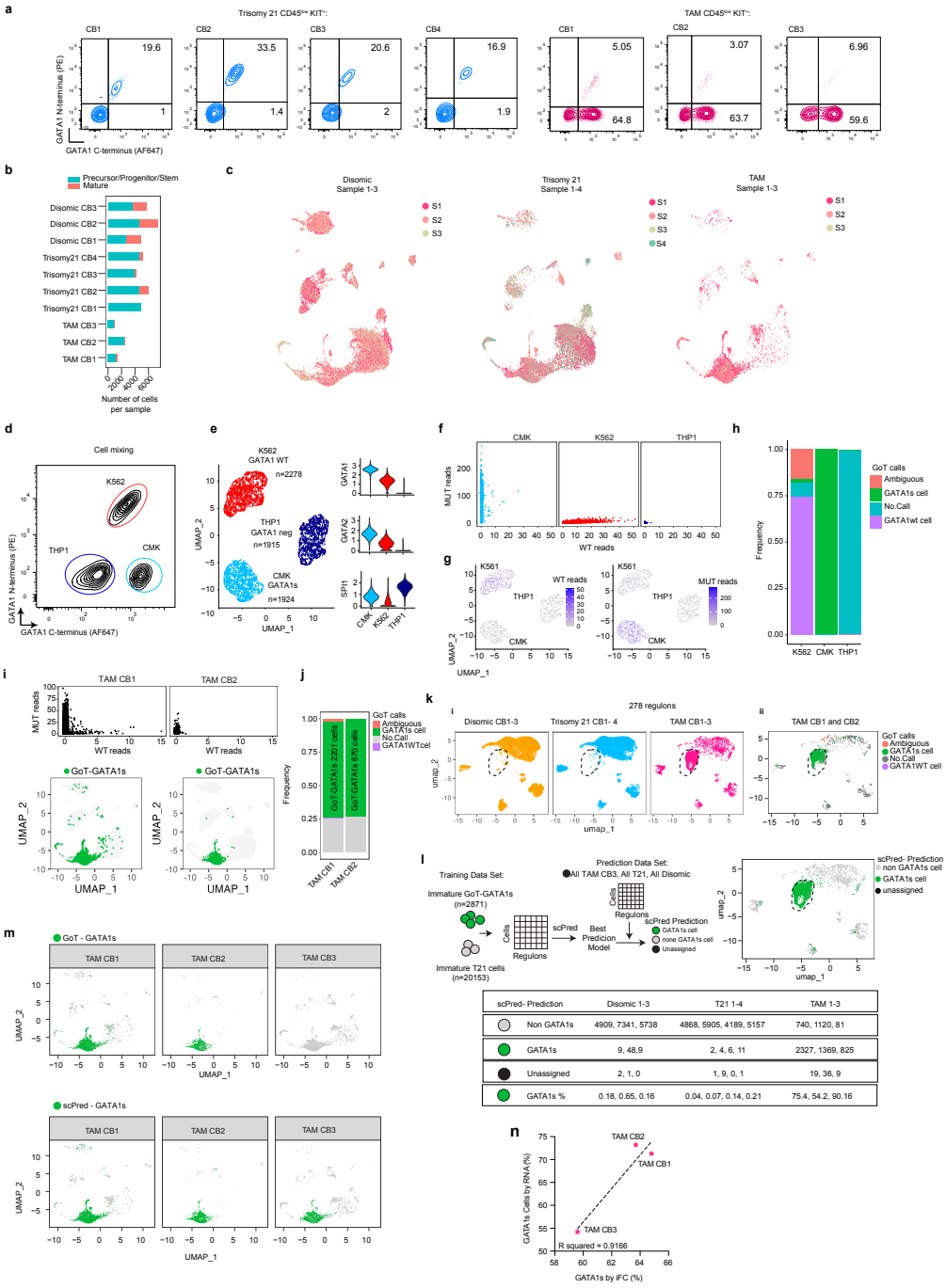
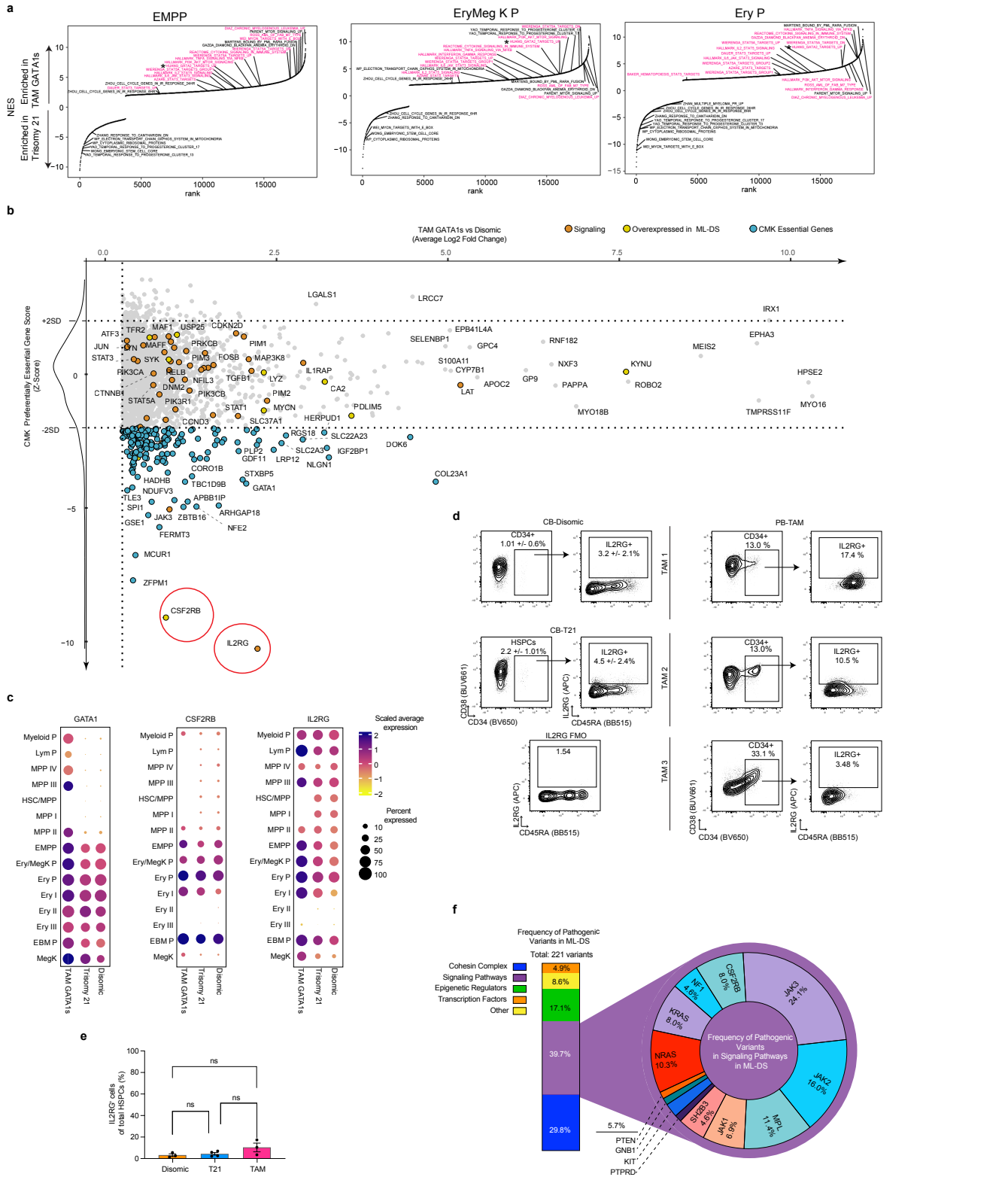


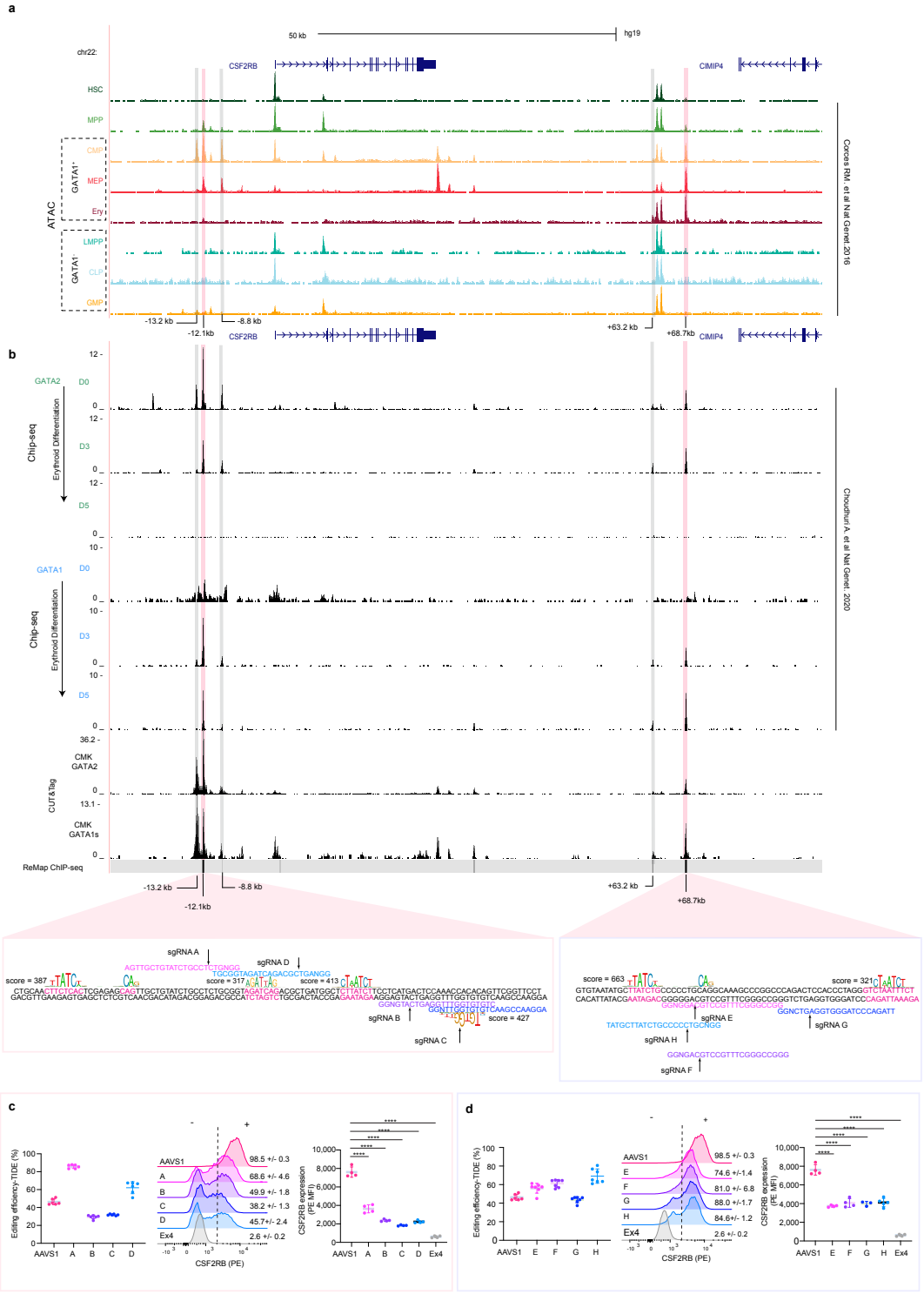
Extended Fig 1. scRNA-seq quality metrics and cluster annotation. a, FACS strategy for scRNA-seq experiments. Sorting gates enrich for CD34^{+} HSPCs and KIT^{+} CD34^{-} precursors. **b**, RNA and gene counts detected per cell in each sample after quality control. **c**, Transcriptional reference UMAP of disomic cells colored by cluster annotation. **d**, UMAP feature plots of *CD34* and *KIT* expression in disomic cells. Red dotted outline denotes mature cells; cyan outline denotes precursor/progenitor/stem cells. **e**, Heatmap of scaled mean expression of lineage-associated genes for each disomic cluster. **f**, Heatmap showing positive GSEA enrichment scores of hematopoietic lineages signatures[1] in disomic cluster to facilitate cluster annotation. White asterisks mark Benjamini-Hochberg (BH) adjusted p value (adj. p) < 0.05. **g**, UMAP feature plots of AUCell scores for cell cycle gene sets (G2M, S-phase) or hematopoietic lineage associated signatures[1].



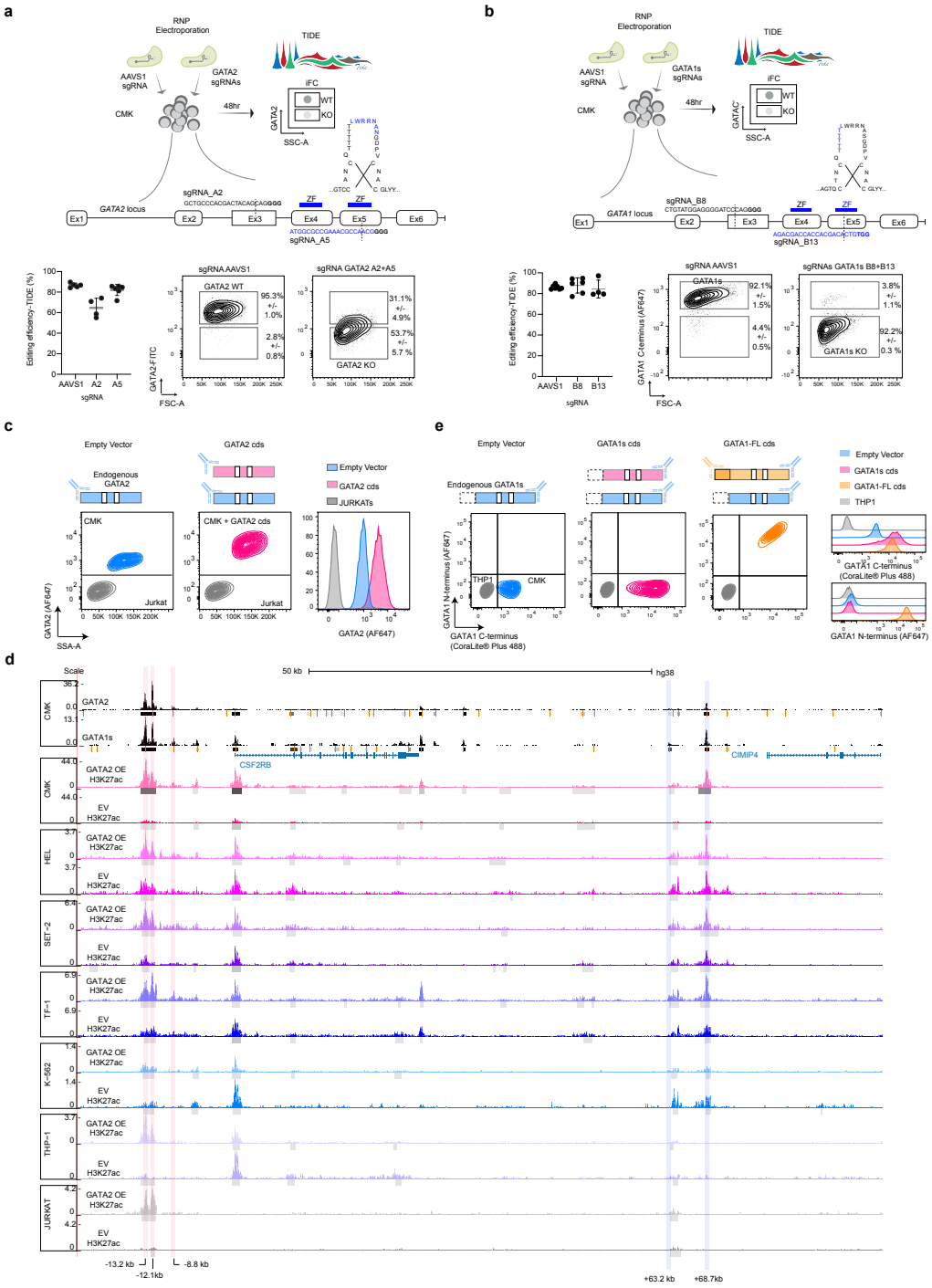
Extended Fig 2. Transcriptional identification of GATA1s-mutant cells. a, Intracellular flow cytometry (iFC) with antibodies against either GATA1 N-terminus or GATA1 C-terminus on live, CD45^{low} KIT^{+}, T21 (blue) or TAM (pink) samples to distinguish GATA1 WT from GATA1s-mutant cells. scRNA-seq was performed on these samples. **b**, Number of ells passing quality control, for each sample. **c**, Integration of disomic (left), T21 (middle) and TAM samples (right). **d**, iFC as in **a** in a mixture of GATA1 WT (K562), GATA1 negative (THP1) or GATA1s-mutant (CMK) cells. **e**, Left, UMAP of the same cell lines mixed at equal ratios. Right, violin plots of *GATA1*, *GATA2* and *SPI* expression. **f**, GATA1s-mutants vs GATA1 WT reads counts quantified by the IronThrone-GoT pipeline (GoT) [2] in each cell line. **g**, UMAP of the cell line mixture colored by GATA1 WT or GATA1s-mutant reads counts. **h**, Frequency of cells classified as either GATA1s-mutant, GATA1 WT, or ambiguous in each cell line. **i**, Top, number of GATA1s-mutants reads and GATA1 WT reads quantified by the GoT in two TAM samples. Bottom, UMAP of TAM cells classified as GATA1s-mutant by GoT (green dots) overlaid on combined disomic and T21 single cells (grey dots). **j**. Frequency of cells classified by GoT as either GATA1s-mutant, GATA1 WT, or ambiguous for each TAM sample. **k**, (i-ii) UMAP of 278 regulons integrating all disomic, trisomy 21 or TAM samples (i); UMAP of 278 regulons depicting only two TAM samples colored by GoT calls (ii). **l**, Left, schematic representation of GATA1s-mutant cell imputation by scPred[3]. Right, UMAP of 278 regulons depicting all disomic, T21 or TAM single cells colored by scPred prediction. Bottom, cell count and frequency of scPred prediction for each donor in each sample type. **m**, UMAP of each TAM sample depicting GATA1s-mutant cells defined by GoT (top) or by scPred (bottom). **n**, Scatter plot of percent of transcriptionally defined GATA1s cells vs percent of GATA1s identified by iFC for each TAM sample.



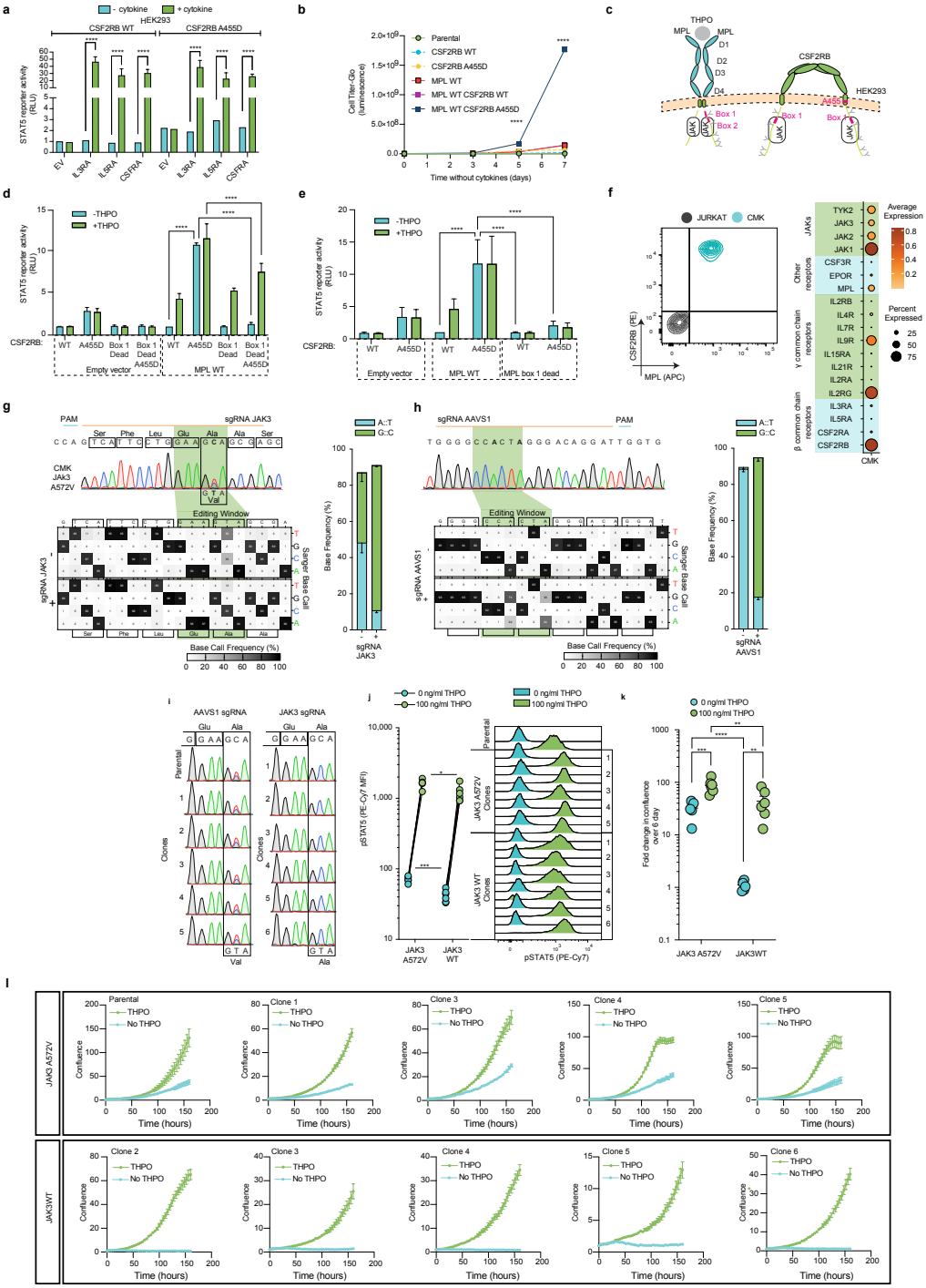
Extended Fig 3. GSEA and differential gene expression. a, GSEA of 5028 ranked significantly (BH adj. p < 0.01) altered signal transduction gene sets in TAM GATA1s-mutant vs to T21 cells. Pink marks leukemic and JAK-STAT related pathways; * marks GATA2 target gene sets. **b**, Intersection of preferentially essential gene scores in CMK cells (DepMap ID: ACH-001036) [4] and upregulated genes in TAM GATA1s-mutant cells compared to disomic cells. Differential gene expression was performed with MAST package[5] in the combined EMPP, Ery/Meg P or Ery P cell clusters. **c**, *GATA1*, *CSF2RB* and *IL2RG* scaled average expression in each cell cluster across sample types. **d**, Flow cytometry analysis of IL2RG expression within the CD34+ compartment in umbilical cord blood (ucb) derived disomic (n=3), ucb T21(n=4) and peripheral blood derived TAM samples (n=3). **e**, Percent of cells expressing IL2RG in the same samples as in **d**; one-way ANOVA with Tukey's correction. **f**, Frequency of pathogenic variants in ML-DS[6]. Data represent mean ± SEM unless otherwise stated. 0.1234 (ns), 0.0332(*),0.0021 (**), 0.0002 (***), < 0.0001(****).



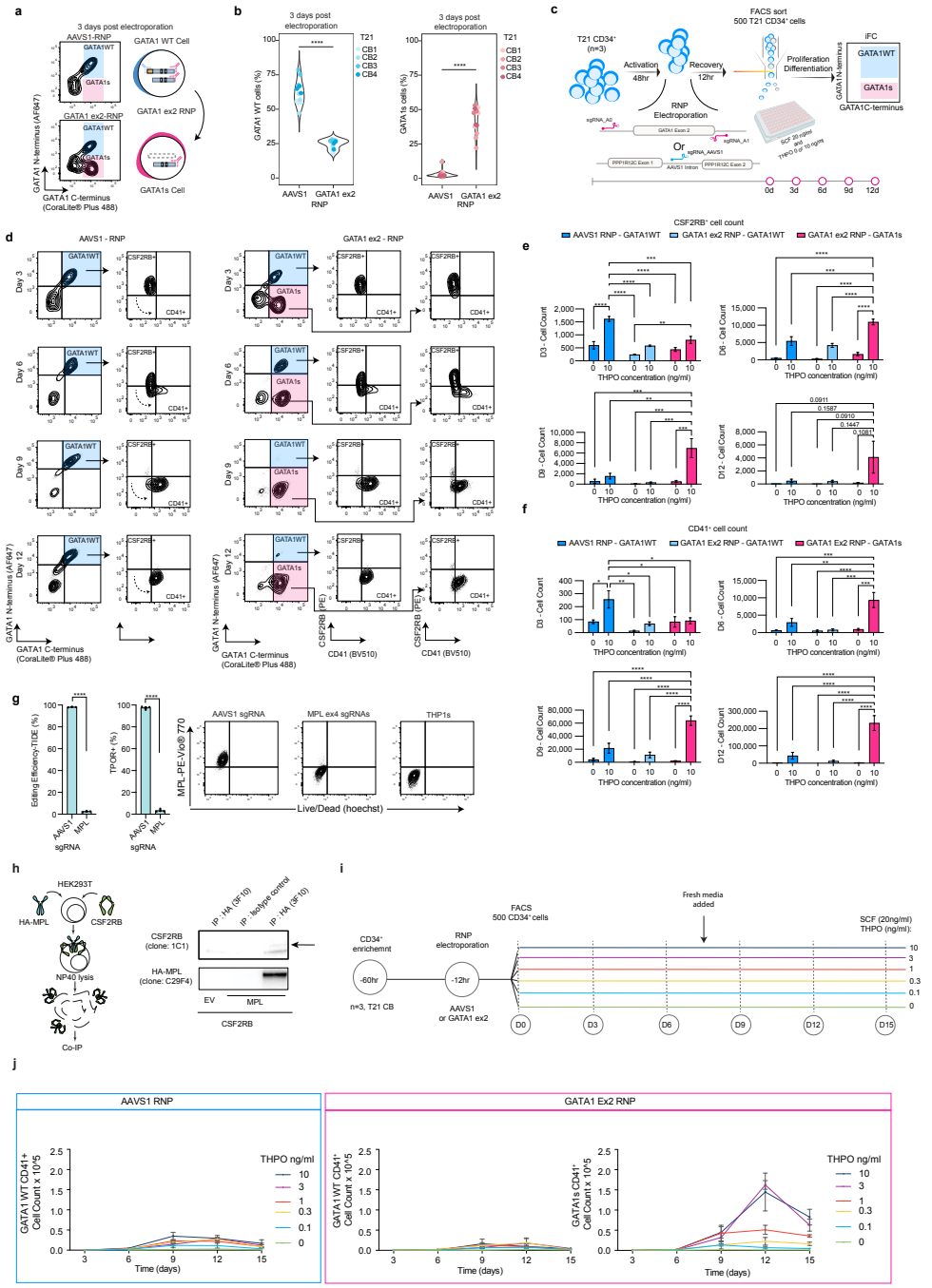
Extended Fig 4. GATA2 and GATA1 biding at the CSF2RB locus. a, UCSC genome browser tracks of the CSF2RB locus depicting publicly available chromatin accessibility in different hematopoietic stem and progenitor cell compartments[7]. **b**, GATA2 and GATA1 binding measured by Chip-Seq at day 0 (D0), day 3 (D3), and day 5 (D5) in *in-vitro* erythroid differentiation of human CD34+[8]; GATA2 (n=2) and GATA1s binding (n=2) in CMK cells measured by CUT&Tag; and GATA2/GATA1 binding from the ReMap ENCODE data[9]. The location of highest ReMap read density are marked relative to the CSF2RB promoter. For the -12.1 kb and +68.7kb cis elements, the GATA, GATA/E-box and Runx1 motifs and their corresponding JASPAR probability scores[10] are shown. The sgRNAs sequences targeting the motifs are shown and the arrows depict the cutting site of spCas9. **c**, Left, gene editing efficiency of each sgRNA as measured by sanger sequencing of PCR amplicons 24 hours post electroporation of the RNP complex targeting the -12.1kb cis element or the *AAVS1*control locus in CMK cells. Middle, representative flow cytometry histograms of CSF2RB expression 24 hours post electroporation of Cas9 and each sgRNA. Dotted line marks positive and negative CSF2RB expression. Right, differential CSF2RB protein expression by flow cytometry (Mean Fluorescence Intensity, MFI). **d**, Like **c** but for the +68.7kb cis element in CMK cells. One-way ANOVA with Dunnett's correction (**c**,**d**). Data represent mean ± SEM unless otherwise stated. 0.1234 (ns), 0.0332(*),0.0021 (**), 0.0002 (***), < 0.0001(****).



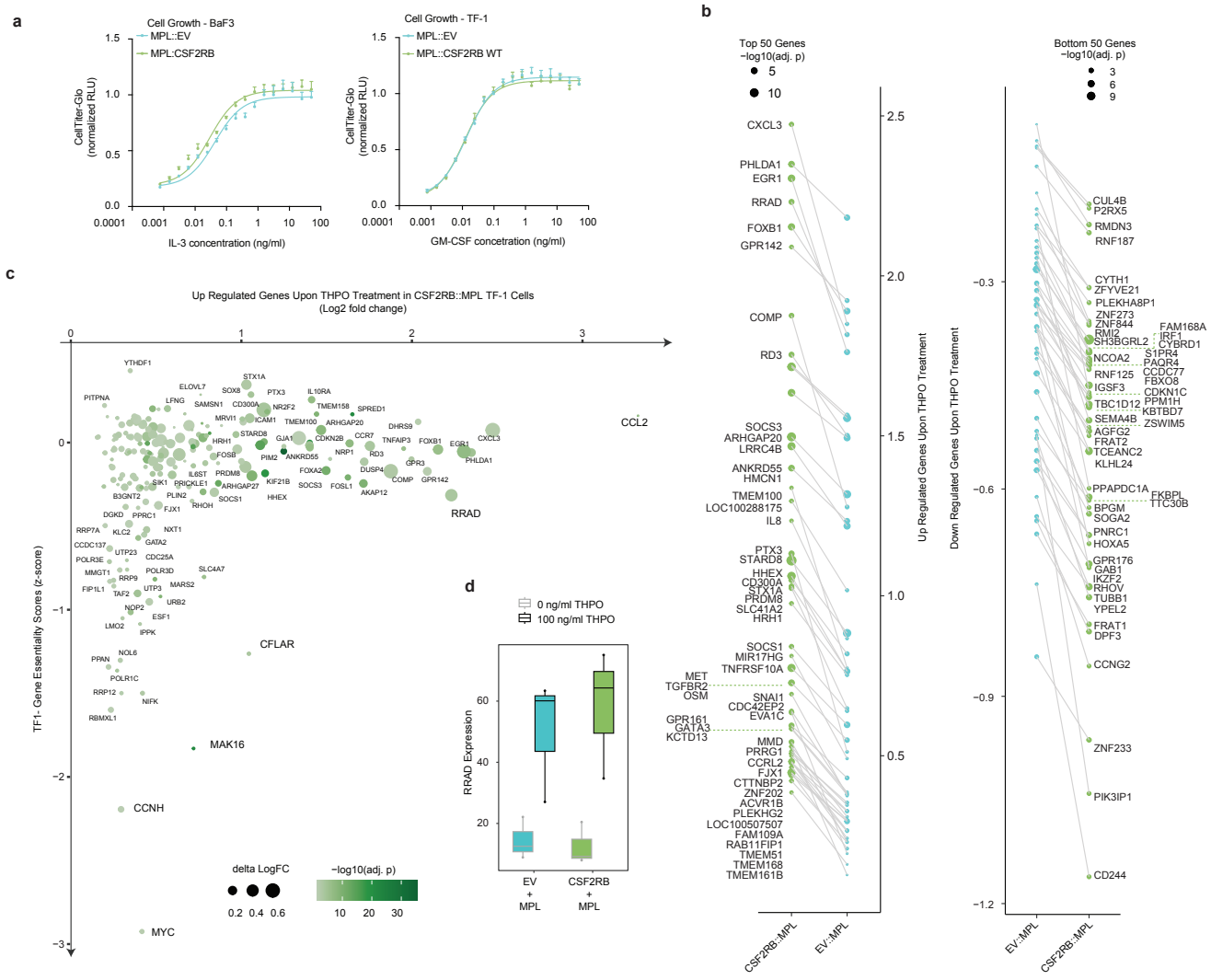
Extended Fig 5. Generation of GATA2 and GATA1s knockout, lentivirus and antibody validation and H3K27ac profiling. **a**, Schematic representation of GATA2 knockout in CMK cells. Two sgRNAs were design targeting either exon 3 or exon 5 of *GATA2* with equal efficiencies as measure by TIDE analysis[11] 48 hours post electroporation. iFC was performed 48 hours after simultaneous electroporation of both sgRNAs to measure loss of GATA2 expression. Gates mark GATA2 WT or GATA2 knockout cells (GATA2 KO). **b**, Schematic representation of GATA1s depletion in CMK cells. Like **a**, iFC was performed 48 hours post simultaneous electroporation of both sgRNAs to measure GATA1s depletion. RNP, ribonucleoprotein; ZF, Zinc Finger. **c**, GATA2 antibody and lenti vector validation. iFC was performed 48 hours post lenti delivery of the empty vector (EV), or the GATA2 coding sequence (cds) in CMK cells. Jurkat cells were used as GATA2 negative controls. **d**, CUT&TAG read density of GATA2 and GATA1s binding in CMK cells; and H3K27ac in 7 cell lines 48 hours post lenti delivery of the empty vector (EV, dark color), or the GATA2 cds (GATA2 OE, light color). **e**, GATA1 antibody and lenti vector validation. iFC was performed 48 hours post lenti delivery of the empty vector (EV), GATA1s, or GATA1full-length (GATA1-FL) cds in CMK cells. THP1 cells were used as GATA1 negative control. For **e**, cells were stained with antibodies against GATA1 C-terminus specific for GATA1s and GATA1-FL or GATA1 N-terminus specific for GATA1-FL only.



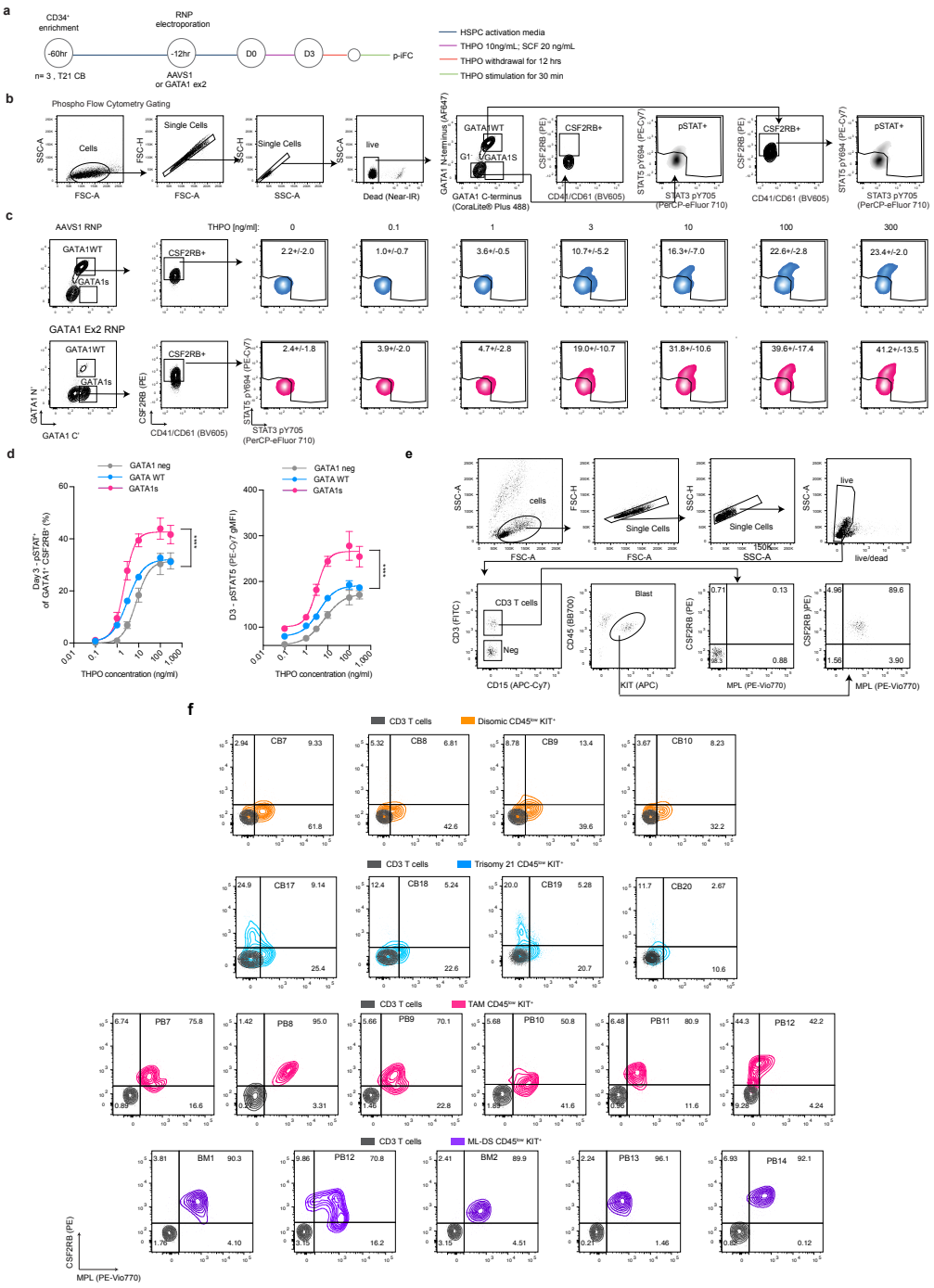
Extended Fig 6. Cellular and biochemical profiling of CSF2RB A455D-MPL axis and JAK3 A572V reversion. a, STAT5 reporter activity of CSF2RB WT or the A455D variant in the presence of canonical α-chains (IL3RA, IL5RA, CSFRA) treated with their corresponding cytokine (mean ± SD, n=4). **b**, Cell proliferation of Ba/F3 cells expressing CSF2RB WT or A455D with or without MPL in the absence of cytokines (mean ± SD, n=3). **c**, Left, schematic of MPL homodimers bound by THPO depicting 4 extracellular domains (D1-D4). Right, schematic of CSF2RB homodimer. For both, MPL and CSF2RB box1 or box2 bound to JAK proteins are shown. **d**, STAT5 reporter activity of CSF2RB WT and A455D, with or without box1 dead co-expressed with empty vector control or MPL with or without THPO. **e**, STAT5 reporter activity of CSF2RB WT or A455D co-expressed with empty vector control or MPL WT or box1 dead with or without THPO. **f**. Left, CSF2RB and MPL expression in CMK cells. Right, gene expression of receptors and tyrosine kinases in CMK cells. **g**-**h**, Adenine base editing[12] of the *JAK3 A572V* variant (**g**) or *AAVS1* locus (**h**) in CMK cells; base editing efficiency as quantified by EditR[13]. **i**, Sanger sequencing at the JAK3 A572V locus in the parental cell line and 5 mutant clones (AAVS1 sgRNA) and 6 WT clones (JAK3 sgRNA). **j**, Intracellular phospho flow cytometry of pSTAT5 in response to THPO (Two-tailed T-test). **k**, Fold change in confluence over a 6-day period of all clones with or without THPO (10ng/ml). **l**, Cell proliferation curves with or without THPO of JAK3 A572V clones (top) or A572 WT clones (bottom). Two-way ANOVA with Holm-Sidak's correction (**a**,**b**,**d**,**e**) or Tukey's correction (**k**). Data represent mean ± SEM unless otherwise stated. 0.1234 (ns), 0.0332(*),0.0021 (**), 0.0002 (***), < 0.0001(****).



Extended Fig 7. Cell proliferation in response to THPO. a, Schematic strategy to engineered and identify GATA1s-mutant cells. GATA1 WT cells express two isoforms of GATA1, and they can be identified by intracellular flow cytometry (iFC) using two antibodies. One antibody against the C-terminus (pink) marks both isoforms while a second antibody against the N-terminus (blue) marks only GATA1-FL[14]. Upon GATA1-FL isoform depletion with CRISPR Cas9 disruption of *GATA1* exon 2 (GATA1 Ex2 RNP), GATA1s-mutant cells are solely marked by the antibody against the C-terminus (pink). **b**, Percent of GATA1 WT or GATA1s-mitant cells defined iFC 3 days post electroporation. (n=4 T21 donors, 4 replicates each; two-tailed T-test). **c**, Schematic to identify GATA1s cell proliferation in response to THPO (n=3 T21 donors). **d**, Per day iFC gating to identify GATA1 WT or GATA1s cells expressing CSF2RB and CD41. **e-f**, Absolute counts of CSF2RB$^+$ cells (**e**) or CD41$^+$ cells (**f**) as determined by CountBright™ Absolute Counting Beads and flow cytometry; two-way ANOVA with Tukey's correction. **g**, CRISPR-Cas9 gene targeting of *MPL* exon 4 to generate MPL knockout CMK cells by Sanger sequencing or flow cytometry (n=3, two-tailed T-test). **h**, Protein co-immunoprecipitation of hemagglutinin (HA) tagged MPL with CSF2RB in HEK293T cells. **i**, Experimental schematic to CSF2RB+ cell proliferation in response to increasing concentrations of THPO. **j**. Proliferation of GATA1 WT or GATA1s CD41+ cells in response to increasing concentration of THPO. Data represent mean ± SEM unless otherwise stated. 0.1234 (ns), 0.0332(*),0.0021 (**), 0.0002 (***), < 0.0001(****).



Extended Fig 8. Transcriptional response of THPO in TF-1 cells co-expressing CSF2RB and MPL. a, Cell proliferation curves of BaF3 or TF-1 cells co expressing MPL and CSF2RB (green) or MPL and empty vector (EV, cyan) treated with increasing concentrations of IL-3 or GM-CSF, respectively. **b**, Top 50 differentially expressed genes upon THPO treatment in TF-1 cells co-expressing CSF2RB and MPL (CSF2RB::MPL, n=3) or empty vector (EV) and MPL (EV::MPL, n=3). Y-axis represents the log2 fold change in gene expression as calculated by the DESeq2[15]. The size of the dot represents the -log10 of the BH adjusted p-value (adj. p). **c**, Intersection of gene essentiality scores (y-axis, z-score) in TF-1 cells (DepMap ID: ACH-000387) [4] and upregulated genes (x-axis) upon THPO (100ng/ml) treatment in TF-1 cells co-expressing MPL and CSF2RB. The size of the dot represents the difference in log2 fold change in TF-1 cells co-expressing CSF2RB and MPL vs empty vector and MPL. **d**, Low quadrant, gene expression of *RRAD*, a Ras-related GTPase, in the TF-1 cell models upon THPO treatment.



Extended Fig 9. Phospho-flow cytometry of GATA1s cells and persistence of CSF2RB$^+$ MPL$^+$ blast cells in ML-DS. a, Experimental design to assess pSTAT in response to THPO. T21 CD34$^+$ cells were gene-edited and cultured for 3 days in SCF + THPO. Cells were rested for 12 hours in SCF alone and stimulated with increasing THPO concentrations for 30 minutes. Intracellular phospho-flow cytometry (phospho-iFC) was performed to detect GATA1, CSF2RB, CD41 and pSTAT1(data not shown), pSTAT3, and pSTAT5 expression. **b**, Representative phospho-iFC gating to identify of pSTAT3 (Y705) and pSTAT5 (Y694) (pSTAT+) in GATA1 negative cells (G1-), GATA1 WT CSF2RB$^+$, or GATA1s-mutant CSF2RB$^+$ cells. **c**, Representative phospho-iFC plots of pSTAT3 (Y705) and pSTAT5 (Y694) (pSTAT+) in GATA1 WT CSF2RB$^+$ cells (blue) or GATA1s-mutant CSF2RB$^+$ cells (pink) treated with increasing doses of THPO at Day 3. **d**, Left, increased efficacy of THPO in GATA1s-mutant CSF2RB$^+$ cells (pink curve) compared to GATA1 WT CSF2RB$^+$ cells (blue curve) or GATA1 negative cells (grey curve) at day 3. Y-axis represents the percent of pSTAT3 and pSTAT5 cells within the CSF2RB$^+$ compartment as gated in **c**. Right, geometric mean fluorescent intensity (gMFI) of pSTAT5 at day 3 in the same samples as in **c**. Sigmoid curves were fitted using a four-parameter logistic regression. Error bars represent ± SEM of 3 T21 donors; statistical test was performed with the extra sum of squares F-test. **e**, Flow cytometry gating to assess CSF2RB and MPL expression in CD3$^+$ T cells and CD45^{low} KIT$^+$ blast cells in disomic, T21 without TAM, TAM and ML-DS samples. **f**, flow cytometry plots of CSF2RB and MPL expression in each disomic cord blood (n=4), T21 cord blood (n=4), TAM peripheral blood (n=6), ML-DS peripheral blood (PB) or bone marrow (BM) (n=5). CD3$^+$ T cells from each sample are mark in gray.

REFERENCE

1. Roy, A., Wang, G., Iskander, D., O'Byrne, S., Elliott, N., O'Sullivan, J., Buck, G., Heuston, E.F., Wen, W.X., Meira, A.R., et al. (2021). Transitions in lineage specification and gene regulatory networks in hematopoietic stem/progenitor cells over human development. Cell Rep *36*, 109698. 10.1016/j.celrep.2021.109698.
2. Nam, A.S., Kim, K.T., Chaligne, R., Izzo, F., Ang, C., Taylor, J., Myers, R.M., Abu-Zeinah, G., Brand, R., Omans, N.D., et al. (2019). Somatic mutations and cell identity linked by Genotyping of Transcriptomes. Nature *571*, 355-360. 10.1038/s41586-019-1367-0.
3. Alquicira-Hernandez, J., Sathe, A., Ji, H.P., Nguyen, Q., and Powell, J.E. (2019). scPred: accurate supervised method for cell-type classification from single-cell RNA-seq data. Genome Biology *20*, 264. 10.1186/s13059-019-1862-5.
4. Tsherniak, A., Vazquez, F., Montgomery, P.G., Weir, B.A., Kryukov, G., Cowley, G.S., Gill, S., Harrington, W.F., Pantel, S., Krill-Burger, J.M., et al. (2017). Defining a Cancer Dependency Map. Cell *170*, 564-576.e516. 10.1016/j.cell.2017.06.010.
5. Finak, G., McDavid, A., Yajima, M., Deng, J., Gersuk, V., Shalek, A.K., Slichter, C.K., Miller, H.W., McElrath, M.J., Prlic, M., et al. (2015). MAST: a flexible statistical framework for assessing transcriptional changes and characterizing heterogeneity in single-cell RNA sequencing data. Genome Biology *16*, 278. 10.1186/s13059-015-0844-5.
6. Labuhn, M., Perkins, K., Matzk, S., Varghese, L., Garnett, C., Papaemmanuil, E., Metzner, M., Kennedy, A., Amstislavskiy, V., Risch, T., et al. (2019). Mechanisms of Progression of Myeloid Preleukemia to Transformed Myeloid Leukemia in Children with Down Syndrome. Cancer Cell *36*, 123-138.e110. 10.1016/j.ccell.2019.06.007.
7. Corces, M.R., Buenrostro, J.D., Wu, B., Greenside, P.G., Chan, S.M., Koenig, J.L., Snyder, M.P., Pritchard, J.K., Kundaje, A., Greenleaf, W.J., et al. (2016). Lineage-specific and single-cell chromatin accessibility charts human hematopoiesis and leukemia evolution. Nat Genet *48*, 1193-1203. 10.1038/ng.3646.
8. Choudhuri, A., Trompouki, E., Abraham, B.J., Colli, L.M., Kock, K.H., Mallard, W., Yang, M.L., Vinjamur, D.S., Ghamari, A., Sporrij, A., et al. (2020). Common variants in signaling transcription-factor-binding sites drive phenotypic variability in red blood cell traits. Nat Genet *52*, 1333-1345. 10.1038/s41588-020-00738-2.
9. Hammal, F., de Langen, P., Bergon, A., Lopez, F., and Ballester, B. (2021). ReMap 2022: a database of Human, Mouse, Drosophila and Arabidopsis regulatory regions from an integrative analysis of DNA-binding sequencing experiments. Nucleic Acids Research *50*, D316-D325. 10.1093/nar/gkab996.
10. Rauluseviciute, I., Riudavets-Puig, R., Blanc-Mathieu, R., Castro-Mondragon, J.A., Ferenc, K., Kumar, V., Lemma, R.B., Lucas, J., Chèneby, J., Baranasic, D., et al. (2024). JASPAR 2024: 20th anniversary of the open-access database of transcription factor binding profiles. Nucleic Acids Res *52*, D174-d182. 10.1093/nar/gkad1059.
11. Brinkman, E.K., Chen, T., Amendola, M., and van Steensel, B. (2014). Easy quantitative assessment of genome editing by sequence trace decomposition. Nucleic Acids Research *42*, e168-e168. 10.1093/nar/gku936.
12. Richter, M.F., Zhao, K.T., Eton, E., Lapinaite, A., Newby, G.A., Thuronyi, B.W., Wilson, C., Koblan, L.W., Zeng, J., Bauer, D.E., et al. (2020). Phage-assisted evolution of an adenine base editor with improved Cas domain compatibility and activity. Nat Biotechnol *38*, 883-891. 10.1038/s41587-020-0453-z.
13. Kluesner, M.G., Nedveck, D.A., Lahr, W.S., Garbe, J.R., Abrahante, J.E., Webber, B.R., and Moriarity, B.S. (2018). EditR: A Method to Quantify Base Editing from Sanger Sequencing. Crispr j *1*, 239-250. 10.1089/crispr.2018.0014.
14. Cruz Hernandez, D., Metzner, M., de Groot, A.P., Usukhbayar, B., Elliott, N., Roberts, I., and Vyas, P. (2020). Sensitive, rapid diagnostic test for transient abnormal myelopoiesis and myeloid leukemia of Down syndrome. Blood *136*, 1460-1465. 10.1182/blood.2020005610.
15. Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol *15*, 550. 10.1186/s13059-014-0550-8.