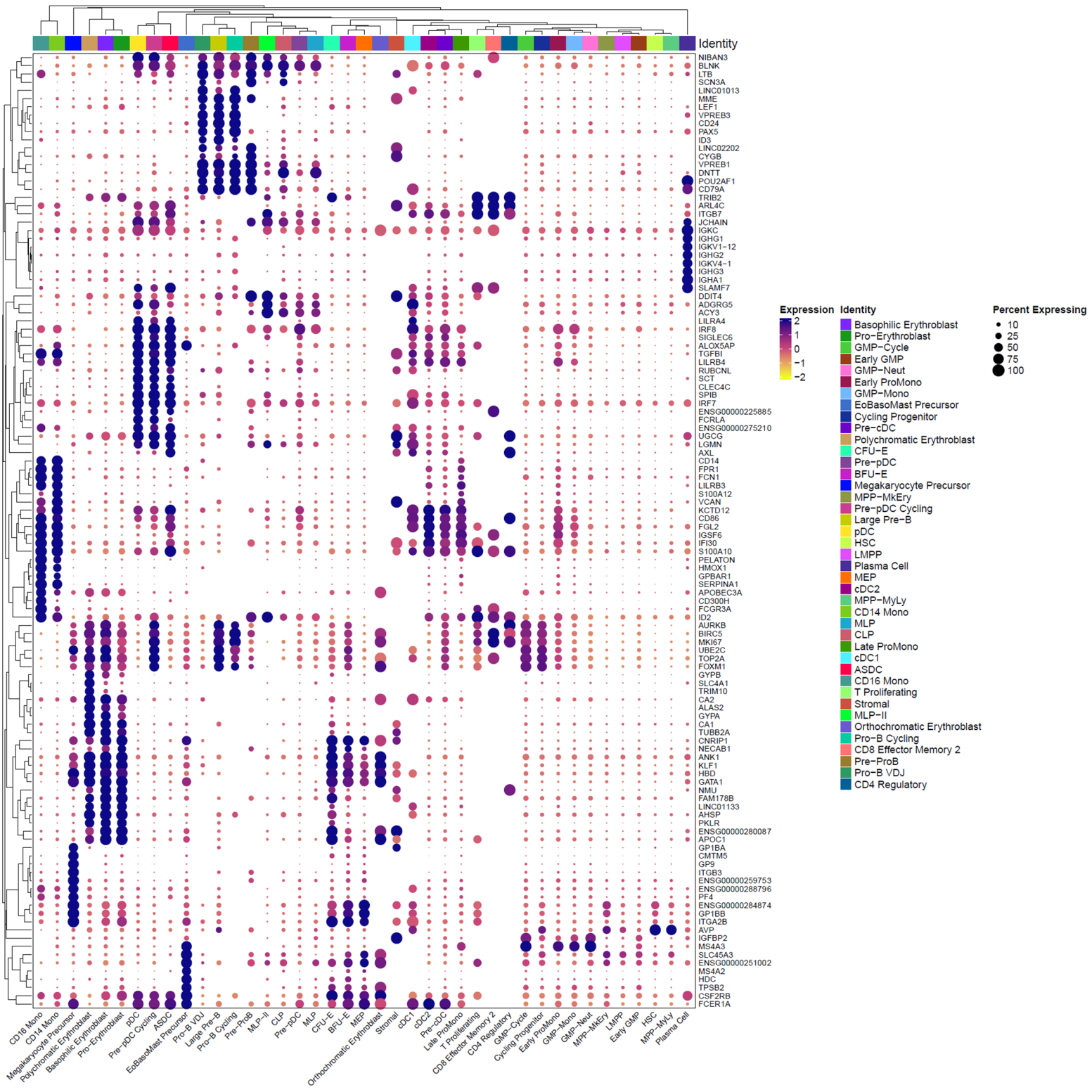
Supplementary table: List of the primer sequences used for p.R525 genotyping.

Primer name	5' - 3'	Purification
i5-UDI0001 primer	AATGATACGGCGACCACCGAGATCTACACAGCGCTAGACACTCTTTCCCTACACG*A*C	OPC
i7-UDI0001 primer	CAAGCAGAAGACGGCATACGAGATAACCGCGGGTGACTGGAGTTCAGACGT*G*T	OPC
P5 generic primer	AATGATACGGCGACCACCGAGATCTACAC	OPC
DDX41 5' distal primer	CCAGCACGTCATCAATTATGAC	OPC
DDX41 5' proximal primer	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTGAGATTGAGAACTATGTACACC	OPC

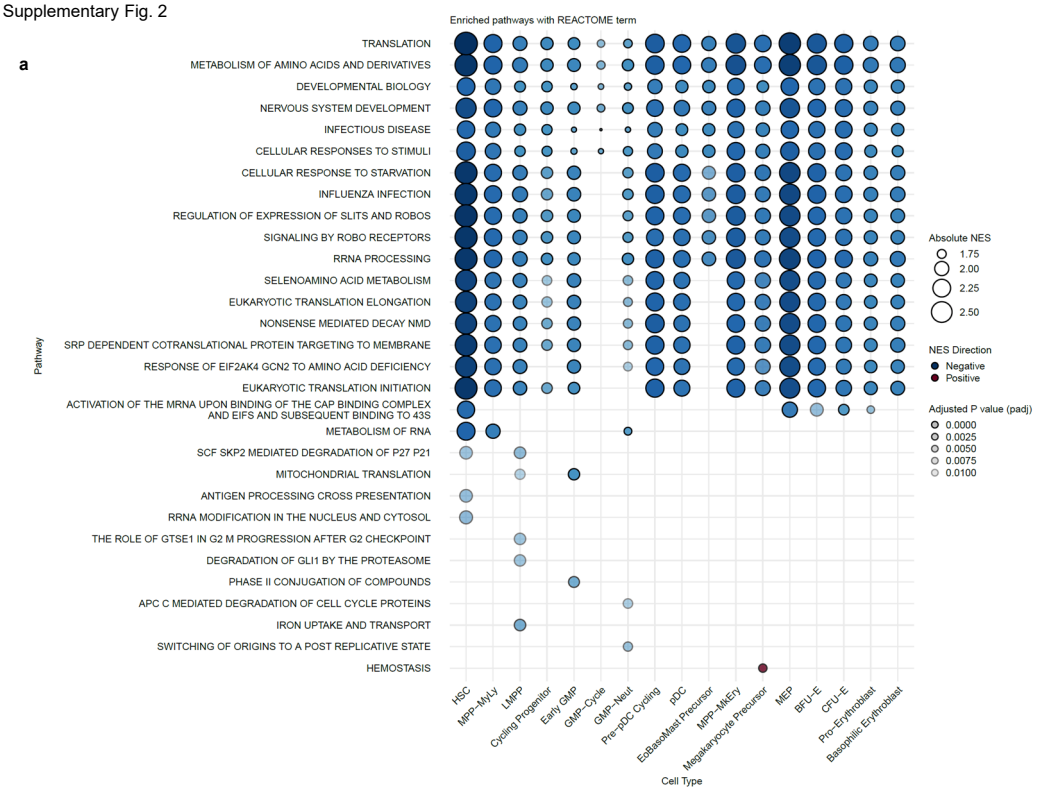
OPC: Oligonucleotide Purification cartridge purification grade

*Phosphorthioate modification

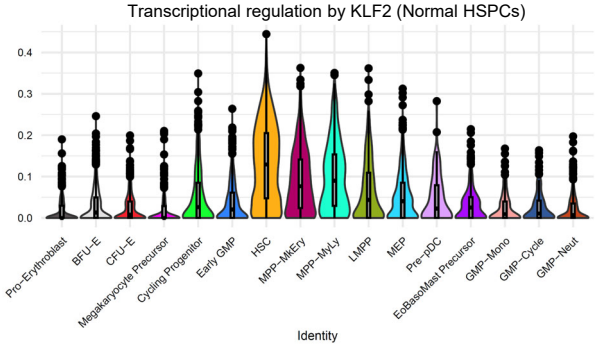


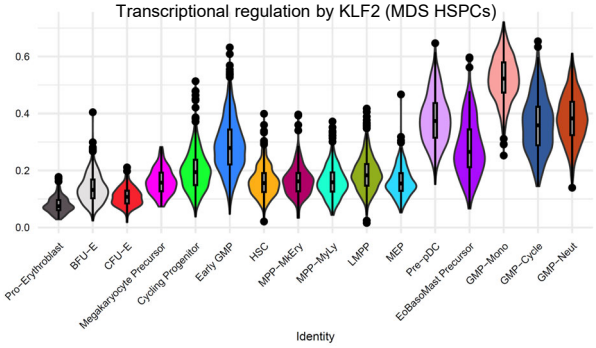
Supplementary Fig. 1: Characterization of CD34-positive bone marrow cells obtained from the studied case.
Gene expression profiles of genes with high cell-type specificity. For each cell type classified by BoneMarrowMap, cell-type-specific genes were identified using the 'FindAllMarkers' function of Seurat, and the top genes per cell type were further analyzed and visualized using the 'Extract_Top_Markers' and 'Clustered_Top_Markers' functions in the scCustomize (2.1.2) R package. The top seven genes per cell type are presented in a combined format. Cell-type abbreviations are provided in the main text and Fig.2.

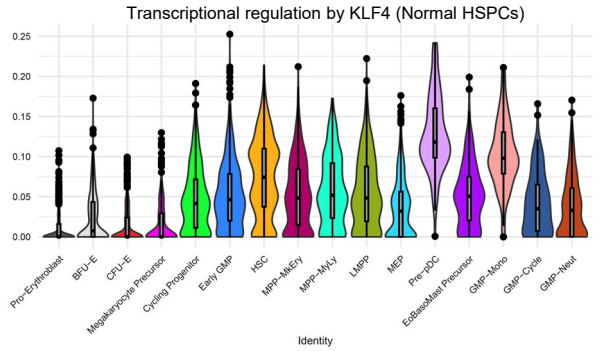
Supplementary Fig. 2

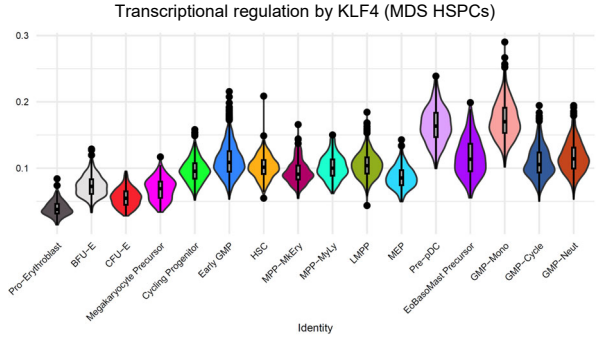


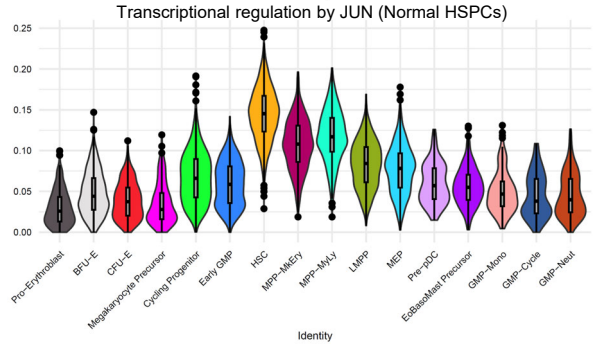
b

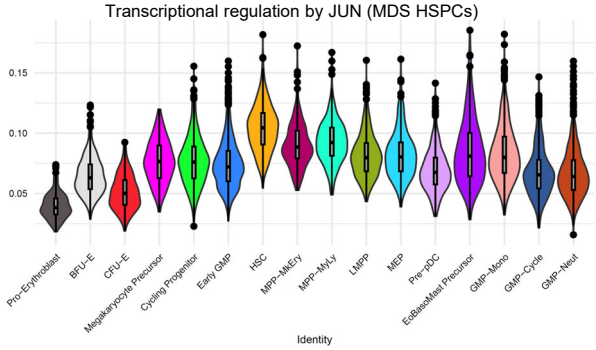


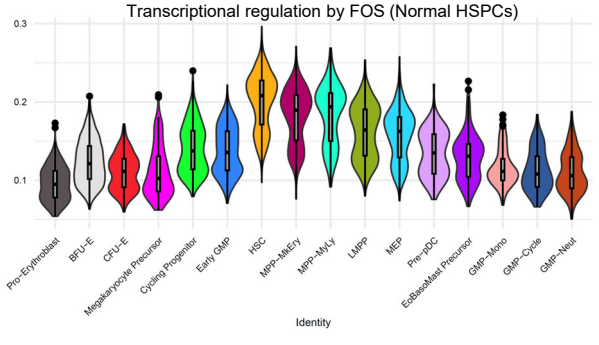


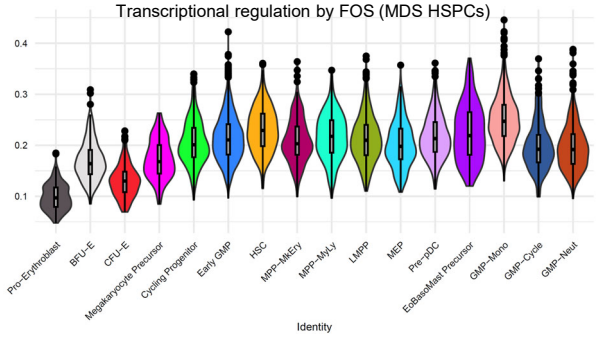




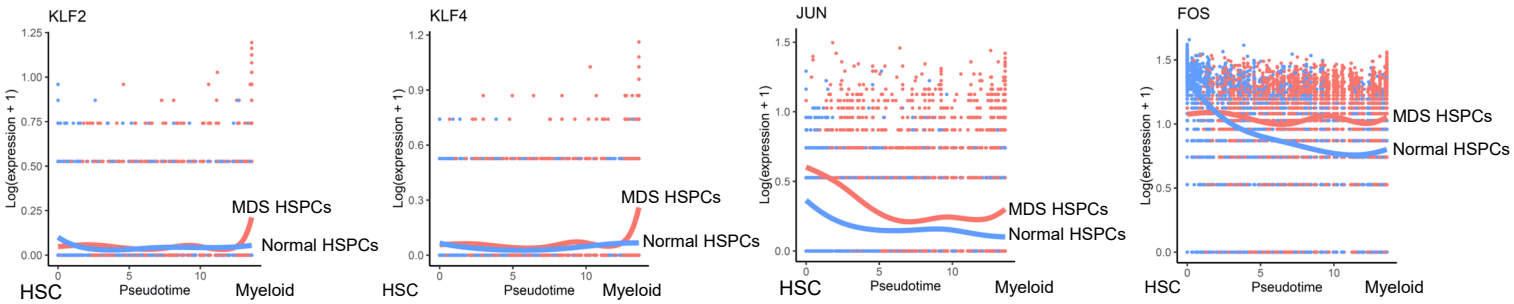






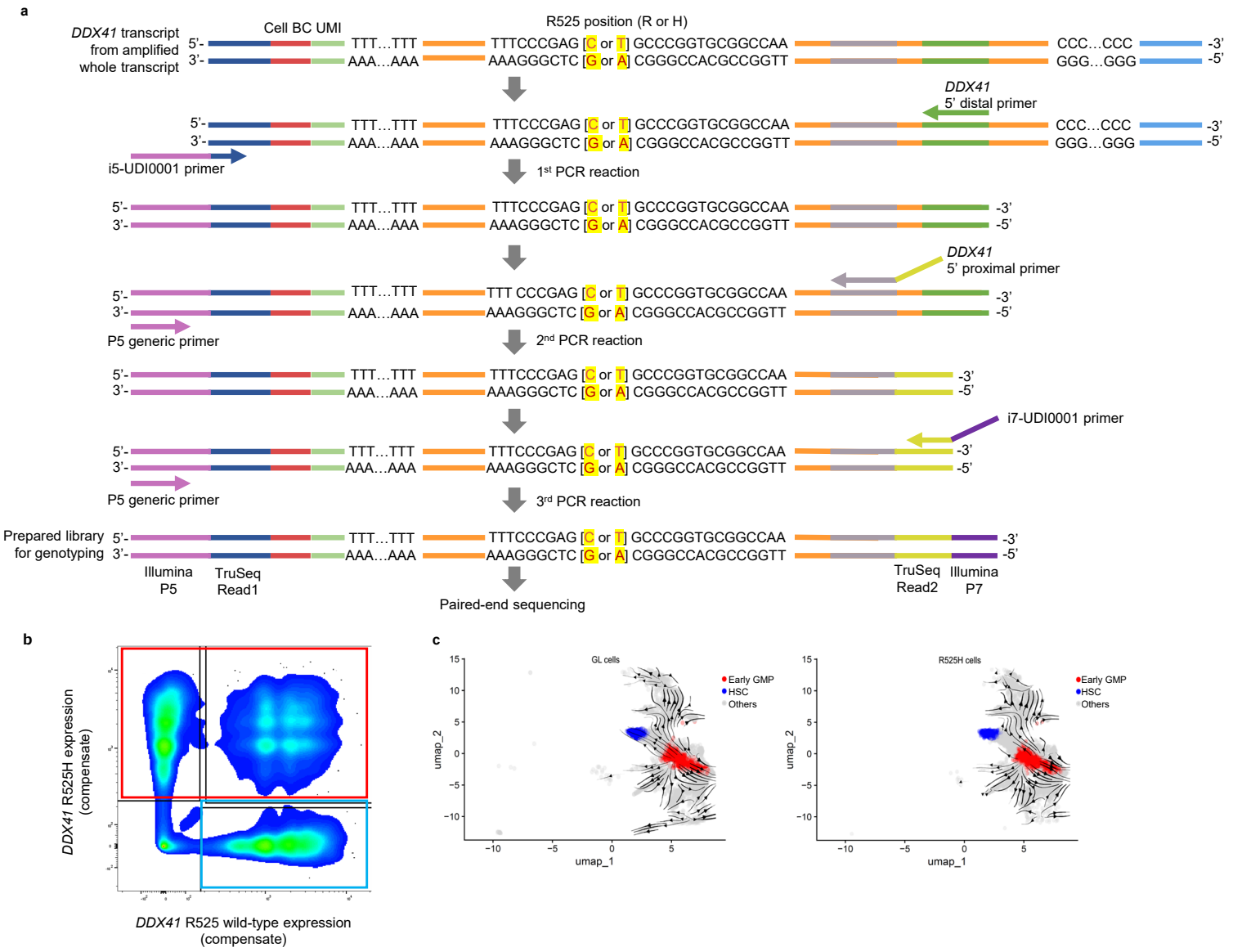


c

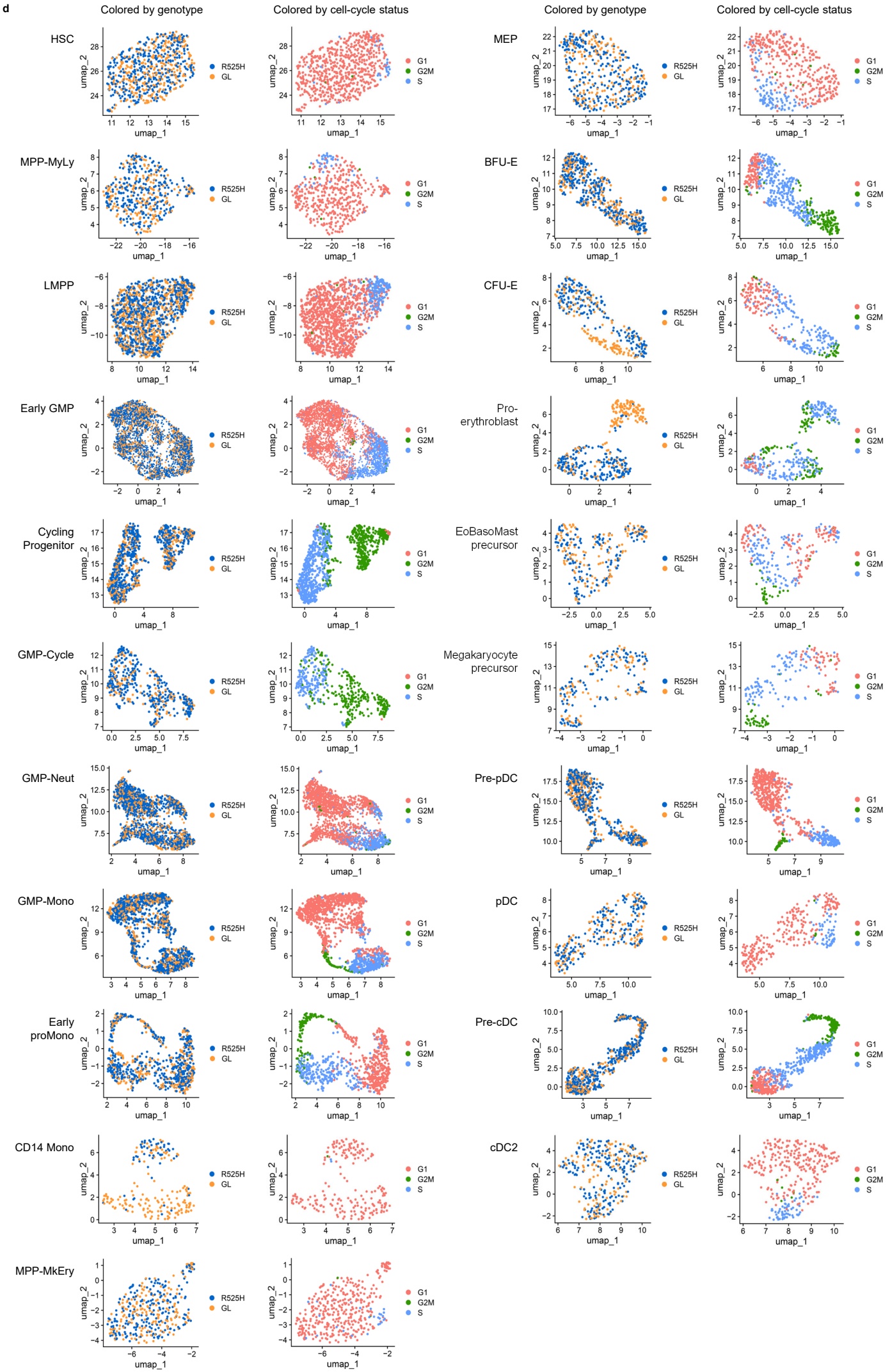


Supplementary Fig. 2: Distinct transcriptional regulations of CD34-positive HSPCs between MDS case and healthy donors.
a Bubble plot presenting GSEA results for each cell type. GSEA was performed as outlined in the Materials and Methods section and Fig. 3a. Shown pathways meet the criteria of *padj* < 0.01 and are classified under the REACTOME category. Pathways with positive and negative enrichment in the MDS sample are depicted in red and blue, respectively. Circle size represents the NES, and color intensity indicates the statistical significance of the *p*-value. **b** Violin plots displaying AUC score distributions by cell type for transcription factors KLF2, KLF4, JUN, and FOS, with box plots overlaid. Cell type ordering is consistent with Fig. 3c. **c** Pseudotime analysis of *KLF2, KLF4, JUN,* and *FOS* expression levels. Pseudotime trajectories of cell differentiation, estimated by Monocle3, are shown on the horizontal axis, with log-transformed expression values [log(expression + 1)] for the transcription factors across temporal series in normal (blue) and MDS (red) HSPCs.
Abbreviations: AUC, area under the curve; GSEA, gene set enrichment analysis; HSPCs, hematopoietic stem and progenitor cells; KLF2, Kruppel-like factor 2; MDS, myelodysplastic neoplasms; NES, Nominal Enrichment Score; *padj*, adjusted *p*-values. Cell-type abbreviations are provided in the main text and Fig.2.

Supplemental Fig. 3

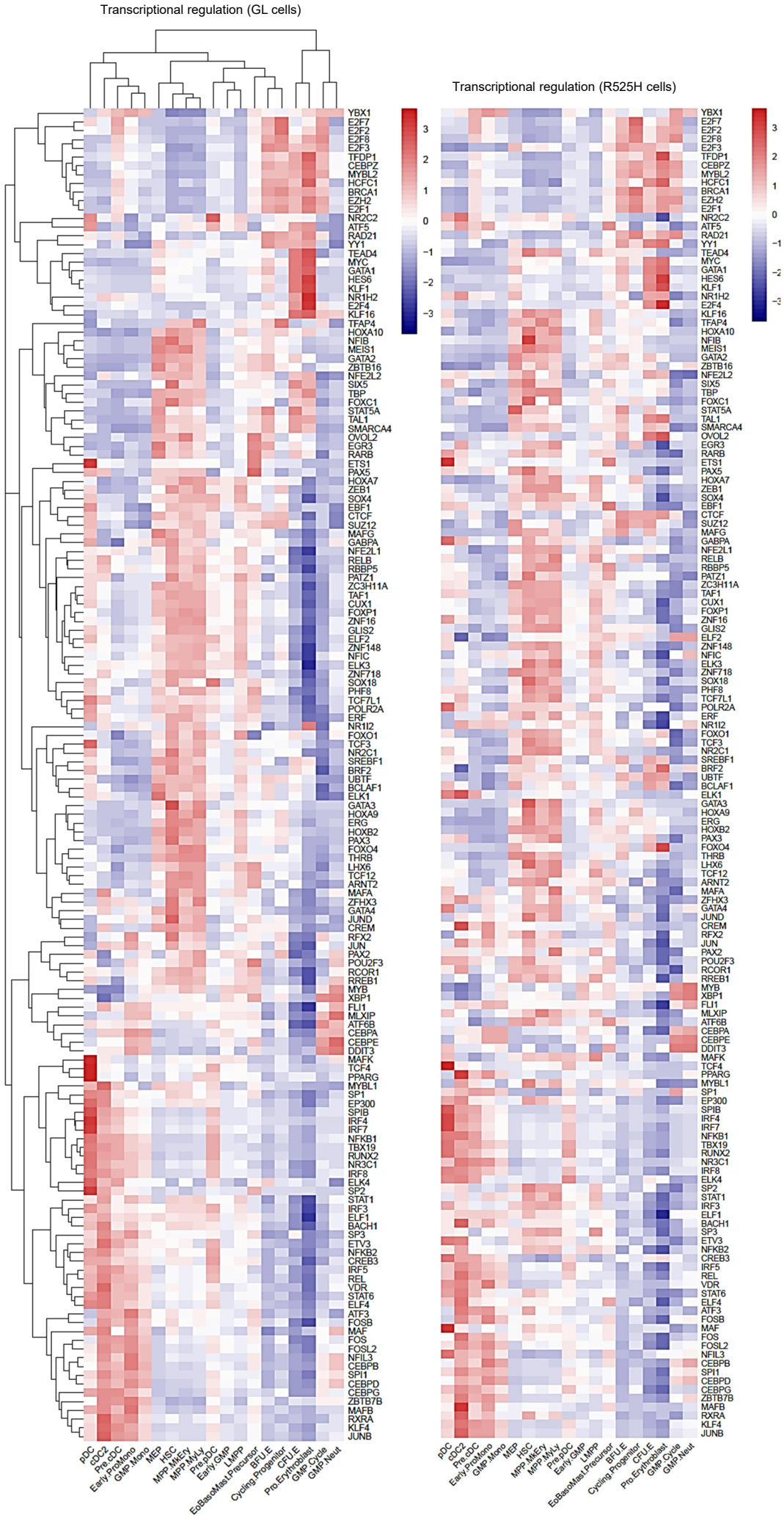


Supplemental Fig. 3 (continued)

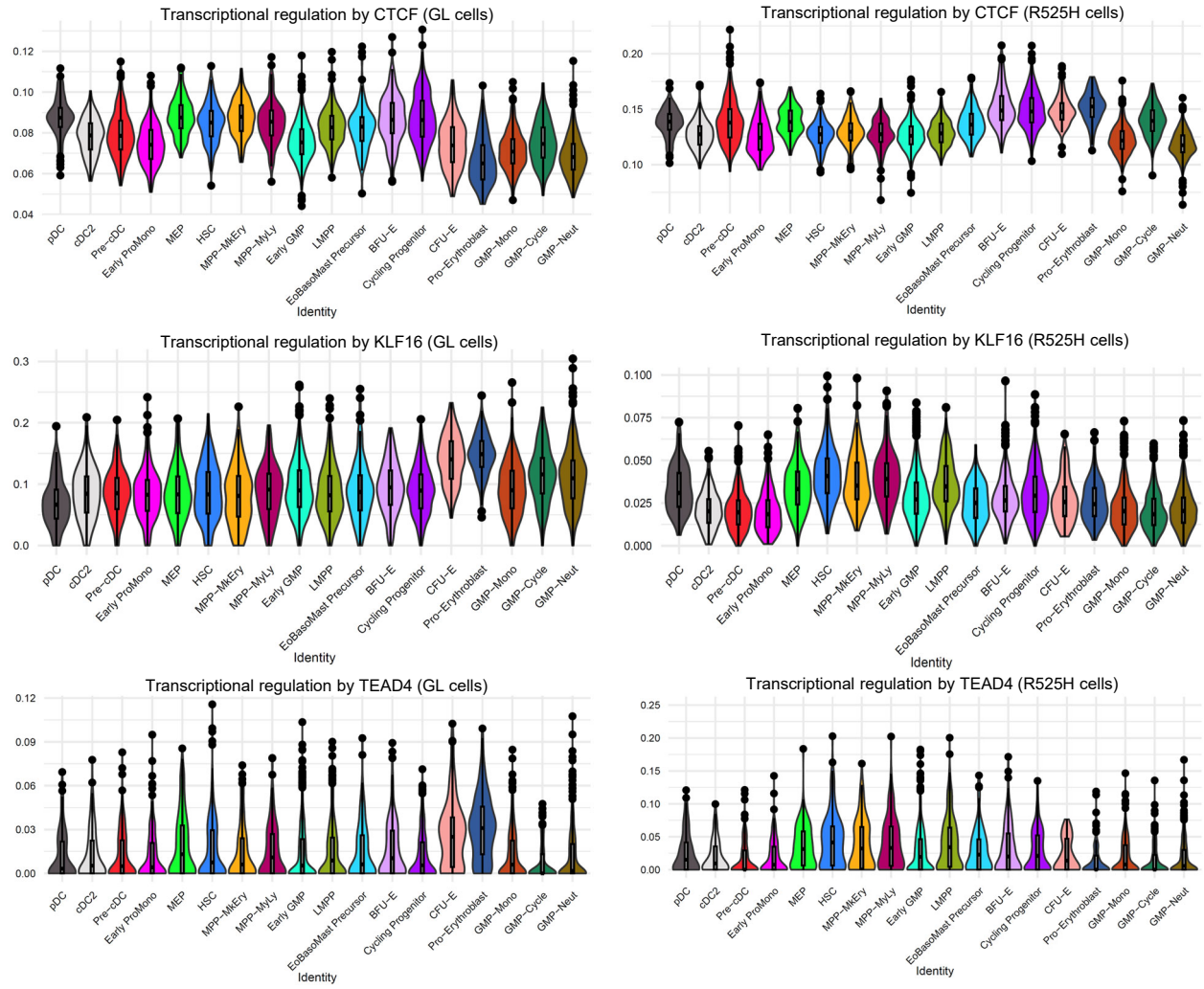


Supplemental Fig. 3 (continued)

e

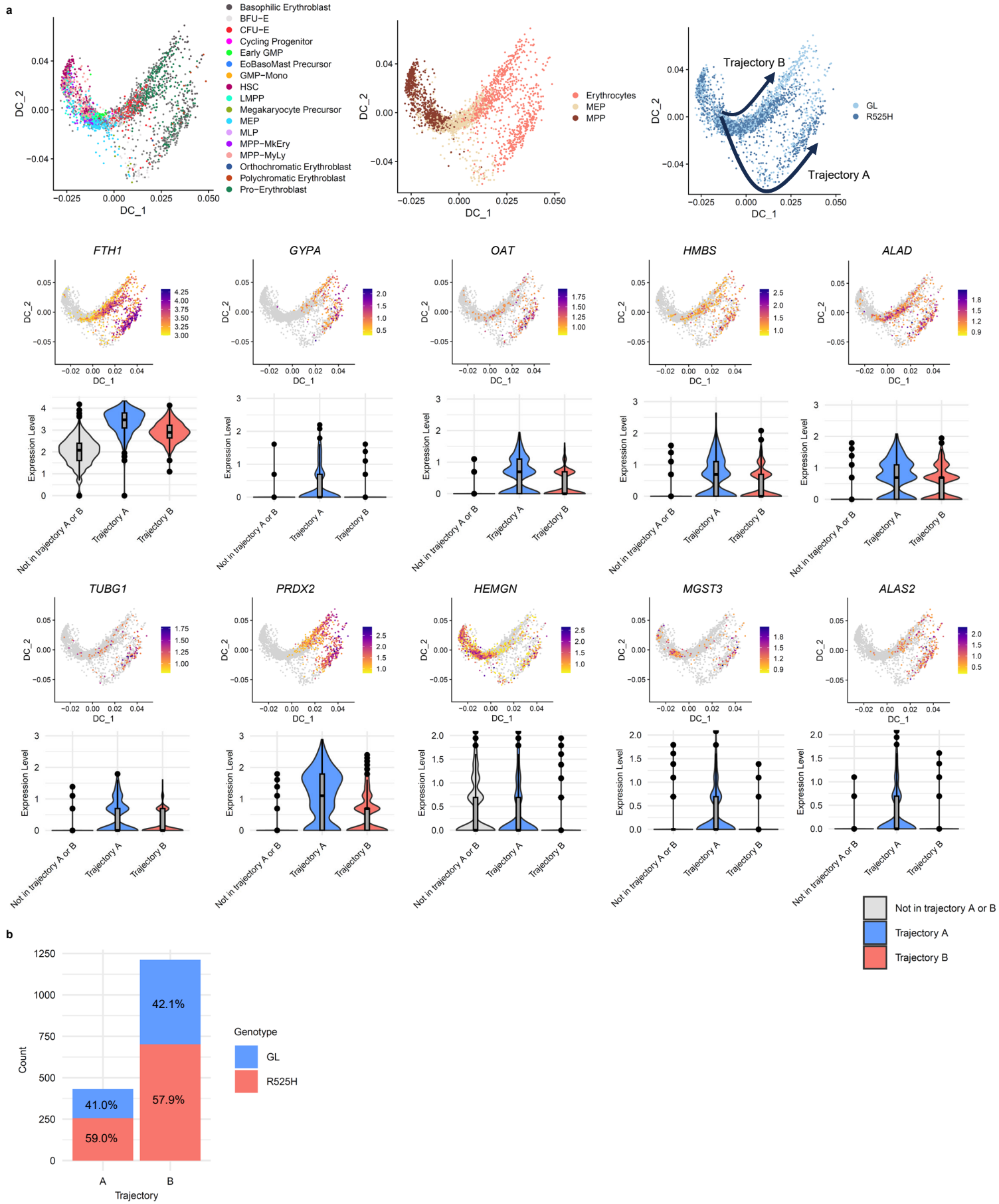


f



Supplementary Fig. 3: Comparison of R525H and GL cells by scRNA-seq analysis combined with single-cell genotyping.
a Schematic of single-cell genotyping methodology: Primers (listed in Supplementary table) were designed to flank the *DDX41* p.R525 site and the 3' end cell barcode, using the amplified whole transcript as the template. Library preparation for genotyping involved three rounds of PCR. **b** Cell-specific genotyping information identified using FlowJo. Cells within the red and blue boxes indicate R525H and GL cells, respectively. **c** UMAP plots generated using the SAM algorithm, where cells are distinguished by genotype (R525H or GL) and by cell cycle phase (G1, S, or G2M) with color-coded labels. **d** Transcriptional regulation profiles across cell types, for both GL and R525H cells. Average AUC scores for transcription factors were calculated per cell type, and results are presented in a heatmap, with the R525H heatmap aligned to match the GL cell order for comparison. **e** Violin plots displaying AUC score distributions by cell type for transcription factors CTCF, KLF16, and TEAD4, with box plots overlaid on each violin plot. Cell type ordering is consistent with panel **d**.
Abbreviations: AUC, area under the curve; BC, barcode; CTCF, CCCTC-binding factor; DDX41, DEAD-box helicase 41; GL, germline; KLF16, Kruppel-like factor 16; SAM, self-assembling manifold; scRNA-seq, single-cell RNA sequencing; TEAD4, TEA domain transcription factor 4; UMAP, Uniform Manifold Approximation and Projection; UMI, unique molecular identifier. Cell-type abbreviations are provided in the main text and Fig. 2.

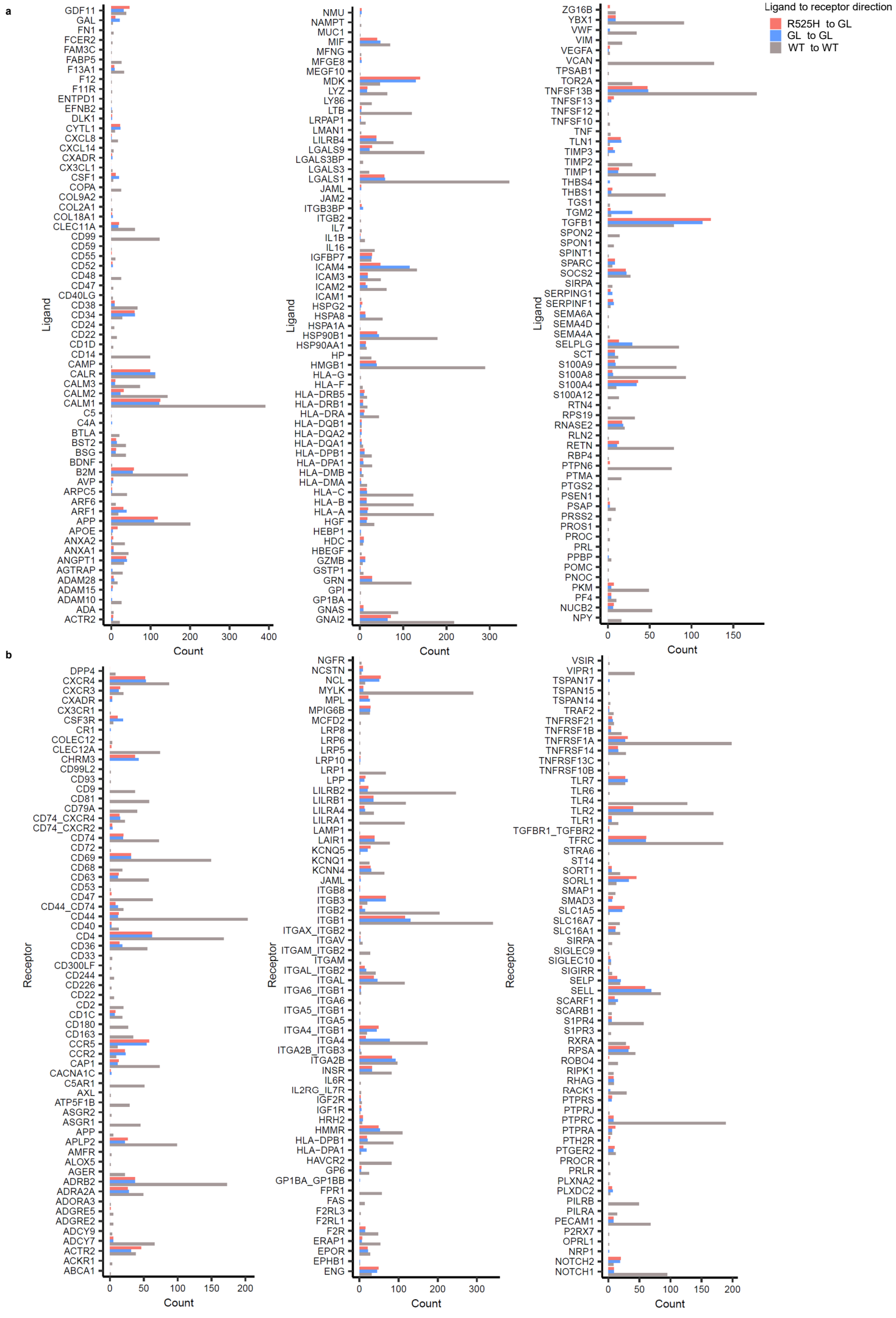
Supplemental Fig. 4



Supplementary Fig. 4: Diffusion map analysis indicating that p.R525H acquisition did not impact the formation of two erythroid differentiation trajectories. **a** Differentiation trajectories of erythroid progenitors identified through diffusion map analysis. Cell types classified using BoneMarrowMap and SingleR are indicated in the top left and top right panels, respectively. Genotype information for each cell is shown at the top right, with indications of the two differentiation trajectories. Gene expression patterns upregulated in trajectory A include *FTH1, GYPA, OAT, HMBS, ALAD, TUBG1, PRDX2, HEMGN, MGST3,* and *ALAS2*. Violin plots display expression differences between trajectories. **b** Ratios of R525H and GL cells across trajectories A and B. Cell counts and ratios for R525H and GL cells were manually determined using Shiny, showing the distribution of each cell type within both trajectories.

Abbreviations: ALAD, aminolevulinate dehydratase; ALAS2, 5'-aminolevulinate synthase 2; DC, diffusion component; GL, germline; FTH1, ferritin heavy chain 1; GYPA, glycophorin A; HEMGN, hemogen; HMBS, hydroxymethylbilane synthase; MGST3, microsomal glutathione S-transferase 3; MPP, multipotent progenitor; OAT, ornithine aminotransferase; PRDX2, peroxiredoxin 2; TUBG1, tubulin gamma 1. Cell-type abbreviations are provided in the main text and Fig. 2.

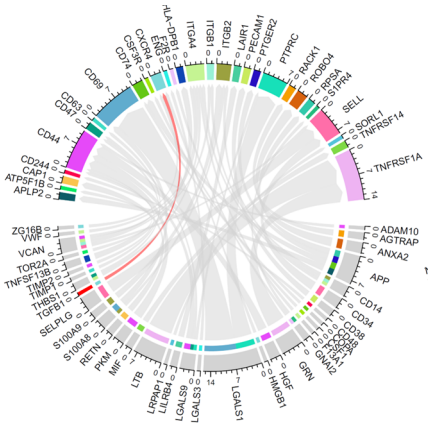
Supplemental Figure 5



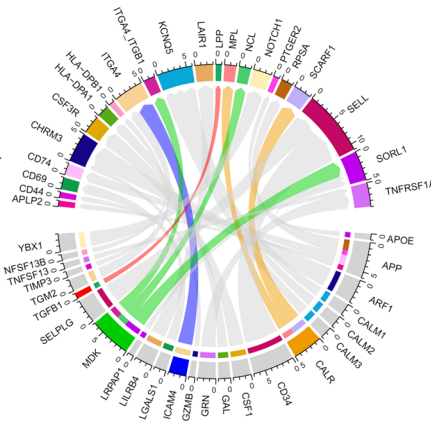
Supplemental Fig. 5 (continued)

c

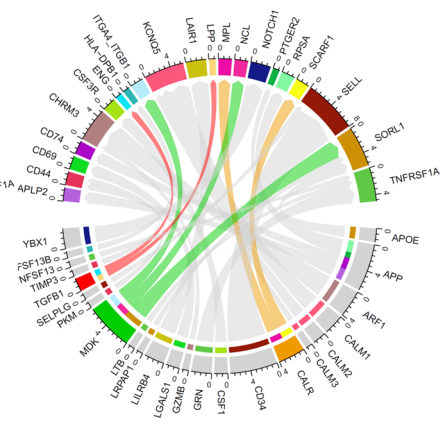
Source cells: normal HSPCs
Target: normal LMPP



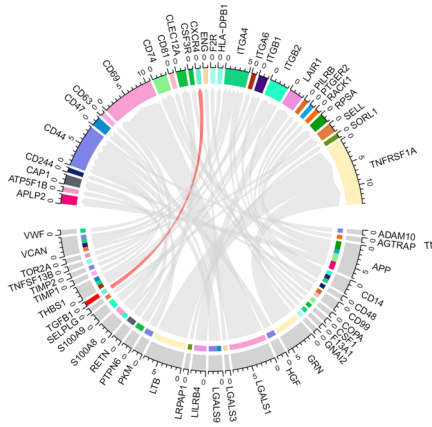
Source cells: GL HSPCs
Target: GL LMPP



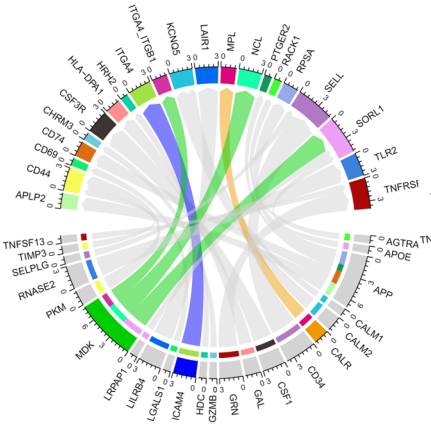
Source cells: R525H HSPCs
Target: GL LMPP



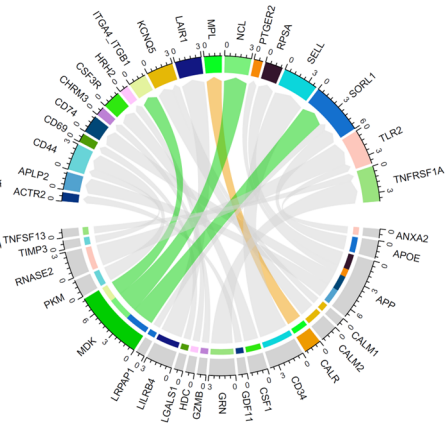
Source cells: normal HSPCs
Target: normal Early GMP



Source cells: GL HSPCs
Target: GL Early GMP

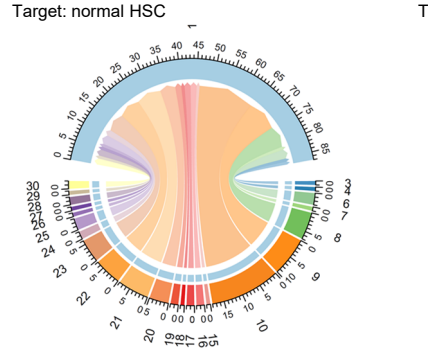


Source cells: R525H HSPCs
Target: GL Early GMP

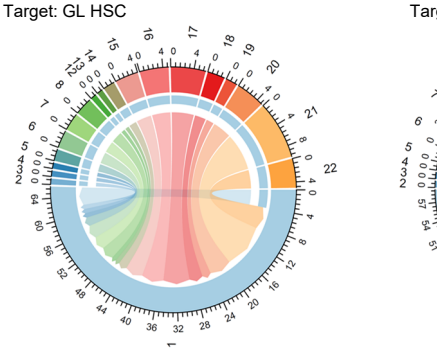


d

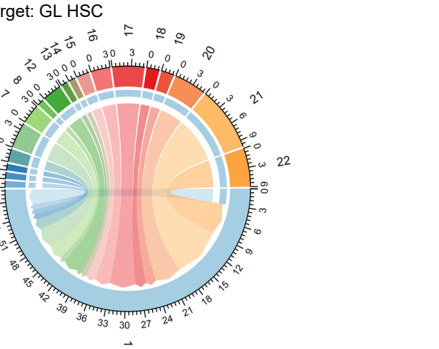
Source cells: normal HSPCs
Target: normal HSC



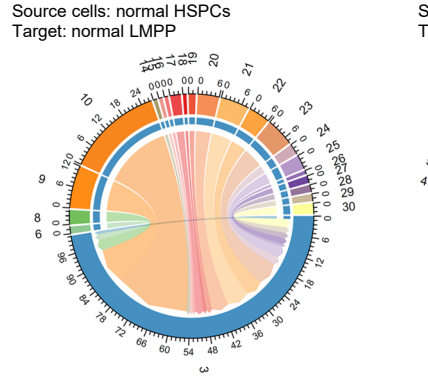
Source cells: GL HSPCs
Target: GL HSC



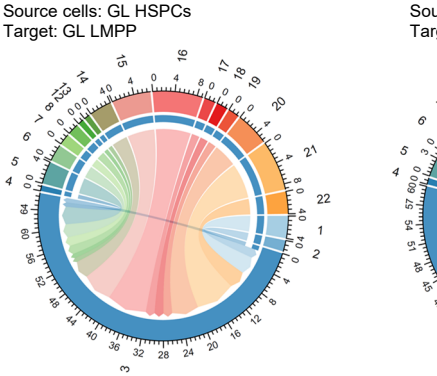
Source cells: R525H HSPCs
Target: GL HSC



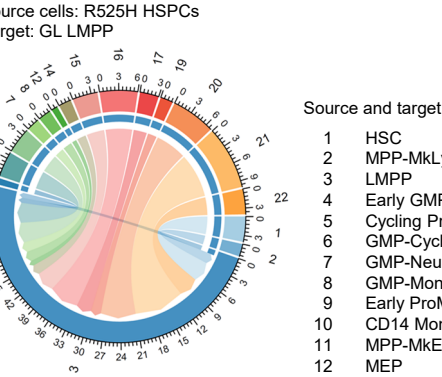
Source cells: normal HSPCs
Target: normal LMPP



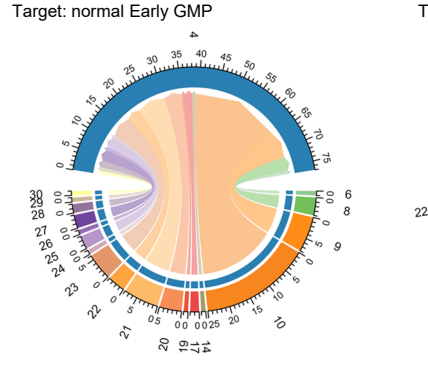
Source cells: GL HSPCs
Target: GL LMPP



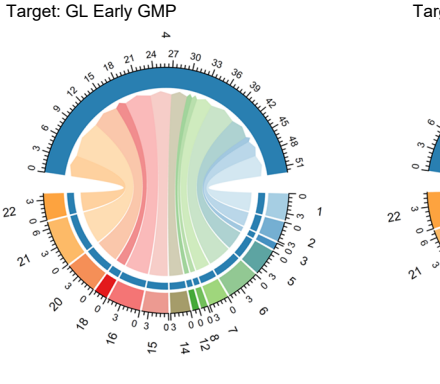
Source cells: R525H HSPCs
Target: GL LMPP



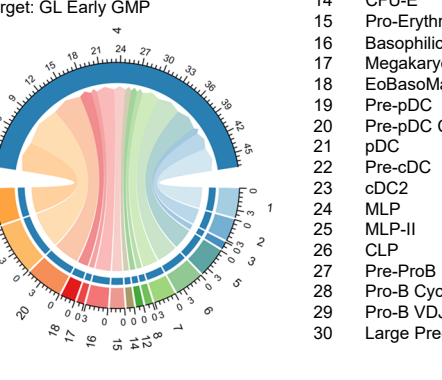
Source cells: normal HSPCs
Target: normal Early GMP



Source cells: GL HSPCs
Target: GL Early GMP



Source cells: R525H HSPCs
Target: GL Early GMP



Source and target cell type

1	HSC
2	MPP-MkLy
3	LMPP
4	Early GMP
5	Cycling Progenitor
6	GMP-Cycle
7	GMP-Neut
8	GMP-Mono
9	Early ProMono
10	CD14 Mono
11	MPP-MkEry
12	MEP
13	BFU-E
14	CFU-E
15	Pro-Erythroblast
16	Basophilic Erythroblast
17	Megakaryocyte Precursor
18	EoBasoMast Precursor
19	Pre-pDC
20	Pre-pDC Cycling
21	pDC
22	Pre-cDC
23	cDC2
24	MLP
25	MLP-II
26	CLP
27	Pre-ProB
28	Pro-B Cycling
29	Pro-B VDJ
30	Large Pre-B

Supplementary Fig. 5: Different ligand-receptor interactions between normal and MDS HSPCs and between R525H and GL cells.
a, b Ligands **a** and receptors **b** involved in interactions among healthy donor-derived normal HSPCs, within GL cells, and from R525H cells (as ligand-presenting cells) to GL cells (as receptor-presenting cells). Interaction frequencies were estimated using LIANA, with counts indicating ligand and receptor appearance frequency. **c, d** Chord diagrams illustrating cell-to-cell interactions within healthy HSPCs, within GL cells, and from R525H cells to GL cells, focusing on LMPP and early GMP as receptor-presenting cell types. In **c**, ligand-receptor interactions are highlighted, with text-referenced connections accentuated by color, while **d** presents cell-type connections specifically to LMPP and early GMP.
Abbreviations: GL, germline; HSPCs, hematopoietic stem and progenitor cells; MDS, myelodysplastic neoplasms; WT, wild-type. Cell-type abbreviations are provided in the main text and Fig.2.