Cell Reports Methods, Volume *2*

Supplemental information

Integrating transcription-factor abundance with chromatin accessibility in human erythroid lineage commitment

Reema Baskar, Amy F. Chen, Patricia Favaro, Warren Reynolds, Fabian Mueller, Luciene Borges, Sizun Jiang, Hyun Shin Park, Eric T. Kool, William J. Greenleaf, and Sean C. Bendall

Supplemental materials

Table S1. Mass cytometry antibody panel, related to STAR methods

Antibody target and metal	Source	Catalog #	Staining concentration (mg/mL)
CD45-89Y	Fluidigm	3089003	2
CD235-113In	Biolegend	349102	1
CD71-115	Biolegend	334102	3
CD61-139La	Biolegend	336402	0.5
CD3-141Pr	BD Biosciences	561416	1
CD19-142Nd	Biolegend	302202	0.25
CD90-143Nd	BD Biosciences	555594	5
CD14-144Nd	Biolegend	301802	4
CD164-145Nd	Biolegend	324802	4
CD34-148Nd	Biolegend	343502	1
CD105-150Nd	Biolegend	323202	4
CD123-151Eu	Biolegend	306002	1
CD10-152Sm	Biolegend	312202	1
FcER1-153Eu	Biolegend	334602	1
CD84-154Sm	Biolegend	326002	1
CD33-158Gd	Biolegend	303402	2
CD11c-159Tb	Biolegend	301602	1
GATA-1-160Gd	Cell Signalling	3535	3
CD7-162Dy	BD Biosciences	555359	1
CD49f-164Dy	Biolegend	555734	3
CD127-165Ho	Biolegend	351302	1
CD66-167Er	BD Biosciences	551354	1
CD38-168Er	Biolegend	303502	2
CD45RA-169Tm	Biolegend	304102	1
CD135-170Er	Biolegend	313302	1.5
CD117-171Yb	Biolegend	313202	2
CD-133-172Yb	Miltenyi Biotec	130-108-062	8
CD172ab-173Yb	Biolegend	323802	3
CD2-174Yb	Biolegend	309202	1
HLA-DR-209Bi	Biolegend	307602	1

Table S2. FACS antibody panel, related to STAR methods

Antibody target and fluorochrome	Source	Catalog #	Clone
CD15-biotinylated	Biolegend	301914	HI98
CD3-biotinylated	Biolegend	300404	UCHT1
CD7-biotinylated	Thermo Fisher Scientific	13-0079-82	124-1D1
CD56-biotinylated	Biolegend	362536	5.1H11
CD34-FITC	Myltenyi	130113178	AC136
CD38-BV421	Biolegend	303526	HIT2
CD45RA-AF700	Biolegend	304120	HI100
CD10-BV650	BD Biosciences	563734	HI10a
CD123-PECy7	Biolegend	306010	6H6
GATA-1-PE	Cell Signalling	13353S	D52H6
CD38-APC/Cy7	Biolegend	303534	HIT2
CD71-PE	Biolegend	334108	CY1G4
CD33-PE/Cy7	Biolegend	303434	WM53
CD84-APC	Biolegend	326010	CD84.1.21

Supplemental figures

Figure S1, related to Figure 1

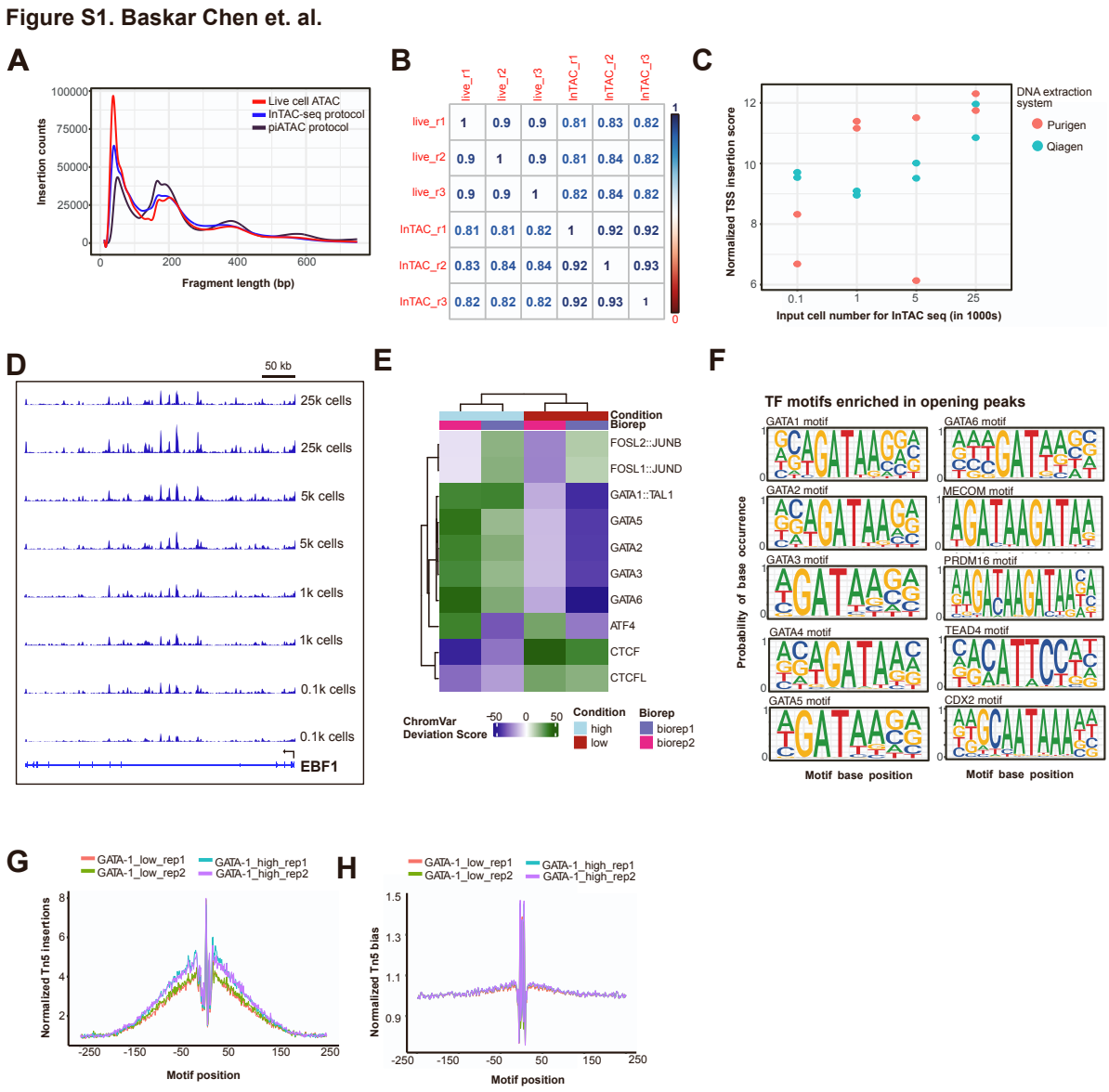

Figure S1. InTAC-seq works robustly on low cell numbers (down to 100 cells)

A. Fragment length distribution of ATAC-seq libraries generated from live, InTAC-seq and published bulk piATAC samples. **B.** Spearman correlation plot of reads in peaks across live and InTAC-seq samples. **C.** Normalized TSS insertion scores for InTAC-seq samples generated from the indicated input cell numbers across replicates and DNA extraction methods. **D.** Genome coverage at EBF1 locus of InTAC-seq data generated from indicated cell numbers. **E.** Heatmap of chromVAR deviation scores across GATA-1 high and GATA-1 low K562 cells for the top 50 most variable TF motifs. **F.** TF motif sequence logos for most significantly enriched motifs in opening peaks as shown in Fig1H. **G.** Bias-corrected, normalized Tn5 insertions across all GATA-1 motif sites in consensus peaks. **H.** Normalized Tn5 bias across all GATA-1 motif sites in consensus peaks.

Figure S2, related to Figure 2

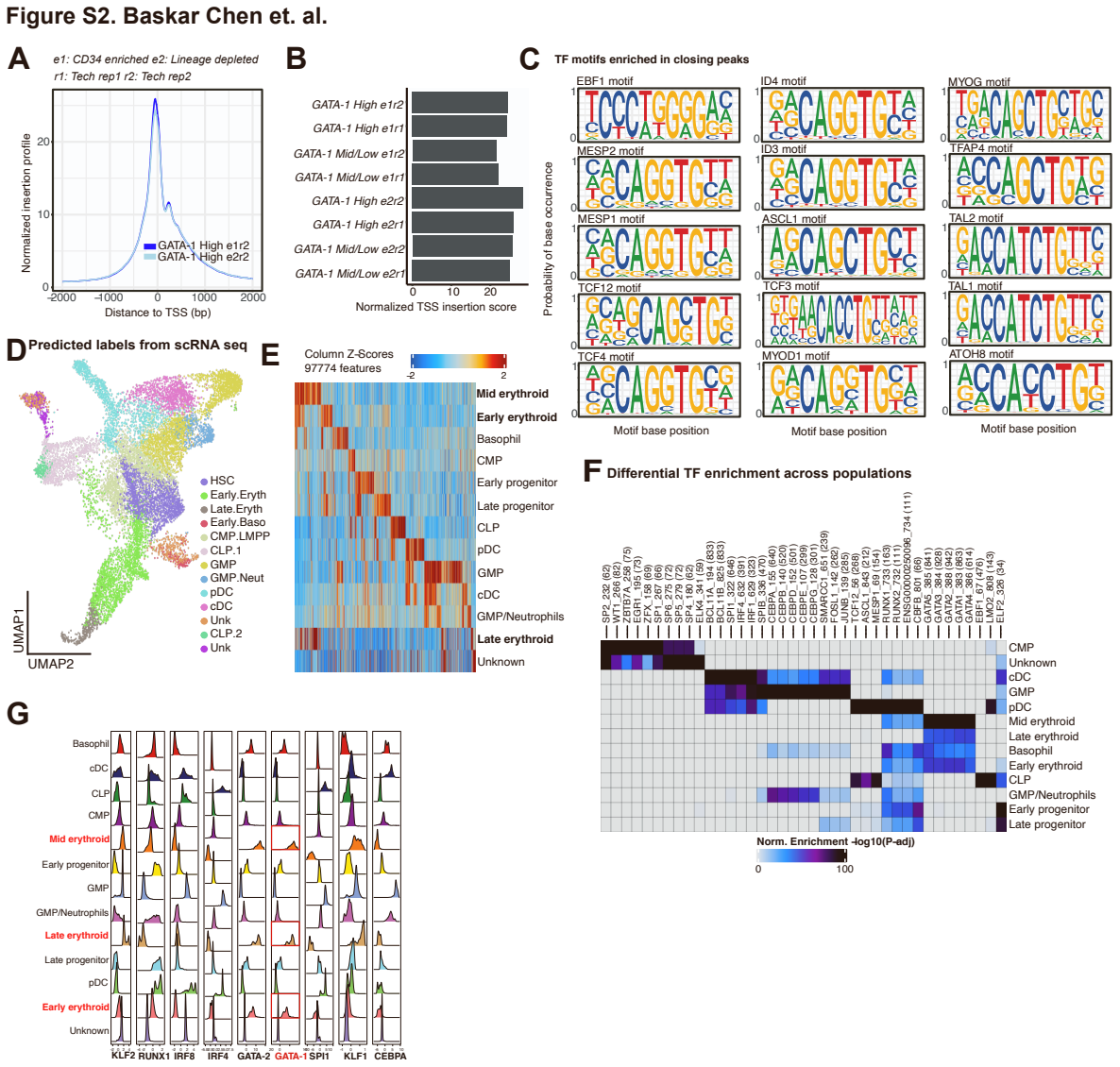

Figure S2. InTAC-seq data can be integrated into scATAC to inform annotation of GATA-1 high BM progenitor cells

A. TSS insertion profiles for GATA-1 high BM samples. **B.** Normalized TSS scores of technical and biological replicates for GATA-1 High and Mid/Low BM progenitor samples. **C.** TF motif sequence logos for motifs most significantly enriched in peaks that are significantly more accessible in GATA-1 mid/low population as shown in Fig2E. **D.** UMAP of published scATAC data predictively labelled by previously annotated scRNA seq reads using Seurat's TransferAnchors single cell genomics integration method. **E.** Heatmap of variable gene score features across scATAC cluster labels. **F.** Heatmap of differential TF enrichment by wilcox testing across annotated BM populations. **G.** Stacked density plots of chromVAR deviation scores of key TFs across scATAC cluster labels.

Figure S3, related to Figure 3

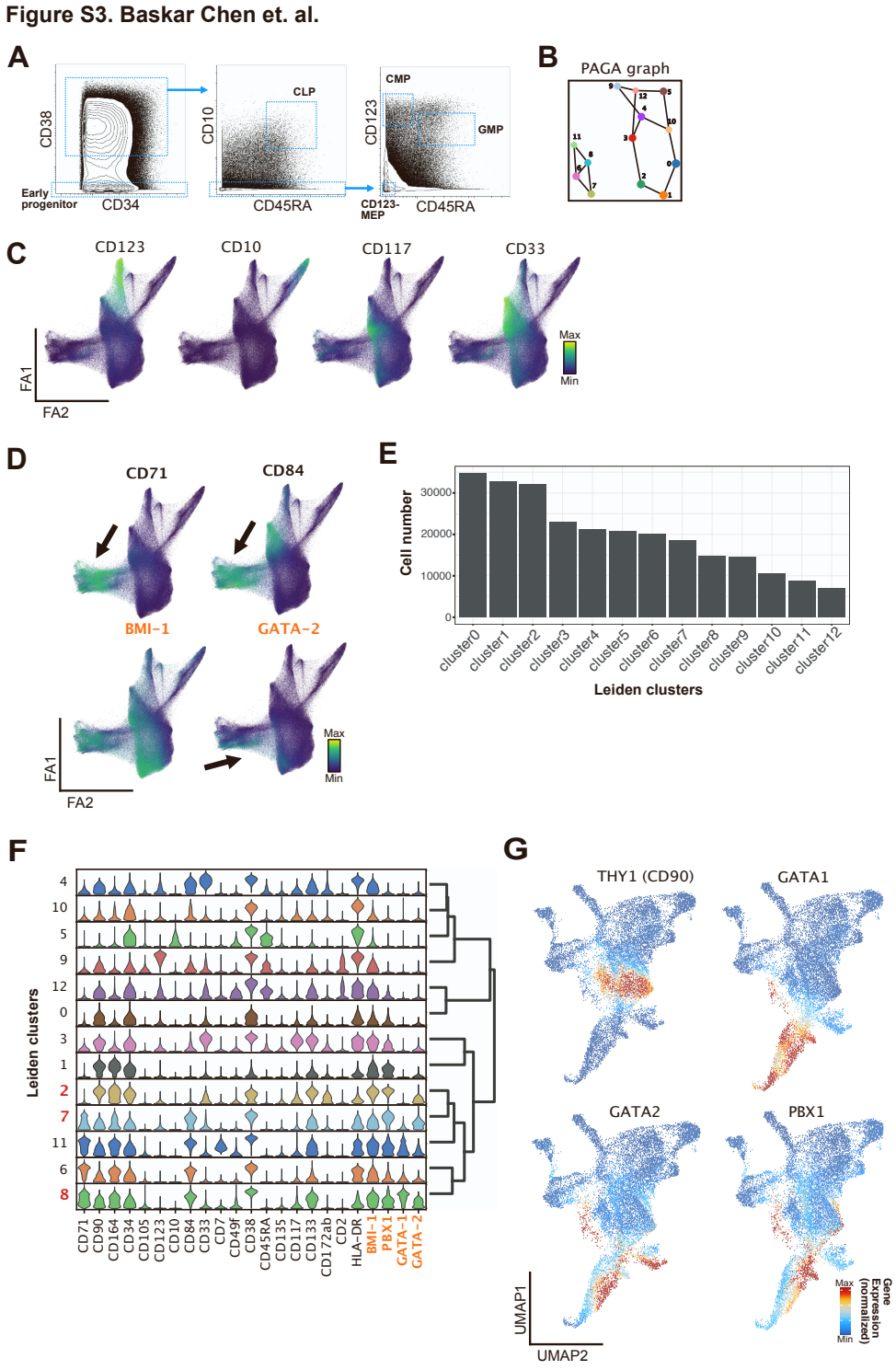

Figure S3. Single-cell proteomic profiling of BM progenitor revealed phenotypic states associated with GATA-1 expression

A. Mass cytometry gating strategy for early progenitors, CLP, GMP, CMP and CD123- MEP respectively **B.** PAGA graph constructed based on nodes defined by Leiden clustering on density-downsampled dataset **C.** Force-directed layout of

downsampled BM progenitors overlaid with key surface marker abundance **D.** Normalized marker expression of relevant erythropoiesis-related surface markers and TF proteins (in orange) across force-directed layout. **E.** Barplot showing number of cells per cluster across Leiden clusters **F.** Stacked violin plot of surface markers and TF proteins across Leiden clusters **G.** scATAC UMAP visualized with integrated gene expression values for markers demarcating early progenitors such as HSC as well as erythroid-primed progenitors.

Figure S4, related to Figure 4

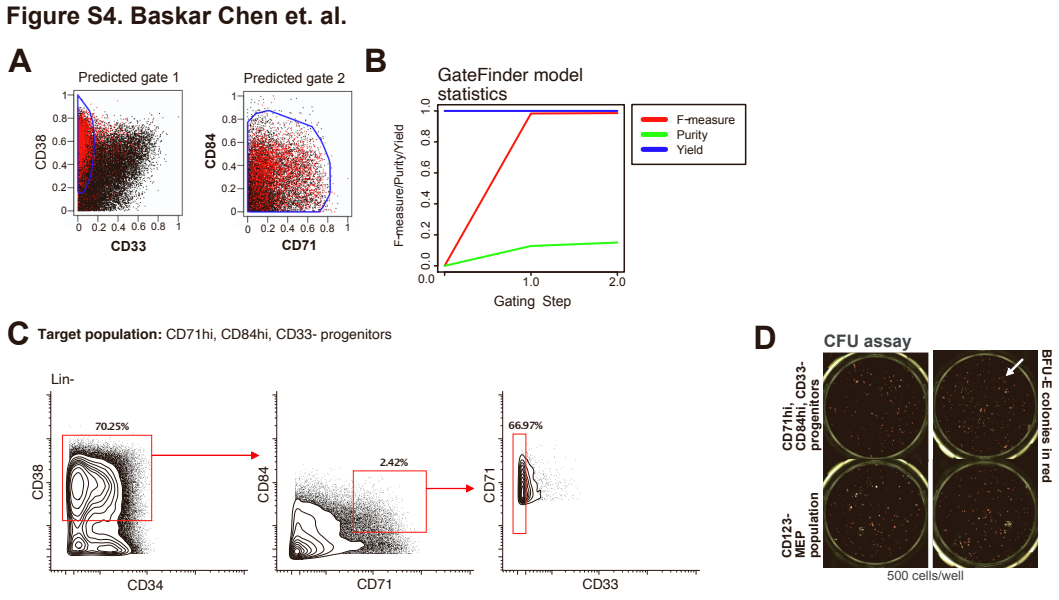

Figure S4. GateFinder identified surface marker gating strategy to enrich for GATA-1 high BM progenitors

A. Predicted gates from GateFinder identifies CD33, CD71 and CD84 (in bold) as putative surface gating surrogates for GATA-1 high BM progenitor cells. **B.** GateFinder algorithm output of gating steps (2) and F-measure, purity and yield of target GATA-1 high BM progenitor population as enriched from all BM progenitors. **C.** Contour plots of BM progenitors gated with predicted strategy. **D.** BFU colonies (in red) in clonal assay from single cell seeded target population (as defined by CD71hi, CD84hi, CD33-) vs CD123- MEP population.

Figure S5, related to Figure 5

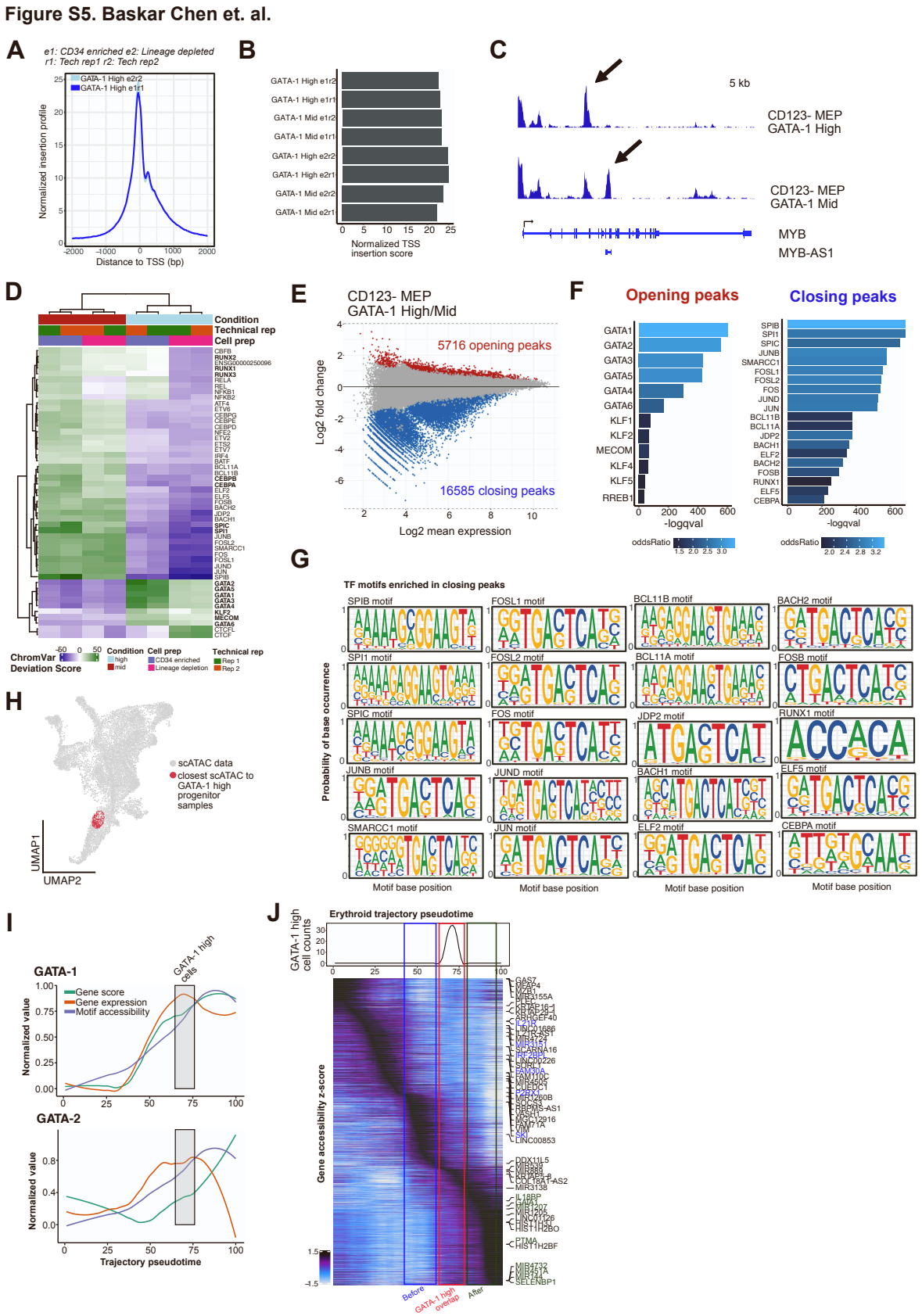

Figure S5. High GATA-1 protein abundance inferred through InTAC-seq overlaps epigenetic switch to erythroid commitment

A. TSS insertion profiles for GATA-1 high CD123- MEP samples. **B.** Normalized TSS scores of technical and biological replicates for GATA-1 high and GATA-1 mid CD123- MEP samples. **C.** InTAC genome coverage plots at MYB locus comparing GATA-1 high and GATA-1 mid CD123- MEP samples. **D.** Heatmap of chromVAR deviation scores across GATA-1 high and mid CD123- MEP cells for top 50 most variable TFs. **E.** MA plot of log2 fold change in accessibility between GATA-1 high and GATA-1 mid CD123- MEPs vs log2 mean number of reads in consensus peaks. Peaks with significant changes in accessibility are highlighted in red or blue. **F.** Most significantly enriched TF motifs in differentially accessible peaks between GATA-1 high and mid CD123- MEP cells calculated using Fisher's test. **G.** TF motif sequence logos of motifs most significantly enriched in peaks that are significantly more accessible in GATA-1 mid population as shown in FigS3F. **H.** Closest 500 scATAC cells to projected bulk GATA-1 high InTAC samples by mahanoblis distance visualized on scATAC UMAP. **I.** Integrated line plot with gene accessibility score (gene score), gene expression score and motif accessibility score (all normalized between 0-1) of GATA-1 and GATA-2 across derived erythoid trajectory with overlaid demarcation of where inferred GATA-1 high scATAC cells are positioned. **J.** Heatmap of top variable gene scores by gene accessibility across constructed erythroid trajectory with overlaid demarcation of before, at and after GATA-1 high point in the trajectory. (Gene colours correspond to the trajectory bins)