

SUPPLEMENTARY INFORMATION

A stem cell differentiation model reveals two alternative fates in CBFA2T3::GLIS2-driven acute megakaryoblastic leukemia initiation

Mohamed R. Shoeb, Dagmar Schinnerl, Lisa E. Shaw, Matthias Farlik, Sabine Strehl, Florian Halbritter,
Klaus Fortschegger

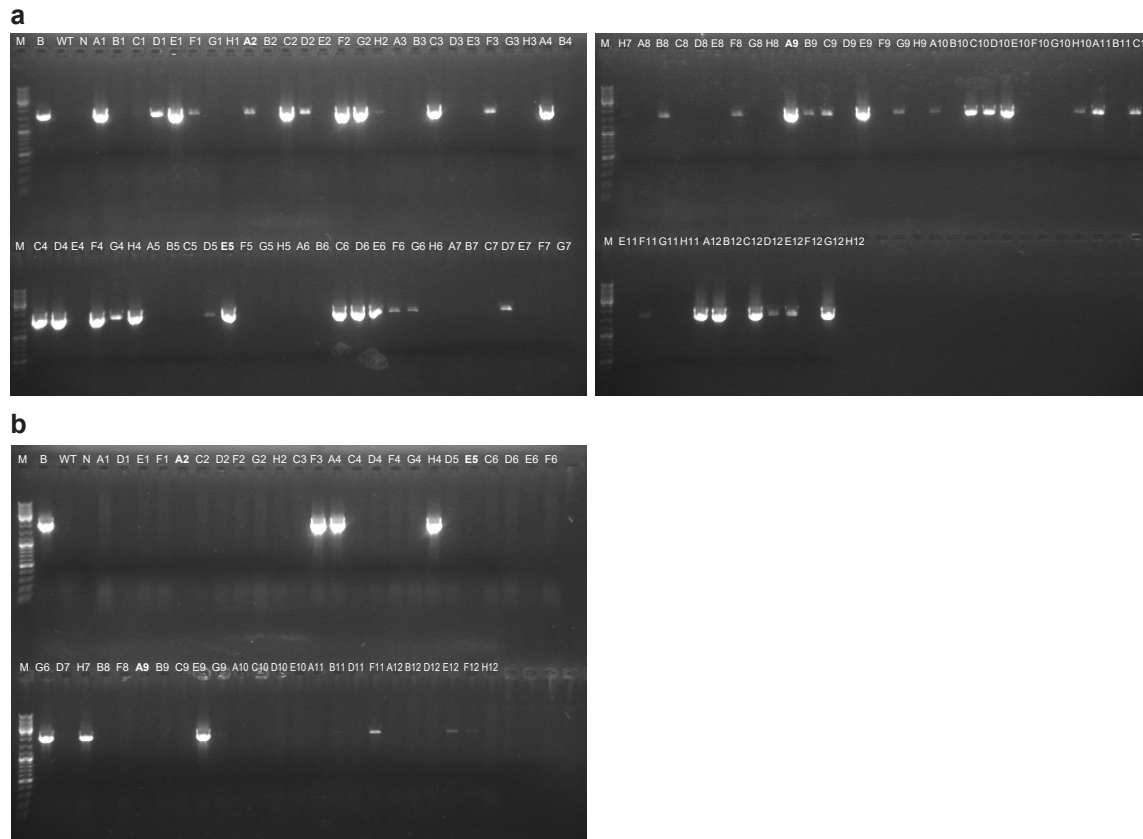
SUPPLEMENTARY FIGURES

Supplementary Fig. 1. Genotyping of CG cell lines
Supplementary Fig. 2. Verification of heterozygous genotypes
Supplementary Fig. 3. Expression of markers of the undifferentiated state in CG hiPSC lines
Supplementary Fig. 4. Engineered hiPSCs express degradable *CBFA2T3::GLIS2* (CG) and give rise to hematopoietic stem and progenitor cells (HSPC)
Supplementary Fig. 5. *CBFA2T3::GLIS2* (CG) long-term cell survival depends on cytokines and the oncogene
Supplementary Fig. 6. Copy number analysis of CG long-term cells
Supplementary Fig. 7. Transcriptional similarity between CG-hiPSC models, CG-AMKL and cell lines
Supplementary Fig. 8. *CBFA2T3::GLIS2*-dTAG-HA protein occupies reported genomic target regions
Supplementary Fig. 9. Opening and closing chromatin regions bound by *CBFA2T3::GLIS2* (CG) are enriched for different TF motifs
Supplementary Fig. 10. Examples of variable chromatin accessibility and transcription at selected genomic loci and its global correlation
Supplementary Fig. 11. Quality control of scRNA-seq data
Supplementary Fig. 12. Control MKs from our model are highly similar to fetal liver MKs
Supplementary Fig. 13. Myeloid cell types, markers, hierarchies, trajectories and MegaCult CFU-Mks of controls
Supplementary Fig. 14. Selection of MK-like cell subsets, examples of single cell gene expression patterns and multi-ome correlations
Supplementary Fig. 15. Degradation of *CBFA2T3::GLIS2* (CG) results in depletion but not differentiation of CG cells
Supplementary Fig. 16. Transcription factor footprint analysis

SUPPLEMENTARY INFORMATION

CG donor vector sequence for *in trans* paired nicking.

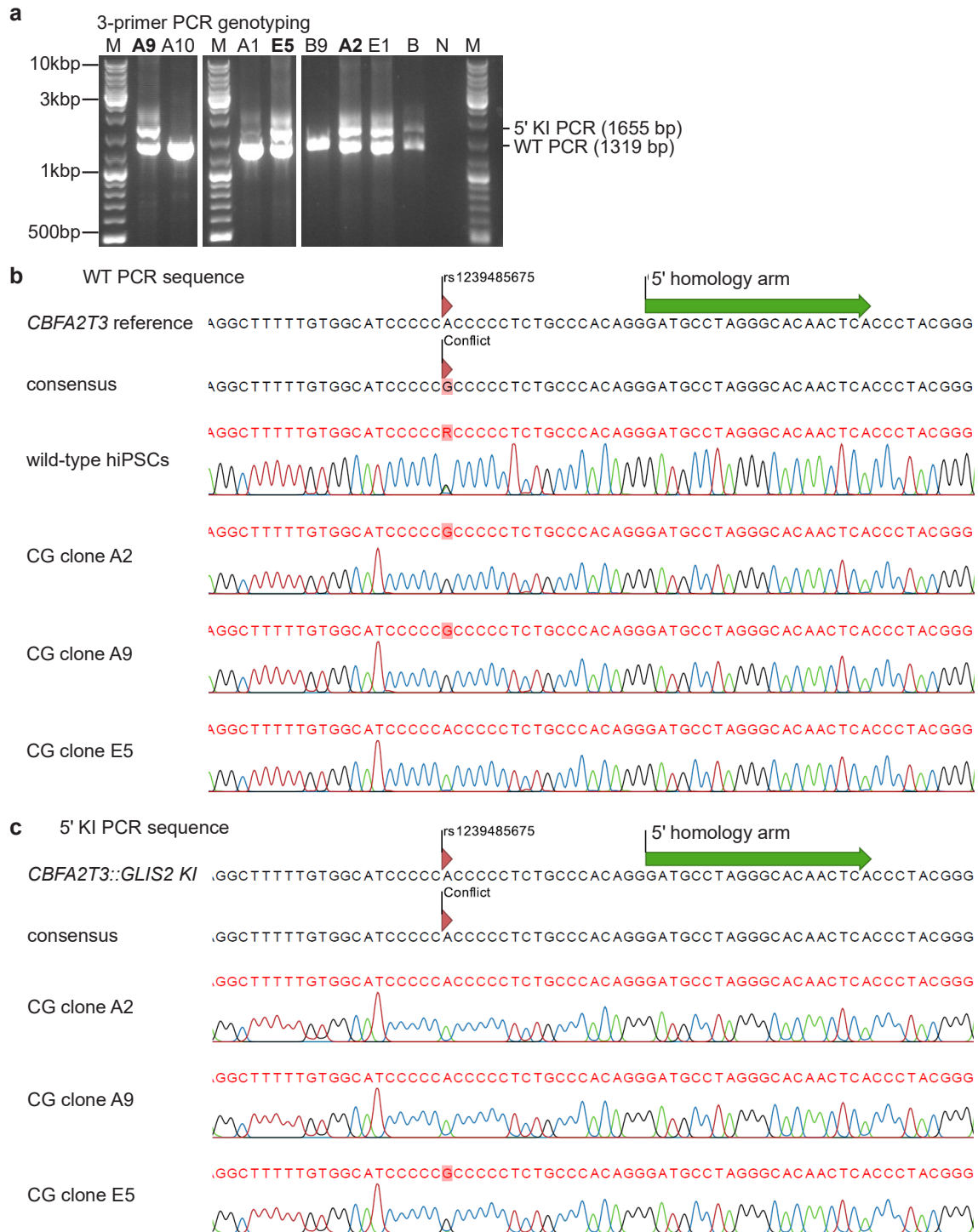
SUPPLEMENTARY REFERENCES



Supplementary Fig. 1. Genotyping of CBFA2T3::GLIS2 (CG) cell lines

a: PCR genotyping of delta (i.e., excised floxed puromycin resistance cassette) CG knock-in (KI) human induced pluripotent stem cell (hiPSC) clones. 96 clones were screened by 3' delta KI PCR and agarose gel electrophoresis. M, molecular weight marker (PeqLab DNA ladder mix); B, bulk of CG knock-in hiPSCs; WT, wild-type hiPSCs; N, no template control; A1 - H12, CG KI clones, pure heterozygous clones in bold. Half of the clones were positive for the expected product (1555 base pair, bp) and further characterized.

b: 3' delta KI PCR positive hiPSC clones were further analyzed with the 3' floxed KI PCR to detect contamination with the non-recombined KI allele. 11 of 48 clones showed traces of 3' floxed KI PCR product (1935 bp) and were discarded.



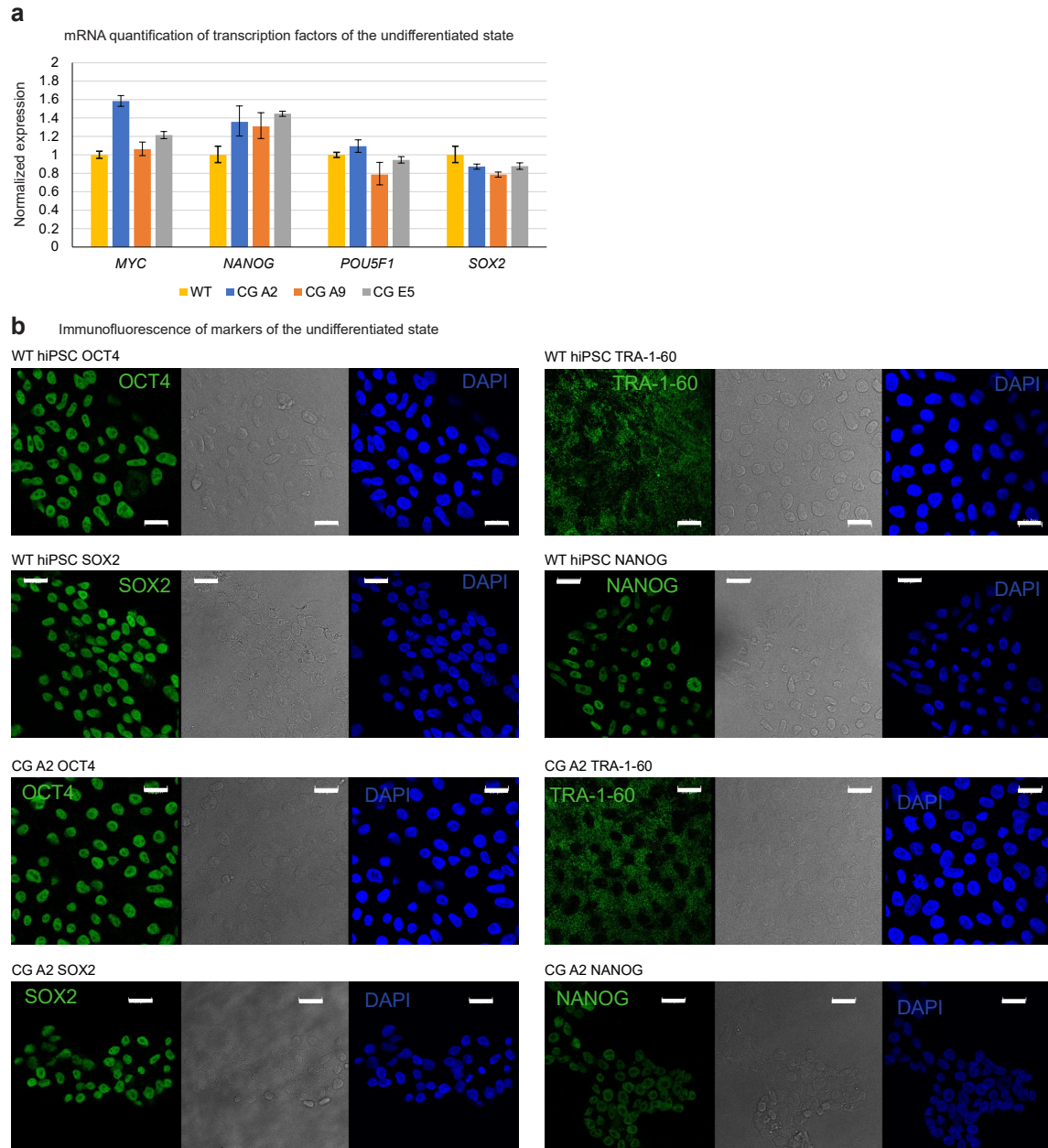
Supplementary Fig. 2 Verification of heterozygous CG genotype

3-primer PCR and Sanger-sequencing were performed with CG KI hiPSC clones to identify pure heterozygous genotypes.

a: Representative agarose gel pictures of 3-primer PCR products. Clones contaminated with wild-type (WT) cells, for example clone A10, were identified by a significantly stronger band for the WT PCR (1319 bp) than for the 5' KI PCR (1655 bp), and by Sanger-sequencing of the PCR products. M, molecular weight marker (PeqLab DNA ladder mix); B, bulk of CG knock-in hiPSCs; N, no template control; A1, ..., E5, CG KI clones; pure heterozygous clones in bold.

b: WT PCR Sanger-sequencing traces of parental hiPSCs and CG clones were aligned to the *CBFA2T3* reference sequence. WT cells were heterozygous for SNP rs1239485675 upstream of the 5' homology arm while clones A2, A9 and E5 were clearly homozygous.

c: 5' KI PCR Sanger-sequencing traces of CG clones were aligned to the anticipated *CBFA2T3::GLIS2* KI locus. Clones A2, A9 and E5 were homozygous for the respective other rs1239485675 SNP.

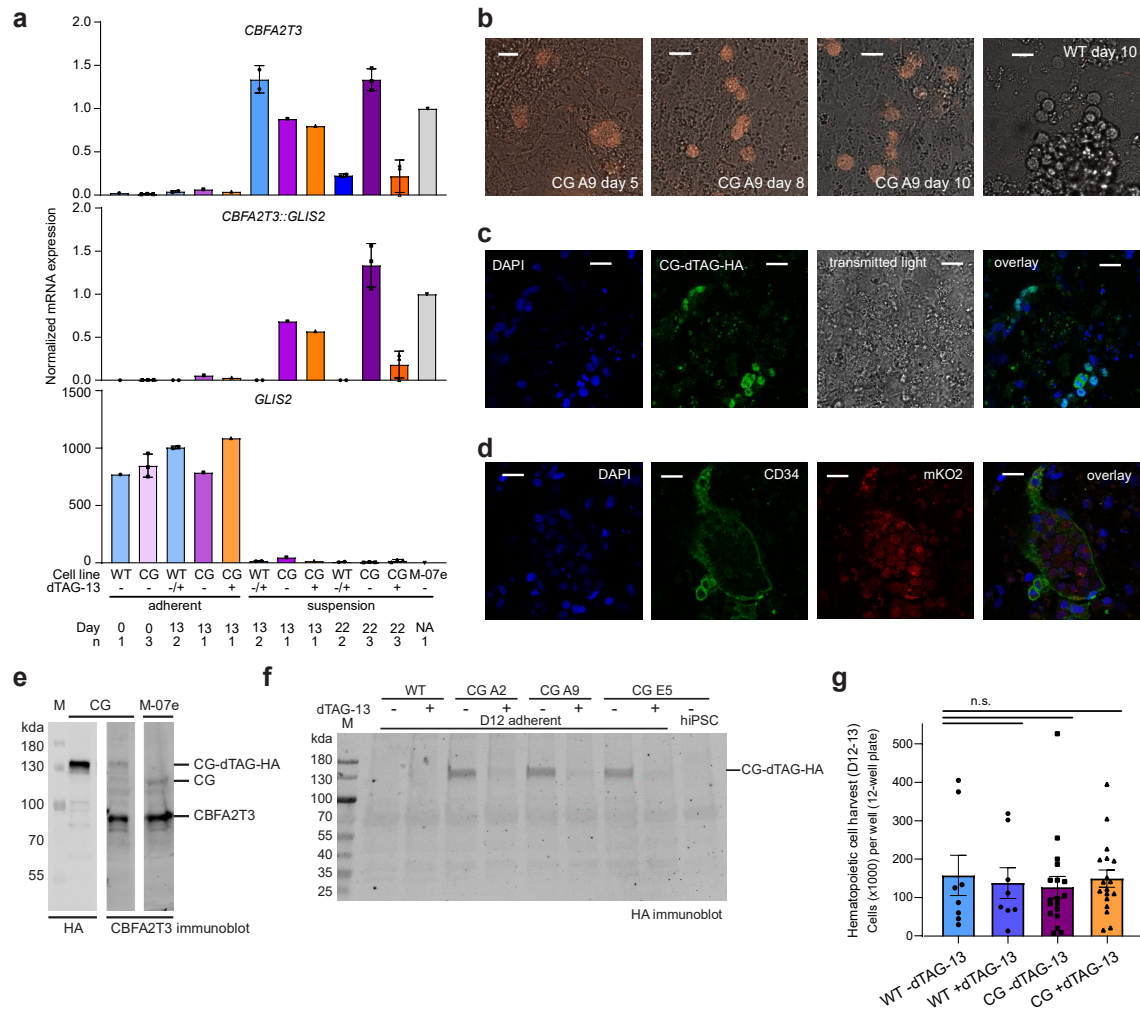


Supplementary Fig. 3. Expression of markers of the undifferentiated state in CG hiPSC lines

Factors of the undifferentiated state were expressed in WT and CG hiPSC clones on RNA (a) and protein (b) level.

a: mRNA expression of *MYC*, *NANOG*, *POU5F1* (*OCT4*), and *SOX2* in CG clones was determined by quantitative reverse transcription PCR after normalization to housekeeping genes *GUSB* and *ABL1* and to WT hiPSCs using the $2^{-\Delta\Delta C_t}$ method (technical triplicates; mean fold change with error bars representing 95% confidence intervals).

b: Nuclear OCT4, SOX2, and NANOG as well as glycosylated podocalyxin (TRA-1-60) cell surface marker (in green) were detected in parental and CG (representative clone A2 shown) by immunofluorescence of fixed hiPSCs with DAPI-counterstained nuclei (blue) and transmitted light pictures in grey. White scale bars = 20 μ m.



Supplementary Fig. 4. Engineered hiPSCs express degradable CG protein and give rise to hematopoietic stem and progenitor cells (HSPC)

a: Expression of *CBFA2T3*, *CG*, and *GLIS2* (top to bottom) transcripts in WT and CG clone A9 (CG) at the undifferentiated stage (day 0), and on days 13 (adherent and suspension fraction) and 22 of differentiation with (+) or without (-) dTAG-13 degrader, and in M-07e cells as indicated (NA, not applicable). Quantitative reverse transcription PCR results were normalized to housekeeping genes *GUSB* and *ABL1*, and to M-07e using the $2^{-\Delta\Delta Ct}$ method (n, biological replicates each analyzed as technical triplicates; mean fold change $\pm$ SD).

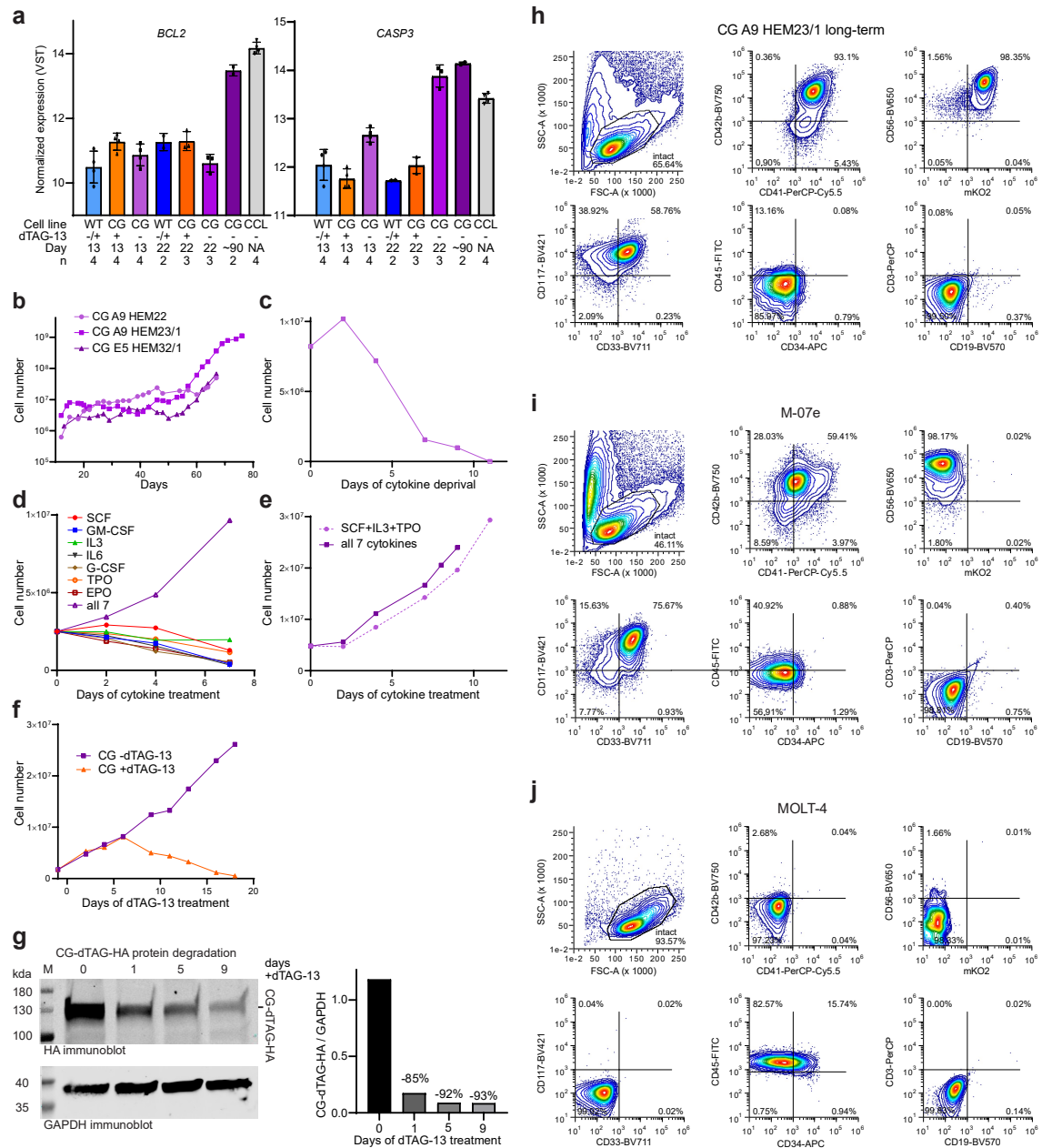
b: Activation of the CG KI locus was monitored by live cell imaging of mKO2 fluorescence. From differentiation day 5 on the reporter was detected in a fraction of cells, while it was not in WT negative controls. White scale bars = 20 μ m.

c,d: Immunofluorescence of CG cells for CG-dTAG-HA (**c**; green) or CD34 (**d**; green), mKO2 (**d**; red) and DAPI (blue). White scale bars = 20 μ m.

e: Western blot with HA and CBFA2T3 (recognizing its middle part included in CG) antibodies showing protein expression of WT CBFA2T3 and CG-dTAG-HA in long-term CG A9 HEM22 or WT CG in M-07e cells. Estimated molecular weights in kilodaltons (kda) are shown for the pre-stained marker (M). **f:** Western blot with fusion specific HA antibody showed expression of CG-dTAG-HA protein in untreated (-) adherent CG cells on day 12 of differentiation, and its efficient degradation upon continuous treatment with 100 nM dTAG-13 (+). CG was also not expressed in differentiated WT controls and CG hiPSCs.

g: HSPC (harvested supernatant and weakly attached live cells) yields of WT and CG lines determined on D12-13 with (+) or without (-) dTAG-13 degrader. No significant (n.s.) difference

(one-way ANOVA with Dunnett's post hoc test) of yields (mean $\pm$ SEM) was observed between WT (n=8 each for + and - dTAG-13) and CG (3 different clones; n=18 each for + and - dTAG-13). On average 1.5×10^5 cells were harvested per well of a 12-well plate.



Supplementary Fig. 5. CBFA2T3::GLIS2 (CG) long-term cell survival depends on cytokines and the oncogene

a: *BCL2* (left) and *CASP3* (right) levels as determined by RNA-seq (log₂ expression normalized by variance-stabilizing transformation, means $\pm$ SD) for wild-type (WT) and CG cells on D13, D22 or in long-term culture $\pm$ dTAG-13. Replicate numbers were n=4 each for WT D13 ($\pm$ dTAG-13 combined), CG D13, CG D13 +dTAG-13 and patient-derived cancer cell lines (CCL; M-07e, M-MOK, WSU-AML, CMS; n=1 each); n=3 each for CG D22 and CG +dTAG-13; and n=2 each for WT D22 ($\pm$ dTAG-13 combined).

b: CG-expressing HSPCs were harvested on D12-13 and continuously cultured for another 55-65 days in StemPro-34 medium containing IL3, IL6, GM-CSF, G-CSF, SCF, EPO and TPO. Growth curves show extrapolated total live cell numbers over culture time. Three outgrowing long-term cell lines (CG A9 HEM22, CG A9 HEM23/1 and CG E5 HEM32/1) were used for further experiments.

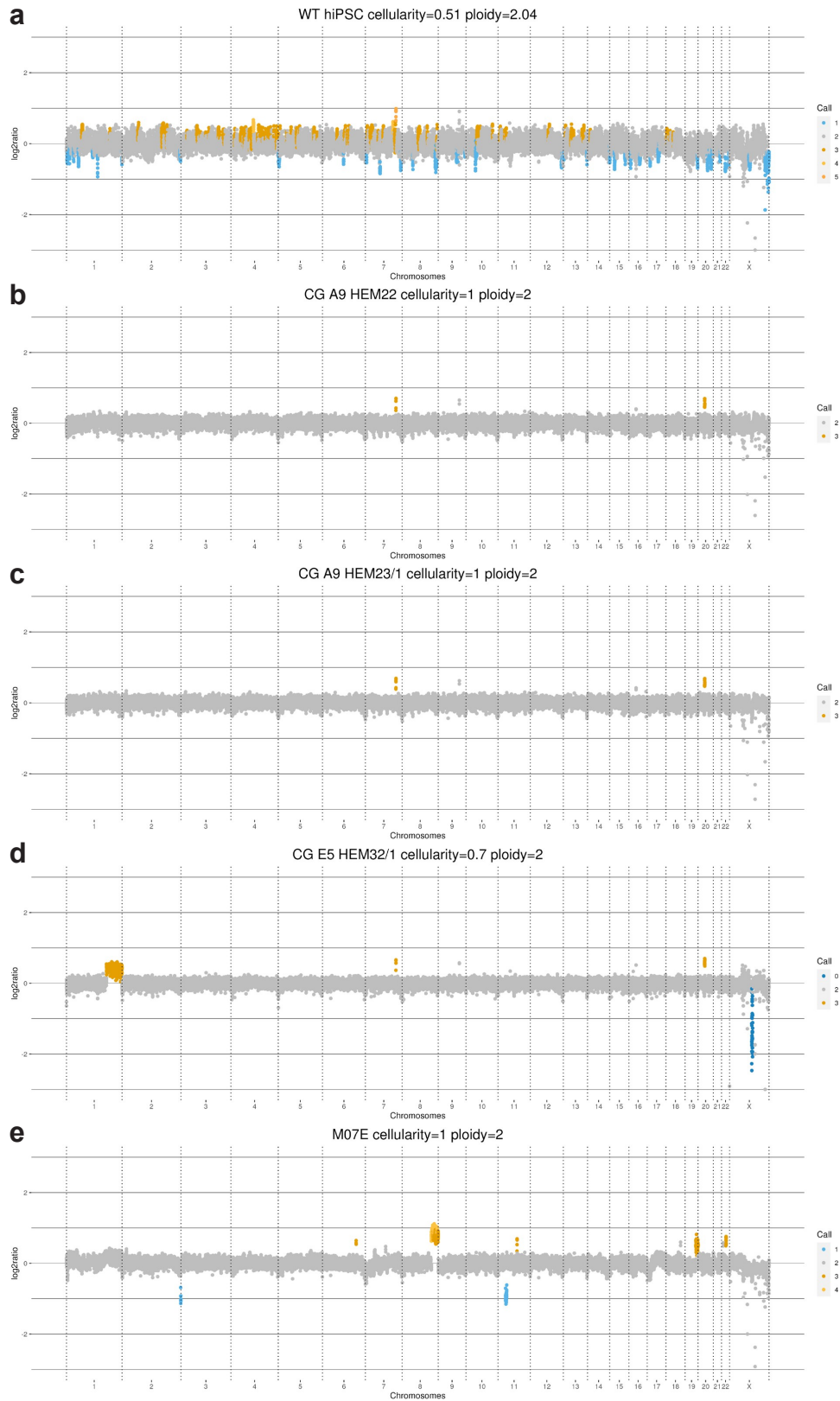
c-f: Live cell numbers after cytokine or CG withdrawal measured by trypan blue exclusion staining and counting (n=1). CG live cell number decreased after cytokine withdrawal (**c**) and no single cytokine (IL3, IL6, GM-CSF, G-CSF, SCF, EPO, or TPO) was able to stimulate growth (**d**). However, single SCF, IL3 or

TPO allowed for survival of some cells (**d**), and their combination stimulated growth albeit at a lower rate than the cocktail of all 7 cytokines (**e**).

f: Treatment of long-term CG cultures with 100 nM dTAG-13 degrader reduced live cell numbers after about one week (one representative of three experiments with similar results is shown).

g: CG-dTAG-HA protein decreased gradually upon dTAG-13 treatment in CG long-term cells according to Western blotting. Fusion and housekeeping protein were detected with HA and GAPDH antibody, respectively (left). GAPDH-normalized CG expression and its percentual decrease were quantified (right).

h-j: Cell surface marker profile determined by flow cytometry for CG A9 HEM23/1 long-term cells (**h**; CD41⁺CD42b⁺CD56⁺CD33⁺CD117⁺CD34⁻CD45^{low}CD19⁻CD3⁻), control CG-AMKL cell line M-07e (**i**; CD41⁺CD42b⁺CD56⁺CD33⁺CD117⁺CD34⁻CD45^{low}CD19⁻CD3⁻) and CG-negative T-ALL cell line MOLT-4 (**j**; CD41⁻CD42b⁻CD56⁻CD33⁻CD117⁻CD34^{low}CD45⁺CD19⁻CD3⁻). The same flow cytometer settings were used; FSC-A, forward scatter area; SSC-A, side scatter area; mKO2, monomeric Kusabira Orange 2.



Supplementary Fig. 6. Copy number analysis of CG long-term cells

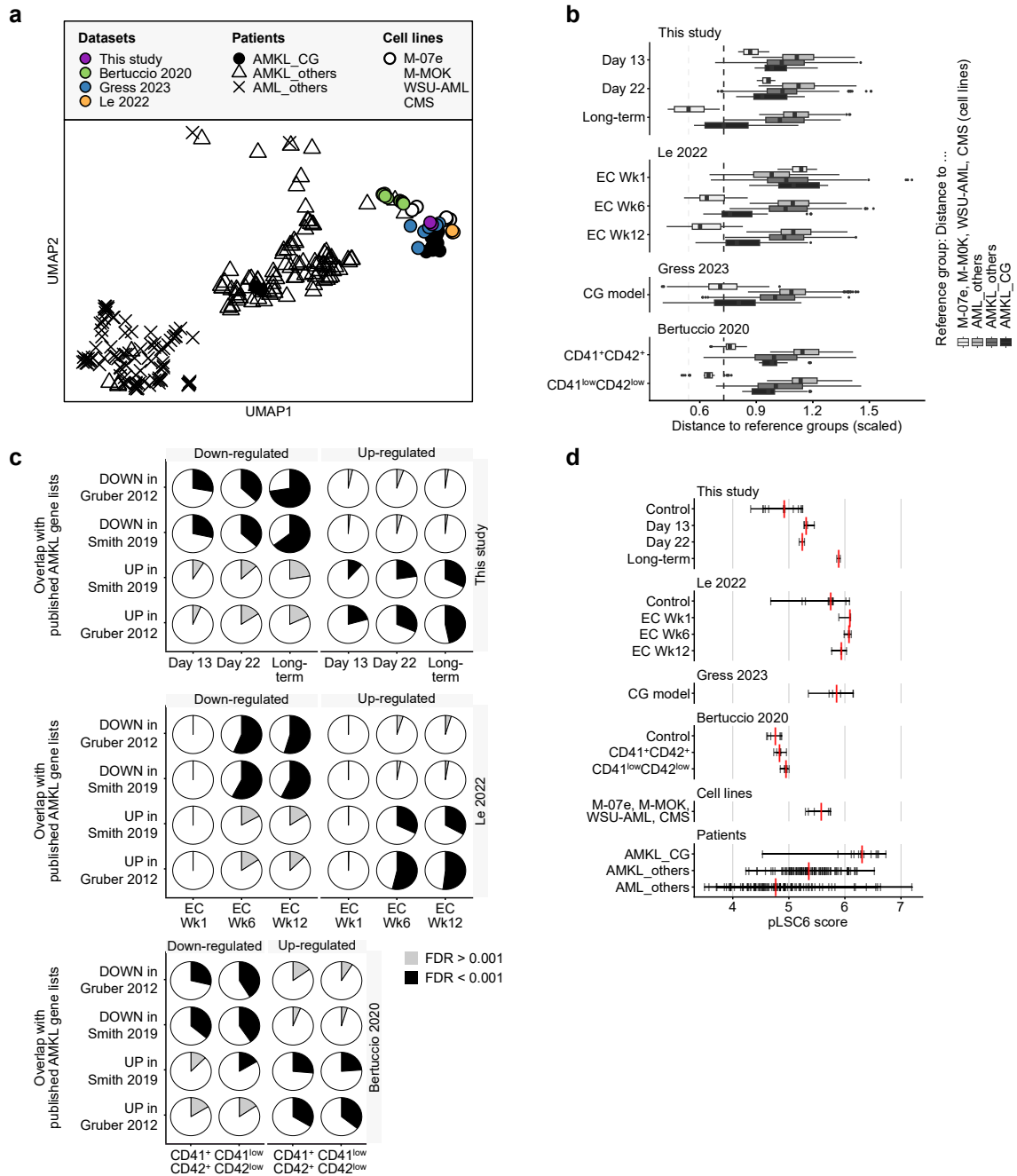
Copy number alterations (CNA) in whole genome sequencing¹ (a) or our ChIP-seq input (b-e) data were analyzed using ACE.² Log₂ ratios related to two expected copies of 100 kbp chromosomal regions are

plotted, significant gains are colored in orange, losses in blue. Parental hiPSC (**a**) and M-07e (**e**) were used as CNA negative and positive controls, respectively.

a: Analysis of published hiPSC data¹ called several CNAs despite an average coverage of 40x. The signal was unexpectedly noisy supposedly due to bias introduced during library construction. CG A9 HEM22 (**b**), CG A9 HEM23/1 (**c**) and CG E5 HEM32/1 (**d**) long-term cell lines exhibited a small gain of an approx. 1 Mbp pericentromeric region on chr20q11.21 including the *BCL2L1* gene. This CNA frequently occurs in pluripotent stem cell lines as well as in AML samples.³ A gain of a small chr7q region was constitutive since already present in parental hiPSCs (**a**).

d: CG E5 HEM32/1 additionally harbored a partial amplification of chr1q including *MDM4* which was recently shown to arise in feeder-free pluripotent stem cell cultures.⁴

e: M-07e harbored reported copy number alterations (e.g., gains of parts of 8q including *MYC* and 19q; https://cog.sanger.ac.uk/cmp/download/SIDM01640_Broad.PureCN_GATK.WES.panel.cnv.pdf). CNAs and involved genes are listed in **Supplementary Data 2**.



Supplementary Fig. 7. Transcriptional similarity between CG-hiPSC models, CG-AMKL and cell lines

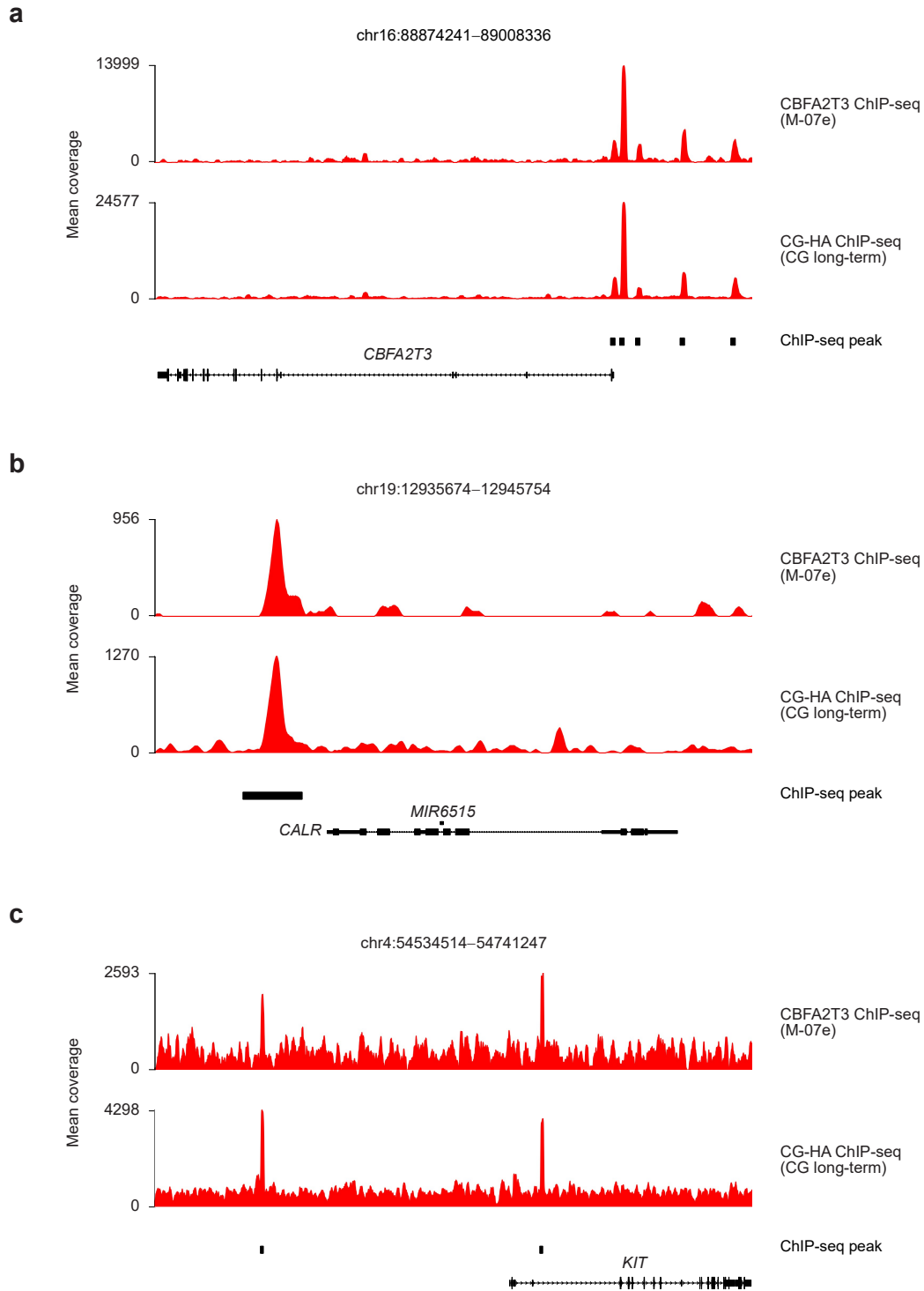
a: UMAP showing RNA-seq data for AMKL patients (CG and other subtypes) and other AML patients (from St. Jude Cloud), CG-AMKL patient-derived cell lines (M-07e, M-MOK, WSU-AML, CMS), long-term CG cells (purple), and three other AMKL CG models (in green, blue, and orange).^{5, 6, 7} Our long-term cells cluster with CG patient samples and cell lines.

b: PCA distance to reference groups AMKL (CG and other subtypes), other AML patients, and CG derived cell lines (boxplots in different shades of grey) was calculated for our D13, D22, and long-term CG cells and for three other AMKL CG models.^{5, 6, 7} Distance is defined as the Euclidean distance over the first 50 principal components, scaled by the root mean square.

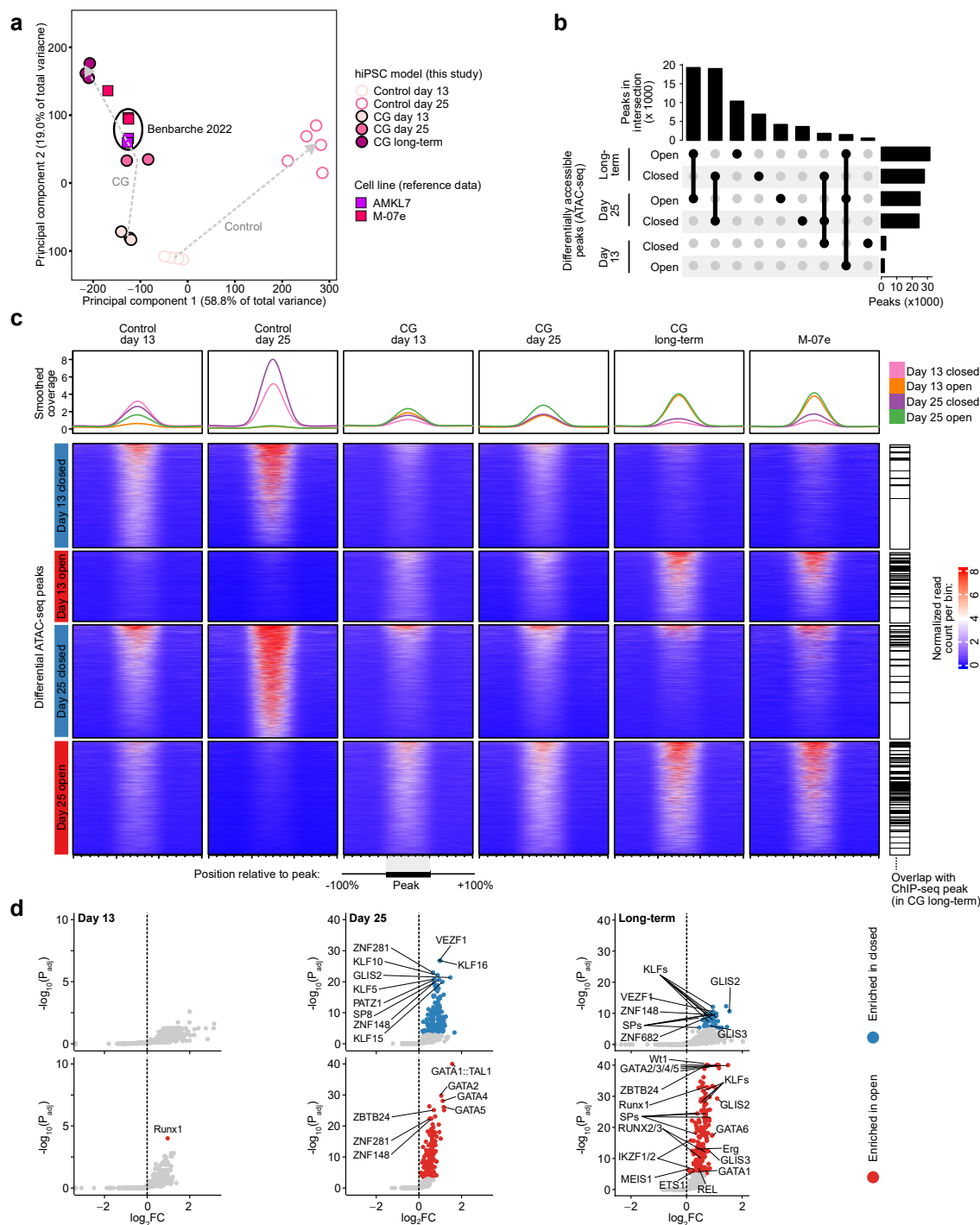
c: Overlap of up- and downregulated genes in D13, D22, or long-term CG cells (compared to controls at D13) in our and two other AMKL-CG models^{5, 7} (compared to their respective controls) from gene sets of CG-AMKL compared to other AML patients of two publications.^{8, 9} The significance of the overlap

was determined by hypergeometric test and significant overlaps (false discovery rate [FDR] ≤ 0.001) are indicated by black pie chart segments.

d: Pediatric leukemia stem cell scores (pLSC6)¹⁰ are shown for controls, D13, D22 and long-term CG cells, and three other AMKL CG models,^{5, 6, 7} AMKL (CG and other subtypes) and other AML patient samples, and CG cell lines. Each segment indicates one sample, and the median is indicated by a red line.



Supplementary Fig. 8. CBFA2T3::GLIS2-dTAG-HA protein occupies reported genomic target regions
 Genomic mean coverage and called peak regions determined by chromatin immunoprecipitation sequencing (ChIP-seq) of M-07e with CBFA2T3 antibody (binding both WT and CG protein, upper graph) and CG long-term cells with HA antibody (binding only tagged CG, lower graph) are shown for the *CBFA2T3* promoter and upstream enhancers (**a**), *CALR* upstream enhancer (**b**), and *KIT* upstream and intronic enhancers (**c**).^{11, 12}



Supplementary Fig. 9. Opening and closing chromatin regions bound by CBFA2T3::GLIS2 (CG) are enriched for different transcription factor (TF) motifs

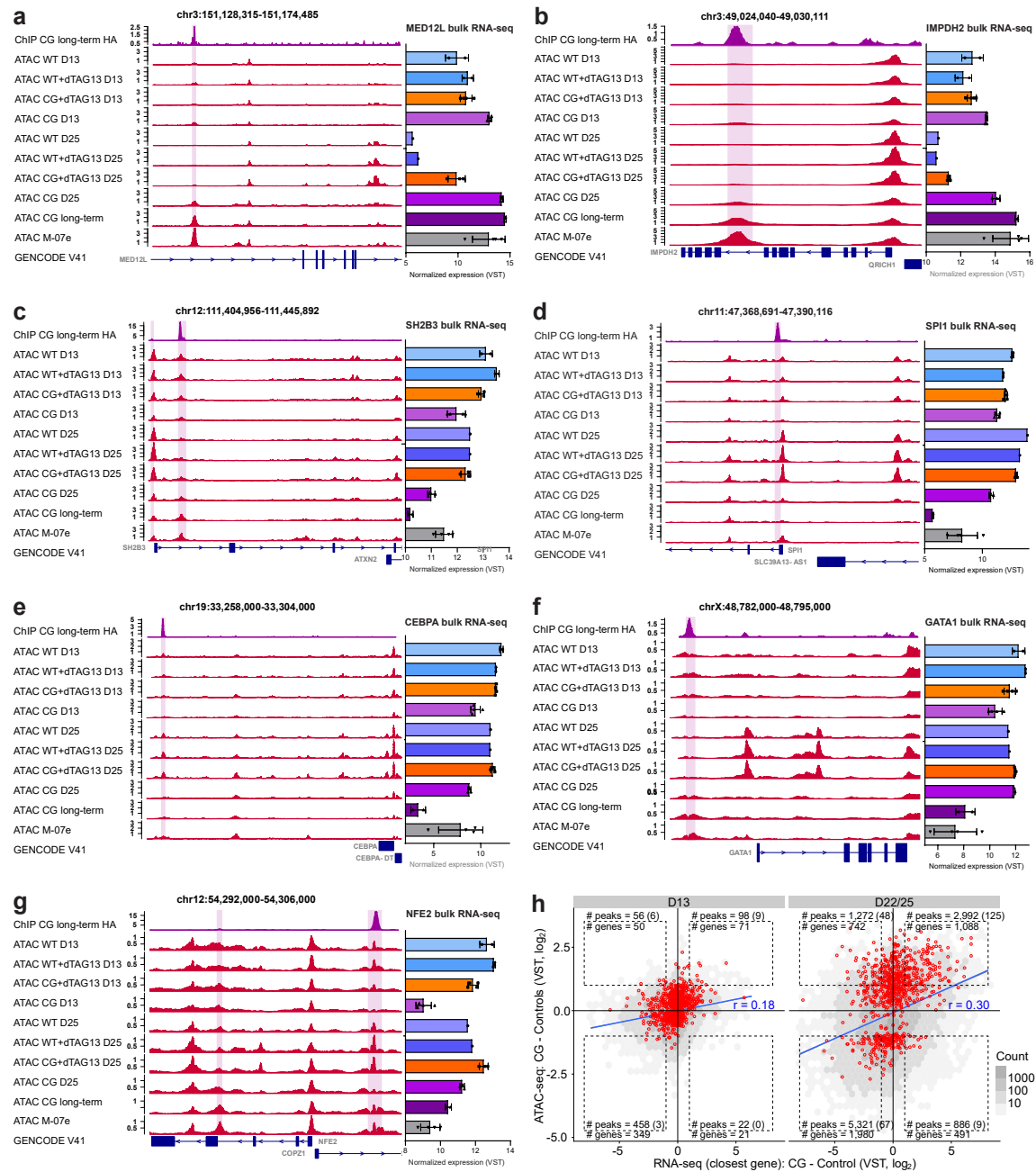
a: PCA plot of ATAC-seq data of D13, D22, or long-term CG cells and controls, and of M-07e from our study, and M-07e and AMKL7 from a previously published study (samples from the other study have been circled in the plot).¹¹ Dashed arrows have been inserted manually to represent the differentiation direction of samples from the same cell line.

b: Plot indicating the number and overlap of differentially accessible regions (DAR) that are significantly more (open) or less (closed) accessible in CG cells at different timepoints compared to controls. Each bar chart reports the number of peaks in the overlap (top) or in any peak dataset (right), and the dots and lines in the middle indicate different combinations of peak datasets. Rows/columns have been

ordered by set size. All peaks and the complete results of the differential accessibility analysis are available in **Supplementary Data 3 and 4**.

c: Heatmaps indicating the coverage (normalized read counts) in genomic bins centered on closed (top) and open (bottom) DARs in D13 or D25 CG cells (top 3 000 regions sorted by p-value; if there were less than 3000 significant DARs, only those DARs are shown) with flanking regions of the same width left and right. Red color indicates high coverage. The plots (from left to right) report the coverage of ATAC-seq data from control (WT±dTAG-13, CG+dTAG-13) and CG (-dTAG-13) at different time points. The line plots on top indicate the mean coverage in each sample across all open (red) and closed (blue) DARs. The barcode-like lines on the right indicate overlaps with HA-ChIP-seq peaks from long-term CG cells. The plots illustrate a gradual change in chromatin accessibility throughout differentiation away from control cells and toward the chromatin state of long-term CG cells. For many opening chromatin regions, there is evidence of direct CG-binding in long-term CG cells.

d: Volcano plot indicating the enrichment of matches with DNA sequence motifs of TFs from the JASPAR database¹³ in closed (top) and open (bottom) DARs in D13, D25 and long-term CG cells. Enrichment was determined by hypergeometric test and the plots indicate the odds ratio and adjusted p-value compared to the background frequency of motif matches in all ATAC-seq peaks. Significantly enriched motifs ($p_{adj} < 0.001$) are indicated by red/blue dots and selected TF motifs have been labeled. At D13, mouse Runx1 was the only significantly enriched TF motif.

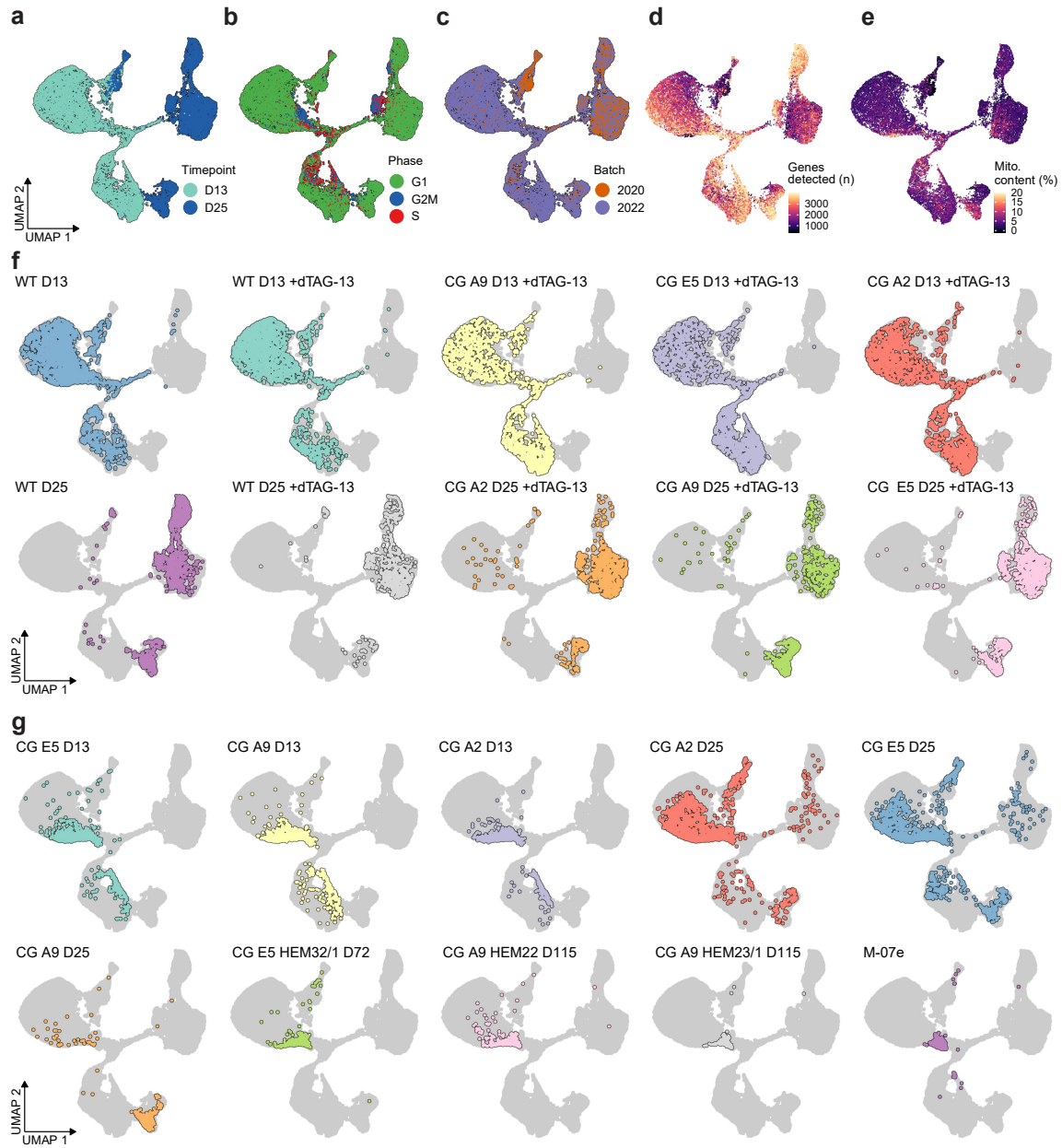


Supplementary Fig. 10. Examples of variable chromatin accessibility and transcription at selected genomic loci and its global correlation

a-g: CG-dTAG-HA ChIP-seq (purple track on top), ATAC-seq (10 conditions, red tracks) and normalized RNA-seq data of same conditions (bar charts to the right; log₂ expression normalized by variance-stabilizing transformation, means ± SD; cancer cell line [M-07e, M-MOK, WSU-AML, CMS] expression in grey) are shown for candidate target loci (exon structure in dark blue on bottom). *MED12L* (a), *IMPDH2* (b), *SH2B3* (c), *SPI1* (d), *CEBPA* (e), *GATA1* (f) and *NFE2* (g) serve as examples. High confidence CG-bound regions have been highlighted with purple overlays. Generated with Gviz and GraphPad Prism.

h: Density scatterplots comparing changes in RNA expression (x-axis) and chromatin accessibility (y-axis) between CG-expressing and control cells at day 13, and at day 22 (RNA-seq) / 25 (ATAC-seq). The density of data points (peak/gene pairs) in the plot is indicated by the shading of hexagons. Peaks overlapping CG-binding peaks (HA-ChIP-seq in long-term CG cells) are highlighted as red circles. The

blue lines represent a linear fit, and the Pearson correlation coefficients (r) are indicated. The dashed boxes enclose peak/gene pairs with an absolute fold change ≥ 2 and we indicated the number of peaks (number in parentheses = peaks overlapping CG binding sites) and genes in each sector. Globally, there is a weak positive correlation between RNA and accessibility.

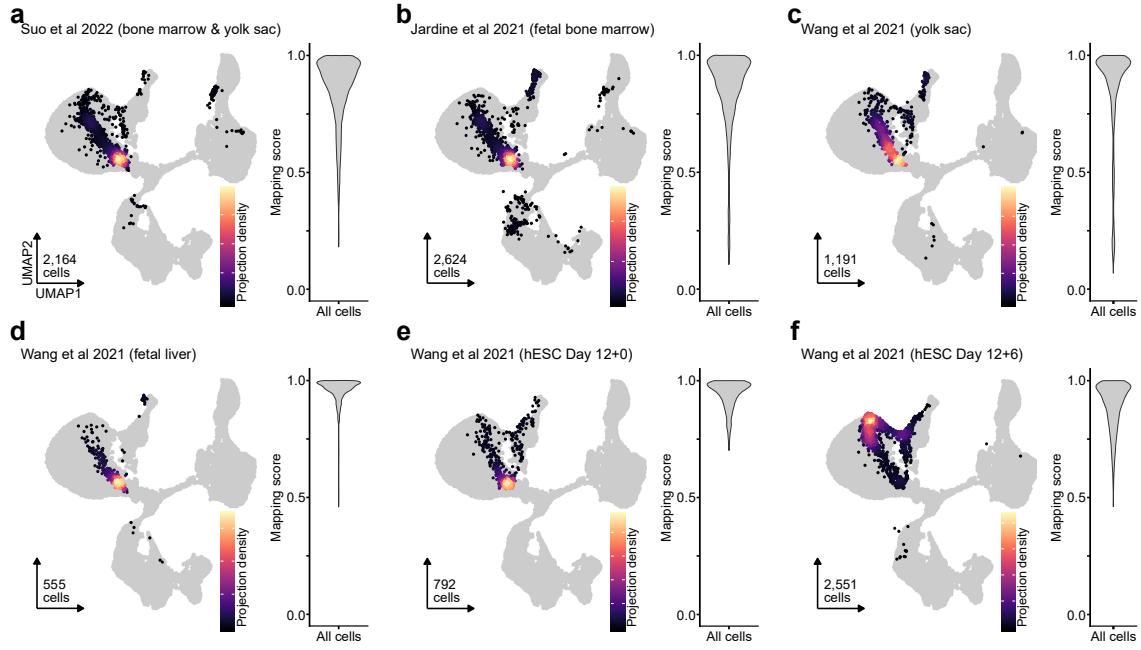


Supplementary Fig. 11. Quality control of scRNA-seq data

a-e: UMAP plots showing the time point of sample collection (**a**), estimated cell cycle phase (**b**), sequencing batch (**c**), number of detected genes (**d**), and percentage of reads from mitochondrial genes (**e**) for each cell in the control dataset (WT, WT+dTAG-13, CG+dTAG-13).

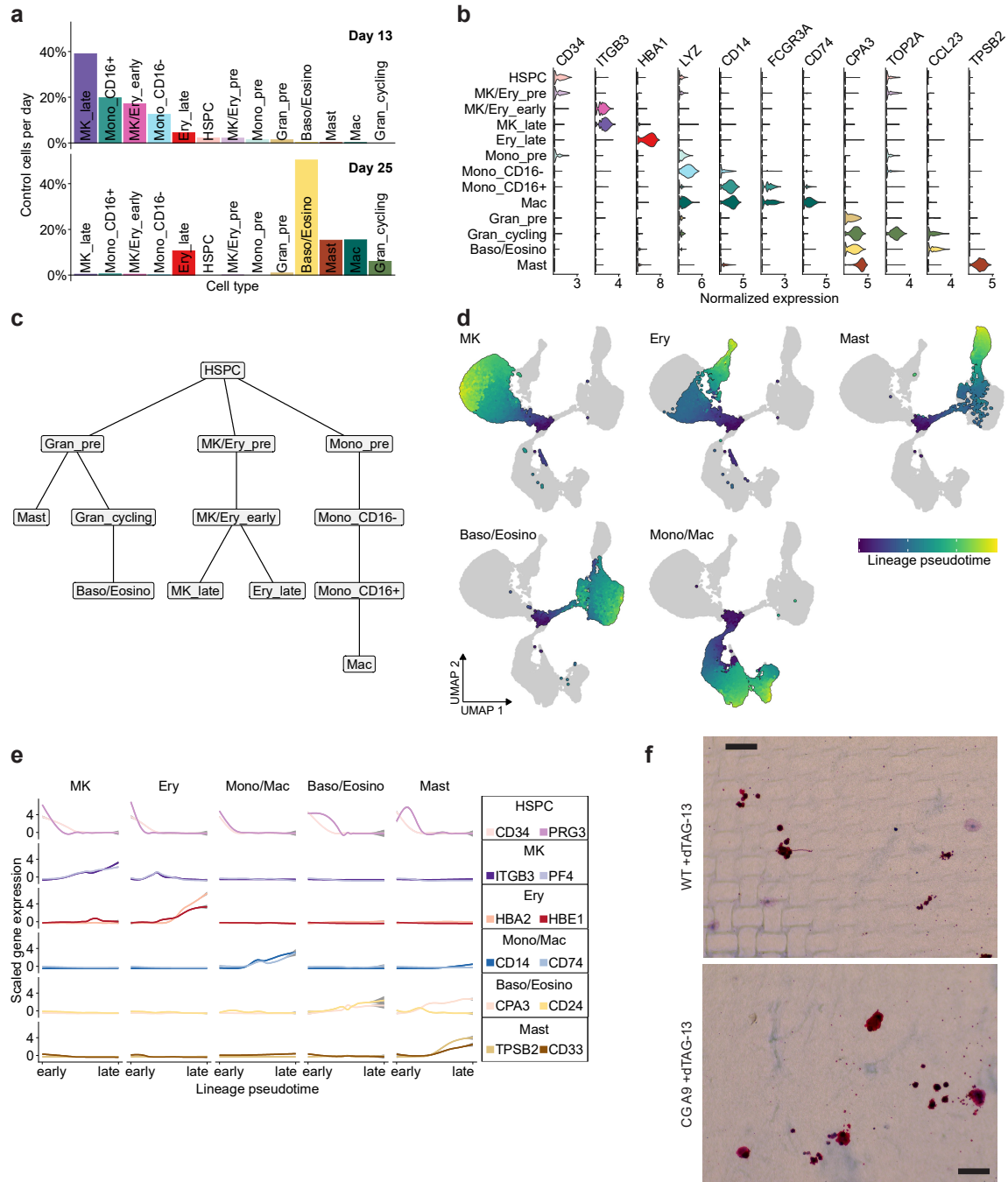
f: UMAP plots showing the distribution of cells from each control sample. An interactive version of this plot is available at https://vitessce.io/#?url=https://fh.ccir.at/zarr/srt_ctrl.h5ad.zarr/config.json.

g: Plots showing the distribution of cells from each untreated CG sample that were projected onto the control UMAP.



Supplementary Fig. 12. Control MKs from our model are highly similar to fetal liver MKs

Published scRNA-seq data of MKs derived from human fetal bone marrow (**a,b**), yolk sac (**a,c**) or liver (**d**), or from human embryonic stem cells differentiated using a distinct protocol for 12 or 18 days^{14, 15, 16} were bioinformatically projected onto our control UMAP (indicated in grey; **Fig. 3a**). The density of cells mapped to each location in the UMAP is shown with brighter colors corresponding to higher density. The distribution of mapping scores is shown as a violin plot on the right.



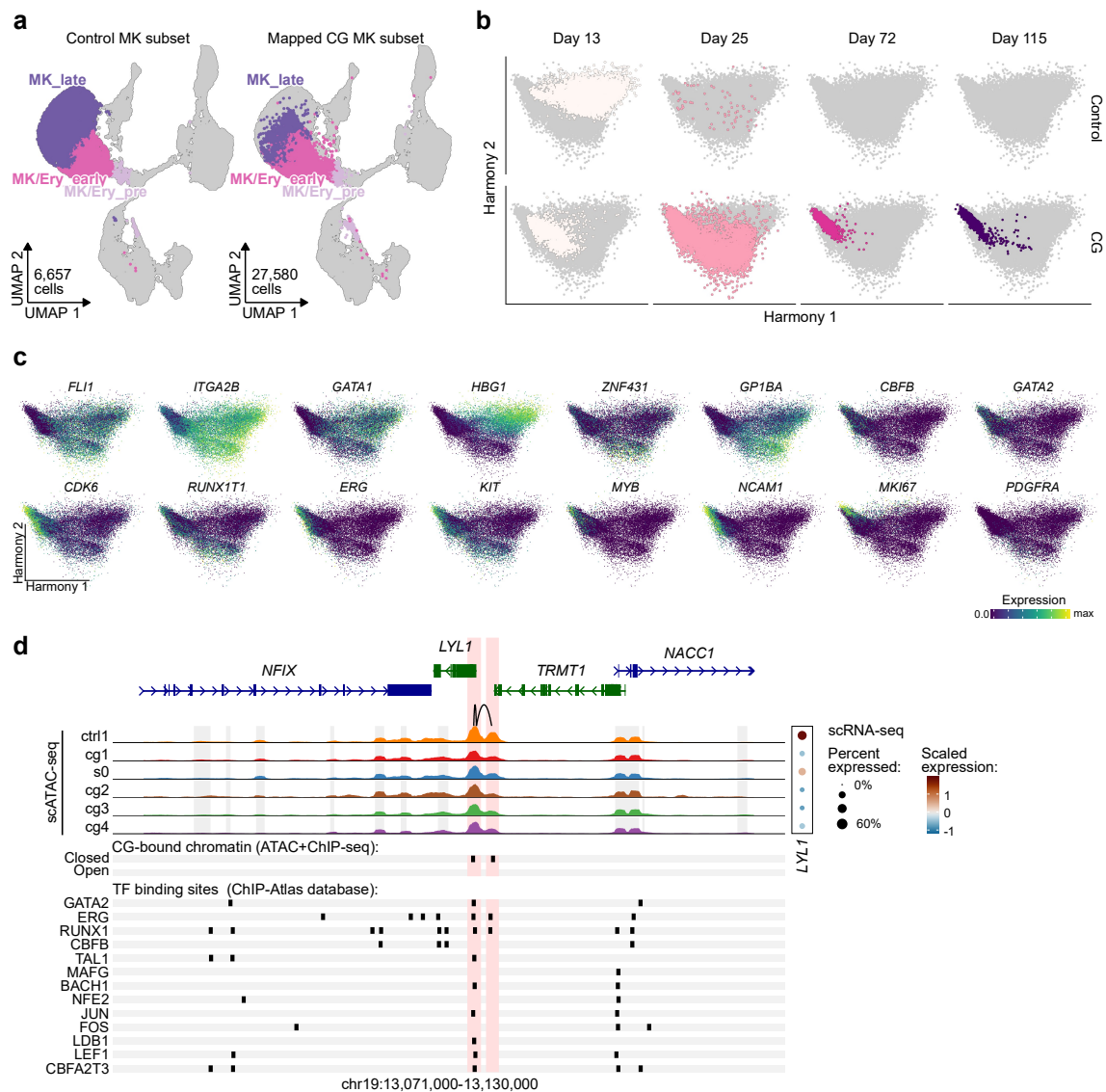
Supplementary Fig. 13. Myeloid cell types, markers, hierarchies, trajectories, and CFU-Mks of controls

a: Bar chart showing the percentage of cells from control samples collected at D13 (top) and D25 (bottom) that were assigned to each of the indicated cell types in **Fig. 3a**.
b: Violin plots indicating the distribution of normalized expression of selected marker genes in different cell types. All cluster markers are available in **Supplementary Data 5**.
c: Inferred cell type hierarchy based on the Minimum Spanning Tree from slingshot¹⁷ connecting underlying clusters. HSPC, hematopoietic stem and progenitor cell; Gran, granulocyte; pre, precursor; Baso/Eosino, basophil/eosinophil; Mast, mast cell; MK, megakaryocyte; Ery, erythrocyte; Mono, monocyte; Mac, macrophage.

d: UMAPs showing the pseudotime score for each lineage-specific trajectory inferred using slingshot.¹⁷ HSPCs are situated in the center of these plots and each trajectory points toward different endpoints.

e: Scaled expression values of selected lineage marker genes along the pseudotime trajectories from panel **d**.

f: Microphotographs showing examples of MegaCult colony forming unit-megakaryocyte (CFU-Mk) for WT and CG (both dTAG-13-treated). CD41⁺ cells are visible in dark red, nuclei in blue. Scale bars = 100 μ m.



Supplementary Fig. 14. Selection of MK-like cell subsets, examples of single-cell gene expression patterns and multi-ome correlations

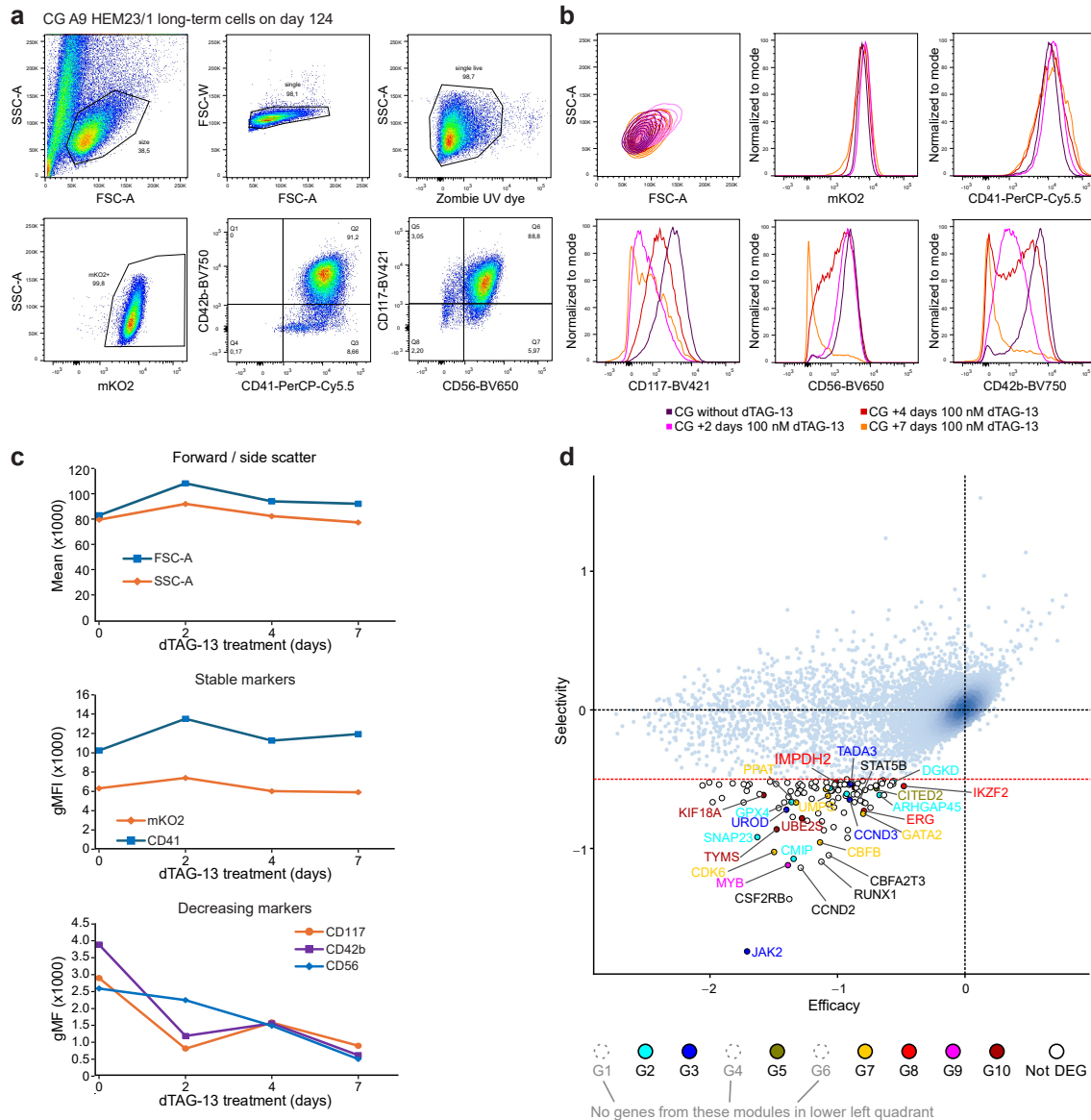
a: UMAP plots indicating the MK-like cell subsets in control and CG samples selected for analysis in Figs. 5-6.

b: Distribution of cells from different input datasets overlaid on the low-dimensional Harmony projection from Fig. 4a.

c: Harmony plots showing examples of expression of genes in modules G1-G10 from Fig. 5a. The normalized expression value of the gene in each cell is indicated by color. An interactive version of this plot is available at https://vitessce.io/#?url=https://fh.ccir.at/zarr/srt_mk.h5ad.zarr/config.json.

d: Representation of genomic coverage of scATAC-seq reads reporting chromatin accessibility at the *LYL1* locus for different cell clusters (s0, ctrl1, cg1-4; aggregated over all cells assigned to the respective cluster). Significant links between peak accessibility (left; scATAC-seq modality) and gene expression (right; scRNA-seq modality) based on Pearson correlation are shown. Binding sites of CG at open and closed chromatin (Fig. 2) and of other transcription factors in leukemia and blood-related cell lines (from the ChIP-Atlas database) are highlighted at the bottom. ATAC-seq peaks overlapping CG binding sites and/or linked to the indicated target genes are highlighted with red overlays. The promoter

chromatin is less accessible in cg clusters, bound by CG and other TFs, and correlates with *LYL1* gene expression.



Supplementary Fig. 15. Degradation of CBFA2T3::GLIS2 (CG) results in depletion but not differentiation of CG cells

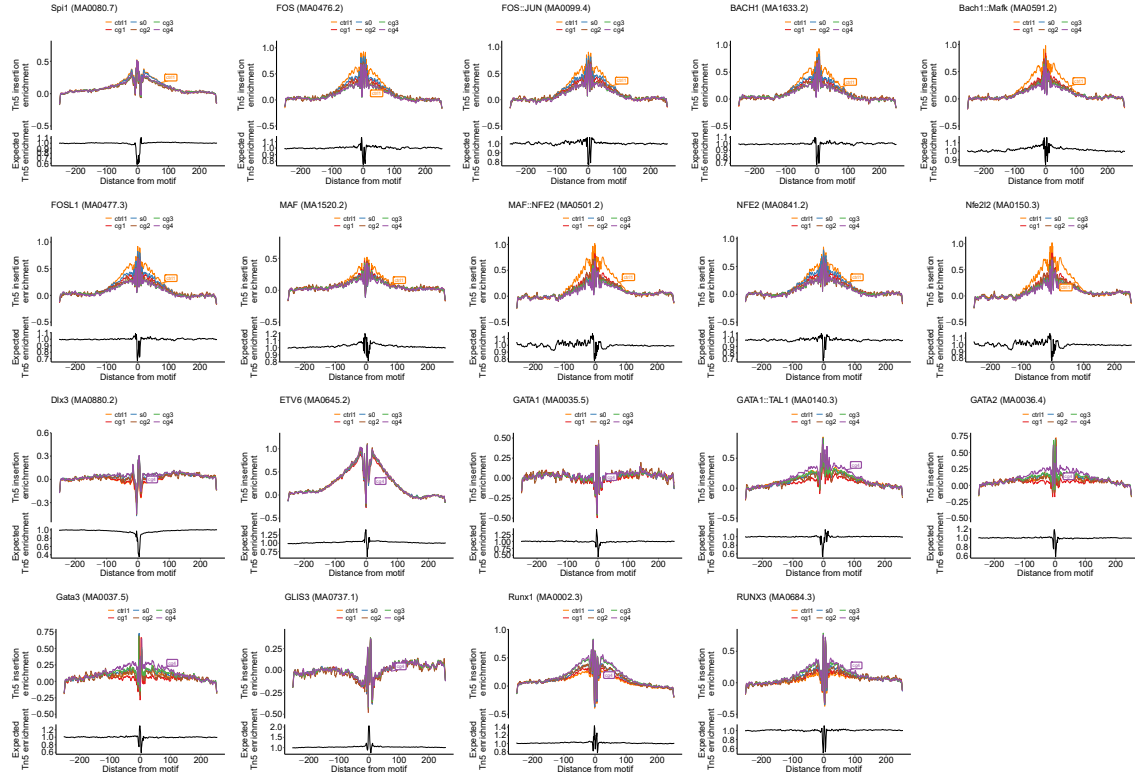
Flow cytometry analysis of CG cells cultured without (0 nM) or with 100 nM dTAG-13 for 2, 4 or 7 days. CG depletion resulted in a decrease of SCF receptor CD117 (KIT), MK marker CD42b (GP1BA) and NK marker CD56 (NCAM1; atypically expressed in CG-AMKL) expression. Results of one of two consistent experiments are shown.

a: Gating strategy according to normal size (forward and side scatter area; FSC-A and SSC-A), single (forward scatter width and area; FSC-W and FSC-A) and live cells (negative for Zombie UV dye dead cell stain). Colored dot plots of untreated CG long-term cells are shown as an example.

b: Overlay of FSC-A / SSC-A contour plots and compensated surface marker signal histograms.

c: Mean FSC-A, SSC-A, and compensated geometric mean fluorescence intensity (gMFI) values are shown for stable (mKO2 and CD41/ITGA2B) and decreasing markers (CD117, CD42b and CD56) of live CG cells during dTAG-13-treatment.

d: Results of CRISPR screens in the CG-AMKL cell line M-07e obtained from the depmap portal.¹⁸ Gene dependencies with selectivity and efficacy scores less than -0.5 have been highlighted and colors indicate genes belonging to modules (G1-G10) from Fig. 5a. Four genes have not been labelled due to lack of space: *LIG1* (G10), *PAICS* (G7), *HEXIM1*, *NAMPT* (both G2).



Supplementary Fig. 16. Transcription factor footprint analysis

Analysis of the chromatin accessibility footprints near selected TF sequence motifs (from the JASPAR¹³ database) aggregated by cell cluster (cp. **Fig. 4**). Each line plot is centered on the match to the indicated motif (x-axis) and reports the enrichment of Tn5 insertion events (ATAC-seq reads) compared to the expected frequency of insertion events (shown in the respective bottom plots).

SUPPLEMENTARY INFORMATION

CG donor vector sequence for *in trans* paired nicking.

CBFA2T3::GLIS2-dTAG-2xHA-IRES-mKO2-floxedPuroR pUC57simple plasmid, circular

Nick site 5' HA (CBFA2T3 intron 11) GLIS2 ex3-6 Linker-dTAG-2xHA IRES mKO2 loxP mPgl1 promoter
PuroR mPgl1 3'UTR 3' HA (CBFA2T3 exon 12) AmpR

TCGCGCGTTTCGGTGATGACGGTGAAAACCTCTGACACATGCAGCTCCCGGAGACGGTCACAGCTTGT
CTGTAAGCGGATGCCGGGAGCAGACAAGCCCGTCAGGGCGCGTCAGCGGGTGTGGCGGGTGTCTGGGG
CTGGCTTAACATATGCGGCATCAGAGCAGATTGTAAGTGCACCATATGCGGTGTGAAATACCGC
ACAGATGCGTAAGGAGAAAATACCGCATCAGGCGCCATTGCGCCATTCAGGCTGCGCAACTGTTGGGAA
GGGCGATCGGTGCGGGCCTCTTCGCTATTACGCCAGCTGGCGAAAGGGGGATGTGCTGCAAGGCGATT
AAGTTGGGTAACGCCAGGGTTTTCCAGTCACGACGTTGTAAAACGACGGCCAGTGAATTGGAGATCG
GTAATTCGCGAATGCGTCGAGATGCGGCCGCTGCAACGCGGCACGCTACTGCG.ATGCCTAGGGCACA
ACTCACCCCTACGGGAAGGGAGCAGAGCCACCACAGGGTGGGGGCAGAGGTGGAGGGGGTGCAGCAGA
GGCAGGGGTGGGGCAGAGGTGGGGAGGGGGACGGGTGGGTGTGGATGCAGCAGAGGCGGATGGTGAGA
TCTCAGCAGCCTCAATCTGCCGGGGCCACGCCCTTTGGTTTTTCGGTTCCTGGAGCCTGGTGAGCTCA
GGAGGGCTTCCTGGAGAGGAGGCAGGGGTGCTGTCCAGTTAGCACAGCGCAGACCTATCTGGGTGGG
CTGGAGCGTGAGACACGGTAGTGGGATGCCGTCCCGCAAATGGCCAGTGAGTGGCTGTATCTGGAG
CATGGCCCCGAGGGCTGTGGGGCTTTGGAGTTTCTCTCTTCTCACCTGGTGATGACGTGGCTGGGTCTG
AATGGCTTGGTTGAGGCGTGTTGGGGATTTCGGGTGGCCAGACCCAGCCCTGACTGGCCCCCTCTGCC
CCCACCCAGGACTTCCAGCCACTGCGCTATTTGGATGGTGTCCCCAGCTCCTTCCAGTTCTTCCTGCC
CCTCGGCTCCGGGGGGGCCCTGCACCTGCCTGCCTCCTCCTTACCCCTCCCAAGGACAAGTGCC
TCTCGCCAGACCTGCCCCTGCCCAAGCAGCTGGTGTGTGCTGGGCCAAGTGTAACCAGCTCTTTGAG
CTCCTGCAAGACCTGGTGGACCATGTCAACGATTACCATGTCAAGCCGAGAAGGATGCGGGGTACTG
CTGCCACTGGGAGGGCTGCGCCCGCCATGGCCGAGTTTCAACGCCAGGTACAAGATGCTCATCCACA
TCCGCACACACACCAACGAGAAGCCACACCGCTGTCCGACCTGCAGCAAGAGCTTCTCCCGCCTGGAG
AACCTGAAGATCCACAACCGGTGCGCACACAGGTGAGAAGCCCTACGTCTGCCCCCTACGAGGGCTGCAA
CAAGCGCTATTCCAACCTCAGTGACCGCTTTAAGCACACGCGCACCCACTATGTGGACAAGCCCTACT
ACTGCAAGATGCCCGGCTGCCACAAGCGCTACACGGACCCAGCTCACTGCGCAAGCACATCAAGGCC
CATGGCCACTTTGTGTCCACAGAGCAGCAAGAGCTCCTGCAGCTGCGCCACCCCCCAAGCCGCCACT
GCCCAGCCCCGACGGCGGGCCCTATGTCAAGTGGGGCCAGATCATCATCCCCAAGCCAGCTGCCCTCT
TTGGAGGCCCTGGCCTGCCCGGCTTACCCCTACCCCTGGCCCCCGGCCCTTGACCTCAGTGCCCTG
GCCTGTGGCAACGGTGGGGGCAGTGGGGGTGGGGGGGCATGGGCCCTGGGCTGCCAGGCCCCGTCTCT
GCCTCTCAATCTGGCCAAGAACCCGCTGCTGCCCTCGCCCTTTGGGGCTGGCGGACTGGGCTTGCCCTG
TGGTCTCCCTCCTTGCTGGCGCAGCTGGTGGCAAGGCCGAGGGGGAGAAGGGGGCTGGGTGCGTGCCC
ACCAGGGCCCTGGGCATGGAGGGCCACAAGACGCCCTTGAAGGACGGAGAGCAGCTGCTCCCGGCC
AAGCCCCGATGGACTCCCCCTGCTGCCAGGCACCGTGTGGACCTGTCCACGGGCGTCAACTCAGCTG
CCAGCAGCCCAGAGGCGTTGGCCCCCTGGCTGGGTGGTTCATCCCGCCGGGCTCGGTGCTGCTCAAACCG
GCTGTGGTGAACGTCGACGGTGGCGGTGGCTCGGGCGGTGGTGGGTGGGTGGCGGCGGATCTGGAGT
GCAGGTGGAAACCATCTCCCCAGGAGACGGGCGCACCTTCCCCAAGCGCGGCCAGACCTGCGTGGTGC
ACTACACCGGGATGCTTGAAGATGGAAAGAAAGTTGATTCCTCCCGGGACAGAAACAAGCCCTTTAAG
TTTATGCTAGGCAAGCAGGAGGTGATCCGAGGCTGGGAAGAAGGGGTGCCCAGATGAGTGTGGGTCA
GAGAGCCAACTGACTATATCTCCAGATTATGCCTATGGTGGCACTGGGCACCCAGGCATCATCCAC
CACATGCCACTCTCGTCTTCGATGTGGAGCTTCTAAACCTGGAAGGCTCTGGCTACCCCTACGACGTG
CCCGACTACGCCGGCTATCCGTATGATGTCCCGGACTATGCAGGCTAGTCGACAAGCTTGCCCCCTCTC
CCTCCCCCCCCCTAACGTTACTGGCCGAGCCGCTTGGAAATAAGGCCGGTGTGCGTTTGTCTATATG
TTATTTTCCACCATATTGCCGTCTTTTGGCAATGTGAGGGCCCGGAAACCTGGCCCTGTCTTCTTGAC
GAGCATTCCTAGGGGTCTTTCCCTCTCGCCAAAGGAATGCAAGGTCTGTTGAATGTCTGTAAGGAAG
CAGTTCTCTGGAAGTTTCTTGAAGACAAACAACGTCTGTAGCGACCCCTTTCAGGCAGCGGAACCCC
CCACCTGGCGACAGGTGCCTCTGCGGCCAAAAGCCACGTGTATAAGATACACCTGCAAGGCGGCACA
ACCCAGTGCCACGTTGTGAGTTGGATAGTTGTGGAAAGAGTCAAATGGCTCTCCTCAAGCGTATTCA
ACAAGGGGGCTGAAGGATGCCCAGAAGGTACCCCATTTGTATGGGATCTGATCTGGGGCCTCGGTGCACA
TGCTTTACATGTGTTTAGTCGAGGTTAAAAAACGTCTAGGCCCCCGAACCACGGGGACGTGGTTTTT
CCTTTGAAAAACACGATGATAATATGGCCACAACCATGGTGAGTGTGATTAAACCAGAGATGAAGATG

AGGTACTACATGGACGGCTCCGTCAATGGGCATGAGTTCACAATTGAAGGTGAAGGCACAGGCAGACC
 TTACGAGGGACATCAAGAGATGACACTACGCGTCACAATGGCCGAGGGCGGGCCAATGCCTTTCGCGT
 TTGACTTAGTGTACACGTGTTCTGTACGGCCACAGAGTATTTACTAAATATCCAGAAGAGATACCA
 GACTATTTCAAACAAGCATTTCCTGAAGGCCTGTCATGGGAAAGGTCGTTGGAGTTCGAAGATGGTGG
 GTCCGCTTCAGTCAGTGCATATAAGCCTTAGAGGAAACACCTTCTACCACAAATCCAAATTTACTG
 GGGTTAACTTTTCCTGCCGATGGTCTATCATGCAAAACCAAAGTGTTGATTGGGAGCCATCAACCGAG
 AAAATTACTGCCAGCGACGGAGTTCTGAAGGGTGATGTTACGATGTACCTAAAACCTGAAGGAGGCGG
 CAATCACAATGCCAAATGAAGACTACTTACAAGGCGGCAAAAGAGATTCTTGAAATGCCAGGAGACC
 ATTACATCGGCCATCGCCTCGTCAGGAAAACCGAAGGCAACATTACTGAGCAGGTAGAAGATGCAGTA
 GCTCATTCCCTAACTAGAAACGATCATATTCAATAACCTTAATATAACTTCGTATAATGTATGCTAT
 ACGAAGTTATTAGGTCTGAAGAGGAGTTTACGTCCAGCCAAGCTTAGGATCTCGACCTCGAAATTCCTA
 CCGGGTAGGGGAGGCGCTTTTCCCAAGGCAGTCTGGAGCATGCGCTTTAGCAGCCCCGCTGGGCACTT
 GCGCTACACAAGTGGCCTCTGGCCTCGCACACATTCCACATCCACCGGTAGGCGCCAACCGGCTCCG
 TTCTTTGGTGGCCCCCTTCGCGCCACCTTCTACTCCTCCCCCTAGTCAGGAAGTTCCCCCCCCGCCGCA
 GCTCGCGTCGTGCAGGACGTGACAAATGGAAGTAGCACGTCTCACTAGTCTCGTGCAGATGGACAGCA
 CCGCTGAGCAATGGAAGCGGGTAGGCCTTTGGGGCAGCGGCCAATAGCAGCTTTGCTCCTTCGCTTTC
 TGGGCTCAGAGGCTGGGAAGGGGTGGGTCCGGGGCGGGCTCAGGGGCGGGCTCAGGGGCGGGCGGG
 CGCCCCGAAGGTCCTCCGAGGCCCCGGCATTCTGCACGCTTCAAAGCGCACGTCTGCCGCGCTGTTCT
 CCTCTTCTCATCTCCGGGCCTTTGACCTGGACCATGACCGAGTACAAGCCCACGGTGCGCCCTCGCC
 ACCCGCGACGACGTCCCCAGGGCCGTACGCACCTCGCCGCCGCTTCGCCGACTACCCGCCACGCG
 CCACACCGTCGATCCGGACCGCCACATCGAGCGGGTCACCGAGCTGCAAGAACTCTTCTCACGCGCG
 TCGGGCTCGACATCGGCAAGGTGTGGGTGCGGACGACGGCGCCGCGGTGGCGGTCTGGACCACGCCG
 GAGAGCGTCGAAGCGGGGGCGGTGTTGCGCGAGATCGGCCCGCGCATGGCCGAGTTGAGCGGTTCCCG
 GCTGGCCGCGCAGCAACAGATGGAAGGCCTCTGGCGCCGACCGGCCCAAGGAGCCCGCTGGTTCC
 TGGCCACCGTCGGCGTCTCGCCGACCACAGGGCAAGGGTCTGGGCAGCGCCGCTCGTGCTCCCCGA
 GTGGAGGCGGCCGAGCGCGCCGGGGTGGCCGCTTCTTGAGACCTCCGCGCCCCGCAACCTCCCCTT
 CTACGAGCGGCTCGGCTTACCGTCACCGCCGACGTGAGGTGCCCGAAGGACCGCGCACCTGGTGCA
 TGACCCGCAAGCCCGGTGCTGACGCCCCCCCCACGACCCGAGCGCCCCGACCGAAAGGAGCGCACGA
 CCCCATGCATCGATGATATCAGATCCCCGGGATGTCAGAAATTGATGATCTATTAAACAATAAAGATGT
 CCACTAAAATGGAAGTTTTTCTGTCTACTTTGTAAAGAAGGGTGAGAACAGAGTACCTACATTTTG
 AATGGAAGGATTGGAGCTACGGGGGTGGGGGTGGGGTGGGATTAGATAAATGCCTGCTCTTTACTGAA
 GGCTCTTTACTATTGCTTTATGATAATGTTTCATAGTTGGATATCATAATTTAAACAAGCAAAACCAA
 ATTAAGGGCCAGCTCATTCCTCCCACTCATGATCTATAGATCTATAGATCTCTCGTGGGATCATTTGT
 TTTCTCTTGATTCCCACTTTGTGGTTCTAAGTACTGTGGTTTCCAAATGTGTGAGTTTCATAGCCTGA
 AGAACGAGATCAGCAGCCTCTGTTCCACATACACTTCATTCTCAGTATTGTTTTGCCAAGTTCTAATT
 CCATCAGAAGCTGGTCGAGATCCGGAACCTTAATATAACTTCGTATAATGTATGCTATACGAAGTTA
 TTAGTCCCTCTCGAGCTACTGCGGGTCTTCTGCCAGCATCGGGACTGGGAGAAGCATCACACAGTG
 TGTGGCCAGAGCCTGCAGGGCCCCACAGCCGTGGTGGCCGACCCGGTGCCTGGACCGCCGAAGCCGC
 CCACAGCCTGGGCCCCCTCCCTGCCTGTGGGTGCTGCCAGCCCCAGCGAAGCCGGCTCTGCGGGGCCCT
 CTCGCCCCGGCTCCCCAGCCCCACCTGGCCCACTGGACACCGTGCCCCGCTGACCCCACTGGCCCCCTG
 GCCTGCCGGACACAGCACCGTGCCAACCCCAACCCAGCTCCAGGCCACCGGATGCTGTGCCTGGCCTC
 CGATGCCCTGGCCTGCCAGACACTGCGCCCCGCTGACCTGGGGGAGCCGACCAATTAGTCACTGCTGC
 TACTGCCCCCTCTCCGAAAGAAGACACAGAACCAACAAAACCGCATTCAAGTGCACCTGCCTCAGCTACC
 TAATGATTCCGCGCGGAGACCTCCTGACAACGTCTCTTCAAGCATCTCAGAAGCCTCGACTGAGCTT
 TGCAACGCGGCACGCTACTGCGCGCCATCGGATGCCGGGACCGACGAGTGCAGAGGCGTGCAAGC
 GAGCTTGGCGTAATCATGGTCATAGCTGTTTCTGTGTGAAATTGTTATCCGCTCACAATTCACACA
 ACATACGAGCCGGAAGCATAAAGTGTAAGCCTGGGGTGCCTAATGAGTGAGCTAACTACATTAATT
 GCGTTGCGCTCACTGCCCGCTTTCCAGTCGGGAAACCTGTCTGTCGAGCTGCATTAATGAATCGGCCA
 ACGCGCGGGGAGAGGCGGTTTGGGTATTGGGCGCTCTTCCGCTTCTCGCTCACTGACTCGCTGCGCT
 CGGTGCTTCCGCTGCGGCGAGCGGTATCAGTCACTCAAAGGCGGTAATACGGTTATCCACAGAATCA
 GGGGATAACGCAGGAAGAACATGTGAGCAAAAAGGCCAGCAAAAGGCCAGGAACCGTAAGGCGCCG
 GTTGCTGGCGTTTTTTCATAGGCTCCGCCCCCTGACGAGCATCAGAAAATCGACGCTCAAGTCAGA
 GGTGGCGAAACCCGACAGGACTATAAAGATACCAGGCGTTTTCCCCCTGGAAGCTCCCTCGTGCGCTCT
 CCTGTTCCGACCCCTGCCGCTTACCGGATACCTGTCCGCTTTCTCCCTTCGGGAAGCGTGCGCTTTC
 TCATAGCTCACGCTGTAGGTATCTCAGTTCCGTGTAGGTCGTTGCTCCAGCTGGGCTGTGTGCACG
 AACCCCCGTTTACGCCCCGACCGCTGCGCCTTATCCGGTAACCTATCGTCTTGAGTCCAACCCGGTAAGA

CACGACTTATCGCCACTGGCAGCAGCCACTGGTAACAGGATTAGCAGAGCGAGGTATGTAGGCGGTGC
TACAGAGTTCTTGAAGTGGTGGCCTAACTACGGCTACACTAGAAGAACAGTATTTGGTATCTGCGCTC
TGCTGAAGCCAGTTACCTTCGGAAAAAGAGTTGGTAGCTCTTGATCCGGCAAACAAACCACCGCTGGT
AGCGGTGGTTTTTTTTGTTTGCAAGCAGCAGATTACGCGCAGAAAAAAGGATCTCAAGAAGATCCTTT
GATCTTTTCTACGGGGTCTGACGCTCAGTGGAACGAAAACTCACGTTAAGGGATTTTGGTCATGAGAT
TATCAAAAAGGATCTTCACCTAGATCCTTTTAAATTAAAAATGAAGTTTTAAATCAATCTAAAGTATA
TATGAGTAAACTTGGTCTGACAGTTACCAATGCTTAATCAGTGAGGCACCTATCTCAGCGATCTGTCT
ATTCGTTTCATCCATAGTTGCCTGACTCCCCGTCGTGTAGATAACTACGATACGGGAGGGCTTACCAT
CTGGCCCCAGTGCTGCAATGATACCGCGAGACCCACGCTCACCGGCTCCAGATTTATCAGCAATAAAC
CAGCCAGCCGGAAGGGCCGAGCGCAGAAGTGGTCCTGCAACTTTATCCGCCTCCATCCAGTCTATTAA
TTGTTGCCGGAAGCTAGAGTAAGTAGTTGCCAGTTAATAGTTTTCGCAACGTTGTTGCCATTGCTA
CAGGCATCGTGGTGTACAGCTCGTCGTTTGGTATGGCTTCATTCAGCTCCGGTTCCCAACGATCAAGG
CGAGTTACATGATCCCCATGTTGTGCAAAAAAGCGGTTAGCTCCTTCGGTCCTCCGATCGTTGTCAG
AAGTAAGTTGGCCGCAGTGTTATCACTCATGGTTATGGCAGCACTGCATAATTCTCTTACTGTCATGC
CATCCGTAAGATGCTTTTCTGTGACTGGTGAGTACTCAACCAAGTCATTCTGAGAATAGTGTATGCGG
CGACCGAGTTGCTCTTGCCCGGCGTCAATACGGGATAATACCGCGCCACATAGCAGAACTTTAAAAGT
GCTCATCATTTGAAAAACGTTCTTCGGGGCGAAAACTCTCAAGGATCTTACCGCTGTTGAGATCCAGTT
CGATGTAACCCACTCGTGCACCCAACTGATCTTCAGCATCTTTTACTTTTACCAGCGTTTCTGGGTGA
GCAAAAACAGGAAGGCAAAATGCCGCAAAAAAGGGAATAAGGGCGACACGGAAATGTTGAATACTCAT
ACTCTTCCTTTTTCAATATTATTGAAGCATTTATCAGGGTTATTGTCTCATGAGCGGATACATATTTG
AATGTATTTAGAAAAATAACAAATAGGGGTTCGCGCACATTTCCCCGAAAAGTGCCACCTGACGTC
TAAGAAACCATTATTATCATGACATTAACCTATAAAAAATAGGCGTATCACGAGGCCCTTTCGTC

SUPPLEMENTARY REFERENCES

1. Mazza MC, Beilina A, Roosen DA, Hauser D, Cookson MR. Generation of iPSC line from a Parkinson patient with PARK7 mutation and CRISPR-edited Gibco human episomal iPSC line to mimic PARK7 mutation. *Stem Cell Res* **55**, 102506 (2021).
2. Poell JB, Mendeville M, Sie D, Brink A, Brakenhoff RH, Ylstra B. ACE: absolute copy number estimation from low-coverage whole-genome sequencing data. *Bioinformatics* **35**, 2847-2849 (2019).
3. Krivec N, Ghosh MS, Spits C. Gains of 20q11.21 in human pluripotent stem cells: Insights from cancer research. *Stem Cell Reports* **19**, 11-27 (2024).
4. Stavish D, *et al.* Feeder-free culture of human pluripotent stem cells drives MDM4-mediated gain of chromosome 1q. *Stem Cell Reports* **19**, 1217-1232 (2024).
5. Bertuccio SN, *et al.* The Pediatric Acute Leukemia Fusion Oncogene ETO2-GLIS2 Increases Self-Renewal and Alters Differentiation in a Human Induced Pluripotent Stem Cells-Derived Model. *Hemasphere* **4**, e319 (2020).
6. Gress V, *et al.* CBFA2T3::GLIS2 pediatric acute megakaryoblastic leukemia is sensitive to BCL-XL inhibition by navitoclax and DT2216. *Blood Adv* **8**, 112-129 (2024).
7. Le Q, *et al.* CBFA2T3-GLIS2 model of pediatric acute megakaryoblastic leukemia identifies FOLR1 as a CAR T cell target. *J Clin Invest* **132**, e157101 (2022).
8. Gruber TA, *et al.* An Inv(16)(p13.3q24.3)-encoded CBFA2T3-GLIS2 fusion protein defines an aggressive subtype of pediatric acute megakaryoblastic leukemia. *Cancer Cell* **22**, 683-697 (2012).
9. Smith JL, *et al.* Comprehensive Transcriptome Profiling of Cryptic CBFA2T3-GLIS2 Fusion-Positive AML Defines Novel Therapeutic Options: A COG and TARGET Pediatric AML Study. *Clin Cancer Res* **26**, 726-737 (2020).
10. Elsayed AH, *et al.* A six-gene leukemic stem cell score identifies high risk pediatric acute myeloid leukemia. *Leukemia* **34**, 735-745 (2020).
11. Benbarche S, *et al.* Screening of ETO2-GLIS2-induced Super Enhancers identifies targetable cooperative dependencies in acute megakaryoblastic leukemia. *Sci Adv* **8**, eabg9455 (2022).
12. Thirant C, *et al.* ETO2-GLIS2 Hijacks Transcriptional Complexes to Drive Cellular Identity and Self-Renewal in Pediatric Acute Megakaryoblastic Leukemia. *Cancer Cell* **31**, 452-465 (2017).
13. Rauluseviciute I, *et al.* JASPAR 2024: 20th anniversary of the open-access database of transcription factor binding profiles. *Nucleic Acids Res* **52**, D174-D182 (2024).

14. Jardine L, *et al.* Blood and immune development in human fetal bone marrow and Down syndrome. *Nature* **598**, 327-331 (2021).
15. Suo C, *et al.* Mapping the developing human immune system across organs. *Science* **376**, eabo0510 (2022).
16. Wang H, *et al.* Decoding Human Megakaryocyte Development. *Cell Stem Cell* **28**, 535-549 e538 (2021).
17. Street K, *et al.* Slingshot: cell lineage and pseudotime inference for single-cell transcriptomics. *BMC Genomics* **19**, 477 (2018).
18. Tsherniak A, *et al.* Defining a Cancer Dependency Map. *Cell* **170**, 564-576 e516 (2017).