

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

Corresponding Author: Dr Klaus Fortschegger

This file contains all reviewer reports in order by version, followed by all author rebuttals in order by version.

Version 0:

Reviewer comments:

Reviewer #1

(Remarks to the Author)

In this study, Shoeb et al. have developed a model to study the CBFA2T3-GLIS2 (CG) fusion, which is responsible for acute megakaryoblastic leukemia, using human induced pluripotent stem cells (iPSCs). They generated three clones in which GLIS2, fused to a degron and an HA tag, was knocked into the endogenous CBFA2T3 locus. This approach allowed for the recreation of the fusion while providing a convenient system for immunoprecipitation or rapid degradation of the fusion protein.

They then used this model to perform a multi-omics analysis, including single-cell RNA sequencing (scRNA-seq), ATAC-seq, and ChIP-seq. Initially, the CG cells did not show significant differences from the control cells; however, after a prolonged culture period, their transcriptome and chromatin profile closely resembled those of patient/patient-derived cells. Single-cell analysis indicated that CG cells form two main populations: one displaying several features of normal megakaryocytic cells but lacking full maturation, and another with a more immature profile.

The study is well-conducted, with robust validation of CG CRISPR iPSC clone generation, extensive analysis of the sequencing data, and integration of these data with previously existing datasets.

However, this work lacks significant novelty. Several studies have already been published in which human models were used to study the CG fusion (PMID 39384814, PMID 39304903, PMID 37729615). More importantly, Boudia et al. (PMID 39656971) recently published an article using the exact same methodology (i.e., iPSCs engineered through CRISPR to recreate the CG fusion, followed by functional assays, transcriptomics, and chromatin analysis). Nonetheless, this study serves as an interesting validation of these previous results.

I only have few minor comments:

1. Multiple iPSC differentiation protocols (2D, 3D, feeder-free, EB, etc.) can lead to different hematopoietic outcomes. While I agree that the differentiation method used here is sufficient to explore the impact of CG fusion, I would like some characterization of the protocol. Can the authors provide phenotypic data for the endothelial-to-hematopoietic transition using common markers such as CD31, CD34, CD43, CD45, etc? This would help confirm the stepwise hematopoietic commitment and clarify why day 13 was chosen as the timepoint for analysis.
2. The authors show that upon CG removal, aMKs are unable to differentiate (Supp. Fig. 15c) and the cells die (Supp. Fig. 5f). While the data clearly support this conclusion, it contradicts findings by Lopez et al. (PMID 31662298), where removal of the fusion fully restored normal hematopoiesis. Can the authors comment on this discrepancy?
3. I am confused about which populations are being compared in Fig. 6b. The populations identified in Fig. 4 include progenitors (s0), normal MKs (ctrl1), aMKs (cg1), and aMKPs (cg2-4). However, in Fig. 6b (right panel), s0 is compared with ctrl1, yet the title refers to MKs vs. aMKPs. Can the authors clarify this?

Reviewer #2

(Remarks to the Author)

In this study, Fortschegger et al have established hiPSCs conditionally expressing the CBFA2T3:GLIS2 fusion from the endogenous CBFA2T3 locus. Using this model system, they have then undertaken extensive single-cell and bulk genomic, transcriptomic and epigenomic studies in order to determine the endogenous gene-regulatory and oncogenic activity at early

and late stage of CG-driven AMKL transformation.

This is a very nice and detailed analysis of CG function that will provide a significant new data resource for the research community. In general experiments have been performed at a high level and results have been described and displayed in sufficient quantity and quality. The findings are novel (although some of the data replicates what has been found in other studies) and these novel findings are of interest to the field by providing additional important insight into the function of the CG fusion in driving AMKL biology. In particular, the findings related to different trajectories of cell development as well as altered regulation by key megakaryocytic and stem cell TFs (and unique TFs) as a result of CG expression are intriguing and of interest. The conclusions are supported by the data.

Some minor comments below:

Figure 2A: This figure is a bit confusing to understand, can there be better annotation of this figure (i.e. is the top row labelling the same as the vertical labelling - to allow pairwise comparisons)? Also how do these data (Suppl. Figure 9a) show that accessibility was mainly disrupted between D13 and D25? It looks like an additional large change in accessibility occurs between D25 and long term culture too no?

Supplementary Figure 9C: It is not clear what this is showing? It is referenced in the article as being related to overlap between CG ChIP-seq peaks and accessibility, but I cannot determine how this shows that? Perhaps this is more related to the previous result of accessibility change from D13 to D25 (see previous comment)? Please fix, explain or reference this more clearly in the article (as appropriate).

Figure 2: ChIP-seq was only performed in long-term cells, yet the authors try to correlate ChIP-seq peaks with accessibility at much earlier timepoints (i.e. D13, D25). How relevant is this? Perhaps it is not surprising that only a small % of open regions correspond to a CG binding site at D13? Would it not be worthwhile to do ChIP-seq at D13 and D25 as well if you wish to make conclusions regarding CG binding and accessibility?

Supplementary Figure 10: This figure just shows a few examples of where CG binding, RNA-seq and ATAC-seq correlate as described in the article body. Can this be demonstrated in a more global way with significance attached? Or have the authors primarily cherry picked examples that support their conclusion? Furthermore, this figure is very hard to decipher and it is not entirely clear that the figures support the conclusion.

Figure 5e: This analysis is entirely based on findings in unrelated cells and/or cancer models. While it may be indicative, it feels incomplete and not of high value. Perhaps these data would have benefited from some follow up confirmation work, using these inhibitors on the AMKL cells themselves?

Reviewer #3

(Remarks to the Author)

The authors present a human iPSC-based model of CBFA2T3::GLIS2 (CG)-driven AMKL, in which CG is conditionally expressed at physiological levels from the endogenous CBFA2T3 locus. This system establishes oncogene dependence and enables precise dissection of CG function. Through single-cell RNA-seq and ATAC-seq, the authors chart the transcriptional and chromatin landscape during early hematopoiesis and show that CG disrupts differentiation via two distinct routes: by maintaining aberrant megakaryocyte progenitors in a self-renewing state and by blocking differentiation through downregulation of some key transcription factors. The study is technically strong and provides comprehensive and modern transcriptional profiling, which in itself warrants publication. However, its primarily descriptive nature limits its broader impact. This should be acknowledged and justified in the discussion. Additionally, while the expression of MKP- and MK-associated transcription factor modules likely reflects their functional activity, a more detailed mechanistic analysis or discussion would strengthen the study's depth.

Version 1:

Reviewer comments:

Reviewer #1

(Remarks to the Author)

The authors have addressed my questions and clarified some specific points. I have no further concerns.

Reviewer #2

(Remarks to the Author)

I thank the authors for addressing my comments and suggestions. I am satisfied with the associated changes and clarifications within the revised manuscript.

Reviewer #3

(Remarks to the Author)

The authors have addressed all points raised and further improved the manuscript, which was already of high quality. I recommend publication in its current form.

Open Access This Peer Review File is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

In cases where reviewers are anonymous, credit should be given to 'Anonymous Referee' and the source.

The images or other third party material in this Peer Review File are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

To view a copy of this license, visit <https://creativecommons.org/licenses/by/4.0/>

Point-by point response to reviewers' comments

Reviewers' comments:

Reviewer #1 (Remarks to the Author):

In this study, Shoeb et al. have developed a model to study the CBFA2T3-GLIS2 (CG) fusion, which is responsible for acute megakaryoblastic leukemia, using human induced pluripotent stem cells (iPSCs). They generated three clones in which GLIS2, fused to a degron and an HA tag, was knocked into the endogenous CBFA2T3 locus. This approach allowed for the recreation of the fusion while providing a convenient system for immunoprecipitation or rapid degradation of the fusion protein.

They then used this model to perform a multi-omics analysis, including single-cell RNA sequencing (scRNA-seq), ATAC-seq, and ChIP-seq. Initially, the CG cells did not show significant differences from the control cells; however, after a prolonged culture period, their transcriptome and chromatin profile closely resembled those of patient/patient-derived cells. Single-cell analysis indicated that CG cells form two main populations: one displaying several features of normal megakaryocytic cells but lacking full maturation, and another with a more immature profile.

The study is well-conducted, with robust validation of CG CRISPR iPSC clone generation, extensive analysis of the sequencing data, and integration of these data with previously existing datasets.

However, this work lacks significant novelty. Several studies have already been published in which human models were used to study the CG fusion (PMID 39384814, PMID 39304903, PMID 37729615). More importantly, Boudia et al. (PMID 39656971) recently published an article using the exact same methodology (i.e., iPSCs engineered through CRISPR to recreate the CG fusion, followed by functional assays, transcriptomics, and chromatin analysis). Nonetheless, this study serves as an interesting validation of these previous results.

We thank the reviewer for the thoughtful and thorough evaluation of our manuscript, and we appreciate the constructive comments. We are pleased that our findings are consistent with recent publications in the field and share the reviewer's enthusiasm for the recent work by Boudia and colleagues, which alongside our study, underscores the value of iPSC models in investigating fusion-driven leukemias.

Our study builds on these advances by introducing a feeder-free, extended culture system combined with time-resolved multi-omics analysis (D13, D25, D72 and D115). This approach provides novel insights into the temporal dynamics of CG-driven AMKL, which have not been explored in previous studies. In addition, our model cell lines allow for the conditional modulation of the CG-fusion protein, offering a versatile platform for mechanistic studies, identifying potential therapeutic targets and preclinical drug testing.

I only have few minor comments:

- 1. Multiple iPSC differentiation protocols (2D, 3D, feeder-free, EB, etc.) can lead to different hematopoietic outcomes. While I agree that the differentiation method used here is sufficient to explore the impact of CG fusion, I would like some characterization of the protocol. Can the authors provide phenotypic data for the endothelial-to-hematopoietic transition using common markers such as CD31, CD34, CD43, CD45, etc? This would help confirm the stepwise hematopoietic commitment and clarify why day 13 was chosen as the timepoint for analysis.*

We fully agree that the choice of differentiation protocol significantly influences hematopoietic outcomes. We employed the commercial STEMdiff hematopoietic kit, as it consistently yielded the most efficient and reproducible results in our preliminary experiments. The cell types obtained by this differentiation protocol at several timepoints have been characterized in our previous work and by others (e.g. PMIDs: 30009677, 30577865, 31710911, 31956083, 33102613, 33177198, 34299194), demonstrating stepwise progression from mesodermal induction to endothelial (CD31, CD34) and arterial/hemogenic specification (CD144, CD43), followed by transition to hematopoietic suspension cells (CD45). To favor definitive hematopoiesis and increase yield, we included a short CHIR99021 treatment, which has been previously shown to affect these parameters (PMIDs: 24837661, 29879440) and the obtained hematopoietic stem and progenitor cells (HSPC) showed robust myeloid potential.

We decided to focus our analyses of the clonal CG lines on D13 onwards for two main reasons: **(a)** At D13, differentiating iPSCs have acquired an HSPC-like phenotype, while earlier stages are dominated by mesodermal and mesenchymal intermediates. Since the CG fusion is thought to arise somatically in HSPCs (PMIDs: 31471945, 39656971), we sought to model primarily the most relevant developmental window. **(b)** Although a nascent expansion of megakaryocyte progenitors was already apparent at D13, at this time-point chromatin accessibility and transcriptional profiles remained largely similar to controls (Figs. 2a, 3d), suggesting that CG-induced changes were only beginning to manifest. This early stage of transformation required detailed single-cell analyses to resolve, as shown in cp. Fig. 6 of our manuscript.

We have clarified the rationale for our timepoint selection and added references to prior protocols for hiPSC differentiation into hematopoietic progenitors in the revised manuscript (lines 98-101, 108-109).

- 2. The authors show that upon CG removal, aMKs are unable to differentiate (Supp. Fig. 15c) and the cells die (Supp. Fig. 5f). While the data clearly support this conclusion, it contradicts findings by Lopez et al. (PMID 31662298), where removal of the fusion fully restored normal hematopoiesis. Can the authors comment on this discrepancy?*

We appreciate that the reviewer is highlighting this important point. We also noted the apparent discrepancy with the findings of Lopez et al. and believe that both technical and

biological differences between the models likely contribute to the divergent outcomes. Instead of a human model, Lopez *et al.* used a murine model and a doxycycline-inducible expression system (rtTA in *Rosa26* locus and ETO2/CBFA2T3::GLIS2 under TRE promoter in *Hprt*) and their analysis focused on either *in vitro* differentiation of uninduced immature CD41+CD11b-Gr1- fetal liver blasts (Fig. S6D-E in PMID 31662298) or secondary hematopoietic engraftment of a heterogeneous population of iEG+ bone marrow cells (Fig. 7H-K in PMID 31662298). These cell populations likely include early hematopoietic progenitors still capable of entering and progressing through normal differentiation trajectories.

In contrast, our long-term cultured human iPSC-derived cells represent a more committed, highly CG-dependent state (aMKP trajectory endpoint; quiescent/cycling cell clusters cg3/4) resembling AMKL cancer cell lines. This interpretation is supported by recent findings in patient-derived M-07e cells, where CRISPR-mediated knockout of CG led to the upregulation of differentiation transcription factor (TF) GATA1, but downregulation of key self-renewal TFs such as ERG, mitogenic RAS/MAPK signaling and anti-apoptotic BCL2 (PMID 37716355). These changes compromised M-07e cell fitness and could be independently reproduced, using, for example, sgRNAs targeting CBFA2T3 (disrupting CG expression), ERG and GATA2 (see: https://depmap.org/portal/cell_line/ACH-000602).

Taken together, we propose that CG fusion protein depletion in late-stage fusion-dependent cells primarily disrupts critical gene regulatory networks (GRNs) governing self-renewal and survival, leading to cell death. We now briefly discuss this discrepancy in the revised manuscript (lines 347-352).

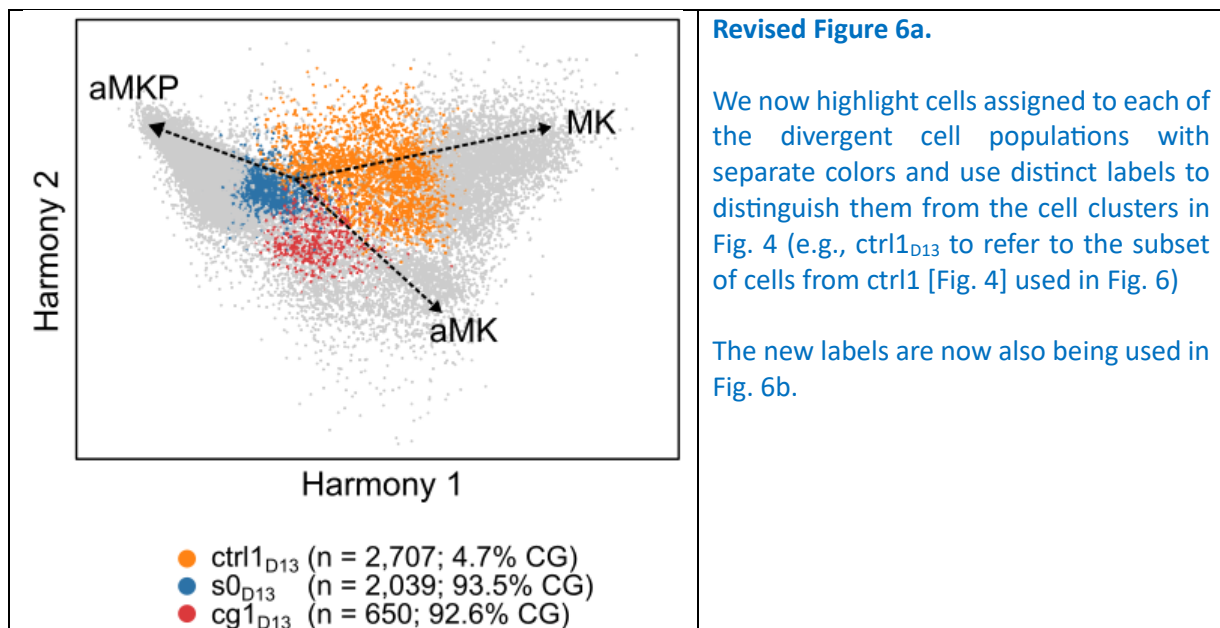
3. *I am confused about which populations are being compared in Fig. 6b. The populations identified in Fig. 4 include progenitors (s0), normal MKs (ctrl1), aMKs (cg1), and aMKPs (cg2-4). However, in Fig. 6b (right panel), s0 is compared with ctrl1, yet the title refers to MKs vs. aMKPs. Can the authors clarify this?*

We acknowledge that our explanations of Fig. 6. were insufficient and may have led to misinterpretation. Our intention in Fig. 6 was to highlight early gene regulatory decisions that guide cells toward three distinct fates described in Figs. 4-5: normal MKs, aberrant MKs ("aMK"), and aberrant MK progenitors ("aMKPs").

As the reviewer correctly noted, the aMKP trajectory ultimately gives rise to cells in clusters cg2-4. However, at day 13, CG-expressing cells are still largely absent from these clusters, and instead are developmentally delayed and remain in cluster s0 as indicated by the relative frequencies of cells in Fig. 4b.

To investigate early gene-regulatory differences, we focused on D13 cells with the highest neighborhood mixing scores – those closest to the inferred “intersection” and used the label “s0” as a proxy for the aMKP trajectory. At this stage, about 93.5% of the selected cells in cluster s0 are CG-expressing. In retrospect, we realized that the use of these labels was misleading, as the cell groups in Fig. 6a/b do not directly correspond to those in other figures.

To address this, we have revised the figure and labeling: We now use time- and condition-specific labels $s0_{D13}$, $ctrl1_{D13}$ and $cg1_{D13}$ and clearly marked the cells assigned to each identity in Fig. 6a. We removed the previous titles and applied the new labels from Fig. 6a also in Fig. 6b as shown in the revised Fig. 6a. To improve clarity, we also modified the Figure legends and the corresponding text in the results section (lines 277-282).



Reviewer #2 (Remarks to the Author):

In this study, Fortschegger et al. have established hiPSCs conditionally expressing the CBFA2T3:GLIS2 fusion from the endogenous CBFA2T3 locus. Using this model system, they have then undertaken extensive single-cell and bulk genomic, transcriptomic and epigenomic studies in order to determine the endogenous gene-regulatory and oncogenic activity at early and late stage of CG-driven AMKL transformation.

This is a very nice and detailed analysis of CG function that will provide a significant new data resource for the research community. In general experiments have been performed at a high level and results have been described and displayed in sufficient quantity and quality. The findings are novel (although some of the data replicates what has been found in other studies) and these novel findings are of interest to the field by providing additional important insight into the function of the CG fusion in driving AMKL biology. In particular, the findings related to different trajectories of cell development as well as altered regulation by key megakaryocytic and stem cell TFs (and unique TFs) as a result of CG expression are intriguing and of interest. The conclusions are supported by the data.

We thank the reviewer for evaluating our study and for the positive feedback regarding its relevance, significance, and the novelty of the findings. We appreciate the recognition of our efforts to advance the understanding of the CBFA2T3:GLIS2 fusion and its implications in the pathogenesis of acute megakaryoblastic leukemia.

Some minor comments below:

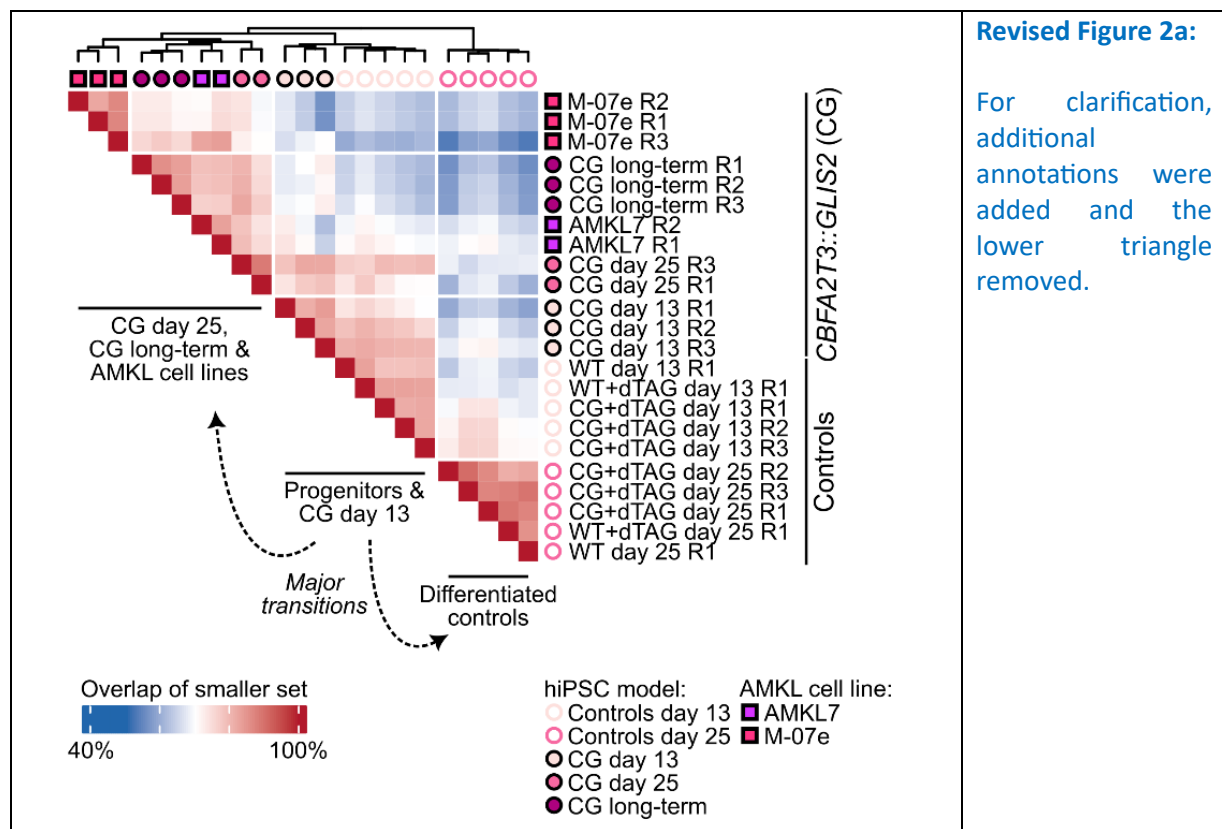
- 1. Figure 2A: This figure is a bit confusing to understand, can there be better annotation of this figure (i.e. is the top row labelling the same as the vertical labelling - to allow pairwise comparisons)? Also how do these data (Suppl. Figure 9a) show that accessibility was mainly disrupted between D13 and D25? It looks like an additional large change in accessibility occurs between D25 and long term culture too no?*

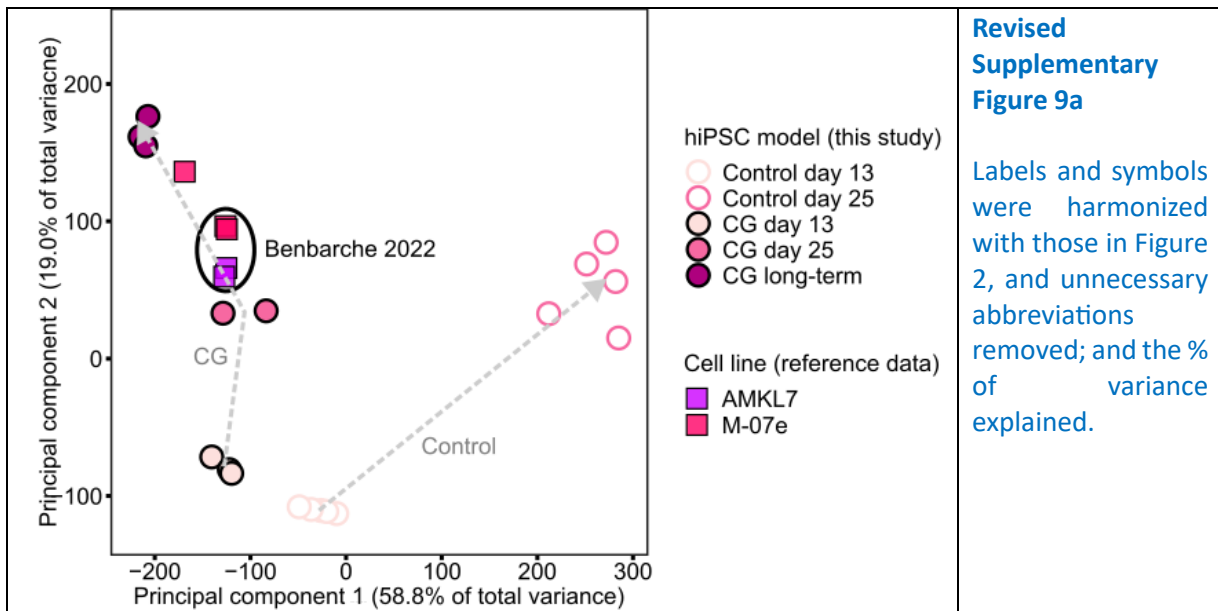
We thank the reviewer for pointing out the confusion in the presentation of the data shown in Figure 2a. We confirm that the vertical and horizontal clustering and the labels in Figure 2a are identical, allowing for pairwise comparisons. The figure shows hierarchical clustering of samples based on the overlap of ATAC-seq peaks, revealing three major clusters: **1.** CG-AMKL cell lines (M-07e & AMKL7) together with our hiPSC-derived CG-expressing cells at day 25 and long-term, **2.** Control cells (WT or CG dTAG treated) as well as early CG-expressing cells at day 13. **3.** Fully differentiated control cells at day 25.

These results indicate a major transition in chromatin architecture between days 13 and 25 in CG-expressing cells. Many peaks that open during this period remain accessible in long-term CG cells and are shared with AMKL cell lines (see Supplementary Fig. 9b for corresponding peak

counts). Over time, these peaks increasingly gain further accessibility and additional peaks emerge in long-term CG cultures, which explains the separation between day 25 and long-term CG samples in the PCA plot in Supplementary Figure 9a. In this plot, PC1 captures the primary variance (58.8%), which is by far the largest source of variation in the data, separating wild-type from CG-expressing cells, while PC2 reflects time in culture explaining the 19.0% of the total variance. The gradual gain of accessibility in all peaks that are eventually opened up in long-term CG cells can be clearly seen in Figure 2b.

To improve clarity, we have modified Figure 2a by removing the lower left triangle of the symmetric heatmap and adding additional annotations to distinguish between samples and sample groups. We also harmonized the labelling between Figure 2a and Supplementary Figure 9, and modified the corresponding figure legends and references in the text (lines 132-140). The revised figure panels are included below for reference.



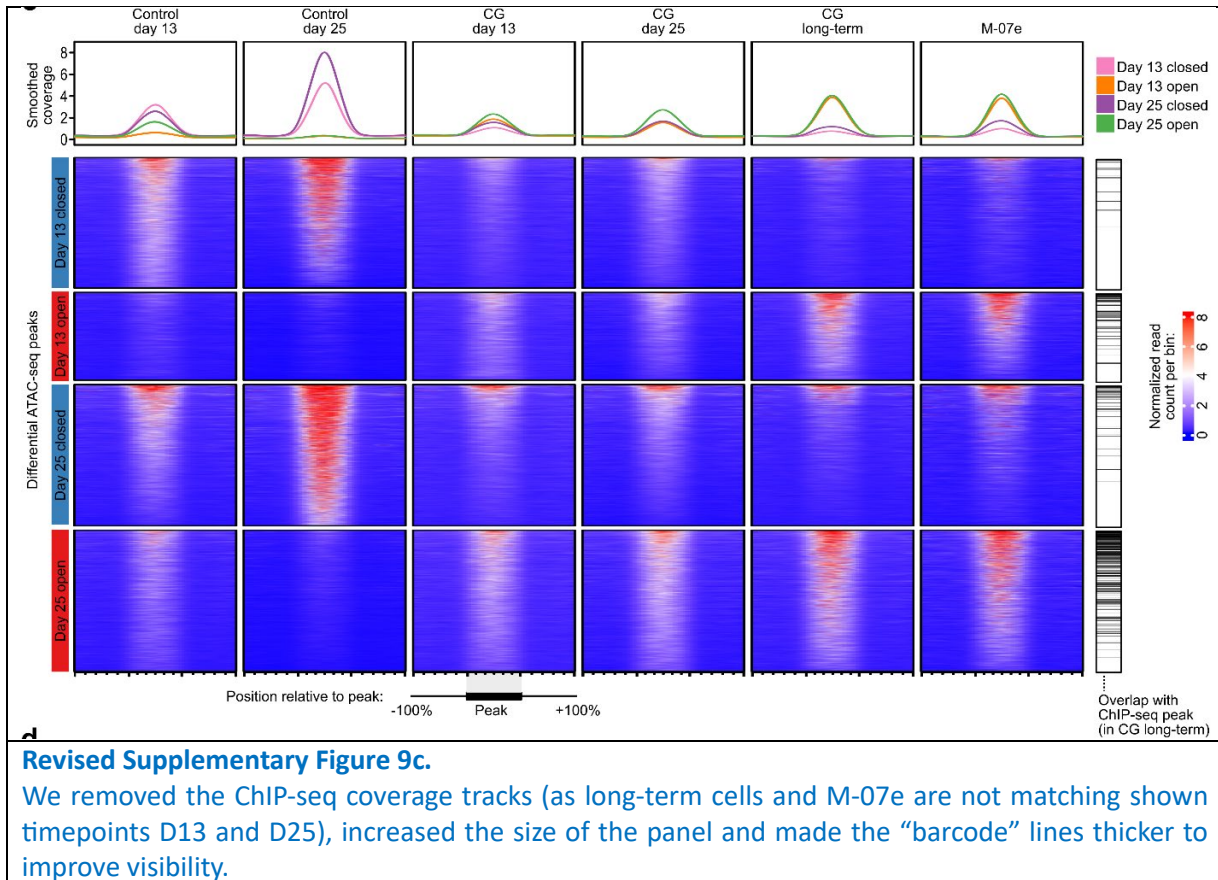


2. *Supplementary Figure 9c: It is not clear what this is showing? It is referenced in the article as being related to overlap between CG ChIP-seq peaks and accessibility, but I cannot determine how this shows that? Perhaps this is more related to the previous result of accessibility change from D13 to D25 (see previous comment)? Please fix, explain or reference this more clearly in the article (as appropriate).*

We thank the reviewer for pointing out the lack of clarity in Supplementary Figure 9c and apologize for the insufficient explanation. This figure illustrates regions with significantly different chromatin accessibility between CG and control cells at days 13 (top heatmaps) and 25 of differentiation (bottom heatmaps). From left to right, each heatmap shows the average accessibility in the selected regions across the various conditions examined, i.e., in controls at different stages of differentiation and in the corresponding CG cells. Each row in the heatmaps represents a differentially accessible peak, sorted by strongest and weakest differences from top to bottom. We had also included the ChIP-seq coverage of the experiments targeting the CG fusion in long-term cells and M-07e for comparison, but we have removed this panel in response to the following point 3. To aid interpretation, we added a “barcode” alongside the heatmaps indicating which of the peaks (rows) overlap with a CG-binding peak. These plots are conceptually equivalent to the plot in Figure 2b, which shows differentially accessible peaks in long-term CG cells.

We acknowledge that these plots are complex and contain dense information. However, we believe that Figure 2b effectively demonstrates how many of the regions that are accessible in long-term CG cells **(a)** become gradually more accessible in day 13 and 25 CG cells, **(b)** remain inaccessible in control cells, and **(c)** are also directly bound by the CG fusion TF.

Similar patterns are observed in regions that begin to change accessibility at earlier stages, which is why we included Supplementary Figure 9b for completeness. However, we agree that this should have been explained more clearly and therefore we have now revised the figure and the accompanying text (lines 141-143) to improve clarity. Please find a copy of the revised figure panel below.



3. Figure 2: ChIP-seq was only performed in long-term cells, yet the authors try to correlate ChIP-seq peaks with accessibility at much earlier timepoints (i.e. D13, D25). How relevant is this? Perhaps it is not surprising that only a small % of open regions correspond to a CG binding site at D13? Would it not be worthwhile to do ChIP-seq at D13 and D25 as well if you wish to make conclusions regarding CG binding and accessibility?

We thank the reviewer for this important point. It is correct that ChIP-seq was only performed with expanded long-term cells, as these provided the necessary cell numbers and homogeneity for robust profiling. We agree that CG binding at earlier stages cannot be directly inferred from long-term ChIP-seq data. However, CG expression starts around D5 (Fig. S4b) and increases over time with its level peaking in long-term cells, suggesting that chromatin binding and remodeling may be initiated as early as day 13, even if most accessibility changes are not yet significant (Fig. 2a-c, Supplementary Fig. 9b). Performing ChIP-seq at earlier timepoints such as

day 13 or day 25 is technically challenging due to limited cell availability and heterogeneity, which would likely compromise data quality; and we would expect only limited new insights from these experiments.

Thus, we refrained from further ChIP-seq experiments, and, as already mentioned above, we removed the ChIP-seq coverage tracks from unmatched timepoints (Supplementary Fig. 9c) to avoid misinterpretation, but we still highlight regions bound by CG in long-term cells.

4. *Supplementary Figure 10: This figure just shows a few examples of where CG binding, RNA-seq and ATAC-seq correlate as described in the article body. Can this be demonstrated in a more global way with significance attached? Or have the authors primarily cherry picked examples that support their conclusion? Furthermore, this figure is very hard to decipher and it is not entirely clear that the figures support the conclusion.*

We thank the reviewer for this valuable comment and the opportunity to clarify the purpose of Supplementary Fig. 10. The regions in this figure were selected as illustrative examples focusing on loci with known relevance to AML, including both activated (enhancers bound and opened by CG) and repressed genes (enhancers/promoters bound but closed by CG). At this point, our intention was not to present a global analysis, but rather to highlight the diversity and complexity of CG's regulatory effects through representative examples. A more global, in-depth analysis was performed later on by enrichment analysis of differentially accessible chromatin regions in relation to the modules of genes linked to different MK-like cell states shown in Fig. 5a-c, or via the multi-omics analyses in Fig. 5d, 6c and Supplementary Fig. 14d.

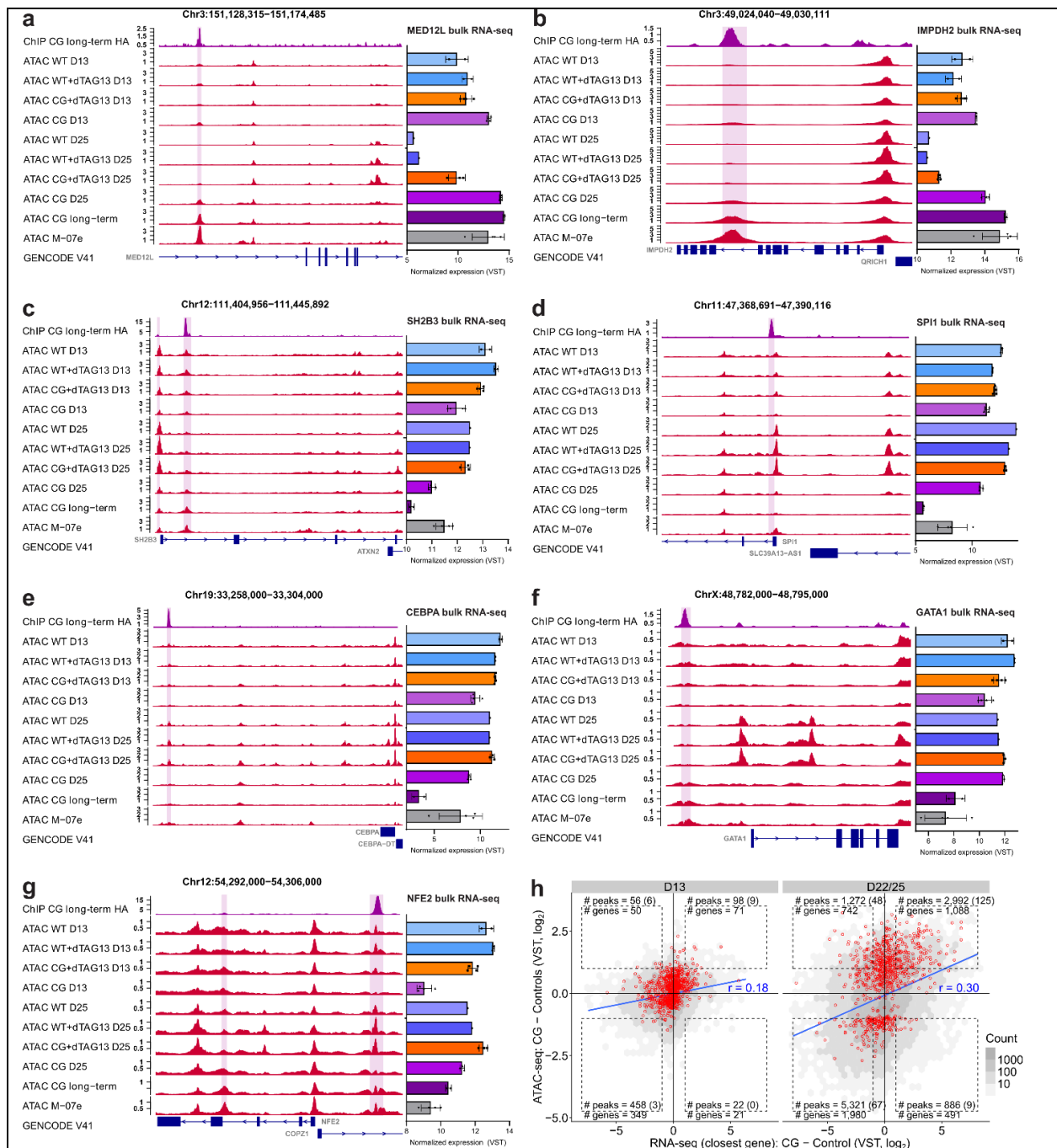
To address the reviewer's concern and avoid any impression of selective reporting, we have clarified the illustrative nature of these examples in the manuscript text (lines 150-153). We have also revised the figure to improve readability by enlarging the panels and clearly marking CG-bound regions in each genomic region.

We also added an additional global analysis directly comparing changes in chromatin accessibility and RNA expression of the next-closest gene (well acknowledging this as an imperfect proxy for regulatory targets) between CG-expressing and control cells during differentiation, which is described in the revised manuscript (lines 156-161) and revised Supplementary Fig. 10h.

While the relationship between chromatin accessibility and expression of the closest genes is generally modest, we observed a clear positive correlation, that increases over time and the effects of the CG fusion expression become more pronounced with Pearson correlation coefficients increasing from $r_{D13} = 0.18$ to $r_{D22/25} = 0.30$. These data suggest that CG's regulatory effects become more prominent as differentiation progresses. Most genes peak/gene pairs fall

into the “upper right” sector of the correlation plot, indicating a positive correlation between accessibility and expression. We additionally highlighted in these plots ATAC-seq peaks which overlap with CG-binding sites (based on CG ChIP-seq data from long-term CG cells). These regions predominantly show increased accessibility and are associated with higher expression of nearby genes, supporting the conclusion that CG binding directly contributes to chromatin remodeling and transcriptional activation at least in a subset of targets.

We hope these improvements address the reviewer's concerns and provide a clearer picture of the correlations between CG binding, RNA-seq, and ATAC-seq data.



Revised Supplementary Figure 10.

We improved panels a-g by making them larger and annotating ChIP-seq peaks and added a new panel (h) with scatterplots directly comparing fold changes in gene expression (x-axis; based on RNA-seq) and chromatin accessibility (y-axis; based on ATAC-seq) between CG-expressing and control cells at D13 (left) and D22/25 (right). Each data point represents one ATAC-seq peak which has been linked with the next-closest gene (hence there may be multiple points for the same gene). The color of the hexagons represents the density of points in each plot region and regions overlapping CG ChIP-seq peaks have additionally been highlighted by red circles. The blue lines represent a linear fit and the Pearson correlation coefficients are indicated. The dashed boxes enclose peak/gene with absolute fold change ≥ 2 and we indicated the number of peaks (number in brackets = peaks overlapping CG binding sites) and genes in each sector.

5. *Figure 5e: This analysis is entirely based on findings in unrelated cells and/or cancer models. While it may be indicative, it feels incomplete and not of high value. Perhaps these data would have benefited from some follow up confirmation work, using these inhibitors on the AMKL cells themselves?*

We thank the reviewer for this thoughtful comment. We fully agree that the gene sets used for the Enrichr analysis shown in Fig. 5e are derived from unrelated cancer cell models, and that without experimental validation in our specific system, the conclusions drawn from this analysis remain speculative. To maintain the clarity and integrity of our manuscript, we have decided to remove Figure 5e and the corresponding text (lines 250-258). We believe this improves the focus of the manuscript and ensures that all presented data are directly supported by our experimental system.

Reviewer #3 (Remarks to the Author):

The authors present a human iPSC-based model of CBFA2T3::GLIS2 (CG)-driven AMKL, in which CG is conditionally expressed at physiological levels from the endogenous CBFA2T3 locus. This system establishes oncogene dependence and enables precise dissection of CG function. Through single-cell RNA-seq and ATAC-seq, the authors chart the transcriptional and chromatin landscape during early hematopoiesis and show that CG disrupts differentiation via two distinct routes: by maintaining aberrant megakaryocyte progenitors in a self-renewing state and by blocking differentiation through downregulation of some key transcription factors. The study is technically strong and provides comprehensive and modern transcriptional profiling, which in itself warrants publication. However, its primarily descriptive nature limits its broader impact. This should be acknowledged and justified in the discussion. Additionally, while the expression of MKP- and MK-associated transcription factor modules likely reflects their functional activity, a more detailed mechanistic analysis or discussion would strengthen the study's depth.

We thank the reviewer for the thoughtful evaluation of our study, and for acknowledging its technical strength and the value of the transcriptional and chromatin profiling we present. We agree that our study is primarily descriptive in nature. Our comprehensive characterization of CG-driven leukemogenesis in a human iPSC-based model provides a critical basis for future mechanistic investigations. By systematically charting the transcriptional and chromatin landscape, we provide a crucial foundation that enables the identification of key regulatory elements and pathways involved in CG-driven leukemogenesis.

As suggested, we have revised the discussion to better reflect the scope and limitations of our work. Specifically, we now emphasize the importance of gene regulatory networks (GRNs) and highlight how our findings lay the groundwork for future mechanistic investigation. At the same time, we point out the limitations of our study to avoid any potential overinterpretation by readers (lines 335-344, 363-364).

We expanded our discussion of the MKP- and MK-associated transcription factor (TF) modules, as we believe that aMKP specific GRNs play an important role in CG-mediated transformation. Several of these TFs (e.g. DLX3, ERG) are directly bound and transcriptionally activated by CG, and their downstream regulon activity aligns with their expression pattern. We now discuss in more detail how these TF modules may contribute to aberrant megakaryopoiesis, referencing recent literature that supports their involvement in leukemogenesis and other malignancies: e.g. CEBPA repression by MECOM in AML (PMID 39803492, DOI 10.1101/2025.01.02.631046); SPI1 repression by RUNX1 and GATA3 (PMID 34180951); distal ERG enhancer activity (PMID 37595278); DLX1 and TMPRSS2::ERG cooperation in prostate cancer (PMID 34493733); involvement of DLX TFs in H3G34 mutated glioma (PMID 39232581).

We hope these clarifications and additions address the concerns raised and justify the approach we have taken in this study.