

CBFA2T3-GLIS2 mediates transcriptional regulation of developmental pathways through a gene regulatory network

Received: 6 March 2023

Accepted: 3 October 2024

Published online: 09 October 2024

 Check for updatesElizabeth A. R. Garfinkle^{1,5}, Pratima Nallagatla^{1,5}, Binay Sahoo¹, Jinjun Dang¹, Mohammad Balood¹, Anitria Cotton², Camryn Franke¹, Sharnise Mitchell³, Taylor Wilson⁴ & Tanja A. Gruber¹ ✉

CBFA2T3-GLIS2 is a fusion oncogene found in pediatric acute megakaryoblastic leukemia that is associated with a poor prognosis. We establish a model of *CBFA2T3-GLIS2* driven acute megakaryoblastic leukemia that allows the distinction of fusion specific changes from those that reflect the megakaryoblast lineage of this leukemia. Using this model, we map fusion genome wide binding that in turn imparts the characteristic transcriptional signature. A network of transcription factor genes bound and upregulated by the fusion are found to have downstream effects that result in dysregulated signaling of developmental pathways including NOTCH, Hedgehog, TGF β , and WNT. Transcriptional regulation is mediated by homo-dimerization and binding of the ETO transcription factor through the nerve homology region 2. Loss of nerve homology region 2 abrogated the development of leukemia, leading to downregulation of JAK/STAT, Hedgehog, and NOTCH transcriptional signatures. These data contribute to the understanding of *CBFA2T3-GLIS2* mediated leukemogenesis and identify potential therapeutic vulnerabilities for future studies.

Pediatric acute megakaryoblastic leukemia (AMKL) comprises 4–15% of pediatric acute myeloid leukemia (AML) patients and is found in patients with Down syndrome (DS-AMKL) and in patients without Down syndrome (non-DS-AMKL)¹. Children <4 years of age with Down syndrome have a 500-fold greater incidence of AMKL than the general population and have an event-free survival (EFS) approaching 90% with standard chemotherapy². Malignant transformation of megakaryoblasts in DS-AMKL is due to trisomy 21 and somatic mutations in the *GATA1* gene, which are rarely found in non-DS-AMKL patients^{3,4}. Prior to next-generation sequencing studies, the genomic alterations driving non-DS-AMKL were not well defined. In 2011, we reported a recurrent cryptic inversion event identified by RNA sequencing on chromosome 16 that produced the coding fusion *CBFA2T3-GLIS2* in a

subset of pediatric non-DS-AMKL patients^{5,6}. A follow-up study in 2017 revealed seven major genomic subtypes in pediatric non-DS-AMKL, five of which contain recurrent chromosomal translocations that code for chimeric transcription factors: *HOX* rearrangements (*HOXr*), *RMB15-MKLL1*, *NUP98-KDMSA*, *KMT2A* rearrangements (*KMT2Ar*), and *CBFA2T3-GLIS2*⁴. *CBFA2T3-GLIS2* was the most prevalent and the most aggressive of the fusions, with overall survival of only 14% despite maximum tolerated high-intensity chemotherapy and stem cell transplant in the first remission⁴. Patients harboring the *CBFA2T3-GLIS2* fusion have significantly fewer cooperating somatic mutations compared to patients with other subtypes of non-DS-AMKL, suggesting *CBFA2T3-GLIS2* is the driver of leukemogenesis⁵. Prior chromatin immunoprecipitation sequencing studies (ChIP-seq) of the

¹Department of Pediatrics, Stanford University School of Medicine, Stanford, CA, USA. ²Division of Experimental Hematology, St. Jude Children's Research Hospital, Memphis, TN, USA. ³Department of Chemical Biology and Therapeutics, St. Jude Children's Research Hospital, Memphis, TN, USA. ⁴Department of Immunology, St. Jude Children's Research Hospital, Memphis, TN, USA. ⁵These authors contributed equally: Elizabeth A. R. Garfinkle, Pratima Nallagatla.

✉ e-mail: tagrub@stanford.edu

CBFA2T3-GLIS2 fusion in engineered cell lines using a GFP tag and a CBFA2T3 antibody identified <2000 binding sites genome-wide with only 20% of target genes differentially expressed and 15% carrying a GLIS2-binding motif⁷. A significant number of peaks were found to have ETS consensus binding sites, and ERG ChIP-seq confirmed overlapping binding in 50% of the identified CBFA2T3-GLIS2 target genes; however, a biochemical association between the fusion and ERG has not been demonstrated. An analysis of super-enhancers identified *ETO*, *ERG*, *PDGFRA*, and *KIT* in cell lines. However, neither of these studies distinguished super-enhancers specific to the fusion as compared to the lineage state of the cell^{7,8}. Given the paucity of identified genomic targets, the lack of differential expression in 80% of these targets, and the genomic complexity of cell lines that carry additional mutations that may interfere with and/or impact results, we sought to establish an optimized model system to more thoroughly understand transcriptional regulation by the fusion. Herein we report on the development of a humanized murine model to identify key transcriptional co-factors and dysregulated signaling pathways that contribute to transformation. Advantages of the model include the isolation of *CBFA2T3-GLIS2* as the only genomic event when compared to fusion negative megakaryoblasts and the ability to compare and contrast results in these two populations. The use of cleavage under targets and release using nuclease (CUT&RUN), which requires a lower input and provides superior sensitivity with a high signal-to-noise ratio, allowing for increased resolution of genes targeted by the fusion in our study⁹. With this approach, we were able to comprehensively map genome-wide binding with accuracy and identify fusion-specific changes to enhance our understanding of the dysregulated signaling and developmental pathways that may be targeted therapeutically.

Results

CBFA2T3-GLIS2 drives the transformation of human megakaryoblasts

Current models to characterize CBFA2T3-GLIS2 positive non-DS-AMKL are limited. Immortalized cell lines derived from patients harboring this fusion, such as MO7e, have additional acquired mutations, such as in *TP53*, that are not represented in the patient population^{4,5}. In contrast to other fusions found in pediatric non-DS-AMKL, *CBFA2T3-GLIS2* gene-modified murine bone marrow cells fail to induce a serially transplantable leukemia in the absence of a cooperating mutation^{10,11}. Several CBFA2T3-GLIS2-positive patient-derived xenograft models are available; however, patient cells are difficult to manipulate for experiments, cell quantity is limited, and they do not grow well in vitro. A reliable model of CBFA2T3-GLIS2-positive AMKL that recapitulates primary patient samples and can be experimentally manipulated is needed.

To develop a model, CD34+ stem cells were isolated from human cord blood and transduced with a lentiviral vector encoding N-terminal tagged *CBFA2T3-GLIS2* and the *GFP* reporter gene. The transduced stem cells were cultured in vitro in the presence of IL1 β and TPO to promote megakaryoblast differentiation, flow-sorted for GFP purity, and transplanted into NOD.scid.II2Ryc^{null} immunodeficient mice with the *IL3*, *CSF2*, and *KITLG* human transgenes (NSG-SGM3)¹². In vivo expanded and recovered cells were defined as mouse CD45 negative and GFP positive (Fig. 1A). The fusion-positive megakaryoblasts induced a serially transplantable leukemia within 200 days compared to empty vector transduced megakaryoblasts lacking the fusion (Fig. 1B, Supplementary Data 1). Immunophenotyping of fusion-positive megakaryoblasts was consistent with patient specimens as determined by proteomics (Supplementary Fig. 2). 39% (12/31) of moribund mice presented with hind leg paralysis, and 45% (14/31) had enlarged spleens (Supplementary Data 1). Histology revealed neoplastic infiltrates in the bone marrow, spleen, and central nervous system, the latter of which caused the hind leg paralysis due to compression of the spinal cord (Supplementary Fig. 1A, B, Supplementary Data 1). One mouse developed a hind leg

tumor (Supplementary Fig. 1A, Supplementary Data 1), and nine mice had neoplastic infiltrates in the lung (Supplementary Fig. 1B, Supplementary Data 1). The latency of disease and moribund phenotype were consistent across serial transplantations, suggesting the model is genomically stable (Fig. 1B, Supplementary Data 1). Expression of the *CBFA2T3-GLIS2* fusion gene was confirmed through RT-PCR (Supplementary Fig. 1C) and Western blot (Supplementary Fig. 1D).

We have previously shown that CBFA2T3-GLIS2-positive AMKL patient specimens ($N=9$) display a distinct gene expression profile compared to other AMKL genomic subsets ($N=35$)^{4,5}. To understand how closely our model mimics the transcriptional program found in patients, we subjected cultured cells as well as primary, secondary, and tertiary leukemias to RNA sequencing (Fig. 1C). An average of 18,660 transcripts were expressed with at least one read in our fusion-positive leukemic megakaryoblasts. The serially transplanted leukemic megakaryoblasts ($N=15$) and fusion-positive megakaryoblasts cultured in vitro ($N=9$) before transplantation were found to recapitulate the gene expression profile of fusion-positive patient specimens (Fig. 1C) with a strong correlation ($R=0.91$, $p<2.2e-16$) (Supplementary Fig. 1E). To assess our model on the translational level, tandem mass tag (TMT)-based quantitative proteomics was performed on cells from two transplanted leukemias and four CBFA2T3-GLIS2-positive patient-derived xenograft specimens¹³. Fusion-negative megakaryoblasts cultured in vitro were used as a control. Consistent with our transcriptome data, we found a strong correlation between transplanted leukemia cells and patient samples ($R=0.98$, $p<2.2e-16$) (Supplementary Fig. 1F, Supplementary Data 2), further supporting that this humanized murine model faithfully replicates primary patient samples phenotypically and can be accurately used to characterize the CBFA2T3-GLIS2 transcriptional complex. Spearman correlation of our transcriptome and proteome data was calculated using paired secondary ($N=1$) and tertiary ($N=1$) transplanted cells that underwent both RNA-sequencing and proteomic analysis¹⁴. The Spearman's correlation coefficient was 0.46 (Supplementary Fig. 1G, Supplementary Data 2), a moderate correlation that compares favorably with prior publications that have looked at the paired transcriptome and proteome data and suggest post-transcriptional modifications and translational regulation that can be investigated in future studies¹⁴.

CBFA2T3-GLIS2 binds at GC-rich promoter regions

To map genome-wide binding of CBFA2T3-GLIS2 in our humanized model, cleavage under targets and release using nuclease with sequencing (CUT&RUN-seq) was done on CBFA2T3-GLIS2-positive leukemic megakaryoblasts from our murine model ($N=2$) using an anti-TY1 antibody to recognize the N-terminal tag engineered on our fusion construct (see the "Methods" section). A total of 28,076 regions were identified, which corresponded to 5256 unique genes bound by the fusion at the promoter. 92% of these genes were expressed, and 55% were differentially expressed when compared to fusion-negative megakaryoblasts (Fig. 2A, Supplementary Data 3). Although a significant portion of peaks were located at gene promoters (52% of total gene regions), 27% and 16% were bound at intron and intergenic regions respectively, suggesting the fusion is also regulating genes within introns and binding at enhancer regions located upstream of promoter sites (Fig. 2A). Of the 55% differentially expressed genes bound by the fusion at promoters, 57% (1517/2657) were downregulated and 43% (1140/2657) were upregulated (Fig. 2B, Supplementary Data 3). Over-representation analysis (ORA), a statistical method to measure the enrichment of certain pathways in a subset of data, revealed a significant upregulation of genes associated with the WNT, TGF β , and Hedgehog signaling pathways (Fig. 2B, Supplementary Data 4), consistent with enriched pathways in CBFA2T3-GLIS2-positive patient samples as previously published^{5,15,16}.

To further characterize the altered gene expression profile of fusion-positive leukemic megakaryoblasts and understand the gene

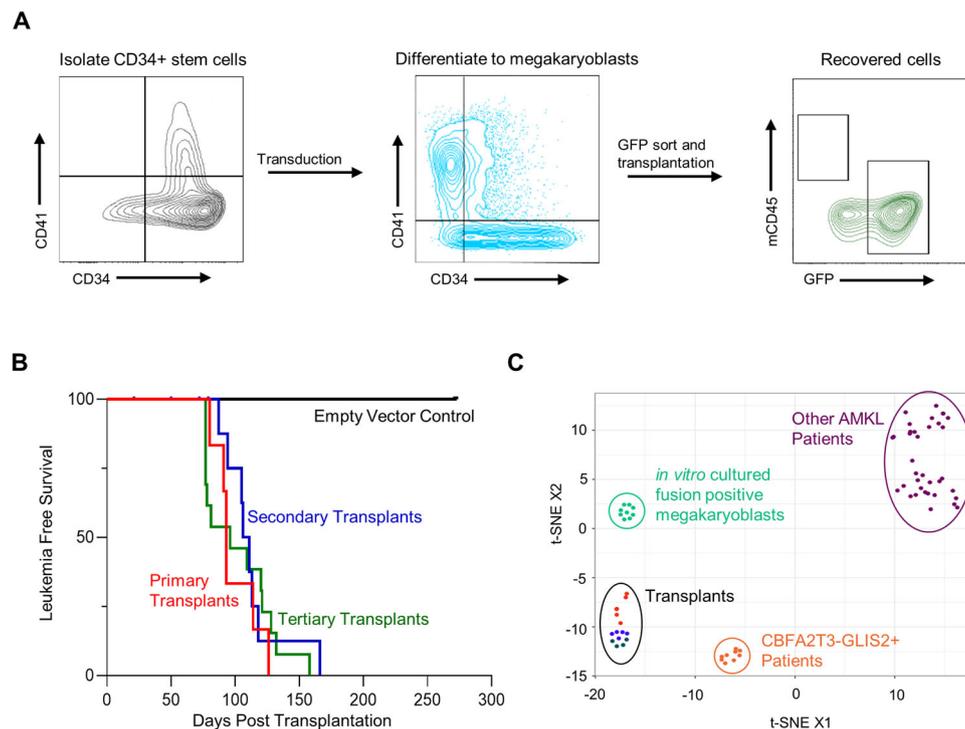


Fig. 1 | CBFA2T3-GLIS2 drives the transformation of primary human megakaryoblasts. **A** Generation of CBFA2T3-GLIS2-positive primary human megakaryoblasts. CD34⁺ stem cells were isolated from human cord blood, transduced with a lentivirus encoding the *CBFA2T3-GLIS2* coding sequence and a GFP reporter, followed by differentiation to megakaryoblasts with human TPO and IL1 β for 7–9 days. Cultures were sorted for GFP⁺ purity and transplanted into immunodeficient NSG-SGM3 mice. Recovered cells from transplantation were depleted for mouse CD45 cells, and GFP expression was confirmed by flow cytometry. **B** Serial transplantation (primary, $N = 6$; secondary, $N = 11$; and tertiary, $N = 14$) of fusion-

positive megakaryoblasts induce AMKL within 200 days compared to empty vector megakaryoblasts ($N = 6$) which do not expand in vivo. Leukemia-free survival is shown. Source data are provided as a Source Data file. **C** t-SNE analysis of the gene expression profiles of in vitro cultured fusion-positive megakaryoblasts (light green), fusion-positive patient samples (orange), and AMKL patient samples harboring a different genetic driver other than the CBFA2T3-GLIS2 fusion (purple). Clusters are based on the top 1% of differentially expressed genes.

regulatory network that leads to dysregulation of genes in the absence of direct CBFA2T3-GLIS2 binding, we performed a global differential gene expression analysis comparing the two transplanted samples and the three fusion negative megakaryoblast samples. This analysis again showed differential expression with 54% of genes (4215) down-regulated, and 46% (3590) upregulated compared to fusion negative megakaryoblasts (Supplementary Fig. 3A, Supplementary Data 3). In total, 68% (5275/7805) of dysregulated genes were not directly bound by CBFA2T3-GLIS2. Given the global-scale, we used gene set enrichment analysis (GSEA) and the Hallmark database to explore enriched pathways and found a consistent upregulation of WNT, TGF β , and Hedgehog signaling-associated genes but also found upregulation of genes in the NOTCH and JAK/STAT pathways (Supplementary Fig. 3A, Supplementary Data 4), suggesting the fusion is dysregulating components of these pathways through binding of transcription factors that mediate these signaling pathways^{17–19}. To model the transcription factor gene regulatory network (GRN), we utilized GeneNetworkBuilder, a pipeline for building regulatory networks using genome-wide binding and gene expression data²⁰. The top 74 nodes of this GRN when ranked by degree centrality (number of connections to and from a node), include transcription factors involved in JAK/STAT signaling (STAT2, STAT4, STAT6, JUN, JUNB), canonical and non-canonical WNT signaling (TCF7L2, NFATC3, NFYA, JUN), TGF β (SMAD2, TGIF1, ZEB1), NOTCH (RBPJ), RUNX (RUNX1, RUNX3), and megakaryocyte development (IRF2, FLII, GFI1B) pathways (Fig. 2C Supplementary Fig. 3B–G, Supplementary Data 5). We did not identify any nodes related to the Hedgehog signaling pathway, consistent with the absence of binding of CBFA2T3-GLIS2 to the *GLI1-3* genes and the

absence of GLIS1-3 in the GeneNetworkBuilder algorithm (Supplementary Data 3). Several transcription factors that are downstream effectors of multiple pathways were found, including JUN, which has been reported to function downstream of JAK/STAT, TGF β , and WNT signaling. Additionally, JUN regulates the expression of members of the RUNX family, including RUNX2 and RUNX3, which play a role in leukemic progression, underscoring the significant cross-talk seen within the transcriptional network mediated by CBFA2T3-GLIS2 (Fig. 2C)²¹. Although the network was dominated by oncogenic and developmental transcription factors, a significant central node, IRF2, was identified as part of the megakaryocyte development pathway (Supplementary Fig. 3G). While IRF2 is primarily known as an antagonist of the tumor suppressor IRF1, studies have also shown that IRF2 expression in myeloid progenitor cells leads to reprogramming towards the megakaryocytic lineage and enables them to respond to thrombopoietin^{22,23}. Collectively, this analysis provides a map of the direct and indirect effects of *CBFA2T3-GLIS2* expression in megakaryoblasts.

Given that the fusion retains the five highly conserved DNA-binding zinc fingers of full-length GLIS2, we hypothesized a majority of genes bound by the fusion would contain a G-C-rich GLIS or the closely related GLI transcription factor consensus sequence. Homer's known motif search was used to identify the top represented consensus sequences from the promoter peak regions (filtered set=10,619) ranging in size from 34 to 11,355 base pairs associated with the 5256 fusion-bound genes. Consistent with previously published data in the MO7e cell line, ETS, GATA, and RUNX consensus sequences were highly represented (Supplementary Fig. 3H)⁷. Although we did not find a GLIS or GLI motif,

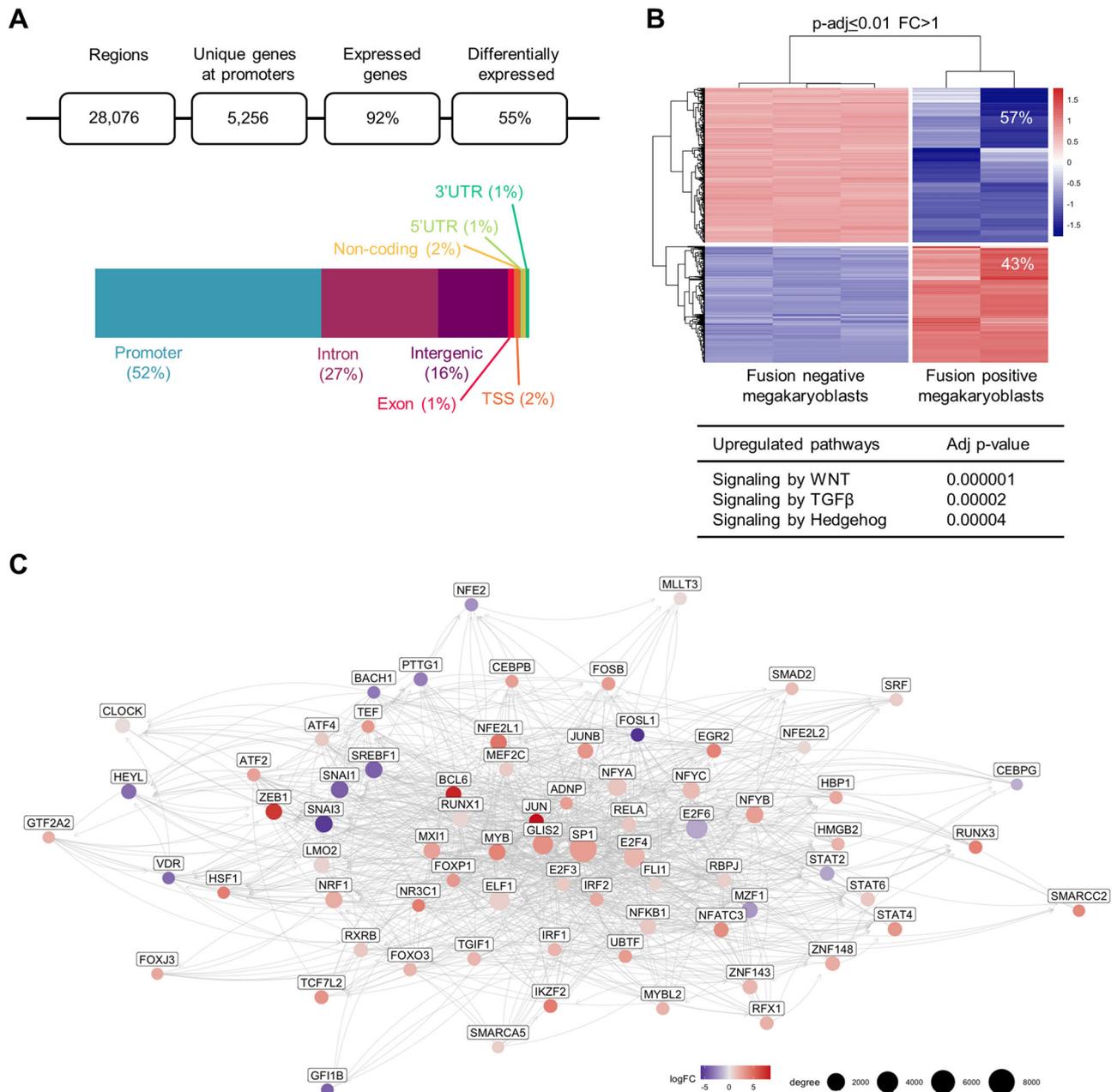


Fig. 2 | CBFA2T3-GLIS2 binds at GC-rich promoter regions to alter gene expression. **A** Summary of CUT&RUN-seq data from fusion immunoprecipitation. The percentage of fusion-bound genes that are expressed and differentially expressed in the cell is shown. Distribution of CUT&RUN-seq peak locations from fusion immunoprecipitation. UTR untranslated region, TSS transcription start site. Source data are provided as a Source Data file. **B** RNA extraction and gene expression profiling were performed on three technical replicates from fusion-negative megakaryoblasts cultured in vitro and fusion-positive leukemic megakaryoblasts recovered from two secondary transplantations. A heatmap of

differentially expressed genes bound by the fusion at promoter regions is shown ($N = 2657$). Log₂ fold change (FC) was calculated using a generalized linear model with a negative binomial distribution in DESeq2 and p -values derived from the Wald t -test and corrected using the Benjamini–Hochberg correction. Over-representation analysis using the Reactome database of differentially expressed genes bound by the fusion. p -values were calculated using a one-sided Fisher's exact and adjusted for multiple comparisons using the Benjamini–Hochberg correction. **C** Top 74 nodes from transcription factor gene regulatory network (GRN) analysis.

we did identify the G-C-rich SP1 motif. To compare the SP1 motif to the GLI superfamily sequences, we performed a neighbor-joining analysis which revealed the SP1 consensus sequence clusters closely with the GLIS1–2 and GLIS1–4 consensus sequences (Supplementary Fig. 3I)²⁴. We hypothesized the lack of GLI motifs in the Homer search was due to a lack of representation of GLI and GLIS consensus sequences in the database. Therefore, to focus the search on the GLI superfamily of transcription factors that share a high degree of homology in the DNA-binding zinc finger domains, a find individual motif occurrences (FIMO)

search was used to determine the abundance of the GLIS and GLI family motifs in our data²⁵. This custom search revealed the GLIS1–2 consensus sequence (DCWHNGTGGGGGGTV) was represented in 39.8% of the promoter regions, suggesting this is the preferential binding sequence. GLIS1–4 (SNSYSYRGGRRGBV) was also represented in 23.6% of regions and SP1 (NRGGGCGGGCCBDGGVSSSGV) in 36.5% (Supplementary Fig. 2J). One or more GLIS1–2, GLIS1–4, and SP1 sites overlapped with fusion-bound sites in 77.5% (4072/5256), 77.8% (4089/5256), and 81% (4246/5256) unique genes, respectively (Supplementary Fig. 3K,

Supplementary Data 6). Fusion preferential binding for GLIS2 consensus sequences is supported by Vasanth et al. who reported full-length GLIS2 binds GLIS binding sequences with a stronger affinity than GLI binding sequences²⁶.

CBFA2T3-GLIS2 alters super-enhancers

To define the super-enhancer landscape in fusion-positive leukemic megakaryoblasts, we first mapped H3K27ac occupancy in fusion-negative and fusion-positive leukemic megakaryoblasts using CUT&RUN-seq. We identified 1537 new H3K27ac peaks in fusion-positive leukemic megakaryoblasts, 56% of which overlap with fusion-bound genes (Supplementary Fig. 4A, Supplementary Data 3). The ranking of super-enhancers (ROSE) algorithm was utilized to identify super-enhancers (SE) in both fusion-negative and fusion-positive leukemic megakaryoblasts. The signals from two biological replicates were averaged together (individual replicates are shown in Supplementary Fig. 3B, C) and identified 263 and 236 unique SE restricted to fusion-negative and fusion-positive leukemic megakaryoblasts, respectively. 74% of SE regions unique to the fusion-positive cells were bound by the fusion (Fig. 3A, B, Supplementary Data 7). As expected, a majority (61%) of the SE identified in fusion-negative megakaryoblasts led to upregulation of their associated genes (Fig. 3C). Similarly, a majority (60%) of the SE identified in fusion-positive leukemic megakaryoblasts were upregulated when compared to fusion negative megakaryoblasts (Fig. 3C) and this pattern held true when restricted to fusion-positive SE that were also bound by the fusion (Fig. 3C).

The 236 unique SE restricted to fusion-positive leukemic megakaryoblasts included gene targets related to pro-survival, hematopoietic differentiation, and developmental pathways (Fig. 3, Supplementary Data 7). Consistent with prior reports, we found SE associated with the *BCL2* and *KIT* genes⁷. 171 SE were present that have not been previously reported, including *TCF7L2*, which encodes a transcription factor that plays a role in the regulation of the WNT signaling pathway (Fig. 3A-D, Supplementary Data 7)²⁸. *TCF7L2* has also been shown to be required for the survival and proliferation of AML cells²⁷. Consistent with a block in differentiation, we found the *TESC* SE present in fusion-negative megakaryoblasts to be lost and expression of this gene, which is required for megakaryoblast differentiation, downregulated in fusion-positive leukemic megakaryoblasts (Fig. 3D)²⁸. Thus, a subset of genes are dysregulated through SE gains and losses in addition to regulation provided via promoter binding.

ETO, CtBP1, and p300 associate with the CBFA2T3-GLIS2

The CBFA2T3 portion of the fusion retains three highly conserved nery homology regions (NHR). NHR1 is a binding site for p300, a transcriptional co-activator and histone acetyltransferase²⁹. NHR2 contains a hydrophobic heptad repeat region that plays a key role in homo- and hetero-oligomerization with the closely related transcriptional repressor ETO (also known as RUNX1T1 and CBFA2T1)³⁰. ETO can also bind p300 thereby suggesting both transcriptional repression and activation can be altered through the hetero-oligomerization of ETO via NHR2³¹. While NHR3 does not contain any highly conserved domains and is not well characterized, it is suggested to support oligomerization³⁰. In addition to the five DNA binding zinc-fingers retained in the GLIS2 portion of the fusion, there is one PXDLS consensus sequence and one degenerate sequence capable of binding CtBP1, a transcriptional repressor^{32,33}. p300, like many histone acetyltransferases, also contains a highly conserved PXDLS sequence in the center of its bromodomain that can bind to CtBP1, which, in turn, represses the transactivation activity of p300³⁴. Based on these data, we hypothesized the CBFA2T3-GLIS2 fusion is able to recruit and associate with ETO, CtBP1, p300 and homo-dimerize to bi-directionally alter gene expression at fusion-bound genes (Supplementary Fig. 5A).

To establish an association of ETO, CtBP1, and p300 with the fusion we conducted co-immunoprecipitation (co-IP) experiments in CBFA2T3-GLIS2 leukemic megakaryoblasts from our murine model (Fig. 4A, Supplementary Fig. 5B). Immunoprecipitation of ETO, CtBP1 and p300 followed by blotting for the fusion using the 3x-TY1 tag confirmed our predicted associations (Fig. 4A, Supplementary Fig. 5B). We then performed CUT&RUN-seq for ETO, CtBP1, and p300 in transplanted leukemic megakaryoblasts and in vitro cultured fusion negative megakaryoblasts to map co-occupancy with the fusion at target genes (Fig. 4B-D, Supplementary Data 3) and assess genome-wide shifts in the binding of these factors in the presence of the fusion (Fig. 4C, Supplementary Data 3). Overall, we found an increase in the absolute number of unique genes bound at promoter regions by ETO and p300 in megakaryoblasts harboring the fusion compared to those lacking the fusion, whereas there were a similar number of genes bound by CtBP1 in fusion-positive and negative megakaryoblasts (Fig. 4C, Supplementary Data 3). We then looked specifically at the bound genes retained, gained, and lost in the fusion-positive megakaryoblasts (Fig. 4C). This demonstrated clear genome-wide shifts for all three factors in the presence of the fusion, with ETO and p300 having a greater number of sites gained compared to lost leading to a net increase of 136% and 303% targets respectively. Although CtBP1 was also found to bind new sites across the genome, there was a significant loss of CtBP1 binding to genes resulting in a net loss of 11% (Fig. 4C). The majority of sites gained, irrespective of the co-factor, were co-bound by the fusion (ETO 64%, CtBP1 52%, and p300 57%) (Supplementary Fig. 5C, Supplementary Data 3). Accordingly, heatmaps demonstrated significant enrichment of these co-factors at fusion-bound genes (Fig. 4D, Supplementary Fig. 5D). Collectively, this data argues in favor of the recruitment of these co-factors to new sites in the genome by the fusion itself.

To gain insight into the impact of these co-factors on the expression of target genes, we evaluated the directionality of expression in the presence of the ETO, CtBP1, and p300. Genes bound solely by the fusion served as a baseline, with 57% of genes downregulated and 43% of genes upregulated. Directionality of expression did not appear to be significantly impacted by an association with ETO, CtBP1, or p300 alone (Supplementary Fig. 5E, Supplementary Data 3). Genes bound by all three factors, however, demonstrated an increase in the proportion of genes upregulated (65%, $p = 0.0028$). An evaluation of the differential expression of target gene subsets bound by the fusion and one or more of these co-factors revealed a contribution of each of these co-factors to the dysregulation of oncogenic pathways. For example, target genes associated with RUNX transcriptional regulation were co-occupied by CtBP1, whereas WNT signaling genes were co-occupied by ETO (Table 1). Most significantly enriched pathways (ORA Reactome analysis), irrespective of the co-factor associated with them, were upregulated with the exception of platelet activation, aggregation, and degranulation downregulation associated with ETO and p300 co-occupancy which are likely a reflection of the block in terminal differentiation (Table 1, Supplementary Data 4). The lack of significant enrichment for downregulated gene sets co-occupied by the fusion and its co-factors was surprising given the proportion of genes whose expression is decreased by the fusion and its associated co-factors. ETO and CtBP1 are well known for their repressor activity, but in the context of co-occupancy with CBFA2T3-GLIS2, our data suggests they play a major role in transformation through the transcriptional activation of key pathways associated with proliferation and development.

Loss of NHR2 disrupts ETO association and dimerization

Having demonstrated binding and genomic co-occupancy of the CBFA2T3-GLIS2 fusion with ETO, CtBP1, and p300, we next sought to determine which structural components of the fusion were essential for recruitment of these co-factors to the transcriptional complex. To do

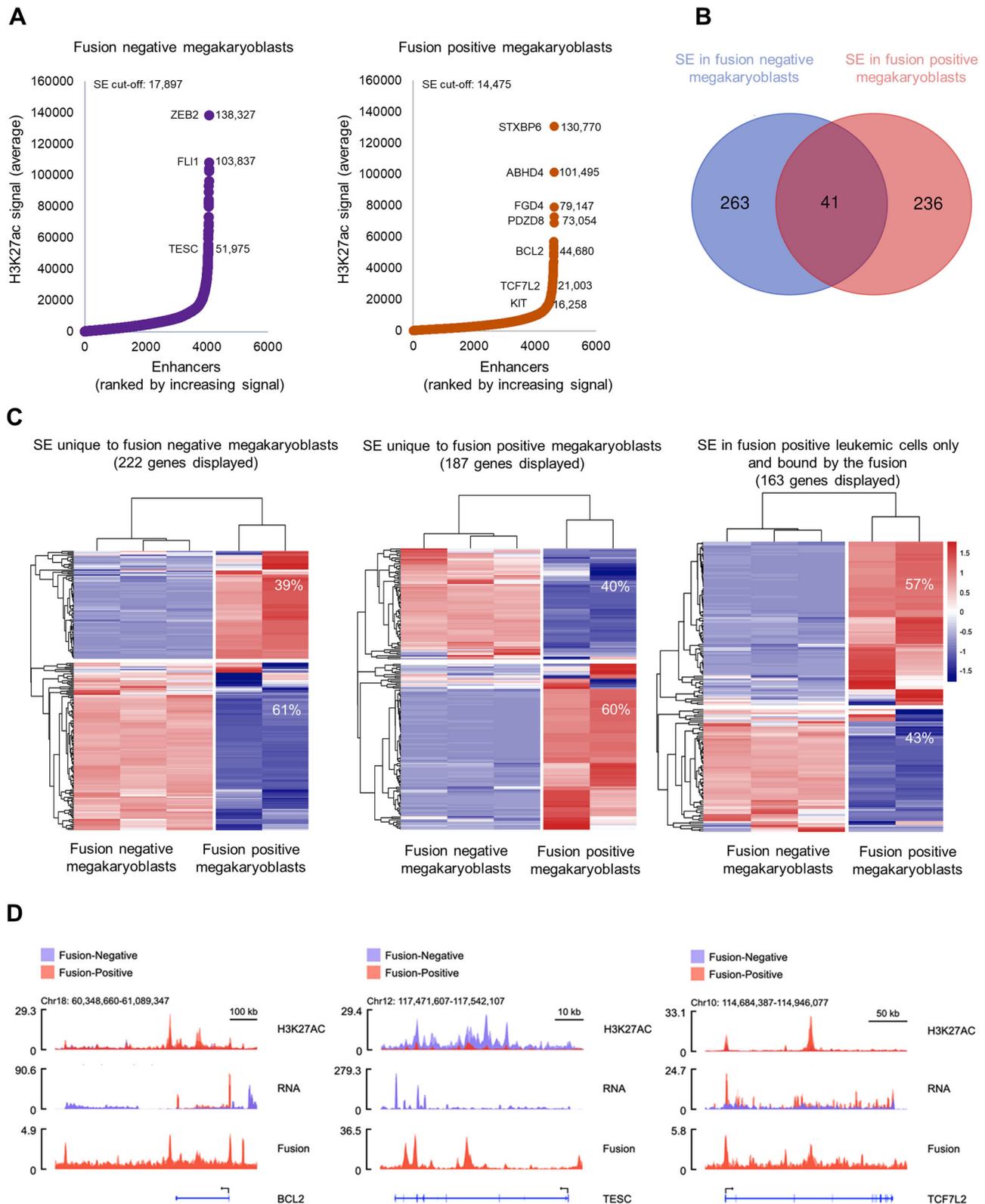


Fig. 3 | CBFA2T3-GLIS2 alters super-enhancers. A Averaged signal-ranking plots of H3K27ac bound genes were identified from CUT&RUN-seq in two fusion-negative megakaryoblast samples and two fusion-positive megakaryoblast samples from secondary transplants. Super-enhancer (SE) cut-off values are displayed on the graphs. Source data are provided as a Source Data file. **B** Venn diagram showing the overlap between SEs in fusion-negative megakaryoblasts (blue) and fusion-positive megakaryoblasts (red). **C** Heatmaps of differentially expressed SEs in fusion-negative megakaryoblasts ($N = 222$), fusion-positive megakaryoblasts

($N = 187$), and fusion-positive megakaryoblasts that are also bound by the fusion ($N = 163$). **D** H3K27ac binding profiles, fusion binding profiles, and corresponding RNA-sequencing tracks at the *TESC*, *BCL2*, and *TCF7L2* SE. The *TESC* SE, essential for megakaryoblast differentiation, is lost in fusion-positive megakaryoblasts. *BCL2*, an anti-apoptotic gene, and *TCF7L2*, a member of the WNT signaling pathway, are upregulated super-enhancers exclusive to fusion-positive megakaryoblasts (generated using SparK v2.6.2)³⁶.

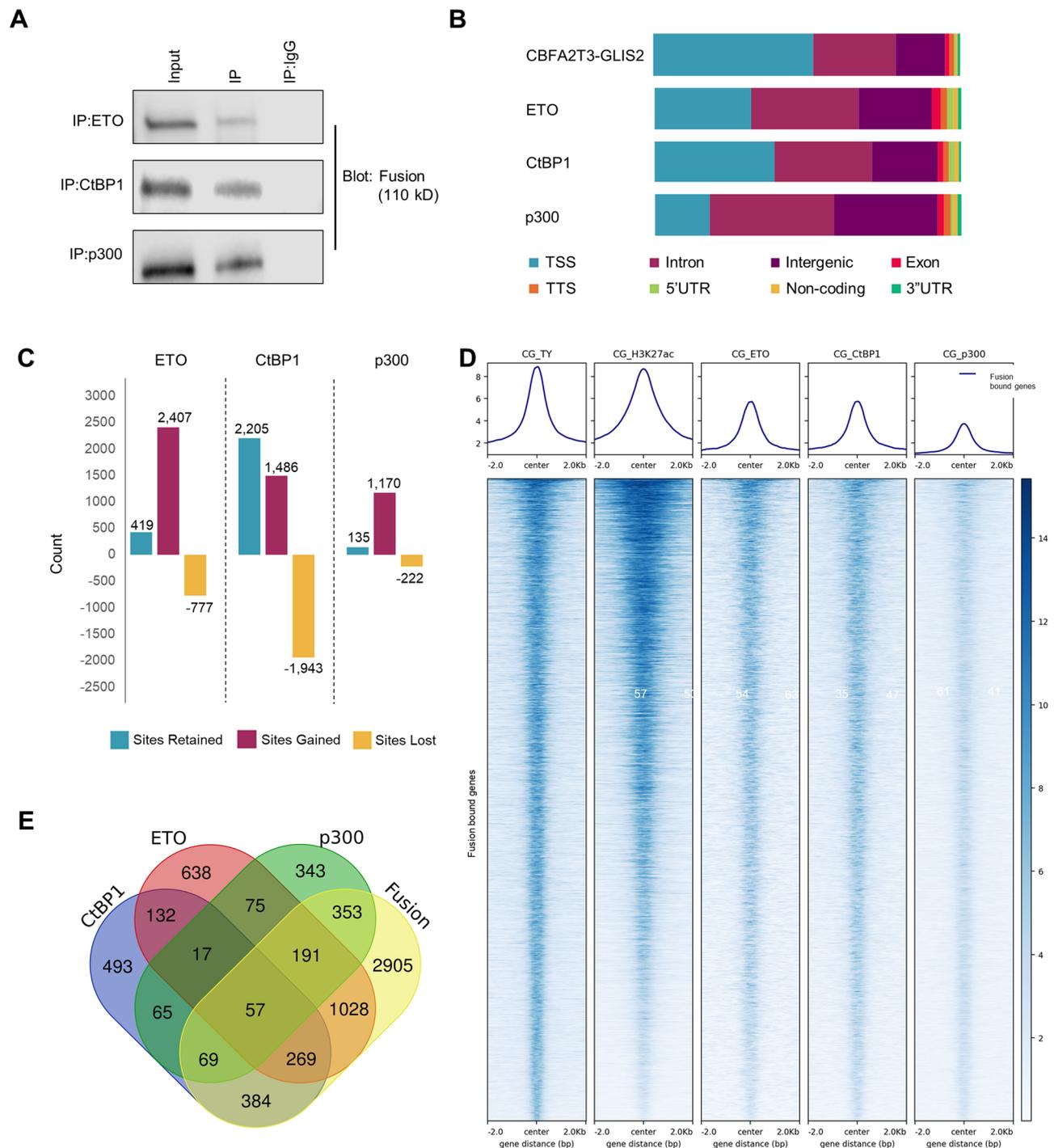


Fig. 4 | ETO, CtBP1, and p300 associated with the CBFA2T3-GLIS2 transcriptional complex. **A** Co-immunoprecipitations using fusion-positive leukemic megakaryoblasts confirm the association of ETO, CtBP1, and p300 with the fusion. All immunoprecipitations were repeated twice, representative blots are shown. Source data are provided as a Source Data file. **B** CUT&RUN-seq has performed in fusion-positive megakaryoblasts from two secondary transplants and fusion negative megakaryoblasts cultured in vitro for ETO, CtBP1, and p300. Distribution of CUT&RUN-seq peak locations for ETO, CtBP1, and p300. CBFA2T3-GLIS2 is included as a comparison. TSS transcription start site, TTS transcription

termination site, UTR untranslated region. Source data are provided as a Source Data file. **C** Summary of ETO, CtBP1, and p300 bound genes retained, gained, and lost in fusion-positive megakaryoblasts compared to fusion-negative megakaryoblasts. Source data are provided as a Source Data file. **D** Heatmaps of enrichment for CBFA2T3-GLIS2 (CG-TY), H3K27ac, ETO, CtBP1, and p300 at genes bound by the fusion. **E** Venn diagram showing the overlap of genes bound at the promoters of ETO, CtBP1, and p300 exclusively in fusion-positive megakaryoblasts and their overlap with fusion-bound genes (Supplementary Data 3).

this, we established a series of mutant *CBFA2T3-GLIS2* constructs for in vitro and in vivo assays based on prior studies of GLIS2 and ETO family members (Supplementary Fig. 5A). In AML1-ETO positive AML, deletion of NHR1 disrupted p300 binding and deletion of NHR2 abrogated ETO

binding^{31,35-38}. We therefore predicted removal of NHR1 from the fusion sequence would lead to a partial loss of p300 due to the continued recruitment of ETO and CtBP1, whereas deletion of NHR2 would abrogate ETO binding and disrupt homo-dimerization. It has

Table 1 | Impact of co-factor binding on over-representation analysis

Pathway	Fusion only*	Fusion p300*	Fusion ETO*	Fusion ETO/p300*	Fusion CtBP1*
Signaling by TGF β	0.004	n.s.	n.s.	n.s.	n.s.
Signaling by WNT	0.05	0.01	0.02	n.s.	n.s.
Signaling by hedgehog	0.05	0.04	n.s.	n.s.	0.05
Signaling by NOTCH4	n.s.	0.03	n.s.	n.s.	n.s.
Transcriptional regulation by RUNX1	n.s.	n.s.	n.s.	n.s.	0.05
Transcriptional regulation by RUNX2	0.01	0.01	n.s.	n.s.	n.s.
Transcriptional regulation by RUNX3	0.03	0.04	n.s.	n.s.	n.s.
Platelet aggregation	n.s.	n.s.	n.s.	0.004	n.s.
Platelet activation	n.s.	n.s.	n.s.	0.01	n.s.
Platelet degranulation	n.s.	n.s.	n.s.	0.01	n.s.

n.s. not significant

*Adjusted *p*-values calculated with a one-sided Fisher's exact test and adjusted for multiple comparisons using the Benjamini–Hochberg correction; all pathways were upregulated with the exception of platelet aggregation, platelet activation, and platelet degranulation, which were downregulated.

been previously demonstrated that mutation of all four PXDLS sequences in the full-length GLIS2 protein abolished CtBP1 binding³⁸. Therefore, we predicted mutation of the two PXDLS residues retained in the fusion (DL380, 487AS) would cause a partial loss of CtBP1 as recruitment of CtBP1 to the complex can still be achieved indirectly through an association of p300 via NHR1 or ETO. To disrupt the association of p300, ETO, and CtBP1, we predicted the fusion would need to have a loss of both NHR1-2 and mutation of the PXDLS residues (Supplementary Fig. 5A).

We went on to perform a series of co-IPs with our mutant constructs in transfected 293T cells, a well-established embryonic kidney cell line that lacks *CBFA2T3-GLIS2* but expresses *ETO*, *CTBP1*, and *EP300*. Cells were transiently transfected for forty-eight hours with a lentiviral construct encoding N-terminal 3xFLAG tagged *CBFA2T3-GLIS2* or one of the several mutant constructs outlined in Fig. 5SA. Consistent with the literature on ETO family members, we found that loss of NHR2 abrogated ETO binding with the fusion (Fig. 5A, Supplementary Fig. 6A). NHR2 also contributes to homo- and hetero-oligomerization. Therefore, to assess the effect of NHR2 deletion on oligomerization, we first co-transfected 293T cells with a lentiviral construct encoding N-terminal 3xTY1-tagged *CBFA2T3-GLIS2* (referred to as wild type) and a lentiviral construct encoding either a wild type N-terminal 3xFLAG tagged fusion or a mutant fusion construct that lacks NHR1, NHR2, or NHR3. We found a significant reduction in, but not complete disruption of, homo-oligomerization upon NHR2 deletion, providing support for the role of NHR3 in oligomerization (Fig. 5B). NHR1 and NHR3 deletion did not have a major impact on the ability to dimerize with the wild type fusion. To confirm our findings, we co-transfected 293T cells with two mutant constructs. In the absence of the wild-type fusion, we again saw a reduction in homo-dimerization only when NHR2 was deleted (Fig. 5C).

In contrast to prior reports on the full-length GLIS2 protein, mutation of the PXDLS motifs retained in the fusion sequence (DL380,487AS) only led to a reduction in CtBP1 binding, not a complete abrogation³⁸. This may be due to the recruitment of CtBP1 to the complex through p300, independent of binding to the motif. However, deletion of NHR1 and 2 in combination with the PXDLS mutations only abrogated CtBP1 association when the fusion was directly immunoprecipitated, suggesting CtBP1 is still able to indirectly associate with the transcriptional complex even in the presence of all three structural mutations (Fig. 5D, Supplementary Fig. 6B). Although IP of p300 demonstrated an association with the fusion, we were unable to show direct binding of the fusion with p300 as when we immunoprecipitated for the fusion and probed for p300 we were unable to detect any bands. This suggests that p300 may only be recruited to the fusion complex via other co-factors, such as ETO and CtBP1 (Fig. 5E, Supplementary Fig. 6C).

Loss of NHR2 abrogates leukemogenesis in vivo

To assess the effects of the structural mutations and the requirement for these co-factors on leukemogenesis in vivo, we transplanted NSG-SGM3 mice with megakaryoblasts transduced with either wild-type *CBFA2T3-GLIS2* (*N*=13) or *CBFA2T3-GLIS2* carrying a deletion of NHR1 (*N*=5), NHR2 (*N*=5), mutation of the PXDLS motif (DL380,487AS) (*N*=5), or all mutations in combination (*N*=5) (Fig. 5F, G, Supplementary Data 1). Loss of NHR2 abrogated leukemogenesis in vivo when monitored for over 250 days (*p*<0.0001, Fig. 5F, Supplementary Data 1), whereas deletion of NHR1 or mutation of the PXDLS motif did not have an effect on leukemia-free survival (Fig. 5G, Supplementary Data 1). A previous study demonstrated an inability to engraft when *CBFA2T3-GLIS2* PDX cells were transduced with NC128, an NHR2 domain competitive peptide, thereby further highlighting the importance of this domain on transformative capability⁷. Our demonstration of the retained ability of the fusion to dimerize, although at significantly reduced levels, even when NHR2 is deleted (Fig. 5B, C), suggests the loss of leukemogenesis in NHR2 deletion mutants cannot be solely due to the loss of oligomerization, arguing in favor of a role for ETO in transformation (Fig. 5A, Supplementary Fig. 6A).

We sought to investigate genome-wide binding and gene expression changes upon deletion of NHR2 to identify key pathways required for leukemogenesis that may reveal therapeutic targets. Due to the lack of expansion in vivo when NHR2 is deleted, we performed CUT&RUN-seq and RNA-sequencing on megakaryoblasts transduced with either the wild-type fusion or a fusion construct lacking NHR2 grown in vitro. Differential gene expression analysis showed 64% (826/1296) of genes were downregulated, and 36% (470/1296) were upregulated when NHR2 was deleted compared to wild-type fusion-positive megakaryoblasts (Fig. 6A, Supplementary Data 3). GSEA found an enrichment of downregulated genes involved in the JAK/STAT, NOTCH, and Hedgehog signaling pathways (Fig. 6A, Supplementary Data 4). CUT&RUN-seq revealed that genome-wide, 61% of previously bound sites were lost, and 20% of sites had a decrease in the amount of fusion binding in NHR2 deletion mutants despite the retention of all five GLIS2 zinc finger DNA binding domains (Fig. 6B, Supplementary Data 8). Loss of binding was preferentially seen at transcriptional start sites, areas where co-factor binding and oligomerization are predicted to be important (Fig. 6C). When assessing gene expression of targets whose binding was lost at the transcriptional start sites, we found a concordant reversal in the signature imparted by the fusion (Fig. 6D). 90% (67/74) of the GRN nodes were no longer bound in NHR2 deletion mutants, so we reassessed the GRN and found a dramatic shift with only 26 nodes reaching significance (Fig. 6E, Supplementary Data 9). Collectively, the data argue in favor of impaired ETO binding and oligomerization of NHR2 deletion mutants, leading to a decrease in binding at

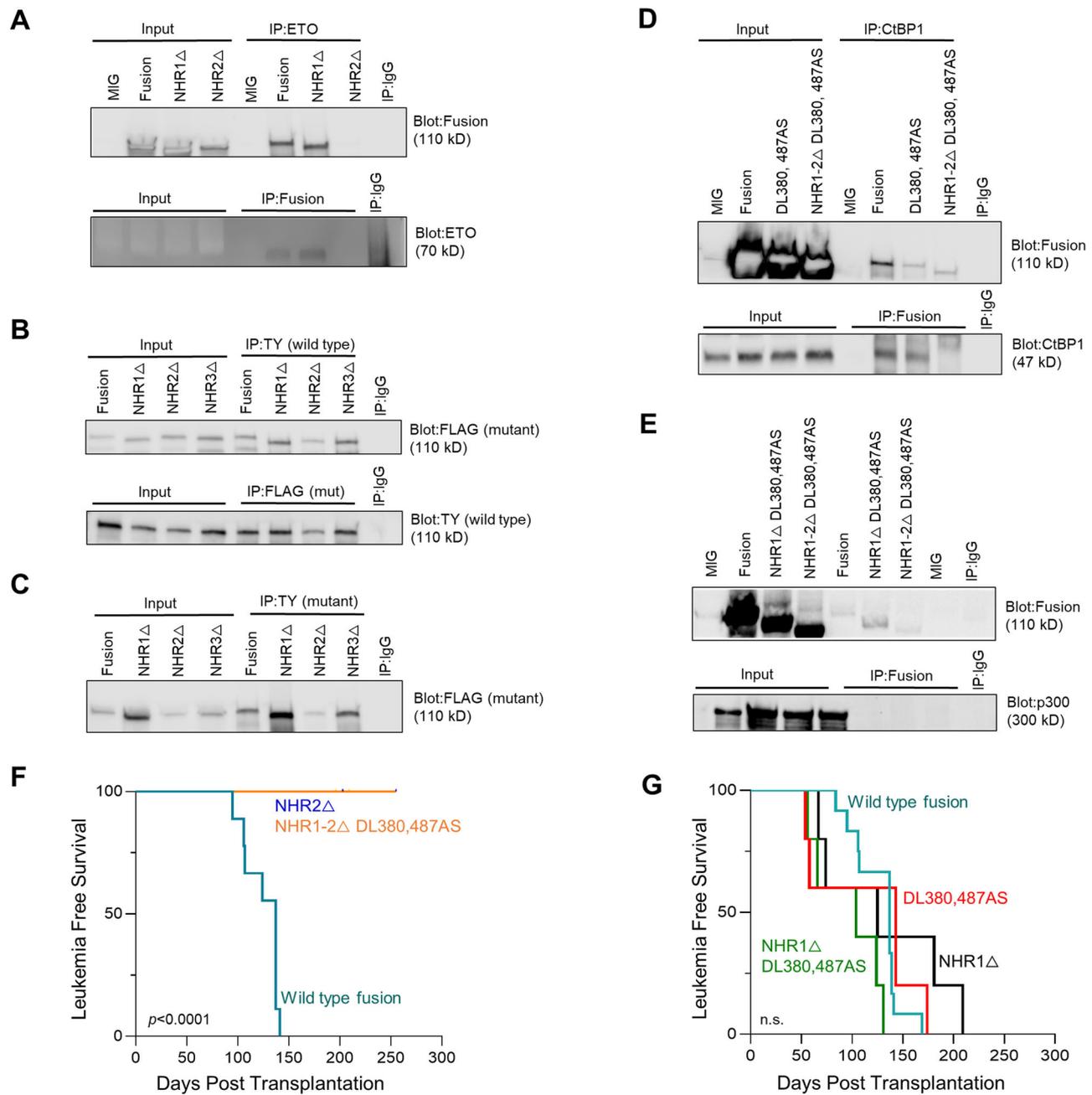


Fig. 5 | Loss of NHR2 disrupts ETO association and dimerization of the CBFA2T3-GLIS2 complex and abrogates leukemogenesis in vivo. **A–E** Co-immunoprecipitation using 293T cells transfected with empty vector (MIG), wild-type fusion, or a fusion containing one or more of the mutations outlined in Fig. 4A. All immunoprecipitations were repeated twice, representative blots are shown. Source data for all blots are provided as a Source Data file. **A** Immunoprecipitation of ETO followed by staining with the TY-1 tag to detect CBFA2T3-GLIS2 and the reciprocal immunoprecipitation of CBFA2T3-GLIS2 via the TY-1 tag followed by staining for ETO is shown. **B** TY-1-tagged wild-type CBFA2T3-GLIS2 and one of three FLAG-tagged mutant constructs were co-transfected into 293T cells to evaluate the ability of the mutant constructs to dimerize with the wild-type fusion. Immunoprecipitation by TY-1 followed by staining for FLAG and the reciprocal immunoprecipitation of FLAG followed by staining for TY-1 is shown. **C** 293T cells were co-transfected with two mutant constructs, one FLAG-tagged and one TY-1-tagged, to

verify the results shown in (B). Immunoprecipitation of TY-1-tagged NHR2 deletion mutant followed by staining for FLAG again revealed a reduction of dimerization. **D** TY-1-tagged wild-type and mutant fusion constructs were transfected into 293T cells. Cells were immunoprecipitated for CtBP1 followed by staining for the fusion via TY-1 and the reciprocal immunoprecipitation of the fusion followed by staining for CtBP1. **E** TY-1-tagged wild-type and mutant fusion constructs were transfected into 293T cells. Cells were immunoprecipitated for p300, followed by staining for the fusion via TY-1, and the reciprocal immunoprecipitation of the fusion followed by staining for p300. **F, G** Transplantation of immunodeficient NSG-SGM3 mice with primary megakaryoblasts transduced with the mutant fusion constructs ($N = 5$ per mutant construct, $N = 9$ wild type in panel **F** and 12 wild type in panel **G**). $p < 0.0001$ for NHR2 deletion and NHR1-2 (DL380,487AS) mutant constructs compared to wild type. p -values determined by log-rank test. Source data are provided as a Source Data file.

transcriptional start sites and the loss of the GRN and key leukemogenic pathways.

To assess the therapeutic potential of several of the key pathways identified in our studies and others, we subjected fusion-positive

megakaryoblasts and cord blood-derived CD34 cells to a series of inhibitors, including cKIT (Ripretinib), JAK/STAT (Ruxolitinib), WNT (iCRT), p300 (A-485), and ETO (7.44) (Supplementary Fig. 7)^{8,39–41}. The greatest sensitivity seen was to ruxolitinib and the p300 inhibitor

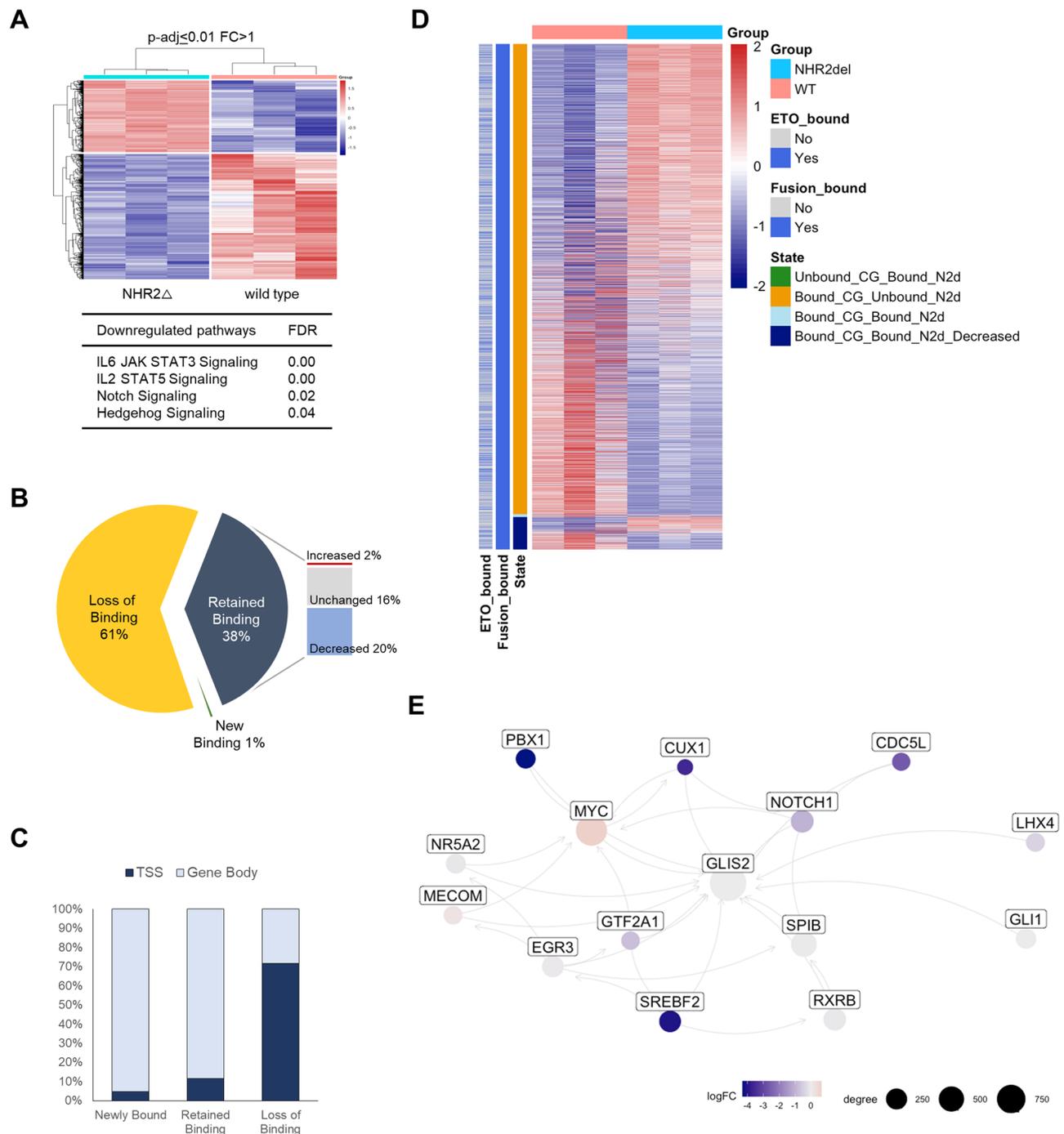


Fig. 6 | Impact of NHR2 deletion on CBFA2T3-GLIS2 transcription. **A** RNA extraction and gene expression profiling were performed in technical triplicate on megakaryoblasts cultured in vitro transduced with either the fusion construct with NHR2 deleted or the wild-type fusion construct. In vitro, culturing was necessary due to the lack of expansion in vivo upon NHR2 deletion. A heatmap of all differentially expressed genes is shown ($N = 1296$). Significantly downregulated pathways, as determined by gene set enrichment analysis, are shown. For a full list of gene sets, please see supplementary data 4. Log₂ fold change (FC) was calculated using a generalized linear model with a negative binomial distribution in DESeq2 and p -values derived from the Wald t -test and corrected using the

Benjamini–Hochberg correction. False discovery rate (FDR) was calculated using the Benjamini–Hochberg correction. **B** CUT&RUN-seq was performed on NHR2 deletion mutants and compared to wild-type fusion cells. The proportion of sites where binding was lost, retained, and newly bound sites is shown. Source data are provided as a Source Data file. **C** Location of binding at newly bound targets, retained targets, and lost targets in NHR2 deletion mutants. Source data are provided as a Source Data file. **D** Gene Expression changes in NHR2 deletion mutants at sites where fusion binding was lost at the transcriptional start sites ($N = 4059$ genes). **E** Gene regulatory network of NHR2 deletion mutants.

A-485 with IC₅₀ values of 70 and 1300 nM, respectively (Supplementary Fig. 7). Only ruxolitinib, however, demonstrated a therapeutic window when taking into account the toxicity of the compounds on CD34 stem cells. We confirmed the hematologic toxicity of p300 loss

by evaluating serial replating capacity in CBFA2T3-GLIS2-modified p300 knockout bone marrow (Supplementary Fig. 8). Confirming the sensitivity seen in vitro, p300 knockout abrogated colony development compared to p300 wild-type cells in the presence or absence of

CBFA2T3-GLIS2 (Supplementary Fig. 8). Consistent with the sensitivity of our model to ruxolitinib *in vitro*, we have previously published the activity of this targeted inhibitor in a CBFA2T3-GLIS2 patient-derived xenograft model¹¹. Combined, our data support a role for JAK/STAT signaling, and its potential as a therapeutic target.

Discussion

The majority of CBFA2T3-GLIS2-positive AMKL patients fail to be cured with the current treatment options that include chemotherapy in combination with stem cell transplant. Due to the relatively low incidence of this malignancy in the broader scope of pediatric cancer, these patients have not been afforded comprehensive research studies toward the development of new therapeutics⁴². We sought to contribute to the characterization of the CBFA2T3-GLIS2 transcriptional complex to enhance our understanding of the mechanism(s) of leukemogenesis and identify biologic pathways that are required in this process. Our data support a model in which the CBFA2T3-GLIS2 fusion facilitates homo-dimerization and recruits the transcriptional co-factors ETO, CtBP1, and p300 to impart differential gene expression at zinc finger bound G-C-rich motifs throughout the genome (Supplementary Fig. 9). Despite the large number of genes bound by the fusion, the impact of the fusion on gene expression extended far beyond fusion bound genes. This led us to map the gene regulatory network, which revealed a large number of well-known transcription factors that are dysregulated by the fusion and, in turn, affect downstream genes. While the fusion-bound genes included several associated with WNT, TGF β , and Hedgehog signaling, NOTCH and the JAK/STAT oncogenic pathways only emerged when we looked at gene expression globally and resulted from upregulation of components of the gene regulatory network. Further, our ability to compare fusion-negative and fusion-positive megakaryoblasts allowed us to confirm the changes that were due to the presence of CBFA2T3-GLIS2 and rule out megakaryoblast lineage-specific changes.

CBFA2T3 and GLIS2 are well-characterized transcription factors which led us to look at the potential role of ETO, CtBP1, and p300 co-factors on transcriptional regulation. We were able to confirm the association of these co-factors biochemically as well as co-localization at promoter sites bound by the fusion genome-wide. The generation of mutant constructs and their functional impact *in vivo* allowed us to narrow down the domain required for transformation to NHR2. When NHR2 is deleted, we found a significant reduction in homo-dimerization, a complete loss of ETO association, and impairment in DNA binding at transcriptional start sites. This had the greatest impact on JAK/STAT and NOTCH signaling pathways and the loss of the GRN at the gene expression level. Abrogation of the development of leukemia *in vivo* suggests that these pathways are important, in contrast to others that likely contribute to the transformation process, such as RUNX but were not affected by NHR2 deletion. Compilation of genomic occupancy mapping and pathway enrichment analyses in combination with NHR2 deletion gene dysregulation data has led us to propose that multiple pathways play an active role in transformation, which is disrupted when NHR2 is deleted. Thus, targeting a single pathway is not predicted to be effective therapeutically. BCL2, a downstream effector of JAK/STAT signaling and a therapeutic target of its own was found to be upregulated in fusion-positive leukemic megakaryoblasts as a result of the acquisition of a SE. The role of BCL-xL in megakaryocyte lineage survival, however, suggests a potential path of resistance to BCL2-specific inhibitors, which has recently been confirmed in CBFA2T3-GLIS2 model systems^{43,44}.

In summary, we have developed a humanized murine model of AMKL that faithfully recapitulates the CBFA2T3-GLIS2 patient profile. We have used this model to show ETO, CtBP1, and p300 are associated with the transcriptional complex and play a role in gene dysregulation that is disrupted when NHR2 is deleted. Our model has the potential to be used for systematic knock-down experiments to identify absolute

dependencies that can be therapeutically targeted and to aid in the pre-clinical testing of these agents.

Methods

Ethical regulations

The research complies with all relevant ethical regulations. Human CD34+ stem cells utilized in this study were commercially purchased, de-identified units. All animal experiments were conducted under IACUC-approved protocols at one of two institutions: St. Jude Children's Research Hospital, where "Animal Care and Use Committee" approved protocol entitled "Establishing Murine Models of Leukemia" 519-100568 (experiments conducted between 11/16/2018 and 8/25/2020); or Stanford University, where "Administrative Panel on Laboratory Animal Care" approved protocol entitled "Murine Modeling of High-Risk Pediatric Acute Leukemias" 33830 (experiments conducted between 08/25/2020 and 08/25/2023). Moribund mice were identified by the presentation of hind leg paralysis, hunched body position, labored breathing, and/or inability to eat or drink.

Generation of primary human megakaryoblasts

CD34+ stem cells were isolated from commercially purchased, de-identified human cord blood units (St. Louis Cord Blood Bank and Cleveland Cord Blood Bank). Cord blood was layered onto Ficoll-Paque (Fisher 17144003), and CD34+ stem cells were positively selected from the mononuclear layer using the human CD34 MicroBead kit (Miltenyi 130046702). CD34+ stem cells were cultured in GMP grade SCGM serum-free medium (CellGenix 208020500) supplemented with 1% penicillin-streptomycin and 100 ng/ μ l each of human TPO (Tonbo 218685), FLT3 ligand (Tonbo 218513), and SCF (PeproTech AF30007) for 48 h. Cultured CD34+ stem cells were then transduced using RetroNectin (Takara Bio 62955748) seeded with lentiviral particles encoding the *CBFA2T3-GLIS2* sequence. Packaged lentiviral particles were produced from transfected 293T cells (ATCC CRL-3216) using a third-generation lentiviral vector (see the section "Constructs" section), pMD2, and PAX2 packing plasmids with 2X BES (BES, NaCl, Na2PO4-Heptahydrate, pH 6.9) and CaCl₂ (pH 5.5). Forty-eight hours post-transduction, cells were removed from the RetroNectin and differentiated towards the megakaryoblast lineage in GMP SCGM serum-free medium supplemented with 1% penicillin-streptomycin and 100 ng/ml human TPO and 10 ng/ml human IL1 β (Tonbo 218018) for 7–9 days which allows for approximately 30-fold expansion⁴⁵. Megakaryoblasts were confirmed by flow cytometry analysis. Cells were pelleted, washed with PBS, blocked with human Fc block (BD 564220) for 10 min at room temperature, and then stained for CD41 (human CD41-PE Life Technologies MHCD4104 clone VIPL3), CD34 (human CD34-APC Miltenyi 130113176 clone AC136), and viability with Ghost Dye Violet 510 (Tonbo 130870) for 20 min in the dark at room temperature. Cells were washed twice with FACS buffer (PBS supplemented with 1% FBS and 0.1 mM EDTA) and then analyzed on a BD Fortessa (St. Jude) or a Cytex DxP10 (Stanford). Megakaryoblasts were then sorted for GFP using fluorescence-activated cell sorting on a BD Biosciences Aria to obtain a pure GFP-positive population for downstream applications such as transplantation, RNA-sequencing, and CUT&RUN-sequencing. Transfection efficiencies averaged 15–30%.

Constructs

Overlap extension PCR was used to clone the N-terminal 3xFLAG tag *CBFA2T3-GLIS2* coding sequence into the pCDH.MSCV.MCS.EF1a-copGFP (pCDH) lentivirus backbone and N-terminal 3xTY1 tag *CBFA2T3-GLIS2* coding sequence into the CSI,pEF1a.IRES.EmGFP (CEI) lentivirus backbone. NHR deletions and PXDLS motif mutations were made by site-directed mutagenesis using the *PfuUltra* high-fidelity DNA polymerase kit (Agilent 600380) as per the manufacturer's instructions for targets <10 kb vector DNA. 250 ng of input DNA was amplified via the following PCR program: Step 1 95 °C for 1 min, Step 2

95 °C for 40 s, Step 3 64 °C for 50 s, Step 4 68 °C for 15 min, repeat steps 2–4 16 times, Step 5 72 °C for 15 min. The following primers were used (5′–3′):

NHR1 Δ fw: CAGCACCTGCCCCAGCCTGCGGGCTGCTGGACGC-CAGCGCTCCTCCCC

NHR1 Δ rv: GGGGAGGAGGCGCTGCGTCCAGCAGCCCC-CAGCTGGGGCAGGTGCTG

NHR2 Δ fw: CGGCAGGAAGAAGTGATCGACCACAAGGACACAAA-GAAGGGCCCCGCTCCC

NHR2 Δ rv: GGGAGCGGGCCCTTCTTTGTGCTCTTGTGGTC-GATCACTTCTCTGCCC

NHR3 Δ fw: ACCCTCACCGGCTACGTGCTGAGG/CGAAGCGGC-AGGCTCCGAGGAC

NHR3 Δ rv: GTCCTCGGAGGCTGCCGCTTCGCCTCAGGCAGG-TAGCCGGTGAGGGT

DL380AS fw: CCTGCTGCCAGGCACCGTGTGGCCTCGTC-CACGGGCGTCAACTCAGCTG

DL380AS rv: CAGCTGAGTTGACGCCCGTGGACGAGGCCAG-CACGGTGCCTGGCAGCAGG

DL487AS fw: ACCCTGGCCCCGGCCCTTGCCTCCAGTGCCTT-GGCCTGTGGCAACG

DL487AS rv: CGTTGCCACAGGCCAGGGCACTGGAGGCAAGGGGG-CCGGGGCCAGGGGT

Plasmids generated in this study are available from the corresponding author.

Humanized murine model

Immunodeficient NSG-SGM3 (NOD.Cg-Prkdc^{scid} Il2rg^{tm1Wjl} ITg(CMV-IL3, CSF2, KITLG) IEav/MloySzj) breeder mice were purchased from the Jackson Laboratory (Strain 013062) and bred in-house maintained on the NOD.Cg-Prkdc^{scid} Il2rg^{tm1Wjl}/Szj background. Engraftment was found to be equal between male and female mice, so both sexes were utilized; see Supplementary Data 1 for the sex of each mouse reported in this study. Mice were treated with prophylaxis Baytril and Amoxicillin to minimize the contraction of bacterial infections. Both male and female mice four to six weeks of age were pre-conditioned with 200 rad (cesium irradiator at St. Jude, x-ray irradiator at Stanford) 24 h prior to megakaryoblast transplantation. Transplantations were done either via tail vein or intrafemoral injection. There was no difference in disease onset and phenotype with either method or based on the gender of recipient mice. All animal work was conducted under IACUC protocols: St. Jude Children's Research Hospital "Establishing Murine Models of Leukemia" 519-100568 (11/16/2018–11/16/2021); Stanford University "Murine Modeling of High-Risk Pediatric Acute Leukemias" 33830 (08/25/2020–08/25/2023). Moribund mice were identified by the presentation of hind leg paralysis, hunched body position, labored breathing, and/or inability to eat or drink. Spleen and bone marrow cells from both hind legs were collected. Red blood cells were lysed with RBC lysis buffer (Invitrogen 501129743) according to the manufacturer's instructions. Mouse CD45 cells were depleted using mouse CD45 microbeads (Miltenyi 130052301) according to the manufacturer's instructions. Mouse CD45-negative cells were stained for mouse CD45 (BioLegend 103132 clone 30-F11) and for viability with Ghost Dye Violet 510 (Tonbo 130870) as described above and analyzed for GFP purity by flow cytometry to confirm recovery of transplanted, fusion-positive megakaryoblasts. Purified cells were used for downstream applications, including RNA-seq, CUT&RUN-seq, and co-IP/Western. Full-body necropsies with histology (hematoxylin and eosin stain) were also performed to assess the engraftment of transplanted megakaryoblasts.

RT-PCR

The iScript cDNA synthesis kit (Bio-Rad 1708890) was used to generate cDNA from primary human megakaryoblasts according to the manufacturer's instructions. The Advantage 2 PCR Enzyme System

(Takara Bio USA 639207) was used to amplify the fusion junction DNA sequence using the following primer sequences: forward (5′CATCTG GAGGAAGGCTGA3′) and reverse (5′CGGTTGTGGATCTTCAGGT3′).

RNA-sequencing

RNA was extracted using TRIzol reagent (Thermo Fisher 155960266) according to the manufacturer's instructions, with all volumes divided in half due to low input (50,000–500,000 cells). Low input library prep (Ovation/NEB) and sequencing on the Illumina NovaSeq6000 (S4) PE150 sequencer were performed. Raw paired-end FASTQ files were filtered to remove low-quality reads and read with adaptor contamination. STAR v2.7.1a was used to align clean reads to the human genome (hg138) and to generate gene counts. STAR output for each sample was combined for differential expression testing using DESeq2_1.32.0R package installed in R v4.1 using default settings. PCA was performed on rlog-transformed data (for sample size <20) or vst-transformed (for sample size ≥ 20) using DESeq2. Gene set enrichment analysis was performed using GSEA v4.2.2¹⁷ with parameters recommended for sample size. For pathway enrichment analysis on subsets of genes, over-representation analysis was performed using bioMART_2.50.3, ReactomePA1_1.38.0, and clusterProfiler_4.2.2R packages. Pearson correlation was calculated using Intensity raw data that was screened for missing values for exclusion before log2 transformation in R. Data was then subset into Transplant and Patient groups, and then rowMean was taken by ID in each group. Cross-correlation plots were generated and correlation was measured using the Pearson method.

Mass spectrometry-based proteomics

100 μ g of protein from 10 million normal primary megakaryoblasts cultured in vitro ($N=2$), recovered transplanted megakaryoblasts ($N=2$), and CBFA2T3-GLIS2 patient samples ($N=4$) were analyzed in technical replicate by tandem mass tag (TMT) mass spectrometry. Proteomic analysis was performed with a protocol of multiplexed tandem mass tag labeling, two-dimensional liquid chromatography and tandem mass spectrometry (TMT-LC/LC-MS/MS)⁴⁶. TMT 10-plex isobaric label reagent set (Thermo Scientific) was used for proteome quantification of 10 samples simultaneously. 100 μ g of protein per sample was used as starting material. Cells were lysed using a buffer that contains 50 mM HEPES, pH 8.5, 8 M urea, and 0.5% sodium deoxycholate. The lysate was split into three aliquots: one to measure the protein concentration, one for a positive control validation of BMP2 by western blotting, and one for proteomic analysis. Protein concentration of sample lysates was quantified by BCA protein assay (Thermo Fisher Scientific) with titrated BSA as a standard. ~100 μ g proteins per sample were first digested with Lys-C (Wako, 1:100 w/w) at room temperature for 2 h, diluted 4 times with 50 mM HEPES, pH 8.5, and then further digested with trypsin (Promega, 1:100 w/w) overnight at room temperature. 1% trifluoroacetic acid was added to quench the digestion reaction, followed by desalting with Sep-Pak C18 cartridge (Waters), and the desalted peptides were dried by speedvac. Samples were then resuspended in 50 mM HEPES, pH 8.5, and were labeled with 10-plex TMT reagents following the manufacturer's instructions. Lastly, 10 isobaric labeled samples were pooled together with equal amount, desalted again by Sep-Pak C18 cartridge and then speedvac dried. TMT-labeled and unlabeled samples were verified by LC-MS/MS. The basic pH reverse phase liquid chromatography peptides pre-fractionation was performed on the Agilent 1220 LC system as previously described⁴⁷. The analysis was carried out based on the St. Jude proteomics core optimized platform as previously described^{47,48}. The dried peptide fractions were reconstituted in loading buffer (5% formic acid), loaded on a reverse phase column (75 μ m \times 50 cm, 1.9 μ m C18 resin (Dr. Maisch GmbH, Germany)), and interfaced with a FUSION or Q Exactive HF mass spectrometer (Thermo Fisher Scientific). Peptides were eluted up to 6 h by a 15–65% gradient of buffer B. (buffer A: 0.2% formic acid, 5% DMSO; buffer B: buffer A plus 65% acetonitrile, flow

rate: 0.25 $\mu\text{l}/\text{min}$). A butterfly portfolio heater (Phoenix S&T) was applied to heat the column at 65 °C to reduce back pressure. The mass spectrometer was operated in data-dependent mode with MS1 settings of 60,000 resolution, 1×10^6 AGC target and 50 ms maximal ion time and top 20 MS/MS high-resolution scans with MS2 settings of $1m/z$ isolation window with offset 0.2, 60,000 resolution, 100 ms maximal ion time, 1×10^5 AGC target, HCD, 33 normalized collision energy, and 40 s dynamic exclusion (35 normalized collision energy and 20 s dynamic exclusion for phosphoproteome). Peptide identification was performed using the JUMP search engine⁴⁹. Commercially available database search engines can be divided into two categories: tag-based De novo sequencing (e.g. PEAKS with limited sensitivity) and pattern-based database search (e.g. SEQUEST, MASCOT). The JUMP software integrates these two methods to score putative peptides, showing significant improvements compared with these commercially available tools¹¹. The JUMP software has previously been utilized in numerous publications⁵⁰. Analysis was done similarly as previously described, MS/MS raw files were first converted into mzXML format and searched against a composite target/decoy database FDR estimation⁴⁸. The target protein database was compiled from the UniProt mouse and human database (Human database: 88,965 protein entries; Mouse database: 52,738 protein entries, downloaded in February 2015), and the decoy database was generated by reversing target protein sequences. Spectra were searched with ± 10 ppm mass tolerance for both precursor ions and product ions with full tryptic restriction, static modification for TMT tag on N-terminus and lysine (+229.16293), dynamic modification for serine, threonine and tyrosine (+79.96633, for phosphoproteome analysis), three maximal modification sites, two maximal missed cleavages, and the assignments of *a*, *b*, and *y* ions. Peptide spectrum matches (PSM) were first filtered by MS mass accuracy (~ 2 ppm, ± 4 standard deviations). PSMs of doubly charged peptides with a JUMP Jscore of >30 were applied for global mass recalibration prior to the filtering. The survived PSMs were first grouped by precursor ion charge state and then further filtered by Jscore and dJn values. Cutoffs were applied to these values and were adjusted until a protein FDR $<1\%$ was achieved. If one peptide was shared by multiple proteins, the protein with the highest PSM will represent the peptide according to the rule of parsimony⁴⁶. For quantitation, an analysis was carried out in the following steps similarly as previously reported⁴⁸. (i) TMT reporter ion intensities of each PSM were extracted; (ii) the raw intensities were corrected according to the isotopic distribution of each labeling reagent; (iii) PSMs with very low reporter ion intensities were excluded (e.g. minimum intensity <1000 and median intensity <5000); (iv) sample loading bias was corrected by normalization with the trimmed median intensity of all PSMs; (v) the mean-centered intensities across samples were calculated; (vi) protein or phosphosite relative intensities were summarized by averaging related PSMs; (vii) protein or phosphosite absolute intensities were derived by multiplying the relative intensities by the grand-mean intensity of the top three most highly abundant PSMs⁴⁸. Protein lysates were digested and desalted and then underwent 16-plex TMT labeling. Samples were analyzed for total proteome by low pH reverse-phase liquid chromatography–tandem mass spectrometry (LC–MS/MS). Retrieved data was searched against the UniProt human database to identify peptides. This was determined by assigning related mass addition to all possible amino acid residues followed by an algorithm that evaluates the accuracy of modification localization. Peptide quantification data was corrected for mixing errors and summarized to derive protein quantification results. Using this approach, about 10,000 proteins are analyzed, which provides a comprehensive proteome analysis.

Spearman's correlation co-efficient

For the paired secondary ($N=1$) and tertiary ($N=1$) transplant cells that underwent RNA-sequencing, transcript per million (TPM) values for

protein-coding genes were log-transformed. Lowly expressed genes were filtered out as background noise (>1 TPM in 50% or more samples). For the transplant cells that underwent proteomics analysis, variance stabilizing normalization (VSN) was applied to the protein intensity data and replicates were combined by mean protein intensity value. 3525 overlapping genes between RNA-Sequencing and proteomics data were used in Spearman's correlation analysis (measured as Spearman's rho). Stringtie (v2.1.4) was used to convert BAM files to TPM values and R (v4.1.2) was used for the remaining analysis.

CUT&RUN-sequencing

500,000 megakaryoblasts recovered from transplantations were cross-linked with 0.1% formaldehyde for one minute at room temperature. The reaction was quenched with 2.5 M glycine. Cross-linked cells were then used with the CUTANA ChIC/CUT&RUN kit (EpiCypher 14-1048) per the manufacturer's instructions. Cells were washed twice with Wash Buffer (pre-wash buffer supplemented with 1% Triton-X and 0.05% SDS) and then bound to ConA beads. Cell-bound beads were collected and re-suspended in Antibody Buffer and 0.5 μg of one of the following antibodies: TY1 (GenScript A01004-40), ETO (Bethyl A303-509A), CtBP1 (ThermoFisher 10972-1-AP), p300 (ThermoFisher PA1848), H3K27ac (Active Motif 39034), and rabbit IgG (Epicyphe 13-0042). Beads were incubated at 4 °C overnight on a nutator. The next day, beads were collected and washed twice with a cold Cell Permeabilization Buffer followed by incubation with pAG-MNase. Beads were collected and washed twice again with Cell Permeabilization Buffer and then 100 mM of calcium chloride was added on ice to activate the enzymatic reaction. Beads were incubated at 4 °C for two hours on a nutator. After the incubation, Stop Buffer was added, and the bound protein-DNA complexes were eluted off the beads at 37 °C. The recovered protein-DNA was collected and de-crosslinked by adding SDS and Proteinase K, followed by an overnight incubation at 55 °C. The next day, DNA was purified using the provided DNA Cleanup Columns and DNA Binding and Wash buffer reagents. DNA concentration was measured using the Qubit dsDNA HS kit (Fisher Q32851). Library prep (NEBNext/ChIPSeq) and sequencing on Illumina NovaSeq600 (s4) PE150 was performed. CUT&RUN-sequencing was performed in biological replications and analyzed with paired RNA-sequencing data. Fastq files were trimmed and aligned to the hg19 genome via Basepair (<https://app.basepairtech.com>). Bam files were extracted and MACS v2.2.71 was used to call peaks in each sample relative to IgG control background (parameters -f BAMPE, -g 2700000000, --p-value 1e-3, --call-summits, -B and -SPMR). To build the single tracks for each sample, macs2 bdgcmp was used to generate fold-enrichment track (parameters -M FE and -p 0.00001), and resulting bedGraphs were fixed for conversion to bigwig files using bdg2bw.sh and bedGraphToBigWig method (bedtools v2.27.1). To calculate correlation between replicates before combination, multiBigWigSummary (bins and -chromosomesToSkip,"chrX", "chrY", "chrM", and "random") and plotCorrelation (-corMethod pearson -skipZeros, --whatToPlot scatterplot and -removeOutliers) from deepTools v3.3.0 were used. Annotation of narrowPeak files was performed using Homer v4.11 and hg19 as references. In the cofactor datasets, peaks were further filtered for p -value $\leq 1\text{E-}4$, q -value ≤ 0.05 and signal >10 th percentile. Genes associated with the promoter regions (in chromosomes 1–22) were intersected between replicates to generate a single gene list. For the CG TY data, further stringency was applied. The optimal signal threshold for CG TY data was determined, and the CG TY bound gene set was finalized based on p -value $\leq 1\text{E-}4$, q -value ≤ 0.05 , and signal >4 filters. The finalized CG TY gene set of 5256 genes (subset to promoter genes and chromosomes 1–22) has 4813 genes (92%) expressed in the bulk RNA-Seq data and 2657 gene IDs (55%) differentially expressed in the bulk RNA-Seq data (q -value ≤ 0.01 and $|\log_2\text{FC}| > 1$).

For the NHR2 deletion binding analysis, the CG TY and NHR2 deletion CUT&RUN data were normalized using DiffChIP by applying

background subtraction, low count filtering, and CPM normalization⁵¹. The resulting peaks and signal information were annotated, grouped by gene, and collapsed by genomic region. Next, binding events were categorized into genes with (1) gained, (2) lost, (3) retained, and (4) decreased binding. To reduce variability between replicates and ensure consistent signals, the interquartile range (IQR) for each group of normalized signal data was calculated, and peaks were filtered based on IQR values. Peaks were classified as “retained” if neither group (N2d or CG) had replicates with zero signal and the maximum difference between any pair of signal values was within the “retained threshold” ($0.5 \times$ standard deviation of the signal values from all four replicates). Peaks were classified as “increased” if the summed signals of N2d replicates exceeded those of CG by more than 1.5, and “decreased” if the opposite was true. Peaks were classified as “lost” if both N2d replicates had zero signal and both CG replicates had signals >1.5 .

Gene regulatory network analysis

For the analysis of direct and indirect interactions in the regulatory network of CBFA2T3-GLIS2, the R package GeneNetworkBuilder (GNB, v1.36.0) was used to create gene regulatory networks²⁰. First, the fusion-bound gene list and GNB’s embedded regulatory network for Homo sapiens (FANTOM Consortium by Riken Omics Science Center 2009) were combined to build the regulatory network. Second, the RNA-seq data showing altered expression were used to filter the network and add annotation to the network nodes, with GLIS2 as the central node. GLIS2 was chosen as the central node because of its ability to bind DNA and the lack of representation of the CBFA2T3-GLIS2 fusion in the FANTOM database. For GRN visualization (R package ggnetwork v0.5.10), degree and stress centralities were calculated using R packages igraph (v1.3.4), sna (v2.7). GRNs were further subset to focus on the top nodes based on degree centrality and to visualize pathways of interest (R v4.1.2)¹⁵. Next, these steps were applied to the NHR2 deletion CUT&RUN and RNA-Seq data to compare the CBFA2T3-GLIS2 fusion GRN to that of the NHR2 deletion. To compare the NHR2 deletion to the same negative sample set, the in vitro cultured NHR2 Δ and wild-type RNA-seq data were batch corrected with Fusion positive and negative megakaryocytes using COMBAT (R package sva v3.52.0) before performing differential expression analysis of NHR2 Δ vs. Fusion negative megakaryocytes (R package limma v3.60.0). The resulting log₂FC values were scaled to achieve a comparable GRN scale, without altering the relative differences or the directionality of the fold change.

Motif analysis

CUT&RUN peaks associated with the finalized CG TY gene set were further processed by merging nearby peaks (<50 bp), intersecting across replicates, and subsetting to a minimum length of 30 bp (bedtools v2.30.0), resulting in 10,619 peaks. Motif discovery was performed using Homer’s findMotifsGenome.pl (v4.11) with peak regions as input and hg19 as reference (using parameters -size given and -mask). Motif scanning was performed using FIMO (v5.5.5), where the peak regions were scanned for individual matches to motifs provided²⁵. The GLIS and GLI family motif sequences and SPI motif were compiled in a custom list and provided as motif input in FIMO²⁴.

Super-enhancer analysis

The Young Lab’s ROSE algorithm (v0.1) was used in the identification and ranking of super-enhancers^{52,53}. Using the main ROSE program (in Python v2.7.5), the density of the BAM sequencing reads in the hg19 GFF regions was calculated (sample density against IgG, default stitching distance, and excluding regions contained within $2000 \pm$ TSS to account for promoter biases). Using ROSE_callSuper.R (R v4.1.2), stitched-enhancer regions were ranked by their densities, and a cut-off was created to separate super-enhancers from typical enhancers. Data were filtered for densities >0 and chr1-22. For annotation,

ROSE_geneMapper.py (Python v3) was adapted to map all enhancer regions to genes. For each replicate, the annotated table of enhancers was merged with signal (density) data to list: enhancer, region, closest gene, and signal. In preparation for merging replicates, genes were collapsed by taking the maximum signal within each replicate. Taking the overlapping genes between replicates, genes were retained if the absolute value of difference in replicate signals was in the 80th percentile, in order to avoid large values skewing the data. Replicates were then merged by average signal to plot super-enhancers (R v4.1.2).

Co-immunoprecipitation and Western blotting

Recovered megakaryoblasts from transplanted mice or 293T cells (American Type Culture Collection, catalog number CRL-3216) transiently transfected with a lentivirus construct, as described above, for forty-eight hours were washed twice with PBS. Megakaryoblasts were lysed in NP40 lysis buffer (Thermo Fisher FNN0021) supplemented with a complete mini protease inhibitor tablet (Sigma 44693124001) and PMSF (Sigma 93482) according to the manufacturer’s instructions. 293T cells were lysed in RIPA buffer (Pierce 89901) supplemented with a complete mini protease inhibitor tablet according to the manufacturer’s instructions. Protein concentration was determined using the Pierce BCA Protein Assay Kit (Thermo Fisher 23225) according to the manufacturer’s instructions. 25 μ g of protein was saved for input and 500 μ g of protein was incubated overnight at 4 °C with 5 μ g of one of the following antibodies: FLAG (Sigma F7425), TY1 (GenScript A01004-40), ETO (Bethyl A303-509A), CtBP1 (ThermoFisher 10972-1-AP), p300 (ThermoFisher PA1948), mouse IgG (Santa Cruz Biotechnology sc2025). The next morning Protein A Dynabeads (Thermo 1002D) were added to the protein lysate and incubated for two hours at 4 °C. Beads were then collected on a magnet and washed twice with low salt buffer (0.1% SDS, 1% Triton-X, 2 mM EDTA, 20 mM Tris-HCl pH 8.0, 150 mM NaCl) followed by two washes with PBS supplemented with 0.02% Tween 20. Bound proteins were eluted off the beads with 2 \times Laemmli Sample Buffer (Bio-Rad 1610737) with β -mercaptoethanol. Samples were heated at 95 °C for 45 min. Eluted proteins were recovered and separated on a 4–20% Mini Protean TGX Stain-Free gel (Bio-Rad 4568094) and run at 100 V for ninety minutes. The gel was transferred using Trans-Blot Turbo Mini 0.2 μ m PVDF Transfer Packs (Bio-Rad 1704156) on the mixed molecular weight setting (high molecular weight setting for p300) using the Bio-Rad Turbo Transfer System. The PVDF membrane was blocked with 5% milk and incubated with 1:1000 of one of the following antibodies in 1% milk overnight at 4 °C on a rocking platform: FLAG (Sigma F3165), TY1 (ThermoFisher MA523513), ETO (Bethyl A303-509A), CtBP1 (ThermoFisher 10972-1-AP), p300 (ThermoFisher PA1948). The next morning membranes were washed with TBS-T (Fisher 28358), incubated with secondary antibody (goat anti-mouse-HRP ThermoFisher A28174 or goat-anti-rabbit-HRP ThermoFisher A27036) at 1:2000 dilution, washed with TBS-T, and developed using Clarity Western ECL Substrate (Bio-Rad 1705061). Membranes were stripped using Restore Western Blot Stripping Buffer (Thermo Fisher 21059) and for re-blotting for immunoprecipitation enrichment confirmation.

Statistics and reproducibility

For differential expression testing, we utilized DESeq2_1.32.0R package installed in R v4.1 using default settings. Log fold changes were calculated using a generalized linear model with a negative binomial distribution, and *p*-values were derived from the Wald *t*-test and adjusted using the Benjamini–Hochberg correction. Gene set enrichment analysis was performed using GSEA v4.2.2 with parameters recommended for sample size. False discovery rates (FDR) were calculated using the Benjamini–Hochberg correction. For pathway enrichment analysis on subsets of genes, over-representation analysis was performed using biomaRt_2.50.3, ReactomePA1_1.38.0, and clusterProfiler_4.2.2R packages. *p*-values were calculated using the

hypergeometric distribution, which corresponds to a one-sided Fisher's exact test, p -values were adjusted for multiple comparisons using the Benjamini–Hochberg correction, and q -values were calculated based on the FDR. Pearson correlation was calculated using Intensity raw data that were screened for missing values for exclusion before log₂ transformation in R. Murine Kaplan–Meier survival curves were plotted in Graph Pad Prism v9 and analyzed by Mantel–Cox log-rank test. No statistical method was used to predetermine sample size, and no data were excluded from the analyses. The experiments were not randomized, and the investigators were not blinded to allocation during experiments and outcome assessment.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

Patient sample gene expression data originally published in Gruber et al., 2012 is available in the NCBI Gene Expression Omnibus (<http://www.ncbi.nlm.nih.gov/geo/>) under [GSE35203](https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE35203)⁵. Additional patient sample gene expression data originally published by de Rooij et al., 2017 is available in the European Genome-phenome Archive (EGA) under [EGAS00001002183](https://ega-archive.org/studies/EGAS00001002183)⁴. CUT&RUN-seq and gene expression data presented in this paper are available in the GEO repository under accession number [GSE219269](https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE219269). The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE partner repository with the dataset identifier [PXDO48474](https://www.ebi.ac.uk/pride/archive/study/PXD048474)^{54,55}. All other relevant data are included in the supplementary materials. Source data are provided with this paper. Correspondence and material requests should be addressed to Tanja A. Gruber: tagrub@stanford.edu. Source data are provided with this paper.

Code availability

Code is available on Zenodo and can be accessed at <https://doi.org/10.5281/zenodo.13826492> and <https://zenodo.org/records/13826492>.

References

- Gruber, T. A. & Downing, J. R. The biology of pediatric acute megakaryoblastic leukemia. *Blood* **126**, 943–949 (2015).
- Taub, J. W. et al. Improved outcomes for myeloid leukemia of Down syndrome: a report from the Children's Oncology Group AAML0431 trial. *Blood* **129**, 3304–3313 (2017).
- Khan, I., Malinge, S. & Crispino, J. Myeloid leukemia in Down syndrome. *Crit. Rev. Oncog.* **16**, 25–36 (2011).
- de Rooij, J. D. et al. Pediatric non-Down syndrome acute megakaryoblastic leukemia is characterized by distinct genomic subsets with varying outcomes. *Nat. Genet.* **49**, 451–456 (2017).
- Gruber TA et al. An Inv(16)(p13.3;q24.3)-encoded CBFA2T3-GLIS2 fusion protein defines an aggressive subtype of pediatric acute megakaryoblastic leukemia. *Cancer cell* **22**, 683–697 (2012).
- Gruber, T. A. et al. Transcriptome sequence analysis of pediatric acute megakaryoblastic leukemia identifies an Inv(16)(p13.3;q24.3)-encoded CBFA2T3-GLIS2 fusion protein as a recurrent lesion in 39% of non-infant cases: a report from the St. Jude Children's Research Hospital—Washington University Pediatric Cancer Genome Project. *Blood* **118**, 757–757 (2011).
- 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).
- Benbarche, S. et al. Screening of ETO2-GLIS2-induced Super Enhancers identifies targetable cooperative dependencies in acute megakaryoblastic leukemia. *Sci. Adv.* **8**, eabg9455 (2022).
- Skene P. J. & Henikoff S. An efficient targeted nuclease strategy for high-resolution mapping of DNA binding sites. *Elife* **6**, e21856 (2017).
- Dang, J. et al. AMKL chimeric transcription factors are potent inducers of leukemia. *Leukemia* **31**, 2228–2234 (2017).
- Drenberg, C. D. et al. A high-throughput screen indicates gemcitabine and JAK inhibitors may be useful for treating pediatric AML. *Nat. Commun.* **10**, 2189 (2019).
- Wunderlich, M. et al. AML xenograft efficiency is significantly improved in NOD/SCID-IL2RG mice constitutively expressing human SCF, GM-CSF and IL-3. *Leukemia* **24**, 1785–1788 (2010).
- Thompson, A. et al. Tandem mass tags: a novel quantification strategy for comparative analysis of complex protein mixtures by MS/MS. *Anal. Chem.* **75**, 1895–1904 (2003).
- Yang, K. C. & Gorski, S. M. Protocol for analysis of RNA-sequencing and proteome profiling data for subgroup identification and comparison. *STAR Protoc.* **3**, 101283 (2022).
- Wieder, C. et al. Pathway analysis in metabolomics: recommendations for the use of over-representation analysis. *PLoS Comput. Biol.* **17**, e1009105 (2021).
- Gillespie, M. et al. The reactome pathway knowledgebase 2022. *Nucleic acids Res.* **50**, D687–D692 (2022).
- Subramanian, A. et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl Acad. Sci. USA* **102**, 15545–15550 (2005).
- Liberzon, A. et al. Molecular signatures database (MSigDB) 3.0. *Bioinformatics* **27**, 1739–1740 (2011).
- Liberzon, A. et al. The Molecular Signatures Database (MSigDB) hallmark gene set collection. *Cell Syst.* **1**, 417–425 (2015).
- Ou, J., Liu, H., Tissenbaum, H. A. & Zhu, L. J. GeneNetworkBuilder: A Bioconductor Package for Building Regulatory Network Using ChIP-chip/ChIP-seq Data and Gene Expression Data. Bioconductor R package version 1.40.0 (2022).
- Zhang, W. et al. Runt-related transcription factor 3 promotes acute myeloid leukemia progression. *Front. Oncol.* **11**, 725336 (2021).
- Choo, A. et al. The role of IRF1 and IRF2 transcription factors in leukemogenesis. *Curr. Gene Ther.* **6**, 543–550 (2006).
- Stellacci, E. et al. Interferon regulatory factor-2 drives megakaryocytic differentiation. *Biochem. J.* **377**, 367–378 (2004).
- Weirauch, M. T. et al. Determination and inference of eukaryotic transcription factor sequence specificity. *Cell* **158**, 1431–1443 (2014).
- Grant, C. E., Bailey, T. L. & Noble, W. S. FIMO: scanning for occurrences of a given motif. *Bioinformatics* **27**, 1017–1018 (2011).
- Vasanth, S., ZeRuth, G., Kang, H. S. & Jetten, A. M. Identification of nuclear localization, DNA binding, and transactivating mechanisms of Kruppel-like zinc finger protein Gli-similar 2 (Glis2). *J. Biol. Chem.* **286**, 4749–4759 (2011).
- Daud, S. S. et al. Identification of the Wnt signalling protein, TCF7L2 as a significantly overexpressed transcription factor in AML. *Blood* **120**, 1281–1281 (2012).
- Levay, K. & Slepak, V. Z. Tescalcin is an essential factor in megakaryocytic differentiation associated with Ets family gene expression. *J. Clin. Investig.* **117**, 2672–2683 (2007).
- Ortega, E. et al. Transcription factor dimerization activates the p300 acetyltransferase. *Nature* **562**, 538–544 (2018).
- Hug, B. A. & Lazar, M. A. ETO interacting proteins. *Oncogene* **23**, 4270–4274 (2004).
- Wang, L. et al. The leukemogenicity of AML1-ETO is dependent on site-specific lysine acetylation. *Science* **333**, 765–769 (2011).
- Quinlan, K. G. et al. Role of the C-terminal binding protein PxDLS motif binding cleft in protein interactions and transcriptional repression. *Mol. Cell. Biol.* **26**, 8202–8213 (2006).
- Kim, S. C., Kim, Y. S., & Jetten, A. M. Kruppel-like zinc finger protein Gli-similar 2 (Glis2) represses transcription through interaction with C-terminal binding protein 1 (CtBP1). *Nucleic Acids Res.* **33**, 6805–6815 (2005).

34. Kim, J. H., Cho, E. J., Kim, S. T. & Youn, H. D. CtBP represses p300-mediated transcriptional activation by direct association with its bromodomain. *Nat. Struct. Mol. Biol.* **12**, 423–428 (2005).
35. Steinauer, N., Guo, C. & Zhang, J. Emerging roles of MTG16 in cell-fate control of hematopoietic stem cells and cancer. *Stem Cells Int.* **2017**, 6301385 (2017).
36. Chen, G. et al. The RUNX1-ETO fusion protein trans-activates c-KIT expression by recruiting histone acetyltransferase P300 on its promoter. *FEBS J.* **286**, 901–912 (2019).
37. Hildebrand, D., Tiefenbach, J., Heinzl, T., Grez, M. & Maurer, A. B. Multiple regions of ETO cooperate in transcriptional repression. *J. Biol. Chem.* **276**, 9889–9895 (2001).
38. Liu, Y. et al. The tetramer structure of the Nrvy homology two domain, NHR2, is critical for AML1/ETO's activity. *Cancer Cell* **9**, 249–260 (2006).
39. Zhang, Y. & Wang, X. Targeting the Wnt/beta-catenin signaling pathway in cancer. *J. Hematol. Oncol.* **13**, 165 (2020).
40. Gopalswamy, M. et al. Biophysical and pharmacokinetic characterization of a small-molecule inhibitor of RUNX1/ETO tetramerization with anti-leukemic effects. *Sci. Rep.* **12**, 14158 (2022).
41. Lasko, L. M. et al. Discovery of a selective catalytic p300/CBP inhibitor that targets lineage-specific tumours. *Nature* **550**, 128–132 (2017).
42. Fornerod, M. et al. Integrative genomic analysis of pediatric myeloid-related acute leukemias identifies novel subtypes and prognostic indicators. *Blood Cancer Discov.* **2**, 586–599 (2021).
43. Neault, M. et al. CBFA2T3-GLIS2-dependent pediatric acute megakaryoblastic leukemia is driven by GLIS2 and sensitive to navitoclax. *Cell Rep.* **42**, 113084 (2023).
44. 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).
45. Tijssen, M. R. et al. Genome-wide analysis of simultaneous GATA1/2, RUNX1, FLI1, and SCL binding in megakaryocytes identifies hematopoietic regulators. *Dev. Cell* **20**, 597–609 (2011).
46. Nesvizhskii, A. I. & Aebersold, R. Interpretation of shotgun proteomic data: the protein inference problem. *Mol. Cell. Proteom.* **4**, 1419–1440 (2005).
47. Bai, B. et al. Deep profiling of proteome and phosphoproteome by isobaric labeling, extensive liquid chromatography, and mass spectrometry. *Methods Enzymol.* **585**, 377–395 (2017).
48. Tan, H. et al. Integrative proteomics and phosphoproteomics profiling reveals dynamic signaling networks and bioenergetics pathways underlying T cell activation. *Immunity* **46**, 488–503 (2017).
49. Wang, X. et al. JUMP: a tag-based database search tool for peptide identification with high sensitivity and accuracy. *Mol. Cell. Proteom.* **13**, 3663–3673 (2014).
50. Gong, J. et al. The *C. elegans* taste receptor homolog LITE-1 is a photoreceptor. *Cell* **168**, 325 (2017).
51. Chen, Y., Chen, S. & Lei, E. P. DiffChIP: a differential peak analysis method for high-throughput sequencing data with biological replicates based on limma. *Bioinformatics* **38**, 4062–4069 (2022).
52. Loven, J. et al. Selective inhibition of tumor oncogenes by disruption of super-enhancers. *Cell* **153**, 320–334 (2013).
53. Whyte, W. A. et al. Master transcription factors and mediator establish super-enhancers at key cell identity genes. *Cell* **153**, 307–319 (2013).
54. Perez-Riverol, Y. et al. The PRIDE database resources in 2022: a hub for mass spectrometry-based proteomics evidences. *Nucleic Acids Res.* **50**, D543–D552 (2022).
55. Deutsch, E. W. et al. The ProteomeXchange consortium at 10 years: 2023 update. *Nucleic Acids Res.* **51**, D1539–D1548 (2023).
56. Kurtenbach, S. & William Harbour, J. Spark: a publication-quality NGS visualization tool. Preprint at *bioRxiv* <https://doi.org/10.1101/845529> (2019).

Acknowledgements

The authors thank all members of St. Jude Children's Research Hospital and the Stanford School of Medicine for their support. We specifically thank the Research Animal Facility (Stanford), the Animal Resource Center (St. Jude), the Shared FACS Facility (Stanford), the Flow Cytometry and Cell Sorting Shared Resource (St. Jude), the Center for Proteomics and Metabolomics (St. Jude), the Hartwell Center for Biotechnology (St. Jude), and the Vanderbilt Technologies for Advanced Genomics (Vanderbilt University Medical Center). Venn diagrams were created with the tool from Bioinformatics and Evolutionary Genomics (<https://bioinformatics.psb.ugent.be/webtools/Venn/>). This work was supported by the American Lebanese Syrian Associated Charities.

Author contributions

Conceptualization: T.A.G.; Methodology: E.A.R.G., P.N., T.A.G.; Investigation: E.A.R.G., P.N., B.S., J.D., M.B., A.C., C.F., S.M., T.W.; Supervision: T.A.G.; Writing—original draft: E.A.R.G., T.A.G.; Writing—review & editing: All authors, funding acquisition: T.A.G.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41467-024-53158-9>.

Correspondence and requests for materials should be addressed to Tanja A. Gruber.

Peer review information *Nature Communications* thanks the anonymous reviewers for their contribution to the peer review of this work. A peer review file is available.

Reprints and permissions information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, 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 licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence 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 licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2024