blood advances®

CBFB-MYH11 fusion transcripts distinguish acute myeloid leukemias with distinct molecular landscapes and outcomes

Benjamin J. Huang,[1,2] Jenny L. Smith,[3] Yi-Cheng Wang,[4] Kassra Taghizadeh,[3] Amanda R. Leonti,[3] Rhonda E. Ries,[3] Yanling Liu,[5] Pandurang Kolekar,[5] Katherine Tarlock,[3,6] Robert Gerbing,[4] Erin Crowgey,[7] Scott N. Furlan,[3,6] Timothy I. Shaw,[5] Kohei Hagiwara,[5] Lisa Wei,[8] Todd M. Cooper,[6] Alan S. Gamis,[9] Richard Aplenc,[10] E. Anders Kolb,[7] Jason E. Farrar,[11] Timothy Triche Jr,[12] Todd A. Alonzo,[13] Xiaotu Ma,[5,*] and Soheil Meshinchi[3,6,*]

[1]Department of Pediatrics, University of California San Francisco, San Francisco, CA; [2]Helen Diller Family Comprehensive Cancer Center, University of California San Francisco, San Francisco, CA; [3]Fred Hutchinson Cancer Research Center, Seattle, WA; [4]Children's Oncology Group, Monrovia, CA; [5]Department of Computational Biology, St. Jude Children's Research Hospital, Memphis, TN; [6]Division of Hematology/Oncology, Seattle Children's Hospital, Seattle, WA; [7]Nemours Center for Cancer and Blood Disorders and Alfred I. DuPont Hospital for Children, Wilmington, DE; [8]Michael Smith Genome Sciences Centre, Vancouver, BC, Canada; [9]Children's Mercy Hospitals and Clinics, Kansas City, MO; [10]Children's Hospital of Philadelphia, Philadelphia, PA; [11]University of Arkansas for Medical Sciences & Arkansas Children's Research Institute, Little Rock, AR; [12]Van Andel Research Institute, Grand Rapids, MI; and [13]Department of Preventive Medicine, University of Southern California, Los Angeles, CA

Key Points

- *CBFB-MYH11* transcripts and *KIT* mutations predict relapse in AML.

- High-risk *CBFB-MYH11* transcripts are associated with distinct transcriptional landscapes and upregulation of early hematopoiesis genes.

Introduction

Although the impacts of *CBFB-MYH11* transcript variants on acute myeloid leukemia (AML) survival outcomes have been reported in adult patients with AML,[1] this finding has not yet been tested in cohorts that include children and adolescents. Furthermore, the underlying reasons for these outcome differences have not been well studied. Therefore, we harnessed transcriptome sequencing performed on diagnostic AML samples of 1776 pediatric patients enrolled on front-line Children's Oncology Group (COG) clinical trials. This comprehensive dataset includes 186 patients with *CBFB-MYH11* fusions (supplemental Table 1; Figure 1A). Our analysis of these transcriptomes demonstrates that fusion transcript subtype predicts relapse in pediatric *CBFB-MYH11* AML and deepens our understanding of how *CBFB-MYH11* fusion transcripts may impact underlying leukemia biology and outcomes.

Methods

Patient samples and RNA-sequencing

AML samples were collected with informed consent from patients (0-28 years of age) enrolled on COG trials CCG-2961 (#NCT00002798),[2] AAML03P1 (#NCT00070174),[3] AAML0531 (#NCT00372593),[4] and AAML1031 (#NCT01371981).[5] Institutional review board approval for each protocol was obtained at each participating institution and submitted to the Cancer Trials Support Unit (CTSU) regulatory office. Each protocol was conducted in accordance with the Declaration of Helsinki. Total RNA derived from peripheral blood or bone marrow diagnostic specimens was purified using AllPrep DNA/RNA/miRNA Universal Kits. Purified RNA samples were then prepared for either strand-specific polyadenylated enriched messenger RNA libraries or strand-specific ribosome RNA-depleted libraries by the British Columbia Genome Sciences Center. Paired-end sequencing was performed on Illumina HiSeq 2000/2500 platforms, and sequence reads were aligned to the GRCh37 reference genome using BWA (v0.5.7).[6] Reads were discarded based on mapping quality or if they failed the Illumina chastity filter, and duplicate reads were marked using Picard (v1.11). Gene level coverage analysis was performed using the British Columbia Genome Sciences Center pipeline v1.1 with Ensembl v69 annotations and was normalized based on RPKM or TPM.

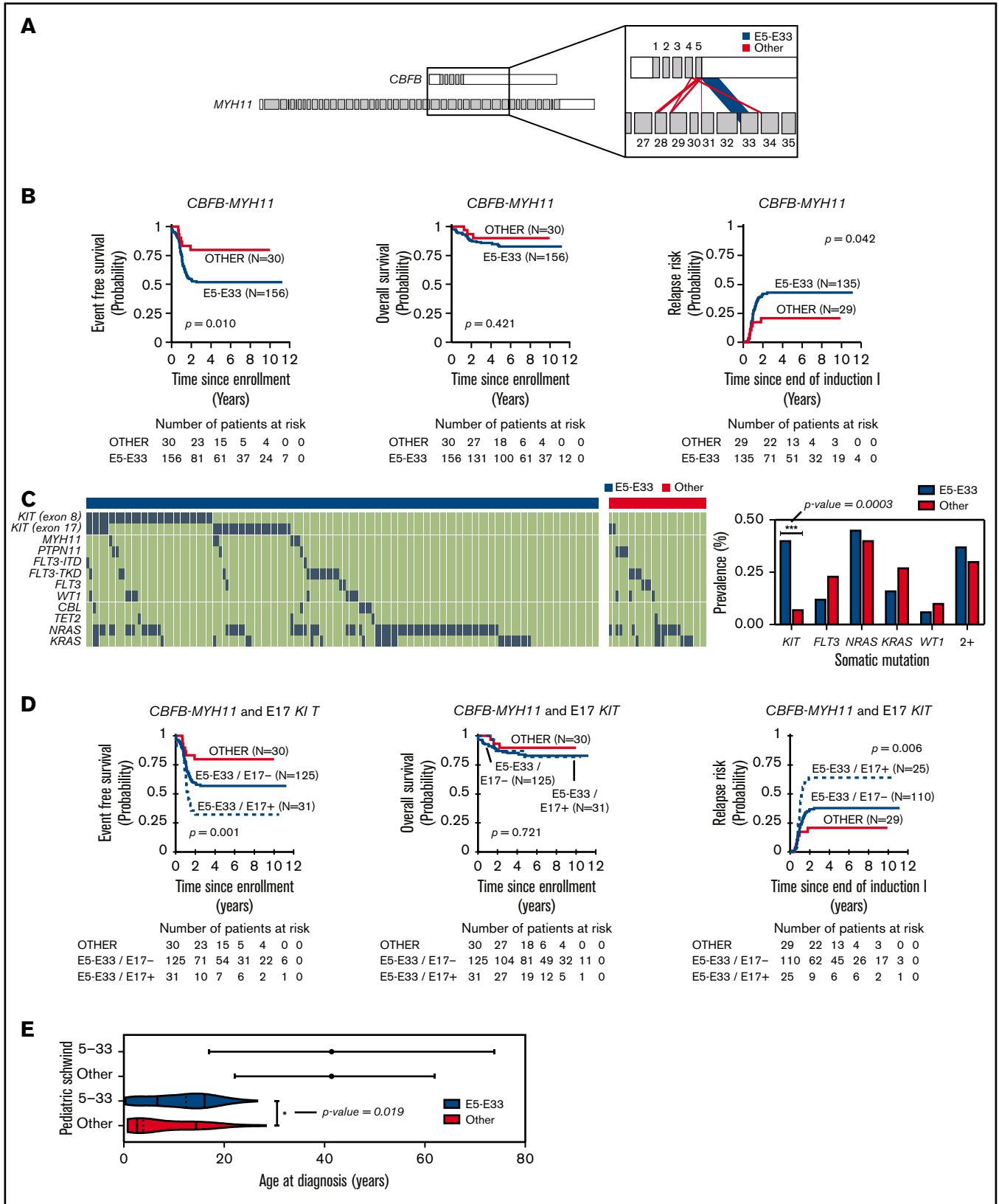The full-text version of this article contains a data supplement.

Figure 1. CBFB-MYH11 fusion transcript subtype predicts relapse in AML. (A) *CBFB-MYH11* fusion transcripts within our cohort with line weights that correspond to the associated prevalence. (B) Kaplan-Meier estimates for the probability of EFS, OS, and RR in patients with *CBFB-MYH11* AML stratified based on fusion transcript (E5-E33 vs all others). (C) Oncoplot with associated somatic driver mutations based on fusion transcript status. *KIT* mutations are more prevalent in E5-E33 *CBFB-MYH11* AMLs compared with others (Fisher's exact test, *P* value = .0003). (D) Kaplan-Meier estimates for the probability of EFS, OS, and RR in patients with *CBFB-MYH11* AML stratified based on fusion transcript and E17 *KIT* mutation status. (E) Age distribution comparing Schwind et al (adults) and patients from our cohort (pediatrics).
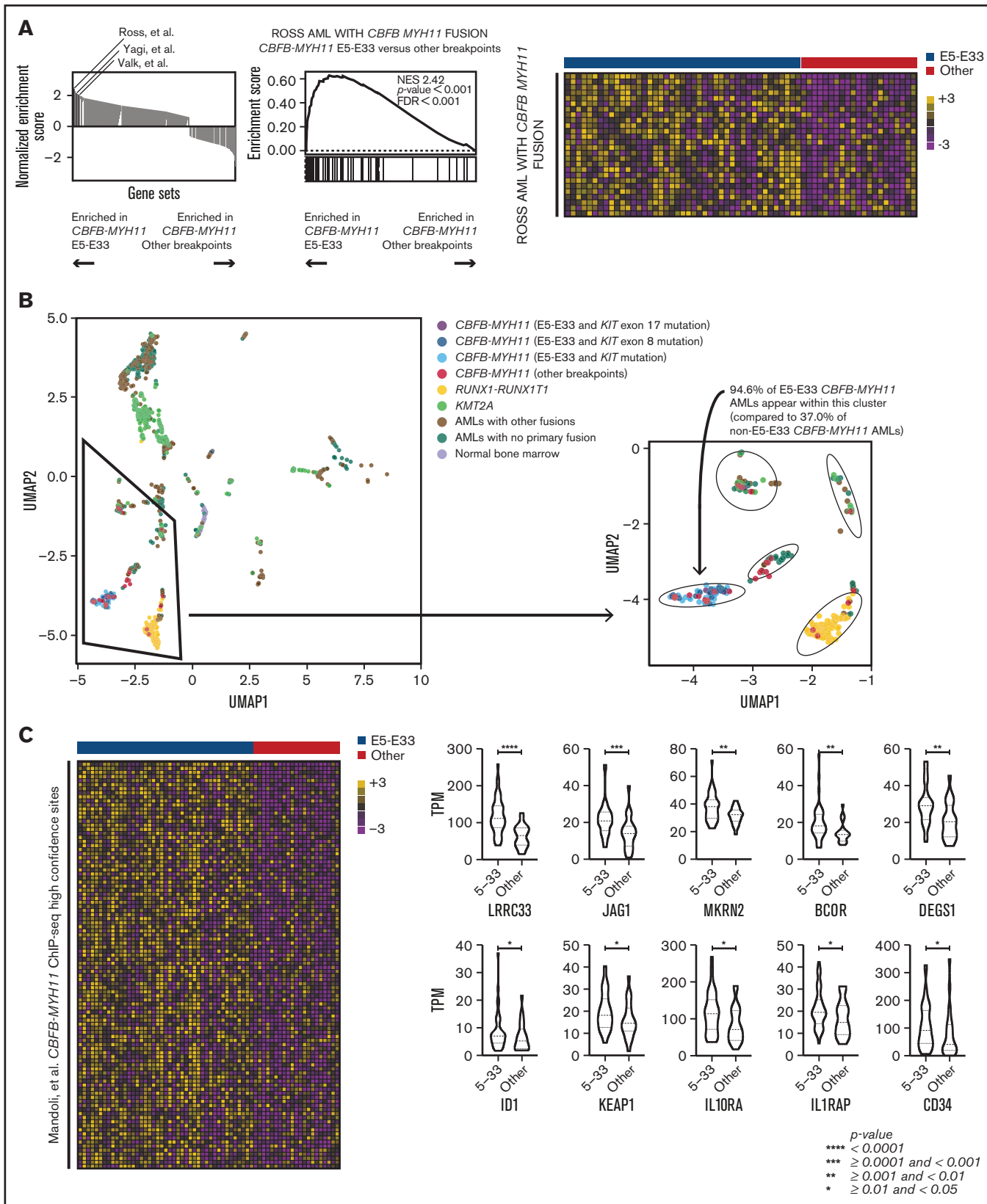
Figure 2. ***CBFB-MYH11* fusion transcript subtypes are associated with distinct transcriptional landscapes.** (A) GSEA comparing E5-E33 *CBFB-MYH11* AMLs with all others, using gene sets available through the Molecular Signature Database (hallmark, curated, and oncogenic signature gene sets). Three gene sets that define *CBFB-MYH11* transcriptional signatures are significantly enriched in E5-E33 *CBFB-MYH11* AMLs. Specifically, the Ross et al gene set is associated with a normalized

Fusion calling

Fusion calls were made using CICERO, STAR-Fusion, and Trans-ABySS as previously described.[7-9] In order to maximize detection sensitivity, *CBFB-MYH11* transcript breakpoints called by one or more fusion callers were accepted after manual examination using Bambino.[10] Overlapping calls between CICERO, STAR-Fusion, and Trans-ABySS for *CBFB-MYH11* fusion transcripts were concordant at the level of exon transcript calling.

Variant calling

Variant calling was performed based on RNA sequencing and targeted DNA sequencing studies. RNA-seq libraries were constructed and sequenced as described above. All computational analyses were performed on a dedicated compute and storage infrastructure designed and implemented at St. Jude as previously described.[11] Briefly, RNA reads were mapped using the StrongARM pipeline.[12] Paired-end reads were aligned to 5 reference human genome databases (including GRCh37) using BWA,[6] and the final BAM file was constructed by selecting the best of the 5 alignments. SNVs and Indels were called using Bambino[10] and RNAIndel,[13] followed by validation and pathogenicity classification as previously described.[14,15]

Statistical analysis

Event-free survival (EFS) and overall survival (OS) analysis was performed using Kaplan-Meier estimates. EFS was defined as the time from enrollment to first event (relapse, induction failure, or death) or last follow-up. OS was defined as the time from study enrollment to death or last follow-up. Relapse risk (RR) was defined as the time since end of Induction I to relapse or last follow-up. Cox proportional hazard regression models were employed to estimate hazard ratios for univariable and multivariable analyses of EFS, OS, and RR. Dierences in EFS, OS, and RR between groups were determined using log-rank testing. The χ^2 test was used to test the significance of observed differences in proportions, and Fisher's exact test was used when data were sparse. Differences in medians were compared by the Mann-Whitney U test. A P value $<.05$ was considered statistically significant.

Gene expression analysis

Uniform manifold approximation and projection (UMAP) was performed using the uwot (v0.1.10) R package. Gene set enrichment analysis (GSEA) was performed using command line tools.[16] In order to ensure a more uniform RNA-seq cohort, only ribosomal RNA-depleted sequenced libraries were included in the gene expression analysis.

Results and discussion

CBFB-MYH11 fusions in the 186 pediatric AML cases analyzed are predominantly composed of transcribed products from the fusion of exon 5 of *CBFB* with exon 33 of *MYH11* (Figure 1A), hereafter referred to as E5-E33 *CBFB-MYH11*. To assess the prognostic significance of the E5-E33 transcript, we evaluated the probability of EFS, OS, and RR in our cohort stratified based on fusion transcript location (E5-E33 vs all others) (Figure 1B). Patients with E5-E33 had an EFS of 52% ± 8.1% (n = 156) at 5 years from diagnosis compared with an EFS of 80% ± 14.7% (n = 30) for those with other fusion transcripts (P value = .01).

We next harnessed variant calls on the same cohort of patients (n = 186) to elucidate the mutational landscape of *CBFB-MYH11* AMLs. *KIT* mutations were significantly enriched in E5-E33 *CBFB-MYH11* AMLs with a prevalence of 42% in the E5-E33 cohort vs 7% in the Other cohort (P value = .0003; Figure 1C; supplemental Table 1). In contrast, the prevalence of *FLT3*, *NRAS*, *KRAS*, and *WT1* mutations was not significantly different between the 2 groups (Figure 1C). We have previously demonstrated that exon 17 (E17) *KIT* mutations are associated with adverse outcomes in patients with core-binding factor AML.[1,17] Therefore, we assessed for their impact on survival within our E5-E33 cohort and found that they were strongly associated with a significantly worse EFS of 32% ± 16.8% for E17 KIT^+ at 5 years from diagnosis compared with an EFS of 56% ± 10.5% for E17 KIT^- patients (P value = .042) (Figure 1D) and was notably independent of E17 *KIT* variant allele frequency (supplemental Figure 1). Conversely, exon 8 (E8) *KIT* mutations were associated with no difference in survival (supplemental Figure 2).

The prognostic impact of E5-E33 transcripts on EFS was maintained in multivariable survival analysis that included age, white blood cell count, and E17 *KIT* mutations (supplemental Tables 2 and 3) and in survival analysis that excluded *KIT* mutant AMLs (supplemental Figure 3). Furthermore, the inferior outcomes associated with E5-E33 *CBFB-MYH11* AMLs that are significantly worse in the presence of an additional E17 *KIT* mutation are particularly relevant because they address a gap in knowledge based on previously studied demographic age groups (Figure 1E) and further support the hypothesis that *CBFB-MYH11* transcript subtypes impact leukemia biology and/or chemoresistance.

Interestingly, and in contrast to Schwind et al,[1] despite adverse EFS in E5-E33 patients, OS did not differ based on fusion transcript (Figure 1B,D) suggesting that relapses in children, adolescents, and young adults with E5-E33 *CBFB-MYH11* AMLs ± E17 *KIT* mutations are highly salvageable. Because hematopoietic stem cell transplantation (HSCT) in first complete remission is deferred to after an initial relapse in favorable risk pediatric patients (with only 1 of 186 patients within our cohort undergoing HSCT in first complete remission), we hypothesize that leukemia cells with *CBFB-MYH11* fusion oncogene are highly susceptible to the allogeneic effect of the HSCT.

Next, we performed unbiased GSEA comparing E5-E33 *CBFB-MYH11* AMLs with all others with alternative breakpoints across the Molecular Signature Database[18,19] (specifically, the hallmark, curated, and oncogenic signature gene sets). We excluded *KIT* mutant AMLs because they have been shown to be associated with a distinct gene expression profile.[20] Intriguingly, 3 previously published gene sets

Figure 2. (continued) enrichment score of 2.42 (P value $<.0001$) with the displayed heatmap of enriched genes. (B) UMAP analysis revealed that E5-E33 *CBFB-MYH11* AMLs occupy a different cluster than most other *CBFB-MYH11* AMLs. (C) GSEA comparing E5-E33 *CBFB-MYH11* AMLs with all others and using a gene set derived by Mandoli et al based on *CBFB-MYH11* chromatin immunoprecipitation sequencing (ChIP-seq) occupancy reveals significant enrichment (normalized enrichment score of 1.37, P value $<.0001$) in E5-E33 *CBFB-MYH11* AMLs (100 most enriched genes shown). Violin plots of representative genes within the Mandoli et al gene set that are associated with hematopoietic stem cells or early hematopoiesis (Mann-Whitney U test).

that established and defined *CBFB-MYH11* AML transcriptional signatures are significantly enriched in E5-E33 *CBFB-MYH11* AMLs compared with non–E5-E33 *CBFB-MYH11* AMLs (Figure 2A), suggesting that non–E5-E33 *CBFB-MYH11* AMLs represent a distinct transcriptional subtype. Indeed, performing UMAP analysis revealed that E5-E33 *CBFB-MYH11* AMLs occupy a distinct cluster from most other *CBFB-MYH11* AMLs ($P < .001$), which are more heterogeneous in their transcriptional clustering (Figure 2B).

To uncover candidate mechanisms to explain the differences in survival outcomes and transcriptional landscapes between E5-E33 *CBFB-MYH11* AMLs and others, we analyzed our transcriptome data in the context of established *CBFB-MYH11* ChIP-seq experimental data.[21] We found that E5-E33 *CBFB-MYH11* AMLs were enriched in the gene set that encompasses all genes located near high-confidence *CBFB-MYH11* ChIP-seq sites (Figure 2C), which included genes implicated in hematopoietic stem cell self-renewal, such as *LRRC33*,[22] *JAG1*,[23] *MKRN2*,[24] *BCOR*,[25] *DEGS1*,[26] *ID1*,[27] *KEAP1*,[28] *IL10RA*,[29] *IL1RAP*,[30] and *CD34*.[31-35] On the basis of their locations in close proximity with experimentally validated *CBFB-MYH11* ChIP-seq sites and their differential expression based on *CBFB-MYH11* transcript subtypes, these genes represent candidate transcriptional markers associated with inferior survival outcomes and present an opportunity for further functional validation and/or therapeutic targeting studies.

Acknowledgments

Authorship

Contribution: B.J.H., X.M., and S.M. conceived and designed the research; J.L.S., Y.-C.W., A.R.L., R.E.R., R.G., and T.A.A. collected and assembled the data; B.J.H., J.L.S., Y.-C.W., K. Taghizadeh., A.R.L., R.E.R., Y.L., P.K., R.G., T.A.A., X.M., and S.M. analyzed and interpreted the data; all authors wrote the paper, had final approval of manuscript, and were accountable for all aspects of the work.

Conflict-of-interest disclosure: The authors declare no competing financial interests.

ORCID profiles: B.J.H., 0000-0001-6996-0833; P.K., 0000-0003-0044-0076; E.C., 0000-0002-2037-0389; T.I.S., 0000-0002-9316-1924; A.S.G., 0000-0003-1513-2893; R.A., 0000-0001-7482-5644; J.E.F., 0000-0003-2148-5839.

Correspondence: Benjamin J. Huang, University of California San Francisco, 550 16th Street, Box 0434, San Francisco, CA 94158; e-mail: ben.huang@ucsf.edu.

References

1. Schwind S, Edwards CG, Nicolet D, et al; Alliance for Clinical Trials in Oncology. inv(16)/t(16;16) acute myeloid leukemia with non-type A CBFB-MYH11 fusions associate with distinct clinical and genetic features and lack KIT mutations. *Blood.* 2013;121(2):385-391.

2. Lange BJ, Smith FO, Feusner J, et al. Outcomes in CCG-2961, a Children's Oncology Group phase 3 trial for untreated pediatric acute myeloid leukemia: a report from the Children's Oncology Group. *Blood.* 2008;111(3):1044-1053.

3. Cooper TM, Franklin J, Gerbing RB, et al. AAML03P1, a pilot study of the safety of gemtuzumab ozogamicin in combination with chemotherapy for newly diagnosed childhood acute myeloid leukemia: a report from the Children's Oncology Group. *Cancer.* 2012;118(3):761-769.

4. Gamis AS, Alonzo TA, Meshinchi S, et al. Gemtuzumab ozogamicin in children and adolescents with de novo acute myeloid leukemia improves event-free survival by reducing relapse risk: results from the randomized phase III Children's Oncology Group trial AAML0531. *J Clin Oncol.* 2014; 32(27):3021-3032.

5. Aplenc R, Meshinchi S, Sung L, et al. Bortezomib with standard chemotherapy for children with acute myeloid leukemia does not improve treatment outcomes: a report from the Children's Oncology Group. *Haematologica.* 2020;105(7):1879-1886.

6. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics.* 2009;25(14):1754-1760.

7. Pliner HA, Packer JS, McFaline-Figueroa JL, et al. Cicero predicts cis-regulatory DNA interactions from single-cell chromatin accessibility data. *Mol Cell.* 2018;71(5):858-871.e8.

8. Robertson G, Schein J, Chiu R, et al. De novo assembly and analysis of RNA-seq data. *Nat Methods.* 2010;7(11):909-912.

9. Tian L, Li Y, Edmonson MN, et al. CICERO: a versatile method for detecting complex and diverse driver fusions using cancer RNA sequencing data. *Genome Biol.* 2020;21(1):126.

10. Edmonson MN, Zhang J, Yan C, Finney RP, Meerzaman DM, Buetow KH. Bambino: a variant detector and alignment viewer for next-generation sequencing data in the SAM/BAM format. *Bioinformatics.* 2011;27(6):865-866.

11. Rusch M, Nakitandwe J, Shurtleff S, et al. Clinical cancer genomic profiling by three-platform sequencing of whole genome, whole exome and transcriptome. *Nat Commun.* 2018;9(1):3962.

12. Wu G, Diaz AK, Paugh BS, et al. The genomic landscape of diffuse intrinsic pontine glioma and pediatric non-brainstem high-grade glioma. *Nat Genet.* 2014;46(5):444-450.

13. Hagiwara K, Ding L, Edmonson MN, et al. RNAIndel: discovering somatic coding indels from tumor RNA-Seq data. *Bioinformatics.* 2020;36(5): 1382-1390.

14. Zhang J, Walsh MF, Wu G, et al. Germline mutations in predisposition genes in pediatric cancer. *N Engl J Med.* 2015;373(24):2336-2346.

15. Ma X, Liu Y, Liu Y, et al. Pan-cancer genome and transcriptome analyses of 1,699 paediatric leukaemias and solid tumours. *Nature.* 2018; 555(7696):371-376.

16. Subramanian A, Tamayo P, Mootha VK, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci USA.* 2005;102(43):15545-15550.

17. Tarlock K, Alonzo TA, Wang YC, et al. Functional properties of *KIT* mutations are associated with differential clinical outcomes and response to targeted therapeutics in CBF acute myeloid leukemia. *Clin Cancer Res.* 2019;25(16):5038-5048.

18. Liberzon A, Subramanian A, Pinchback R, Thorvaldsdóttir H, Tamayo P, Mesirov JP. Molecular signatures database (MSigDB) 3.0. *Bioinformatics.* 2011;27(12):1739-1740.

19. Liberzon A, Birger C, Thorvaldsdóttir H, Ghandi M, Mesirov JP, Tamayo P. The Molecular Signatures Database (MSigDB) hallmark gene set collection. *Cell Syst.* 2015;1(6):417-425.

20. Lück SC, Russ AC, Du J, et al. KIT mutations confer a distinct gene expression signature in core binding factor leukaemia. *Br J Haematol.* 2010; 148(6):925-937.

21. Mandoli A, Singh AA, Jansen PW, et al. CBFB-MYH11/RUNX1 together with a compendium of hematopoietic regulators, chromatin modifiers and basal transcription factors occupies self-renewal genes in inv(16) acute myeloid leukemia. *Leukemia.* 2014;28(4):770-778.

22. Ma W, Qin Y, Chapuy B, Lu C. LRRC33 is a novel binding and potential regulating protein of TGF-β1 function in human acute myeloid leukemia cells. *PLoS One.* 2019;14(10):e0213482.

23. Varnum-Finney B, Purton LE, Yu M, et al. The Notch ligand, Jagged-1, influences the development of primitive hematopoietic precursor cells. *Blood.* 1998;91(11):4084-4091.

24. Lee KY, Chan KY, Tsang KS, et al. Ubiquitous expression of MAKORIN-2 in normal and malignant hematopoietic cells and its growth promoting activity. *PLoS One.* 2014;9(3):e92706.

25. Cao Q, Gearhart MD, Gery S, et al. BCOR regulates myeloid cell proliferation and differentiation. *Leukemia.* 2016;30(5):1155-1165.

26. Xie SZ, Garcia-Prat L, Voisin V, et al. Sphingolipid modulation activates proteostasis programs to govern human hematopoietic stem cell self-renewal. *Cell Stem Cell.* 2019;25(5):639-653.e7.

27. Jankovic V, Ciarrocchi A, Boccuni P, DeBlasio T, Benezra R, Nimer SD. Id1 restrains myeloid commitment, maintaining the self-renewal capacity of hematopoietic stem cells. *Proc Natl Acad Sci USA.* 2007;104(4):1260-1265.

28. Murakami S, Suzuki T, Harigae H, Romeo PH, Yamamoto M, Motohashi H. NRF2 activation impairs quiescence and bone marrow reconstitution capacity of hematopoietic stem cells. *Mol Cell Biol.* 2017;37(19):e00086-17.

29. Kang YJ, Yang SJ, Park G, et al. A novel function of interleukin-10 promoting self-renewal of hematopoietic stem cells. *Stem Cells.* 2007;25(7): 1814-1822.

30. Zhao K, Yin LL, Zhao DM, et al. IL1RAP as a surface marker for leukemia stem cells is related to clinical phase of chronic myeloid leukemia patients. *Int J Clin Exp Med.* 2014;7(12):4787-4798.

31. Servida F, Soligo D, Caneva L, et al. Functional and morphological characterization of immunomagnetically selected CD34+ hematopoietic progenitor cells. *Stem Cells.* 1996;14(4):430-438.

32. Huss R. Isolation of primary and immortalized CD34-hematopoietic and mesenchymal stem cells from various sources. *Stem Cells.* 2000;18(1):1-9.

33. Majeti R, Park CY, Weissman IL. Identification of a hierarchy of multipotent hematopoietic progenitors in human cord blood. *Cell Stem Cell.* 2007; 1(6):635-645.

34. Oguro H, Ding L, Morrison SJ. SLAM family markers resolve functionally distinct subpopulations of hematopoietic stem cells and multipotent progenitors. *Cell Stem Cell.* 2013;13(1):102-116.

35. Pietras EM, Reynaud D, Kang YA, et al. Functionally distinct subsets of lineage-biased multipotent progenitors control blood production in normal and regenerative conditions. *Cell Stem Cell.* 2015;17(1):35-46.