Germline *RUNX1* variation and predisposition to childhood acute lymphoblastic leukemia

Yizhen Li,[1] Wentao Yang,[1] Meenakshi Devidas,[2] Stuart S. Winter,[3] Chimene Kesserwan,[4] Wenjian Yang,[1] Kimberly P. Dunsmore,[5] Colton Smith,[1] Maoxiang Qian,[6] Xujie Zhao,[1] Ranran Zhang,[1] Julie M. Gastier-Foster,[7] Elizabeth A. Raetz,[8] William L. Carroll,[8] Chunliang Li,[9] Paul P. Liu,[10] Karen R. Rabin,[11] Takaomi Sanda,[12,13] Charles G. Mullighan,[14] Kim E. Nichols,[15] William E. Evans,[1,16] Ching-Hon Pui,[15,16] Stephen P. Hunger,[17] David T. Teachey,[17] Mary V. Relling,[1,16] Mignon L. Loh,[18] and Jun J. Yang[1,15,16]

[1]Department of Pharmaceutical Sciences and [2]Department of Global Pediatric Medicine, St. Jude Children's Research Hospital, Memphis, Tennessee, USA. [3]Children's Minnesota Research Institute, Children's Minnesota, Minneapolis, Minnesota, USA. [4]Center for Cancer Research, Genetics Branch, National Cancer Institute, Bethesda, Maryland, USA. [5]Children's Hematology and Oncology, Carilion Clinic and Virginia Tech Carilion School of Medicine, Roanoke, Virginia, USA. [6]Institute of Pediatrics and Department of Hematology and Oncology, Children's Hospital of Fudan University, Institutes of Biomedical Sciences, Shanghai, China. [7]Baylor College of Medicine and Texas Children's Hospital, Houston, Texas, USA. [8]Division of Pediatric Hematology and Oncology, Perlmutter Cancer Center, New York University Langone Health, New York, New York, USA. [9]Tumor Cell Biology, St. Jude Children's Research Hospital, Memphis, Tennessee, USA. [10]Oncogenesis and Development Section, National Human Genome Research Institute, NIH, Bethesda, Maryland, USA. [11]Texas Children's Cancer and Hematology Centers, Baylor College of Medicine, Houston, Texas, USA. [12]Cancer Science Institute of Singapore, and [13]Department of Medicine, Yong Loo Lin School of Medicine, National University of Singapore, Singapore. [14]Department of Pathology, [15]Department of Oncology, and [16]Hematological Malignancies Program, St. Jude Children's Research Hospital, Memphis, Tennessee, USA. [17]Department of Pediatrics and Center for Childhood Cancer Research, Children's Hospital of Philadelphia and Perelman School of Medicine, University of Pennsylvania, Philadelphia, Pennsylvania, USA. [18]Department of Pediatrics, Benioff Children's Hospital and the Helen Diller Family Comprehensive Cancer Center, UCSF, San Francisco, California, USA.

Genetic alterations in the *RUNX1* gene are associated with benign and malignant blood disorders, particularly of megakaryocyte and myeloid lineages. The role of RUNX1 in acute lymphoblastic leukemia (ALL) is less clear, particularly in terms of how germline genetic variation influences the predisposition to this type of leukemia. Sequencing DNA of 4836 children with B cell ALL (B-ALL) and 1354 with T cell ALL (T-ALL), we identified 31 and 18 germline *RUNX1* variants, respectively. *RUNX1* variants in B-ALL consistently showed minimal damaging effects. In contrast, 6 T-ALL–related variants resulted in drastic loss of RUNX1 activity as a transcription activator in vitro. Ectopic expression of dominant-negative *RUNX1* variants in human CD34$^+$ cells repressed differentiation into erythroid cells, megakaryocytes, and T cells, while promoting myeloid cell development. Chromatin immunoprecipitation sequencing of T-ALL models showed distinctive patterns of RUNX1 binding by variant proteins. Further whole-genome sequencing identified the *JAK3* mutation as the most frequent somatic genomic abnormality in T-ALL with germline *RUNX1* variants. Cointroduction of *RUNX1* variant and *JAK3* mutation in hematopoietic stem and progenitor cells in mice gave rise to T-ALL with the early T cell precursor phenotype. Taken together, these results indicate that *RUNX1* is an important predisposition gene for T-ALL and point to biology of RUNX1-mediated leukemogenesis in the lymphoid lineages.

Introduction

Acute lymphoblastic leukemia (ALL) is the most common cancer in children. The exact cause of ALL is incompletely understood, although somatic genomic abnormalities affecting a wide range of signaling pathways are well documented (1). There is also growing evidence of inherited susceptibility to ALL. For example, common genetic polymorphisms in genes, such as *IKZF1* (2), *ARID5B* (3), *CDKN2A* (4), *GATA3* (5, 6), *CEBPE* (7), *PIP4K2A* (8), and *TCF3-PBX1* (9), are associated with the risk of ALL in an age- and subtype-dependent manner. On the other hand, rare germline variants have been linked to familial predisposition to childhood ALL; this is particularly notable in *TP53* (10), *ETV6* (11), and *IKZF1* (2, 9). These findings point to a strong genetic basis of interindividual variability in ALL risk.

The RUNX1 protein plays key roles in definitive hematopoiesis (12). RUNX1 functions as a transcription factor by forming a heterodimer with core binding factor β (CBFβ). RUNX1 consists of a Runt homology domain (RHD) responsible for DNA binding and cofactor interaction (13) and the C-terminal transcriptional activation domain (TAD), which recruits coactivators and activates the expression of RUNX1 target genes (14). *RUNX1* germline variants are associated with familial platelet disorder (FPD). Many patients with FPD develop leukemia later in life, predominately acute myeloid leukemia (AML) and myelodysplastic syndrome (MDS) (15–18). Somatic *RUNX1* mutations are also recurrent in B- and T-ALL; they mostly occur in RHD and TAD and are more common in T-ALL (19, 20). *RUNX1* mutations also are related to poor prognosis in T-ALL (19) and particularly enriched in those with early T

▶ Related Commentary: https://doi.org/10.1172/JCI152464

Authorship note: YL and Wentao Yang contributed equally to this work.
Conflict of interest: The authors have declared that no conflict of interest exists.

Submitted: January 20, 2021; **Accepted:** June 22, 2021; **Published:** September 1, 2021.
Reference information: *J Clin Invest*. 2021;131(17):e147898.
https://doi.org/10.1172/JCI147898.

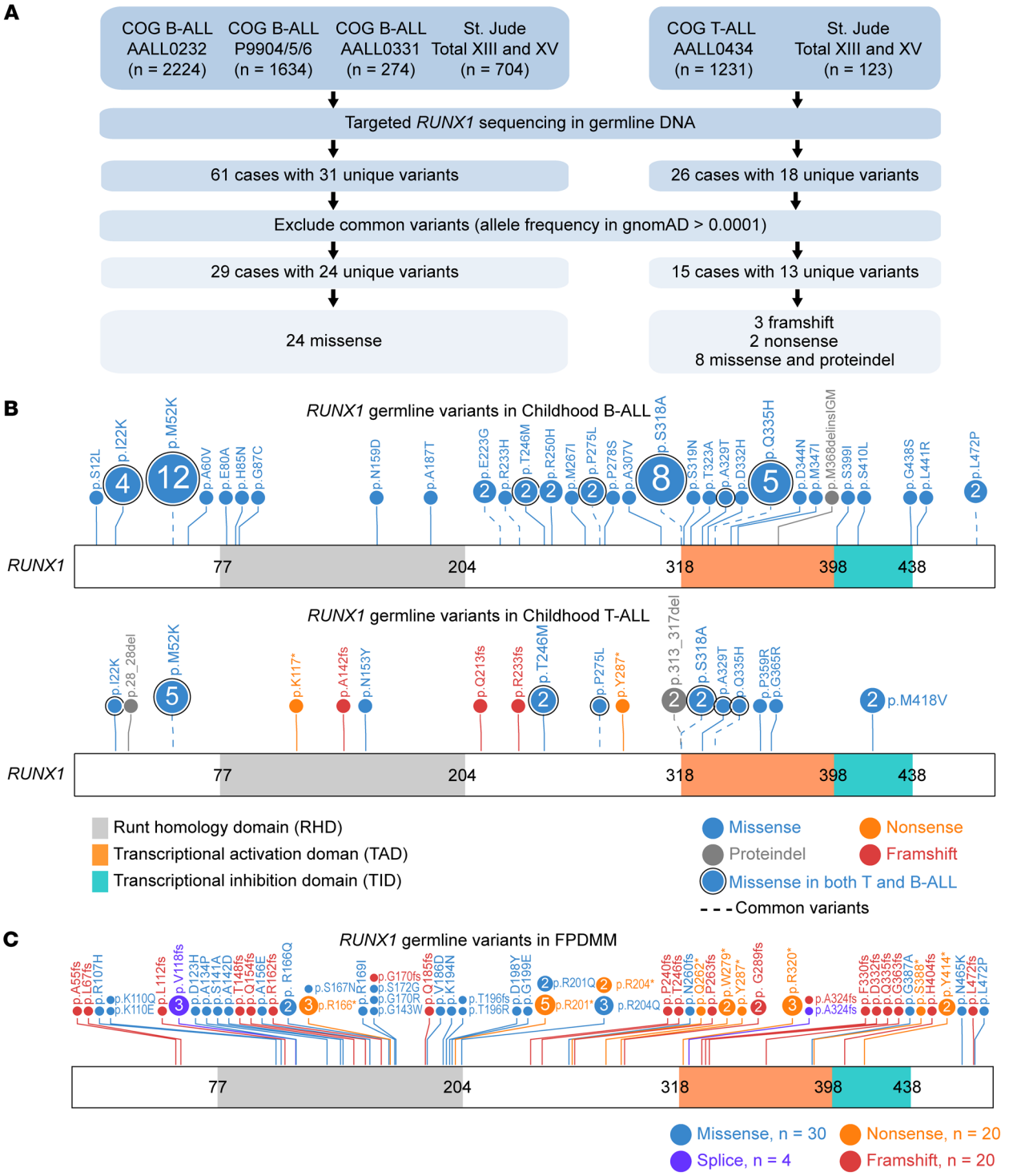


Figure 1. Germline *RUNX1* variants in childhood B- and T-ALL. (**A**) CONSORT diagram of the COG and St. Jude patients included in this study. (**B**) Protein domain plot of RUNX1 and the amino acid substitutions predicted to result from the germline *RUNX1* variants identified in this study. The upper panel shows germline *RUNX1* variants in B-ALL patients, and the lower panel shows those in T-ALL patients. Numbers in circles indicate the number of patients in our cohort that harbor the variant of interest. (**C**) Protein domain plot of RUNX1 and germline RUNX1 variants identified previously in FPDMM. Data were retrieved from recently published paper (22).

cell precursor immunophenotype (ETP) (21). The role of germline *RUNX1* variants in the pathogenesis of ALL is much less known, and their pattern, prevalence, and functional consequences in this lymphoid malignancy have not been comprehensively investigated.

Here, we report results from targeted germline sequencing of 6190 children with B- or T-ALL enrolled in the frontline Children's Oncology Group (COG) and St. Jude Children's Research Hospital (St. Jude) ALL trials. We observed a lineage-

Table 1. Germline RUNX1 variants in pediatric ALL patients

Protein (NM_001754)	CDS (NM_001754)	Variation type	ALL subtype (number of patients)	Allele frequency in gnomAD (allele count)	REVEL score	ACMG classification
p.S12L	c.C35T	Missense	B-ALL (*n* = 1)	(7.95×10^{-06}) 2	0.305	VUS
p.I22K	c.T65A	Missense	B-ALL (*n* = 4) and T-ALL (*n* = 1)	(1.99×10^{-04}) 56	0.271	LB
p.28_28del	c.82_84del	Nonframeshift_deletion	T-ALL (*n* = 1)	(4.62×10^{-05}) 13	.	VUS
p.M52K	c.T155A	Missense	B-ALL (*n* = 12) and T-ALL (*n* = 5)	(2.46×10^{-04}) 65	0.845	VUS
p.A60V	c.C179T	Missense	B-ALL (*n* = 1)	(7.40×10^{-05}) 17	0.284	B
p.E80A	c.A239C	Missense	B-ALL (*n* = 1)	(1.22×10^{-05}) 3	0.856	VUS
p.H85N	c.C253A	Missense	B-ALL (*n* = 1)	(3.59×10^{-05}) 10	0.852	LB
p.G87C	c.G259T	Missense	B-ALL (*n* = 1)	.	0.918	VUS
p.K117*	c.A349T	Stopgain	T-ALL (*n* = 1)	.	.	LP
p.A142fs	c.422_423insAACC	Frameshift_insertion	T-ALL (*n* = 1)	.	.	LP
p.N153Y	c.A457T	Missense	T-ALL (*n* = 1)	.	0.956	VUS
p.N159D	c.A475G	Missense	B-ALL (*n* = 1)	.	0.956	VUS
p.A187T	c.G559A	Missense	B-ALL (*n* = 1)	(3.98×10^{-06}) 1	0.96	VUS
p.Q213fs	c.637delC	Frameshift_deletion	T-ALL (*n* = 1)	.	.	LP
p.E223G	c.A668G	Missense	B-ALL (*n* = 2)	(1.24×10^{-04}) 31	0.799	VUS
p.R233fs	c.696delG	Frameshift_deletion	T-ALL (*n* = 1)	.	.	LP
p.R233H	c.G698A	Missense	B-ALL (*n* = 1)	(1.67×10^{-04}) 47	0.514	LB
p.T246M	c.C737T	Missense	B-ALL (*n* = 2) and T-ALL (*n* = 2)	(4.61×10^{-05}) 13	0.4	VUS
p.R250H	c.G749A	Missense	B-ALL (*n* = 2)	(3.91×10^{-05}) 11	0.488	VUS
p.M267I	c.G801A	Missense	B-ALL (*n* = 1)	(3.93×10^{-05}) 11	0.283	VUS
p.P275L	c.C824T	Missense	B-ALL (*n* = 2) and T-ALL (*n* = 1)	(2.03×10^{-04}) 51	0.146	B
p.P278S	c.C832T	Missense	B-ALL (*n* = 1)	(1.19×10^{-05}) 3	0.275	VUS
p.Y287*	c.C861G	Stopgain	T-ALL (*n* = 1)	.	.	LP
p.A307V	c.C920T	Missense	B-ALL (*n* = 1)	.	0.098	VUS
p.313_317del	c.939_950del	Nonframeshift_deletion	T-ALL (*n* = 2)	(1.73×10^{-04}) 49	.	LB
p.S318A	c.T952G	Missense	B-ALL (*n* = 8) and T-ALL (*n* = 2)	(8.13×10^{-04}) 230	0.093	B
p.S319N	c.G956A	Missense	B-ALL (*n* = 1)		0.127	VUS
p.T323A	c.A967G	Missense	B-ALL (*n* = 1)	(1.59×10^{-05}) 4	0.063	LB
p.A329T	c.G985A	Missense	B-ALL (*n* = 1) and T-ALL (*n* = 1)	.	0.067	LB
p.D332H	c.G994C	Missense	B-ALL (*n* = 1)	.	0.391	VUS
p.Q335H	c.G1005T	Missense	B-ALL (*n* = 5) and T-ALL (*n* = 1)	(1.70×10^{-04}) 43	0.4	B
p.D344N	c.G1030A	Missense	B-ALL (*n* = 1)	(6.20×10^{-05}) 16	0.082	LB
p.M347I	c.G1041A	Missense	B-ALL (*n* = 1)	(8.79×10^{-06}) 2	0.374	VUS
p.P359R	c.C1076G	Missense	T-ALL (*n* = 1)	.	0.62	VUS
p.G365R	c.G1093C	Missense	T-ALL (*n* = 1)	.	0.776	LP
p.M368delinsIGM	c.1103_1104insCGGCAT	Nonframeshift_insertion	B-ALL (*n* = 1)	(7.34×10^{-05}) 17	.	VUS
p.S399I	c.G1196T	Missense	B-ALL (*n* = 1)	(2.89×10^{-05}) 5	0.116	VUS
p.S410L	c.C1229T	Missense	B-ALL (*n* = 1)	(1.56×10^{-05}) 3	0.339	VUS
p.M418V	c.A1252G	Missense	T-ALL (*n* = 2)	(9.79×10^{-05}) 3	0.14	LB
p.G438S	c.G1312A	Missense	B-ALL (*n* = 1)	.	0.23	VUS
p.L441R	c.T1322G	Missense	B-ALL (*n* = 1)	(7.28×10^{-06}) 1	0.74	VUS
p.L472P	c.T1415C	Missense	B-ALL (*n* = 2)	(1.99×10^{-04}) 23	0.414	B

Additional variant annotations (e.g., chromosomal positions) can be found in Supplemental Table 10. CDS, coding region; VUS, variant of unknown significance; LB, likely benign; LP, likely pathogenic.

specific pattern of germline variation in the *RUNX1* gene, with deleterious variants exclusively present in T-ALL patients. Furthermore, we experimentally characterized *RUNX1* variants for their effects on transcription factor activity, subcellular localization, cofactor interaction, in vitro hematopoiesis, and genome-wide RUNX1-binding profile. Finally, we examined the somatic genomic landscape of T-ALL arising from *RUNX1* germline variants and modeled *RUNX1*-mediated leukemogenesis in mice.

Results

Identification of germline RUNX1 variants in pediatric ALL. To comprehensively characterize inherited *RUNX1* variations in ALL, we performed targeted sequencing in germline DNA of 4836 patients with newly diagnosed B cell ALL (B-ALL) and 1354 patients with T-ALL enrolled in COG and St. Jude frontline trials (Figure 1A and Table 1). We identified 31 unique variants in 61 B-ALL patients and 18 unique variants in 26 T-ALL patients. Seven of these variants were found in both B- and T-ALL (Figure 1A and Table 1).

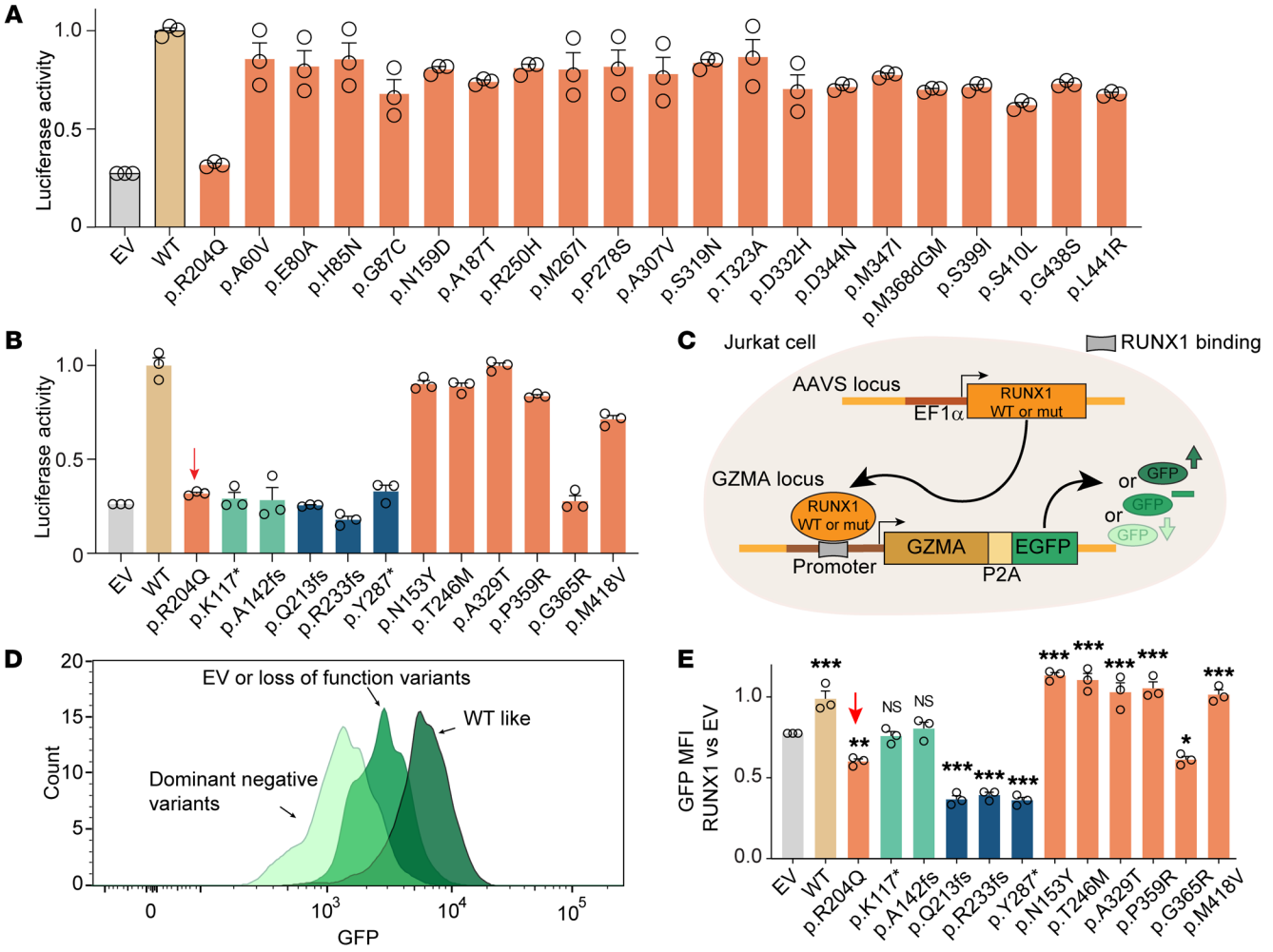


Figure 2. Germline *RUNX1* variants influence transcription factor activity, subcellular localization, and CBFβ interaction. (**A** and **B**) Luciferase reporter gene assay (driven by the *PU.1* promoter in HeLa cells) showed minimal effects on transcription factor activity by missense *RUNX1* variants identified in B-ALL (**A**) and T-ALL (**B**). A previously reported damaging variant p.R204Q was included as the reference for luciferase assay (red arrow). (**C**) Design of the Jurkat landing-pad system to measure RUNX1 variant activity in T-ALL. *RUNX1* (either WT or variant) was inserted at the AAVS locus. EGFP coding sequence was knocked at the 3′ end of *GZMA*, a RUNX1 target gene. RUNX1 transcription factor activity was determined by flow cytometry of GFP signal, which reflects RUNX1-driven *GZMA* transcription. (**D**) Flow cytometry analysis of Jurkat cells expressing different *RUNX1* variants. Cells harboring dominant-negative, loss-of-function, and WT-like RUNX1 variants exhibited the lowest, moderate, and highest GFP signals, respectively. (**E**) GFP signals from Jurkat cells expressing each *RUNX1* variant (relative to WT) are shown in a bar graph. Data represent mean ± SEM ($n = 3$). The difference of each variant relative to EV was evaluated using Dunnett's test. p.R204Q was used as control. $^*P < 0.05$; $^{**}P < 0.01$; $^{***}P < 0.001$.

Of the 31 variants in B-ALL, 6 were not observed in the general population (Genome Aggregation Database [gnomAD], $n = 15,496$), 18 were rare, with a maximum allele frequency of 0.00122%, and the remaining 7 were considered common variants, with allele frequency of greater than 0.01% (Figure 1A and Table 1). All variants in B-ALL except one were missense; most of these were in the C terminus distal to the DNA-binding RHD (Figure 1B). Of the 18 variants in T-ALL patients, 8 were absent in the gnomAD data set, 5 were rare, with a maximum allele frequency of 0.00239%, and the remaining 5 were common variants (Figure 1A and Table 1). Five of the T-ALL–related variants were frameshift or nonsense, including (a) p.K117* and p.A142fs, which truncated both RHD and TAD (Figure 1B and Supplemental Figure 1; supplemental material available online with this article; https://doi.org/10.1172/JCI147898DS1), and (b) p.Q213fs, p.R233fs, and p.Y287*, which resulted in the loss of TAD only (Supplemental Figure 1). Seven missense and 1 in-frame deletion variants in T-ALL were distributed across *RUNX1*. This is significantly different from the pattern of missense *RUNX1* variants in FPD with associated myeloid malignancy (FPDMM) (Figure 1C and ref. 22), which are largely restricted to the DNA-binding domain ($P = 8.35 \times 10^{-5}$, Fisher's exact test).

Effects of RUNX1 variants on transcriptional regulation, cellular localization, and protein-protein interaction. To understand how germline *RUNX1* variants affect gene function, we first examined their transcription activator activity using the luciferase reporter assay in HeLa cells. With *SPI1* as the RUNX1 target gene (23), none of the germline variants identified in B-ALL showed a significant impact on reporter gene transcription compared with the WT protein, and therefore, these were not studied further (Figure 2A and Supplemental Figure 1A). Among *RUNX1* alleles seen in T-ALL, all frame-

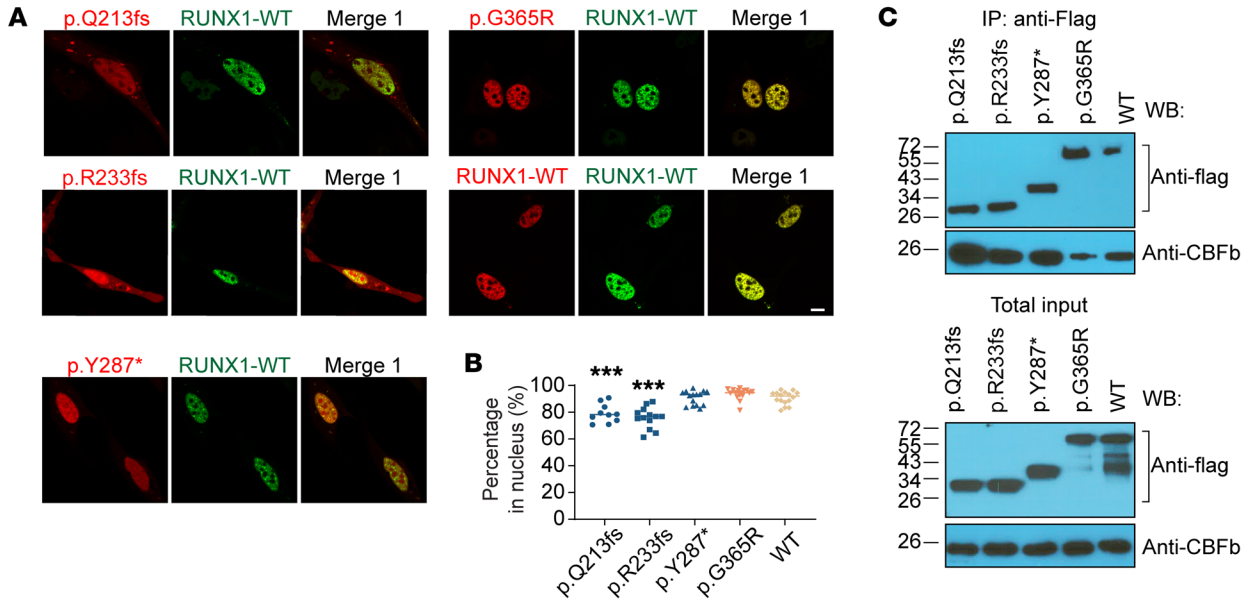


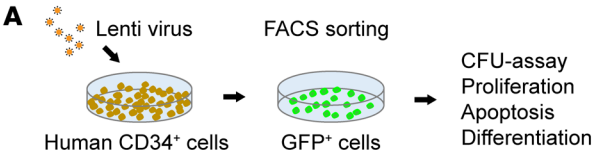
Figure 3. Dominant-negative *RUNX1* variants retain nuclear localization and CBFβ interaction. (**A**) Immunofluorescence microscopy shows subcellular localization of mCherry-tagged variant proteins and EGFP-tagged WT RUNX1. Variant and WT RUNX1 were fused to mCherry and EGFP and expressed transiently in HEK293T cells, which were then subjected to imaging analyses. Scale bar: 10 μm. (**B**) Summary of the percentages of mCherry signals in nucleus is shown as dot plot. The difference of each variant relative to WT was evaluated using Dunnett's test. p.Q213fs, $n = 10$; p.R233fs, $n = 13$; p.Y287*, $n = 15$; p.G365R, $n = 16$; WT, $n = 16$. (**C**) Coimmunoprecipitation assay was performed to determine RUNX1-CBFβ interaction for each deleterious variant. Experiments were performed in HEK293T cells. RUNX1 proteins were pulled down using anti–FLAG-M2 beads, and the presence or absence of CBFβ in the pellet was determined by immunoblotting.

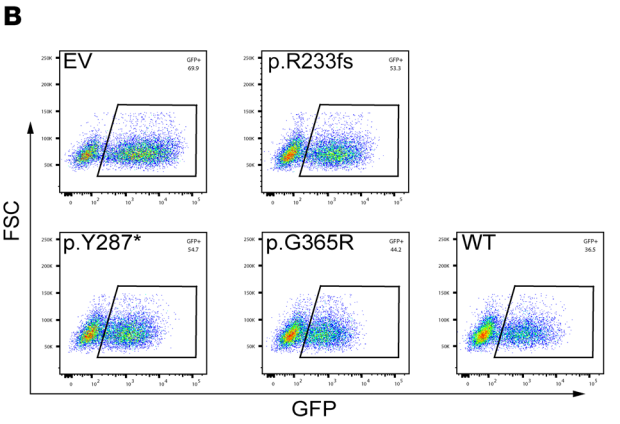
shift and nonsense variants (p.K117*, p.A142fs, p.S213fs, p.R233fs, and p.Y287*) and also missense variant G365R caused a significant reduction of RUNX1 activity in this assay (Figure 2B and Supplemental Figure 1B). As previously reported, dominant-negative variant p.R204Q was used as a reference (24, 25). Notably, p.K117* and p.A142fs proteins were only modestly expressed compared with WT and other variants, suggesting that the lack of RUNX1 activity may be due to reduced translation and/or the instability of truncated protein. To further characterize these *RUNX1* variants in a more relevant cellular context, we engineered the Jurkat T-ALL cell line in which each *RUNX1* variant of interest was individually inserted into the safe-harbor *AAVS1* locus (Supplemental Figure 2 and ref. 26) and a GFP tag was added to the C terminus of RUNX1 target gene *GZMA* (ref. 27, Figure 2C, and Supplemental Figure 3). Using this model system, RUNX1 transactivation activity could be directly measured as the GFP intensity in *RUNX1* variant knockin cells in the presence of endogenous *RUNX1* (Figure 2D). As shown in Figure 2E, the introduction of WT *RUNX1* as well as most missense variants (p.N153Y, p.T246M, p.A329T, p.P359R, and p.M418V) led to robust GFP signals relative to cells with no *RUNX1* insertion at the AAVS locus (empty vector [EV]), confirming WT-like transcription activator activity. In contrast, cells with K117* and A142fs showed no expression of variant RUNX1 protein and thus only baseline GFP signals (Figure 2E and Supplemental Figure 1C). Insertion of the p.Q213fs, p.R233fs, p.Y287*, and p.G365R variants resulted in the lowest GFP intensity (despite robust protein expression; Figure 2E and Supplemental Figure 1C), suggesting these variants not only lost their transcription activator activity, but also repressed endogenous RUNX1 in a plausibly dominant-negative manner.

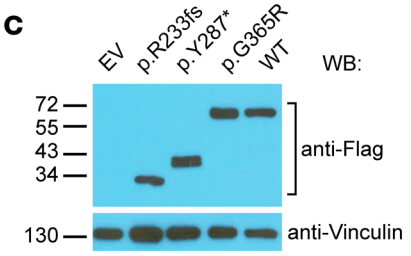
We next analyzed subcellular localization and CBFβ cofactor interaction of deleterious variants that are readily expressed, including p.Q213fs, p.R233fs, p.Y287*, and p.G365R (Supplemental Figure 4). Fluorescence microscopy of HEK293T cells ectopically expressing *RUNX1* variants showed that p.Q213fs, p.R233fs, p.Y287*, and p.G365R proteins mostly remained in the nucleus and retained the ability to interact with CBFβ (Figure 3).

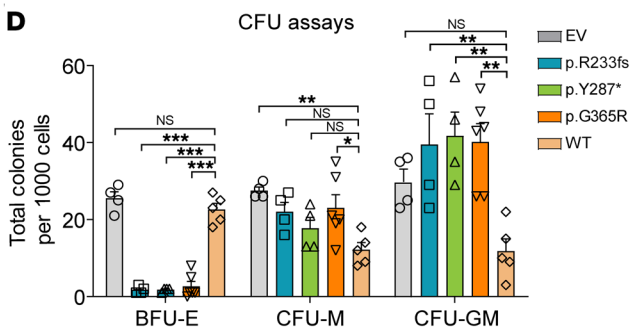
Effects of RUNX1 variants on differentiation and proliferation of human cord blood CD34⁺ cells in vitro. We sought to examine the effects of *RUNX1* variants on hematopoietic differentiation in vitro using human cord blood CD34⁺ cell as the model system. Because p.K117* and p.A142fs resulted in complete loss of function with no dominant-negative effects, we chose not to further characterize them. For the remaining deleterious variants, we selected p.R233fs, p.Y287*, and p.G365R to represent frameshift, nonsense, and missense variants, respectively. *RUNX1* variants were ectopically expressed in human CD34⁺ cells, which were then subjected to differentiation, proliferation, and apoptosis assays in vitro (Figure 4, A–C).

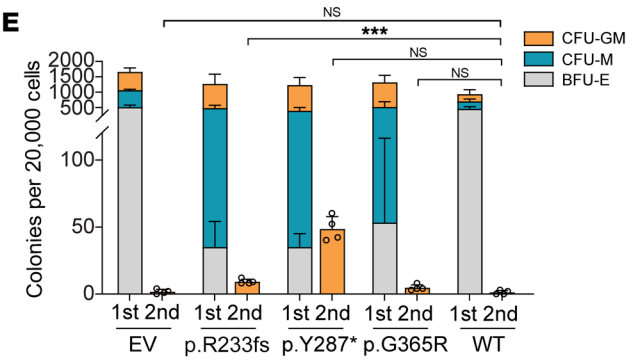
In colony-formation assays conditioned for erythroid and myeloid progenitor cell growth, the expression of p.R233fs, p.Y287*, and p.G365R significantly repressed burst-forming unit erythroid (BFU-E) and increased CFU granulocyte-macrophage (CFU-GM) colonies compared with CD34⁺ cells transduced with WT RUNX1 (Figure 4D and Supplemental Figure 5, A and B). The immunophenotype of these progenitor cells was confirmed by flow cytometry (Supplemental Figure 6A), with cell identities also examined using Giemsa staining (Supplemental Figure 5C). Upon replating, *RUNX1* variants (especially p.Y287*) gave rise to more CFU-GM colonies relative to WT, suggesting their potential effects on hematopoietic cell

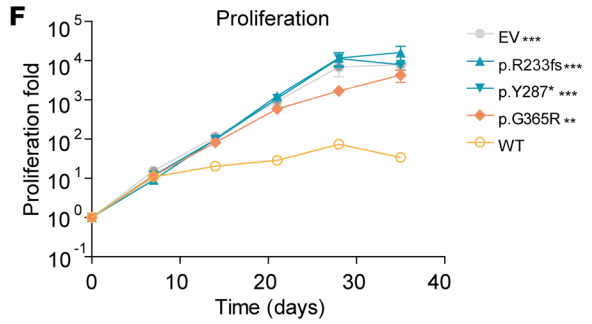

A
Lenti virus
FACS sorting
Human CD34+ cells
GFP+ cells
CFU-assay
Proliferation
Apoptosis
Differentiation

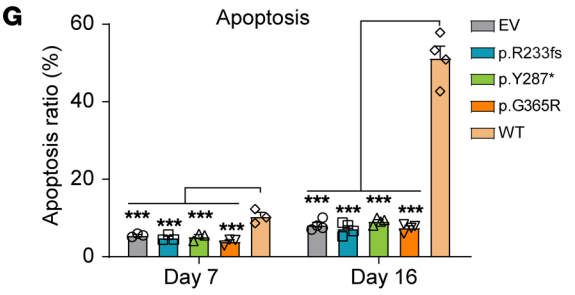

B
EV
p.R233fs
p.Y287*
p.G365R
WT
FSC
GFP

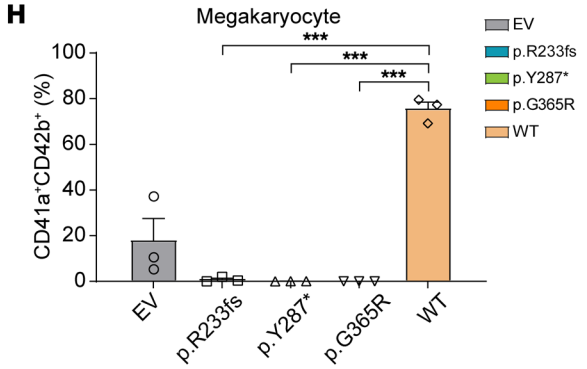

C
EV
p.R233fs
p.Y287*
p.G365R
WT
WB:
72
55
43
34
anti-Flag
130
anti-Vinculin


D
CFU assays
Total colonies per 1000 cells
BFU-E
CFU-M
CFU-GM
EV
p.R233fs
p.Y287*
p.G365R
WT


E
Colonies per 20,000 cells
CFU-GM
CFU-M
BFU-E
1st 2nd
EV
p.R233fs
p.Y287*
p.G365R
WT


F
Proliferation
Proliferation fold
Time (days)
EV
p.R233fs
p.Y287*
p.G365R
WT


G
Apoptosis
Apoptosis ratio (%)
Day 7
Day 16
EV
p.R233fs
p.Y287*
p.G365R
WT


H
Megakaryocyte
CD41a+CD42b+ (%)
EV
p.R233fs
p.Y287*
p.G365R
WT

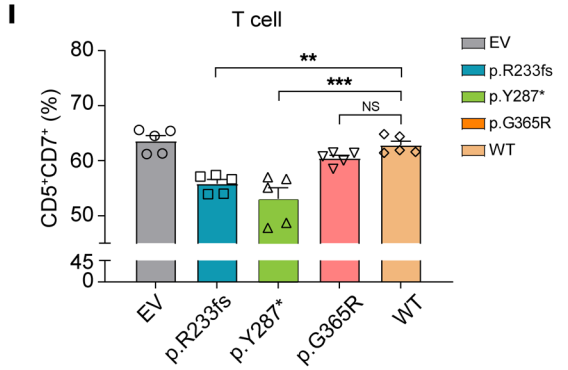

I
T cell
CD5+CD7+ (%)
EV
p.R233fs
p.Y287*
p.G365R
WT

Figure 4. *RUNX1* variants affect in vitro differentiation of human cord blood CD34⁺ cells. (**A** and **B**) Schematic showing in vitro hematopoietic differentiation assay. *RUNX1* variants (with IRES-GFP) were lentivirally introduced into human cord blood CD34⁺ cells. Successfully transduced cells were sorted by flow cytometry (**B**), processed for CFU assays, and assessed for cell proliferation and apoptosis, as appropriate. (**C**) Western blot was used to confirm RUNX1 expression, with vinculin as the internal control. (**D**) 1000 RUNX1-expressing CD34⁺ cells were plated in MethoCult H4034. The *y* axis shows the count of colonies for each lineage: BFU-E, CFU-macrophage (CFU-M), and CFU-GM. EV, p.R233fs, and p.Y287*, $n = 3$; p.G365R, $n = 6$; WT, $n = 5$. (**E**) Colony number of CFU assays and CFU-replating assays ($n = 4$). (**F**) Proliferation of RUNX1-expressing CD34⁺ cells was monitored for 5 weeks in IMDM containing 20% BIT9500, 10 ng/mL FLT-3 ligand, TPO, SCF, IL-3, and IL-6. The number of cells was counted every week for 5 weeks. $n = 4$. (**G**) Apoptosis of RUNX1-transduced CD34⁺ cells after 7 ($n = 3$) and 16 ($n = 4$) days of culture (same culture medium as in **F**) was measured by flow cytometry using annexin V and DAPI antibodies. (**H** and **I**) CD34⁺ cells ectopically expressing RUNX1 variants were also subjected to in vitro differentiation assays for megakaryocyte ($n = 3$) or T cell ($n = 4$) lineages. Following RUNX1 transduction, cells were cultured in the presence of SFEMII-containing megakaryocyte expansion supplement or T cell progenitor differentiation supplement for 2 weeks. Megakaryocyte (**H**) was identified as CD41a⁺CD42b⁺, and T cells (**I**) were defined as CD5⁺CD7⁺ by flow cytometry. Data are represented as mean ± SEM. *P* values were estimated by using Dunnett's test. $*P < 0.05$; $**P < 0.01$; $***P < 0.001$.

self-renewal (Figure 4E). Long-term culture showed that the *RUNX1* variant–transduced CD34⁺ cells proliferated significantly faster with concomitant reduction in apoptosis compared with WT *RUNX1* cells (Figure 4, F and G, and Supplemental Figure 6B).

With culture conditions for megakaryocyte differentiation, expression of *RUNX1* variants consistently resulted in a significant reduction of CD41a⁺CD42b⁺ population compared with that in WT (Figure 4H). These variants also repressed the generation of CD5⁺CD7⁺ T cells from the CD34⁺ population (Figure 4I). Collectively, these results suggest that *RUNX1* variants promoted myeloid differentiation while repressing megakaryocyte and T cell differentiation in vitro.

RUNX1 variants have highly distinctive patterns of DNA binding and are associated with altered posttranslational modifications. To understand the molecular effects of *RUNX1* variants, we comprehensively profiled RUNX1 binding across the genome using ChIP-Seq. We first engineered 3 isogenic Jurkat cell lines in which each *RUNX1* variant (p.R233fs, p.Y287*, and p.G365R) was individually knocked in at the endogenous locus in a hemizygous fashion to represent heterozygous genotypes seen in patients (Figure 5, A and B). In these models, we introduced the HA and TY1 epitope tags at the 3′ end of the coding exon on the variant and WT *RUNX1* alleles, respectively (Supplemental Figures 7–10 and Figure 5C). This enabled us to separately profile variant or WT RUNX1 binding using HA or TY1 antibodies (Figure 5D). To control for differences in ChIP-Seq efficiency between HA and TY1 antibodies, we also generate Jurkat cells in which WT *RUNX1* alleles were tagged with HA or TY1, respectively (Supplemental Figures 9 and 10).

ChIP-Seq showed both similarities and differences in RUNX1 binding between variant and WT proteins in the T-ALL genome (Figure 5, E and F). Calculating a pairwise Pearson's correlation coefficient in ChIP-Seq signal intensity at all identified binding sites, we observed the highest concordance between WT and p.G365R and also between p.Y287 and p.R233fs (Figure 5F and Supplemental Figure 11). In contrast, the truncating variants showed a very distinct pattern of RUNX1 binding compared with missense variant or WT (Figure 5F and Supplemental Figure 11). For each RUNX1 variant, we also identified binding sites unique to variant protein versus those shared with WT RUNX1 (Supplemental Figure 12A). In total, we defined 59,151 peaks shared by WT and variant RUNX1, and 2026, 782, and 93 ChIP-Seq signals specific to p.R233fs, p.Y287*, and p.G365R, respectively. While WT RUNX1-binding sites were found both within and outside of promoter regions, variant-specific sites were almost exclusively in introns and intergenic regions (Supplemental Figure 12B). Our RUNX1 ChIP-Seq results were generally consistent with results published previously for hematopoietic tissues (refs. 27, 28, and Supplemental Table 1).

In T-ALL, RUNX1 often functions through a transcription factor complex with multiple components (e.g., GATA3, TAL1, E2A, HEB, and LMO1; ref. 27). A question arises as to whether or how this might be affected by RUNX1 variants. Comparing our results with previously published ChIP-Seq data of RUNX1 complex members (27), we noted that WT RUNX1-binding sites overlap with GATA3, TAL1, E2A, HEB, and LMO1 binding, but they were largely absent within ChIP-Seq peaks unique to RUNX1 variants, with the exception of CTCF (ref. 29 and Supplemental Figure 12C). This was further confirmed by de novo motif analyses (Supplemental Table 2). Furthermore, we performed gene-set enrichment analysis to identify pathways that were preferentially affected by RUNX1 variants and observed a preponderance of IL-2–STAT5– and TGF-β–signaling genes (Supplemental Figure 12D). At last, we identified 402, 424, and 136 differentially expressed genes between variants (p.R233fs, p.Y287*, or p.G365R) and WT Jurkat cells by RNA-Seq, respectively, most of which also harbor RUNX1-binding sites as identified by ChIP-Seq (Supplemental Figure 12E and Supplemental Tables 3–5).

Interestingly, the p.G365R variant gave rise to a methylation site in RUNX1, with mono- or dimethylation of the arginine residue confirmed by mass spectrometry and Western blot analysis (Figure 5, G and H). Immunoprecipitation–mass spectrometry results suggested that RUNX1 protein methylation at this site may disrupt its interaction with TUBB family proteins (TUBB2A, TUBB2B, TUBB4B, TUBB5, TUB8, et al.) and heat shock proteins, but with an increase of CBFβ binding (Supplemental Table 6).

Somatic genomic abnormalities in T-ALL with germline RUNX1 variants. To characterize the somatic genomic landscape of T-ALL with germline *RUNX1* variants, we analyzed whole-genome sequencing of 6 patients with p.K117*, p.A142fs, p.Q213fs, p.R233fs, p.Y287*, and p.G365R variants, which were contrasted with 263 T-ALL with somatic mutations in *RUNX1* or WT genotypes (30). Five of 6 T-ALL (83.3%) with germline *RUNX1* variants had a somatic *JAK3* mutation, which was significantly higher compared with the frequency of *JAK3* mutation percentages in T-ALL patients without germline variants in *RUNX1* (7.6%, $P = 2.59 \times 10^{-5}$) or T-ALL patients with somatic mutations in *RUNX1* (27.3%, $P = 0.05$; Figure 6A). *JAK3* mutations in T-ALL patients with germline *RUNX1* variants were located in either the pseudo-kinase domain (M511I and R657Q) or in the kinase domain (L950V; Supplemental Figure 13 and Supplemental Tables 7 and 8). Of interest, the patient with a germline *RUNX1*-p.R233fs variant also subsequently acquired a somatic *RUNX1* mutation (R169_E5splice_region).

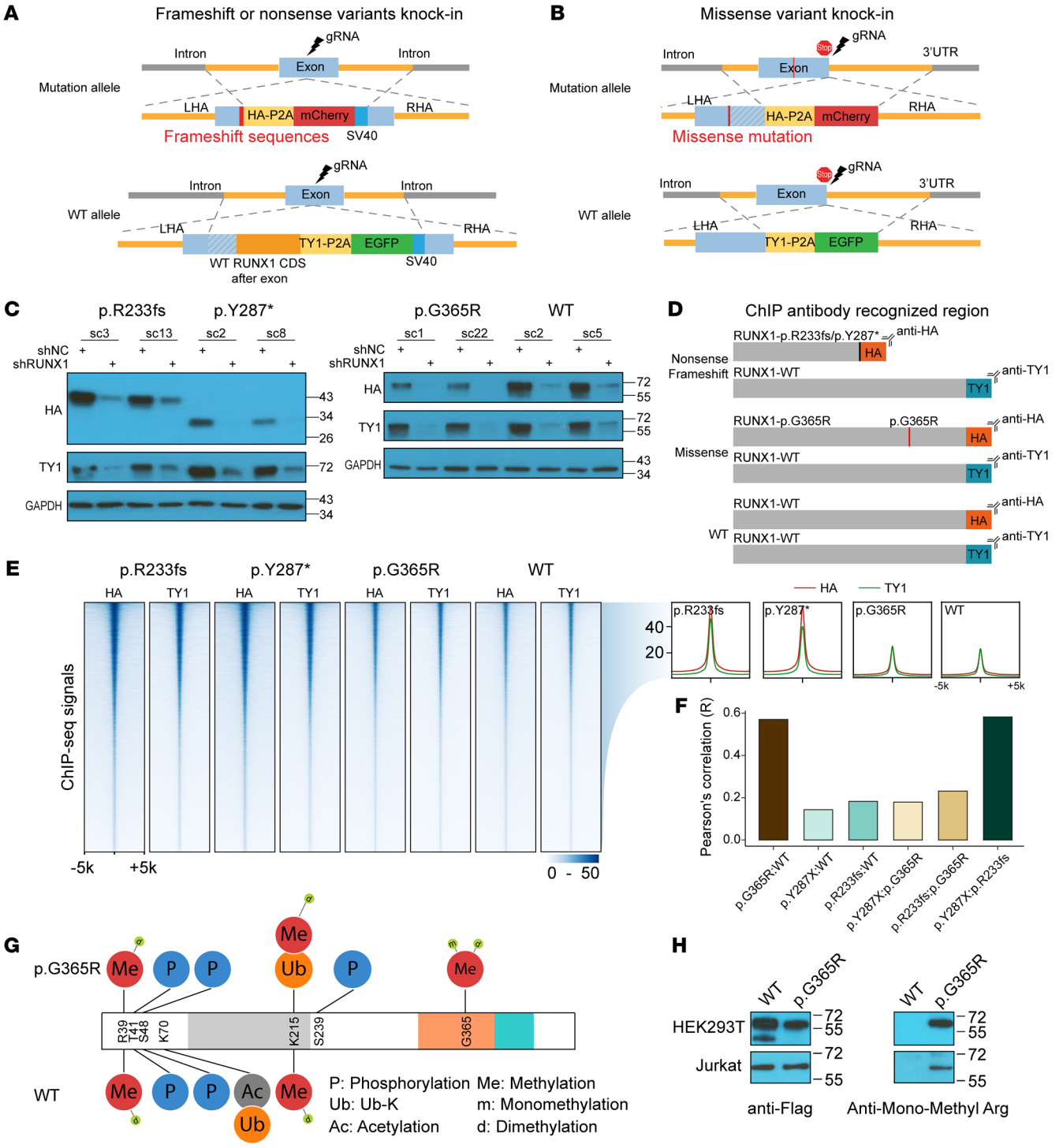


Figure 5. *RUNX1* variants are associated with distinctive DNA-binding patterns and altered posttranslational modifications. (**A** and **B**) Schematic representation of Jurkat cell models for RUNX1 ChIP-Seq profiling. p.R233fs, p.Y287*, and p.G365R were knocked in using CRISPR-Cas9 at the endogenous locus in a heterozygous fashion. TY1 and HA epitopes were inserted in the coding sequence of WT and variant RUNX1, respectively. This design enables ChIP assay of each protein simultaneously using 2 different antibodies. Detailed information can be found in Supplemental Figures 7–9. Sequence of the HDR templates is provided in Supplemental Methods. (**C**) Immunoblot confirmed that the expression of HA- and TY1-tagged RUNX1 and the identity of these bands were verified using *RUNX1* shRNA. (**D**) Schematic representation of ChIP-Seq strategy. To control for the difference in ChIP efficiency using HA versus TY1 antibodies, we also included Jurkat cells with each copy of the WT RUNX1 tagged with HA and TY1, respectively. (**E**) Heatmap shows the ChIP-Seq signals of variants and WT RUNX1 in Jurkat cells (left panel). Each row represents a genomic locus centered around the summit of the ChIP-Seq peak, with the color indicating the signal intensity. The aggregation plot (right panel) showed the average of ChIP-Seq signal intensity for all peaks identified with the HA-tagged (red line) versus TY1-tagged RUNX1 proteins (green line). (**F**) Across all ChIP-Seq peaks identified, we derived HA/TY1 signal ratios to quantify variant RUNX1-binding normalized against WT proteins. Comparing this ratio at each peak between different variants and also with WT, we calculated Pearson's correlation of each pair to indicate the overall similarity in the RUNX1-binding pattern. (**G**) Immunoprecipitation–mass spectrometry assays identified p.G365R-specific arginine methylation at this residue in HEK293T cells. (**H**) p.G365R methylation was confirmed by immunoblotting using an anti–mono-methyl arginine antibody in both HEK293T and Jurkat cells.

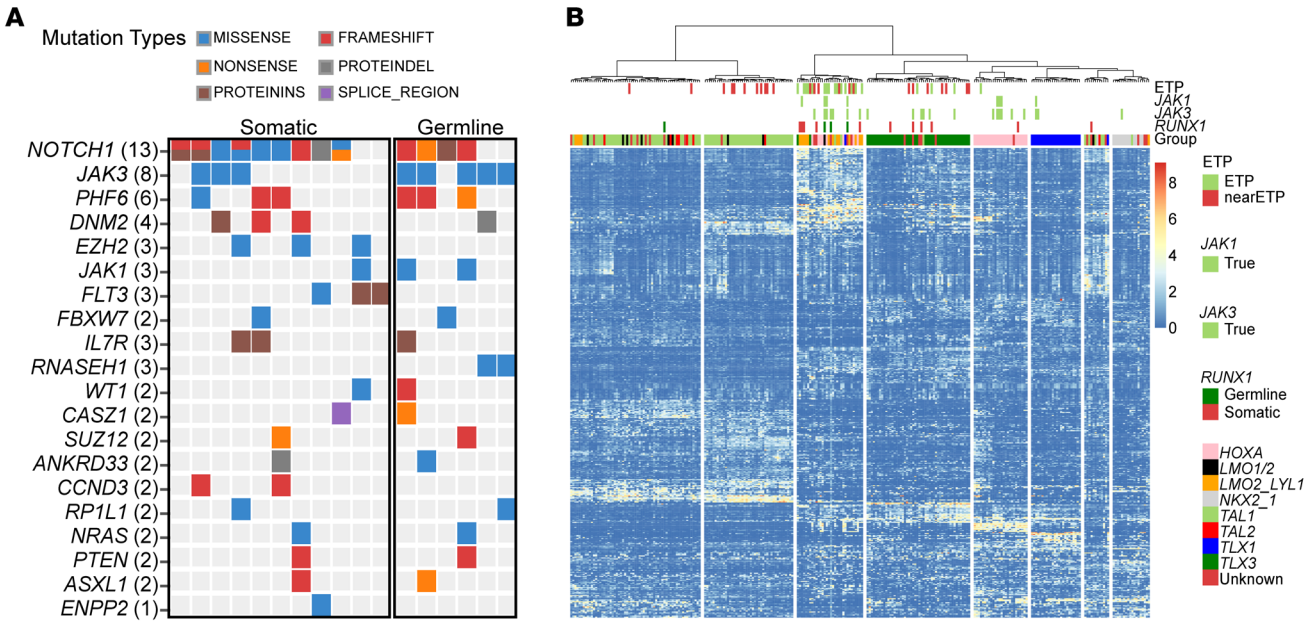


Figure 6. Somatic *JAK3* mutations cooccur in T-ALL with germline deleterious *RUNX1* variants. (**A**) Somatic *JAK3* mutations were significantly enriched in T-ALL patients with germline *RUNX1* variants. Whole-genome sequencing of remission samples for 17 T-ALL patients, 6 and 11 with germline variants or somatic mutations in *RUNX1*, respectively. (**B**) RNA-Seq was analyzed for 267 T-ALL patients, including 252, 4, and 11 subjects with WT RUNX1 carrying germline variants or somatic mutations in this gene. Unsupervised clustering shows that RUNX1-variant patients, either germline or somatic, clustered tightly with T-ALL with ETP and near-ETP immunophenotypes.

Mutation signature analyses showed highly significant enrichment of the SBS1, SBS5, SBS8, SBS9, and SBS18 patterns (COSMIC Mutational Signatures, version 3; synapse.org ID syn12009743; ref. 31 and Supplemental Figure 14) in *RUNX1* variant T-ALL samples. This is of interest because SBS5 was previously associated with the process of hematopoietic cell divisions and one of the main contributors to mutagenesis during T-ALL evolution (32).

We also performed RNA-Seq of T-ALL with germline *RUNX1* variants and compared the expression profile with patients with germline *RUNX1* variants (p.A142fs, p.R233fs, p.Y287*, and p.G365R), somatic *RUNX1* mutations, or WT *RUNX1* (n = 4, 11, and 252, respectively). Based on hierarchical clustering of global-expression profiles, *RUNX1*-variant patients (either germline or somatic) consistently clustered with T-ALL with ETP or near-ETP patients (Figure 6B and Supplemental Table 9). These results are consistent with previous reports of the preponderance of *RUNX1* variants in ETP T-ALL (33).

RUNX1 and JAK3 mutation induced ETP T-ALL in mice. To model *RUNX1*-related T-ALL leukemogenesis, especially in conjunction with somatic *JAK3* mutation, we introduced different combinations of *RUNX1* and *JAK3* mutations ($RUNX1^{R233fs}$ and $JAK3^{M511I}$) into mouse hematopoietic progenitor cells ($Lin^-Sca\text{-}1^+C\text{-}Kit^+$) and monitored leukemia development in vivo after transplantation. We hereafter refer to recipient mice with LSK cells transduced with EV, $RUNX1^{R233fs}$, $JAK3^{M511I}$, and $JAK3^{M511I}/RUNX1^{R233fs}$ as control, $RUNX1^M$, $JAK3^M$, and $JAK3^MRUNX1^M$ mice, respectively. At 4 months, peripheral leukocyte counts of $JAK3^M$ and $JAK3^MRUNX1^M$ mice ($41.78 \pm 44.3 \times 10^3$ cells/μL and $14.93 \pm 3.42 \times 10^3$ cells/μL, respectively) were significantly higher than those for control mice ($8.84 \pm 2.00 \times 10^3$ cells/μL), and the lowest peripheral leukocyte counts were seen in $RUNX1^M$ mice ($6.10 \pm 2.03 \times 10^3$ cells/μL; Figure 7A). Flow cytometry analysis at this time point showed a significant increase of $CD8^+$ T cells in $JAK3^M$ mice (Figure 7, B and C) compared with control mice. In contrast, $JAK3^MRUNX1^M$ mice showed an increase in $Mac1^+$ population and lower T cell population, suggesting an outgrowth of cells with ETP immunophenotype (Figure 7, B and C).

At 6 to 10 months after transplantation, both $JAK3^MRUNX1^M$ and $JAK3^M$ mice developed overt leukemia presenting with leukocytosis and splenomegaly, with a penetrance of 66.7% and 100%, respectively (Figure 7, D–F, and Supplemental Figure 15). The thymus of $JAK3^MRUNX1^M$ mice showed a significant increase of $CD4^-CD8^-$ (DN) T cells, particularly DN1 cells, as compared with that found in $JAK3^M$ mice (Figure 7, G and H). Circulating leukemic cells of $JAK3^MRUNX1^M$ mice showed a markedly higher $Mac1^+$ population, but lower lymphoid surface marker as compared with those of $JAK3^M$ mice (Figure 7I). Also, flow analysis of spleen and bone marrow showed a leukemia immunophenotype similar to that seen in peripheral blood (Figure 7I). These results indicate that $JAK3^MRUNX1^M$ induced the ETP T-ALL phenotype in vivo. There was also a trend for higher $Mac1^+$ cells with a lower level of $CD3^+$ cells in the peripheral blood of $RUNX1^M$ mice, but these mice never developed leukemia within this time frame (Figure 7A).

Discussion

RUNX1 plays important roles in definitive hematopoiesis by regulating the differentiation of myeloid, megakaryocyte, and lymphoid lineages. In this study, we comprehensively investigated *RUNX1* variation in ALL patients and identified highly deleterious germline variants in T-ALL, most of which were frameshift

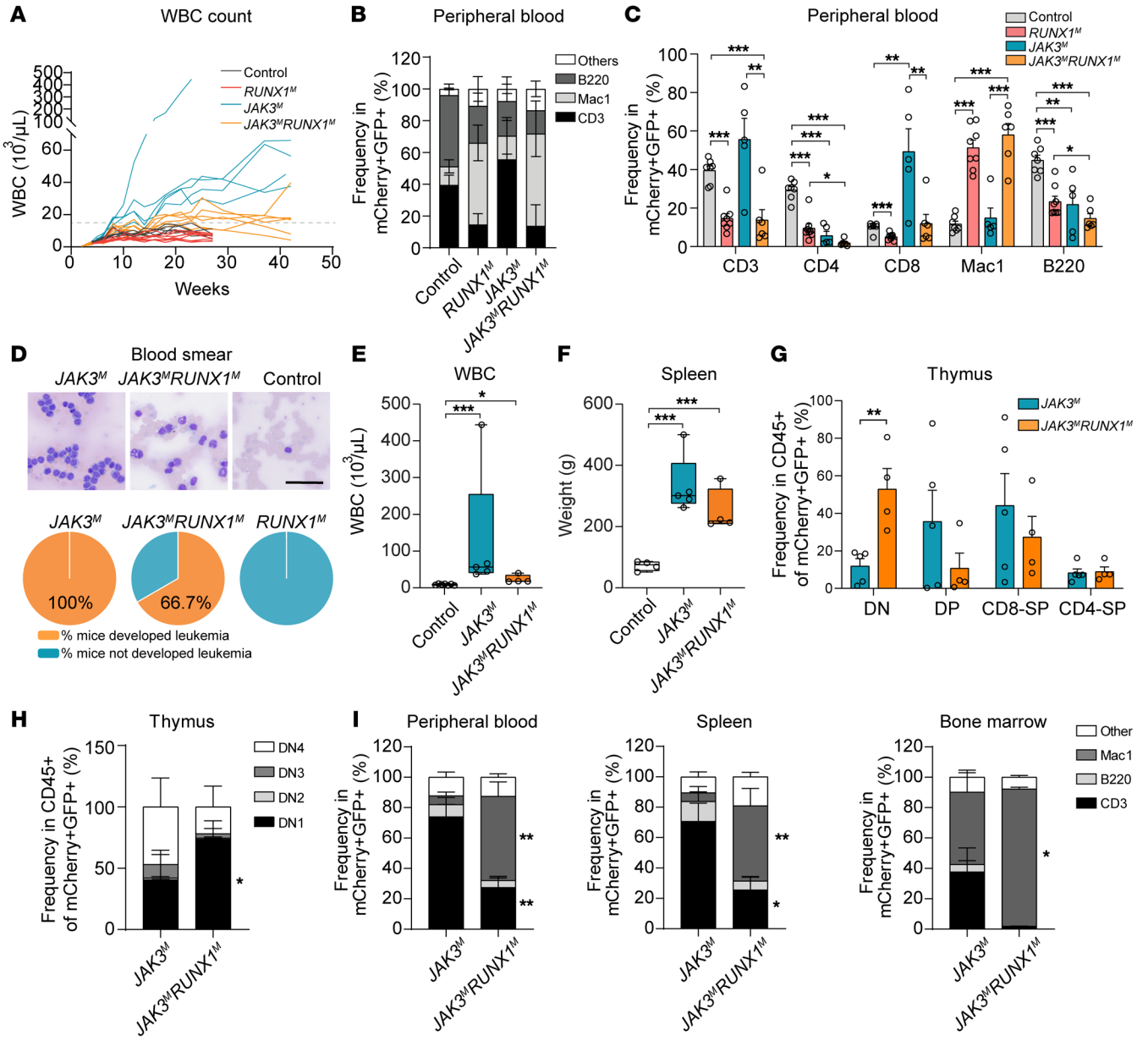


Figure 7. *RUNX1* variation and *JAK3* mutation jointly drive ETP phenotype in murine bone marrow transplantation model. (**A**) Peripheral leukocyte count of *RUNX1*^{M}, *JAK3*^{M}, *JAK3*^{M}*RUNX1*^{M}, and control mice. Mouse hematopoietic stem and progenitor cells were lentivirally transduced with *RUNX1*^{M} and *JAK3*^{M} constructs or EV and injected into recipient 8-week-old female mice. Peripheral blood count was monitored biweekly. (**B** and **C**) There was a significant increase in the CD8$^+$ population in *JAK3*^{M} mice compared with control mice. *RUNX1*^{M} and *JAK3*^{M}*RUNX1*^{M} mice showed increases of Mac1$^+$ cells and fewer CD3$^+$ cells than control or *JAK3*^{M} mice at 4 months. *C*ontrol, $n = 7$; *RUNX1*^{M}, $n = 8$; *JAK3*^{M}, $n = 5$; *JAK3*^{M}*RUNX1*^{M}, $n = 6$. (**D**) Upper panel: blood smear of *JAK3*^{M} and *JAK3*^{M}*RUNX1*^{M} mice at the time of sacrifice and control mice at 4 months. Scale bar: 50 μm. Lower panel: percentages of mice that developed leukemia in each group. *JAK3*^{M}, 100%; *JAK3*^{M}*RUNX1*^{M}, 66.7%. (**E**) Peripheral leukocyte count of *JAK3*^{M}*RUNX1*^{M} ($n = 4$) and *JAK3*^{M} ($n = 5$) mice at time of sacrifice and control ($n = 7$) mice after 4 months of transplantation. (**F**) Spleen weight of *JAK3*^{M} ($n = 5$) and *JAK3*^{M}*RUNX1*^{M} ($n = 4$) mice at time of sacrifice and control mice ($n = 4$) after 4 months of transplantation. (**G** and **H**) Thymocyte immunophenotype of *JAK3*^{M} ($n = 5$) and *JAK3*^{M}*RUNX1*^{M} ($n = 4$) mice at time of sacrifice. Coexpression of *RUNX1*^{M} and *JAK3*^{M} resulted in a drastic increase in DN1 population compared with that in mice receiving LSK cells expressing JAK3^M only. (**I**) In peripheral blood ($n = 3$), bone marrow ($n = 4$), and spleen ($n = 4$), *JAK3*^{M}*RUNX1*^{M} mice showed a significant increase in Mac1$^+$ population and a reduction of the CD3$^+$ population compared with *JAK3*^{M} ($n = 5$) mice. For **E** and **F**, *P* values were estimated by using Dunnett's test. For **B**, **C**, **G**, **H**, and **I**, data represent mean ± SEM and *P* values were generated by *t* test. $^*P < 0.05$; $^{**}P < 0.01$; $^{***}P < 0.001$.

or nonsense variations. By multilayer functional experiments and comprehensive epigenomic and genomic profiling analyses, we systematically characterized *RUNX1* variant function and identified *JAK3* mutations as the predominant cooperating somatic lesions in T-ALL. Furthermore, *RUNX1* variant, in conjunction with mutant *JAK3*, directly gave rise to ETP T-ALL in vivo. These findings advance our understanding of the role of *RUNX1* in the predisposition to childhood ALL.

RUNX1 is one of the most frequent target genes of chromosomal translocation, mutation, and copy number alteration in different

hematopoietic diseases and leukemia. *RUNX1* germline variants are associated with FPDMM (OMIM #601399; refs. 15, 16, 18, 34). Although most patients with FPD progress to myeloid malignancies, ALL has been reported in a minority of patients (16, 17). In MDS and AML with germline *RUNX1* variants, somatic *RUNX1* mutations are the most common acquired genomic alterations, suggesting that they are key cooperating events for leukemia progression (22, 35). Other studies identified somatic mutations in *CDC25C*, *GATA2*, *BCOR*, *PHF6*, *JAK2*, *DNMT3A*, *TET*, and *ASXL1*, albeit with lower frequencies. (22, 35, 36). In contrast, we identified *JAK3* mutations as the predominant cooccurring event with *RUNX1* germline variants in T-ALL, which consistently drove an ETP phenotype in patients and in mouse models of T-ALL. Therefore, we postulate that, while germline *RUNX1* variants disrupt normal hematopoiesis and generally increase the risk of leukemia, the lineage specification of these hematological malignancies is mostly dictated by secondary mutations acquired later in life.

Activating *JAK3* mutations have been reported in T-ALL (33). In vivo studies using a murine bone marrow transplantation model showed that *JAK3* mutations in the pseudo-kinase domain caused T cell lymphoproliferative disease that progressed to T-ALL, mainly by increasing the $CD8^+$ cell population (37, 38). This is in line with our observation that $JAK3^M$ mice exhibited a marked accumulation of $CD8^+$ cells in thymus, peripheral blood, spleen, and bone marrow. However, $JAK3^M RUNX1^M$ mice developed lymphoid leukemia with a distinctive phenotype that recapitulated human ETP T-ALL features and is similar to that of previously reported mouse models (e.g., circulating leukemic cells expressed the myeloid cell marker Mac1, but not the lymphoid markers CD8/CD3; refs. 39, 40). Also, the $CD4^-CD8^-$ (DN), especially the DN1 population, was particularly enriched in thymocytes from $JAK3^M$-$RUNX1^M$ mice as compared with $JAK3^M$ mice. Alongside genomic findings in T-ALL patients, these in vivo experiments indicate that *RUNX1* dominant-negative variants plus *JAK3*-activating mutations most likely result in the ETP T-ALL. The segregation of *RUNX1* variants with ETP T-ALL is of interest because historically this subtype is associated with inferior treatment outcomes (41). However, the survival gap between ETP and other types of T-ALL is no longer significant with contemporary treatment regimens. For these reasons, children with ETP T-ALL are not routinely taken to allogeneic transplantation. Currently, there are insufficient data to inform the treatment of patients with ETP T-ALL who harbor germline *RUNX1* variations. In theory, these children remain at risk for persistent thrombocytopenia and development of secondary AML and/or MDS. Allogeneic hematopoietic transplantation using a *RUNX1* WT donor could potentially cure both of these problems. That said, decisions regarding allogeneic transplantation are complex and must take into account the condition and age of the patient, donor availability, and preferences of the patient, treating physicians, and transplant specialists.

A recent study by Brown et al. comprehensively described the genomic landscape of *RUNX1*-related FPD and myeloid malignancy in 130 families (22). In this cohort, missense and truncating germline *RUNX1* variants were equally represented. While truncating variants occurred in both RHD and TAD, missense variants were largely restricted to the former. This pattern is significantly different from that in the lymphoid malignancies as described here. In T-ALL, deleterious *RUNX1* variants were predominantly nonsense or frameshift and the only missense variant resided in the activation domain. Unfortunately, we do not have family history for children with T-ALL carrying *RUNX1* germline variants and therefore cannot ascertain the exact penetrance of leukemia or FPD. However, given the profound effects on RUNX1 activity and a range of phenotypes in vitro and in vivo, these variants are defined as likely pathogenic according to the American College of Medical Genetics (ACMG) classification system (Table 1 and Supplemental Table 10). In fact, the p.Y287* variant seen in our T-ALL cohort has been previously linked to FPD, and functional characterization indicated that this variant causes defective megakaryocyte differentiation in the induced pluripotent stem cell (iPSC) model (42). In B-ALL, almost all variants were missense and localized outside of the RHD domain. Although these variants showed little effect on RUNX1 transcriptional activity level, some of them were predicted to be damaging variants by REVEL (ref. 43 and Table 1). More comprehensive functional assays might be needed to definitively determine the effects of these variants. Finally, we found that RUNX1 variants dramatically alter the colony formation of BFU-E and CFU-GM in a pattern similar to that seen with *GATA2* mutation (44). *GATA2* (or *GATA1*) expression was not influenced by *RUNX1* variants in our human $CD34^+$ cell models (data not shown). However, GATA2 can colocalize on chromatin with RUNX1 (45), and these factors plausibly cooperate with each other to regulate hematopoietic differentiation.

Genome-wide patterns of RUNX1 binding have been investigated extensively using ChIP-Seq assays (27, 46, 47), but there are a paucity of studies directly examining target genes of variant RUNX1. When this was attempted in the past, variant RUNX1 was either ectopically expressed in iPSC or cord blood $CD34^+$ cells, raising the possibility of false positives due to artificially high levels of RUNX1 (48). This is also hindered by the lack of antibodies that specifically recognize WT but not variant RUNX1. To overcome these issues, we engineered Jurkat cells with heterozygous knockin of *RUNX1* variants (p.R233fs/WT, p.Y287*/WT, and p.G365R/WT) using the CHASE-KI method (49). In this model, we also introduced the HA and TY1 epitope tags at the 3′ end of the coding exon on the variant and WT *RUNX1* allele, respectively. Our model recapitulated *RUNX1* variant status in patients with T-ALL and enabled us to profile variant or WT RUNX1 binding using different antibodies. Moreover, our ChIP-Seq result indicated that the C-terminal truncating variants p.R233fs and p.Y287* indeed exhibited a distinct binding pattern compared with the full-length p.G365R variant and WT RUNX1.

Although our studies comprehensively characterized *RUNX1* variant functions using a variety of orthogonal methods, our model systems and assays were not without limitations. For example, ectopic expression of the *RUNX1* variant can artificially increase gene activity that is not seen under physiological conditions, and this can be problematic for studying loss-of-function variants when endogenous *RUNX1* is also present. We tried to mitigate these issues by including WT *RUNX1* as control, and our assays indeed confirmed the effects of known pathogenic variants (Figure 2, A, B, and E; p.R204Q). Nonhematopoietic cells such as HEK293T and HeLa have been used extensively to study RUNX1 function (24, 50, 51), but one could easily argue against their tissue

relevance in the context of RUNX1-related leukemia, which is why we performed extensive experiments to confirm *RUNX1* function in T-ALL cell lines and also in human cord blood $CD34^+$ cells. Recent development of gene-editing techniques in iPSCs provides an exciting system for this type of work (42, 52). For example, CRISPR-Cas9 gRNA–mediated homology-directed repair (HDR) or base editing in the iPSC system can precisely introduce the *RUNX1* variant at the endogenous locus and in the presence of the WT allele, and the hematopoietic differentiation potential of these engineered progenitor cells can be directly characterized in vitro.

In summary, we comprehensively describe *RUNX1* germline variants in childhood ALL. Using multiple functional assays, we identified highly deleterious germline variants in T-ALL and their biochemical and cellular effects. In addition, we characterized somatic genomic alterations associated with *RUNX1* germline variations in T-ALL, illustrating the interplay between acquired and inherited genetic variants in the context of leukemia pathogenesis.

Methods

Patients. A total of 6190 ALL patients were included for *RUNX1*-targeted sequencing: 4132 children with newly diagnosed B-ALL enrolled in the COG AALL0232 ($n = 2224$), P9904/5/6 ($n = 1634$), and AALL0331 ($n = 274$) protocols; 704 children with newly diagnosed B-ALL enrolled in the St. Jude Total XIII and XV protocols; 1231 children with newly diagnosed T-ALL enrolled in the COG AALL0434 protocols; and 123 children with newly diagnosed T-ALL enrolled in the St. Jude Total XIII and XV protocols (Figure 1A and refs. 53–56). Family histories were not available for patients in the COG studies. Peripheral blood or bone marrow from children with ALL during remission was collected as a source of genomic DNA. Due to the leukemia-free status of these samples, we consider them as "germline like," and the variants identified in these samples are provisionally of germline origin.

For targeted *RUNX1* sequencing in the ALL cohort, Illumina dual-indexed libraries were created from the germline DNA of 6190 children, and pooled in sets of 96 before hybridization with customized Roche NimbleGene SeqCap EZ probes (Roche NimbleGen) to capture the *RUNX1* genomic region. Quantitative PCR was used to define the appropriate capture product titer necessary to efficiently populate an Illumina HiSeq 2000 flow cell for paired end 2×100 bp sequencing. Coverage of at least 20-fold depth was achieved across the targeted *RUNX1* locus for 99.2% of samples. Sequence reads in FASTQ format were mapped and aligned using the Burrows-Wheeler Aligner (BWA) (57, 58), and genetic variants were called using the GATK pipeline (version 3.1; ref. 58), as previously described, and annotated using the ANNOVAR program (59) with the annotation databases including GRCh37/hg19, GRCh38/hg38 RefSeq (60), and REVEL (43). Variant classification was done following ACMG guidelines (61). Noncoding, and synonymous coding variants were excluded from further consideration for this study.

Genomic analysis of patient samples. Whole-genome sequencing and RNA-Seq were performed for T-ALL patients with germline RUNX1 variants whenever available samples were identified. Whole-genome sequencing was done for matched germline and leukemia samples, whereas RNA-Seq was done only for leukemia samples. Briefly, DNA was purified using the QIAamp DNA Blood Mini Kit (QIAGEN, 51104), and RNA was purified using the RiboPure RNA Purification Kit (Thermo Fisher Scientific, AM1928). DNA (250–1000 ng) and RNA (500–1000 ng) were sent to the St. Jude Hartwell Center for sequencing.

Details for functional experiments, leukemia modeling in mouse, genomic analyses, other experiments, and further details regarding Methods can be found in Supplemental Methods.

Data access. ChIP-Seq and RNA-Seq data were deposited in the NCBI's Gene Expression Omnibus database (GEO GSE178239). Whole-genome sequencing data were deposited in the European Genome-Phenome Archive (EGAS00001005403).

Statistics. Statistical analyses were performed using Student's *t* test, Dunnett's test, G-test, or Fisher's exact test. Multiple comparisons were accounted for using Dunnett's test for analysis when applicable. The choice of statistical test was based on data distribution, as noted in the figure legends. All tests were 2-tailed. Statistical significance was defined as $P < 0.05$.

Study approval. Human subject research was approved by Institutional Review Boards at St. Jude– and COG-affiliated institutions. Informed consent was obtained from parents, guardians, or patients, and assent was obtained from patients, as appropriate. All animal experiments were conducted according to the protocols approved by the St. Jude Institutional Animal Care and Use Committee.

Author contributions

JJY initiated and led the project. JJY and YL designed the functional experiments. YL, XZ, and RZ performed in vitro experiments. CL and YL designed the CRISPR-Cas9 experiments. YL performed in vivo experiments. Wentao Yang analyzed the genomic data. CK performed ACMG annotation. MD, SSW, KPD, JMGF, EAR, WLC, KRR, CGM, WEE, CHP, SPH, DTT, MVR, and MLL contributed to data gathering. JJY, Wenjian Yang, CS, CHP, SPH, KEN, and CGM interpreted data. PPL, TS, and KEN provided relevant intellectual input and edited the manuscript. MQ performed RUNX1-targeted sequencing analysis when working at St Jude Children's Research Hospital. JJY and YL wrote the manuscript. All authors reviewed and commented on the manuscript.

Acknowledgments

We thank the Hartwell Center for Bioinformatics and Biotechnology, the Flow Cytometry and Cell Sorting Core, the Cytogenetic Shared Resource, and the Animal Research Center of St. Jude Children's Research Hospital for their technical support in performing experiments included in this study. This work was supported in part by P50GM115279 (to MVR, WEE, CGM, and JJY), R01CA241452 (KEN and JJY), and P30CA21765 (CHP, MVR, JJY, WEE, KEN, and CGM), and the American Lebanese Syrian Associated Charities.

Address correspondence to: Jun J. Yang, 262 Danny Thomas Place, MS313, Memphis, Tennessee, 38105, USA. Phone: 901.595.2517; Email: jun.yang@stjude.org.

1. Iacobucci I, Mullighan CG. Genetic basis of acute lymphoblastic leukemia. *J Clin Oncol*. 2017;35(9):975–983.
2. Churchman ML, et al. Germline genetic IKZF1 variation and predisposition to childhood acute lymphoblastic leukemia. *Cancer Cell*. 2018;33(5):937–948.
3. Trevino LR, et al. Germline genomic variants associated with childhood acute lymphoblastic leukemia. *Nat Genet*. 2009;41(9):1001–1005.
4. Xu H, et al. Inherited coding variants at the CDKN2A locus influence susceptibility to acute lym-

phoblastic leukaemia in children. *Nat Commun*. 2015;6:7553.
5. Perez-Andreu V, et al. Inherited GATA3 variants are associated with Ph-like childhood acute lymphoblastic leukemia and risk of relapse. *Nat Genet*. 2013;45(12):1494–1498.
6. Perez-Andreu V, et al. A genome-wide association study of susceptibility to acute lymphoblastic leukemia in adolescents and young adults. *Blood*. 2015;125(4):680–686.
7. Papaemmanuil E, et al. Loci on 7p12.2, 10q21.2 and 14q11.2 are associated with risk of childhood acute lymphoblastic leukemia. *Nat Genet*. 2009;41(9):1006–1010.
8. Migliorini G, et al. Variation at 10p12.2 and 10p14 influences risk of childhood B-cell acute lymphoblastic leukemia and phenotype. *Blood*. 2013;122(19):3298–3307.
9. Lee SHR, et al. Genome-wide association study of susceptibility loci for TCF3-PBX1 acute lymphoblastic leukemia in children [published online September 3, 2020]. *J Natl Cancer Inst*. https://doi.org/10.1093/jnci/djaa133.
10. Qian M, et al. TP53 germline variations influence the predisposition and prognosis of B-cell acute lymphoblastic leukemia in children. *J Clin Oncol*. 2018;36(6):591–599.
11. Moriyama T, et al. Germline genetic variation in ETV6 and risk of childhood acute lymphoblastic leukaemia: a systematic genetic study. *Lancet Oncol*. 2015;16(16):1659–1666.
12. Wang Q, et al. Disruption of the Cbfa2 gene causes necrosis and hemorrhaging in the central nervous system and blocks definitive hematopoiesis. *Proc Natl Acad Sci U S A*. 1996;93(8):3444–3449.
13. Huang G, et al. Dimerization with PEBP2beta protects RUNX1/AML1 from ubiquitin-proteasome-mediated degradation. *EMBO J*. 2001;20(4):723–733.
14. Kanno T, et al. Intrinsic transcriptional activation-inhibition domains of the polyomavirus enhancer binding protein 2/core binding factor alpha subunit revealed in the presence of the beta subunit. *Mol Cell Biol*. 1998;18(5):2444–2454.
15. Owen CJ, et al. Five new pedigrees with inherited RUNX1 mutations causing familial platelet disorder with propensity to myeloid malignancy. *Blood*. 2008;112(12):4639–4645.
16. Preudhomme C, et al. High frequency of RUNX1 biallelic alteration in acute myeloid leukemia secondary to familial platelet disorder. *Blood*. 2009;113(22):5583–5587.
17. Nishimoto N, et al. T cell acute lymphoblastic leukemia arising from familial platelet disorder. *Int J Hematol*. 2010;92(1):194–197.
18. Song WJ, et al. Haploinsufficiency of CBFA2 causes familial thrombocytopenia with propensity to develop acute myelogenous leukaemia. *Nat Genet*. 1999;23(2):166–175.
19. Grossmann V, et al. Prognostic relevance of RUNX1 mutations in T-cell acute lymphoblastic leukemia. *Haematologica*. 2011;96(12):1874–1877.
20. Della Gatta G, et al. Reverse engineering of TLX oncogenic transcriptional networks identifies RUNX1 as tumor suppressor in T-ALL. *Nat Med*. 2012;18(3):436–440.
21. Quentin S, et al. Myelodysplasia and leukemia of Fanconi anemia are associated with a specific pattern of genomic abnormalities that includes cryptic RUNX1/AML1 lesions. *Blood*. 2011;117(15):e161–e170.
22. Brown AL, et al. RUNX1-mutated families show phenotype heterogeneity and a somatic mutation profile unique to germline predisposed AML. *Blood Adv*. 2020;4(6):1131–1144.
23. Huang G, et al. PU.1 is a major downstream target of AML1 (RUNX1) in adult mouse hematopoiesis. *Nat Genet*. 2008;40(1):51–60.
24. Harada H, et al. Implications of somatic mutations in the AML1 gene in radiation-associated and therapy-related myelodysplastic syndrome/acute myeloid leukemia. *Blood*. 2003;101(2):673–680.
25. Osato M, et al. Biallelic and heterozygous point mutations in the runt domain of the AML1/PEBP2alphaB gene associated with myeloblastic leukemias. *Blood*. 1999;93(6):1817–1824.
26. Matreyek KA, et al. A platform for functional assessment of large variant libraries in mammalian cells. *Nucleic Acids Res*. 2017;45(11):e102.
27. Sanda T, et al. Core transcriptional regulatory circuit controlled by the TAL1 complex in human T cell acute lymphoblastic leukemia. *Cancer Cell*. 2012;22(2):209–221.
28. Beck D, et al. Genome-wide analysis of transcriptional regulators in human HSPCs reveals a densely interconnected network of coding and noncoding genes. *Blood*. 2013;122(14):e12–e22.
29. Cuddapah S, et al. Global analysis of the insulator binding protein CTCF in chromatin barrier regions reveals demarcation of active and repressive domains. *Genome Res*. 2009;19(1):24–32.
30. Liu Y, et al. The genomic landscape of pediatric and young adult T-lineage acute lymphoblastic leukemia. *Nat Genet*. 2017;49(8):1211–1218.
31. Bamford S, et al. The COSMIC (Catalogue of Somatic Mutations in Cancer) database and website. *Br J Cancer*. 2004;91(2):355–358.
32. Sentis I, et al. The evolution of relapse of adult T cell acute lymphoblastic leukemia. *Genome Biol*. 2020;21(1):284.
33. Zhang J, et al. The genetic basis of early T-cell precursor acute lymphoblastic leukaemia. *Nature*. 2012;481(7380):157–163.
34. Michaud J, et al. In vitro analyses of known and novel RUNX1/AML1 mutations in dominant familial platelet disorder with predisposition to acute myelogenous leukemia: implications for mechanisms of pathogenesis. *Blood*. 2002;99(4):1364–1372.
35. Antony-Debre I, et al. Somatic mutations associated with leukemic progression of familial platelet disorder with predisposition to acute myeloid leukemia. *Leukemia*. 2016;30(4):999–1002.
36. Yoshimi A, et al. Recurrent CDC25C mutations drive malignant transformation in FPD/AML. *Nat Commun*. 2014;5:4770.
37. Degryse S, et al. JAK3 mutants transform hematopoietic cells through JAK1 activation, causing T-cell acute lymphoblastic leukemia in a mouse model. *Blood*. 2014;124(20):3092–3100.
38. Cornejo MG, et al. Constitutive JAK3 activation induces lymphoproliferative syndromes in murine bone marrow transplantation models. *Blood*. 2009;113(12):2746–2754.
39. Booth CAG, et al. Ezh2 and Runx1 mutations collaborate to initiate lympho-myeloid leukemia in early thymic progenitors. *Cancer Cell*. 2018;33(2):274–291.
40. Wang C, et al. Ezh2 loss propagates hypermethylation at T cell differentiation-regulating genes to promote leukemic transformation. *J Clin Invest*. 2018;128(9):3872–3886.
41. Teachey DT, O'Connor D. How I treat newly diagnosed T-cell acute lymphoblastic leukemia and T-cell lymphoblastic lymphoma in children. *Blood*. 2020;135(3):159–166.
42. Connelly JP, et al. Targeted correction of RUNX1 mutation in FPD patient-specific induced pluripotent stem cells rescues megakaryopoietic defects. *Blood*. 2014;124(12):1926–1930.
43. Ioannidis NM, et al. REVEL: an ensemble method for predicting the pathogenicity of rare missense variants. *Am J Hum Genet*. 2016;99(4):877–885.
44. Katsumura KR, et al. Human leukemia mutations corrupt but do not abrogate GATA-2 function. *Proc Natl Acad Sci U S A*. 2018;115(43):E10109–E10118.
45. Chacon D, et al. BloodChIP: a database of comparative genome-wide transcription factor binding profiles in human blood cells. *Nucleic Acids Res*. 2014;42(database issue):D172–D177.
46. Giambra V, et al. NOTCH1 promotes T cell leukemia-initiating activity by RUNX-mediated regulation of PKC-θ and reactive oxygen species. *Nat Med*. 2012;18(11):1693–1698.
47. Wilkinson AC, et al. RUNX1 is a key target in t(4;11) leukemias that contributes to gene activation through an AF4-MLL complex interaction. *Cell Rep*. 2013;3(1):116–127.
48. Gerritsen M, et al. RUNX1 mutations enhance self-renewal and block granulocytic differentiation in human in vitro models and primary AMLs. *Blood Adv*. 2019;3(3):320–332.
49. Hyle J, et al. Acute depletion of CTCF directly affects MYC regulation through loss of enhancer-promoter looping. *Nucleic Acids Res*. 2019;47(13):6699–6713.
50. Decker M, et al. Functional classification of RUNX1 variants in familial platelet disorder with associated myeloid malignancies [published online March 21, 2021]. *Leukemia*. https://doi.org/10.1038/s41375-021-01200-w.
51. Imai Y, et al. Mutations of the AML1 gene in myelodysplastic syndrome and their functional implications in leukemogenesis. *Blood*. 2000;96(9):3154–3160.
52. Estevez B, et al. RUNX1 haploinsufficiency causes a marked deficiency of megakaryocyte-biased hematopoietic progenitor cells. *Blood*. 2021;137(19):2662–2675.
53. Larsen EC, et al. Dexamethasone and high-dose methotrexate improve outcome for children and young adults with high-risk B-acute lymphoblastic leukemia: a report from Children's Oncology Group study AALL0232. *J Clin Oncol*. 2016;34(20):2380–2388.
54. Bowman WP, et al. Augmented therapy improves outcome for pediatric high risk acute lymphocytic leukemia: results of Children's Oncology Group trial P9906. *Pediatr Blood Cancer*. 2011;57(4):569–577.
55. Pui CH, et al. Improved outcome for children with acute lymphoblastic leukemia:

results of Total Therapy Study XIIIB at St Jude Children's Research Hospital. *Blood*. 2004;104(9):2690–2696.
56. Pui CH, et al. Rationale and design of Total Therapy Study XV for newly diagnosed childhood acute lymphoblastic leukemia. *Ann Hematol*. 2004;83 Suppl 1:S124–S126.
57. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*. 2009;25(14):1754–1760.
58. McKenna A, et al. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res*. 2010;20(9):1297–1303.
59. Wang K, et al. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res*. 2010;38(16):e164.
60. Pruitt KD, et al. RefSeq: an update on mammalian reference sequences. *Nucleic Acids Res*. 2014;42(D1):D756–D763.
61. Richards S, et al. Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. *Genet Med*. 2015;17(5):405–424.