Supplementary Materials

Supplementary Methods

Genetic association analyses

ARID5B variant genotype in ALL cases was determined by direct targeted sequencing as described in Methods. Because of the prevalence of childhood ALL in the general population is exceedingly low, we used subjects in the UK10K as non-ALL controls, and variants within the same genomic region on Chromosome 10 were identified using publicly available whole-genome sequencing data of this cohort. and we only used variants that passed quality control as implemented by UK10K[1].

For the genetic association analyses, we first estimated genetic ancestry (European, African, Amerindian, and East Asian) for each ALL case and also controls in the UK10K cohort, using genome-wide SNP genotype or whole-genome sequencing, respectively. This was done using STRUCTURE (version 2.3.4), with HapMap CEU, YRI, CHB/JPT and indigenous Amerindians as the reference populations, respectively, as we described previously[2-4]. Then we compared genotype frequency between 4,671 cases and 3,644 controls, with % genetic ancestry as covariate, using an additive logistic regression model in PLINK (version 1.9).

In addition, we also re-tested each variant in cases and controls of European descent (>95% European ancestry, N=2,629 and 3,592, respectively), to further minimize the effects of potential population stratification.

Cell culture

Nalm6, REH, and CEM cell lines were purchased from ATCC. GM12878 cell line was purchased from the NIGMS Human Genetic Cell Repository (Coriell Institute). All cells (except for

GM12878) are grown in culture in RPMI 1640 medium (GIBCO, Life Technologies) supplemented with 10% heat-inactivated FBS and 2 mM L-glutamine at 37°C in 5% CO_2. GM12878 is grown in culture in RPMI 1640 medium (GIBCO, Life Technologies) supplemented with 20% heat-inactivated FBS and 2 mM L-glutamine at 37°C in 5% CO_2. Lenti-X cells (Takara #632180) were cultured with 90% Dulbecco's Modified Eagle's Medium (DMEM) with high glucose (4.5 g/L), 4 mM L-glutamine, and sodium bicarbonate (Sigma-Aldrich, D5796); 10% Fetal Bovine Serum (FBS); 100 units/mL penicillin G sodium, and 100 µg/mL streptomycin sulfate. Lenti-X cells were used to generate all the high-titer lentivirus used in this study. All cell lines were validated by STR three months ago.

ARID5B-mCherry reporter cell line

The all-in-one Lenti-Cas9-T2A-mCherry-sgRNA (V6) was derived from a previous study [5]. The sgRNA targeting stop codon of *ARID5B* was designed using online software (http://crispr.mit.edu/). A two-step cloning protocol was performed to generate the homology arm (HA) donor vectors. Primers were designed to amplify the 800bp 5' HA flanking the stop codon of *ARID5B*. Overhangs of 23 bp including the target sgRNA (GCACCCCAGTACAAAACTGT) and PAM sequences were included at the 5' and 3' end of the HA sequence. The HAs were amplified from Nalm6 cell genomic DNA (5'HR arm forward: GCACCCCAGTACAAAACTGTAGGCACAAACCTACCGGCAAGGTC, reverse: CAGTTTTGTACTGGGGTGCAC; 3'HR arm forward: GCTCAGCTCTGCCCAGCAGTC, reverse: CCTACAGTTTTGTACTGGGGTGCCAATAACTTTACCACCCAAAG). The P2A-mCherry was amplified from pEGFP-C1 (Clontech #632470). Snapgene software (Snapgene) was used to design all primers for in-fusion cloning. The PCR reactions were performed using CloneAmp polymerase (Clontech #639298). First, the amplified 5' HA for *ARID5B* knock-in was cloned into pCR-Blunt IITOPO using the Zero-Blunt cloning kit (Invitrogen # K2875J10), and the amplified 5' HA for *ARID5B* was cloned into pBluescript-SK by using the TA cloning kit

(Invitrogen #K204001). The DNA was purified from colonies and screened by Sanger sequencing with the primers of M13F and M13R. The pCR-Blunt II-TOPO or pBluescript-SK-sgRNA-PAM-5'HA was then linearized by unique restriction enzyme digestion and ligated with P2A-mCherry and the 3'HA-sgRNA-PAM through in-fusion cloning. Sanger sequencing was performed to ensure that the knock-in DNA was cloned in-frame to the HAs. The Lenti-dCas9-KRAB-blasticidin was purchased from Addgene (Plasmid #89567).

2.5 μg of the donor plasmid and 2.5 μg of the CRISPR/Cas9-*ARID5B*-C-terminus-sgRNA all-in-one plasmid were delivered into 5 million Nalm6 cells using the Nucleofector-2b device with Kit V (Lonza #VCA-1003) and program X-001. Twenty-four hours after transfection, cells were sorted for the mCherry fluorescent marker linked to Cas9 expression vector to enrich the transfected cell population. After the sorted cells were recovered after culture for up to 3 weeks, a second sort was performed to select cells for successful knock-in by sorting for mCherry positive cells. Two weeks later, a third sort was repeated for mCherry positive cells.

To confirm mCherry knock-in in Nalm6 cells, DNA from single-cell-derived clones was extracted with a DNeasy Blood & Tissue Kit (QIAGEN #69506). Two sets of primers were designed to recognize the 5′ and 3′ knock-in boundaries were used with the following PCR cycling conditions: 98°C for 2 mins, followed by 40 cycles of 98°C for 15 s and 72°C for 60 s. After electrophoresis, DNA fragments with expected size were purified and submitted for Sanger-sequencing. Validation primer for 5' knock-in: forward: CTTCCCACAGACACCAAGAAA, reverse: CACGTCTCCTGCTTGCTTTA; for 3' knock-in: forward: CGCCTACAACGTCAACATCA, reverse: AGTTGCGGATCTACAAAGGAAC.

ARID5B **intron 3 enhancer deletion using CRISPR-Cas9 editing**

Two sgRNAs (GTATATCATAGGTGCTCCGT; TAATGACTCTACGGGCACGT) were designed against the two ends of target region. In order to increase the deletion efficiency, we need to make sure that the two ends were cut at the same time. To this purpose, these two

sgRNAs were cloned into two Tet-on inducible sgRNA vectors: FgH1tUTG-GFP (Addgene #70183) and FgH1tUTG-CFP which was engineered from FgH1tUTG-GFP using QuikChange Lighting mutagenesis kit (Agilent #210518). Nalm6 cells with stable expression of Cas9 were infected with these two sgRNAs and sorted for GFP/CFP double positive cells, followed by doxycycline treatment (1 mg/mL). Genomic PCR (forward: AGTAGGGAGTTCTGCATATTGTT, reverse: TTGAGTTGAGTCCATTGGTAGAG) and Sanger sequencing were performed to confirm the deletion in single cell clones.

We performed fluorescence in situ hybridization to confirm enhancer deletion. An 800-bp purified P2A-mCherry DNA fragment was labeled with a red-dUTP (AF594, Molecular Probes) by nick translation and an *ARID5B* BAC clone (CH17-412I12/7p15.2) was labeled with a green-dUTP (AF488, Molecular Probes). Both labeled probes were combined with sheared human DNA and independently hybridized to fix the interphase and metaphase nuclei derived from each sample by using routine cytogenetic methods in a solution containing 50% formamide, 10% dextran sulfate, and 2 × SSC. The cells were then stained with DAPI and analyzed. One hundred interphase nuclei were scored for the pattern of hybridization (pairing of red and green signals) and for random P2A-mCherry integration signals.

ARID5B knock-out using CRISPR-Cas9 editing

CRISPR/Cas9 was used to knock out *MEF2C* in Nalm6 cells. sgRNA (GACAACGAGCCGCATGAGAGC) targeting *MEF2C* was designed using online software (http://crispr.mit.edu/). Expressions of *MEF2C* (forward: GACTGTGAGATTGCGCTGAT, reverse: TTCAATGCCTCCACGATGTC) and *ARID5B* (forward: AATCTTGTCCCTTGGCGACT, reverse: GACGGCGGGCTGTTATTGTTTCAT) were then confirmed by RT-PCR. S18 (forward: AGGAGCGATTTGCTGGTGTG, reverse: GCTACCAGGGCCTTTGAGAT) was used as an internal control.

Analyses of publicly available epigenomic profiling data

GM12878 ChIA-PET data were reanalyzed using UCSC browser.

Chromatin state was annotated using ChIP-seq data of H3K4me1, H3K4me3, and H3K27ac from GM12878 cells generated by ENCODE project (GSE170245)[6]. Chromatin accessibility (based on ATAC-seq) in 13 human primary blood cell types was generated from UCSC Genome Browser Track Hub, https://s3-us-west-1.amazonaws.com/chang-public-data/2016_NatGen_ATAC-AML/hub.txt. ATAC-seq data from ALL cell lines (Nalm6, REH, UOC-B1, and 697) are accessible through GEO series accession number: GSE129066[7]. ATAC-seq was performed following the Fast-ATAC protocol. Paired-end sequencing (2 × 100 bp) was performed using the Illumina HISeq 2500 platform. Paired-end reads were first applied on cutadapt (version 1.18)[8] for adaptor trimming and then mapped to the GRCh38 human genome reference by Bowtie2 (version 2.2.9)[9]. Peak calling was conducted by MACS[10] with default parameters on each sample. Peaks from all samples were merged by bedtools (version 2.25.0)[11] to retain non-overlapped regions which were used to generate differentially enriched peaks by ABSSeq under the aFold model with read count from HTSeq[12]. Regions with adjusted *P* values < 0.05 and log2 fold change >=2 were nominated. Enriched regions were then mapped to the nearest gene in GRCh38 by Homer[13]. Footprint analysis was performed using TOBIAS method to explore TF motif at intron 3 enhancer of *ARID5B*[14].

TF-enrichment analysis

TF-binding enrichment analysis was performed according to the method of Cowper-Sal. lari, et al[15]. Motif enrichment analysis for MEF2C ChIP-seq data was performed using Homer. TF-binding sites were filtered for those with a MACS peak Q-value >100.

Quantitative real-time PCR and Western blotting

Total RNA was collected by using RNeasy Miniprep Kit (QIAGEN #74106). cDNA synthesis was performed using SuperScript IV One-Step RT-PCR System (Thermo Fisher

#12594100). RT- PCR was conducted using FAST SYBR Green Master Mix (Applied Biosystems #4385612) in accordance with the manufacturer's instructions. Relative gene expression was determined by using the ΔΔ-CT method.

Cells were treated with RIPA buffer, and then the lysates were subjected to SDS-PAGE (Invitrogen) and transferred to a PVDF membrane (GE) following the manufacturer's protocols at a constant 100 V for 1 hour. After incubation with 5% non-fat milk in TBS-T (10 mM Tris, pH 8.0, 150 mM NaCl, 0.5% Tween-20) for 1 hour at room temperature, the membrane was incubated with antibodies against GAPDH (Abcam, ab9485, 1:1000), ARID5B (NOVUS Biologicals NBP1-83622, 1:1000), and mCherry (Abcam ab167453, 1:1000) at 4°C for 48 hours with gentle shaking. Membranes were washed three times for 30 mins and incubated with a 1:2000 dilution of horseradish peroxidase–conjugated secondary antibodies for 2 hours at room temperature. Blots were washed with TBS-T three times for 30 mins and developed with the ECL system (Amersham Biosciences) in accordance with the manufacturer's protocols.

Luciferase reporter assay to characterize the effects of *ARID5B* variants

According to CRISPR/dCas9-KRAB library screening and CRISPR/Cas9 deletion results, intron 3 enhancer containing SNP rs7090445 (1.2 kb) was cloned (cloning primer forward: CCTAACTGGCCGGTACCATCTGATAGGGATACTGTAGGATTTAATAACATATAGAATCC, reverse: CCATTATATACCCTCTAGTGTCTAAGCTTCTCATTTTACCCAGGACTCAGTTGT) into a pGL4.23 (luc2/miniP) luciferase vector (Promega). Site directed mutagenesis was performed to generate plasmids containing each SNP allele (Agilent Technologies QuikChange II Site-Directed Mutagenesis Kit #200521). For luciferase assay, 5×10^6 GM12878 cells cultured in 12-well plate were transiently transfected with 6 μg of pGL4 construct (luciferase gene with *ARID5B* enhancer with reference allele (T) or risk allele (C)) and 1 μg of pGL-TK (Renilla luciferase) using Nucleofector-2b device (Lonza) with Kit V (#VCA-1003) and program X-001. Firefly luciferase activity was measured 24 hours post-transfection and normalized to *Renilla* luciferase activity.

Relative luciferase activity indicates the ratio normalized to the value from empty pGL4.23 vector. All experiments were performed in triplicate and repeated three times.

MEF2C ChIP-seq and ChIP-quantitative PCR at the *ARID5B* locus

ChIP was performed using the Active Motif ChIP-IT Express kit (Active Motif, La Hulpe Belgium). The procedure was briefly described below. Cells were fixed with a specially formulated formaldehyde buffer, which cross-links and preserves protein/DNA interactions. Chromatin is then sheared into small fragments using sonication and incubated with 1 mg antibodies against MEF2C (Cell Signaling #5030S) or IgG2b (Thermo Fisher Scientific #MA5-14447). The antibody-bound protein/DNA complexes were immunoprecipitated using Protein G agarose beads and washed via gravity filtration. Following immunoprecipitation, the DNA cross-links were reversed, the proteins are removed by Proteinase K and the DNA is recovered and purified. ChIP enriched DNA was either subjected to next-generation sequencing for ChIP-seq or used as template for ChIP-quantitative PCR. Paired-end reads of 100bp were mapped human genome hg38(GRCh38-lite) by BWA (version 0.7.12-r1039, default parameter) [16] after trimming for adapters by fastp (version 0.20.0, paired-end mode, parameters "--trim_poly_x --cut_by_quality5 --cut_by_quality3 --cut_mean_quality 15 --length_required 20 --low_complexity_filter --complexity_threshold 30 --detect_adapter_for_pe")[17], duplicated reads were then marked with biobambam2 (version v2.0.87)[18] and only non-duplicated reads have been kept by samtools (parameter "-q 1 -F 1804" version 1.2)[19]. We then generated bigwig files using the center 80bp of fragments smaller than 2000bp and normalized to 10M fragments. Peaks were called by MACS2 (version 2.1.1.20160309, "--extsize 200 --nomodel --keep-dup all")[20]. Three biological replicates were performed using separate chromatin preparations. Interpolated quantities for each target were normalized to an input DNA sample, without antibody pull-down. ChIP-qPCR primer for *ARID5B* intron three enhancer: forward: ATCGATAGCTTTGAGACCTTCTG, reverse: GCAGTTACACTTCTGAGCCTAT; for *ARID5B*

promoter: forward: CGGACGCAGACGTTATGAAA, reverse: GAAGTTAATACACCCGGCAGAG; for negative control region: forward: CTAGTCCCAGCTACTCAAGA, reverse: GAGTGCAGAGGCACAATAA. ChIP-seq data are accessible through GEO series accession number: GSE195831.

MEF2C Co-Immunoprecipitation

Total protein from 20 million cells were harvested using Pierce RIPA Buffer (Thermo Scientific #89900) supplemented with Halt Protease and Phosphatase Inhibitor Cocktail (Thermo Scientific #78440). MEF2C protein polymers were incubated with anti-MEF2C antibody (Abcam #ab211493) and pulled down by Protein G DynaBeads (Thermo Scientific #10007D). The presence of RUNX1 and MEF2C protein in the pulled down lysate was examined by Western blotting using antibodies against RUNX1 (CST #4334S) and MEF2C (CST #5030S) respectively.

Statistical analyses

Four ALL patient cohorts (DCOG, St. Jude TOTAL XIIIA/XIIIB/XV, COG P9906, and Ma-Spore ALL 2010) were used to analyze the correlation between *ARID5B* and *MEF2C* expressions. The expression datasets of these four ALL patient cohorts are accessible through GEO series accession number: GSE13351[21], GSE66702[22], GSE11877[23], and EGA accession number: EGAS00001001858[24], respectively. Correlation coefficient was calculated by Pearson's correlation. The meta-analysis of correlations between *MEF2C* and *ARID5B* in four ALL cohorts was based on a random effect model using inverse-variance method implemented in R "meta" package[25]. R statistical software (version 0.98.1091) was used to analyze and generate the plot[26]. For CRISPRi library screening, *t*-test was used to nominate *cis*-regulatory elements. All statistical tests are 2-sided.

References

1. Liam Abbott SB, Claire Churchhouse, Andrea Ganna, Daniel Howrigan, Duncan Palmer, Ben Neale, Raymond Walters, Caitlin Carey, The Hail team. *UK BIOBANK GWAS*. http://www.nealelab.is/uk-biobank/ (2018; date last accessed).

2. Qian M, Cao X, Devidas 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.

3. Perez-Andreu V, Roberts KG, Harvey RC*, et al.* Inherited GATA3 variants are associated with Ph-like childhood acute lymphoblastic leukemia and risk of relapse. Nat Genet 2013;45(12):1494-8.

4. Xu H, Yang WJ, Perez-Andreu V*, et al.* Novel Susceptibility Variants at 10p12.31-12.2 for Childhood Acute Lymphoblastic Leukemia in Ethnically Diverse Populations. Jnci-Journal of the National Cancer Institute 2013;105(10):733-742.

5. Vo BT, Li C, Morgan MA*, et al.* Inactivation of Ezh2 Upregulates Gfi1 and Drives Aggressive Myc-Driven Group 3 Medulloblastoma. Cell Rep 2017;18(12):2907-2917.

6. Consortium EP. An integrated encyclopedia of DNA elements in the human genome. Nature 2012;489(7414):57-74.

7. Corces MR, Buenrostro JD, Wu B*, et al.* Lineage-specific and single-cell chromatin accessibility charts human hematopoiesis and leukemia evolution. Nat Genet 2016;48(10):1193-203.

8. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biology 2014;15(12).

9. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. Nature Methods 2012;9(4):357-U54.

10. Zhang Y, Liu T, Meyer CA*, et al.* Model-based Analysis of ChIP-Seq (MACS). Genome Biology 2008;9(9).

11. Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics 2010;26(6):841-842.

12. Yang WT, Rosenstiel PC, Schulenburg H. ABSSeq: a new RNA-Seq analysis method based on modelling absolute expression differences. Bmc Genomics 2016;17.

13. Anders S, Pyl PT, Huber W. HTSeq-a Python framework to work with high-throughput sequencing data. Bioinformatics 2015;31(2):166-169.

14. Bentsen M, Goymann P, Schultheis H*, et al.* ATAC-seq footprinting unravels kinetics of transcription factor binding during zygotic genome activation. Nat Commun 2020;11(1):4267.

15. Cowper-Sal lari R, Zhang X, Wright JB*, et al.* Breast cancer risk-associated SNPs modulate the affinity of chromatin for FOXA1 and alter gene expression. Nat Genet 2012;44(11):1191-8.

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

17. Martin M. Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet.journal 2011;17(1):pp.-10-12.

18. Tischler G, Leonard S. biobambam: tools for read pair collation based algorithms on BAM files. Source Code for Biology and Medicine 2014;9(1):13.

19. Li H, Handsaker B, Wysoker A*, et al.* The Sequence Alignment/Map format and SAMtools. Bioinformatics 2009;25(16):2078-2079.

20. Zhang Y, Liu T, Meyer CA*, et al.* Model-based analysis of ChIP-Seq (MACS). Genome Biology 2008;9(9):R137.

21. Den Boer ML, van Slegtenhorst M, De Menezes RX*, et al.* A subtype of childhood acute lymphoblastic leukaemia with poor treatment outcome: a genome-wide classification study. Lancet Oncol 2009;10(2):125-34.

22. Paugh SW, Bonten EJ, Savic D*, et al.* NALP3 inflammasome upregulation and CASP1 cleavage of the glucocorticoid receptor cause glucocorticoid resistance in leukemia cells. Nat Genet 2015;47(6):607-14.

23. Harvey RC, Mullighan CG, Wang X*, et al.* Identification of novel cluster groups in pediatric high-risk B-precursor acute lymphoblastic leukemia with gene expression profiling: correlation with genome-wide DNA copy number alterations, clinical characteristics, and outcome. Blood 2010;116(23):4874-84.

24. Qian MX, Zhang H, Kham SKY*, et al.* Whole-transcriptome sequencing identifies a distinct subtype of acute lymphoblastic leukemia with predominant genomic abnormalities of EP300 and CREBBP. Genome Research 2017;27(2):185-195.

25. Balduzzi S, Rucker G, Schwarzer G. How to perform a meta-analysis with R: a practical tutorial. Evid Based Ment Health 2019;22(4):153-160.

26. Tsui ASM, Erickson LC, Mallikarjunn A*, et al.* Dual language statistical word segmentation in infancy: Simulating a language-mixing bilingual environment. Dev Sci 2020; 10.1111/desc.13050:e13050.

Supplementary table 1. Characteristics of four ALL cohorts					
		COG AAL0232	COG P9904/5/6	COG AALL0331	Total XIII/XV
		n=2,224	n=1,634	n=321	n=829
		n(%)	n(%)	n(%)	n(%)
Gender	Female	989 (44.5)	765 (46.8)	161 (50.2)	375 (45.2)
	Male	1235 (55.5)	869 (53.2)	160 (49.8)	454 (54.8)
Age	< 1 year	0 (0)	0 (0)	0 (0)	8 (0.96)
	1 - 10 year	775 (34.8)	1350 (82.6%)	321 (100)	633 (76.3)
	> 10 year	1449 (65.2)	284 (17.4)	0 (0)	188 (22.7)
Population and ancestry*	African	117 (5.3)	80 (4.9)	2 (0.7)	108 (13.0)
	Admixed American	557 (25.0)	315 (19.3)	52 (16.2)	86 (10.4)
	Asian	50 (2.2)	30 (1.83)	2 (0.6)	7 (0.84)
	European	729 (32.8)	983 (60.2)	62 (19.3)	437 (52.7)
	Other	729 (32.8)	212 (13)	74 (23.0)	39 (4.7)
	Unknown	42 (1.9)	14 (0.86)	129 (40)	152 (18.3)
Subtype	ETV6-RUNX1	287 (12.9)	405 (24.8)	NA	120 (14.5)
	BCR-ABL1	62 (2.8)	NA	NA	12 (1.4)
	KMT2A-rearrangement	77 (3.5)	14 (0.86)	NA	9 (1.1)
	TCF3-PBX1	NA	80 (4.9)	NA	37 (4.5)
	Hyperdiploid	326 (14.6)	450 (27.5)	73 (22.7)	150 (18.1)
	Other	1472 (66.4)	685 (41.9)	248 (77.3)	501 (60.4)

Data are presented as number(%) of patients unless otherwise indicated; NA, data not available

*Ethnicity was determined by using STRUCTURE (version 2.3.4), with the sum of these four ancestries being 100% for any given subject. European American (EA), African American (AA), and Asian were defined as having more than 95% European genetic ancestry, more than 70% African ancestry, and more than 90% Asian ancestry, respectively. Hispanics were individuals for whom Native American ancestry was more than 10% and greater than African ancestry, and the remaining subjects were grouped as "Other", as previously described (Perez-Andreu, V. et al. Nat Genet 2013;45(12):1494-8; Xu, H. et al. J Natl Cancer Inst 2013;105(10):733-42; Qian, M. et al. J Clin Oncol. 2018;36(6):591-599.)

Supplementary Table 2. Sequences of sgRNA library

This table is available for separate download as a .xls file.

Supplementary table 3 . 54 variants with significant association discovered from target-sequencing.

UID	CHROM	POS	REF	ALT	SNP	MAFCase	MAFCtrl	OR*	P.adj.*	European only OR	European only P.adj.	Previous GWAS SNP
1	10	63719739	C	T	rs4948492	0.4891	0.3375	1.717	4.07E-45	1.7100	4.65E-09	
2	10	63721176	C	T	rs7090445	0.4875	0.3382	1.712	5.57E-45	1.7100	4.65E-09	Yes
3	10	63723577	C	T	rs10821936	0.4878	0.3375	1.703	1.35E-44	1.7050	4.65E-09	Yes
4	10	63723909	C	T	rs10821937	0.4871	0.3378	1.701	1.68E-44	1.7050	4.65E-09	
5	10	63722895	C	T	rs4245595	0.4885	0.338	1.699	1.75E-44	1.7070	4.65E-09	Yes
6	10	63724390	A	G	rs7896246	0.4654	0.3378	1.692	7.46E-44	1.6940	6.05E-09	
7	10	63725942	G	A	rs4245597	0.4827	0.3428	1.686	2.64E-43	1.6820	9.72E-09	
8	10	63753482	G	A	rs9415635	0.4838	0.3411	1.672	5.34E-42	1.6790	1.44E-08	
9	10	63752159	G	T	rs7089424	0.4846	0.3414	1.67	7.25E-42	1.6710	1.78E-08	Yes
10	10	63754024	A	T	rs9414758	0.5008	0.3675	1.613	4.32E-37	1.6370	6.08E-08	
11	10	63725424	C	G	rs4948494	0.4329	0.5785	1.5728	4.73E-34	1.6240	1.55E-07	
12	10	63724773	A	C	rs10821938	0.5637	0.4211	1.5704	6.94E-34	1.5390	2.15E-06	
13	10	63725372	G	A	rs4948493	0.4339	0.5781	1.5667	1.29E-33	1.5360	2.29E-06	
14	10	63723440	C	A	rs7923074	0.4351	0.5781	1.5637	2.32E-33	1.5300	2.50E-06	
15	10	63723336	T	C	rs7908445	0.5626	0.4219	1.5630	2.44E-33	1.5270	2.51E-06	
16	10	63725862	A	G	rs4245596	0.421	0.573	1.5593	5.86E-33	1.5070	5.80E-06	
17	10	63751748	A	G	rs10821939	0.4314	0.5727	1.5312	2.31E-30	1.4650	3.57E-05	
18	10	63699895	A	C	rs7073837	0.5209	0.4013	1.4879	3.13E-26	1.4210	1.67E-04	
19	10	63721595	T	C	rs10430495	0.4253	0.5421	1.4881	4.74E-26	1.4640	3.69E-05	
20	10	63700228	A	T	rs4360655	0.521	0.403	1.4859	4.91E-26	1.4490	6.91E-05	
21	10	63700245	A	G	rs4562771	0.5211	0.4034	1.4817	1.07E-25	1.4220	1.67E-04	
22	10	63718039	C	T	rs4131566	0.4334	0.5497	1.4789	1.49E-25	1.4520	4.94E-05	
23	10	63697616	C	G	rs2893880	0.43	0.3002	1.4850	2.35E-24	1.3940	6.74E-04	
24	10	63721616	T	C	rs10430496	0.4296	0.5417	1.4526	8.07E-23	1.4280	1.46E-04	
25	10	63778606	G	T	rs2278308	0.4609	0.4931	0.7865	1.64E-10	0.7654	7.91E-03	
26	10	63721516	C	T	rs10994989	0.05596	0.08795	0.7061	2.69E-06	0.6393	0.035470655	
27	10	63722686	A	G	rs138664335	0.02312	0.01866	1.648	9.19E-05	1.6050	0.219446471	
28	10	63736231	T	C	rs79872566	0.02229	0.01894	1.602	0.000246	1.5860	0.232434857	
29	10	63700935	T	G	rs61852292	0.03724	0.06778	0.7342	0.000306	0.9675	0.919234513	
30	10	63717717	A	C	rs77856151	0.03284	0.05708	0.7186	0.000372	0.6865	0.206563125	
31	10	63658086	G	A	rs7893397	0.155	0.1022	1.226	0.001316	1.2470	0.2712765	
32	10	63656715	A	T	rs34990179	0.09511	0.105	0.8004	0.001316	0.8838	0.692338356	
33	10	63829539	T	C	rs76296217	0.04873	0.08425	0.787	0.002047	1.1140	0.734263291	
34	10	63660689	A	G	rs6415872	0.4615	0.5011	0.8845	0.00234	0.7896	0.021433889	
35	10	63663395	G	A	rs74868128	0.101	0.1063	0.8282	0.007336	0.8503	0.548105263	
36	10	63664397	T	C	rs12412757	0.09784	0.1055	0.8301	0.008309	0.8551	0.566008475	
37	10	63665247	G	A	rs4589241	0.09794	0.1057	0.8311	0.008571	0.8551	0.566008475	
38	10	63661340	C	T	rs4509706	0.1011	0.1057	0.8341	0.009705	0.8779	0.683034375	
39	10	63661035	G	C	rs4405263	0.108	0.1058	0.8341	0.009705	0.8898	0.692338356	
40	10	63723517	A	G	rs77918077	0.04885	0.04967	1.258	0.010551	1.2480	0.498485455	
41	10	63664036	T	G	rs36082697	0.1122	0.1051	0.8442	0.017064	0.8706	0.643114286	
42	10	63720965	A	C	rs139936628	0.0147	0.02497	0.7154	0.024965	0.9388	0.919234513	
43	10	63701011	C	T	rs78385480	0.04268	0.06092	0.8139	0.024965	0.7464	0.297113333	
44	10	63753401	G	A	rs151146136	0.01636	0.02676	0.7269	0.024965	0.8307	0.744222353	
45	10	63752510	G	A	rs149590676	0.0166	0.02676	0.7269	0.024965	0.8307	0.744222353	
46	10	63753120	T	C	rs147534635	0.01637	0.02676	0.7271	0.024965	0.8307	0.744222353	
47	10	63755722	C	CA	rs201836069	0.01614	0.02676	0.7274	0.024965	0.8307	0.744222353	
48	10	63701737	C	T	rs75526350	0.0428	0.06078	0.8175	0.027603	0.7462	0.297113333	
49	10	63701536	T	C	rs112779912	0.04768	0.07286	0.8337	0.031238	1.0170	0.949063866	
50	10	63798430	C	A	rs112734707	0.0351	0.03622	1.252	0.031775	1.2280	0.621348387	
51	10	63747914	T	C	rs74914138	0.02336	0.02648	0.7388	0.032158	0.8407	0.759385227	
52	10	63663741	C	T	rs956002	0.2888	0.3672	0.9117	0.032158	0.8284	0.126253548	
53	10	63746820	T	G	rs78068393	0.02356	0.02648	0.7447	0.037761	0.8439	0.763014607	
54	10	63663562	G	A	rs72833334	0.2887	0.3662	0.9153	0.041252	0.8291	0.126253548	

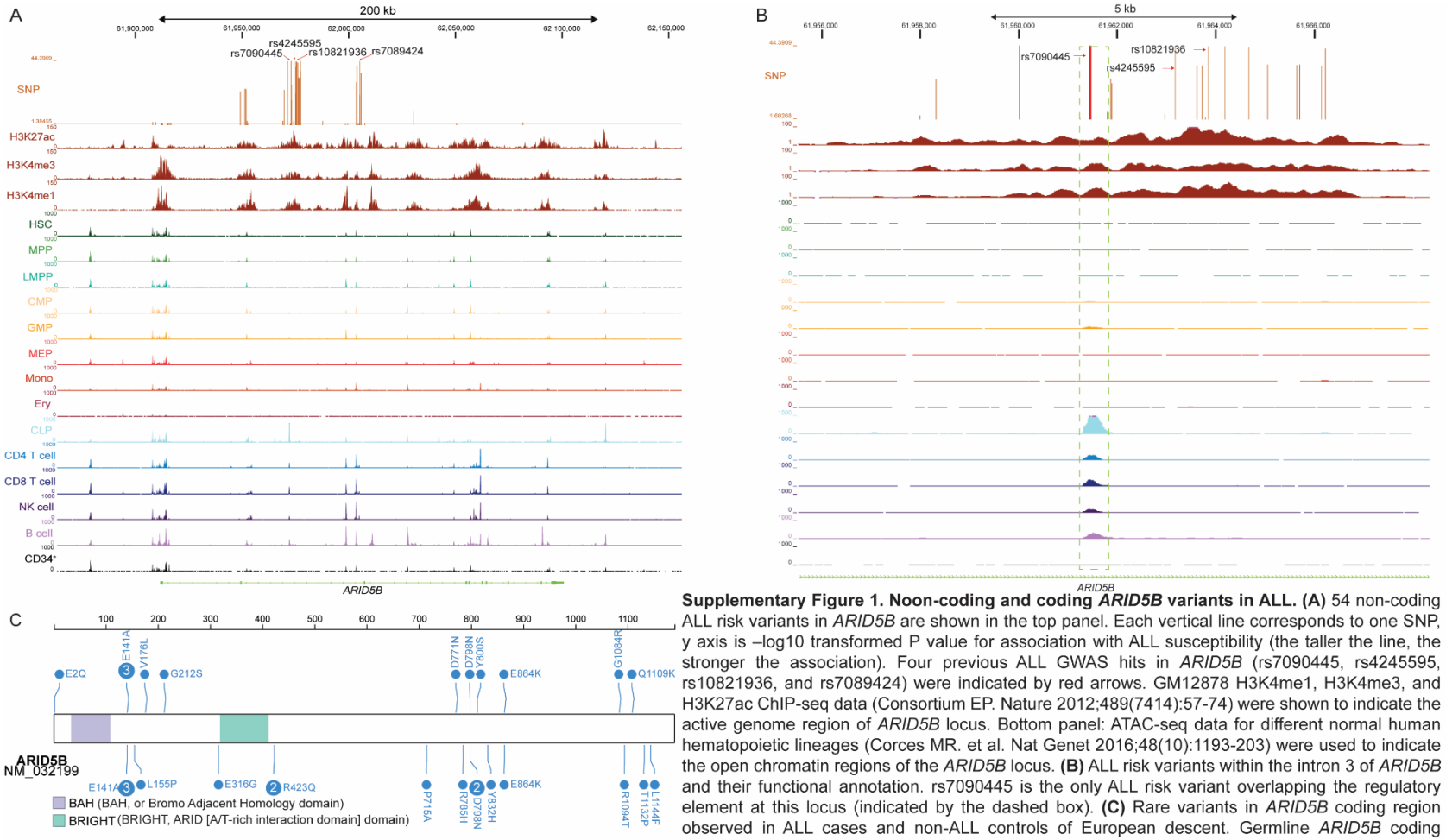
Genetic ancestry (European, African, Amerindian, and East Asian) of four ALL cohorts was first estimated, and 54 leukemia risk variants were identified using an additive logistic regression model with genetic ancestry as covariates by comparing genotype frequency in 4,671 ALL cases and 3,644 UK10K non-ALL controls.

**P* values and odds ratios (OR) were estimated by the additive logistic regression test adjusting genetic ancestry

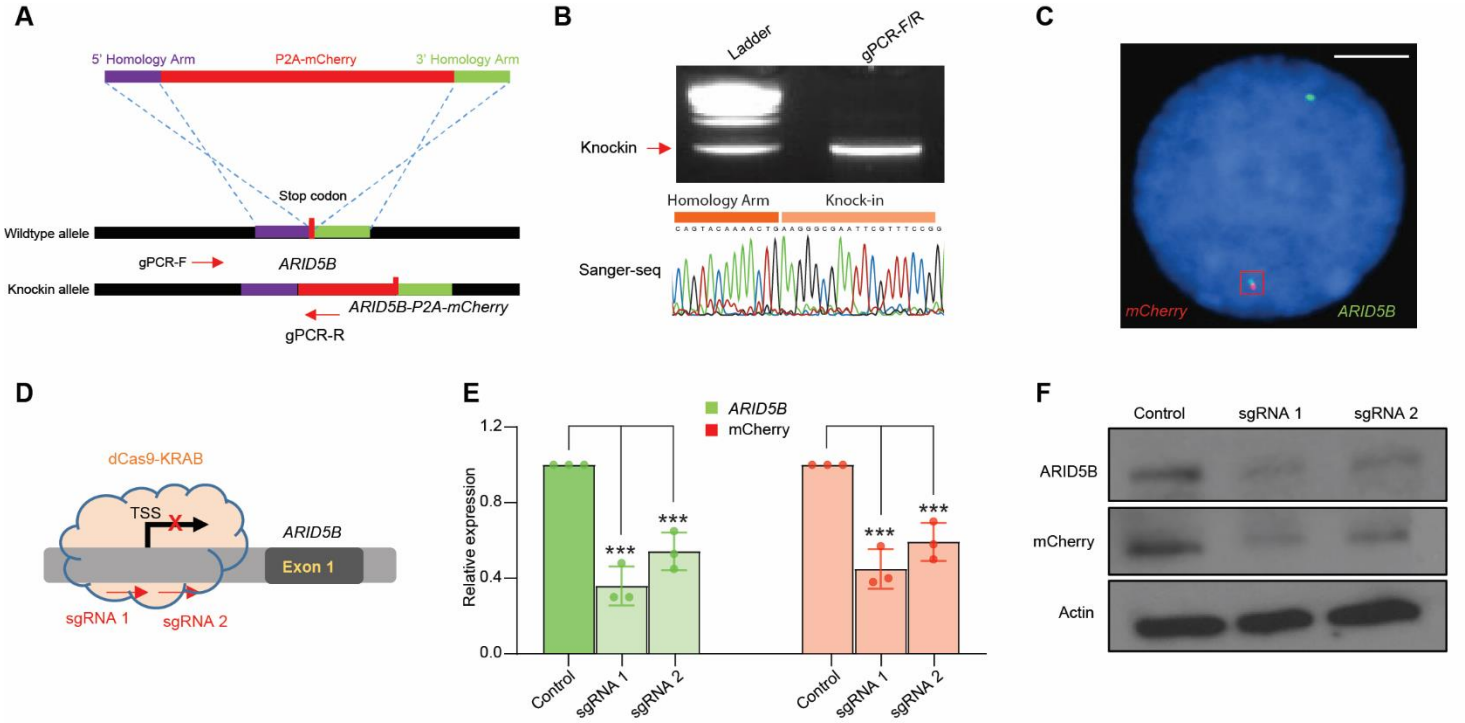
Supplmentary table 4. Regulatory elements identified based on CRISPRi screen by *t*-test				
Chromosome	Start	End	P value	Segment
chr10	61750605	61753229	1	1
chr10	61765397	61766241	1	2
chr10	61779821	61780790	1	3
chr10	61785602	61787583	1	4
chr10	61804403	61804804	1	5
chr10	61808455	61809652	0.070996886	6
chr10	61866724	61868146	1	7
chr10	61896746	61907434	4.32E-13	8
chr10	61936760	61946300	4.19E-19	9
chr10	61957118	61967424	1.39E-80	10
chr10	61984718	62003907	5.79E-19	11
chr10	62015452	62024895	4.27E-07	12
chr10	62037614	62056787	4.64E-26	13
chr10	62069711	62069795	1	14
chr10	62075759	62088987	3.59E-20	15
chr10	62104034	62111946	0.262208007	16
chr10	62165713	62167960	1	17
chr10	62170212	62171651	1	18
chr10	62182171	62186948	0.009123961	19
chr10	62239747	62246306	1	20
chr10	62267453	62269924	0.01209644	21

In total, sgRNA library falls into 21 segments from the upstream to the downstream of *ARID5B* gene. Seven of them are significant ($P<0.0001$) by *t*-test via comparing each segment of *ARID5B*^{mCherry} knock-in to that of random knock-in. Of the seven, No.8 is the promoter region (N.O. 8) and the other six are enhancers (N.O. 9-13, 15).

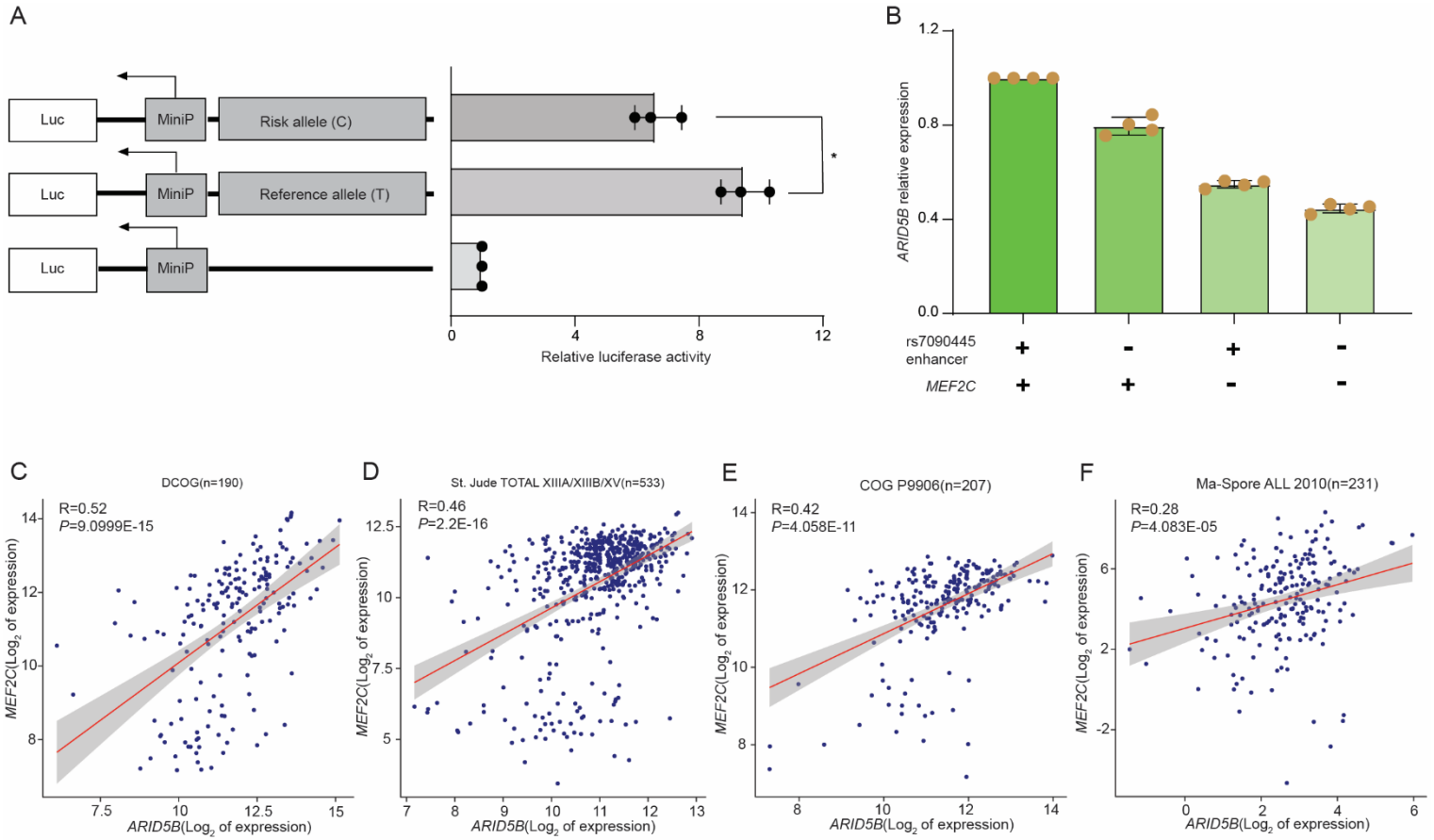
Supplmentary table 5. Transcription factor motif analysis using ENCODE ChIP-seq data				
TF	PWM Score	Alt.Motif	PWM Score	Ref.Motif
	Risk allele (C)		Reference allele (T)	
ATF2	-21.4607	ATCGATAA	-21.4607	ATCAATAA
CUX2	-9.61468	CTAACCTAGGTTATCGAT	-12.9908	CTAACCTAGGTTATTGAT
EP300	-7.16364	GTTATCG	-12.6667	GTTATTG
FOXM1	-8.93976	GGTCTCAAAGCTATCGATAACC	-6.00602	GGTCTCAAAGCTATCAATAACC
HOXD8	-1.41085	GTTATCGATA	3.02326	GTTATTGATA
MEF2A	-32.4268	CTCAAAGCTATCGATAACCTAG	-20.5976	CTCAAAGCTATCAATAACCTAG
MEF2C	2.26012	AGGTTATCGATAGC	7.30058	AGGTTATTGATAGC
NFIC	-32.1633	TATCGATAACCTAG	-23.1735	TATCAATAACCTAG
RELA	-18.3596	TCGATAACCT	-23.8989	TCAATAACCT
RUNX3	-21.5577	AAGCTATCGA	-19.7308	AAGCTATCAA
10 transcription factors were identified with ChIP-seq peak overlapping with *ARID5B* rs709445.				
Alt.Motif: motif with risk allele C; Ref.Motif: motif with reference allele T				



Supplementary Figure 1. Noon-coding and coding *ARID5B* variants in ALL. (A) 54 non-coding ALL risk variants in *ARID5B* are shown in the top panel. Each vertical line corresponds to one SNP, y axis is –log10 transformed P value for association with ALL susceptibility (the taller the line, the stronger the association). Four previous ALL GWAS hits in *ARID5B* (rs7090445, rs4245595, rs10821936, and rs7089424) were indicated by red arrows. GM12878 H3K4me1, H3K4me3, and H3K27ac ChIP-seq data (Consortium EP. Nature 2012;489(7414):57-74) were shown to indicate the active genome region of *ARID5B* locus. Bottom panel: ATAC-seq data for different normal human hematopoietic lineages (Corces MR. et al. Nat Genet 2016;48(10):1193-203) were used to indicate the open chromatin regions of the *ARID5B* locus. **(B)** ALL risk variants within the intron 3 of *ARID5B* and their functional annotation. rs7090445 is the only ALL risk variant overlapping the regulatory element at this locus (indicated by the dashed box). **(C)** Rare variants in *ARID5B* coding region observed in ALL cases and non-ALL controls of European descent. Germline *ARID5B* coding variants were identified by targeted sequencing of 2,629 patients with ALL with European ancestry (top) and by whole-genome sequencing in 3,592 subjects from the UK10K cohort as non-ALL controls (bottom). Only variants with CADD score >15 are shown and analyzed because they are considered as deleterious. Functional domains of the ARID5B protein are indicated by color. Each circle represents a unique individual who carries the indicated variant, except for variants that recur in more than one individual, the exact frequency is indicated as a number in the circle. No statistically significant difference was observed in the prevalence of ARID5B deleterious variants between ALL cases and non-ALL controls (UK10K), using the SKAT method (SNP-set [Sequence] Kernel Association Test).



Supplementary Figure 2. The development and validation of the ALL reporter cell line for *ARID5B* enhancer screen. **(A)** Schematic diagram of CRISPR/Cas9-mediated P2A-mCherry knock-in: a donor template with P2A-mCherry flanked by homology arms for *ARID5B*, and P2A-mCherry was successfully knocked in right after the *ARID5B* stop codon. **(B)** Genomic PCR (gPCR) and Sanger-seq were performed to confirm the successful knock-in. **(C)** FISH was also applied to confirm the insertion of mCherry coding sequence. Red: mCherry; green: ARID5B. Scale bar:5 μm. **(D)** Schematic diagram of CRISPRi with two sgRNAs targeting the *ARID5B* TSS region to disrupt *ARID5B* transcription. **(E, F)** When ARID5B expression was downregulated by CRISPRi, the expression of mCherry also decreased, as shown by RT-PCR (**E**) and Western blotting (**F**). RT-PCR results were from 3 independent experiments. Western blotting results were from one of the three independent experiments. ***: $P<0.001$. Error bar: mean + SD. P value was estimated by two-sided t-test.



Supplementary Figure 3. MEF2C-mediated regulation of *ARID5B* expression is affected by rs7090445. (A) A Luciferase assay was performed to validate the allele-biased enhancer activity for the CRE encompassing rs7090445. The risk allele (C) significantly decreased enhancer activity compared to reference allele (T). *: *P*=0.01. Error bar: mean + SD. *P* value was estimated by two-sided *t*-test. **(B)** Knocking out *MEF2C* in Nalm6 cells with deletion of the intron 3 enhancer of *ARID5B*. While deletion of the intron 3 enhancer of *ARID5B* led to downregulation of *ARID5B* expression, concurrent knock-out of *MEF2C* further downregulated *ARID5B* expression. **(C-F)** Correlation between the expression of *MEF2C* and *ARID5B* in primary ALL blasts. Expression of respective genes was extracted from previously published gene expression array datasets. A positive correlation between *MEF2C* and *ARID5B* was confirmed in four ALL cohorts (DCOG [**C**], St. Jude TOTAL XIIIA/XIIIB/XV [**D**], COG P9906 [**E**], and Ma-Spore ALL 2010 [**F**]) by Pearson's correlation.