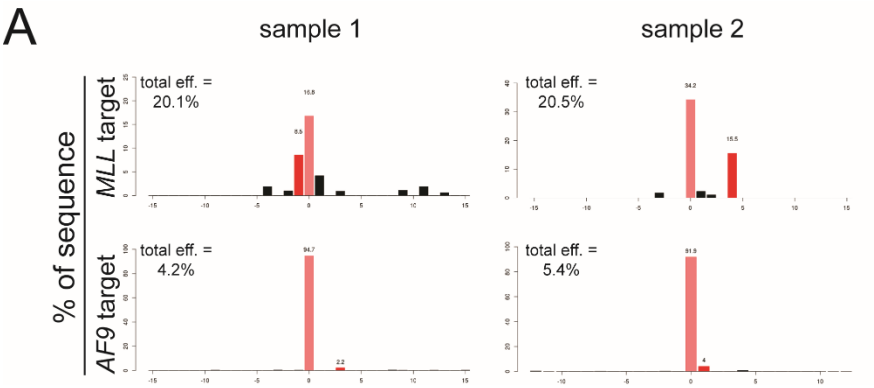
Supplementary Materials

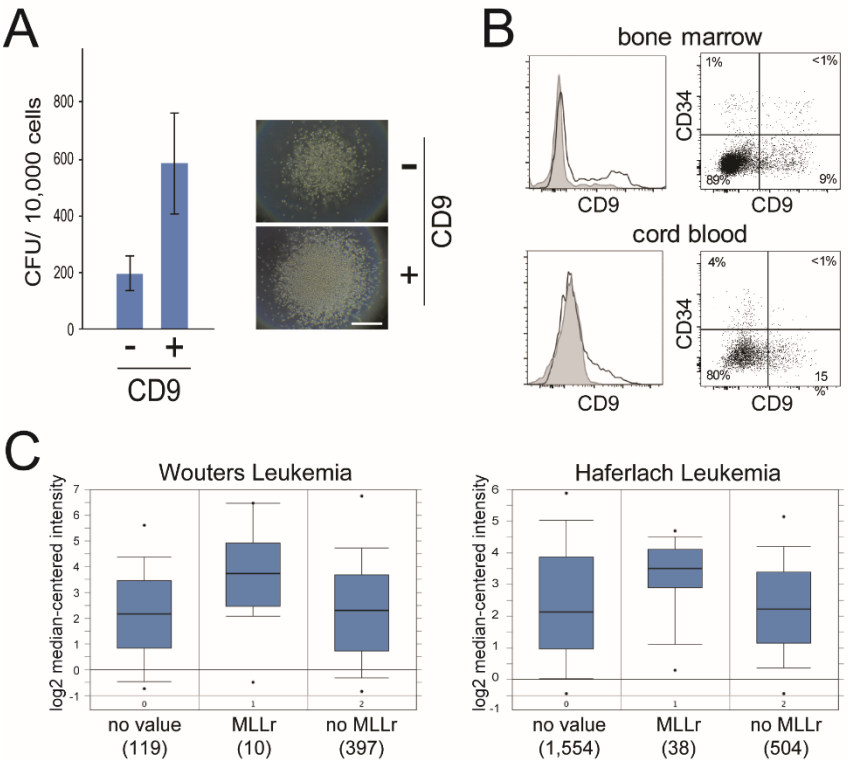
MLL leukemia induction by t(9;11) chromosomal translocation in human hematopoietic stem cells using genome editing

Corina Schneidawind, Johan Jeong, Dominik Schneidawind, In-Suk Kim, Jesús Duque-Afonso', Stephen Hon Kit Wong, Masayuki Iwasaki, Erin H. Breese, James L. Zehnder, Matthew Porteus and Michael L. Cleary

Supplementary Figures

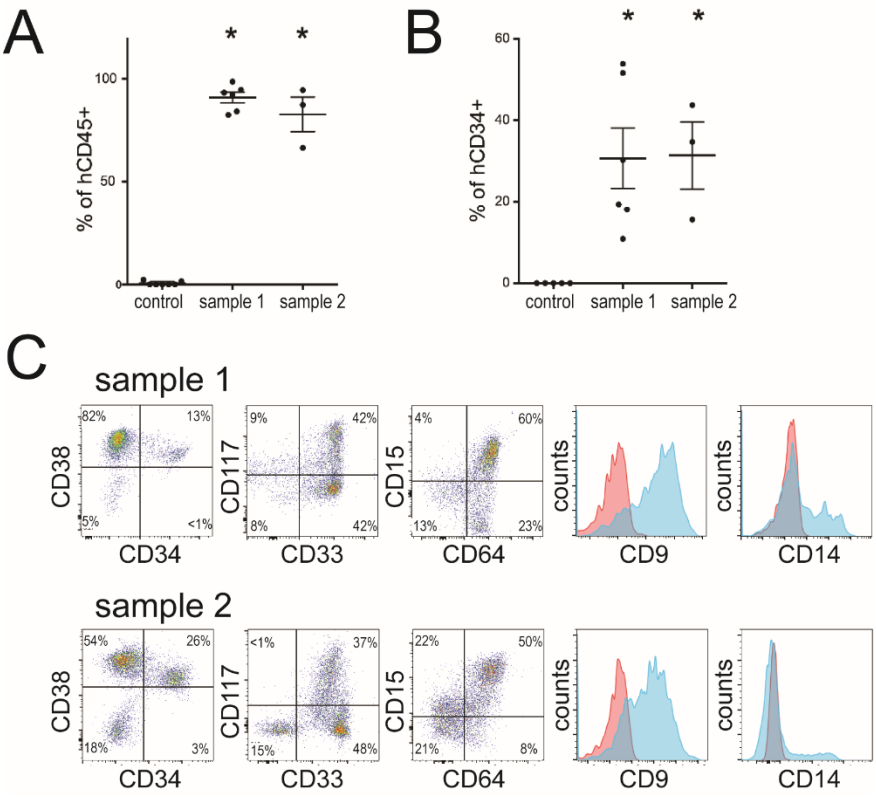


Supplementary Figure S1. TALEN-induced DNA double strand break. (A) Indel frequencies in the *MLL* and *AF9* alleles not involved in the translocation were measured and compared to the control using the online Tracking of Indels by Decomposition (TIDE) software (http://tide.nki.nl).

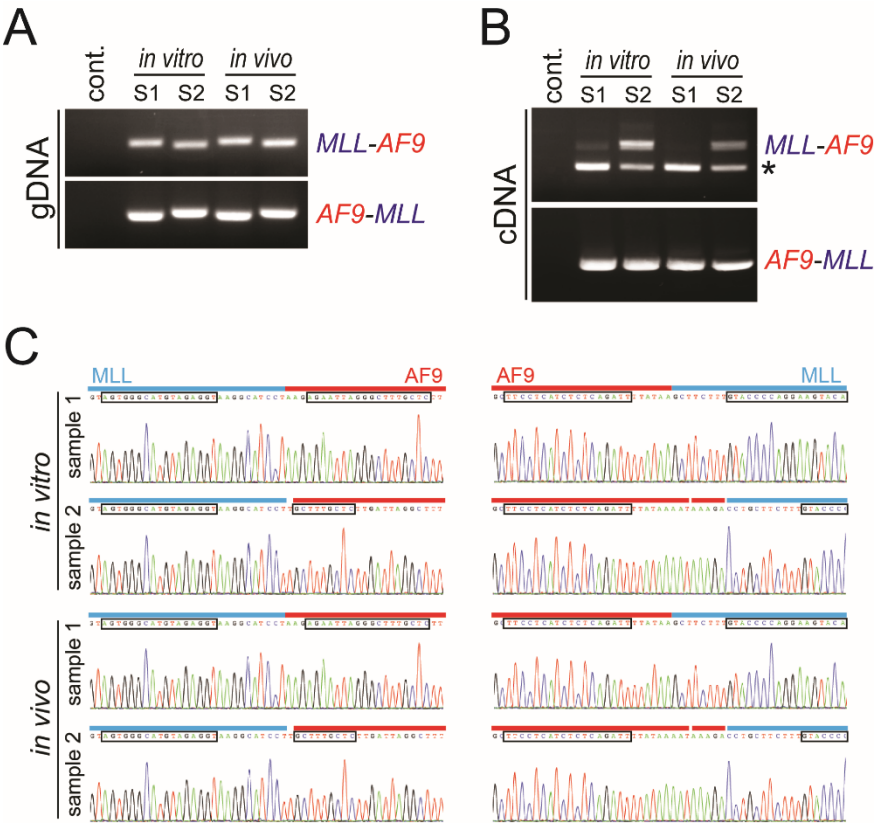


Supplementary Figure S2. Enrichment of clonogenic cells in the CD9-positive population, and CD9 expression in healthy controls and patients with *MLL*r leukemia. (A) Left, bar graph shows the frequency of colony-forming cells for FACS-sorted CD9-negative and CD9-positive cells on day 90 plated in semi-solid media for two weeks. Results are from triplicate analyses (±SEM). Right, representative colony morphologies of indicated cells. Scale bar, 400μm. (B) Representative flow cytometry plots of CD9 expression on MNCs from bone marrow and cord blood from healthy donors. Grey shading indicates control (FMO); black line denotes expression of indicated marker. (C) CD9 expression in *MLLr* AML patient samples compared to that in non-*MLLr* leukemia patients (data from Oncomine.org)

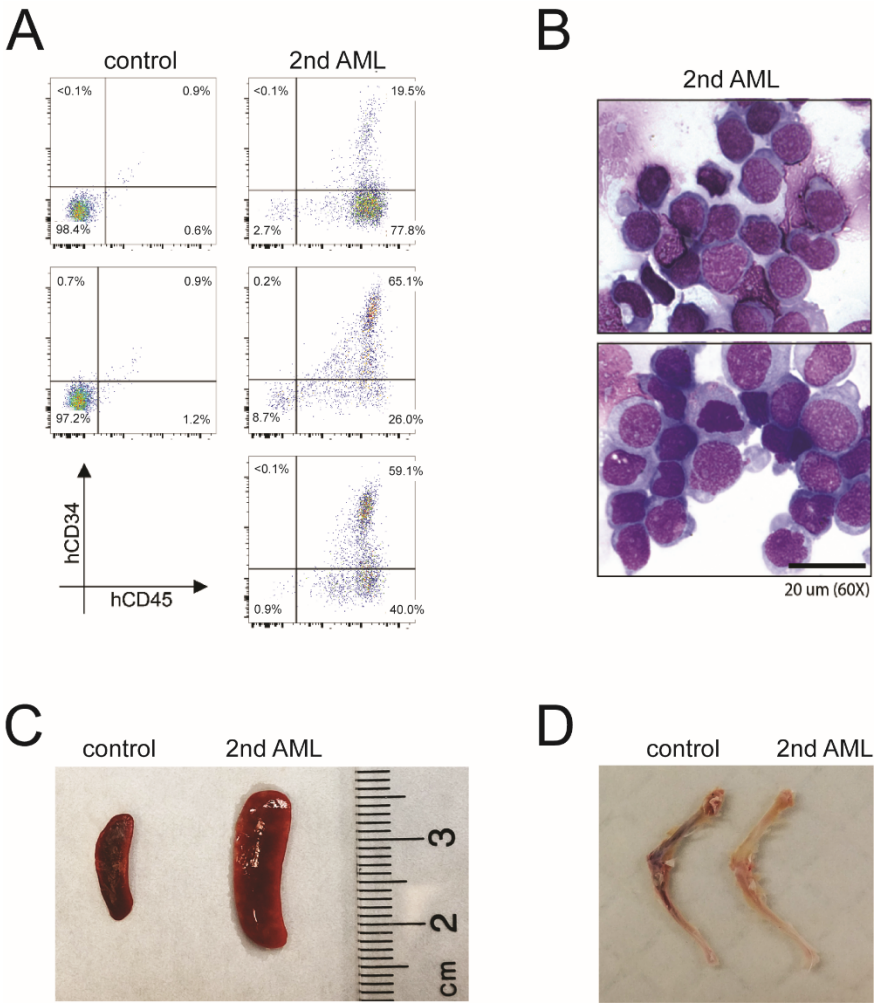
.



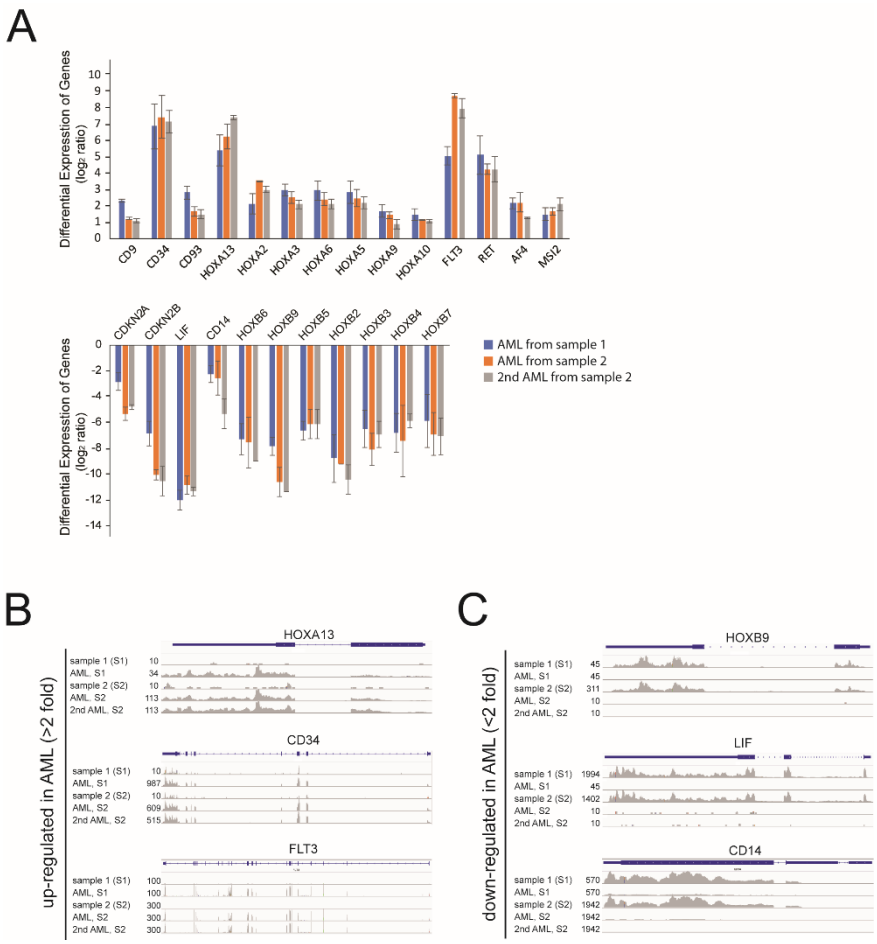
Supplementary Figure S3. Immature myelomonocytic phenotypes of the xenotransplanted leukemic bone marrow cells. (A) Plots show proportions of human cells expressing hCD45 or (B) hCD34 in the bone marrow of transplanted mice, **p*<0.0001 and ***p*<0.005. (C) Flow cytometry analysis of bone marrow cells in leukemic mice for indicated cell surface proteins. Red shading indicates control (FMO); blue shading denotes expression of indicated marker.



Supplementary Figure S4. Monoclonal expansion of *MLL-AF9* chromosomal-translocated cells in mouse leukemia. (A) PCR products are shown for *MLL-AF9* and *AF9-MLL* breakpoint junctions present in gDNA of cultured genome-edited cells (*in vitro*) or corresponding leukemic mouse bone marrow cells (*in vivo*). Control is *GFP*-transduced $CD34^+$ cord blood cells; S1, sample 1; S2, sample 2. (B) RT-PCR products are shown for the *MLL-AF9* and reciprocal *AF9-MLL* fusion transcripts expressed in the indicated cells. Asterisk indicates alternative *MLL-AF9* fusion transcript, which lacks MLL exon 11. (C) DNA sequences are shown for PCR products in (A) to compare translocation breakpoints present in the indicated cells. Boxed sequences indicate TALEN binding sites.



Supplementary Figure S5. Pathologic features of disease in secondary recipient mice transplanted with genome-edited AML cells. (A) Flow cytometry analysis shows relative proportions of engrafted human CD45-positive cells in the bone marrows of secondary transplanted leukemic mice. (B) Morphologies of the leukemic cells in mouse bone marrows are shown by May-Grünwald-Giemsa staining of cytospin preparations. Scale bar, 20 μm. (C and D) Representative image of enlarged spleen (C) and pale leg bone (D) from secondary transplanted mouse compared to control.



Supplementary Figure S6. AML signature gene expression changes in the *MLL-AF9* AMLs. (A) Gene expression changes of *MLL*-leukemia associated genes in primary and secondary transplanted leukemic mice compared to their respective injected cells. Up, up-regulated genes and down, down-regulated genes. (B and C) Representative IGV images showing normalized RNA-seq coverage of genes up- (B) and down-regulated (C) in leukemic bone marrow cells compared to the injected cells.

Supplementary Tables

Supplementary Table S1. Primers used for PCR and RT-PCR.

Name	Sequence (5’ to 3’)	Note
gMLL-AF9, M	TTTTATGCTTTTCATCCTTATTTTTCC	genomic *MLL-AF9, MLL*
gMLL-AF9, A	AGTTCTTGAATGGAATTAAAAGTGCT	genomic *MLL-AF9, AF9*
gAF9-MLL, A	AAGGCTGTTTCGTCTACATAGAAAAT	genomic *AF9-MLL, AF9*
gAF9-MLL, M	ACTTCAAAACATTTCTTTAGCTGGTT	genomic *AF9-MLL, MLL*
cMLL-AF9, M224	ATCCCTGTAAAACAAAAACCAAAA	cDNA *MLL-AF9, MLL*
cMLL-AF9, A225	TTATAGACCTCAAAGGACCTTGTTG	cDNA *MLL-AF9, AF9*
cAF9-MLL, A226	CCATCACCAGTGGACAAGATAA	cDNA *AF9-MLL, AF9*
cAF9-MLL, M227	AGAGGGCAGAAGTTTCCTTTAG	cDNA *AF9-MLL, MLL*
cMLL-AF9, M368	CGCCCAAGTATCCCTGTAAA	cDNA *MLL-AF9, MLL*
cMLL-AF9, A369	TGTCGTTATCCTCCACTTCATC	cDNA *MLL-AF9, AF9*
cAF9-MLL, A372	GATGGCCTTCAAGGAACCTAA	cDNA *AF9-MLL, AF9*
cAF9-MLL, M373	GCTGTTTCGGCACTTATTACAC	cDNA *AF9-MLL, MLL*

Supplementary Table S2. Mutations found in RNA-seq.

Gene	Chromosome	Position	REF	ALT	Consequence
NREP	chr5	111091536	T	A	Intron/transcript variant
DBF4	chr7	87516147	G	C	NMD
ENY2	chr8	110346882	T	C	NMD
APIP	chr11	34912186	T	A	NMD
OS9	chr12	58089822	G	A	NMD
CNOT2	chr12	70723195	T	C	splice region variant
BCKDHA	chr19	41916827	G	C	splice acceptor variant
SNX5	chr20	17933230	C	G	splice donor variant
RUNX1	chr21	36193982	C	T	Ala246Thr
RBX1	chr22	41360050	G	T	splice acceptor variant

* REF, reference sequence; ALT, altered sequence; NMD, nonsense-mediated decay

Supplementary Methods

Colony-forming cell assays

$CD34^+$ cells nucleofected with *MLL* and *AF9* TALENs or control (*GFP* or *MLL* TALENs alone) were removed from liquid cultures after 60-70 days and seeded in triplicate (10,000 cells/dish) in Methocult H4230 methylcellulose medium (StemCell Technologies) supplemented with SCF, TPO, FLT3L (100 ng/µl each), IL-6, IL-3 and G-CSF (50 ng/µl each), SR1 and UM729 (0.75 µM). Colony-forming cell (CFC) assays were performed as previously described.[1,2] Total cell counts were determined after each round of plating.

Histopathology, fluorescence *in situ* hybridization and karyotype analyses

Cytospins were prepared and stained by Wright-Giemsa (Sigma). Fluorescence *in situ* hybridization (FISH) and G-Banding analyses were performed by the Cytogenetics Laboratory of Stanford Hospital. For FISH, cells were fixed and hybridized with the Vysis *MLL* Dual Color Break Apart Rearrangement Probe (5' green, 3' red). 200 cells were analyzed per sample for the presence or absence of an *MLL* rearrangement. Representative images were obtained for each sample. For karyotyping, metaphase cells were prepared by standard cytogenetics methods. Karyotypes were described according to the International System for Human Cytogenomic Nomenclature.

Western blots

Cells (200,000 per sample) were lysed in 2X SDS-protein sample buffer and boiled for 5 minutes. WT MLL and MLL-AF9 fusion proteins were separated in 4-15% TGX gradient

gel (Bio-Rad) and visualized by Western blot using anti-MLL antibody (Bethyl Laboratories, A300-086A). Anti-GAPDH antibody (SIGMA, G9545) was used for loading control. To quantify relative MLL-AF9 expression to WT MLL^N, band intensities were measured, normalized to GAPDH, and compared using ImageJ and Excel, respectively.

Next generation sequencing (NGS)

For targeted exome sequencing, genomic DNA was extracted from cultured cells and mouse bone marrow cells using the DNeasy Blood & Tissue kit (Qiagen, Hilden, Germany) and analyzed using the TruSight myeloid sequencing (Illumina, CA, USA). Paired-end sequencing runs were performed on a MiSeq (Illumina) with reagent kit v3 according to manufacturer's instructions. Paired sequences obtained from each sample were mapped to human genome reference GRCh37/hg19. Variant calls were annotated by Variant Effect Predictor (VEP)[3] and then manually examined with Integrative Genomics Viewer (IGV).[4,5] For detection of *FLT3* internal tandem duplication (ITD), additional variant callers were used specifically for the region chr13:28607161–28609590: (1) Pindel version 0.2.5a7 with the insert size configured, and (2) a novel algorithm ITDseek developed in this study. Variant calls were first annotated by Ensembl Variant Effect Predictor version 75 and then manually examined by at least two individuals. *FLT3* ITD mutations were confirmed in patients using PCR fragment analysis by capillary electrophoresis (primer sequences available upon request) and analysis software Peak Scanner version 1.0 (http://www.appliedbiosystems.com). Variants were described according to the recommendations of Human Genome Variation Society (HGVS). Variant descriptions were checked by Mutalyzer Name Checker (http://mutalyzer.nl).

For RNA-sequencing, RNA was purified from leukemic mouse bone marrow and cultured cells using Trizol (Invitrogen) followed by RNeasy kit (Qiagen), and sample quality was evaluated using Agilent RNA 6000 nano chip and Agilent 2100 Bioanalyzer (Agilent, St. Laurent, QC, Canada). Single-ended 50 bp reads (59 million per sample on average) were generated using a BGISEQ-500 platform (BGI America Co., Cambridge, MA). Somatic SNP/Indel calling and annotation were performed using VarScan 2[6] followed by VEP[3] in DNAnexus cloud platform (dnanexus.com). Human genome reference GRCh37/hg19, minimum coverage of 50, and *p*-value threshold of 0.05 were used. SNV/Indels were manually confirmed using IGV. ClinVar database was used to search pathogenic mutations and GRCm38.p4 mouse genome reference used to filter out mouse sequence.[7]

NGS data have been deposited in NCBI's Gene Expression Omnibus (GEO)[8] and are accessible through accession number GSE103811 (https://www.ncbi.nlm.nih.gov/geo/guery/acc.cgi?acc=GSE103811).

Statistical and bioinformatics analyses

All statistical analyses were performed with the Student's t test unless otherwise indicated. $p<0.05$ was considered statistically significant. Gene set enrichment analyses were performed using GSEA software (http://www.broad.mit.edu.laneproxy.stanford.edu/gsea) with a t test metric for gene ranking and 1,000 data permutations. Gene sets of the 104 most differentially expressed genes in LSC-positive versus LSC-negative cells were used.[9] Hierarchical cluster analysis was performed using Cluster 3.0[10] and visualized by JavaTreeView (ver. 1.1).[11]

Supplementary References

1. Breese EH, Dawson C, Buechele C, Breese MR, Cleary ML, Porteus MH. Using Genome Engineering To Prospectively Investigate The Pathogenesis Of MLL Translocations In Infant Acute Lymphoblastic Leukemia. *Blood* 2013;122(21):3722.
2. Lavau C, Szilvassy SJ, Slany R, Cleary ML. Immortalization and leukemic transformation of a myelomonocytic precursor by retrovirally transduced HRX-ENL. *EMBO J*. 1997;16(14):4226-4237.
3. McLaren W, Gil L, Hunt SE, et al. The Ensembl Variant Effect Predictor. *Genome Biology*. 2016;17(1):122.
4. Robinson JT, Thorvaldsdottir H, Winckler W, et al. Integrative genomics viewer. *Nat Biotech*. 2011;29(1):24-26.
5. Thorvaldsdóttir H, Robinson JT, Mesirov JP. Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Briefings in Bioinformatics*. 2013;14(2):178-192.
6. Koboldt DC, Zhang Q, Larson DE, et al. VarScan 2: Somatic mutation and copy number alteration discovery in cancer by exome sequencing. *Genome Research*. 2012;22(3):568-576.
7. Landrum MJ, Lee JM, Riley GR, Jang W, Rubinstein WS, Church DM. ClinVar: public archive of relationships among sequence variation and human phenotype. *Nucleic Acids Res*. 2014;42.
8. Barrett T, Wilhite SE, Ledoux P, et al. NCBI GEO: archive for functional genomics data sets—update. *Nucleic Acids Research*. 2013;41(Database issue):D991-D995.
9. Ng SWK, Mitchell A, Kennedy JA, et al. A 17-gene stemness score for rapid determination of risk in acute leukaemia. *Nature*. 2016;540(7633):433-437.
10. de Hoon MJ, Imoto S, Nolan J, Miyano S. Open source clustering software. *Bioinformatics*. 2004;20(9):1453-1454.
11. Saldanha AJ. Java Treeview—extensible visualization of microarray data. *Bioinformatics*. 2004;20(17):3246-3248.