***eLife’s* transparent reporting form**

We encourage authors to provide detailed information *within their submission* to facilitate the interpretation and replication of experiments. Authors can upload supporting documentation to indicate the use of appropriate reporting guidelines for health-related research (see [EQUATOR Network](http://www.equator-network.org/%20)), life science research (see the [BioSharing Information Resource](https://biosharing.org/%22%20%5Ct%20%22_blank)), or the [ARRIVE guidelines](http://www.plosbiology.org/article/info%3Adoi/10.1371/journal.pbio.1000412) for reporting work involving animal research. Where applicable, authors should refer to any relevant reporting standards documents in this form.

If you have any questions, please consult our Journal Policies and/or contact us: editorial@elifesciences.org.

**Sample-size estimation**

* You should state whether an appropriate sample size was computed when the study was being designed
* You should state the statistical method of sample size computation and any required assumptions
* If no explicit power analysis was used, you should describe how you decided what sample (replicate) size (number) to use

Please outline where this information can be found within the submission (e.g., sections or figure legends), or explain why this information doesn’t apply to your submission:

The chosen sample sizes are based on the numbers used for previous publications. In transplantation assay, donor cells underwent gene transduction/modification were injected into 3 –6 mice/group. This procedure was duplicated, usually giving final sample size of N = 7 ­–12/group. In some cases, larger sample size was used to precisely determine the penetrance of certain types of leukemia (Figures 3D and 4E), or to further confirm the difference between two groups (Figure 7A,B). Smaller sample size (n = 4 or 5) was used in secondary transplantation assay conducted only for confirmation of enrichment of leukemia cells harvested from primary leukemia mice, and those for technical validation of on-target effect of sgRNA (Figures 4E and 7D)

**Replicates**

* You should report how often each experiment was performed
* You should include a definition of biological versus technical replication
* The data obtained should be provided and sufficient information should be provided to indicate the number of independent biological and/or technical replicates
* If you encountered any outliers, you should describe how these were handled
* Criteria for exclusion/inclusion of data should be clearly stated
* High-throughput sequence data should be uploaded before submission, with a private link for reviewers provided (these are available from both GEO and ArrayExpress)

Please outline where this information can be found within the submission (e.g., sections or figure legends), or explain why this information doesn’t apply to your submission:

Numbers of experiments and whether they are biological or technical replicates are indicated in the figure legends. All of the IP -western, ChIP-qPCR, RT-qPCR experiments were performed at least twice and confirmed their reproducibility. Myeloid progenitor transformation assay was performed at least three times (≥3 biological replicates). ChIP-seq analysis was performed once, and its reproducibility was confirmed by ChIP-qPCR analysis on selected targets.

ChIP-seq data have been deposited to the DDBJ archive and have been published.

RNA-seq analyses of murine cells were performed for 3 biological replicates of MYC-, HOXA9-, and MLL-AF10-ICs and single RNA samples of c-kit-positive cells, HOXA9/MYC LCs, and HOXA9/MEIS1 LCs.

RNA-seq analyses of human cell lines were performed for single RNA samples or two different preparations of HB1119 cells.

**Statistical reporting**

* Statistical analysis methods should be described and justified
* Raw data should be presented in figures whenever informative to do so (typically when N per group is less than 10)
* For each experiment, you should identify the statistical tests used, exact values of N, definitions of center, methods of multiple test correction, and dispersion and precision measures (e.g., mean, median, SD, SEM, confidence intervals; and, for the major substantive results, a measure of effect size (e.g., Pearson's r, Cohen's d)
* Report exact p-values wherever possible alongside the summary statistics and 95% confidence intervals. These should be reported for all key questions and not only when the p-value is less than 0.05.

Please outline where this information can be found within the submission (e.g., sections or figure legends), or explain why this information doesn’t apply to your submission:

Statistical methods used are indicated in the figure legends. Data are represented as mean ± SD with the number of replicates. For multiple comparisons for in vitro experiments, ANOVA followed by Dunnett’s test was used. For in vivo experiments, log-rank test followed by Bonferroni correction was used.

(For large datasets, or papers with a very large number of statistical tests, you may upload a single table file with tests, Ns, etc., with reference to sections in the manuscript.)

**Group allocation**

* Indicate how samples were allocated into experimental groups (in the case of clinical studies, please specify allocation to treatment method); if randomization was used, please also state if restricted randomization was applied
* Indicate if masking was used during group allocation, data collection and/or data analysis

Please outline where this information can be found within the submission (e.g., sections or figure legends), or explain why this information doesn’t apply to your submission:

In transplantation assay, irradiated recipient mice were allocated into experimental groups at random. After injection, mice were kept in a cage of the same experimental group. If needed to increase the sample size of certain groups, more recipient mice were newly allocated only for the groups. No blinding was used.

**Additional data files (“source data”)**

* We encourage you to upload relevant additional data files, such as numerical data that are represented as a graph in a figure, or as a summary table
* Where provided, these should be in the most useful format, and they can be uploaded as “Source data” files linked to a main figure or table
* Include model definition files including the full list of parameters used
* Include code used for data analysis (e.g., R, MatLab)
* Avoid stating that data files are “available upon request”

Please indicate the figures or tables for which source data files have been provided: