High-resolution characterization of gene function using single-cell CRISPR tiling screen

Lu Yang
Beckman Research Institute of City of Hope
Anthony Chan
City Of Hope National Medical Center
Kazuya Miyashita
City of Hope
Christopher Delaney
Dana-Farber Cancer Institute
Xi Wang
Dana-Farber Cancer Institute
Hongzhi Li
City of Hope National Medical Center
Sheela Pokharel
City Of Hope National Medical Center
Sandra Li
City Of Hope National Medical Center
Mingli Li
City Of Hope National Medical Center
Xiaobao Xu
City Of Hope National Medical Center
Wei Lu
City of Hope
Qiao Liu
City of Hope
Nicole Mattson
City of Hope
Kevin Chen
Dana-Farber Cancer Institute
Jinhui Wang
City Of Hope National Medical Center
Yate-Ching Yuan
City of Hope Comprehensive Cancer Center and Beckman Research Institute
David Horne

Beckman Research Institute at the City of Hope
Steven Rosen
City of Hope
Yadira Soto-Feliciano
Rockefeller University
Zhaohui Feng
Dana-Farber Cancer Institute
Takayuki Hoshii
Graduate School of Medicine, Chiba University https://orcid.org/0000-0002-5418-1981
Gang Xiao
Zhejiang University School of Medicine
Markus Müschen
Department of Systems Biology https://orcid.org/0000-0002-6064-8613
Jianjun Chen
Department of Systems Biology, Beckman Research Institute of City of Hope https://orcid.org/0000-0003-3749-2902
Scott Armstrong
Dana-Farber Cancer Institute https://orcid.org/0000-0002-9099-4728
Chun-Wei Chen (cweichen@coh.org)
City of Hope https://orcid.org/0000-0002-8737-6830

Article

Keywords: Functional Domains, Regulatory Mechanisms, High-density Mutagenesis, Single-cell Droplet RNA Sequencing

Posted Date: January 15th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-135816/v1

Version of Record: A version of this preprint was published at Nature Communications on July 1st, 2021. See the published version at https://doi.org/10.1038/s41467-021-24324-0.

High-resolution characterization of gene function using single-cell CRISPR tiling screen

Authors:

Lu Yang[1]†, Anthony K.N. Chan[1]†, Kazuya Miyashita[1]†, Christopher D. Delaney[2], Xi Wang[2], Hongzhi Li[3], Sheela Pangeni Pokharel[1], Sandra Li[1], Mingli Li[1], Xiaobao Xu[1], Wei Lu[1], Qiao Liu[1], Nicole Mattson[1], Kevin Yining Chen[2], Jinhui Wang[3], Yate-Ching Yuan[3], David Horne[3], Steven T. Rosen[3], Yadira Soto-Feliciano[4], Zhaohui Feng[2], Takayuki Hoshii[2], Gang Xiao[1,5], Markus Müschen[1,6], Jianjun Chen[1,3], Scott A. Armstrong[2,*], Chun-Wei Chen[1,2,3,#]

Affiliations:

[1]Department of Systems Biology, Beckman Research Institute, City of Hope, Duarte, CA, USA

[2]Department of Pediatric Oncology, Dana-Farber Cancer Institute - Harvard Medical School, Boston, MA, USA

[3]City of Hope Comprehensive Cancer Center, Duarte, CA, USA

[4]Rockefeller University, New York, NY, USA

[5]Department of Immunology, Zhejiang University School of Medicine, Hangzhou, China

[6]Yale School of Medicine, New Haven, CT, USA

† These authors contributed equally to this work.

[#] Lead Contact and Correspondence:

Chun-Wei Chen, Ph.D.

Beckman Research Institute, City of Hope National Medical Center

1500 E. Duarte Rd, Duarte, CA 91010, USA

e-mail: cweichen@coh.org

*** Co-correspondence:**

Scott A. Armstrong, M.D., Ph.D.

Dana-Farber Cancer Institute - Harvard Medical School

450 Brookline Avenue, Boston, MA 02215, USA

e-mail: scott_armstrong@dfci.harvard.edu

Introductory Paragraph

Identification of novel functional domains and characterization of detailed regulatory mechanisms in cancer-driving genes is critical for advanced cancer therapy[1]. To date, CRISPR (clustered, regularly interspaced, short palindromic repeats) gene editing has primarily been applied to defining the role of individual genes[2,3]. Recently, high-density mutagenesis via CRISPR tiling of gene-coding exons has been demonstrated to identify functional regions in genes[4-8]. Furthermore, breakthroughs in combining CRISPR library screens with single-cell droplet RNA sequencing (sc-RNAseq) platforms have revealed the capacity to monitor gene expression changes upon genetic perturbations at single-cell resolution (e.g., Perturb-seq, CRISP-seq, CROP-seq)[9-11]. Here, we present "sc-Tiling," which integrates a CRISPR gene-tiling screen with single-cell transcriptomic and protein structural analyses. Distinct from other reported single-cell CRISPR screens focused on observing gene function and gene-to-gene/enhancer-to-gene regulation[11-16], sc-Tiling enables the capacity to identify regulatory mechanisms within a gene-coding region that dictate gene activity and therapeutic response.

Main Text

Recent achievements in cancer epigenetics include discovery of a central role for the histone 3 lysine 79 (H3K79) methyltransferase DOT1L in maintaining mixed lineage leukemia gene-rearranged (MLL-r) leukemia, an aggressive malignancy recognized in 5–10% of human acute leukemia cases[17,18]. A selective DOT1L-inhibitor, EPZ5676 (Pinometostat)[19], has demonstrated proof-of-principle clinical benefits via induction of differentiation of MLL-r leukemic cells in a Phase I clinical trial[20]. However, the variable responses of patients with MLL-r in this trial underscore the need for novel mechanistic insights into functional regions of DOT1L to improve therapeutic efficacy and trial designs for DOT1L-targeted therapy.

To achieve high-resolution characterization of DOT1L's function, we developed a single-cell CRISPR gene-tiling approach named sc-Tiling, which utilizes a capture sequence (CS1: 5'-GCTTTAAGGCCGGTCCTAGCA-3') at the end of each single guide RNA (sgRNA) for direct-capture by

the Chromium Next GEM Single Cell 3′ Kit v3.1 (Fig. 1a and S1a-c)[21]. We cloned a pool of 602 sgRNAs that target most of the "NGG" protospacer adjacent motifs (PAM) within the mouse *Dot1l* coding exons (average targeting density 7.7 bp per sgRNA; Fig. S2a,b and S3). We then delivered this CRISPR library into Cas9-expressing mouse MLL-AF9 transduced leukemic cells (MLL-AF9-Cas9^{+}; Fig. S2c,d), a well-established murine leukemia model that mimics human *MLL*-r conditions[22,23]. Three days after transduction, the cells carrying library constructs were subjected to droplet single-cell barcoding and mRNA/sgRNA library preparation using the 10X Chromium workflow (Fig. 1a). Subsequent single-cell transcriptomic analysis revealed an average of 26,350 reads per cell and a median of 2,935 genes detected per cell (Fig. S4). To avoid contamination by doublets and multi-sgRNA infected cells, we filtered out any single cells carrying more than one sgRNA sequence. Finally, 88.2% of single cells (4,362 out of 4,943) passed the quality control (QC) filter (Fig. 1b), giving an average library coverage of 7.1 cells per sgRNA.

Single-cell projections using Uniform Manifold Approximation and Projection (UMAP)[24] of DOT1L-dependent genes[23] identified seven cell clusters (Fig. 1c). Gene expression annotation revealed distinct distributions of cells expressing leukemia-associated genes (Meis1, Hoxa9, Myc; clustered toward the right) vs. myeloid-differentiation markers (Cd11b, Gr1, Ltf; clustered toward the left) (Fig. 1d). Cells expressing sgRNAs targeting the functionally essential lysine methyltransferase (KMT) core (residues M127–P332; total 56 sgRNA) of DOT1L[25,26] clustered to regions that overlap with the differentiated myeloid population (Fig. 1e). On the contrary, the sgRNAs targeting a non-essential region of DOT1L (the C-terminal end 100 amino acids of DOT1L; total 54 sgRNA) behaved similarly to spiked-in negative control sgRNAs (targeting *Firefly luciferase* [Luc], *Renilla luciferase* [Ren], *Green fluorescent protein* [GFP], *Red fluorescent protein* [RFP], and *Rosa26* coding sequences; Fig. S3), with both cluster to the region representing undifferentiated leukemia (Fig. S5a). Trajectory analysis (pseudo-time)[27] correlated closely with the expression of these marker genes, with leukemia-associated genes being gradually reduced while myeloid-differentiation markers increased along the pseudo-time trajectory (Fig. 1f, right to left; and S5b). These results indicate efficient CRISPR editing of DOT1L in cells expressing the CS1 direct-capturable sgRNA library.

To evaluate the resolution of sc-Tiling for detecting functional elements within a protein domain, we summarized the overall behavior of neighboring sgRNAs using a local-smoothing strategy[6] (Fig. 1g), and mapped the smoothened pseudo-time score to a cryo-electron microscopy (EM) structure of the DOT1L KMT core in an "active state" interacting with a histone H2B-ubiquitinylated nucleosome (Fig. 1h) [28,29]. Our results revealed that within the KMT core domain, the resolution of sc-Tiling allowed recognition of all the amino acid residues that directly contacted the enzymatic substrate S-adenosyl methionine (SAM) (Fig. S6a), including the D1 loop (residues P133–T139)[26]. This method also detected the critical regions within the KMT core domain that mediate its chromatin interaction. These include the W22–D32 loop (Fig. S6b; interacts with histone H4 tail), R282 loop (Fig. S6c; interacts with the histone H2A/H2B acidic patch), and T320–K330 helix (Fig. S6d; interacts with the ubiquitin conjugated to histone H2BK120)[28,29]. Taken together, sc-Tiling clearly distinguished the functional regions of KMT from the non-essential region (residues A33–T100) that is not involved in substrate/ligand interaction, revealing the capacity of single-cell CRISPR gene-tiling to pinpoint functional elements at a sub-domain resolution.

To identify novel functional elements that modulate DOT1L activity, we utilized the top 100 genes affected by DOT1L inhibitor[23] to develop a high-resolution transcriptomic correlation heatmap across DOT1L protein (Fig 2a). This method revealed two functionally distinct segments of DOT1L, i.e., the N-module (residues M1–T900) and the C-module (residues P901–N1537). The strong correlation of the sgRNAs targeting the C-module with the negative control sgRNAs (Fig S5a) indicates a lack of essential components in the C-terminal portion of DOT1L. On the other hand, we observed several functional regions of DOT1L within the N-module, including the KMT core (black dashed triangle)[26,28,29] and the AF9-binding motif (green dashed box; residues T863–T900)[30]. Whereas the AF9-binding motif showed a moderate correlation (Pearson score ~0.75) with the KMT core, we identified a novel region (cyan dashed box; designated as the "R domain") located in the center of the N-module that exhibited a higher correlation (Pearson score > 0.8) with the KMT core in the transcriptional signature. Based on this observation, we presumed that disrupting the function of the R domain would impair the survival of MLL-AF9 leukemia cells, similar to inhibition of the KMT core. To test this, we utilized the DOT1L-tiling CRISPR library to

perform pooled survival screens[5,7] in MLL-AF9-Cas9$^+$ cells and examined the cell survival by comparing the frequencies of each integrated sgRNA sequence before vs. after 3-, 6-, 9-, or 12-day cultures using high-throughput sequencing (Fig. 2b). Our results revealed a progressive depletion of clusters of sgRNAs (and smoothed CRISPR scan scores) targeting the KMT core, AF9-binding motif, and the first half of the R domain (designated as the "R1 element;" residues F460–G555). To confirm these results, we chose three sgRNAs each targeting the KMT core, AF9-binding motif, and the R1 element for functional validation (guide sequence shown in Fig. S7). Using an RFP flow cytometric growth competition assay (Fig. S8)[4] and immunoblotting, we observed that comparing to the sgRNAs targeting AF9-binding motif, expression of sgRNAs targeting the R1 element resulted in a more drastic suppression of cell proliferation (Fig. 2c) and impaired histone H3K79 methyltransferase activity (Fig. 2d), resembling the effects of sgRNAs targeting the KMT core. In addition to a similar UMAP distribution between cells expressing sgRNAs targeting the KMT core and the R1 element (Fig. 2e), sc-RNAseq revealed significantly overlapped gene regulation between these two sgRNA-targeted populations (Fig. 2f,g). These results indicate functional coordination between the DOT1L KMT core and R1 element for histone modification.

To investigate whether the R domain mediates the response of MLL-AF9 leukemia cells to DOT1L-inhibitory treatment, we compared a pair of pooled survival tiling screens conducted under control (DMSO) vs. DOT1L-inhibited (1 uM EPZ5676) conditions (Fig. 3a). Consistent with the results of the sc-Tiling, we observed that a cluster of 27 sgRNAs targeting the R1 region (residues F460–G555) sensitized the MLL-AF9-Cas9$^+$ cells to DOT1L inhibition (Fig. 3b and S9). By contrast, a cluster of 36 sgRNAs targeting the residues A558–C662 (designated as the "R2 element") exhibited a significantly increased CRISPR score only in the DOT1L-inhibited condition (Fig. 3a,b). The expression of individual sgRNAs targeting the R2 element exhibited minimal impact on the proliferation of MLL-AF9-Cas9$^+$ cells (Fig. 3c) but increased the resistance index to the DOT1L inhibitor (Fig. 3d and S9), confirming the EPZ5676-resistant phenotype we observed in the CRISPR gene body scans. Computational modeling of the R domain (residues F460–C662) revealed a consensus "coiled-coil" structure consisting of four alpha-helixes (Fig. 3e), which is capable of interacting with the KMT core domain of DOT1L (Fig. 3f). Within the R domain,

the R1 element (consisting of CC0 and CC1) overlaps with an area previously reported to interact with AF10[31,32], a co-activator of DOT1L required for methyltransferase activation. On the other hand, the R2 element (consisting of CC2 and CC3) is predicted to interact with the DOT1L KMT core and masks the R282 loop (Fig. 3f), thereby interrupting the DOT1L-nucleosome interaction and methyltransferase activity of the KMT core. This model suggests the R domain mediates the transition from a "closed" to an "open" state of DOT1L (Fig. 3g; left to right), which is required before the engagement of the KMT core with nucleosomes for H3K79 methylation (Fig. 3g; blue area summarized in Fig. 1h).

To evaluate the impact of this novel self-regulatory mechanism on DOT1L-targeted therapy, we queried the cBioPortal database[33] and focused on the R2 element (residues A558–C662) that exerted a robust EPZ5676-resistant phenotype in the CRISPR scan. Out of a total of 54,510 patient samples, we found nineteen DOT1L variant alleles to exist in this 105-amino acid region (Fig. S10a). Compared to expression of wild-type-DOT1L, expression of several mutant-DOT1L constructs (each harbors a single amino acid missense mutation) in MLL-AF9 cells resulted in an increased resistance to EPZ5676 treatment (Fig. 3h and S10b). We then focused on the top three drug-resistant variants (Q584P, L626P, and C637G) and found that these mutant-DOT1L led to an elevated H3K79me2 and required a higher dosage of EPZ5676 to suppress their activity compared to wild-type-DOT1L (Fig. 3i,j). Computational modeling of these drug-resistant variants indicates that mutations at these residues may destabilize alpha-helix bundles and lead to dissociation of the R domain from the KMT core, resulting in increased kinetic activity and tolerance to DOT1L-inhibitory therapy (Fig S11).

Discussion

The integration of CRISPR with next-generation sequencing technology for high-throughput genetic screens has been established as a powerful tool for discovering novel functional genes in various cellular contexts[2,3]. In contrast, the potential of CRISPR technology to investigate gene function at a sub-gene (i.e., protein domain or sub-domain) resolution has not been fully explored. Furthermore, traditional

pooled CRISPR screens limit the ability to identify functional elements associated with cell killing/proliferation phenotypes (i.e., by observing the depletion or enrichment of specific sgRNA). The requirement for significant changes in cell number in survival CRISPR screens (which typically take 2-4 weeks of culture) prohibits the determination of causal mechanisms induced by CRISPR perturbation.

To overcome this obstacle, our study integrated a CRISPR gene-tiling screen with a recently available direct-capture Perturb-seq workflow[21] to develop the single-cell CRISPR gene body-scan pipeline sc-Tiling. Using this approach, we provide the first high-resolution transcriptomic correlation map across DOT1L, an epigenetic therapeutic candidate essential to MLL-r leukemia[17,23,25]. Although the limitations of CRISPR genome editing (e.g., variable cutting efficiency, potential for off-targeting, and the mosaic effect [i.e., generation of random mutations]) remain concerns in the CRISPR sc-Tiling approach, by considering multiple sgRNAs clustered in a peptide region via a local-smoothing strategy, we significantly increased the statistical confidence and minimized the impact of noise associated with individual sgRNAs. Importantly, the use of single-cell transcriptional profiling in sc-Tiling could predict functional elements and corresponding gene regulations that led to a cellular survival phenotype after prolonged culture, and provided superior resolution in detecting sub-domain functional elements than survival CRISPR gene-tiling screens using pooled sequencing (i.e., Fig 1g vs. 2b; KMT core). Furthermore, when we coupled sc-Tiling with three-dimensional structural modeling, we discovered a previously unrecognized self-regulatory R domain in DOT1L that modulates chromatin interaction, enzymatic activation, and therapeutic sensitivity in MLL-r leukemia. To our knowledge, this is the first characterization of an intragenic regulatory module that mediates switching between a “closed” and an “open” state of an epigenetic enzyme.

Finally, our study demonstrates the utility of combining sc-Tiling with consortium genomic databases (e.g., cBioPortal) for *de novo* identification of therapeutically relevant alleles in the human population (Fig. 3h). We propose that single-cell CRISPR gene tiling may complement the rapidly growing multi-omics databases to provide new insights that bridge functional genomics, structural biology, and clinical investigation. We envision this approach will accelerate the recognition of clinically impactful

variants within the human genome and has the potential to direct more precise clinical trials and therapeutic decisions.

Acknowledgments:

This work was supported by the American Society of Hematology (to C.-W.C.), Alex's Lemonade Stand Foundation (to C.-W.C.), Leukemia & Lymphoma Society (to J.C. and S.T.R.), and National Institutes of Health Grants CA176745, CA206963 (to S.A.A.), CA197498, CA233691, CA236626 (to C.-W.C.). The sequencing and structural computational studies were supported by the National Institutes of Health P30 award CA033572 (City of Hope). We thank Dr. Sarah Wilkinson for editing the manuscript.

Author contributions:

L.Y., A.K.N.C., K.M., C.D.D., X.W., S.P.P., M.L., X.X., Q.L., N.M., K.Y.C, J.W., Y.S.-F., Z.F., and G.X. performed the experiments; L.Y., A.K.N.C., H.L., S.L., W.L., Y.-C.Y., D.H., and C.-W.C. analyzed the data; D.H., S.T.R., T.H., M.M., J.C., S.A.A., and C.-W.C. provided conceptual input; L.Y., A.K.N.C., K.M., H.L., S.A.A., and C.-W.C. wrote the paper; S.A.A. and C.-W.C. conceived and supervised the study.

Declaration of interests:

The authors declare no competing interests.

Data availability statement

The data that support the findings of this study are available from the corresponding author upon reasonable request.

Availability of biological materials:

Direct-capturable lentiviral sgRNA vector will be deposited to Addgene. Cas9-expressing MLL-AF9 leukemia cells will be available upon request. CRISPR gene-tiling library for mouse Dot1l will be available

upon request. All other biological materials are commercially available.

Code availability statement

The computational codes/tool packages used in this study are available through developers and venders, including Genetic Perturbation Platform (Broad Institute)[34], Seurat v3.0[24], Monocle[27], Gaussian kernel smoothing in R[35], CLC Main Workbench version 8.1 (QIAGEN), Bowtie2[36], edgeR R package[37], PSIPRED v3.3 server[38], MultAlin v5.4.1 server[39], I-TASSER server[40], ZDOCK v3.0.2 software[41], PyMOL v1.8.6 software (Schrödinger, LLC.), UCSF Chimera[42], Research Collaboratory for Structural Bioinformatics Protein Data Bank (RCSB PDB)[43], and cBioPortal database[33].

Figure 1

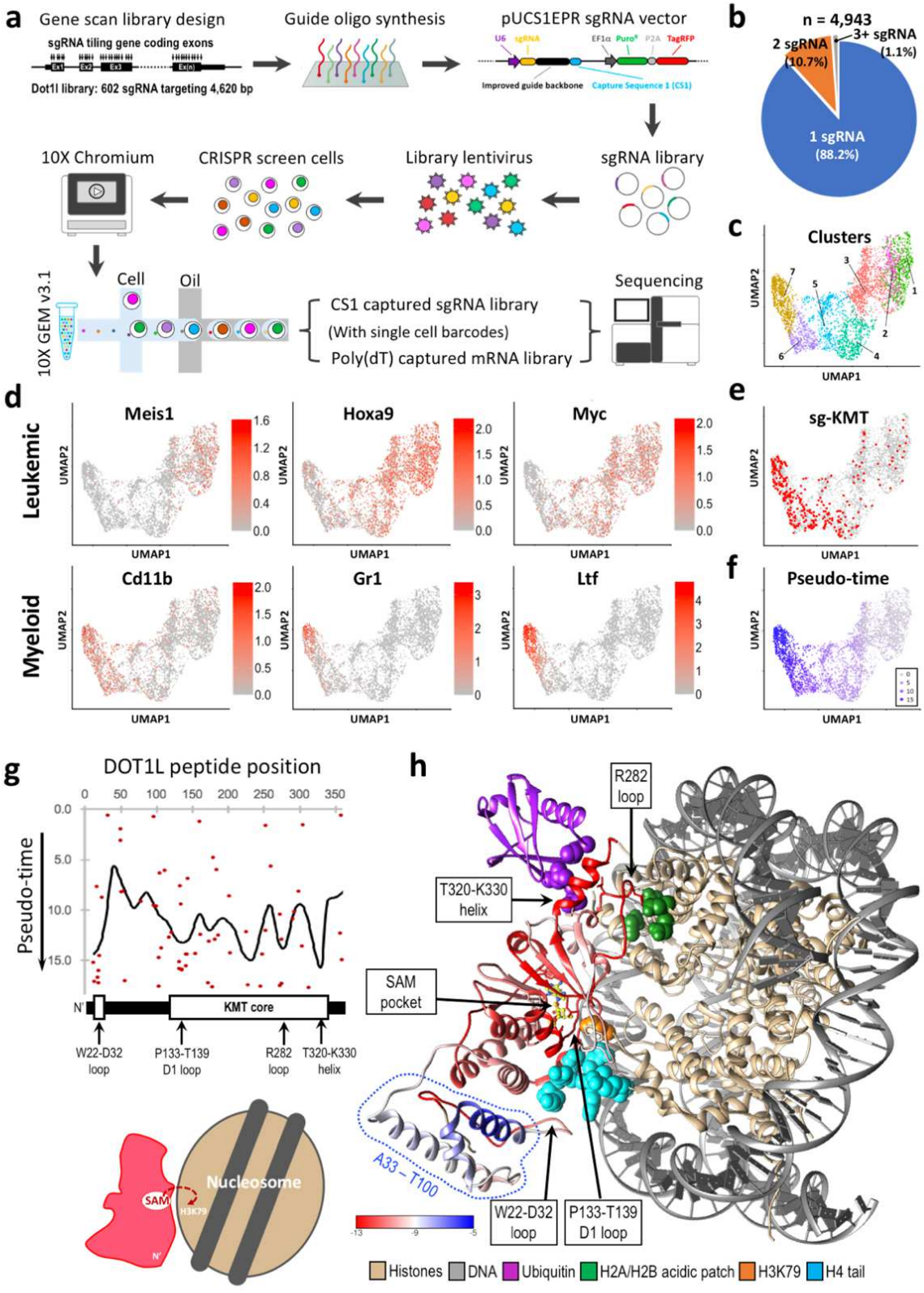

a
Gene scan library design
sgRNA tiling gene coding exons
Dot1l library: 602 sgRNA targeting 4,620 bp
Guide oligo synthesis
pUCS1EPR sgRNA vector
U6
sgRNA
EF1α
Puro^R
P2A
TagRFP
Improved guide backbone
Capture Sequence 1 (CS1)
sgRNA library
Library lentivirus
CRISPR screen cells
10X Chromium
10X GEM v3.1
Cell
Oil
CS1 captured sgRNA library
(With single cell barcodes)
Poly(dT) captured mRNA library
Sequencing
b
n = 4,943
2 sgRNA (10.7%)
3+ sgRNA (1.1%)
1 sgRNA (88.2%)
c
Clusters
UMAP1
UMAP2
d
Leukemic
Meis1
Hoxa9
Myc
Myeloid
Cd11b
Gr1
Ltf
e
sg-KMT
f
Pseudo-time
g
DOT1L peptide position
Pseudo-time
N'
KMT core
W22-D32 loop
P133-T139 D1 loop
R282 loop
T320-K330 helix
SAM
H3K79
Nucleosome
h
R282 loop
T320-K330 helix
SAM pocket
A33 – T100
W22-D32 loop
P133-T139 D1 loop
Histones
DNA
Ubiquitin
H2A/H2B acidic patch
H3K79
H4 tail

Figure 1. Single-cell CRISPR gene tiling of DOT1L. (a) Schematic outline of sc-Tiling library construction and screening in MLL-AF9-Cas9^{+} cells. (b) Assignment rates for direct-capture sgRNA. The total number of cells and fraction of cells assigned a single guide, two guides, or more than two guides are indicated. (c) Two-dimensional projection (UMAP) of cell clusters based on sc-RNAseq of DOT1L-dependent genes. The transcriptionally distinguishable cell populations (1 to 7) are color labelled. (d) Annotation of leukemia-associated (Meis1, Hoxa9, Myc) and myeloid-differentiation (Cd11b, Gr1, Ltf) gene expression on UMAP. (e-f) Annotation of (e) cells harboring sgRNAs targeting the DOT1L KMT core (red) and (f) pseudo-time value (purple gradient) on UMAP. (g) Median pseudo-time of each sgRNA construct (dots) and the smoothed pseudo-time score (line) of the KMT core. (h) Three-dimensional annotation of smoothed pseudo-time score relative to a cryo-EM structural model of "active state" DOT1L (residues M1–P332) bound to a ubiquitylated nucleosome (PDB ID: 6NQA; and a simplified scheme shown on the bottom-left)[28]. Histones (gold; including H2A, H2B, H3 and H4), DNA (grey), ubiquitin (purple; conjugated to histone H2BK120; DOT1L contact points on ubiquitin are labeled as purple spheres), histone H4 N-terminal tail (cyan spheres), the enzymatic substrate SAM (colored sticks), histone H3K79 (orange spheres), and an H2A/H2B acidic patch (green spheres) are shown. Enlarged images shown in Fig. S6.

Figure 2

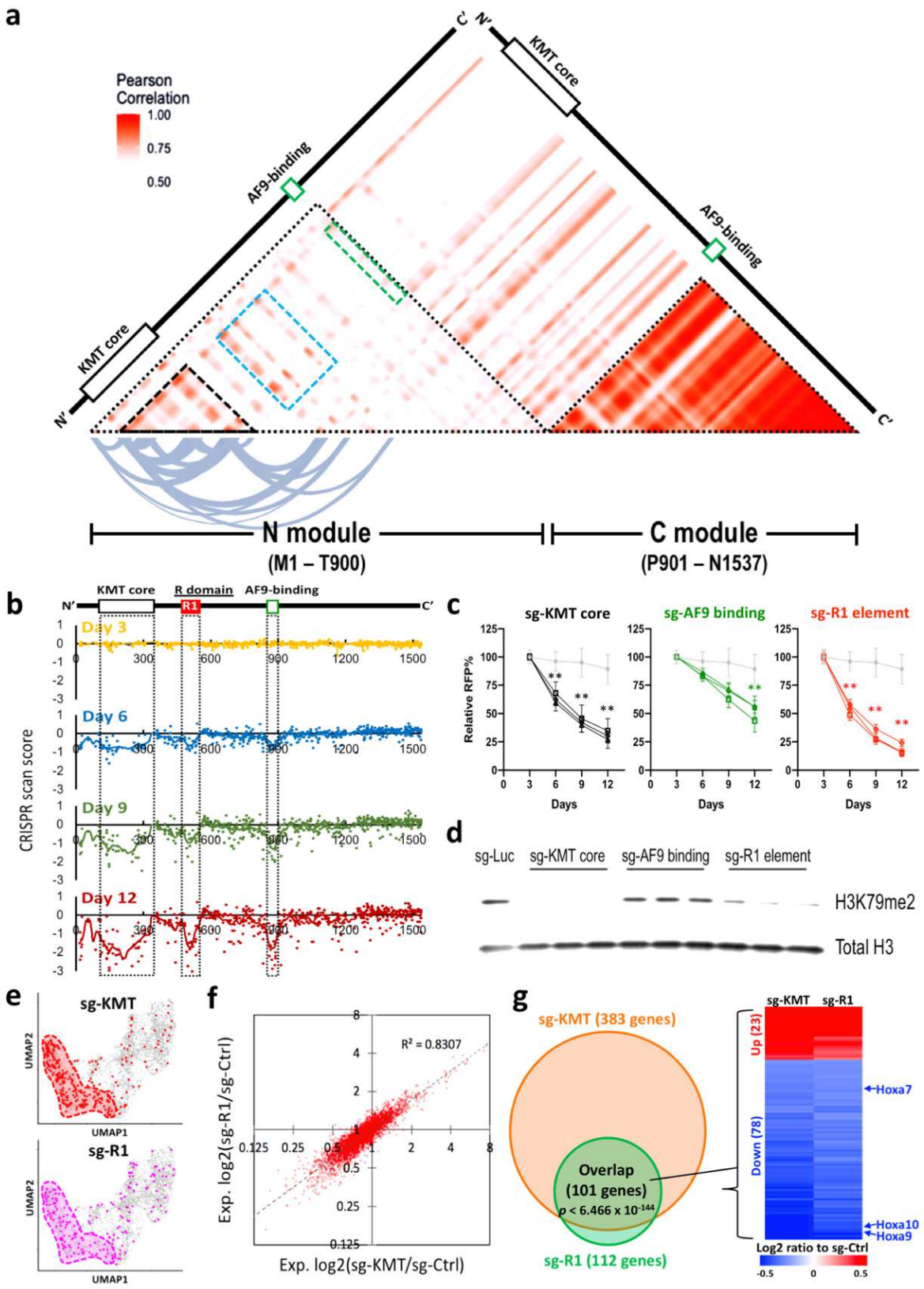

a
Pearson Correlation
1.00
0.75
0.50
KMT core
AF9-binding
AF9-binding
KMT core
N module
(M1 – T900)
C module
(P901 – N1537)
b
KMT core
R domain
AF9-binding
R1
Day 3
Day 6
Day 9
Day 12
CRISPR scan score
c
sg-KMT core
sg-AF9 binding
sg-R1 element
Relative RFP%
Days
d
sg-Luc
sg-KMT core
sg-AF9 binding
sg-R1 element
H3K79me2
Total H3
e
sg-KMT
sg-R1
UMAP1
UMAP2
f
R² = 0.8307
Exp. log2(sg-R1/sg-Ctrl)
Exp. log2(sg-KMT/sg-Ctrl)
g
sg-KMT (383 genes)
Overlap
(101 genes)
p < 6.466 x 10-144
sg-R1 (112 genes)
sg-KMT
sg-R1
Up (23)
Down (78)
Hoxa7
Hoxa10
Hoxa9
Log2 ratio to sg-Ctrl

Figure 2. sc-Tiling pinpoints novel functional elements in DOT1L. (a) Heatmap depicts Pearson correlations between sgRNAs targeting different positions across the DOTL1 protein. The curved lines indicate highly correlative (Pearson score > 0.8) residue pairs in the N-module of DOT1L. (b) CRISPR scan score of each sgRNA (dots) and smoothed score (line) of the DOT1L-tiling survival screen in MLL-AF9-Cas9$^+$ leukemia at indicated number of days in culture. (c) Effect of individual sgRNAs targeting the KMT core (black), AF9-binding motif (green), or R1 element (red) of DOT1L on the proliferation of MLL-AF9-Cas9$^+$ leukemia. Data represent mean ± 95% confidence interval of a quadruplicate experiment. **P < 0.001 by two-sided Student's t-test compared to an sgRNA targeting Luciferase (sg-Luc; grey). (d) Western blot of H3K79me2 and total histone H3 in MLL-AF9-Cas9$^+$ cells expressing indicated sgRNAs. (e) Annotation of cells harboring sgRNAs targeting the KMT core (red) or R1 element (pink) on UMAP. (f) Correlation of gene expression changes induced by sgRNAs targeting the KMT core (x-axis) and R1 element (y-axis) summarized from the sc-Tiling of DOT1L. (g) Overlap of differentially expressed genes in cells harboring sgRNAs targeting the KMT core (orange) and R1 element (green), including the known DOT1L-driven leukemia genes Hoxa7, Hoxa9, and Hoxa10.

Figure 3

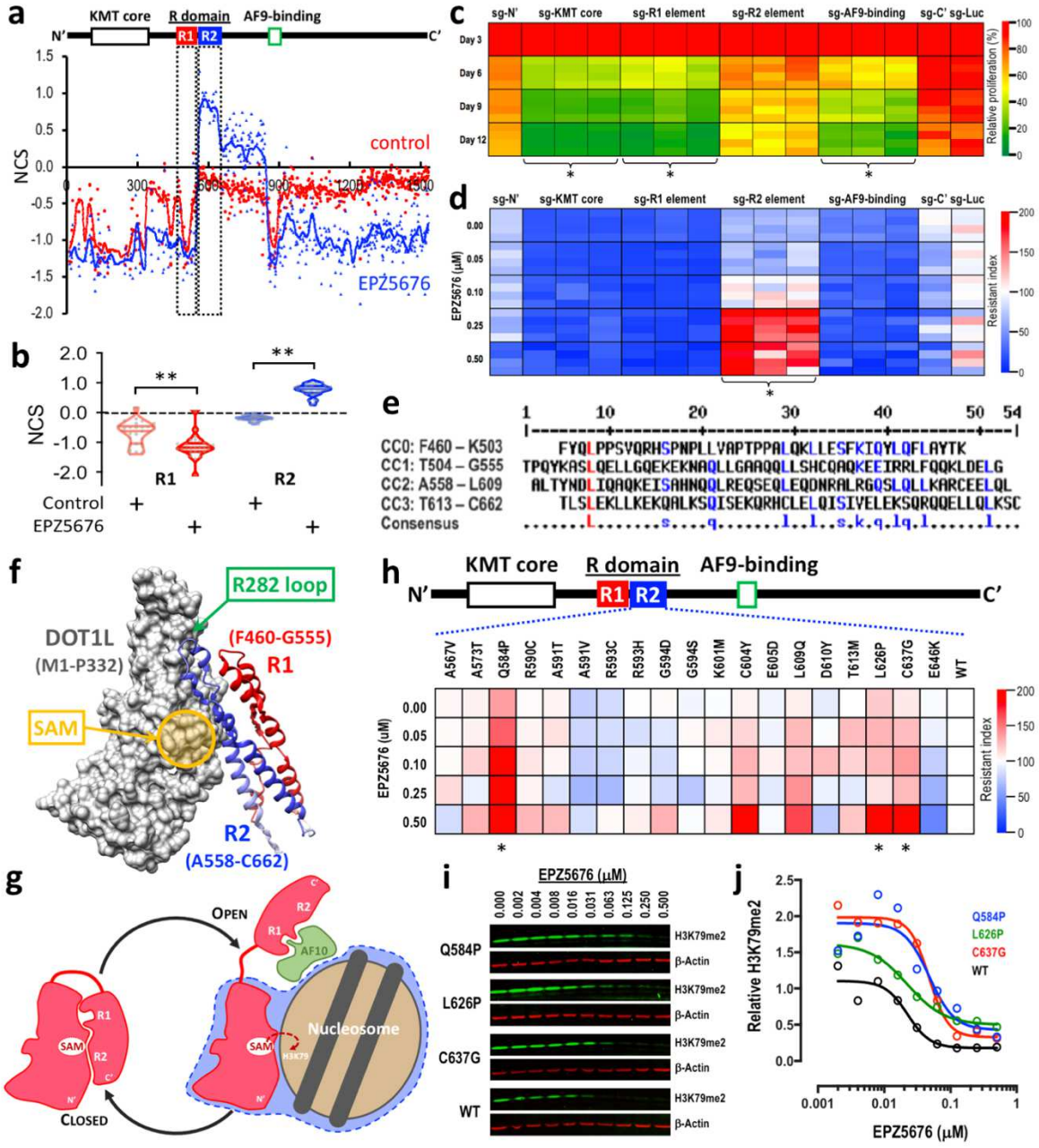


Figure 3. sc-Tiling identifies noncanonical EPZ5676-resistant alleles in human population. (a) Normalized CRISPR score (NCS) of each sgRNA construct (dots) and the smoothed score (line) of the pooled DOT1L-tiling survival screen before vs. after 12 days of treatment in control (red) or 1 μM EPZ5676-treated (blue) MLL-AF9-Cas9^{+} leukemia cells. Data represent the average of a triplicate

experiment. (b) Violin dot plots showing the NCS of each sgRNA targeting the R1 (red; 27 sgRNAs) and R2 (blue; 36 sgRNAs) elements in control (DMSO) or 1 μM EPZ5676-treated MLL-AF9-Cas9$^+$ leukemia cells. **P < 0.001 by two-sided Student's t-test. (c) Heatmap showing the effect of individual sgRNAs targeting the indicated areas of DOT1L (Fig. S7) on the proliferation of MLL-AF9-Cas9$^+$ leukemia cells on days 3, 6, 9, and 12. Data represent the observed values of a quadruplicate experiment. *Significantly (P < 0.01 by two-sided Student's t-test) more depletion compared to the sgRNA targeting a non-essential N' region (sg-N') on day 12. (d) Heatmap showing EPZ5676 resistance index of MLL-AF9-Cas9$^+$ leukemia cells transduced with sgRNAs targeting the indicated areas of DOT1L (Fig. S7). Data represent the observed values of a quadruplicate experiment. *Significantly (P < 0.01 by two-sided Student's t-test) higher resistance compared to sg-N' at 0.5 μM EPZ5676. (e) Peptide sequence alignment of the R domain (residues F460–C662) in human DOT1L. The predicted alpha-helixes in this coiled-coil domain are designated CC0–CC3 and the consensus residues between the helixes are noted. (f) Computationally modeled structure of the human DOT1L R1 (red) and R2 (blue) coiled-coil domains interacting with the KMT core domain (grey; PDB ID: 3UWP)[44]. (g) Cartoon representation of the R1/R2 self-regulatory module mediating the closed (left) vs. open (right) states of DOT1L. (h) Heatmap showing EPZ5676 resistance index of MLL-AF9 leukemia cells transduced with human DOT1L cDNA harboring clinically observed variants (from cBioPortal database) in the R2 element. Data represent the averaged values of a quadruplicate experiment. *Significantly (P < 0.01 by two-sided Student's t-test) higher resistance compared to wild-type at 0.5 μM EPZ5676. (i) Western blot images of H3K79me2 (green) and β-Actin (red) and (j) quantitative measurement of relative H3K79me2 level (normalized to β-Actin) in MLL-AF9 leukemia cells transduced with wild-type (WT; black), Q584P (blue), L626P (green), or C637G (red) human DOT1L cDNA. Cells were treated with EPZ5676 for three days.

Online Methods

Cas9-expressing MLL-AF9 leukemic cell culture

Mouse MLL-AF9 leukemic cells were generated by transformation of mouse bone marrow $Lin^{-}Sca1^{+}cKit^{+}$ (LSK) cells with a MIG (MSCV-IRES-GFP) retrovirus expressing the MLL-AF9 fusion protein and transplanted into sublethally irradiated recipient mice as described previously[22]. Leukemic blasts were subsequently harvested from the diseased mice and cultured *in vitro* in Iscove's Modified Dulbecco's Medium (IMDM; Gibco) plus 15% FBS (Gibco) supplemented with 20 ng/ml mouse SCF (PeproTech), 10 ng/ml mouse IL-3 (PeproTech), 10 ng/ml mouse IL-6 (PeproTech), penicillin (100 units/ml; Gibco), streptomycin (100 μg/ml; Gibco), and plasmocin (5 μg/ml; InvivoGen). Cas9-expressing MLL-AF9 cells were established through lentiviral transduction of LentiCas9-Blast (Addgene)[45] followed by Blasticidin S (10 mg/ml; Gibco) selection, single-cell cloning, and CRISPR editing efficiency test (Fig. S2C,D).

CRISPR gene-tiling screens

sgRNA sequences targeting the coding regions of mouse *Dot1l* (Fig. S3) were designed using the Genetic Perturbation Platform (Broad Institute)[34]. Briefly, sgRNA oligonucleotides were synthesized via microarray (CustomArray) and cloned into the pUCS1EPR lentiviral sgRNA vector (Fig. S1A) using BsmBI (NEB)[34]. The sgRNA library was packaged by HEK293 cells (ATCC) co-transfected with psPAX2 (Addgene) and pMD2.G (Addgene) to produce lentiviral particles, and pre-titrated to obtain 10–20% infection (monitored by flow cytometry for red fluorescent protein [tagRFP] expression) in the MLL-AF9-$Cas9^{+}$ cells. Each screen culture was calculated to maintain at least 1,000x the number of constructs in each library. For sc-Tiling, library transduced cultures were selected using puromycin (2.5 μg/ml; Gibco) for 3 days and subjected to single-cell separation and barcoding using a Chromium Controller (10X Genomics). For survival CRISPR gene tiling, the sgRNA library-transduced cells were subcultured every 3 days for a total of 12 days. At each designated time point, the number of cells from cultures that covered at least 1,000x the number of constructs in the library were collected for analysis.

Single-cell CRISPR gene tiling (sc-Tiling) data analysis

Using the Next GEM Single Cell 3' Kit v3.1 and a Chromium Controller (10X Genomics), CS1-captured sgRNA and the poly(dT)-captured mRNA from each single cell were converted to next-generation sequencing libraries (Fig. S1C), and sequenced (paired-end 150 base pair) using Illumina HiSeqX (Novogene Inc.). Sequencing QC and data preprocessing were performed using Seurat v3.0[24]. Low-quality single cells with abnormal gene numbers (less than 200 or more than 4,500) or significant mitochondrial RNA contamination (greater than 10% reads) were removed (Fig. S4A). The normalized expression data from selected single cells then underwent dimensionality reduction by PCA and UMAP embeddings for visualization and clustering. Cells were clustered based on the poly(dT)-captured transcriptome information and simultaneously annotated by CS1-captured sgRNA. Single cells with more than one detected sgRNA sequence (due to multiple sgRNA transductions or multiple cells in a single-cell droplet) were excluded. Pseudo-time trajectory analysis of the DOT1L inhibitor-affected genes was performed on single-cell transcriptomic data using Monocle[27]. Position-ordered Pearson correlation matrix across the *Dot1l* gene body was calculated based on the top 100 genes affected by DOT1L inhibition.

Three-dimensional protein structural annotation of sc-Tiling

First, the median value pseudo-time projection generated from sc-Tiling was summarized for each sgRNA. To depict the pseudo-time score over regions with no sgRNA coverage, we interpolated the signal via Gaussian kernel smoothing in R[35]. The bandwidth was defined by the maximum gap length of the non-covered regions for local smoothing due to regional uneven sgRNA densities. To map the smoothed pseudo-time score to peptide positions, the average pseudo-time score over the trinucleotide codons was calculated for each peptide position. Pairwise alignments of primary amino acid sequences were performed using CLC Main Workbench version 8.1 (QIAGEN) to ensure functional annotations of the smoothed pseudo-time scores of mouse Dot1l sc-Tiling data onto human DOT1L protein structures. Atomic data of macromolecular structures were retrieved from the Research Collaboratory for Structural Bioinformatics Protein Data Bank (RCSB PDB at https://www.rcsb.org)[43] in Protein Data Bank (PDB) file format. The

PDB files were visualized and analyzed using UCSF Chimera (version 1.14 build 42000)[42]. Subsequently, the smoothed pseudo-time scores were mapped onto three-dimensional protein structures using the "Defined Attribute" and "Render by Attribute" functionalities in UCSF Chimera[42].

Survival CRISPR gene tiling data analysis

Genomic DNA from survival screen cell pellets was harvested, PCR-amplified (NEBNext Ultra II Q5; NEB) using primers DCF01 5'-CTTGTGGAAAGGACGAAACACCG-3' and CS1_R01 5'-TGCTAGGACCGGCCTTAAAGC-3' (Fig. S1A), and subjected to high-throughput sequencing (NextSeq550, Illumina). To quantify sgRNA reads in the library, we first extracted 20-nucleotide sequences that matched the sgRNA backbone structure (5' prime CACCG and 3' prime GTTT) from raw fastq reads. Extracted reads were then mapped to a reference database built from corresponding sgRNA library sequences using Bowtie2[36]. Only reads that perfectly matched to the reference database were counted. The frequency for individual sgRNAs was calculated as the read counts of each sgRNA divided by the total read counts matched to the library. Individual sgRNAs with read counts less than 5% of the expected frequency were excluded from downstream analysis. A CRISPR score was defined as a log10-fold change in the frequency of individual sgRNAs between early (day 0) and late (defined time points) of the screened samples, calculated using the edgeR R package[37] based on the negative binomial distribution of sgRNA read count data. To obtain a CRISPR scan score over regions with no sgRNA coverage, we interpolated the signal via Gaussian kernel smoothing in R[35]. Bandwidth was defined as the maximum gap length of the non-covered regions for local smoothing due to regional uneven sgRNA densities. To map CRISPR scan scores to peptide positions, the average CRISPR scan score over the trinucleotide codons was calculated for each peptide position. To compare survival screens performed in different culture conditions (e.g., control vs. EPZ5676-treated), the smoothed CRISPR scan score was further normalized by the median CRISPR score of the negative control sgRNA (defined as 0.00; sgRNA targeting Luc, Ren, GFP, RFP and Rosa26) and the median CRISPR score of the positive control sgRNA (defined as -1.00; sgRNA targeting mRpa3)[4] within the screen data.

Computational structural modeling

Four helices (CC0–CC3) of the R domain were predicted using the PSIPRED v3.3 server[38]. Sequence alignment of the helical regions (Fig. 3E) was produced using the MultAlin v5.4.1 server[39]. The model of the coiled-coil domain was predicted using the I-TASSER server[40]. The complex model of the R domain and KMT core domain (PDB ID: 3UWP)[44] was picked from 5,000 complex models generated using ZDOCK v3.0.2 software[41]. The best model (Fig. 3F) was selected based on the largest number of hydrophobic contact residue pairs between the KMT core and R domain. The structures were visualized using PyMOL v1.8.6 software (Schrödinger, LLC.) and UCSF Chimera[42].

Generation of human DOT1L variant cDNA expression constructs

A MIY (MSCV-IRES-YFP) retroviral construct expressing wild-type human DOT1L and yellow fluorescent protein (YFP) was obtained from Dr. Yi Zhang[26]. The initial wild-type human DOT1L cDNA (MIY-DOT1L-WT) was then point-mutated to obtain 19 clinically observed DOT1L variants (Fig. S10) using the Q5 Site-Directed Mutagenesis Kit (NEB). The mutated DOT1L cDNA fragments were confirmed using Sanger sequencing (Eton Bioscience).

Western blotting

Cells were harvested and lysed in LDS sample buffer (Invitrogen) at $5x10^6$ cells/mL, separated electrophoretically using Bolt 4–12% Bis-Tris plus gels (Invitrogen), and transferred onto PVDF membranes (0.2 μm pore size, low fluorescence) using PVDF Mini Stacks and iBlot 2 (Invitrogen). Membranes were probed with rabbit anti-H3K79me2 antibody (D15E8, Cell Signaling Technology; 1:1,000), rabbit anti-histone H3 (ab1791, Abcam; 1:10,000) and mouse anti-β Actin antibody (ab8226, Abcam; 1:1,000) at 4°C overnight. After washing, the membranes were incubated with HRP-linked goat anti-rabbit IgG antibody (CST7074, Cell Signaling Technology; 1:10,000), donkey anti-rabbit IgG antibody conjugated with Alexa Fluor 488 (ab150061, Abcam; 1:10,000) or donkey anti-mouse IgG antibody

conjugated with Cy3 (AP192C, Sigma-Aldrich; 1:10,000) at room temperature for 1 hour. Chemiluminescent signals were developed using the SuperSignal West Femto Substrate (Cat# 34095, ThermoFisher). The chemiluminescent and fluorescent signals on Western blot membranes were detected using a ChemiDoc imaging system (Bio-Rad). Signal intensity from image files was analyzed using ImageJ software (National Institutes of Health).

Growth competition assay

Cas9-expressing MLL-AF9 cells were virally transduced with the designated constructs (RFP^+ ipUSEPR lentiviral sgRNA constructs listed in Fig. S7; YFP^+ MIY retroviral DOT1L variant cDNA constructs listed in Fig. S10) in 96-well plates at ~50% infection and monitored using flow cytometry for RFP or YFP (FP). At each time point, live cell counts and the percentage of FP^+ cells (*FP%*) were obtained by high-throughput flow cytometry and 4',6-diamidino-2-phenylindole (DAPI; Invitrogen) dye exclusion using an Attune NxT flow cytometer with an autosampler (ThermoFisher).

The *Relative Proliferation (RP)* of FP^+ (sgRNA- or DOT1L cDNA-expressing) vs. FP^- (non-transduced) cells was defined as:

$$Relative\ proliferation\ (RP) = \frac{[N(t)\ \times\ FP\%(t)]\ \times\ \left[N(d3)\ \times\ \left(100 - FP\%(d3)\right)\right]}{[N(d3)\ \times\ FP\%(d3)]\ \times\ \left[N(t)\ \times\ \left(100 - FP\%(t)\right)\right]}$$

where *N(t)* and *FP%(t)* are the observed live cell number and FP^+% at time point *t*; *d3* denotes the day-3 time point.

The *Resistance Index* was defined as:

$$Resistance\ Index = \frac{RP(x,m)}{RP(con,m)}\ \times\ 100\%$$

where *RP(x,m)* is the *RP* of cells expressing sgRNA or DOT1L cDNA variant *x* under *m* μM of EPZ5676 (Selleck Chemicals) on day-9; *con* denotes the sg-Luc or wild-type DOT1L cDNA.

Supplemental Information:

Figure S1

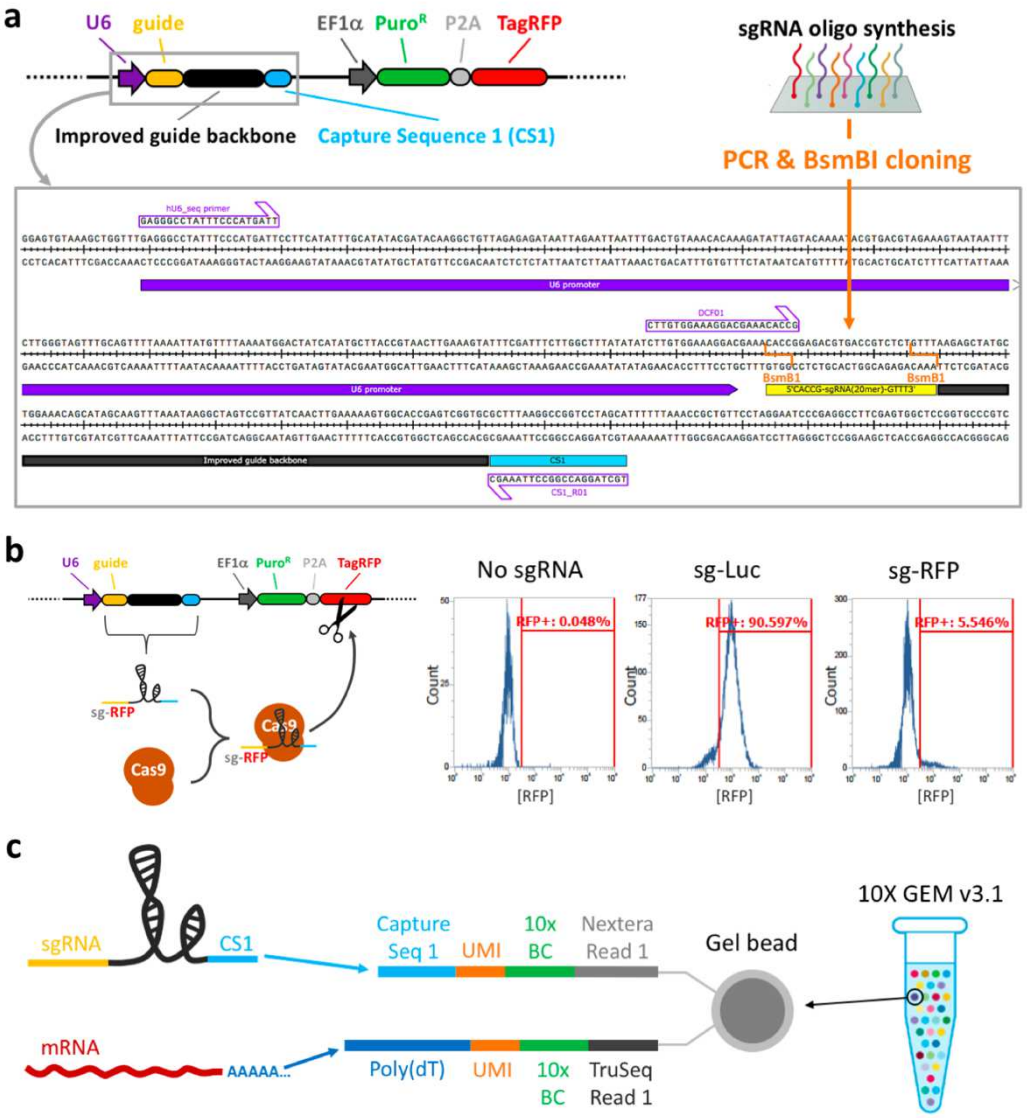


Figure S1. Schematic outline of the single-cell CRISPR screen system. (a) Sequence map of the pUCS1EPR vector expressing a CS1-capturable sgRNA together with a puromycin-resistant gene ($Puro^R$) and a TagRFP fluorescent protein. Primers for Sanger (hU6_seq) and Illumina (DCF01 and CS1_R03) sequencing are listed. (b) Validation of CRISPR editing efficiency of the CS1-capturable sgRNA in Cas9-expressing cells using an RFP inactivation assay. (c) Cartoon illustration of the CS1 and poly(dT) capture mechanisms for single-cell CRISPR and mRNA sequencing.

Figure S2

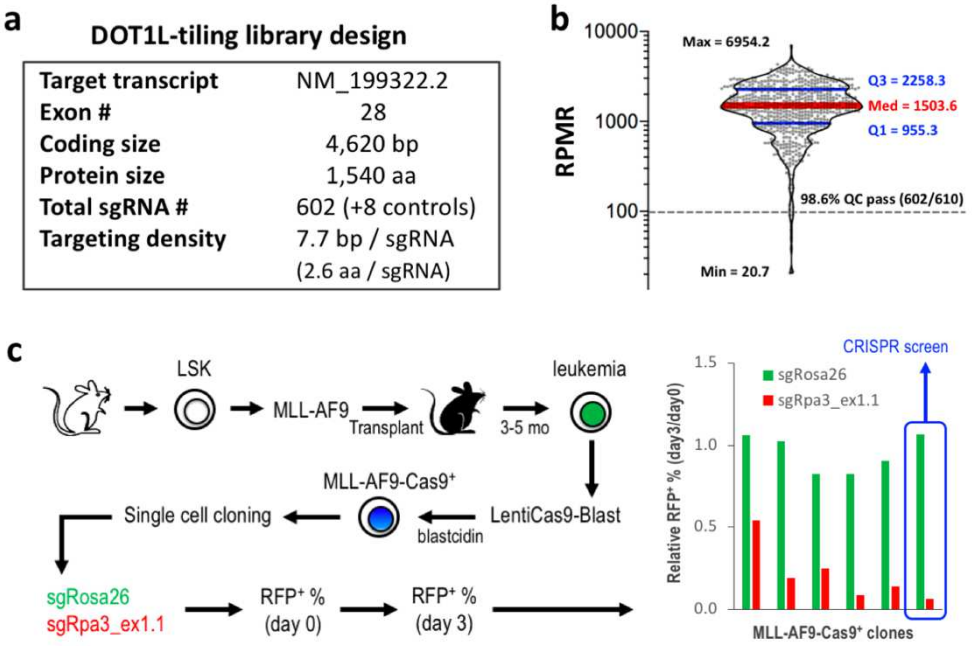
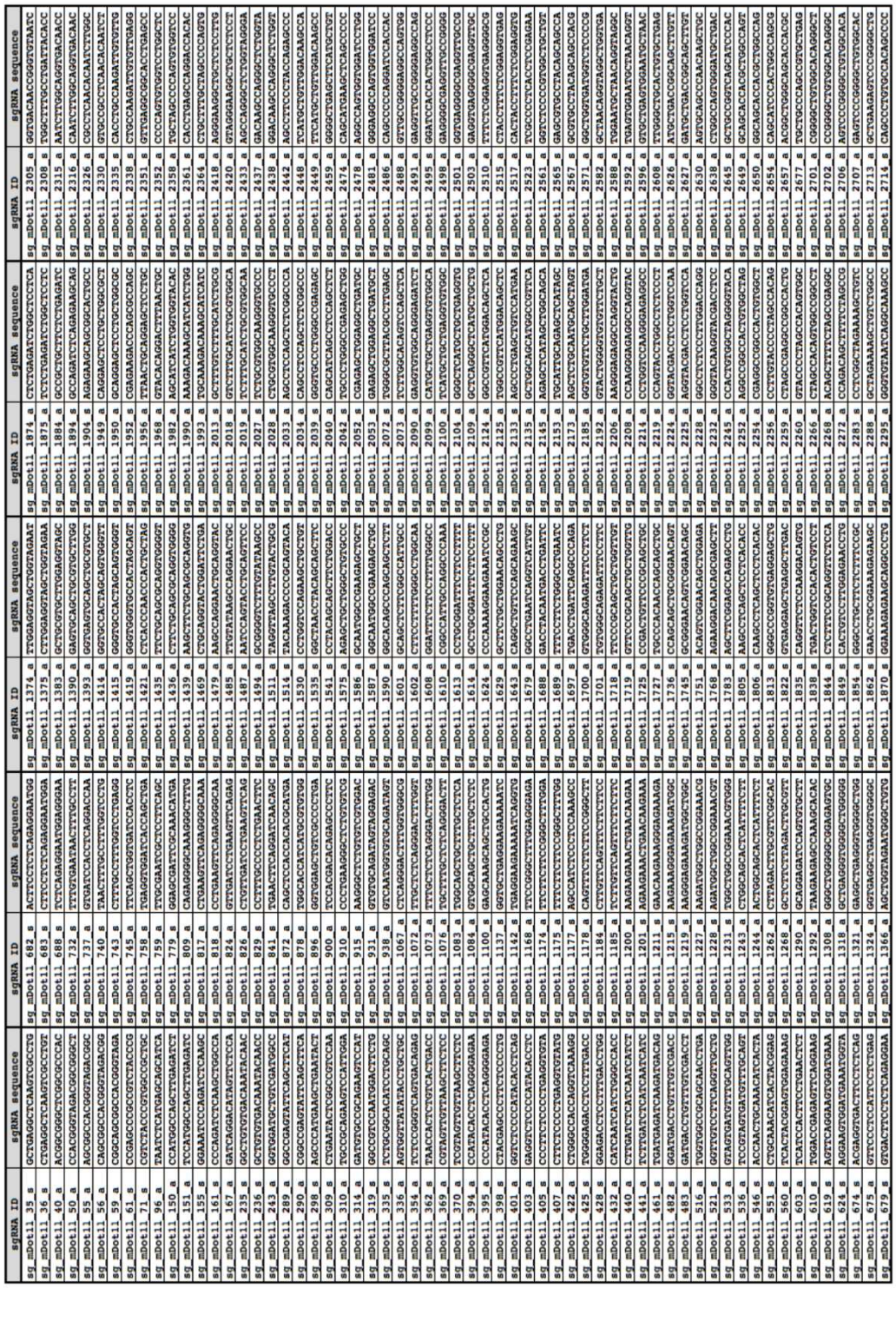


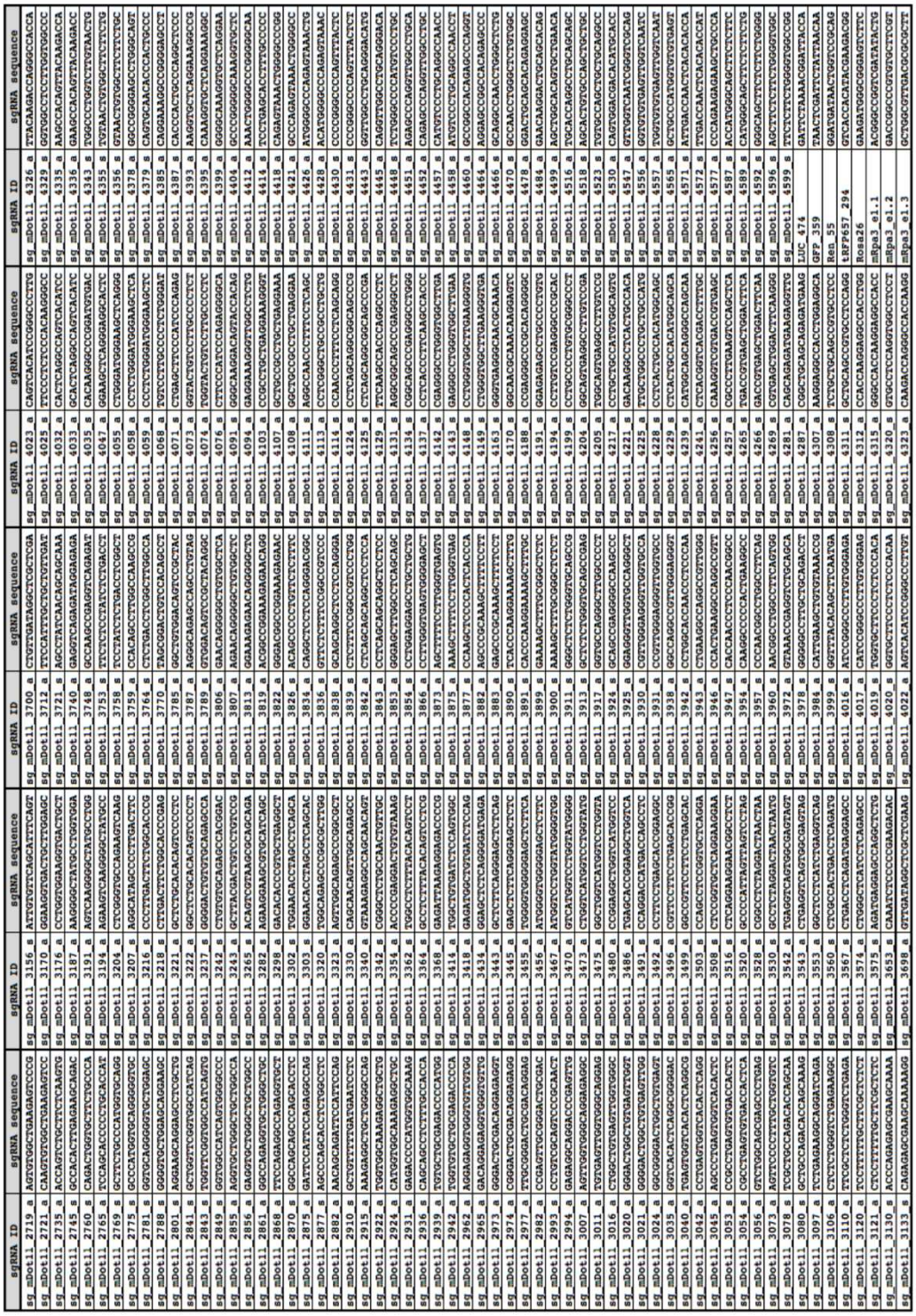
Figure S2. CRISPR gene tiling screen in MLL-AF9 leukemia cells. (a) Specifications of the DOT1L-tiling CRISPR library targeting the coding exons of mouse *Dot1l*. (b) Distribution of individual sgRNA frequencies (reads per million reads; RPMR) in the DOT1L-tiling CRISPR library. 98.6 % of sgRNA passed the QC by exhibiting greater than 100 RPMR. (c) Establishment of monoclonal Cas9-expressing MLL-AF9 leukemia cells with high CRISPR editing efficiency. The sgRNA targeting *Rosa26* (green; negative control) and *Rpa3* (red; positive control) were reported by Shi et al.[4] (Fig. S3).

Figure S3



sgRNA ID	sgRNA sequence	sgRNA ID	sgRNA sequence	sgRNA ID	sgRNA sequence	sgRNA ID	sgRNA sequence	sgRNA ID	sgRNA sequence
sg_mDot11_35_s	GCTGAGGCTCAAGTCGCCTG	sg_mDot11_682_s	ACTTCCTCTCAGAGGAATGG	sg_mDot11_1374_s	TTGGAGGTAGCTGGTAGAAT	sg_mDot11_1874_s	CTCTGAGATCTGGCTCCTCA	sg_mDot11_2305_s	GGTGACAACCGGGTGTAATC
sg_mDot11_36_s	CTGAGCTCAAGTCGCCTGT	sg_mDot11_683_s	CTTCCTCTCAGAGGAATGGA	sg_mDot11_1375_s	CTTGGAGGTAGCTGGTAGAA	sg_mDot11_1875_s	TCTCTGAGATCTGGCTCCTC	sg_mDot11_2308_s	TGGCTTTGCCTGATACACC
sg_mDot11_40_s	ACGGCGGCCTCGGCCCCAC	sg_mDot11_688_s	TCTCAGAGGAATGGAGGAAA	sg_mDot11_1383_s	GCTGCTGCTCTCTGAGATC	sg_mDot11_1884_s	GCCGCTGCTCTCTGAGATC	sg_mDot11_2315_s	AATCTTGGCAGGTGACAACC
sg_mDot11_50_s	CCACGGGTAGACGGCGGGCT	sg_mDot11_732_s	TTTGTGAATAACTTGCCTT	sg_mDot11_1390_s	GAGTGCAGCTGCGTGCTTGG	sg_mDot11_1894_s	GCCAGATCTCAGAGAAGCAG	sg_mDot11_2316_s	CAATCTTGGCAGGTGACAAC
sg_mDot11_55_s	AGCGGCCACGGGTAGACGGC	sg_mDot11_737_s	GTGATCCACCTCAGACCAA	sg_mDot11_1393_s	GGTGAGTGCAGCTGCGTGCT	sg_mDot11_1904_s	AGAGAAGCAGCGGCCACTGC	sg_mDot11_2326_s	CGCCTCAACAACAATCTTGC
sg_mDot11_56_s	CAGCGGCCACGGGTAGACGG	sg_mDot11_740_s	TAACTTTGCCTTTGGTCCTG	sg_mDot11_1414_s	GGTGCCACTAGCAGTGGGTT	sg_mDot11_1949_s	CAGGAGCTCCTGCTGGCGCT	sg_mDot11_2330_s	GTGCCGCCTCAACACAATCT
sg_mDot11_59_s	CCGAGCCGGCCACGGGTAGA	sg_mDot11_743_s	CTTTGCCTTTGGTCCTGAGG	sg_mDot11_1419_s	GGGTGGGTGCCACTAGCAGT	sg_mDot11_1950_s	GCAGGAGCTCCTGCTGGCGC	sg_mDot11_2338_s	CTGCCAAGATGTGTGTGAGG
sg_mDot11_61_s	CGTCTACCCGTGGCCGCTGC	sg_mDot11_758_s	TGAGGTGGATCACCAGCTGA	sg_mDot11_1421_s	CTCACCCAACCCACTGCTAG	sg_mDot11_1956_s	TTAACTGCAGGAGCTCCTGC	sg_mDot11_2351_s	GTTGAGGCGGCACCTGAGCC
sg_mDot11_96_s	TAATCTATGAGCAGCATCA	sg_mDot11_759_s	TTGCGAATCCTCCTCAGC	sg_mDot11_1435_s	TTCTGCAGCGCAGGTGGGGT	sg_mDot11_1968_s	GTACACAGGACTTTAACTGC	sg_mDot11_2352_s	CCCCAGTGTGGTCCTGCTC
sg_mDot11_150_s	CCATGGCCAGCTTGAGATCT	sg_mDot11_779_s	GGAGCGATTCTGCAAACATGA	sg_mDot11_1436_s	CTTCTGCAGCGCAGGTGGGG	sg_mDot11_1982_s	AGCATCATCTGGTGGTACAC	sg_mDot11_2358_s	TGCTAGCCCCCAGTGTGTCC
sg_mDot11_151_s	TCCATGGCCAGCTTGAGATC	sg_mDot11_809_s	CAGAGGGGCAAAGGGCTTTG	sg_mDot11_1439_s	AAGCTCTGCAGCGCAGGTG	sg_mDot11_1990_s	AAAGACAAAGCATCATCTGG	sg_mDot11_2361_s	CACCTGAGCCCAGGACCACAC
sg_mDot11_155_s	GGAAATCCCAGATCTCAAGC	sg_mDot11_817_s	CTGAAGTTCAGAGGGGCAAA	sg_mDot11_1469_s	CTGCAGGTACTGGATTCTGA	sg_mDot11_1993_s	TGCAAAGACAAAGCATCATC	sg_mDot11_2364_s	CTGCTTTGCTAGCCCCAGTG
sg_mDot11_161_s	CCCAGATTCTCAAGCTGGCCA	sg_mDot11_818_s	CCTGAAGTTCAGAGGGGCAA	sg_mDot11_1479_s	AAGCCAGGAACTGCAGGTAC	sg_mDot11_2013_s	GCTTTGTCTTGCATCTGCG	sg_mDot11_2418_s	AGGGAAGGCTGCTCTCTTG
sg_mDot11_167_s	GATCAAGCATAGTTCTCCA	sg_mDot11_824_s	GTAGAATCCTGAAGTTCAGA	sg_mDot11_1485_s	TTGTATAAGCCAGGAACTGC	sg_mDot11_2018_s	GTCTTTGCATCTGCGTGGCA	sg_mDot11_2420_s	GTAGGGAAGGCTGCTCTCTT
sg_mDot11_235_s	GGCTGTGTGACAAATACAAC	sg_mDot11_826_s	CTGTTGATCCTGAAGTTCAG	sg_mDot11_1487_s	AATCCAGTACCTGCAGTCCC	sg_mDot11_2019_s	TCTTTGCATCTGCGTGGCAA	sg_mDot11_2433_s	AGCCAGGGCCTCTGGTAGGGA
sg_mDot11_236_s	GCTGTGTGACAAATACAACC	sg_mDot11_829_s	CCTTTGCCCCTCTGAATTC	sg_mDot11_1494_s	GCGGGGTCTTTGTAATAAGC	sg_mDot11_2027_s	TCTGCGTGGCAAGGGTGCCC	sg_mDot11_2437_s	GACAAGCCAGGGCTCTGGTA
sg_mDot11_243_s	GGTGGATGCTGTCGATGGCC	sg_mDot11_841_s	TGAACTTCAGATCACAGC	sg_mDot11_1511_s	TAGGTTAGCCTTGTACTGCG	sg_mDot11_2028_s	CTGCGTGGCAAGGGTGCCCT	sg_mDot11_2438_s	GGACAAGCCAGGGCTCTGGT
sg_mDot11_289_s	GGCCGAGTATTCAGCTTCAT	sg_mDot11_872_s	CAGCTCCACCACACGCATGA	sg_mDot11_1514_s	TACAAAGACCCCGCAGTACA	sg_mDot11_2033_s	AGCCTCCAGCTCCTGGCCCA	sg_mDot11_2442_s	AGCCTCCCCTACCAGAGCCC
sg_mDot11_298_s	AGCCCATGAAGCTGATACT	sg_mDot11_896_s	GGTGGAGCTGTCGCCCCTGA	sg_mDot11_1535_s	GGCTAACCTACAGCAGCTTC	sg_mDot11_2039_s	GGGTGCCCTGGGCCGAGAGC	sg_mDot11_2449_s	TTCATGCTGTTGGACAAGCC
sg_mDot11_309_s	CTGAATACTGCGCGTCCAA	sg_mDot11_900_s	TCCACGACAGAGCCCTTC	sg_mDot11_1541_s	CCTACAGCAGCTTCTGACC	sg_mDot11_2040_s	CAGCATCAGCCTCCTGACCT	sg_mDot11_2459_s	GGGGCTGAGCTCATGTGT
sg_mDot11_310_s	TGCCGCAGAAGTCCATTGGA	sg_mDot11_910_s	CCCTGAAGGGCTCTGTGTCG	sg_mDot11_1575_s	AGAGCTGCTGGGCCTGTGCC	sg_mDot11_2042_s	TGCCCTGGGCCGAGAGCTGG	sg_mDot11_2474_s	CAGCATGAAGCTCAGCCCCC
sg_mDot11_314_s	GATGTGCCGCAGAAGTCCAT	sg_mDot11_915_s	AAGGGCTCTGTGTCGTGGAC	sg_mDot11_1586_s	GCAATGGCCGAAGAGCTGCT	sg_mDot11_2052_s	CGAGAGCTGGAGGTGTGATC	sg_mDot11_2478_s	AGGCCAGTGGTGGATCCTGG
sg_mDot11_319_s	GGCCGTCCCAATGGACTTCG	sg_mDot11_931_s	GTGTGCAGATAGTAGGAGAC	sg_mDot11_1587_s	GGCAATGGCCGAAGAGCTGC	sg_mDot11_2053_s	GAGAGCTGGAGGCTGATGCT	sg_mDot11_2481_s	GGGAGGCCAGTGGTGGATCC
sg_mDot11_335_s	TCTGCGGCACATCCTGCAGC	sg_mDot11_938_s	GTCAATGGTGTGCAGATAGT	sg_mDot11_1590_s	GGCACAGCCCAGCAGCTCTT	sg_mDot11_2072_s	TGGGCGCTTACGCCTTGAGC	sg_mDot11_2486_s	CAGCCCCCCAGGATCCACCAC
sg_mDot11_336_s	AGTGGTTATATACCTGCTGC	sg_mDot11_1067_s	CTCAGGGACTTTGTGGGCG	sg_mDot11_1601_s	GCAGCTCTTCGGCCATTGCC	sg_mDot11_2073_s	TCTTGGCACATCCAGCTCCA	sg_mDot11_2488_s	GTTGCCGGGGAGGCCAGGTG
sg_mDot11_354_s	TCTCCGGGTCAGTGACAGAG	sg_mDot11_1072_s	TTGCTCTCAGGGACTTTGGT	sg_mDot11_1602_s	CTTCCTTTTGGCCCTGGCAA	sg_mDot11_2090_s	GAGGTGTGCCGGGGAGGCCA	sg_mDot11_2491_s	GAGGTGCCGGGGAGGCCAG
sg_mDot11_362_s	TAACCACTCTGTCACTGACC	sg_mDot11_1073_s	TTTGCTCTCAGGGACTTTGG	sg_mDot11_1608_s	GGATTCTCTTCTTTTGGGCC	sg_mDot11_2099_s	CATGCTGCTGAGGTGTGGCA	sg_mDot11_2495_s	GGATCCACCACTGCCTCCC
sg_mDot11_369_s	CGTAGTTGTTAAGCTCTCC	sg_mDot11_1076_s	TGCTTTGCTCTCAGGGACTT	sg_mDot11_1610_s	CGGCCATTGCCAGGCCCCAA	sg_mDot11_2100_s	TCATGCTGCTGAGGTGTGGC	sg_mDot11_2498_s	GAGGGGCGAGGTTGCCGGGG
sg_mDot11_370_s	TCGTAGTTGTTAAGCTCTCC	sg_mDot11_1083_s	TGGCAGCTGCTTTGCTCTCA	sg_mDot11_1613_s	CCTCGCGATTTCTTCTCTTT	sg_mDot11_2104_s	GGGCTCATGCTGCTGAGGTG	sg_mDot11_2501_s	GGTGAGGGGCGAGGTTGCCG
sg_mDot11_394_s	CCATACACCTCAGGGAGAA	sg_mDot11_1084_s	GTGGCAGCTGCTTTGCTCTC	sg_mDot11_1614_s	GCTCGCGATTTCTTCTCTTT	sg_mDot11_2109_s	GCTCAGGGCTCATGCTGCTG	sg_mDot11_2503_s	GAGGTGAGGGGCGAGGTTGC
sg_mDot11_395_s	CCCATACACCTCAGGGAGAA	sg_mDot11_1100_s	GAGCAAAGCAGCTGCCACTG	sg_mDot11_1624_s	CCCAAAAGGAAGAAATCCGC	sg_mDot11_2124_s	GGCCGTTCATGGACAGCTCA	sg_mDot11_2510_s	TTTCTCGGAGGTGAGGGGCG
sg_mDot11_398_s	CTACGAGCCCTTCCCCCTG	sg_mDot11_1137_s	GGTGCTGAGGAAGAAAAATC	sg_mDot11_1629_s	GCTTCTGCTGGAACAGCCTG	sg_mDot11_2125_s	TGGCCGTTCATGGACAGCTC	sg_mDot11_2515_s	CTACCTTTCCTGCGAGGTGA
sg_mDot11_401_s	GGTCTCCCCATACACCTCAG	sg_mDot11_1142_s	TGAGGAAAGAAAAATCAGGTG	sg_mDot11_1643_s	CAGGCTGTTCCAGCAGAAGC	sg_mDot11_2133_s	AGCCCTGAGCTGTCCATGAA	sg_mDot11_2517_s	CACTACCTTTCCTGCGAGGT
sg_mDot11_403_s	GAGGTCTCCCCATACACCTC	sg_mDot11_1168_s	TTCCGGGCTTTGGAGGGAGA	sg_mDot11_1679_s	GGCCTGAATCCAGGCCTTCA	sg_mDot11_2135_s	GCTGGCAGCATGGCCCGTTC	sg_mDot11_2523_s	TCGCCCCCTCACCTCGGAGA
sg_mDot11_405_s	CCCTTCTCCCCTGAGGTGTA	sg_mDot11_1174_s	TTCTTCTTCCCGGGCTTTGG	sg_mDot11_1688_s	GACCTACAATGACCTGATTC	sg_mDot11_2145_s	AGAGCTCATAGCTGGCAGCA	sg_mDot11_2561_s	GGTCTCCCCGTGGCTGCTGT
sg_mDot11_407_s	CTTCTCCCCTGAGGTGTATG	sg_mDot11_1175_s	TTTCTTCTTCCCGGGCTTTG	sg_mDot11_1689_s	TTTCCTTCTGGCCCTGAATC	sg_mDot11_2153_s	TGCATTGCAGAGCTCATAGC	sg_mDot11_2565_s	GAGCGTGCCTACAGCAGCCA
sg_mDot11_422_s	CTGGGCCACATGGACTTCTG	sg_mDot11_1177_s	AGCCACCCCTCCCAAAGCC	sg_mDot11_1697_s	TGACCTGATTCAGGCCCAGA	sg_mDot11_2173_s	AGCTCTGCAATGCAGCTAGT	sg_mDot11_2567_s	GCGTGCCTACAGCAGCCACG
sg_mDot11_425_s	TGGGGAGACCTCCTTTGACC	sg_mDot11_1178_s	CAGTTTCTCTTCCGGGCTTT	sg_mDot11_1700_s	GTGGGCAGATGATTCCTTCT	sg_mDot11_2185_s	GGTGTGTCTCTGCTTGGATGA	sg_mDot11_2571_s	GGCTGGTGATGGTCTCCCCG
sg_mDot11_428_s	GGAGACCTCCTTTGACCTGG	sg_mDot11_1184_s	CTTGTTCAGTTTCTCTTCTC	sg_mDot11_1701_s	TGTGGGCAGATGATTCCTTC	sg_mDot11_2192_s	GTACTGGGGTGTGTCTGCTT	sg_mDot11_2582_s	GCTAACAGGTAGGCTGGTGA
sg_mDot11_432_s	CATCAATCATCTGGGCCACC	sg_mDot11_1185_s	TCTTGTTCAGTTTCTCTTCT	sg_mDot11_1718_s	TTCCCGCAGCTGCTGGTTGT	sg_mDot11_2206_s	AAGGGAGAGGCCAGGTACTG	sg_mDot11_2588_s	TGGAATGCTAACAGGTAGGC
sg_mDot11_440_s	CTTGATCTCATCAATCATCT	sg_mDot11_1200_s	AAGAAGAAACTGAACAAGAA	sg_mDot11_1719_s	GTTCCCGCAGCTGCTGGTTG	sg_mDot11_2208_s	CCAAGGGAGAGGCCAGGTAC	sg_mDot11_2592_s	TGATGTGGAATGCTAACAGG
sg_mDot11_441_s	TCTTGATCTCATCAATCATC	sg_mDot11_1201_s	AGAAGAAACTGAACAAGAAA	sg_mDot11_1725_s	CCGACTGTTCCCGCAGCTGC	sg_mDot11_2214_s	CCTGGTCCAAGGGAGAGGCC	sg_mDot11_2596_s	GTGCTGATGTGGAATGCTAA
sg_mDot11_461_s	TGATGAGATCAAGATGACAG	sg_mDot11_1211_s	GAACAAGAAGGGAGAAAGA	sg_mDot11_1727_s	TGCCCACAACCAGCAGCTGC	sg_mDot11_2219_s	CCAGTACCTGGCCCTCTCCT	sg_mDot11_2608_s	TTGGGCTGCACTGTGCTGAG
sg_mDot11_482_s	GGATGCCTGTTGTCGACC	sg_mDot11_1215_s	AAGAAAGGGAGAAAGATGGC	sg_mDot11_1736_s	CCAGCAGCTGCGGGAACAGT	sg_mDot11_2224_s	GGTACGACCCTCCTCCAGA	sg_mDot11_2626_s	ATGCTGACCGGCAGCTTGTT
sg_mDot11_483_s	GATGACCTGTTTGTCGACCT	sg_mDot11_1219_s	AAGGGAGAAAGATGGCTGGC	sg_mDot11_1745_s	GCGGGAACAGTCGGAACAGC	sg_mDot11_2225_s	AGGTACGACCCTCCTGGCCA	sg_mDot11_2627_s	GATGCTGACCGGCAGCTTGT
sg_mDot11_516_s	TGGTGGCCGCAGCAACCTGA	sg_mDot11_1227_s	AAGATGGCTGGCCGGAAACG	sg_mDot11_1751_s	ACAGTCGGAACAGCAGCTGG	sg_mDot11_2228_s	GGCCTCTCCCCTTGGACCAG	sg_mDot11_2630_s	AGTCGACGCCCAACAAGCTGC
sg_mDot11_521_s	GGTTGTCCTCTCAGGTTCTG	sg_mDot11_1228_s	AGATGGCTGGCCGGAAACGT	sg_mDot11_1760_s	AGAAGGACCACAGCGACGCT	sg_mDot11_2232_s	GGGTACAAGTGGTCATCGCT	sg_mDot11_2638_s	CTGGCCAGGTCAGCATCCCAC
sg_mDot11_533_s	GTAGTGATGTTTGCAGTTGG	sg_mDot11_1231_s	TGGCTGGCCGGAAACGTGGG	sg_mDot11_1783_s	AGCTTCGGAGCCAGAGCCTG	sg_mDot11_2245_s	CCACTGTGGCTAGGGGTACA	sg_mDot11_2645_s	GCTGCCGGTCAGCATCCCAC
sg_mDot11_536_s	TCCGTAGTGATGTTTGCAGT	sg_mDot11_1243_s	GCTGGCAGCACTCATTTCTT	sg_mDot11_1805_s	AAGCCTCAGCTCCTCACACC	sg_mDot11_2252_s	AGGCCGGCCACTGTGGCTAG	sg_mDot11_2649_s	GCAGCACCAGCTGCTGGCCAG
sg_mDot11_546_s	ACCAACTGCAAACATCACTA	sg_mDot11_1244_s	ACTGGCAGCACTCATTTTCT	sg_mDot11_1806_s	CAAGCCTCAGCTCCTCACAC	sg_mDot11_2254_s	CGAGGCCGGCCACTGTGGCT	sg_mDot11_2650_s	GGCAGCACCAGCTGCTGGCCA
sg_mDot11_551_s	CTGCAAACATCACTACGGAG	sg_mDot11_1262_s	CTTAGACTTGCGTCCGGCAC	sg_mDot11_1813_s	GGGCCCGGTGTGGAGGACTG	sg_mDot11_2256_s	CCTTGTACACCCTACTGGCCAGG	sg_mDot11_2654_s	CAGCATCCCACTGGCCAGCG
sg_mDot11_560_s	TCACTACGGAGTGGAGAAAG	sg_mDot11_1268_s	GCTCTTCTTAGACTTGCGTT	sg_mDot11_1822_s	GTGAGGAGCTGAGCCTGTAC	sg_mDot11_2259_s	CTGACCGAGGCCGGCCACTG	sg_mDot11_2657_s	ACGGCTGGGCAGCATCACGC
sg_mDot11_603_s	TCATCCACTCCCTGAACTCT	sg_mDot11_1290_s	GCAGGAGATCCAGTGTGCTT	sg_mDot11_1835_s	CAGGTTCTCCAAGGACAGTG	sg_mDot11_2260_s	GTACCCCTAGCCACAGTGGC	sg_mDot11_2677_s	TGCTGCCCAGCCGTGCTGAG
sg_mDot11_610_s	TGGACCGAGAGTTCAAGAAG	sg_mDot11_1292_s	TAAAAGAAGCCAAAGACAC	sg_mDot11_1838_s	TGACTGCCCTCCACACTGCC	sg_mDot11_2265_s	CTAGCCACAGTGGCCGGCCT	sg_mDot11_2701_s	CGGGGCTGTGGCACAGGGCT
sg_mDot11_619_s	AGTTCAGGAAGTGGATGAAA	sg_mDot11_1308_s	GGGCTGGGGGCGGAGATGC	sg_mDot11_1844_s	CTCTTTCCGCAGGTTCTCCA	sg_mDot11_2268_s	ACAGCTTTTCTAGCCGAGGC	sg_mDot11_2702_s	CCGGGGCTGTGGCACAGGGC
sg_mDot11_624_s	AGGAAGTGGATGAAATGGTA	sg_mDot11_1310_s	GCTGAGGGTGAAAGTGTG	sg_mDot11_1849_s	CACTGTCCCTTGGAGAACCTG	sg_mDot11_2272_s	CCAGACAGCTTTTCTAGCCG	sg_mDot11_2706_s	GGGGGCTGTGGCACAGGGCT
sg_mDot11_674_s	ACGAGGTGACTTCCTCTCAG	sg_mDot11_1321_s	GAGGCTGAGGGTGGGGCTGG	sg_mDot11_1854_s	GGGCCTGCTCCTCTTCCCGC	sg_mDot11_2283_s	CCTCGGCTAGAAAAGCTGTC	sg_mDot11_2707_s	GAGTCCCCGGGGCTGTGGCAC
sg_mDot11_675_s	GTTCCCTCCATTCCTCTGAG	sg_mDot11_1324_s	GGTCAGGCTGAGGGTGGGGC	sg_mDot11_1862_s	GAACCTGCGGAAAGGAAGC	sg_mDot11_2288_s	GCTAGAAAAGCTGTCCTGGC	sg_mDot11_2713_s	GCTGAAGAGTCCCGGGGCTG
sg_mDot11_679_s	GTGACTTCCTCTCAGAGGAA	sg_mDot11_1366_s	AGCTGGTAGAATGGGCTGTG	sg_mDot11_1870_s	GGAAAGGAAGCAGGCCCTG	sg_mDot11_2295_s	GGGTGTAATCAGGCAAAGCC	sg_mDot11_2714_s	CAGCCCTGTGCCACAGCCCC

Figure S3 (continued)



sgRNA ID	sgRNA sequence	sgRNA ID	sgRNA sequence	sgRNA ID	sgRNA sequence	sgRNA ID	sgRNA sequence	sgRNA ID	sgRNA sequence	sgRNA ID	sgRNA sequence
sg_mDot1l_2719_a	AGTGTGGTCGAAGAGTCCCG	sg_mDot1l_3156_a	ATTGTGTTCAGCATTTCAGT	sg_mDot1l_3700_a	CTGTTGATAGCCTCGCTCGA	sg_mDot1l_4023_a	CAGTCACATCCGGGCCCTTG	sg_mDot1l_4326_a	TTACAAGACCAGGGCCACCA
sg_mDot1l_2721_a	CAAGTGTGGCTGAAGAGTCC	sg_mDot1l_3170_a	GGAAGGTGACTGCTTGGAGC	sg_mDot1l_3712_a	TTCCATTTGCTGCTGTTGAT	sg_mDot1l_4025_a	TTCCCTCTCCCACAAGGGCC	sg_mDot1l_4329_a	GGTGGCCTCCTTGGTGGCCC
sg_mDot1l_2735_a	ACCAGTCTGCTTCTCAAGTG	sg_mDot1l_3176_a	CCTGGTGGAAGGTGACTGCT	sg_mDot1l_3721_a	AGCCTATCAACAGCAGCAAA	sg_mDot1l_4032_a	CACTCAGGCCAGTCACATCC	sg_mDot1l_4335_a	AAGCCACAGTTACAAGACCA
sg_mDot1l_2745_a	GCCACACTTGAGAAGCAGAC	sg_mDot1l_3187_a	AAGGGGCTATGCCTGGTGGA	sg_mDot1l_3740_a	GAGGTCAGAGATAGGAGAGA	sg_mDot1l_4033_a	GCACTCAGGCCAGTCACATC	sg_mDot1l_4336_a	GAAGCCACAGTTACAAGACC
sg_mDot1l_2760_a	CAGATGGTGCTCTCGCCCA	sg_mDot1l_3191_a	AGTCAAGGGCCTATGCCTGG	sg_mDot1l_3748_a	GCCAAGCCGAGGTCAGAGAT	sg_mDot1l_4035_a	CACAAGGCCCGGATGTGAC	sg_mDot1l_4343_a	TGGCCCTGGTCTGTGTAACTG
sg_mDot1l_2765_a	TCCAGCACCCCCCTGCACCAT	sg_mDot1l_3194_a	AGAAGTCAAGGGGCTATGCC	sg_mDot1l_3753_a	TTCTCTCCTATCTCCTGACCT	sg_mDot1l_4047_a	GGAAGCTCAGGGAGGCACTC	sg_mDot1l_4355_a	TGTAACTGTGGCTCTCTCTG
sg_mDot1l_2775_a	GCTTCTGCCCAATGGTCCAG	sg_mDot1l_3204_a	AGGCATGCCCCTTGACTTC	sg_mDot1l_3758_a	TCCTATCTCTGCACCTCGCT	sg_mDot1l_4055_a	CTGGGATGGGAAGCTCAGA	sg_mDot1l_4356_a	GTAACTGTGGCTCTCTCTGC
sg_mDot1l_2781_a	GGTGCAGGGGGTGCTGGAGC	sg_mDot1l_3216_a	CCCTTGACTTCTGGCACCCG	sg_mDot1l_3764_a	CCACAGCCTTGGCCAAGCCG	sg_mDot1l_4058_a	CCTCTGGGATGGGAAGCTCA	sg_mDot1l_4378_a	GGCCGGGGAGCCCTGGGCAGT
sg_mDot1l_2788_a	GGGTGCTGCAGGAGCAGAAGC	sg_mDot1l_3218_a	CTGACTTCTGGCACCGGAG	sg_mDot1l_3770_a	TAGCGGACTGTCCACAGCCT	sg_mDot1l_4059_a	CCCTCTGGGATGGGAAGCTC	sg_mDot1l_4379_a	CAGTGCAACACCCACTGCCC
sg_mDot1l_2801_a	AGGAAGCAGGAGCCTCGCTG	sg_mDot1l_3221_a	GCTCTGCACACAGTCCCCTC	sg_mDot1l_3785_a	GGCTGTGGACAGTCCGCTAC	sg_mDot1l_4068_a	TGTCCTGTCCCCCTCTGGGAT	sg_mDot1l_4385_a	CAGGAAAGGCCGGGGAGCCT
sg_mDot1l_2841_a	GCTGGTTCGGTGGCCATCAG	sg_mDot1l_3222_a	GGCTCTGCACACAGTCCCCT	sg_mDot1l_3787_a	AGGGCAGAGCCAGCCTGTAG	sg_mDot1l_4071_a	GGTACTGTCCTGCCCCTCT	sg_mDot1l_4387_a	CACCCACTGCCCAGGGCTCC
sg_mDot1l_2843_a	TGGTTCGGTGGCCATCAGTG	sg_mDot1l_3237_a	GGGGACTGTGTGCAGAGCCA	sg_mDot1l_3789_a	GTGGACAGTCCGCTACAGGC	sg_mDot1l_4073_a	GGTACTGTCCTGCCCCTCTG	sg_mDot1l_4393_a	AAGGTGCTCAGAAAAGGCCG
sg_mDot1l_2849_a	GGTGGCCATCAGTGGGGCCC	sg_mDot1l_3242_a	CTGTGTGCAGAGCCACGGAC	sg_mDot1l_3806_a	GAACAGGGGGCTGTGGCTCA	sg_mDot1l_4076_a	CTTCCCATCCCAGAGGGGCA	sg_mDot1l_4399_a	GGGGCAAAGGTGCTCAGAAA
sg_mDot1l_2855_a	GCTTCTGCCCAATGGTCCAG	sg_mDot1l_3243_a	GCTTACGACTGCTCGTCCCG	sg_mDot1l_3807_a	AGAACAGGGGGCTGTGGCTC	sg_mDot1l_4091_a	GGGCAAGGACAGTACCACAG	sg_mDot1l_4404_a	GCCCGGGGGCAAAGGTGCTC
sg_mDot1l_2856_a	GAGGTGCTGGGCTGCTGCCC	sg_mDot1l_3265_a	ACAGTCGTAAGCGCAGCAGA	sg_mDot1l_3813_a	GGAAAGGAACAGGGGCTGG	sg_mDot1l_4094_a	GAGGAAAGGGTTGGCCTCTG	sg_mDot1l_4412_a	AAACTGGGGCCCGGGGGCAA
sg_mDot1l_2861_a	GGCCAGAGGTGCTGGGCTGC	sg_mDot1l_3282_a	AGAAGGAGCGTCGCATCAGC	sg_mDot1l_3819_a	ACGGCCGGAAAGAGGAACAG	sg_mDot1l_4103_a	CCGCCTGTGAGGAAAGGGGT	sg_mDot1l_4414_a	TCCTGAGCACCTTTGCCCCC
sg_mDot1l_2868_a	TTCCAGAGGCCAGAGTGCT	sg_mDot1l_3298_a	GACACACCCGTGCTGAGGCT	sg_mDot1l_3822_a	GGGACGGCCGGAAAGAGAAC	sg_mDot1l_4107_a	GCTGCCGCCTGCTGAGGAAA	sg_mDot1l_4418_a	CAGAGTAAACTGGGGGCCCG
sg_mDot1l_2870_a	GGCCAGCAGCCCAGCACCTC	sg_mDot1l_3302_a	TGGAACACCTAGCCTCAGCA	sg_mDot1l_3826_a	ACAGCCCCCTGTCTCTTTC	sg_mDot1l_4108_a	GGCTGCCGCCTGCTGAGGAA	sg_mDot1l_4421_a	GCCCAGAGTAAACTGGGGCC
sg_mDot1l_2875_a	GATCCCATTCCAGAGCCCAG	sg_mDot1l_3314_a	GGAACACCTAGCCTCAGCAC	sg_mDot1l_3829_a	CAGCCTCCTCCCAGGGACGC	sg_mDot1l_4111_a	AGGCCAACCCTTTCCTCAGC	sg_mDot1l_4426_a	ATGGGGCCCCAGAGTAAACT
sg_mDot1l_2877_a	AGCCCAGCACCTCTGGCCTC	sg_mDot1l_3320_a	TGGCAGAGCCCGGCGCTTGG	sg_mDot1l_3836_a	GTTCTCTTTCCGGCCGTCCC	sg_mDot1l_4113_a	CCTCGGGCTGCCGCCTGCTG	sg_mDot1l_4428_a	ACATGGGGCCCCAGAGTAAAC
sg_mDot1l_2882_a	AACAGCAGATTCCATTCCAG	sg_mDot1l_3323_a	AGTTGGCAGAGCCCGGCGCT	sg_mDot1l_3838_a	GCAGCAGCTCCTCCCAGGGA	sg_mDot1l_4114_a	CCAACCCTTTCCTCAGCAGG	sg_mDot1l_4430_a	CCCGGGCCCCCAGTTTACTCT
sg_mDot1l_2910_a	GCTGTTTTTGATGAATCCTC	sg_mDot1l_3330_a	CAGCAACAGTTGGCAGAGCC	sg_mDot1l_3839_a	CTCTTTCCGGCCGTCCCTGG	sg_mDot1l_4124_a	CCTCAGCAGGCGGCAGCCCG	sg_mDot1l_4431_a	CCCGGGCCCCCAGTTTACTCT
sg_mDot1l_2915_a	AAAGAGCCTGCTGGGGCCAG	sg_mDot1l_3340_a	GTAAAGAGCCAGCACACAGT	sg_mDot1l_3842_a	CTCAGCAGCAGGCTCCTCCA	sg_mDot1l_4125_a	CTCAGCAGCGGCAGCCCGA	sg_mDot1l_4443_a	GGTGGCCTGCAGGGACATG
sg_mDot1l_2922_a	TGGTGGCAAAGAGCTCGCTG	sg_mDot1l_3342_a	CGGGCTCGCCAACTGTTGC	sg_mDot1l_3843_a	CCTCAGCAGCAGGCTCCTCC	sg_mDot1l_4129_a	TTCAAGCCACCCAGGCCCTC	sg_mDot1l_4445_a	CAGGTTGGCCTGCAGGGACA
sg_mDot1l_2924_a	CATGGTGGCAAAGAGCTGC	sg_mDot1l_3354_a	ACCCCGAGGACTGTGTAAAG	sg_mDot1l_3853_a	GGGAGCTTGGCCTCAGCAGC	sg_mDot1l_4131_a	AGGCGGCAGCCCGAGGGCCT	sg_mDot1l_4448_a	TCTGGGCCCCATGTCCCTGC
sg_mDot1l_2931_a	GAGACCCCATGGTGGCAAGG	sg_mDot1l_3362_a	TGGCCTCTTTACACAGTCCT	sg_mDot1l_3854_a	CCTGGAGGAGCCTGCTGCTG	sg_mDot1l_4134_a	CGGCAGCCGAGGGCCTGGG	sg_mDot1l_4451_a	AGAGCCCAGGTTGGCCTGCA
sg_mDot1l_2936_a	CAGCAGCCTCTTTGCCACCA	sg_mDot1l_3364_a	GCCTCTTTACACAGTCCTCG	sg_mDot1l_3866_a	CCTGGGTGAGTGGGGAGCCT	sg_mDot1l_4137_a	CCTCACCCTTCAAGCCACCC	sg_mDot1l_4452_a	CAGGAGCCCCAGGTGTGGCCTGC
sg_mDot1l_2939_a	TGTGCTGCGAGACCCCATGG	sg_mDot1l_3368_a	GAGATTAAGAGGAGACCCCG	sg_mDot1l_3873_a	AGCTTTTCCTTGGGTGAGTG	sg_mDot1l_4142_a	CGAGGGCCTGGGTGCCTTGA	sg_mDot1l_4457_a	CATGTCCCTGCAGGCCAACC
sg_mDot1l_2942_a	TGGTGTGCTGCGAGACCCCA	sg_mDot1l_3414_a	TGGCTGTGATACTCCCAGTGGC	sg_mDot1l_3875_a	AAAGCTTTTCCTTGGGTGAG	sg_mDot1l_4143_a	GAGGGCCTGGGTGGCTTGAA	sg_mDot1l_4458_a	ATGTCCCTGCAGGCCAACCT
sg_mDot1l_2962_a	AGGAAGGTGGGTGTTGTGG	sg_mDot1l_3418_a	GAGATGGCTGTGATACTCCAG	sg_mDot1l_3877_a	CCAAGCTCCCCACTCACCCA	sg_mDot1l_4148_a	CCTGGGTGGCTTGAAGGGTG	sg_mDot1l_4460_a	GCCGGCCACAGAGCCCAGGT
sg_mDot1l_2965_a	GACAGGAGAGTGGGGTGTTG	sg_mDot1l_3424_a	GGAGCTCTCAGGGAGTGAGA	sg_mDot1l_3882_a	AGCCCGCAAAGCTTTTCCTT	sg_mDot1l_4149_a	CTGGGTGGCTTGAAGGGTGA	sg_mDot1l_4464_a	AGGAGCCGGCCACAGAGCCC
sg_mDot1l_2973_a	GGGACTGCGACAGGAGAGGT	sg_mDot1l_3443_a	GCTCTTCAGGGAGCTCTCAG	sg_mDot1l_3883_a	GAGCCCGCAAAGCTTTTCCT	sg_mDot1l_4163_a	GGGTGAGGGCAACGCAAACA	sg_mDot1l_4466_a	GCAGGCCAACCTGGGCTCTG
sg_mDot1l_2974_a	GGGACTGCGACAGGAGAGGT	sg_mDot1l_3445_a	GAGCTCTCAGGGAGCTCTCA	sg_mDot1l_3890_a	TCACCCAAGGAAAAAGCTTTG	sg_mDot1l_4170_a	GGCAACGCAAACAAGGATCC	sg_mDot1l_4470_a	GCCAACCTGGGCTCTGTGGC
sg_mDot1l_2977_a	TTGCGGGACTGCGACAGGAG	sg_mDot1l_3455_a	TGGGGTGGGGGAGCTCTTCA	sg_mDot1l_3891_a	CACCCAAGGAAAAAGCTTTGC	sg_mDot1l_4188_a	CCGAGGGCCCGCACAGGGGC	sg_mDot1l_4478_a	GGACTGCAGCACAGAGGAGC
sg_mDot1l_2982_a	CCGAGTTGCGGGACTCGAC	sg_mDot1l_3467_a	ATGGGGTGGGGGAGCTCTTC	sg_mDot1l_3899_a	AAAAGCTTTGCGGCCTCTCC	sg_mDot1l_4191_a	CCTTGTCCGAGGGCCCGCAC	sg_mDot1l_4484_a	GAACAAGGACTGCAGCACAG
sg_mDot1l_2993_a	CCTGTCCGCAGTCCCGCACT	sg_mDot1l_3470_a	GTCATGGTCCCTGGTATGGGG	sg_mDot1l_3900_a	AAAAGCTTTGCGGCCTCTCT	sg_mDot1l_4194_a	CCTTGTCCGAGGGCCCGCAC	sg_mDot1l_4499_a	AGCTGGCACAGTGCTGAACA
sg_mDot1l_2994_a	GAGAGGCAGGACCCGAGTTG	sg_mDot1l_3473_a	CTGGTCATGGTCCTGGTATG	sg_mDot1l_3911_a	GGGCTCTCTGGGTGCAGCCG	sg_mDot1l_4199_a	CCTGCCCCTGTGCGGGCCCT	sg_mDot1l_4516_a	TGCACCAGGCCTGCAGCAGC
sg_mDot1l_3007_a	AGTTGGTGGGCAGAGAAGGC	sg_mDot1l_3475_a	GGCTGGTCATGGTCCTGGTA	sg_mDot1l_3913_a	GCTCTCTGGGTGCAGCCGAG	sg_mDot1l_4204_a	GGCAGTGAGGCCTGTGCCGA	sg_mDot1l_4518_a	AGCACTGTGCCAGCTGCTGC
sg_mDot1l_3011_a	TGTGAGTTGGTGGGCAGGAG	sg_mDot1l_3480_a	CCGGAGGCTGGTCATGGTCC	sg_mDot1l_3917_a	GGTGCCAGGGCTTGGCCCCT	sg_mDot1l_4205_a	TGGCAGTGAGGCCTGTGCCG	sg_mDot1l_4523_a	TGTGCCAGCTGCTGCAGGCC
sg_mDot1l_3016_a	CTGCCTGGAGTTGGTGGGC	sg_mDot1l_3486_a	TGAGCACCGGAGGCTGGTCA	sg_mDot1l_3924_a	GCAGCCGAGGGGCCCAAGCC	sg_mDot1l_4217_a	CCTGCTGCCATGTGGCAGTG	sg_mDot1l_4530_a	CAGTGGACACACATGCACCC
sg_mDot1l_3020_a	GGGACTGGCTGTGATGTTGT	sg_mDot1l_3491_a	CCAGGACCATGACCAGCCTC	sg_mDot1l_3925_a	GGAGGGTGGTGTCCCAGGGCT	sg_mDot1l_4221_a	GACAAGGCCTCACTGCCACA	sg_mDot1l_4547_a	GTTGGTCAATGGGGTGCGAG
sg_mDot1l_3021_a	GGGGACTGGCTGTGATGTTG	sg_mDot1l_3492_a	CCTTCCTGAGCACCGGAGGC	sg_mDot1l_3930_a	CGTTGGGAGGGTTGTGTGCCA	sg_mDot1l_4225_a	TTGCTGGCCCTGCTGCCATG	sg_mDot1l_4556_a	GGTGTGTGAGTTGTGTCAATC
sg_mDot1l_3024_a	GGCGGGGACTGGCTGTGATG	sg_mDot1l_3496_a	CGTTCCTTCCTGAGCACCGG	sg_mDot1l_3931_a	CCGTTGGGAGGGTTGTGTGC	sg_mDot1l_4228_a	CCTCACTGCCACATGGCAGCA	sg_mDot1l_4557_a	TGGTGTGAGTTGTGTCAATC
sg_mDot1l_3035_a	GGTCACACTCAGGCGGGGAC	sg_mDot1l_3499_a	GGCCGTCTCCTCCTGAGCAC	sg_mDot1l_3938_a	GGCCAGGCCGTTGGGAGGGT	sg_mDot1l_4229_a	CTCACTGCCACATGGCAGCA	sg_mDot1l_4565_a	GCTGCCCATGGTGTGTGAGT
sg_mDot1l_3040_a	TGAGTGGTCACACTCCAGGG	sg_mDot1l_3503_a	CCAGCCTCCGGTGCTCAGGA	sg_mDot1l_3942_a	CCTGGCACCAACCCTCCCAA	sg_mDot1l_4239_a	CATGGCAGCAGGCCAGCAA	sg_mDot1l_4571_a	ATGACCAACTCACACCACAT
sg_mDot1l_3042_a	CCTGAGTGGTCACACTCAGG	sg_mDot1l_3508_a	CTCCGGTGCTCAGGAAGAA	sg_mDot1l_3943_a	CTGAAGGCCAGCCGTTGGG	sg_mDot1l_4241_a	CTCACGGTCACGACCTTTGC	sg_mDot1l_4572_a	TTGACCAACTCACACCACAT
sg_mDot1l_3045_a	AGCCCTGAGTGGTCACACTC	sg_mDot1l_3516_a	CTCCGGTGCTCAGGAAGAAA	sg_mDot1l_3946_a	CCACTGAAGGCCAGCCGTTG	sg_mDot1l_4256_a	CAAAGGTCGTGACCGTGAGC	sg_mDot1l_4577_a	CCCAGAAGAAGCAAGCTGCCCA
sg_mDot1l_3053_a	CGCCTGAGTGTGACCACTCA	sg_mDot1l_3520_a	GCCCCATTAGTTAGTCCTAG	sg_mDot1l_3954_a	CAAGGCCCCCACTGAAGGCC	sg_mDot1l_4265_a	TGACCGTGAGCTGGACTTCA	sg_mDot1l_4589_a	CATGGGCAGCTTCTCTCTTG
sg_mDot1l_3054_a	CGCCTGAGTGTGACCACTCA	sg_mDot1l_3528_a	CGGCCTCTAGGACTAACTAA	sg_mDot1l_3957_a	CCCAACGGCCTGGCCTTCAG	sg_mDot1l_4266_a	GACCGTGAGCTGGACTTCAA	sg_mDot1l_4592_a	GGGCAGCTTCTCTCTTGTGG
sg_mDot1l_3073_a	AGTTCCCCTTTGCTGGTGTC	sg_mDot1l_3530_a	GCCTCTAGGACTAACTAATG	sg_mDot1l_3960_a	AACGGCCTGGCCTTCAGTGG	sg_mDot1l_4269_a	CGTGAGCTGGACTTCAAGGG	sg_mDot1l_4596_a	AGCTTCTCTCTTGTGGTGGC
sg_mDot1l_3078_a	TCGCTGCCAGACACCAGCAA	sg_mDot1l_3542_a	TGAGTCCAGTGGCGAGTAGT	sg_mDot1l_3972_a	GTAAACCGAGGTCTGCAGCA	sg_mDot1l_4281_a	CAGCAGAGATGAAGAGGTTG	sg_mDot1l_4599_a	TCTCTCTTGTGGTGGCCGCG
sg_mDot1l_3080_a	GCTGCCAGACACCAGCAAAG	sg_mDot1l_3543_a	CTGAGTCCAGTGGCGAGTAG	sg_mDot1l_3978_a	GGGGGCCTGCTGCCAGACCT	sg_mDot1l_4287_a	CGGCTGCACAGAGATGAAGG	UJC_474	GATTCTAAAACGGATTACCA
sg_mDot1l_3097_a	TCTGAGAAGGCAGATCAGA	sg_mDot1l_3553_a	GGCTCCTCATCTGAGGTCAG	sg_mDot1l_3986_a	CATTGAAGCTGTGTAAACCG	sg_mDot1l_4307_a	AAGGAGGCCACCTGGAGCA	GFP_359	TAACTCGATTCTATTAACAA
sg_mDot1l_3106_a	GTCTGGCTCTGAGAAGGC	sg_mDot1l_3560_a	CTCGCCACTGACCTCAGAGG	sg_mDot1l_3999_a	TCTGCTTACACAGCTTAATGA	sg_mDot1l_4311_a	TCTGCTGCAGCCGTGCCTCC	Rsa_55	GGATGATAACTGGTCCGCAG
sg_mDot1l_3110_a	TTCGCTCTCTGGGTCTGAGA	sg_mDot1l_3567_a	CTGACCTCAGAGTGAGGAGCC	sg_mDot1l_4016_a	ATCCGGGCCCTTGTGGGAGA	sg_mDot1l_4312_a	GCTGCAGCCGTGCCTCCAGG	tRFP657_294	GTCACCACATACGAAGACGG
sg_mDot1l_3120_a	TCCTTTTTGCTGGTCTGAGA	sg_mDot1l_3574_a	CTCGCCACTGACCTCAGAGG	sg_mDot1l_4017_a	ATCCGGGCCCTTGTGGGAGA	sg_mDot1l_4315_a	CCACCAAGGAGGCCACCTGG	Rosa26	GAAGATGGGCGGGAGTCTTC
sg_mDot1l_3121_a	CTCCTTTTTGCTGGTCTGCT	sg_mDot1l_3575_a	AGATGGAGGAGCCAGGCTCTG	sg_mDot1l_4019_a	TGGTGCTCCCTCTCCCACAG	sg_mDot1l_4315_a	GGGCCACCAAGGAGGCCACC	mRpa3_e1.1	ACGGGCCGGTCGATATACTG
sg_mDot1l_3130_a	ACCCAGAGAGCGAAGCAAAA	sg_mDot1l_3653_a	CAAATCTCCCCCGAAGGACAC	sg_mDot1l_4020_a	GGTGCTCCCTCTCCCACAAG	sg_mDot1l_4320_a	GTGCCTCCAGGTGGCCTCCT	mRpa3_e1.2	GACCGGCCCGTGTGCTTCGT
sg_mDot1l_3133_a	CAGAGACGAAGCAAAAAGG	sg_mDot1l_3698_a	GTTGATAGGCCTCGCTCGAAG	sg_mDot1l_4022_a	AGTCACATCCGGGCCCTTGT	sg_mDot1l_4323_a	CAAGACCAGGGCCACCAAGG	mRpa3_e1.3	GCTGGCGTGTGACGCGCGCTT

Figure S3. Sequence of individual sgRNAs in the DOT1L-tiling CRISPR library. sgRNAs are named as "sg_mDot1l_(cut position in cDNA)_(sense/antisense)".

Figure S4

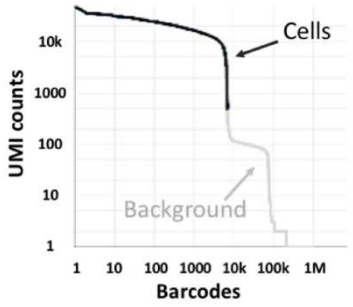
a

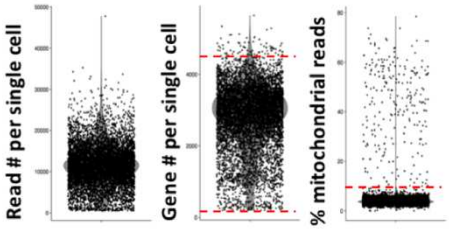
Poly(dT) captured mRNA library

Number of reads	185,270,701
Valid barcodes	96.2%
Q30 – Barcode	95.9%
Q30 – mRNA	92.7%
Q30 – Index	94.2%
Q30 – UMI	93.3%
Mean reads per cell	26,350
Median genes per cell	2,935
Total genes detected	15,921
Median UMI # per cell	11,341

Alignment (mm10)

Genome	92.0%
Exons	76.0%
Introns	12.2%
Intergenic region	3.8%
Antisense	2.0%





b

CS1 captured sgRNA library

Number of reads	8,910,124
Valid barcodes	84.2%
Q30 – Barcode	96.7%
Q30 – sgRNA	94.1%
Q30 – Index	91.7%
Q30 – UMI	96.7%
Mean reads per cell	1,267
Cells with 1 sgRNA	4,362
Cells with 2 sgRNA	527
Cells with 3+ sgRNA	54

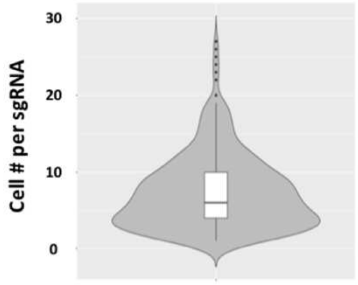


Figure S4. Summary of sequencing specifications of sc-Tiling mouse *Dot1l* gene. Single-cell sequencing QC report of (a) poly(dT)-captured mRNA library and (b) CS1-captured sgRNA library.

Figure S5

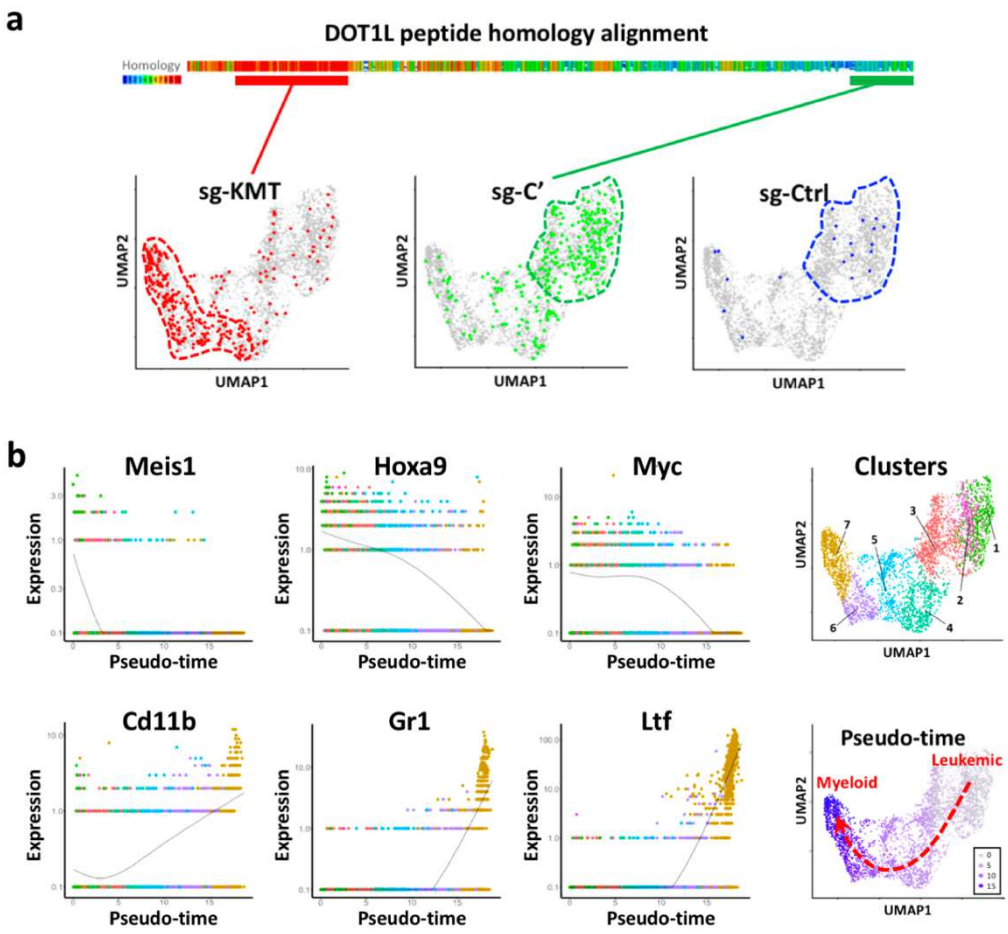


Figure S5. Distribution and gene expression pattern of sc-Tiling single cells. (a) Distribution of single cells harboring sgRNAs targeting the KMT core (red; 56 sgRNAs), C-terminus (green; 54 sgRNAs), or negative controls (blue; 5 sgRNAs) on UMAP. Peptide sequence homology was analyzed using PRALINE multiple sequence alignment[46] at http://www.ibi.vu.nl/programs/pralinewww. (b) Association of the pseudo-time projection with the progressive reduction of leukemia-associated genes (Meis1, Hoxa9, Myc) or induction of myeloid-differentiation genes (Cd11b, Gr1, Ltf). Colored dots indicate single cells in each cell cluster (1–7) identified on UMAP (upper-right). Annotation of pseudo-time projection (bottom-right; purple gradient) indicates clustering of the leukemic (right) vs. myeloid (left) cells on UMAP.

Figure S6

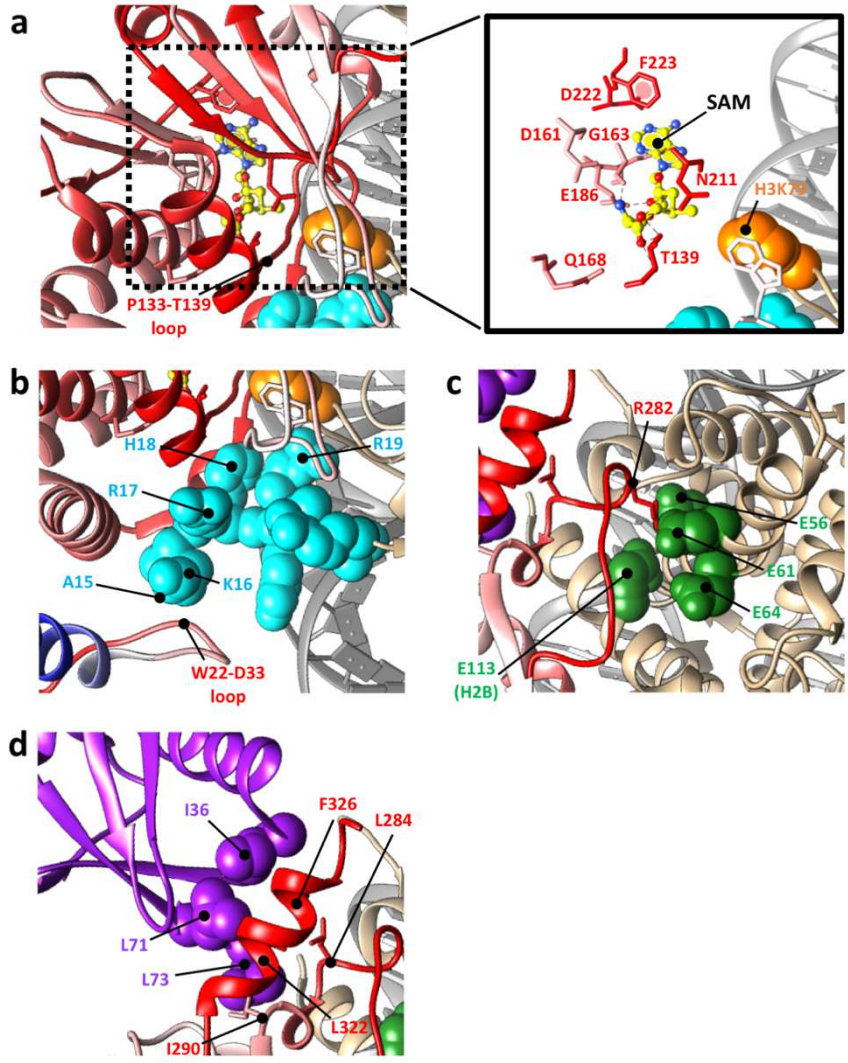
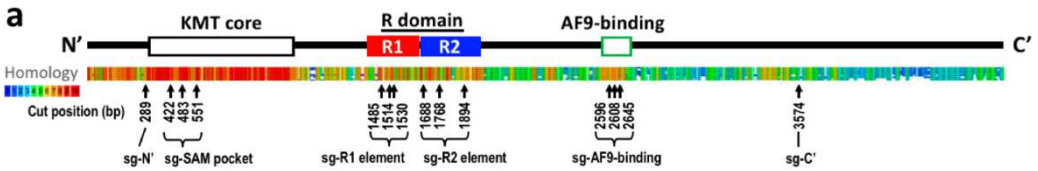


Figure S6. Three-dimensional annotation of DOT1L sc-Tiling. Mapping of smoothed pseudo-time score to a cryo-EM structural model of "active state" DOT1L (residues M1–P332) bound to a ubiquitylated nucleosome (PDB ID: 6NQA)[28], as shown in Fig. 1H. (a) DOT1L catalytic pocket nearby histone H3K79 (orange spheres; methylation target of DOT1L) including the DOT1L residues directly interacting with the enzymatic substrate SAM (colored sticks). (b) DOT1L W22–D33 loop stabilizes the interaction with histone H4 N-terminal tail (cyan spheres; H4 residues A15–R19 are labeled). (c) DOT1L R282 loop interacts with an H2A/H2B acidic patch (green spheres; histone H2A E56/E61/E64 and H2B E113). (d) DOT1L T320–K330 helix (L322 and F326) and nearby residues (L284 and I290) interact with H2BK120

conjugated to ubiquitin (purple). The DOT1L contact points I36/L71/L73 on ubiquitin are labeled as purple spheres. Histones (gold; including H2A, H2B, H3, and H4) and DNA (grey) are shown.

Figure S7

a



b

sgRNA name	Target gene	ID	sgRNA sequence	CRISPR cut position	
				cDNA (bp)	Peptide (aa)
sg-N'	Mouse Dot1l	sg_mDot1l_289_a	GGCCGAGTATTCAGCTTCAT	289	97
sg-SAM pocket	Mouse Dot1l	sg_mDot1l_422_a	CTGGGCCACCAGGTCAAAGG	422	141
	Mouse Dot1l	sg_mDot1l_483_s	GATGACCTGTTTGTCGACCT	483	161
	Mouse Dot1l	sg_mDot1l_551_s	CTGCAAACATCACTACGGAG	551	184
sg-R1 element	Mouse Dot1l	sg_mDot1l_1485_a	TTGTATAAGCCAGGAACTGC	1485	495
	Mouse Dot1l	sg_mDot1l_1514_s	TACAAAGACCCCGCAGTACA	1514	505
	Mouse Dot1l	sg_mDot1l_1530_a	CCTGGTCCAGAAGCTGCTGT	1530	510
sg-R2 element	Mouse Dot1l	sg_mDot1l_1688_s	GACCTACAATGACCTGATTC	1688	563
	Mouse Dot1l	sg_mDot1l_1768_s	AGAAGGACAACAGCGAGCTT	1768	590
	Mouse Dot1l	sg_mDot1l_1894_s	GCCAGATCTCAGAGAAGCAG	1894	632
sg-AF9 binding	Mouse Dot1l	sg_mDot1l_2596_a	GTGCTGAGTGGAATGCTAAC	2596	866
	Mouse Dot1l	sg_mDot1l_2608_a	TTGGGCTGCACTGTGCTGAG	2608	870
	Mouse Dot1l	sg_mDot1l_2645_s	GCTGCCGGTCAGCATCCCAC	2645	882
sg-C'	Mouse Dot1l	sg_mDot1l_3574_a	CTGGGCTCATCCTCAGAGCC	3574	1192
sgLuc	Firefly Luciferase	sg_Luc_474	GATTCTAAAACGGATTACCA	-	-

Figure S7. Individual sgRNAs targeting mouse Dot1l coding regions for functional validation. (a) Schematic outline of the mouse DOT1L protein domains, peptide homology (by PRALINE multiple sequence alignment), and select sgRNA targeting positions. (b) Specifications of individual sgRNAs targeting indicated regions of mouse *Dot1l* and *Firefly luciferase*.

Figure S8

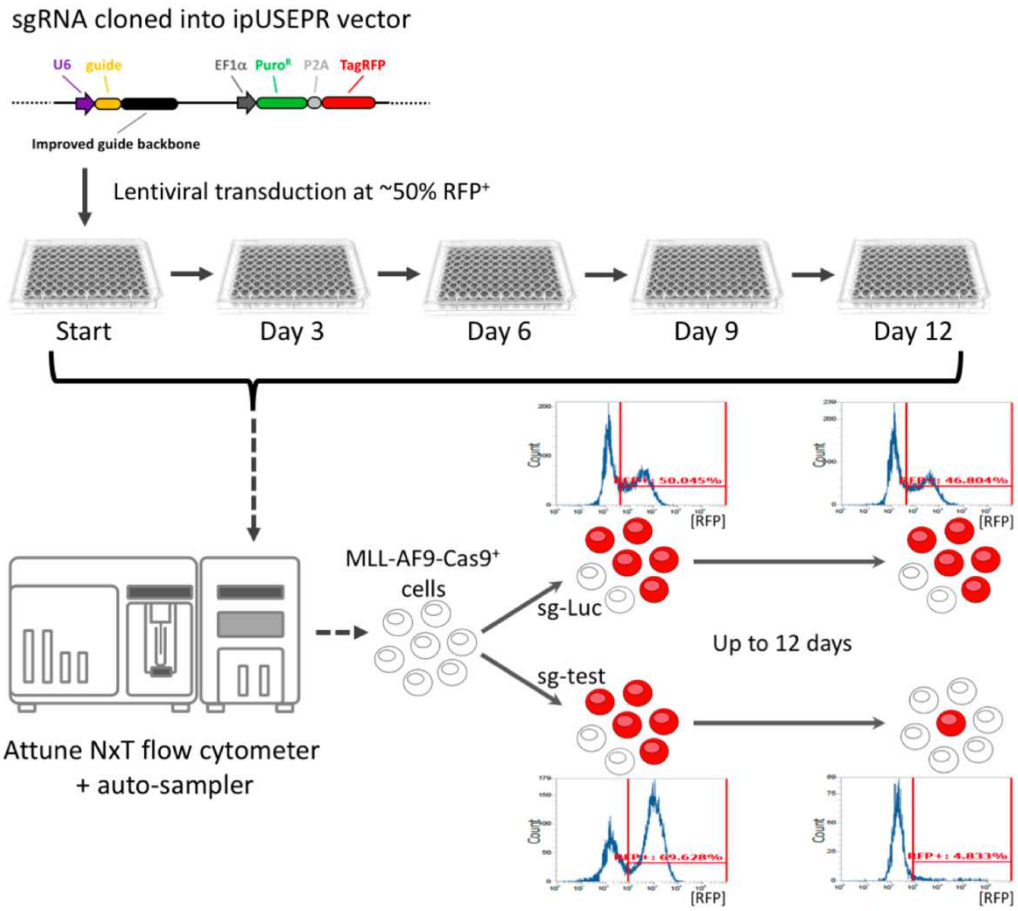


Figure S8. Schematic outline of the RFP flow cytometric growth competition assay.

Figure S9

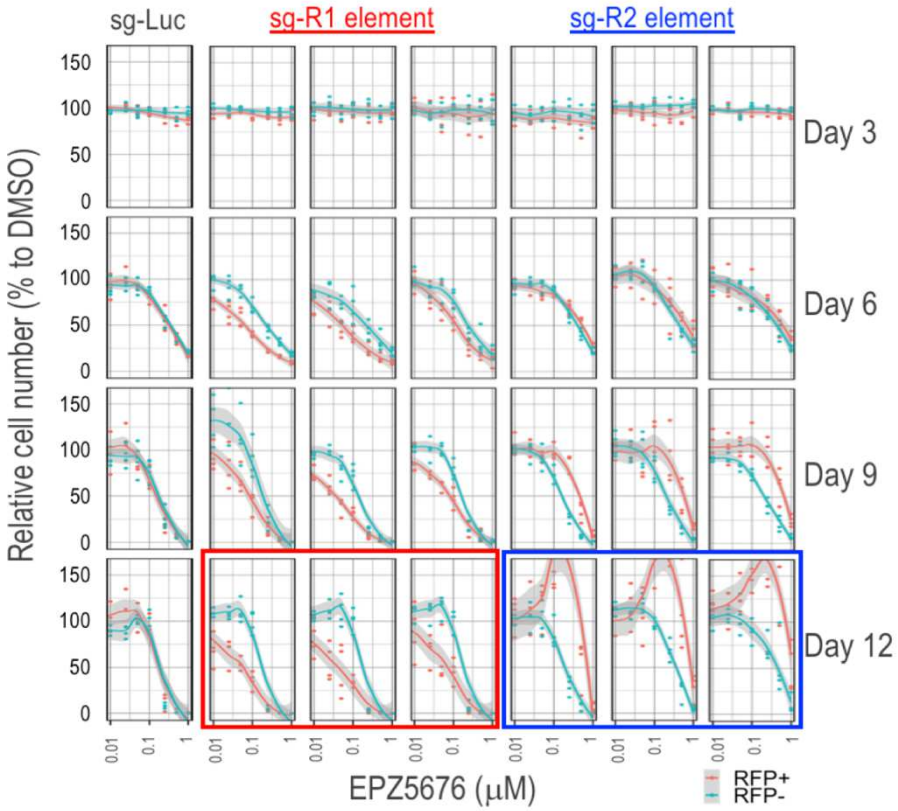
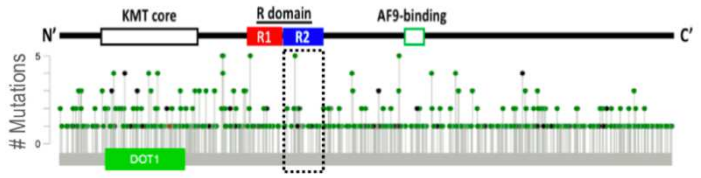


Figure S9. Effect of individual sgRNAs targeting the R1 or R2 elements of DOT1L in response to EPZ5676 in MLL-AF9-Cas9^{+} leukemia cells. Number of RFP^{+} (sgRNA transduced) and RFP^{-} (non-transduced) cells were measured by flow cytometer (Attune NxT, ThermoFisher). Data represent the observed values and mean ± 95% confidence interval (n = 4).

Figure S10

a



Protein Change	Study	Sample ID	Cancer Type	Mutation Type	Copy #	Allele Freq (T)
A567V	Uterine Corpus E...	TCGA-B5-A3FC-01	Uterine Endometrioid Carcinoma	Missense	Diploid	0.29
A573T	MSK-IMPACT Clini...	P-0010275-T01-IM5	Rectal Adenocarcinoma	Missense	Diploid	0.15
Q584P	MSK-IMPACT Clini...	P-0006571-T01-IM5	Head and Neck Squamous Cell Ca...	Missense	Diploid	0.15
R590C	Colon Adenocarci...	TCGA-CK-4951-01	Mucinous Adenocarcinoma of the...	Missense	Diploid	0.15
A591T	Lung Adenocarcin...	TCGA-62-8395-01	Lung Adenocarcinoma	Missense	ShallowDel	0.29
A591V	Stomach Adenocar...	TCGA-CG-5733-01	Stomach Adenocarcinoma	Missense	Diploid	0.37
R593C	Colon Adenocarci...	TCGA-DM-A28F-01	Colon Adenocarcinoma	Missense	Gain	0.36
R593H	Colon Adenocarci...	TCGA-CM-4750-01	Colon Adenocarcinoma	Missense	Diploid	0.58
G594D	MSK-IMPACT Clini...	P-0004051-T01-IM5	Colon Adenocarcinoma	Missense	Diploid	0.29
G594S	MSK-IMPACT Clini...	P-0001851-T01-IM3	Adrenocortical Carcinoma	Missense	Diploid	0.49
K601M	MSK-IMPACT Clini...	P-0009591-T01-IM5	Lung Adenocarcinoma	Missense	Diploid	
C604Y	MSK-IMPACT Clini...	P-0001448-T01-IM3	Pleomorphic Liposarcoma	Missense	Diploid	0.12
E605D	Recurrent and Me...	P-0000573-T01-IM3	Oral Cavity Squamous Cell Carc...	Missense	Diploid	0.2
L609Q	MSK-IMPACT Clini...	P-0000226-T01-IM3	Alveolar Rhabdomyosarcoma	Missense	Diploid	0.44
D610Y	MSK-IMPACT Clini...	P-0010130-T02-IM5	Small Cell Lung Cancer	Missense	Diploid	0.27
T613M	Head and Neck Sq...	TCGA-CV-7097-01	Head and Neck Squamous Cell Ca...	Missense	ShallowDel	0.37
L626P	MSK-IMPACT Clini...	P-0007296-T01-IM5	Colon Adenocarcinoma	Missense	Diploid	0.37
C637G	Lung Adenocarcin...	TCGA-69-7974-01	Lung Adenocarcinoma	Missense	Diploid	0.06
E646K	MSK-IMPACT Clini...	P-0000138-T01-IM3	Breast Mixed Ductal and Lobula...	Missense	Diploid	0.28

b

Protein Change	Wild type	Mutant
A567V	GCG	GTG
A573T	GCC	ACC
Q584P	CAG	CCG
R590C	CGC	TGC
A591T	GCG	ACG
A591V	GCG	GTG
R593C	CGC	TGC
R593H	CGC	CAC
G594D	GGC	GAC
G594S	GGC	AGC
K601M	AAG	ATG
C604Y	TGC	TAC
E605D	GAG	GAC
L609Q	CTG	CAG
D610Y	GAC	TAC
T613M	ACG	ATG
L626P	CTG	CCG
C637G	TGC	GGC
E646K	GAG	AAG

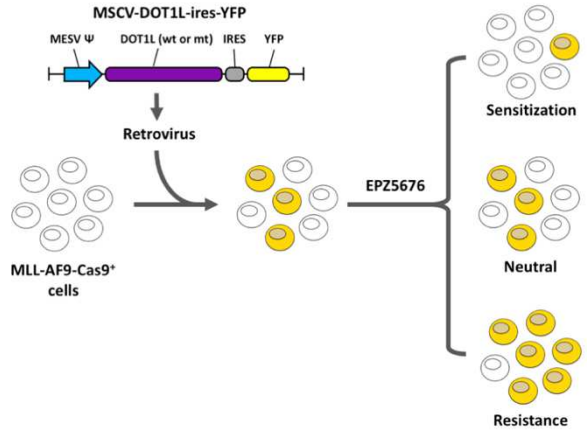


Figure S10. Clinically observed *DOT1L* variant alleles in the R2 element. (a) A total of 779 missense DOT1L variants were identified in cBioPortal database (missense mutations labeled as green in the lolliplot). Table shows a total of 19 missense variants observed in the R2 element (residues A558–C662). (b) Mutation of DOT1L cDNAs and schematic outline of the YFP flow cytometric growth competition assay.

Figure S11

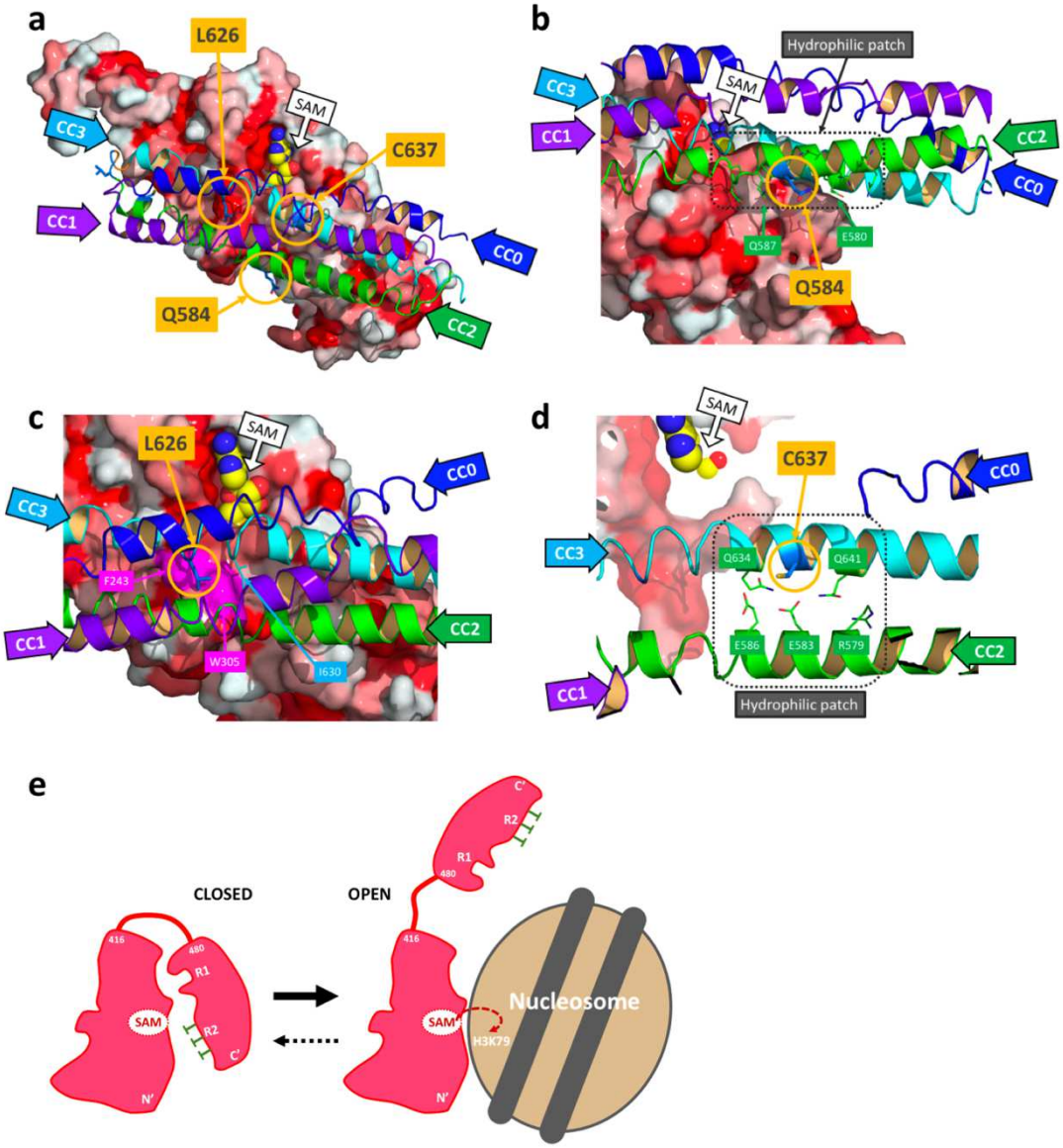


Figure S11. Binding model of the DOT1L R domain with the KMT domain and drug-resistant mutations. (a) The DOT1L KMT domain (residues M1–P332; PDB ID 3QOW) is shown as surface with hydrophobic residues colored in red. The enzymatic substrate SAM is displayed as spheres. The R domain is depicted as a cartoon shown with CC0 (blue), CC1 (purple-blue), CC2 (green), and CC3 (cyan) helices. The residues identified as drug-resistant mutations (Q584P, L626P, and C637G) in the R2 element (CC2 and CC3) are shown as colored sticks and their contact residues are displayed as lines. (b) Q584 is located

in the center of a hydrophilic patch (dashed-line box) on both CC2 (Q576, Q577, R579, E580, Q581, E583, Q587, D588, N589) and CC3 (Q634, Q641) helices. Q584 stabilizes the CC2 helical structure via hydrophilic interactions with its neighboring residues E580 and Q587. Mutation of Q584 to a hydrophobic proline (Q584P), which will either break or kink the helix, will destabilize the CC2 helix structure and weaken the interaction between CC2 and CC3. (c) L626 is located in the middle of the CC3 helix and is a key residue by which the R domain interacts with the KMT domain (F243 and W305; shown as a pink patch) via hydrophobic interaction. It also stabilizes the CC3 helix by hydrophobic interaction with I630. As proline tends to break helical structures, the L626P mutation may disrupt the CC3 helical structure and weaken the interaction between the R domain and KMT domain. (d) C637 is located in the center of a hydrophilic patch composed of Q634/C637/Q641 in CC3 and R579/E583/E586 in CC2 (dashed-line box) that stabilizes the helical structure and establishes the CC2/CC3 interaction. Mutation of C637 to a glycine (C637G) may disrupt the CC3 helical structure (due to the high conformational flexibility of glycine) and the interaction between CC2 and CC3. (e) Cartoon representation of the missense mutations (green) in the DOT1L R2 element, which favor the transition from a closed (left) to an open (right) state of the DOT1L KMT domain, thereby facilitating DOT1L-to-nucleosome interactions and H3K79 methylation.

References:

1 Klein, S., McCormick, F. & Levitzki, A. Killing time for cancer cells. *Nat Rev Cancer* **5**, 573-580, doi:10.1038/nrc1651 (2005).

2 Shalem, O., Sanjana, N. E. & Zhang, F. High-throughput functional genomics using CRISPR-Cas9. *Nat Rev Genet* **16**, 299-311, doi:10.1038/nrg3899 (2015).

3 Tsherniak, A. *et al.* Defining a Cancer Dependency Map. *Cell* **170**, 564-576 e516, doi:10.1016/j.cell.2017.06.010 (2017).

4 Shi, J. *et al.* Discovery of cancer drug targets by CRISPR-Cas9 screening of protein domains. *Nat Biotechnol* **33**, 661-667, doi:10.1038/nbt.3235 (2015).

5 Munoz, D. M. *et al.* CRISPR Screens Provide a Comprehensive Assessment of Cancer Vulnerabilities but Generate False-Positive Hits for Highly Amplified Genomic Regions. *Cancer Discov* **6**, 900-913, doi:10.1158/2159-8290.CD-16-0178 (2016).

6 Schoonenberg, V. A. C. *et al.* CRISPRO: identification of functional protein coding sequences based on genome editing dense mutagenesis. *Genome Biol* **19**, 169, doi:10.1186/s13059-018-1563-5 (2018).

7 He, W. *et al.* De novo identification of essential protein domains from CRISPR-Cas9 tiling-sgRNA knockout screens. *Nat Commun* **10**, 4541, doi:10.1038/s41467-019-12489-8 (2019).

8 Ipsaro, J. J. *et al.* Rapid generation of drug-resistance alleles at endogenous loci using CRISPR-Cas9 indel mutagenesis. *PLoS One* **12**, e0172177, doi:10.1371/journal.pone.0172177 (2017).

9 Adamson, B. *et al.* A Multiplexed Single-Cell CRISPR Screening Platform Enables Systematic Dissection of the Unfolded Protein Response. *Cell* **167**, 1867-1882 e1821, doi:10.1016/j.cell.2016.11.048 (2016).

10 Jaitin, D. A. *et al.* Dissecting Immune Circuits by Linking CRISPR-Pooled Screens with Single-Cell RNA-Seq. *Cell* **167**, 1883-1896 e1815, doi:10.1016/j.cell.2016.11.039 (2016).

11 Datlinger, P. *et al.* Pooled CRISPR screening with single-cell transcriptome readout. *Nat Methods* **14**, 297-301, doi:10.1038/nmeth.4177 (2017).

12 Giladi, A. *et al.* Single-cell characterization of haematopoietic progenitors and their trajectories in homeostasis and perturbed haematopoiesis. *Nat Cell Biol* **20**, 836-846, doi:10.1038/s41556-018-0121-4 (2018).

13 Gasperini, M. *et al.* A Genome-wide Framework for Mapping Gene Regulation via Cellular Genetic Screens. *Cell* **176**, 377-390 e319, doi:10.1016/j.cell.2018.11.029 (2019).

14 McFaline-Figueroa, J. L. *et al.* A pooled single-cell genetic screen identifies regulatory checkpoints in the continuum of the epithelial-to-mesenchymal transition. *Nat Genet* **51**, 1389-1398, doi:10.1038/s41588-019-0489-5 (2019).

15 Tian, R. *et al.* CRISPR Interference-Based Platform for Multimodal Genetic Screens in Human iPSC-Derived Neurons. *Neuron* **104**, 239-255 e212, doi:10.1016/j.neuron.2019.07.014 (2019).

16 Norman, T. M. *et al.* Exploring genetic interaction manifolds constructed from rich single-cell phenotypes. *Science* **365**, 786-793, doi:10.1126/science.aax4438 (2019).

17 Chen, C. W. & Armstrong, S. A. Targeting DOT1L and HOX gene expression in MLL-rearranged leukemia and beyond. *Exp Hematol* **43**, 673-684, doi:10.1016/j.exphem.2015.05.012 (2015).

18 Chan, A. K. N. & Chen, C. W. Rewiring the Epigenetic Networks in MLL-Rearranged Leukemias: Epigenetic Dysregulation and Pharmacological Interventions. *Front Cell Dev Biol* **7**, 81, doi:10.3389/fcell.2019.00081 (2019).
19 Daigle, S. R. *et al.* Potent inhibition of DOT1L as treatment of MLL-fusion leukemia. *Blood* **122**, 1017-1025, doi:10.1182/blood-2013-04-497644 (2013).
20 Stein, E. M. *et al.* The DOT1L inhibitor pinometostat reduces H3K79 methylation and has modest clinical activity in adult acute leukemia. *Blood* **131**, 2661-2669, doi:10.1182/blood-2017-12-818948 (2018).
21 Replogle, J. M. *et al.* Combinatorial single-cell CRISPR screens by direct guide RNA capture and targeted sequencing. *Nat Biotechnol* **38**, 954-961, doi:10.1038/s41587-020-0470-y (2020).
22 Krivtsov, A. V. *et al.* Transformation from committed progenitor to leukaemia stem cell initiated by MLL-AF9. *Nature* **442**, 818-822, doi:10.1038/nature04980 (2006).
23 Chen, C. W. *et al.* DOT1L inhibits SIRT1-mediated epigenetic silencing to maintain leukemic gene expression in MLL-rearranged leukemia. *Nat Med* **21**, 335-343, doi:10.1038/nm.3832 (2015).
24 Butler, A., Hoffman, P., Smibert, P., Papalexi, E. & Satija, R. Integrating single-cell transcriptomic data across different conditions, technologies, and species. *Nat Biotechnol* **36**, 411-420, doi:10.1038/nbt.4096 (2018).
25 Bernt, K. M. *et al.* MLL-rearranged leukemia is dependent on aberrant H3K79 methylation by DOT1L. *Cancer Cell* **20**, 66-78, doi:10.1016/j.ccr.2011.06.010 (2011).
26 Min, J., Feng, Q., Li, Z., Zhang, Y. & Xu, R. M. Structure of the catalytic domain of human DOT1L, a non-SET domain nucleosomal histone methyltransferase. *Cell* **112**, 711-723 (2003).
27 Trapnell, C. *et al.* The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. *Nat Biotechnol* **32**, 381-386, doi:10.1038/nbt.2859 (2014).
28 Worden, E. J., Hoffmann, N. A., Hicks, C. W. & Wolberger, C. Mechanism of Cross-talk between H2B Ubiquitination and H3 Methylation by Dot1L. *Cell* **176**, 1490-1501 e1412, doi:10.1016/j.cell.2019.02.002 (2019).
29 Valencia-Sanchez, M. I. *et al.* Structural Basis of Dot1L Stimulation by Histone H2B Lysine 120 Ubiquitination. *Mol Cell* **74**, 1010-1019 e1016, doi:10.1016/j.molcel.2019.03.029 (2019).
30 Kuntimaddi, A. *et al.* Degree of Recruitment of DOT1L to MLL-AF9 Defines Level of H3K79 Di- and Tri-methylation on Target Genes and Transformation Potential. *Cell Rep* **11**, 808-820, doi:10.1016/j.celrep.2015.04.004 (2015).
31 Deshpande, A. J. *et al.* AF10 regulates progressive H3K79 methylation and HOX gene expression in diverse AML subtypes. *Cancer Cell* **26**, 896-908, doi:10.1016/j.ccell.2014.10.009 (2014).
32 Zhang, H. *et al.* Structural and functional analysis of the DOT1L-AF10 complex reveals mechanistic insights into MLL-AF10-associated leukemogenesis. *Genes Dev* **32**, 341-346, doi:10.1101/gad.311639.118 (2018).
33 Gao, J. *et al.* Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. *Sci Signal* **6**, pl1, doi:10.1126/scisignal.2004088 (2013).
34 Doench, J. G. *et al.* Optimized sgRNA design to maximize activity and minimize off-target effects of CRISPR-Cas9. *Nat Biotechnol* **34**, 184-191, doi:10.1038/nbt.3437 (2016).

35 Canver, M. C. *et al.* BCL11A enhancer dissection by Cas9-mediated in situ saturating mutagenesis. *Nature* **527**, 192-197, doi:10.1038/nature15521 (2015).
36 Langmead, B., Trapnell, C., Pop, M. & Salzberg, S. L. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol* **10**, R25, doi:10.1186/gb-2009-10-3-r25 (2009).
37 Robinson, M. D., McCarthy, D. J. & Smyth, G. K. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**, 139-140, doi:10.1093/bioinformatics/btp616 (2010).
38 Buchan, D. W., Minneci, F., Nugent, T. C., Bryson, K. & Jones, D. T. Scalable web services for the PSIPRED Protein Analysis Workbench. *Nucleic Acids Res* **41**, W349-357, doi:10.1093/nar/gkt381 (2013).
39 Corpet, F. Multiple sequence alignment with hierarchical clustering. *Nucleic Acids Res* **16**, 10881-10890 (1988).
40 Yang, J. *et al.* The I-TASSER Suite: protein structure and function prediction. *Nat Methods* **12**, 7-8, doi:10.1038/nmeth.3213 (2015).
41 Pierce, B. G., Hourai, Y. & Weng, Z. Accelerating protein docking in ZDOCK using an advanced 3D convolution library. *PLoS One* **6**, e24657, doi:10.1371/journal.pone.0024657 (2011).
42 Pettersen, E. F. *et al.* UCSF Chimera--a visualization system for exploratory research and analysis. *J Comput Chem* **25**, 1605-1612, doi:10.1002/jcc.20084 (2004).
43 Burley, S. K. *et al.* RCSB Protein Data Bank: biological macromolecular structures enabling research and education in fundamental biology, biomedicine, biotechnology and energy. *Nucleic Acids Res* **47**, D464-D474, doi:10.1093/nar/gky1004 (2019).
44 Yu, W. *et al.* Catalytic site remodelling of the DOT1L methyltransferase by selective inhibitors. *Nat Commun* **3**, 1288, doi:10.1038/ncomms2304 (2012).
45 Sanjana, N. E., Shalem, O. & Zhang, F. Improved vectors and genome-wide libraries for CRISPR screening. *Nat Methods* **11**, 783-784, doi:10.1038/nmeth.3047 (2014).
46 Simossis, V. A. & Heringa, J. PRALINE: a multiple sequence alignment toolbox that integrates homology-extended and secondary structure information. *Nucleic Acids Res* **33**, W289-294, doi:10.1093/nar/gki390 (2005).

Figures

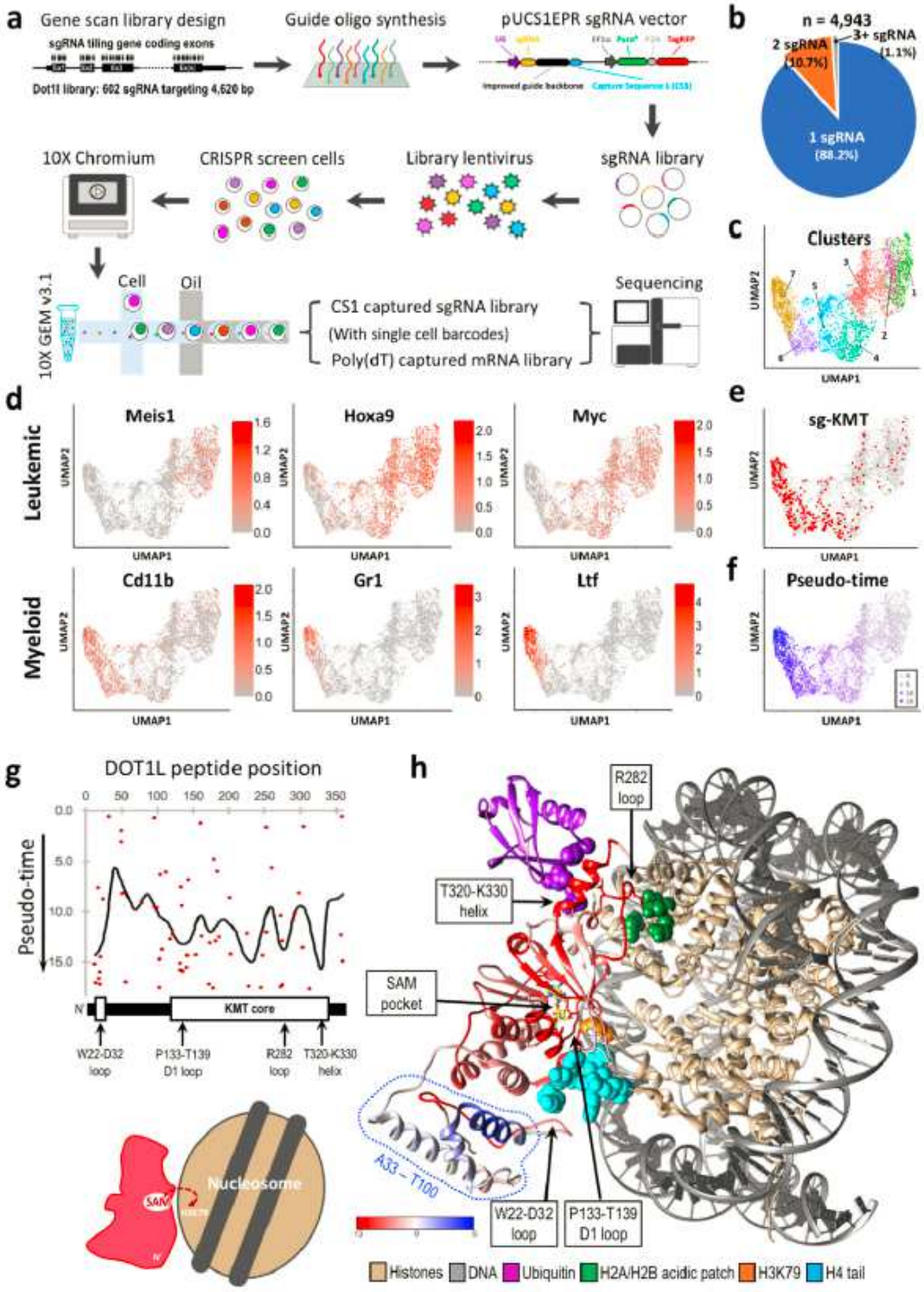


Figure 1

Single-cell CRISPR gene tiling of DOT1L. (a) Schematic outline of sc-Tiling library construction and screening in MLL-AF9-Cas9+ cells. (b) Assignment rates for direct-capture sgRNA. The total number of cells and fraction of cells assigned a single guide, two guides, or more than two guides are indicated. (c)

Two-dimensional projection (UMAP) of cell clusters based on sc-RNAseq of DOT1L-dependent genes. The transcriptionally distinguishable cell populations (1 to 7) are color labelled. (d) Annotation of leukemia-associated (Meis1, Hoxa9, Myc) and myeloid-differentiation (Cd11b, Gr1, Ltf) gene expression on UMAP. (e-f) Annotation of (e) cells harboring sgRNAs targeting the DOT1L KMT core (red) and (f) pseudo-time value (purple gradient) on UMAP. (g) Median pseudo-time of each sgRNA construct (dots) and the smoothed pseudo-time score (line) of the KMT core. (h) Three-dimensional annotation of smoothed pseudo-time score relative to a cryo-EM structural model of “active state” DOT1L (residues M1–P332) bound to a ubiquitylated nucleosome (PDB ID: 6NQA; and a simplified scheme shown on the bottom-left)28. Histones (gold; including H2A, H2B, H3 and H4), DNA (grey), ubiquitin (purple; conjugated to histone H2BK120; DOT1L contact points on ubiquitin are labeled as purple spheres), histone H4 N-terminal tail (cyan spheres), the enzymatic substrate SAM (colored sticks), histone H3K79 (orange spheres), and an H2A/H2B acidic patch (green spheres) are shown. Enlarged images shown in Fig. S6.

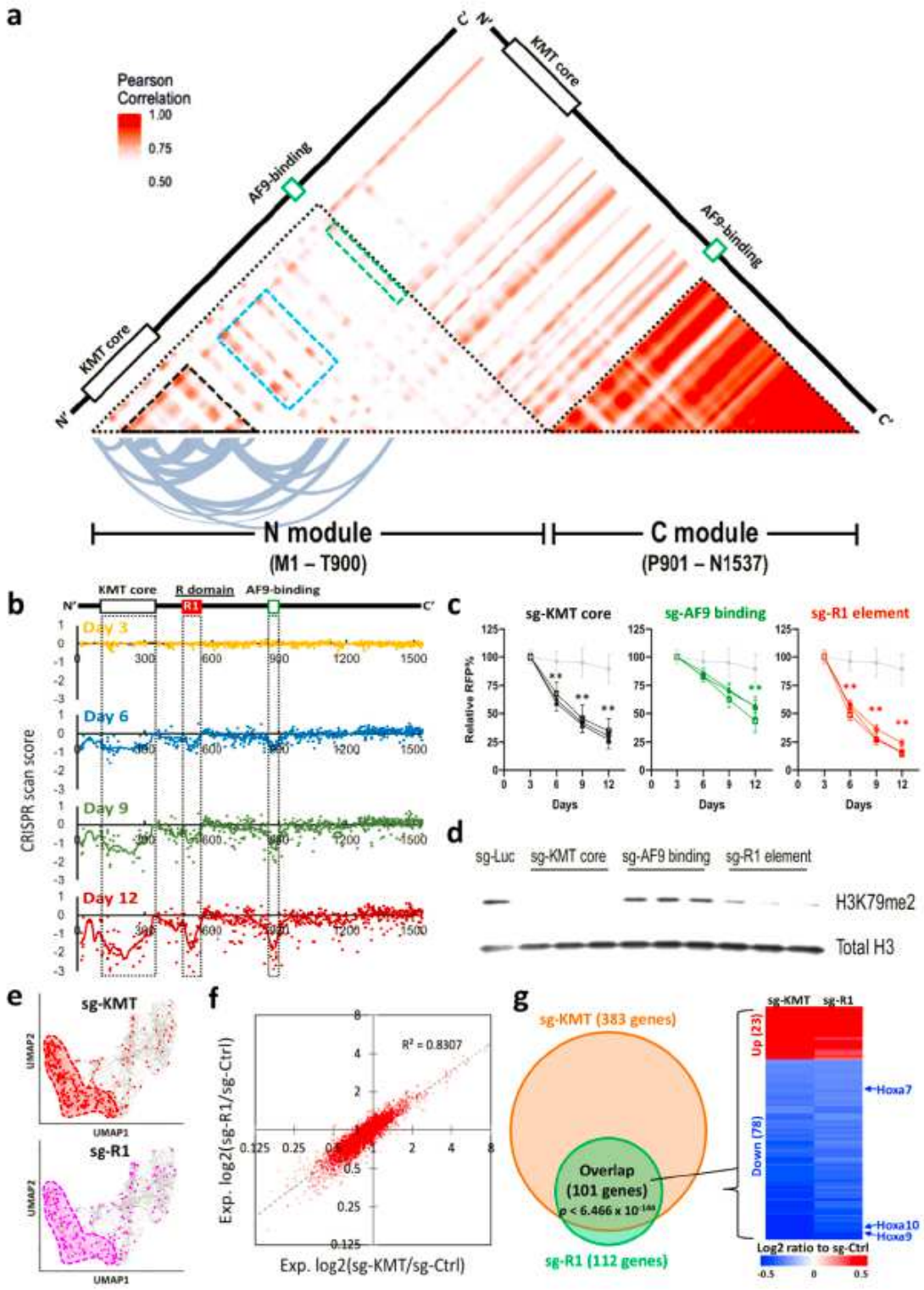


Figure 2

sc-Tiling pinpoints novel functional elements in DOT1L. (a) Heatmap depicts Pearson correlations between sgRNAs targeting different positions across the DOTL1 protein. The curved lines indicate highly correlative (Pearson score > 0.8) residue pairs in the N-module of DOT1L. (b) CRISPR scan score of each sgRNA (dots) and smoothed score (line) of the DOT1L-tiling survival screen in MLL-AF9-Cas9+ leukemia at indicated number of days in culture. (c) Effect of individual sgRNAs targeting the KMT core (black),

AF9-binding motif (green), or R1 element (red) of DOT1L on the proliferation of MLL-AF9-Cas9+ leukemia. Data represent mean ± 95% confidence interval of a quadruplicate experiment. **P < 0.001 by two-sided Student's t-test compared to an sgRNA targeting Luciferase (sg-Luc; grey). (d) Western blot of H3K79me2 and total histone H3 in MLL-AF9-Cas9+ cells expressing indicated sgRNAs. (e) Annotation of cells harboring sgRNAs targeting the KMT core (red) or R1 element (pink) on UMAP. (f) Correlation of gene expression changes induced by sgRNAs targeting the KMT core (x-axis) and R1 element (y-axis) summarized from the sc-Tiling of DOT1L. (g) Overlap of differentially expressed genes in cells harboring sgRNAs targeting the KMT core (orange) and R1 element (green), including the known DOT1L-driven leukemia genes Hoxa7, Hoxa9, and Hoxa10.

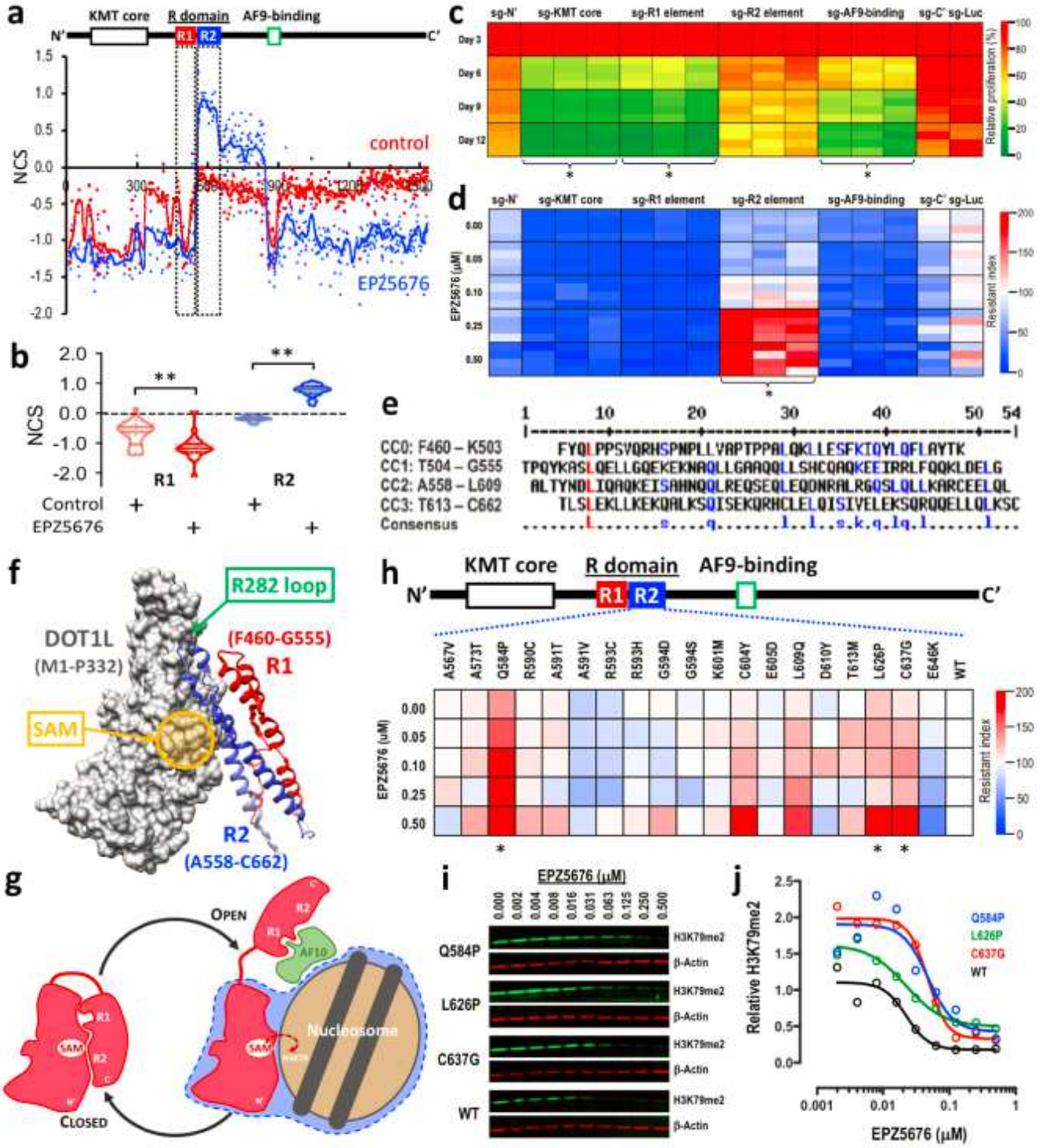


Figure 3

sc-Tiling identifies noncanonical EPZ5676-resistant alleles in human population. (a) Normalized CRISPR score (NCS) of each sgRNA construct (dots) and the smoothed score (line) of the pooled DOT1L-tiling survival screen before vs. after 12 days of treatment in control (red) or 1 μM EPZ5676-treated (blue) MLL-AF9-Cas9+ leukemia cells. Data represent the average of a triplicate experiment. (b) Violin dot plots showing the NCS of each sgRNA targeting the R1 (red; 27 sgRNAs) and R2 (blue; 36 sgRNAs) elements in

control (DMSO) or 1 μM EPZ5676-treated MLL-AF9-Cas9+ leukemia cells. **P < 0.001 by two-sided Student's t-test. (c) Heatmap showing the effect of individual sgRNAs targeting the indicated areas of DOT1L (Fig. S7) on the proliferation of MLL-AF9-Cas9+ leukemia cells on days 3, 6, 9, and 12. Data represent the observed values of a quadruplicate experiment. *Significantly (P < 0.01 by two-sided Student's t-test) more depletion compared to the sgRNA targeting a non-essential N' region (sg-N') on day 12. (d) Heatmap showing EPZ5676 resistance index of MLL-AF9-Cas9+ leukemia cells transduced with sgRNAs targeting the indicated areas of DOT1L (Fig. S7). Data represent the observed values of a quadruplicate experiment. *Significantly (P < 0.01 by two-sided Student's t-test) higher resistance compared to sg-N' at 0.5 μM EPZ5676. (e) Peptide sequence alignment of the R domain (residues F460–C662) in human DOT1L. The predicted alpha-helixes in this coiled-coil domain are designated CC0–CC3 and the consensus residues between the helixes are noted. (f) Computationally modeled structure of the human DOT1L R1 (red) and R2 (blue) coiled-coil domains interacting with the KMT core domain (grey; PDB ID: 3UWP)44. (g) Cartoon representation of the R1/R2 self-regulatory module mediating the closed (left) vs. open (right) states of DOT1L. (h) Heatmap showing EPZ5676 resistance index of MLL-AF9 leukemia cells transduced with human DOT1L cDNA harboring clinically observed variants (from cBioPortal database) in the R2 element. Data represent the averaged values of a quadruplicate experiment. *Significantly (P < 0.01 by two-sided Student's t-test) higher resistance compared to wild-type at 0.5 μM EPZ5676. (i) Western blot images of H3K79me2 (green) and b-Actin (red) and (j) quantitative measurement of relative H3K79me2 level (normalized to b-Actin) in MLL-AF9 leukemia cells transduced with wild-type (WT; black), Q584P (blue), L626P (green), or C637G (red) human DOT1L cDNA. Cells were treated with EPZ5676 for three days.