Supplementary Information

DNA Targeting Specificity of the RNA-guided Cas9 Nuclease

Patrick D. Hsu[1,2,3*], David A. Scott[1,2*], Joshua A. Weinstein[1,2‡], F. Ann Ran[1,2,3‡], Silvana Konermann[1,2], Vineeta Agarwala[1], Yinqing Li[1,2], Eli J. Fine[4], Xuebing Wu[5], Ophir Shalem[1,2], Thomas J. Cradick[4], Luciano A. Marraffini[6], Gang Bao[4], and Feng Zhang[1,2†]

[1] Broad Institute of MIT and Harvard
7 Cambridge Center
Cambridge, MA 02142, USA

[2] McGovern Institute for Brain Research
Department of Brain and Cognitive Sciences
Department of Biological Engineering
Massachusetts Institute of Technology
Cambridge, MA 02139, USA

[3] Department of Molecular and Cellular Biology
Harvard University
Cambridge, MA 02138, USA

[4] Department of Biomedical Engineering
Georgia Institute of Technology and Emory University
Atlanta, Georgia 30332, USA

[5] Koch Institute for Integrative Cancer Research
Massachusetts Institute of Technology
Cambridge, MA 02139, USA

[6] Laboratory of Bacteriology
The Rockefeller University
1230 York Ave.
New York, NY 10065, USA

[*]These authors contributed equally to this work.
[†]To whom correspondence should be addressed: zhang@broadinstitute.org.

Nature Biotechnology: doi:10.1038/nbt.2647

SUPPLEMENTARY METHODS

Cell culture and transfection

Human embryonic kidney (HEK) cell line 293FT (Life Technologies) was maintained in Dulbecco's modified Eagle's Medium (DMEM) supplemented with 10% fetal bovine serum (HyClone), 2mM GlutaMAX (Life Technologies), 100U/mL penicillin, and 100μg/mL streptomycin at 37°C with 5% CO_2 incubation.

293FT cells were seeded onto 6-well plates, 24-well plates, or 96-well plates (Corning) 24 hours prior to transfection. Cells were transfected using Lipofectamine 2000 (Life Technologies) at 80-90% confluency following the manufacturer's recommended protocol. For each well of a 6-well plate, a total of 1 ug of Cas9+sgRNA plasmid was used. For each well of a 24-well plate, a total of 500ng Cas9+sgRNA plasmid was used unless otherwise indicated. For each well of a 96-well plate, 65 ng of Cas9 plasmid was used at a 1:1 molar ratio to the U6-sgRNA PCR product.

Human embryonic stem cell line HUES9 (Harvard Stem Cell Institute core) was maintained in feeder-free conditions on GelTrex (Life Technologies) in mTesR medium (Stemcell Technologies) supplemented with 100ug/ml Normocin (InvivoGen). HUES9 cells were transfected with Amaxa P3 Primary Cell 4-D Nucleofector Kit (Lonza) following the manufacturer's protocol.

SURVEYOR nuclease assay for genome modification

293FT and HUES9 cells were transfected with DNA as described above. Cells were incubated at 37ºC for 72 hours post-transfection prior to genomic DNA extraction. Genomic DNA was extracted using the QuickExtract DNA Extraction Solution (Epicentre) following the manufacturer's protocol. Briefly, pelleted cells were resuspended in QuickExtract solution and incubated at 65ºC for 15 minutes, 68°C for 15 minutes, and 98ºC for 10 minutes.

The genomic region flanking the CRISPR target site for each gene was PCR amplified (primers listed in Supplementary Table 2), and products were purified using QiaQuick Spin Column (Qiagen) following the manufacturer's protocol. 400ng total of the purified PCR products were mixed with 2□l 10X Taq DNA Polymerase PCR buffer (Enzymatics) and ultrapure water to a final volume of 20□l, and subjected to a re-annealing process to enable heteroduplex formation: 95°C for 10min, 95°C to 85°C ramping at −2°C/s, 85°C to 25°C at −

0.25ºC/s, and 25ºC hold for 1 minute. After re-annealing, products were treated with SURVEYOR nuclease and SURVEYOR enhancer S (Transgenomics) following the manufacturer's recommended protocol, and analyzed on 4-20% Novex TBE poly-acrylamide gels (Life Technologies). Gels were stained with SYBR Gold DNA stain (Life Technologies) for 30 minutes and imaged with a Gel Doc gel imaging system (Bio-rad). Quantification was based on relative band intensities. Indel percentage was determined by the formula, 100 x (1 - (1 - (b + c) / (a + b + c))$^{1/2}$), where a is the integrated intensity of the undigested PCR product, and b and c are the integrated intensities of each cleavage product.

Northern blot analysis of tracrRNA expression in human cells

Northern blots were performed as previously described[1]. Briefly, RNAs were extracted using the mirPremier microRNA Isolation Kit (Sigma) and heated to 95ºC for 5 min before loading on 8% denaturing polyacrylamide gels (SequaGel, National Diagnostics). Afterwards, RNA was transferred to a pre-hybridized Hybond N+ membrane (GE Healthcare) and crosslinked with Stratagene UV Crosslinker (Stratagene). Probes were labeled with [gamma-32P] ATP (Perkin Elmer) with T4 polynucleotide kinase (New England Biolabs). After washing, membrane was exposed to phosphor screen for one hour and scanned with phosphorimager (Typhoon).

Bisulfite sequencing to assess DNA methylation status

Genomic DNA from 293FT cells was isolated with the DNeasy Blood & Tissue Kit (Qiagen) and bisulfite converted with EZ DNA Methylation-Lightning Kit (Zymo Research). Bisulfite PCR was conducted using KAPA2G Robust HotStart DNA Polymerase (KAPA Biosystems) with primers designed using the Bisulfite Primer Seeker (Zymo Research, Supplementary Table 2). Resulting PCR amplicons were gel-purified, digested with EcoRI and HindIII, and ligated into a pUC19 backbone prior to transformation. Individual clones were then Sanger sequenced to assess DNA methylation status.

In vitro transcription and cleavage assay

Whole cell lysates from 293FT cells were prepared with lysis buffer (20 mM HEPES, 100 mM KCl, 5 mM MgCl2, 1 mM DTT, 5% glycerol, 0.1% Triton X-100) supplemented with Protease Inhibitor Cocktail (Roche). T7-driven sgRNA was transcribed *in vitro* using custom oligos (Supplementary Sequences) and HiScribe T7 *In Vitro* Transcription Kit (NEB), following the manufacturer's recommended protocol. To prepare methylated target sites, pUC19 plasmid was methylated by M.*Sss*I and tested by digestion with HpaII. Unmethylated and successfully methylated pUC19 plasmids were linearized by NheI. The *in vitro* cleavage assay was performed as follows: for a 20 uL cleavage reaction, 10 uL of cell lysate was incubated with 2 uL cleavage buffer (100 mM HEPES, 500 mM KCl, 25 mM MgCl2, 5 mM DTT, 25% glycerol), 1 ug *in vitro* transcribed RNA, and 300 ng pUC19 plasmid DNA.

Deep sequencing to assess targeting specificity

HEK 293FT cells plated in 96-well plates were transfected with Cas9 plasmid DNA and single guide RNA (sgRNA) PCR cassette 72 hours prior to genomic DNA extraction (Supplementary Fig. 4). The genomic region flanking the CRISPR target site for each gene was amplified (Supplementary Fig. 6, Supplementary Table 5, Supplementary Sequences) by a fusion PCR method to attach the Illumina P5 adapters as well as unique sample-specific barcodes to the target amplicons (schematic described in Supplementary Figure 5). PCR products were purified using EconoSpin 96-well Filter Plates (Epoch Life Sciences) following the manufacturer's recommended protocol.

Barcoded and purified DNA samples were quantified by Quant-iT PicoGreen dsDNA Assay Kit or Qubit 2.0 Fluorometer (Life Technologies) and pooled in an equimolar ratio. Sequencing libraries were then sequenced with the Illumina MiSeq Personal Sequencer (Life Technologies).

Sequencing data analysis and indel detection

MiSeq reads were filtered by requiring an average Phred quality (Q score) of at least 23, as well as perfect sequence matches to barcodes and amplicon forward primers. Reads from on- and off-target loci were analyzed by first performing Smith-Waterman alignments against amplicon sequences that included 50 nucleotides upstream and downstream of the target site (a

total of 120 bp). Alignments, meanwhile, were analyzed for indels from 5 nucleotides upstream to 5 nucleotides downstream of the target site (a total of 30 bp). Analyzed target regions were discarded if part of their alignment fell outside the MiSeq read itself, or if matched base-pairs comprised less than 85% of their total length.

Negative controls for each sample provided a gauge for the inclusion or exclusion of indels as putative cutting events. For each sample, an indel was counted only if its quality score exceeded $\mu - \sigma$, where μ was the mean quality-score of the negative control corresponding to that sample and σ was the standard deviation of the same. This yielded whole target-region indel rates for both negative controls and their corresponding samples. Using the negative control's per-target-region-per-read error rate, q, the sample's observed indel count n, and its read-count R, a maximum-likelihood estimate for the fraction of reads having target-regions with true-indels, p, was derived by applying a binomial error model, as follows.

Letting the (unknown) number of reads in a sample having target regions incorrectly counted as having at least 1 indel be E, we can write (without making any assumptions about the number of true indels)

$$\text{Prob}(E|p) = \binom{R(1-p)}{E} q^E (1-q)^{R(1-p)-E}$$

since $R(1-p)$ is the number of reads having target-regions with no true indels. Meanwhile, because the number of reads observed to have indels is n, $n = E + Rp$, i.e. the number of reads having target-regions with errors but no true indels *plus* the number of reads whose target-regions *correctly* have indels. We can then re-write the above

$$\text{Prob}(E|p) = \text{Prob}(n = E + Rp|p) = \binom{R(1-p)}{n-Rp} q^{n-Rp} (1-q)^{R-n}$$

Taking all values of the frequency of target-regions with true-indels p to be equally probable *a priori*, $\text{Prob}(n|p) \propto \text{Prob}(p|n)$. The maximum-likelihood estimate (MLE) for the frequency of target regions with true-indels was therefore set as the value of p that maximized $\text{Prob}(n|p)$. This was evaluated numerically.

In order to place error bounds on the true-indel read frequencies in the sequencing libraries themselves, Wilson score intervals[2] were calculated for each sample, given the MLE-estimate for true-indel target-regions, Rp, and the number of reads R. Explicitly, the lower bound l and upper bound u were calculated as

$$l = \left(Rp + \frac{z^2}{2} - z\sqrt{Rp(1-p) + z^2/4} \right) / (R + z^2)$$

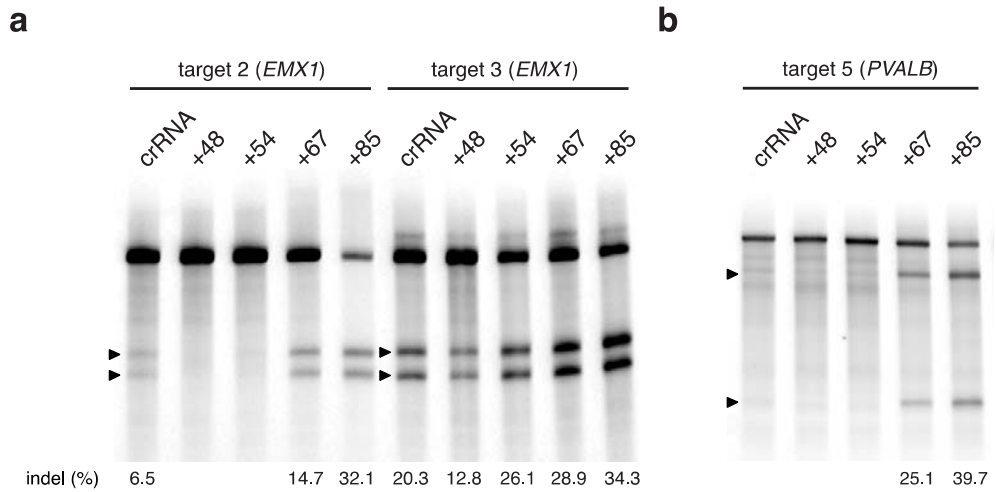
$$u = \left(Rp + \frac{z^2}{2} + z\sqrt{Rp(1-p) + z^2/4} \right) / (R + z^2)$$

where z, the standard score for the confidence required in normal distribution of variance 1, was set to 1.96, meaning a confidence of 95%. The maximum upper bounds and minimum lower bounds for each biological replicate are listed in Supplementary Tables 5-8.

qRT-PCR analysis of relative Cas9 and sgRNA expression

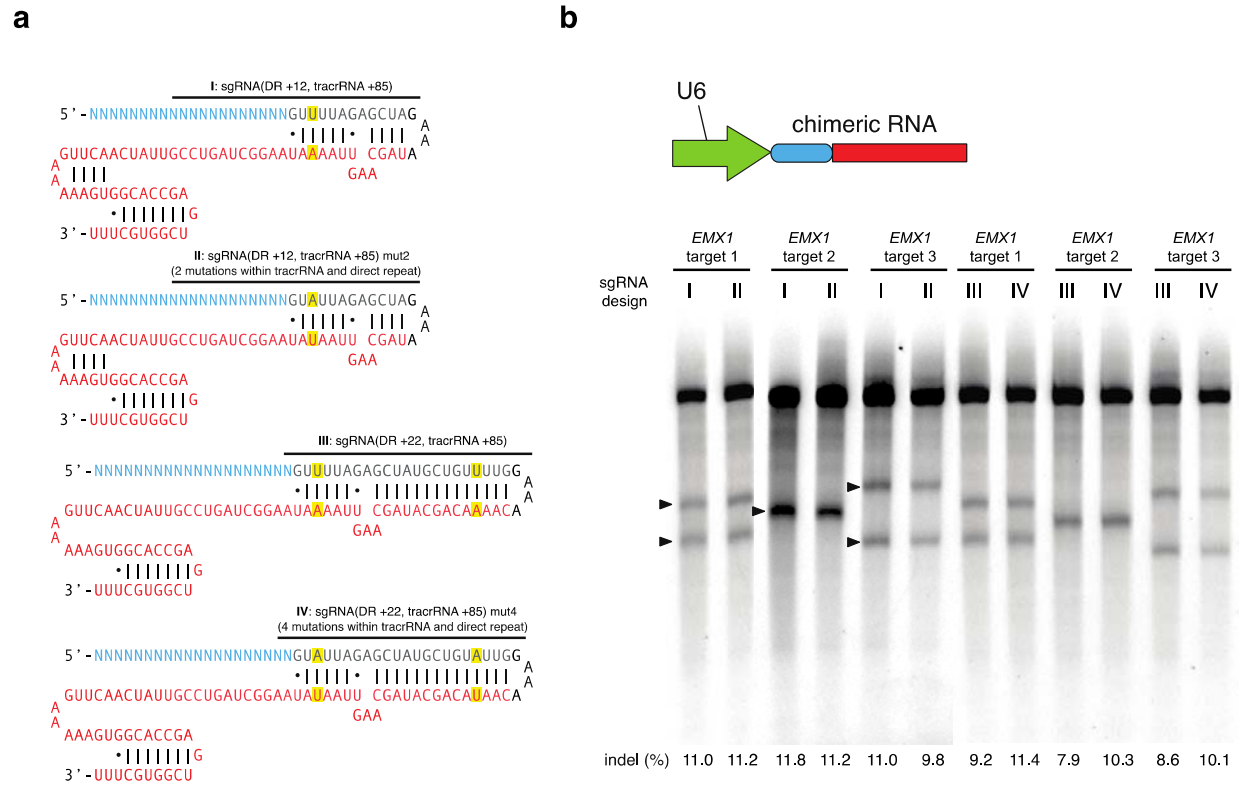
72 hours post-transfection, total RNA from 293FT cells was harvested with miRNeasy Micro Kit (Qiagen). Reverse-strand synthesis for sgRNAs was performed with qScript Flex cDNA kit (VWR) and custom first-strand synthesis primers (Supplementary Table 2). qPCR analysis was performed with Fast SYBR Green Master Mix (Life Technologies) and custom primers (Supplementary Table 2), using GAPDH as an endogenous control. Relative quantification was calculated by the ΔΔCT method.

SUPPLEMENTARY FIGURE 1



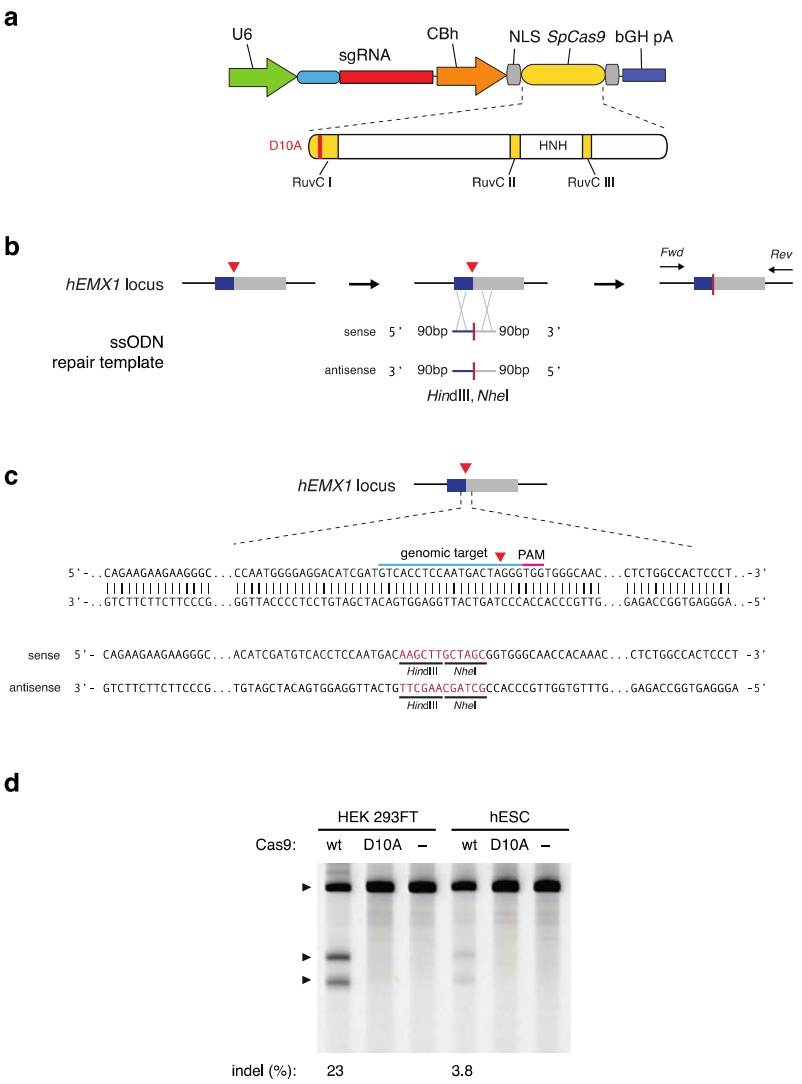
Supplementary Figure 1 | Modification efficiencies of CRISPR-Cas system for additional human genomic targets. DNA expression vectors carrying SpCas9 and crRNA-tracrRNA pair or single guide RNA (sgRNA) are co-transfected into 293FT cells. Cleavage efficiency (% indel) is assessed using the SURVEYOR nuclease assay as described[1]. Modification efficiencies at **a,** 2 *EMX1* loci and **b,** 1 *PVALB* locus are shown. All target site sequences are listed in **Supplementary Table 1**. Arrows indicate the expected SURVEYOR fragments.

SUPPLEMENTARY FIGURE 2



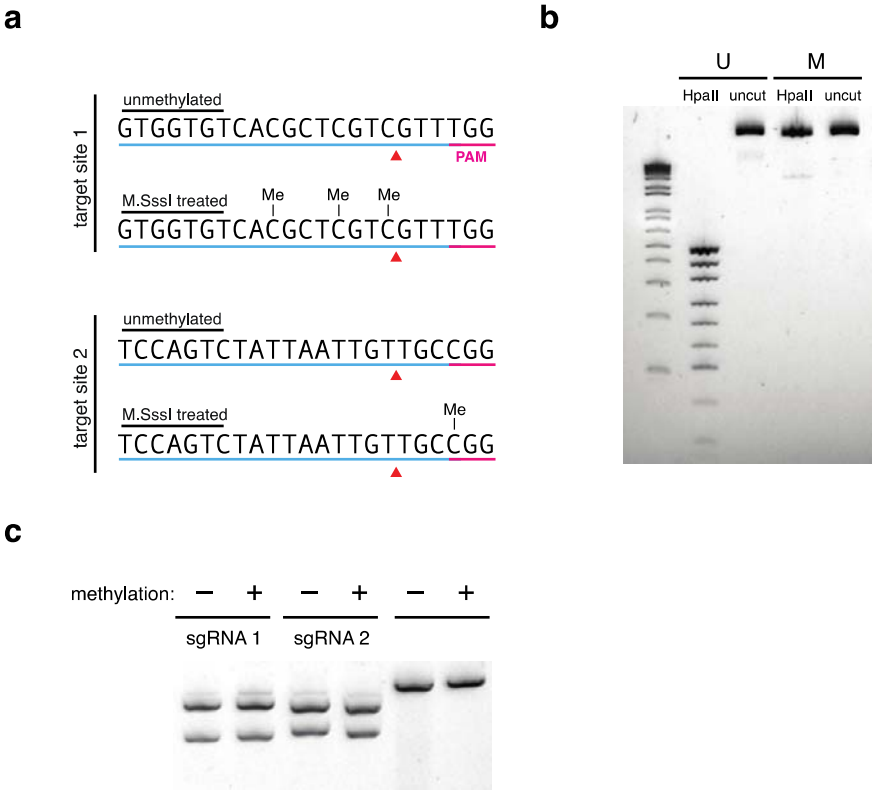
Supplementary Figure 2 | Further optimization of CRISPR-Cas sgRNA architecture. a, Schematic of four additional sgRNA architectures, I-IV. Each consists of a 20-nt guide sequence (blue) joined to the direct repeat (DR, grey), which base-pairs with the tracrRNA (red). The DR-tracrRNA hybrid is truncated at +12 or +22, as indicated, with an artificial GAAA stem loop. tracrRNA truncation positions are numbered according to the previously reported transcription start site for tracrRNA (Supplementary Figure 11 of reference)[3]. sgRNA architectures II and IV carry mutations within their poly-U tracts, which could serve as premature transcriptional terminators for U6 promoter. **b,** SURVEYOR assay for SpCas9-mediated indels at the human *EMX1* locus for target sites 1-3. Arrows indicate the expected SURVEYOR fragments ($n = 3$).

SUPPLEMENTARY FIGURE 3



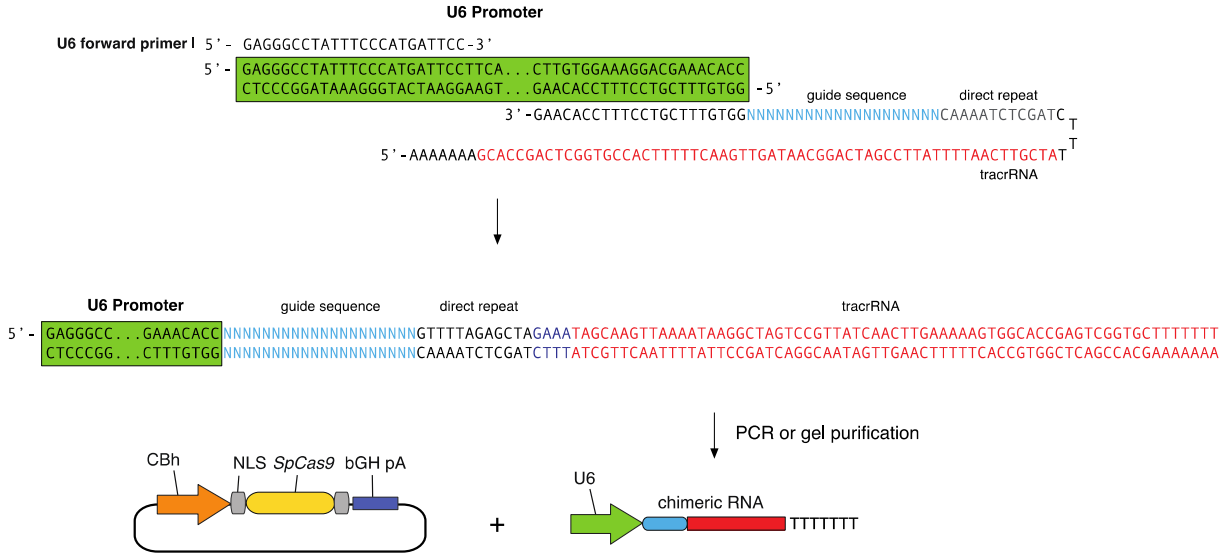
Supplementary Figure 3 | Genome editing via homologous recombination. a, Schematic of SpCas9 nickase, with D10A mutation in the RuvC I catalytic domain. **b,** Schematic representing homologous recombination (HR) at the human *EMX1* locus using either sense or antisense single stranded oligonucleotides as repair templates. Red arrow above indicates sgRNA cleavage site; PCR primers for genotyping (Supplementary Table 2) are indicated as arrows in right panel. **c,** Sequence of region modified by HR. **d,** SURVEYOR assay for wildtype (wt) and nickase (D10A) SpCas9-mediated indels at the *EMX1* target 1 locus ($n = 3$). Arrows indicate positions of expected fragment sizes.

SUPPLEMENTARY FIGURE 4



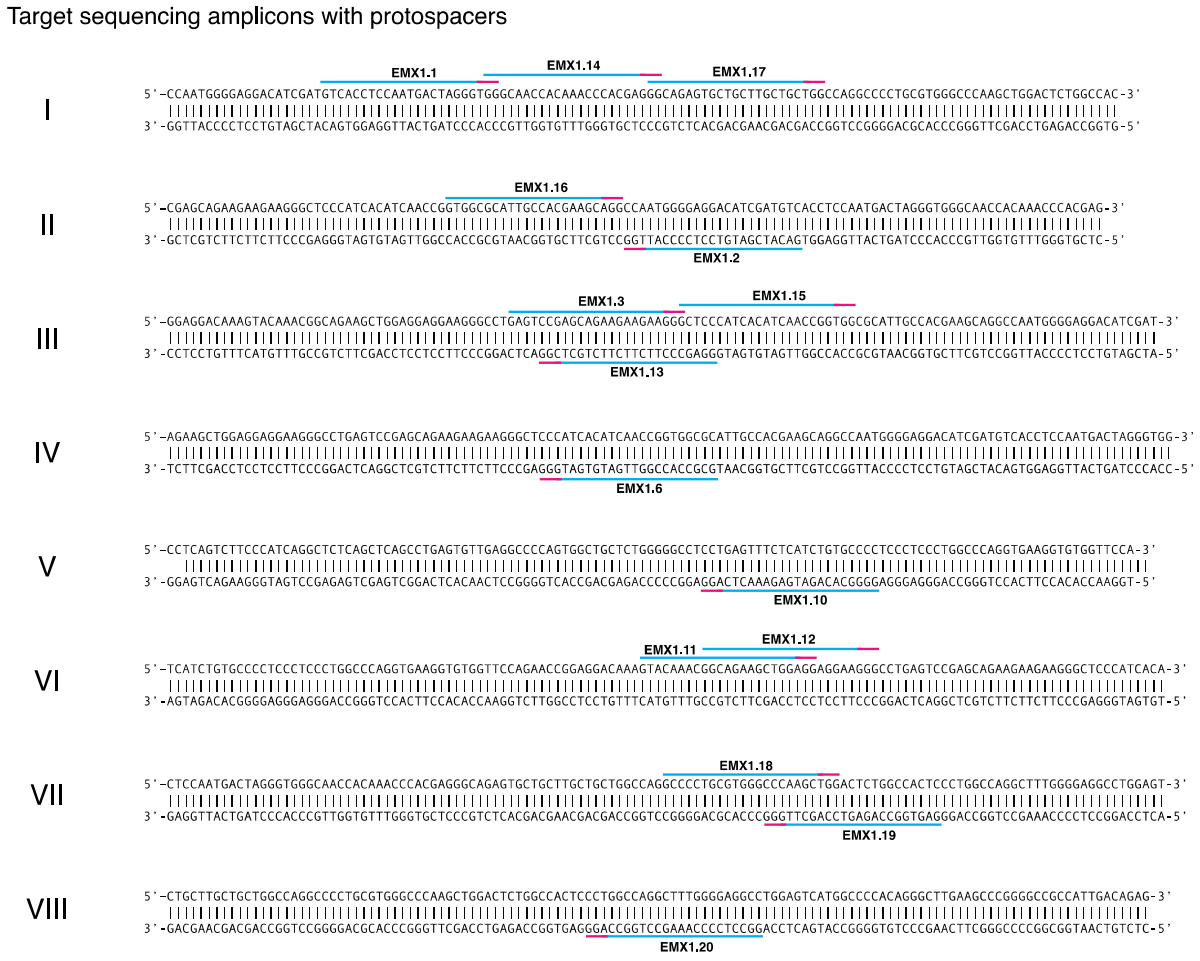
Supplementary Figure 4 | SpCas9 cleaves methylated targets *in vitro*. a, Sequence of CpG dinucleotide-containing targets in pUC19 plasmid methylated *in vitro* by M.*SssI*. Methyl-CpGs in either the target sequence or PAM are indicated; arrows indicate expected cleavage site. **b,** Unmethylated (U) or methylated (M) pUC19 was subjected to restriction digest by the methylation-sensitive restriction enzyme HpaII. Unmethylated pUC19 is digested into a ladder while M.SssI-treated pUC19 is protected from HpaII digestion. **c,** Cleavage of either unmethylated or methylated targets 1 and 2 on linearized pUC19 by SpCas9. No sgRNAs are present in negative control lanes.

SUPPLEMENTARY FIGURE 5



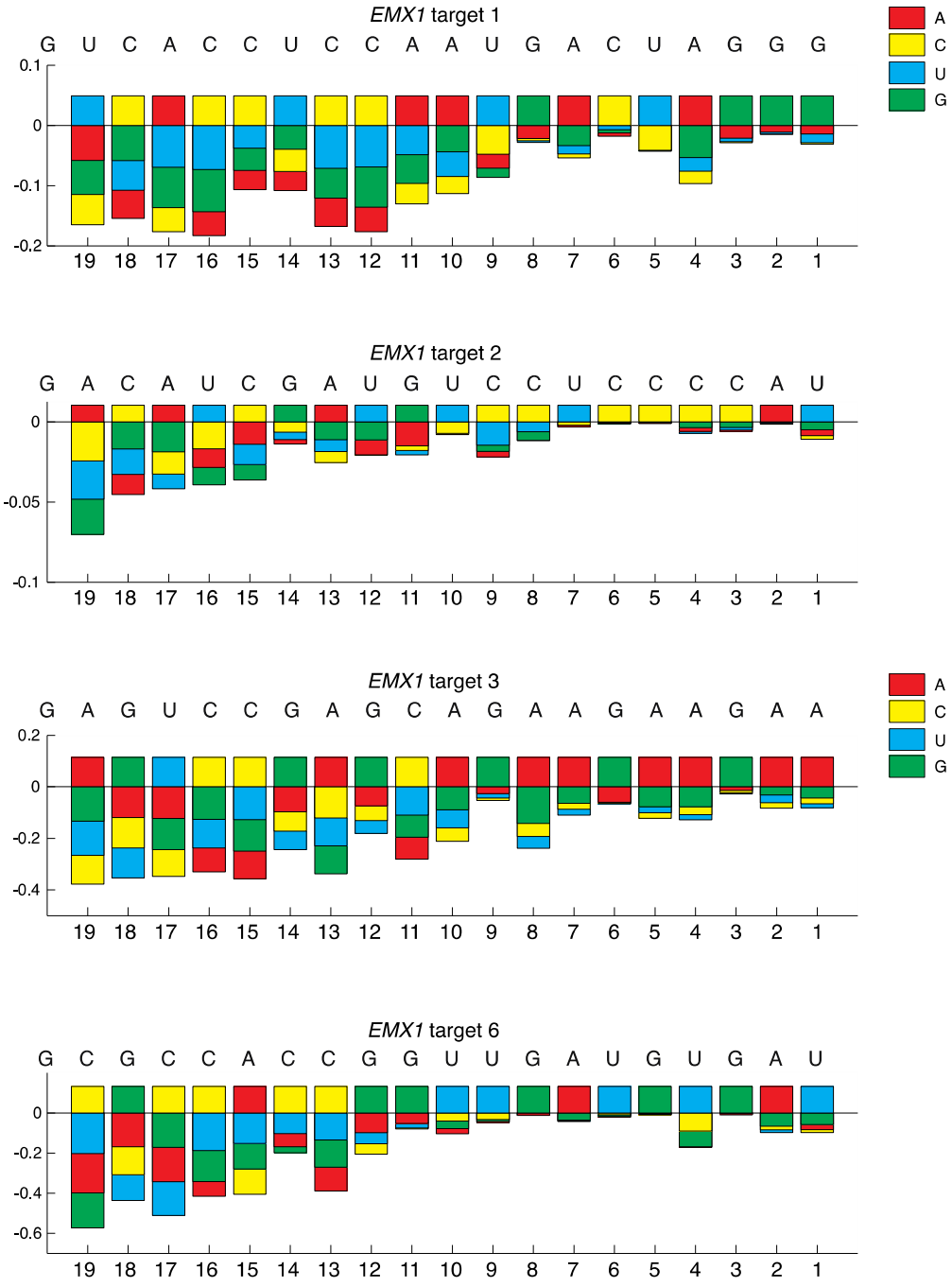
Supplementary Figure 5 | PCR cassette for sgRNA expression. a, Schematic of a PCR-based method for rapid and efficient CRISPR targeting in mammalian cells. A plasmid containing the human RNA polymerase III promoter U6 is PCR-amplified using a U6-specific forward primer and a reverse primer carrying the reverse complement of part of the U6 promoter, the sgRNA(+85) scaffold with guide sequence, and 7 T nucleotides for transcriptional termination. The resulting PCR product is purified and co-delivered with a plasmid carrying Cas9 driven by the CBh promoter.

SUPPLEMENTARY FIGURE 6



Supplementary Figure 6 | The human EMX1 locus with target sites. Schematic of the human *EMX1* locus showing the location of 15 target DNA sites, indicated by blue lines with corresponding PAM in magenta.

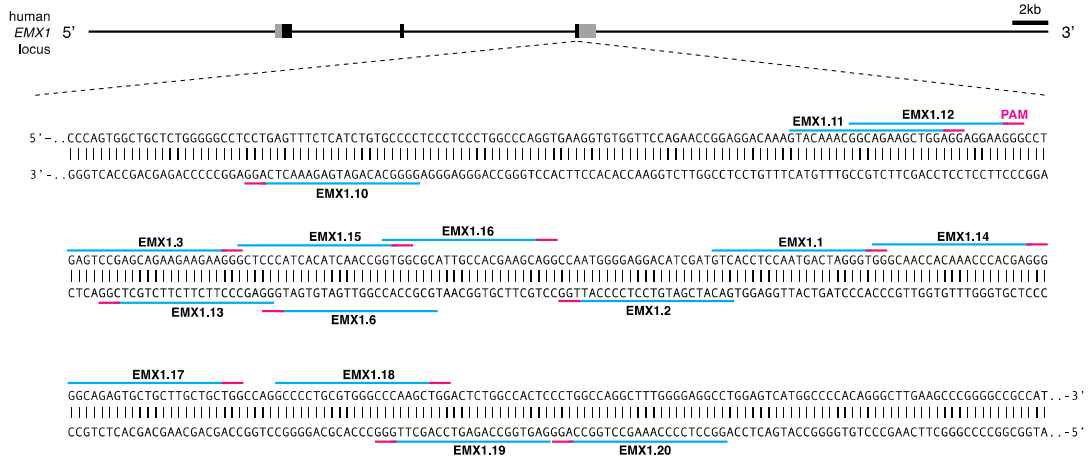
SUPPLEMENTARY FIGURE 7



Supplementary Figure 7 | Base frequency plots of relative SpCas9 cleavage efficiency for four *EMX1* target sites. Relative contribution of each base per guide sequence position to SpCas9 cleavage efficiency for *EMX1* targets 1, 2, 3, and 6. Modification efficiencies are normalized to cleavage levels mediated by the original guide sequence (Supplementary Table 5).

SUPPLEMENTARY FIGURE 8

a



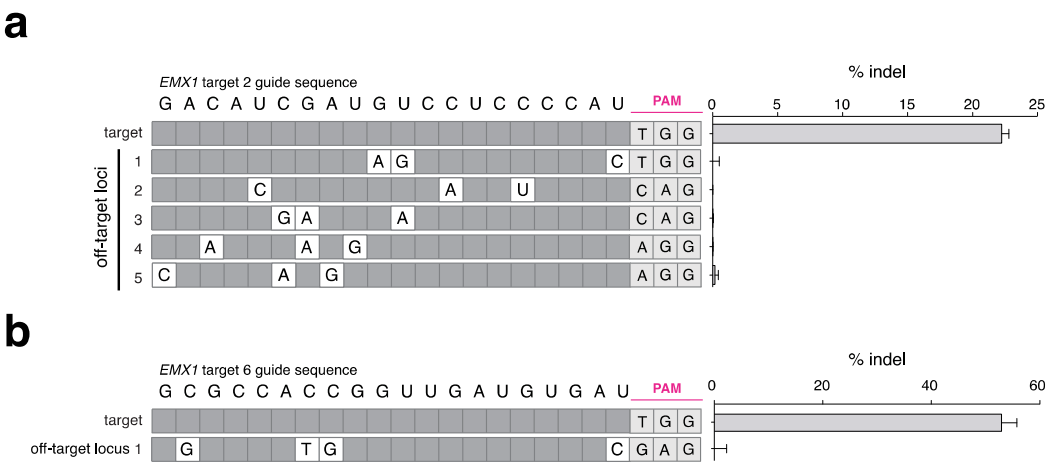
b

target species	gene	protospacer ID	target site (5' to 3')	PAM	strand
Homo sapiens	*EMX1*	1	GTCACCTCCAATGACTAGGG	TGG	+
	EMX1	2	GACATCGATGTCCTCCCCAT	TGG	–
	EMX1	3	GAGTCCGAGCAGAAGAAGAA	GGG	+
	EMX1	6	GCGCCACCGGTTGATGTGAT	GGG	–
	EMX1	10	GGGGCACAGATGAGAAACTC	AGG	–
	EMX1	11	GTACAAACGGCAGAAGCTGG	AGG	+
	EMX1	12	GGCAGAAGCTGGAGGAGGAA	GGG	+
	EMX1	13	GGAGCCCTTCTTCTTCTGCT	CGG	–
	EMX1	14	GGGCAACCACAAACCCACGA	GGG	+
	EMX1	15	GCTCCCATCACATCAACCGG	TGG	+
	EMX1	16	GTGGCGCATTGCCACGAAGC	AGG	+
	EMX1	17	GGCAGAGTGCTGCTTGCTGC	TGG	+
	EMX1	18	GCCCCTGCGTGGGCCCAAGC	TGG	+
	EMX1	19	GAGTGGCCAGAGTCCAGCTT	GGG	–
	EMX1	20	GGCCTCCCCAAAGCCTGGCC	AGG	–

Supplementary Figure 8 | Target amplicon sequencing for assessing CRISPR-Cas modification efficiency. a, Schematic of the human *EMX1* locus and a target site of interest. Target amplicons are PCR-amplified by a fusion PCR method to add sequencing adapters for the deep sequencing. Each sample is uniquely barcoded for multiplexed sequencing and pooled in an equimolar ratio into a sequencing library. **b,** target sequencing amplicons with associated target sites. Experimental samples are tracked by their barcode and amplicon identities.

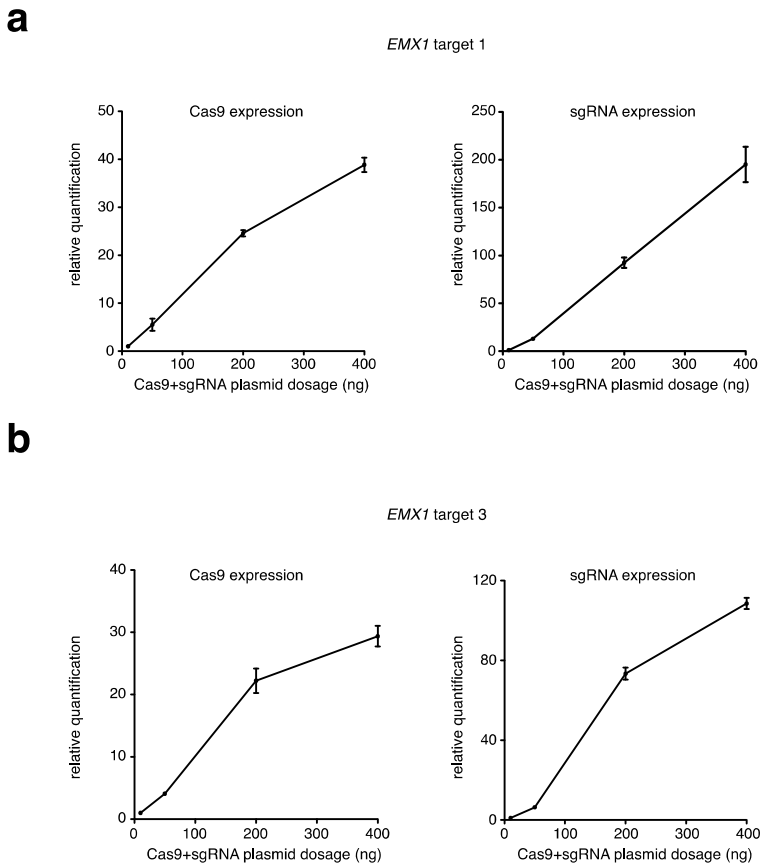
Nature Biotechnology: doi:10.1038/nbt.2647

SUPPLEMENTARY FIGURE 9



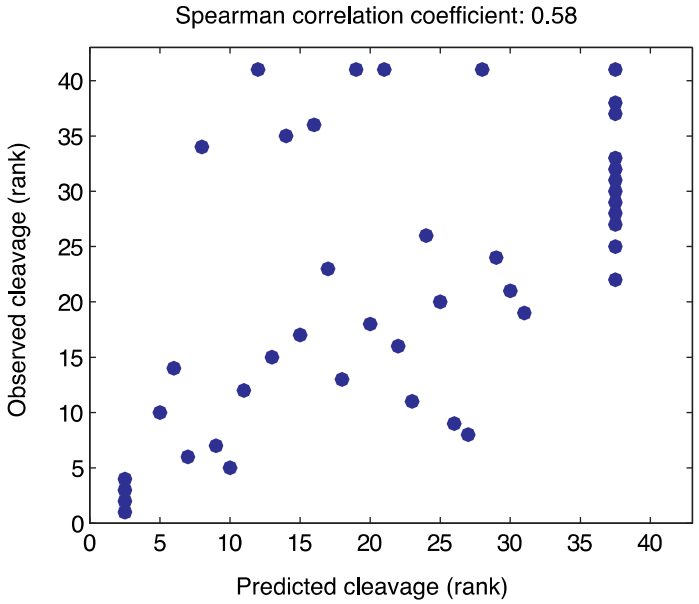
Supplementary Figure 9 | Additional genomic off-target site analysis. Cleavage levels at candidate genomic off-target loci (white cells indicating mismatches) for **a,** *EMX1* target 2 and **b,** *EMX1* target 6 were analyzed by deep sequencing. All indel frequencies are absolute and analyzed by deep sequencing from 2 biological replicates. Error bars indicate Wilson confidence intervals (Supplementary Methods).

SUPPLEMENTARY FIGURE 10



Supplementary Figure 10 | qRT-PCR analysis of Cas9 and sgRNA expression. Relative mRNA expression levels of hSpCas9 and sgRNA(+85) at different DNA transfection dosages for **a,** *EMX1* target 1 and **b,** *EMX1* target 3. The DNA transfected contains both hSpCas9 as well as the corresponding sgRNA on a single plasmid (guide sequence cloned into pX330). Data are plotted as the mean of independent biological replicates ($n = 3$) and relative quantification over the lowest SpCas9 dosage.

SUPPLEMENTARY FIGURE 11



Supplementary Figure 11 | Predicted and observed cutting frequency-ranks among genome-wide targets. Predicted cutting frequencies for genome-wide targets were calculated by multiplying, in series: $f_{est} = f(1)g(N_1, N_1') \times f(2)g(N_2, N_2') \times \cdots f(19)g(N_{19}, N_{19}') \times h$ with values $f(i)$ and $g(N_i, N_i')$ at position i corresponding, respectively, to the aggregate position- and base-mismatch cutting frequencies for positions and pairings indicated in Fig. 2c. Each frequency was normalized to range from 0 to 1, such that $f \rightarrow (f - f_{\min})/(f_{\max} - f_{\min})$. In case of a match, both were set equal to 1. The value h meanwhile re-weighted the estimated frequency by the minimum pairwise distance between consecutive mismatches in the target sequence. This distance value, in base-pairs, was divided by 18 to give a maximum value of 1 (in cases where fewer than 2 mismatches existed, or where mismatches occurred on opposite ends of the 19 bp target-window). Observed and predicted cutting-frequency rank (the highest frequency set to 1) for genome-wide samples having a read-count of at least 10,000 (*n=43*) were plotted. Those tied in rank were given a rank-average. The Spearman correlation coefficient, 0.58, indicated that the estimated frequencies recapitulated 58% of the rank-variance for the observed cutting frequencies. Comparing f_{est} with the cutting frequencies directly yielded a Pearson correlation of 0.89. While dominated by the highest-frequency gRNA/target pairs, this value indicated that nearly 90% of all cutting-frequency variance was explained by the predictions above.

Supplementary Table 1 | Target site sequences. Tested target sites for *S. pyogenes* type II CRISPR system with the requisite PAM. Cells were transfected with Cas9 and either crRNA-tracrRNA duplex or chimeric sgRNA for each target.

Target site ID	genomic target	Target site sequence (5' to 3')	PAM	strand
1	*EMX1*	GTCACCTCCAATGACTAGGG	TGG	+
2	*EMX1*	GACATCGATGTCCTCCCCAT	TGG	-
3	*EMX1*	GAGTCCGAGCAGAAGAAGAA	GGG	+
6	*EMX1*	GCGCCACCGGTTGATGTGAT	GGG	-
10	*EMX1*	GGGGCACAGATGAGAAACTC	AGG	-
11	*EMX1*	GTACAAACGGCAGAAGCTGG	AGG	+
12	*EMX1*	GGCAGAAGCTGGAGGAGGAA	GGG	+
13	*EMX1*	GGAGCCCTTCTTCTTCTGCT	CGG	-
14	*EMX1*	GGGCAACCACAAACCCACGA	GGG	+
15	*EMX1*	GCTCCCATCACATCAACCGG	TGG	+
16	*EMX1*	GTGGCGCATTGCCACGAAGC	AGG	+
17	*EMX1*	GGCAGAGTGCTGCTTGCTGC	TGG	+
18	*EMX1*	GCCCCTGCGTGGGCCCAAGC	TGG	+
19	*EMX1*	GAGTGGCCAGAGTCCAGCTT	GGG	-
20	*EMX1*	GGCCTCCCCAAAGCCTGGCC	AGG	-
4	*PVALB*	GGGGCCGAGATTGGGTGTTC	AGG	+
5	*PVALB*	GTGGCGAGAGGGGCCGAGAT	TGG	+
1	*SERPINB5*	GAGTGCCGCCGAGGCGGGGC	GGG	+
2	*SERPINB5*	GGAGTGCCGCCGAGGCGGGG	CGG	+
3	*SERPINB5*	GGAGAGGAGTGCCGCCGAGG	CGG	+

Supplementary Table 2 | Primer sequences

SURVEYOR assay

primer name	genomic target	primer sequence (5' to 3')
Sp-EMX1-F1	*EMX1*	AAAACCACCCTTCTCTCTGGC
Sp-EMX1-R1	*EMX1*	GGAGATTGGAGACACGGAGAG
Sp-EMX1-F2	*EMX1*	CCATCCCCTTCTGTGAATGT
Sp-EMX1-R2	*EMX1*	GGAGATTGGAGACACGGAGA
Sp-PVALB-F	*PVALB*	CTGGAAAGCCAATGCCTGAC
Sp-PVALB-R	*PVALB*	GGCAGCAAACTCCTTGTCCT

qRT-PCR for Cas9 and sgRNA expression

primer name	primer sequence (5' to 3')
sgRNA reverse-strand synthesis	AAGCACCGACTCGGTGCCAC
EMX1.1 sgRNA qPCR F	TCACCTCCAATGACTAGGGG
EMX1.1 sgRNA qPCR R	CAAGTTGATAACGGACTAGCCT
EMX1.3 sgRNA qPCR F	AGTCCGAGCAGAAGAAGAAGTTT
EMX1.3 sgRNA qPCR R	TTTCAAGTTGATAACGGACTAGCCT
Cas9 qPCR F	AAACAGCAGATTCGCCTGGA
Cas9 qPCR R	TCATCCGCTCGATGAAGCTC
GAPDH qPCR F	TCCAAAATCAAGTGGGGCGA
GAPDH qPCR R	TGATGACCCTTTTGGCTCCC

Bisulfite PCR and sequencing

primer name	primer sequence (5' to 3')
Bisulfite PCR F (*SERPINB5* locus)	GAGGAATTCTTTTTTTGTTYGAATATGTTGGAGGTTTTTTGGAAG
Bisulfite PCR R (*SERPINB5* locus)	GAGAAGCTTAAATAAAAAACRACAATACTCAACCCAACAACC
pUC19 sequencing	CAGGAAACAGCTATGAC

Supplementary Table 3 | Sequences for primers to test sgRNA architecture. Primers hybridize to the reverse strand of the U6 promoter unless otherwise indicated. The U6 priming site is in green, the guide sequence is in blue, the direct repeat sequence is highlighted in grey, and the tracrRNA sequence is in red. The secondary structure of each sgRNA architecture is shown in **Supplementary Fig. 3**.

primer name	primer sequence (5' to 3')
U6-Forward	GCCTCTAGAGGTACCTGAGGGCCTATTTCCCATGATTCC
I: sgRNA(DR +12, tracrRNA +85)	ACCTCTAGAAAAAAAGCACCGACTCGGTGCCACTTTTTCAAGTTGATAACGGA CTAGCCTTATTTTAACTTGCTATTTCTAGCTCTAAAACNNNNNNNNNNNNNNN NNNNNGGTGTTTCGTCCTTTCCACAAG
II: sgRNA(DR +12, tracrRNA +85) mut2	ACCTCTAGAAAAAAAGCACCGACTCGGTGCCACTTTTTCAAGTTGATAACGGA CTAGCCTTATATTAACTTGCTATTTCTAGCTCTAATACNNNNNNNNNNNNNNN NNNNNGGTGTTTCGTCCTTTCCACAAG
III: sgRNA(DR +22, tracrRNA +85)	ACCTCTAGAAAAAAAGCACCGACTCGGTGCCACTTTTTCAAGTTGATAACGGA CTAGCCTTATTTTAACTTGCTATGCTGTTTTGTTTCCAAAACAGCATAGCTCT AAAACNNNNNNNNNNNNNNNNNNNGGTGTTTCGTCCTTTCCACAAG
IV: sgRNA(DR +22, tracrRNA +85) mut4	ACCTCTAGAAAAAAAGCACCGACTCGGTGCCACTTTTTCAAGTTGATAACGGA CTAGCCTTATATTAACTTGCTATGCTGTATTGTTTCCAATACAGCATAGCTCT AATACNNNNNNNNNNNNNNNNNNNGGTGTTTCGTCCTTTCCACAAG

Supplementary Table 4 | Target sites with alternate PAMs for testing PAM specificity of Cas9. All target sites for PAM specificity testing are found within the human *EMX1* locus.

Target site sequence (5' to 3')	PAM
AGGCCCCAGTGGCTGCTCT	NAA
ACATCAACCGGTGGCGCAT	NAT
AAGGTGTGGTTCCAGAACC	NAC
CCATCACATCAACCGGTGG	NAG
AAACGGCAGAAGCTGGAGG	NTA
GGCAGAAGCTGGAGGAGGA	NTT
GGTGTGGTTCCAGAACCGG	NTC
AACCGGAGGACAAAGTACA	NTG
TTCCAGAACCGGAGGACAA	NCA
GTGTGGTTCCAGAACCGGA	NCT
TCCAGAACCGGAGGACAAA	NCC
CAGAAGCTGGAGGAGGAAG	NCG
CATCAACCGGTGGCGCATT	NGA
GCAGAAGCTGGAGGAGGAA	NGT
CCTCCCTCCCTGGCCCAGG	NGC
TCATCTGTGCCCCTCCCTC	NAA
GGGAGGACATCGATGTCAC	NAT
CAAACGGCAGAAGCTGGAG	NAC
GGGTGGGCAACCACAAACC	NAG
GGTGGGCAACCACAAACCC	NTA
GGCTCCCATCACATCAACC	NTT
GAAGGGCCTGAGTCCGAGC	NTC
CAACCGGTGGCGCATTGCC	NTG
AGGAGGAAGGGCCTGAGTC	NCA
AGCTGGAGGAGGAAGGGCC	NCT
GCATTGCCACGAAGCAGGC	NCC
ATTGCCACGAAGCAGGCCA	NCG
AGAACCGGAGGACAAAGTA	NGA
TCAACCGGTGGCGCATTGC	NGT
GAAGCTGGAGGAGGAAGGG	NGC

Supplementary Table 5: All sequencing data for Figure 2. Please see attached Excel file.

Supplementary Table 6: All sequencing data for Figure 3. Please see attached Excel file.

Supplementary Table 7: All sequencing data for Figure 4. Please see attached Excel file.

Supplementary Table 8: All sequencing data for expanded set of candidate genomic off-target loci for EMX1 targets 1, 2, 3, and 6. Please see attached Excel file.

Nature Biotechnology: doi:10.1038/nbt.2647

SUPPLEMENTARY SEQUENCES

All sequences are in the 5' to 3' direction. For U6 transcription, the string of underlined Ts serve as the transcriptional terminator.

> U6-short tracrRNA (*Streptococcus pyogenes* SF370)
gagggcctatttcccatgattccttcatatttgcatatacgatacaaggctgttagagagataattggaa
ttaatttgactgtaaacacaaagatattagtacaaaatacgtgacgtagaaagtaataatttcttgggta
gtttgcagttttaaaattatgttttaaaatggactatcatatgcttaccgtaacttgaaagtatttcgat
ttcttggctttatatatcttgtggaaaggacgaaacaccGGAACCATTCAAAACAGCATAGCAAGTTAAA
ATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCTTTTTTT
(tracrRNA sequence is highlighted in red)

>U6-DR-guide sequence-DR (*Streptococcus pyogenes* SF370)
gagggcctatttcccatgattccttcatatttgcatatacgatacaaggctgttagagagataattggaa
ttaatttgactgtaaacacaaagatattagtacaaaatacgtgacgtagaaagtaataatttcttgggta
gtttgcagttttaaaattatgttttaaaatggactatcatatgcttaccgtaacttgaaagtatttcgat
ttcttggctttatatatcttgtggaaaggacgaaacaccgggttttagagctatgctgttttgaatggtc
ccaaaacNNNNNNNNNNNNNNNNNNNNNNNNNNNNgttttagagctatgctgttttgaatggtcccaa
aacTTTTTTT
(direct repeat sequence is highlighted in gray and the guide sequence is highlighted in blue)

> sgRNA containing +48 tracrRNA (*Streptococcus pyogenes* SF370)
gagggcctatttcccatgattccttcatatttgcatatacgatacaaggctgttagagagataattggaa
ttaatttgactgtaaacacaaagatattagtacaaaatacgtgacgtagaaagtaataatttcttgggta
gtttgcagttttaaaattatgttttaaaatggactatcatatgcttaccgtaacttgaaagtatttcgat
ttcttggctttatatatcttgtggaaaggacgaaacaccNNNNNNNNNNNNNNNNNNNgttttagagct
agaaatagcaagttaaaataaggctagtccgTTTTTTT
(guide sequence is highlighted in blue and the tracrRNA fragment is highlighted in red)

> sgRNA containing +54 tracrRNA (*Streptococcus pyogenes* SF370)
gagggcctatttcccatgattccttcatatttgcatatacgatacaaggctgttagagagataattggaa
ttaatttgactgtaaacacaaagatattagtacaaaatacgtgacgtagaaagtaataatttcttgggta
gtttgcagttttaaaattatgttttaaaatggactatcatatgcttaccgtaacttgaaagtatttcgat
ttcttggctttatatatcttgtggaaaggacgaaacaccNNNNNNNNNNNNNNNNNNNgttttagagct
agaaatagcaagttaaaataaggctagtccgttatcaTTTTTTTT
(guide sequence is highlighted in blue and the tracrRNA fragment is highlighted in red)

> sgRNA containing +67 tracrRNA (*Streptococcus pyogenes* SF370)
gagggcctatttcccatgattccttcatatttgcatatacgatacaaggctgttagagagataattggaa
ttaatttgactgtaaacacaaagatattagtacaaaatacgtgacgtagaaagtaataatttcttgggta
gtttgcagttttaaaattatgttttaaaatggactatcatatgcttaccgtaacttgaaagtatttcgat
ttcttggctttatatatcttgtggaaaggacgaaacaccNNNNNNNNNNNNNNNNNNNgttttagagct
agaaatagcaagttaaaataaggctagtccgttatcaacttgaaaaagtgTTTTTTT
(guide sequence is highlighted in blue and the tracrRNA fragment is highlighted in red)

> sgRNA containing +85 tracrRNA (*Streptococcus pyogenes* SF370)

gagggcctatttcccatgattccttcatatttgcatatacgatacaaggctgttagagagataattggaa
ttaatttgactgtaaacacaaagatattagtacaaaatacgtgacgtagaaagtaataatttcttgggta
gtttgcagttttaaaattatgttttaaaatggactatcatatgcttaccgtaacttgaaagtatttcgat
ttcttggctttatatatcttgtggaaaggacgaaacaccNNNNNNNNNNNNNNNNNNNgttttagagct
agaaatagcaagttaaaataaggctagtccgttatcaacttgaaaaagtggcaccgagtcggtgcTTTTT
TT

(guide sequence is highlighted in blue and the tracrRNA fragment is highlighted in red)

> CBh-NLS-SpCas9-NLS

CGTTACATAACTTACGGTAAATGGCCCGCCTGGCTGACCGCCCAACGACCCCCGCCCATTGACGTCAATA
ATGACGTATGTTCCCATAGTAACGCCAATAGGGACTTTCCATTGACGTCAATGGGTGGAGTATTTACGGT
AAACTGCCCACTTGGCAGTACATCAAGTGTATCATATGCCAAGTACGCCCCCTATTGACGTCAATGACGG
TAAATGGCCCGCCTGGCATTATGCCCAGTACATGACCTTATGGGACTTTCCTACTTGGCAGTACATCTAC
GTATTAGTCATCGCTATTACCATGGTCGAGGTGAGCCCCACGTTCTGCTTCACTCTCCCCATCTCCCCCC
CCTCCCCACCCCCAATTTTGTATTTATTTATTTTTTAATTATTTTGTGCAGCGATGGGGGCGGGGGGGGG
GGGGGGGCGCGCGCCAGGCGGGGCGGGGCGGGGCGAGGGGCGGGGCGGGGCGAGGCGGAGAGGTGCGGCG
GCAGCCAATCAGAGCGGCGCGCTCCGAAAGTTTCCTTTTATGGCGAGGCGGCGGCGGCGGCCCTATA
AAAAGCGAAGCGCGCGGCGGGCGGGAGTCGCTGCGACGCTGCCTTCGCCCCGTGCCCCGCTCCGCCGCCG
CCTCGCGCCGCCCGCCCCGGCTCTGACTGACCGCGTTACTCCCACAGGTGAGCGGGCGGGACGGCCCTTC
TCCTCCGGGCTGTAATTAGCTGAGCAAGAGGTAAGGGTTTAAGGGATGGTTGGTTGGTGGGGTATTAATG
TTTAATTACCTGGAGCACCTGCCTGAAATCACTTTTTTTCAGGTTGGaccggtgccaccATGGACTATAA
GGACCACGACGGAGACTACAAGGATCATGATATTGATTACAAAGACGATGACGATAAGATGGCCCCAAAG
AAGAAGCGGAAGGTCGGTATCCACGGAGTCCCAGCAGCCGACAAGAAGTACAGCATCGGCCTGGACATCG
GCACCAACTCTGTGGGCTGGGCCGTGATCACCGACGAGTACAAGGTGCCCAGCAAGAAATTCAAGGTGCT
GGGCAACACCGACCGGCACAGCATCAAGAAGAACCTGATCGGAGCCCTGCTGTTCGACAGCGGCGAAACA
GCCGAGGCCACCCGGCTGAAGAGAACCGCCAGAAGAAGATACACCAGACGGAAGAACCGGATCTGCTATC
TGCAAGAGATCTTCAGCAACGAGATGGCCAAGGTGGACGACAGCTTCTTCCACAGACTGGAAGAGTCCTT
CCTGGTGGAAGAGGATAAGAAGCACGAGCGGCACCCCATCTTCGGCAACATCGTGGACGAGGTGGCCTAC
CACGAGAAGTACCCCACCATCTACCACCTGAGAAAGAAACTGGTGGACAGCACCGACAAGGCCGACCTGC
GGCTGATCTATCTGGCCCTGGCCCACATGATCAAGTTCCGGGGCCACTTCCTGATCGAGGGCGACCTGAA
CCCCGACAACAGCGACGTGGACAAGCTGTTCATCCAGCTGGTGCAGACCTACAACCAGCTGTTCGAGGAA
AACCCCATCAACGCCAGCGGCGTGGACGCCAAGGCCATCCTGTCTGCCAGACTGAGCAAGAGCAGACGGC
TGGAAAATCTGATCGCCCAGCTGCCCGGCGAGAAGAAGAATGGCCTGTTCGGCAACCTGATTGCCCTGAG
CCTGGGCCTGACCCCCAACTTCAAGAGCAACTTCGACCTGGCCGAGGATGCCAAACTGCAGCTGAGCAAG
GACACCTACGACGACGACCTGGACAACCTGCTGGCCCAGATCGGCGACCAGTACGCCGACCTGTTTCTGG
CCGCCAAGAACCTGTCCGACGCCATCCTGCTGAGCGACATCCTGAGAGTGAACACCGAGATCACCAAGGC
CCCCCTGAGCGCCTCTATGATCAAGAGATACGACGAGCACCACCAGGACCTGACCCTGCTGAAAGCTCTC
GTGCGGCAGCAGCTGCCTGAGAAGTACAAAGAGATTTTCTTCGACCAGAGCAAGAACGGCTACGCCGGCT
ACATTGACGGCGGAGCCAGCCAGGAAGAGTTCTACAAGTTCATCAAGCCCATCCTGGAAAAGATGGACGG
CACCGAGGAACTGCTCGTGAAGCTGAACAGAGAGGACCTGCTGCGGAAGCAGCGGACCTTCGACAACGGC
AGCATCCCCCACCAGATCCACCTGGGAGAGCTGCACGCCATTCTGCGGCGGCAGGAAGATTTTTACCCAT
TCCTGAAGGACAACCGGGAAAAGATCGAGAAGATCCTGACCTTCCGCATCCCCTACTACGTGGGCCCTCT
GGCCAGGGGAAACAGCAGATTCGCCTGGATGACCAGAAAGAGCGAGGAAACCATCACCCCCTGGAACTTC
GAGGAAGTGGTGGACAAGGGCGCTTCCGCCCAGAGCTTCATCGAGCGGATGACCAACTTCGATAAGAACC
TGCCCAACGAGAAGGTGCTGCCCAAGCACAGCCTGCTGTACGAGTACTTCACCGTGTATAACGAGCTGAC

CAAAGTGAAATACGTGACCGAGGGAATGAGAAAGCCCGCCTTCCTGAGCGGCGAGCAGAAAAAGGCCATC
GTGGACCTGCTGTTCAAGACCAACCGGAAAGTGACCGTGAAGCAGCTGAAAGAGGACTACTTCAAGAAAA
TCGAGTGCTTCGACTCCGTGGAAATCTCCGGCGTGGAAGATCGGTTCAACGCCTCCCTGGGCACATACCA
CGATCTGCTGAAAATTATCAAGGACAAGGACTTCCTGGACAATGAGGAAAACGAGGACATTCTGGAAGAT
ATCGTGCTGACCCTGACACTGTTTGAGGACAGAGAGATGATCGAGGAACGGCTGAAAACCTATGCCCACC
TGTTCGACGACAAAGTGATGAAGCAGCTGAAGCGGCGGAGATACACCGGCTGGGGCAGGCTGAGCCGGAA
GCTGATCAACGGCATCCGGGACAAGCAGTCCGGCAAGACAATCCTGGATTTCCTGAAGTCCGACGGCTTC
GCCAACAGAAACTTCATGCAGCTGATCCACGACGACAGCCTGACCTTTAAAGAGGACATCCAGAAAGCCC
AGGTGTCCGGCCAGGGCGATAGCCTGCACGAGCACATTGCCAATCTGGCCGGCAGCCCCGCCATTAAGAA
GGGCATCCTGCAGACAGTGAAGGTGGTGGACGAGCTCGTGAAAGTGATGGGCCGGCACAAGCCCGAGAAC
ATCGTGATCGAAATGGCCAGAGAGAACCAGACCACCCAGAAGGGACAGAAGAACAGCCGCGAGAGAATGA
AGCGGATCGAAGAGGGCATCAAAGAGCTGGGCAGCCAGATCCTGAAAGAACACCCCGTGGAAAACACCCA
GCTGCAGAACGAGAAGCTGTACCTGTACTACCTGCAGAATGGGCGGGATATGTACGTGGACCAGGAACTG
GACATCAACCGGCTGTCCGACTACGATGTGGACCATATCGTGCCTCAGAGCTTTCTGAAGGACGACTCCA
TCGACAACAAGGTGCTGACCAGAAGCGACAAGAACCGGGGCAAGAGCGACAACGTGCCCTCCGAAGAGGT
CGTGAAGAAGATGAAGAACTACTGGCGGCAGCTGCTGAACGCCAAGCTGATTACCCAGAGAAAGTTCGAC
AATCTGACCAAGGCCGAGAGAGGCGGCCTGAGCGAACTGGATAAGGCCGGCTTCATCAAGAGACAGCTGG
TGGAAACCCGGCAGATCACAAAGCACGTGGCACAGATCCTGGACTCCCGGATGAACACTAAGTACGACGA
GAATGACAAGCTGATCCGGGAAGTGAAAGTGATCACCCTGAAGTCCAAGCTGGTGTCCGATTTCCGGAAG
GATTTCCAGTTTTACAAAGTGCGCGAGATCAACAACTACCACCACGCCCACGACGCCTACCTGAACGCCG
TCGTGGGAACCGCCCTGATCAAAAAGTACCCTAAGCTGGAAAGCGAGTTCGTGTACGGCGACTACAAGGT
GTACGACGTGCGGAAGATGATCGCCAAGAGCGAGCAGGAAATCGGCAAGGCTACCGCCAAGTACTTCTTC
TACAGCAACATCATGAACTTTTTCAAGACCGAGATTACCCTGGCCAACGGCGAGATCCGGAAGCGGCCTC
TGATCGAGACAAACGGCGAAACCGGGGAGATCGTGTGGGATAAGGGCCGGGATTTTGCCACCGTGCGGAA
AGTGCTGAGCATGCCCCAAGTGAATATCGTGAAAAAGACCGAGGTGCAGACAGGCGGCTTCAGCAAAGAG
TCTATCCTGCCCAAGAGGAACAGCGATAAGCTGATCGCCAGAAAGAAGGACTGGGACCCTAAGAAGTACG
GCGGCTTCGACAGCCCCACCGTGGCCTATTCTGTGCTGGTGGTGGCCAAAGTGGAAAAGGGCAAGTCCAA
GAAACTGAAGAGTGTGAAAGAGCTGCTGGGGATCACCATCATGGAAAGAAGCAGCTTCGAGAAGAATCCC
ATCGACTTTCTGGAAGCCAAGGGCTACAAAGAAGTGAAAAAGGACCTGATCATCAAGCTGCCTAAGTACT
CCCTGTTCGAGCTGGAAAACGGCCGGAAGAGAATGCTGGCCTCTGCCGGCGAACTGCAGAAGGGAAACGA
ACTGGCCCTGCCCTCCAAATATGTGAACTTCCTGTACCTGGCCAGCCACTATGAGAAGCTGAAGGGCTCC
CCCGAGGATAATGAGCAGAAACAGCTGTTTGTGGAACAGCACAAGCACTACCTGGACGAGATCATCGAGC
AGATCAGCGAGTTCTCCAAGAGAGTGATCCTGGCCGACGCTAATCTGGACAAAGTGCTGTCCGCCTACAA
CAAGCACCGGGATAAGCCCATCAGAGAGCAGGCCGAGAATATCATCCACCTGTTTACCCTGACCAATCTG
GGAGCCCCTGCCGCCTTCAAGTACTTTGACACCACCATCGACCGGAAGAGGTACACCAGCACCAAAGAGG
TGCTGGACGCCACCCTGATCCACCAGAGCATCACCGGCCTGTACGAGACACGGATCGACCTGTCTCAGCT
GGGAGGCGACTTTCTTTTTCTTAGCTTGACCAGCTTTCTTAGTAGCAGCAGGACGCTTTAA
(NLS-hSpCas9-NLS is highlighted in blue)

> Sequencing amplicon for EMX1 guides 1.1, 1.14, 1.17
CCAATGGGGAGGACATCGATGTCACCTCCAATGACTAGGGTGGGCAACCACAAACCCACGAGGGCAGAGT
GCTGCTTGCTGCTGGCCAGGCCCCTGCGTGGGCCCAAGCTGGACTCTGGCCAC

> Sequencing amplicon for EMX1 guides 1.2, 1.16
CGAGCAGAAGAAGAAGGGCTCCCATCACATCAACCGGTGGCGCATTGCCACGAAGCAGGCCAATGGGGAG
GACATCGATGTCACCTCCAATGACTAGGGTGGGCAACCACAAACCCACGAG

> Sequencing amplicon for EMX1 guides 1.3, 1.13, 1.15
GGAGGACAAAGTACAAACGGCAGAAGCTGGAGGAGGAAGGGCCTGAGTCCGAGCAGAAGAAGAAGGGCTC
CCATCACATCAACCGGTGGCGCATTGCCACGAAGCAGGCCAATGGGGAGGACATCGAT

> Sequencing amplicon for EMX1 guides 1.6
AGAAGCTGGAGGAGGAAGGGCCTGAGTCCGAGCAGAAGAAGAAGGGCTCCCATCACATCAACCGGTGGCG
CATTGCCACGAAGCAGGCCAATGGGGAGGACATCGATGTCACCTCCAATGACTAGGGTGG

> Sequencing amplicon for EMX1 guides 1.10
CCTCAGTCTTCCCATCAGGCTCTCAGCTCAGCCTGAGTGTTGAGGCCCCAGTGGCTGCTCTGGGGGCCTC
CTGAGTTTCTCATCTGTGCCCCTCCCTCCCTGGCCCAGGTGAAGGTGTGGTTCCA

> Sequencing amplicon for EMX1 guides 1.11, 1.12
TCATCTGTGCCCCTCCCTCCCTGGCCCAGGTGAAGGTGTGGTTCCAGAACCGGAGGACAAAGTACAAACG
GCAGAAGCTGGAGGAGGAAGGGCCTGAGTCCGAGCAGAAGAAGAAGGGCTCCCATCACA

> Sequencing amplicon for EMX1 guides 1.18, 1.19
CTCCAATGACTAGGGTGGGCAACCACAAACCCACGAGGGCAGAGTGCTGCTTGCTGCTGGCCAGGCCCCT
GCGTGGGCCCAAGCTGGACTCTGGCCACTCCCTGGCCAGGCTTTGGGGAGGCCTGGAGT

> Sequencing amplicon for EMX1 guides 1.20
CTGCTTGCTGCTGGCCAGGCCCCTGCGTGGGCCCAAGCTGGACTCTGGCCACTCCCTGGCCAGGCTTTGG
GGAGGCCTGGAGTCATGGCCCCACAGGGCTTGAAGCCCGGGGCCGCCATTGACAGAG

>T7 promoter F primer for annealing with target strand
GAAATTAATACGACTCACTATAGGG

>oligo containing pUC19 target site 1 for methylation (T7 reverse)
AAAAAAGCACCGACTCGGTGCCACTTTTTCAAGTTGATAACGGACTAGCCTTATTTTAACTTGCTATTTC
TAGCTCTAAAACAACGACGAGCGTGACACCACCCTATAGTGAGTCGTATTAATTTC

>oligo containing pUC19 target site 2 for methylation (T7 reverse)
AAAAAAGCACCGACTCGGTGCCACTTTTTCAAGTTGATAACGGACTAGCCTTATTTTAACTTGCTATTTC
TAGCTCTAAAACGCAACAATTAATAGACTGGACCTATAGTGAGTCGTATTAATTTC

SUPPLEMENTARY REFERENCES

1. Cong, L. et al. Multiplex genome engineering using CRISPR/Cas systems. *Science* **339**, 819-823 (2013).
2. Wilson, E.B. Probable inference, the law of succession, and statistical inference. *J Am Stat Assoc* **22**, 209-212 (1927).
3. Deltcheva, E. et al. CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III. *Nature* **471**, 602-607 (2011).

Nature Biotechnology: doi:10.1038/nbt.2647