nature biotechnology

DNA targeting specificity of RNA-guided Cas9 nucleases

Patrick D Hsu[1–3,9], David A Scott[1,2,9], Joshua A Weinstein[1,2], F Ann Ran[1–3], Silvana Konermann[1,2], Vineeta Agarwala[1,4,5], Yinqing Li[1,2], Eli J Fine[6], Xuebing Wu[7], Ophir Shalem[1,2], Thomas J Cradick[6], Luciano A Marraffini[8], Gang Bao[6] & Feng Zhang[1,2]

The *Streptococcus pyogenes* Cas9 (SpCas9) nuclease can be efficiently targeted to genomic loci by means of single-guide RNAs (sgRNAs) to enable genome editing[1–10]. Here, we characterize SpCas9 targeting specificity in human cells to inform the selection of target sites and avoid off-target effects. Our study evaluates >700 guide RNA variants and SpCas9-induced indel mutation levels at >100 predicted genomic off-target loci in 293T and 293FT cells. We find that SpCas9 tolerates mismatches between guide RNA and target DNA at different positions in a sequence-dependent manner, sensitive to the number, position and distribution of mismatches. We also show that SpCas9-mediated cleavage is unaffected by DNA methylation and that the dosage of SpCas9 and sgRNA can be titrated to minimize off-target modification. To facilitate mammalian genome engineering applications, we provide a web-based software tool to guide the selection and validation of target sequences as well as off-target analyses.

The bacterial type II clustered, regularly interspaced, short palindromic repeats (CRISPR) system from *S. pyogenes* can be reconstituted in mammalian cells using three minimal components[1]: the CRISPR-associated nuclease Cas9 (SpCas9), a specificity-determining CRISPR RNA (crRNA), and an auxiliary trans-activating crRNA (tracrRNA)[11]. Following crRNA and tracrRNA hybridization, SpCas9 is targeted to genomic loci matching a 20-nt guide sequence within the crRNA, immediately upstream of a required 5′-NGG protospacer adjacent motif (PAM)[11]. crRNA and tracrRNA duplexes can also be fused to generate a chimeric sgRNA[12] that mimics the natural crRNA-tracrRNA hybrid. Both crRNA-tracrRNA duplexes and sgRNAs can be used to target SpCas9 for multiplexed genome editing in eukaryotic cells[1,3].

Although an sgRNA design consisting of a truncated crRNA and tracrRNA had been previously shown to mediate efficient cleavage *in vitro*[12], it failed to achieve detectable cleavage at several loci that were efficiently modified by crRNA-tracrRNA duplexes bearing identical guide sequences[1]. Because the major difference between this sgRNA design and the native crRNA-tracrRNA duplex is the length of the tracrRNA sequence, we tested whether extension of the tracrRNA tail would improve SpCas9 activity.

We generated a set of sgRNAs targeting multiple sites within the human *EMX1* and *PVALB* loci with different tracrRNA 3′ truncations (**Fig. 1a**). Using the SURVEYOR nuclease assay[13], we assessed the ability of each Cas9-sgRNA complex to generate indels in human embryonic kidney (HEK) 293FT cells through the induction of DNA double-stranded breaks (DSBs) and subsequent nonhomologous end joining (NHEJ) DNA damage repair (Online Methods). sgRNAs with +67 or +85 nucleotide (nt) tracrRNA tails mediated DNA cleavage at all target sites tested, with up to fivefold higher levels of indels than the corresponding crRNA-tracrRNA duplexes (**Fig. 1b** and **Supplementary Fig. 1a**). Furthermore, both sgRNA designs efficiently modified *PVALB* loci that were previously not targetable using crRNA-tracrRNA duplexes[1] (**Fig. 1b** and **Supplementary Fig. 1b**). For all five tested targets, we observed a consistent increase in modification efficiency with increasing tracrRNA length. We performed northern blot analyses for the guide RNA truncations and found increased levels of expression for the longer tracrRNA sequences, suggesting that improved target cleavage was at least partially due to higher sgRNA expression or stability (**Fig. 1c**). Taken together, these data indicate that the tracrRNA tail is important for optimal SpCas9 expression and activity *in vivo*.

We further investigated the sgRNA architecture by extending the duplex length from 12 to the 22 nt found in the native crRNA-tracrRNA duplex (**Supplementary Fig. 2a**). We also mutated the sequence encoding the sgRNAs to abolish any poly-T tracts that could serve as premature transcriptional terminators for U6-driven transcription[14]. We tested these new sgRNA scaffolds on three targets within the human *EMX1* gene (**Supplementary Fig. 2b**) and observed only modest changes in modification efficiency. Thus, we established sgRNA(+67) as a minimum effective SpCas9 guide RNA architecture and for all subsequent studies we used the most active sgRNA(+85) architecture.

[1]Broad Institute of MIT and Harvard, Cambridge, Massachusetts, USA. [2]McGovern Institute for Brain Research, Department of Brain and Cognitive Sciences, Department of Biological Engineering, Massachusetts Institute of Technology, Cambridge, Massachusetts, USA. [3]Department of Molecular and Cellular Biology, Harvard University, Cambridge, Massachusetts, USA. [4]Program in Biophysics, Harvard University, Cambridge, Massachusetts, USA. [5]Harvard-MIT Division of Health Sciences and Technology, MIT, Cambridge, MA 02139, USA. [6]Department of Biomedical Engineering, Georgia Institute of Technology and Emory University, Atlanta, Georgia, USA. [7]Computational and Systems Biology Graduate Program, Koch Institute for Integrative Cancer Research, Massachusetts Institute of Technology, Cambridge, Massachusetts, USA. [8]Laboratory of Bacteriology, The Rockefeller University, New York, New York, USA. [9]These authors contributed equally to this work. Correspondence should be addressed to F.Z. (zhang@broadinstitute.org).

Received 30 March; accepted 30 June; published online 21 July 2013; corrected online 28 August 2013; doi:10.1038/nbt.2647

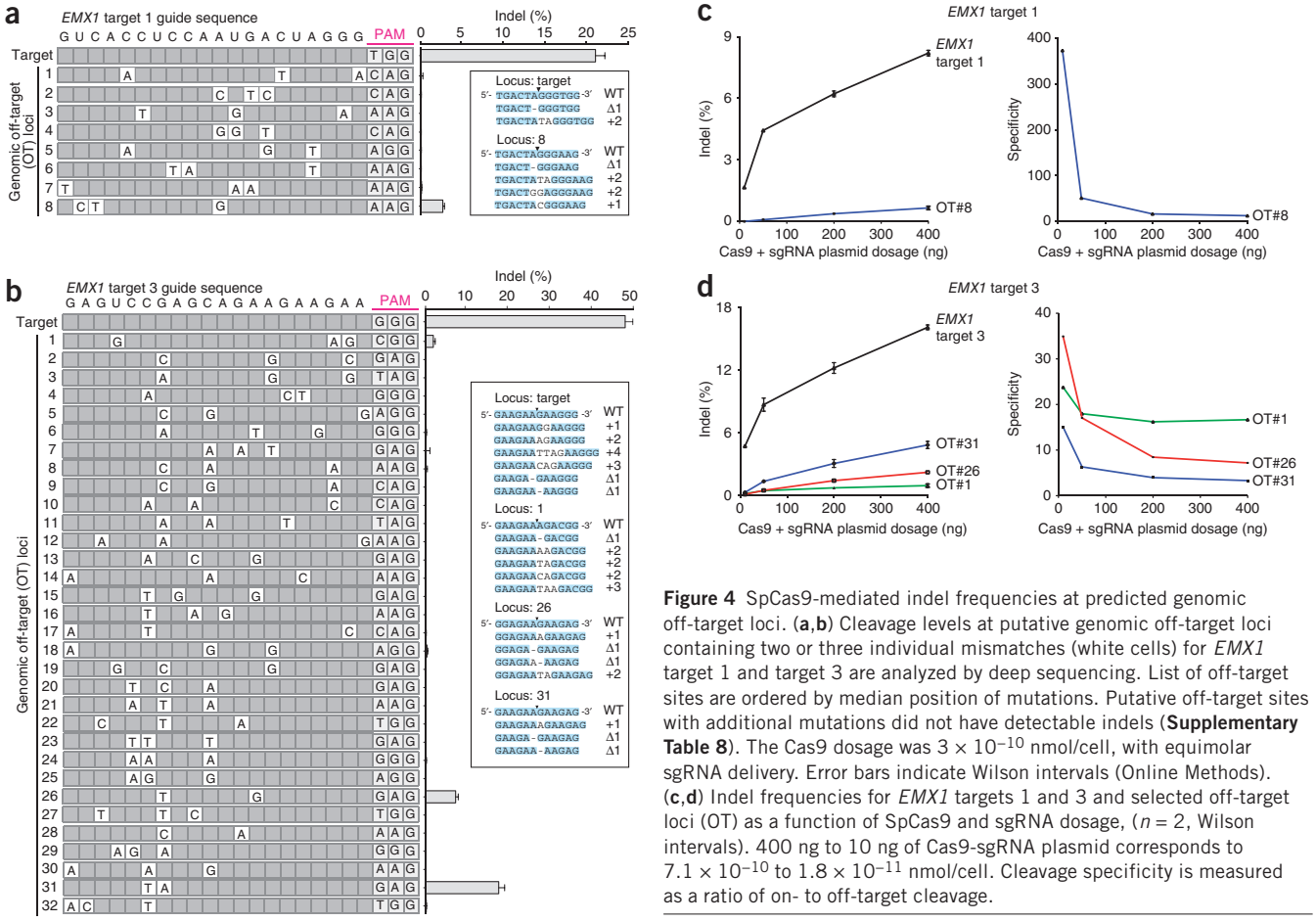

Figure 4 SpCas9-mediated indel frequencies at predicted genomic off-target loci. (**a**,**b**) Cleavage levels at putative genomic off-target loci containing two or three individual mismatches (white cells) for *EMX1* target 1 and target 3 are analyzed by deep sequencing. List of off-target sites are ordered by median position of mutations. Putative off-target sites with additional mutations did not have detectable indels (**Supplementary Table 8**). The Cas9 dosage was 3×10^{-10} nmol/cell, with equimolar sgRNA delivery. Error bars indicate Wilson intervals (Online Methods). (**c**,**d**) Indel frequencies for *EMX1* targets 1 and 3 and selected off-target loci (OT) as a function of SpCas9 and sgRNA dosage, ($n = 2$, Wilson intervals). 400 ng to 10 ng of Cas9-sgRNA plasmid corresponds to 7.1×10^{-10} to 1.8×10^{-11} nmol/cell. Cleavage specificity is measured as a ratio of on- to off-target cleavage.

observations from the single base-pair mismatch data (**Fig. 2c**). This effect is particularly salient in guide sequences bearing a small number of total mismatches, whether those are consecutive (**Fig. 3a**) or interspaced (**Fig. 3b**). Additionally, guide sequences with mismatches spaced four or more bases apart also mediated SpCas9 cleavage in some cases (**Fig. 3c**). Thus, together with the identity of mismatched base-pairing, we observed that many off-target cleavage effects can be explained by a combination of mismatch number and position.

Given these mismatched guide RNA results, we expected that for any particular sgRNA, SpCas9 may cleave genomic loci that contain small numbers of mismatched bases. For the four *EMX1* targets described above, we computationally selected 117 candidate off-target sites in the human genome that are followed by a 5′-NRG PAM and meet any of the following additional criteria: (i) up to five mismatches, (ii) short insertions or deletions or (iii) mismatches only in the PAM-distal region. Additionally, we assessed off-target loci of high sequence similarity without the PAM requirement. The majority of off-target sites tested for each sgRNA (30/31, 23/23, 48/51 and 12/12 sites for *EMX1* targets 1, 2, 3 and 6, respectively) exhibited modification efficiencies at least 2 magnitudes lower than that of corresponding on-targets (**Fig. 4a**,**b**, **Supplementary Fig. 9** and **Supplementary Tables 7** and **8**). Of the four off-target sites that exhibit substantial modification efficiencies, three contained only mismatches in the PAM-distal region, consistent with our multiple mismatch sgRNA observations (**Fig. 3**). Notably, these three loci were followed by 5′-NAG PAMs, demonstrating that off-target analyses of SpCas9 must include 5′-NAG as well as 5′-NGG candidate loci.

Enzymatic specificity and activity strength are often highly dependent on reaction conditions, which at high enzyme concentration might amplify off-target activity[28,29]. One potential strategy for minimizing nonspecific cleavage is to limit the enzyme concentration, namely the level of SpCas9-sgRNA complex. Cleavage specificity, measured as the ratio of on- to off-target cleavage, increased dramatically as we decreased the equimolar amounts of SpCas9 and sgRNA transfected into 293FT cells (**Fig. 4c**,**d**) from 7.1×10^{-10} to 1.8×10^{-11} nmol/cell (400 ng to 10 ng of Cas9-sgRNA plasmid). qRT-PCR assay confirmed that the level of hSpCas9 mRNA and sgRNA decreased proportionally to the amount of transfected DNA (**Supplementary Fig. 10**). Whereas specificity increased gradually by nearly fourfold as we decreased the transfected DNA amount from 7.1×10^{-10} to 9.0×10^{-11} nmol/cell (400 ng to 50 ng plasmid), we observed a notable additional sevenfold increase in specificity upon further decreasing transfected DNA from 9.0×10^{-11} to 1.8×10^{-11} nmol/cell (50 ng to 10 ng plasmid; **Fig. 4c**). These findings suggest that we can minimize the level of off-target activity by titrating the amount of SpCas9 and sgRNA DNA delivered. However, increasing specificity by reducing the amount of transfected DNA also leads to a reduction in on-target cleavage. These measurements enable quantitative integration of specificity and efficiency criteria into dosage choice to optimize SpCas9 activity for different applications. Additional work to explore modifications in SpCas9 and sgRNA design may improve SpCas9-intrinsic specificity without sacrificing cleavage efficiency.

The ability to program SpCas9 to target specific sites in the genome by simply designing a short guide RNA complementary to the desired target site holds enormous potential for applications throughout biology and medicine. Our results demonstrate that the specificity of SpCas9-mediated DNA cleavage is sequence- and locus-dependent and