

RESEARCH

Open Access



# NmeCas9 is an intrinsically high-fidelity genome-editing platform

Nadia Amrani<sup>1†</sup>, Xin D. Gao<sup>1†</sup>, Pengpeng Liu<sup>3,4</sup>, Alireza Edraki<sup>1</sup>, Aamir Mir<sup>1</sup>, Raed Ibraheim<sup>1</sup>, Ankit Gupta<sup>3,8</sup>, Kanae E. Sasaki<sup>1,9</sup>, Tong Wu<sup>3</sup>, Paul D. Donohoue<sup>6</sup>, Alexander H. Settle<sup>6,10</sup>, Alexandra M. Lied<sup>6</sup>, Kyle McGovern<sup>6,7</sup>, Chris K. Fuller<sup>6</sup>, Peter Cameron<sup>6</sup>, Thomas G. Fazio<sup>2,3</sup>, Lihua Julie Zhu<sup>2,3,5</sup>, Scot A. Wolfe<sup>3,4</sup> and Erik J. Sontheimer<sup>1,2\*</sup> 

## Abstract

**Background:** The development of CRISPR genome editing has transformed biomedical research. Most applications reported thus far rely upon the Cas9 protein from *Streptococcus pyogenes* SF370 (SpyCas9). With many RNA guides, wildtype SpyCas9 can induce significant levels of unintended mutations at near-cognate sites, necessitating substantial efforts toward the development of strategies to minimize off-target activity. Although the genome-editing potential of thousands of other Cas9 orthologs remains largely untapped, it is not known how many will require similarly extensive engineering to achieve single-site accuracy within large genomes. In addition to its off-targeting propensity, SpyCas9 is encoded by a relatively large open reading frame, limiting its utility in applications that require size-restricted delivery strategies such as adeno-associated virus vectors. In contrast, some genome-editing-validated Cas9 orthologs are considerably smaller and therefore better suited for viral delivery.

**Results:** Here we show that wildtype NmeCas9, when programmed with guide sequences of the natural length of 24 nucleotides, exhibits a nearly complete absence of unintended editing in human cells, even when targeting sites that are prone to off-target activity with wildtype SpyCas9. We also validate at least six variant protospacer adjacent motifs (PAMs), in addition to the preferred consensus PAM (5'-N<sub>4</sub>GATT-3'), for NmeCas9 genome editing in human cells.

**Conclusions:** Our results show that NmeCas9 is a naturally high-fidelity genome-editing enzyme and suggest that additional Cas9 orthologs may prove to exhibit similarly high accuracy, even without extensive engineering.

**Keywords:** Cas9, CRISPR, sgRNA, Protospacer adjacent motif, Off-target, *Neisseria meningitidis*

## Background

Over the past decade, clustered, regularly interspaced, short palindromic repeats (CRISPRs) have been revealed as genomic sources of small RNAs (CRISPR RNAs (crRNAs)) that specify genetic interference in many bacteria and most archaea [1–3]. CRISPR sequences include “spacers,” which often match sequences of previously encountered invasive nucleic acids such as phage genomes and plasmids. In conjunction with CRISPR-associated (Cas) proteins, crRNAs recognize target nucleic acids

(DNA, RNA, or both, depending on the system) by base pairing, leading to their destruction. The primary natural function of CRISPR-Cas systems is to provide adaptive immunity against phages [4, 5] and other mobile genetic elements [6]. CRISPR-Cas systems are divided into two main classes: class 1, with large, multi-subunit effector complexes, and class 2, with single-protein-subunit effectors [7]. Both CRISPR-Cas classes include multiple types based primarily on the identity of a signature effector protein. Within class 2, the type II systems are the most abundant and the best characterized. The interference function of type II CRISPR-Cas systems requires the Cas9 protein, the crRNA, and a separate non-coding RNA known as the trans-activating crRNA (tracrRNA) [8–10]. Successful interference also requires that the DNA target (the “protospacer”) be highly complementary to the spacer portion of

\* Correspondence: [erik.sontheimer@umassmed.edu](mailto:erik.sontheimer@umassmed.edu)

<sup>†</sup>Nadia Amrani and Xin D. Gao contributed equally to this work.

<sup>1</sup>RNA Therapeutics Institute, University of Massachusetts Medical School, 368 Plantation Street, Worcester, MA 01605, USA

<sup>2</sup>Program in Molecular Medicine, University of Massachusetts Medical School, 368 Plantation Street, Worcester, MA 01605, USA

Full list of author information is available at the end of the article



the crRNA and that the PAM consensus be present at neighboring base pairs [11, 12].

Following the discovery that type II interference occurs via double-strand breaks (DSBs) in the DNA target [9], the Cas9 protein was shown to be the only Cas protein required for type II interference, to be manually reprogrammable via engineered CRISPR spacers, and to be functionally portable between species that diverged billions of years ago [10]. Biochemical analyses with purified Cas9 confirmed its role as a crRNA-guided, programmable nuclease that induces R-loop formation between the crRNA and one dsDNA strand, and that cleaves the crRNA-complementary and noncomplementary strands with its HNH and RuvC domains, respectively [13, 14]. In vitro cleavage reactions also showed that the tracrRNA is essential for DNA cleavage activity and that the naturally separate crRNA and tracrRNA could retain function when fused into a single-guide RNA (sgRNA) [14]. Several independent reports then showed that the established DSB-inducing activity of Cas9 could be elicited not only in vitro but also in living cells, both bacterial [15] and eukaryotic [16–20]. As with earlier DSB-inducing systems [21], cellular repair of Cas9-generated DSBs by either non-homologous end joining (NHEJ) or homology-directed repair (HDR) enabled live-cell targeted mutagenesis, and the CRISPR-Cas9 system has now been widely adopted as a facile genome-editing platform in a wide range of organisms [22–24]. In addition to genome editing, catalytically inactivated Cas9 (“dead” Cas9, dCas9) retains its sgRNA-guided DNA binding function, enabling fused or tethered functionalities to be delivered to precise genomic loci [25, 26]. Similar RNA-guided tools for genome manipulations have since been developed from type V CRISPR-Cas systems that use the Cas12a (formerly Cpf1) enzyme [27].

Type II CRISPR-Cas systems are currently grouped into three subtypes (II-A, II-B, and II-C) [7, 28]. The vast majority of Cas9 characterization has been done on a single type II-A ortholog, SpyCas9, in part due to its consistently high genome-editing activity. SpyCas9’s sgRNAs typically contain a 20-nt guide sequence (the spacer-derived sequence that base pairs to the DNA target [8, 14]). The PAM requirement for SpyCas9 is 5′-NGG-3′ (or, less favorably, 5′-NAG-3′), after the 3′ end of the protospacer’s crRNA-noncomplementary strand [8, 14]. Based on these and other parameters, many sgRNAs directed against potentially targetable sites in a large eukaryotic genome also have near-cognate sites available to it that lead to unintended, “off-target” editing. Indeed, off-target activity by SpyCas9 has been well-documented with many sgRNA-target combinations [29, 30], prompting the development of numerous approaches to limit editing activity at unwanted sites [31–36]. Although these strategies have been shown to minimize off-targeting to various degrees, they do not

always abolish it and they can also reduce on-target activity, at least with some sgRNAs. Furthermore, each of these approaches has required extensive testing, validation, and optimization and in some cases [33, 37, 38] depended heavily upon prior high-resolution structural characterization [39–42].

Thousands of other Cas9 orthologs have been documented [7, 28, 43, 44], providing tremendous untapped potential for additional genome-editing capabilities beyond those offered by SpyCas9. Many Cas9 orthologs will provide distinct PAM specificities, increasing the number of targetable sites in any given genome. Many pair-wise Cas9 combinations also have orthogonal guides that load into one ortholog but not the other, facilitating multiplexed applications [44–46]. Finally, some Cas9 orthologs (especially those from subtype II-C) are hundreds of amino acids smaller than the 1368 amino acid SpyCas9 [7, 43, 44] and are therefore more amenable to combined Cas9/sgRNA delivery via a single-size-restricted vector such as adeno-associated virus (AAV) [47, 48]. Finally, there may be Cas9 orthologs that exhibit additional advantages such as greater efficiency, natural hyper-accuracy, distinct activities, reduced immunogenicity, or novel means of control over editing. Deeper exploration of the Cas9 population could therefore enable expanded or improved genome engineering capabilities.

We have used *N. meningitidis* (strain 8013) as a model system for the interference functions and mechanisms of type II-C CRISPR-Cas systems [49–52]. In addition, we and others previously reported that the type II-C Cas9 ortholog from *N. meningitidis* (NmeCas9) can be applied as a genome engineering platform [46, 53, 54]. At 1082 amino acids, NmeCas9 is 286 residues smaller than SpyCas9, making it nearly as compact as SauCas9 (1053 amino acids) and well within range of all-in-one AAV delivery. Its spacer-derived guide sequences are longer (24 nt) than those of most other Cas9 orthologs [51], and like SpyCas9, it cleaves both DNA strands between the third and fourth nucleotides of the protospacer (counting from the PAM-proximal end). NmeCas9 also has a longer PAM consensus (5′-N<sub>4</sub>GATT-3′, after the 3′ end of the protospacer’s crRNA-noncomplementary strand) [44, 46, 51–54], leading to a lower density of targetable sites compared to SpyCas9. Considerable variation from this consensus is permitted during bacterial interference [46, 52], and a smaller number of variant PAMs can also support targeting in mammalian cells [53, 54]. Unlike SpyCas9, NmeCas9 has been found to cleave the DNA strand of RNA-DNA hybrid duplexes in a PAM-independent fashion [52, 55] and can also catalyze PAM-independent, spacer-directed cleavage of RNA [56]. Recently, natural Cas9 inhibitors (encoded by bacterial mobile elements) have been identified and validated in *N. meningitidis* and other bacteria with type II-C

systems, providing for genetically encodable off-switches for NmeCas9 genome editing [57, 58]. These “anti-CRISPR” (Acr) proteins [59] enable temporal, spatial, or conditional control over the NmeCas9 system. Natural inhibitors of type II-A systems have also been discovered in *Listeria monocytogenes* [60] and *Streptococcus thermophilus* [61], some of which are effective at inhibiting SpyCas9.

The longer PAM consensus, longer guide sequence, or enzymological properties of NmeCas9 could result in a reduced propensity for off-targeting, and targeted deep sequencing at bioinformatically predicted near-cognate sites is consistent with this possibility [54]. A high degree of genome-wide specificity has also been noted for the dNmeCas9 platform [62]. However, the true, unbiased accuracy of NmeCas9 is not known, since empirical assessments of genome-wide off-target editing activity (independent of bioinformatics prediction) have not been reported for this ortholog. Here we define and confirm many of the parameters of NmeCas9 editing activity in mammalian cells including PAM sequence preferences, guide length limitations, and off-target profiles. Most notably, we use two empirical approaches (GUIDE-seq [63] and SITE-Seq™ [64]) to define NmeCas9 off-target profiles and find that NmeCas9 is a high-fidelity genome-editing platform in mammalian cells, with far lower levels of off-targeting than SpyCas9. These results further validate NmeCas9 as a genome engineering platform and suggest that continued exploration of Cas9 orthologs could identify additional RNA-guided nucleases that exhibit favorable properties, even without the extensive engineering efforts that have been applied to SpyCas9 [31, 34, 35].

## Results

### Co-expressed sgRNA increases NmeCas9 accumulation in mammalian cells

Previously, we demonstrated that NmeCas9 (derived from *N. meningitidis* strain 8013 [51]) can efficiently edit chromosomal loci in human stem cells using either dual RNAs (crRNA + tracrRNA) or a sgRNA [53]. To further define the efficacy and requirements of NmeCas9 in mammalian cells, we first constructed an all-in-one plasmid (pEJS15) that delivers both NmeCas9 protein and a sgRNA in a single transfection vector, similar to our previous all-in-one dual-RNA plasmid (pSimple-Cas9-TracrRNA; Addgene #47868) [53]. The pEJS15 plasmid expresses NmeCas9 fused to a C-terminal single-HA epitope tag and nuclear localization signal (NLS) sequences at both N- and C-termini under the control of the elongation factor-1 $\alpha$  (EF1 $\alpha$ ) promoter. The sgRNA cassette (driven by the U6 promoter) includes two *BsmBI* restriction sites that are used to clone a spacer of interest from short, synthetic oligonucleotide duplexes. First, we cloned three different bacterial spacers (spacers 9, 24, and 25) from the

endogenous *N. meningitidis* CRISPR locus (strain 8013) [51, 52] to express sgRNAs that target protospacer (ps) 9, ps24 or ps25, respectively (Additional file 1: Figure S1A). None of these protospacers have cognate targets in the human genome. We also cloned a spacer sequence to target an endogenous genomic NmeCas9 target site (NTS) from chromosome 10 that we called NTS3 (Table 1). Two of the resulting all-in-one plasmids (spacer9/sgRNA and NTS3/sgRNA), as well as a plasmid lacking the sgRNA cassette, were transiently transfected into HEK293T cells for 48 h, and NmeCas9 expression was assessed by anti-HA western blot (Fig. 1a). As a positive control, we also included a sample transfected with a SpyCas9-expressing plasmid (triple-HA epitope-tagged, and driven by the cytomegalovirus (CMV) promoter) [65] (Addgene #69220). Full-length NmeCas9 was efficiently expressed in the presence of both sgRNAs (lanes 3 and 4). However, the abundance of the protein was much lower in the absence of sgRNA (lane 2). A different type II-C Cas9 (*Corynebacterium diphtheria* Cas9, CdiCas9) was shown to be dramatically stabilized by its cognate sgRNA when subjected to proteolysis in vitro [55]; if similar resistance to proteolysis occurs with NmeCas9 upon sgRNA binding, it could explain some or all of the sgRNA-dependent increase in cellular accumulation.

### Efficient editing in mammalian cells by NmeCas9

To establish an efficient test system for NmeCas9 activity in mammalian cells, we used a co-transfected fluorescent reporter carrying two truncated, partially overlapping GFP fragments that are separated by a cloning site [66] into which we can insert target protospacers for NmeCas9 (Additional file 1: Figure S1B). Cleavage promotes a single-strand-annealing-based repair pathway that generates an intact GFP open reading frame (ORF), leading to fluorescence [66] that can be scored after 48 h by flow cytometry. We generated reporters carrying three validated bacterial protospacers (ps9, ps24, and ps25, as described above) [51, 52] for transient cotransfection into HEK293T cells along with the corresponding NmeCas9/sgRNA constructs. Figure 1b shows that all three natural protospacers of NmeCas9 can be edited in human cells and the efficiency of GFP induction was comparable to that observed with SpyCas9 (Fig. 1b).

Next, we reprogrammed NmeCas9 by replacing the bacterially derived spacers with a series of spacers designed to target 13 human chromosomal sites (Fig. 1c, d and e) with an N<sub>4</sub>GATT PAM (Table 1). These sgRNAs induced insertion/deletion (indel) mutations at all sites tested, except NTS10 (Fig. 1c, lanes 23–25), as determined by T7 Endonuclease 1 (T7E1) digestion (Fig. 1c). The editing efficiencies ranged from 5% for NTS1B site to 47% in the case of NTS33 (Fig. 1d), though T7E1 tends to underestimate the

**Table 1** NmeCas9 or SpyCas9 guide and target sequences used in this study. NTS, NmeCas9 target site; STS, SpyCas9 target site. The sgRNA spacer sequences (5'→3') are shown with their canonical lengths, and with a 5'-terminal G residue; non-canonical lengths are described in the text and figures. Target site sequences are also 5'→3' and correspond to the DNA strand that is non-complementary to the sgRNA, with PAM sequences underlined. The names of sites that showed at least 3% editing with NmeCas9 are indicated in bold

NmeCas9 Site	Gene or locus	Spacer Sequence	Target site, with PAM
<b>Ps9</b>	<i>N.A. (reporter)</i>	GCAUCCUCAGAUUUAGUAUUCAGA	ATGACATCCTCAGATTTAGTATTTCAGAAATATGATTTTC
<b>Ps24</b>	<i>N.A. (reporter)</i>	GUCGCGCAACAAAACCCCGCUACU	GTCGTGCGCAACAAAACCCCGCTACTCGGCGATTTTC
<b>Ps25</b>	<i>N.A. (reporter)</i>	GCAUGCGCGGCGCAUACCUUUAC	TCCTCATGCGCGGCGCATTACCTTTACGCCGGATTGCC
<b>NTS1B</b>	<i>SLC9A9</i>	GGGCAUCAUGAUUUUGAACUCCCU	CCTTGGCATCATGATTTTGAACCTCCATATGTGATTCTA
<b>NTS1C</b>	<i>SLC9A9</i>	GUGGUCUGGGGUACAGCCUUGGCA	TACTUGGTCTGGGTACAGCCTTGGCATCATGATTTTG
<b>NTS1C-OT1</b>	<i>PHKG2</i>	GCGGUGAGGUACAGCCUUGGCA	TAATCGGTGTGAGGTACAGCCTTGGCATCAGGATTTCT
<b>NTS3</b>	<i>AL158836</i>	GAUGCUCAGAAAGAGGAAGCUGGU	GGGGATGCTCAGAAAAGAGGAAGCTGGTTTATGATTGGA
<b>NTS4B</b>	<i>FLJ00328</i>	GGACAGGAGUCGCCAGAGGCCGGU	GCAGGACAGGAGTCGCCAGAGGCCGGTGGTGGATTTC
<b>NTS4C</b>	<i>FLJ00328</i>	GGGGCUGGCUCACGUCGCGCCGC	TGCGGGGCTGGCTCCACGTGCGCCGCGCGGTTTGGG
<b>NTS5</b>	<i>AF064860</i>	GAAACAGACUCGCAAGACUUCAGA	GACAAAACAGACTCGCAAGACTTCAGATACAGATTCCA
<b>NTS7</b>	<i>LOC100505797</i>	GAGGGAGAGGUGAGCGGAUGAA	GCAAAGGAGAGAGGTGAGCGGATGAAGGGAGATTGGT
<b>NTS8</b>	<i>ESPN</i>	GGACGCAAUUCAGAGGUGAUGGG	CGGCGACGCAATTCAGAGGTGATGGGGAGTGATTGTC
<b>NTS9</b>	<i>ZNHIT2</i>	GGCGCUGUUUUCGCAAAGCUUC	CGGCGCGCTGTGTTTTCGCAAAGCTTCCGAGGATTCTC
NTS10	<i>HHLA1</i>	GCAGCCAAGUUUGAGAACUGCUGU	TGTGCAGCCAAGTTTGAGAACTGCTGTTACAGATTTC
<b>NTS11</b>	<i>SMARCB1</i>	GUUCCAGUUGGAAGGGCCAGUGC	TAGATTCCAGTTGGGAAGGGCCAGTGCCCTCCGATTCCA
NTS20	<i>CH251-396E2</i>	GAUACCAGUCAGAUUUGUGCAAUG	TCTAATACCAGTCAGATTTGTGCAATGTTGGGTTTGGGA
<b>NTS21</b>	<i>TNNC1</i>	GCCAGAGCUGCCGCCAGACAGUGA	CAGTCCAGAGCTGCCGCCAGACAGTGATGCTGTCTTGG
NTS22	<i>ANO1</i>	GUUUGGCUGGGCGCUCUGAAACU	TGCTTTTGGCTGGGCGCTCCTGAAACTCTGGGCTTTCA
NTS23	<i>LOC103240889</i>	GCGGCCCGGGGCGGGCCGCGGG	GCGACGGGCCCGGGGCGGGCCGCGGGCGGAGACTGAC
NTS24	<i>AC097382</i>	GCUGCCUCAUGGUGUGUGCUGAGG	GCTGCTGCCTCATGGTGTGTGCTGAGGCTCAGACATCA
<b>NTS25</b>	<i>AC193513</i>	GGUUUCUAUCCUGUCUUCUGCCU	CCGCGTTTCTCATCTGTCTTCTGCCTAGTGGATATGT
<b>NTS26</b>	<i>LOC105378512</i>	GUUCAAAGUAGCGGGCGCUAGGC	GTACTTCAAAGTAGCGGGCGTAGGCGGGTGTCTTCTG
<b>NTS27</b>	<i>TIE1</i>	GUUCUCAAGCCUCGGACCUCGU	CGGCTTCTCAAGCCCTCGGACCTCGTGGGCGTCTTCT
NTS28	<i>AL138885</i>	GCAGGGGCAGGGCGCAUCAGCUG	GCGGCAGGGGCAGGGCGCATCAGCTGTCCGGCTTGGGA
<b>NTS30</b>	<i>NEK8</i>	GGGGCUCGGAGCCCACCCAGGA	CGCGGGGCTCCGGAGCCCACCCAGGACCAGGACTTAG
NTS31	<i>POC1A</i>	GUGGGAAGUGUAGCUCCACCUUCC	ATGTTGGGAAGTGTAGCTCCACCTTCTCCAGACATAG
<b>NTS32</b>	<i>VEGFA</i>	GCCCCGGCUCUGGCUAAGAGGGA	CACACCCCGGCTCTGGCTAAAGAGGGAATGGGCTTTGG
<b>NTS33</b>	<i>VEGFA</i>	GCGGGGAGAAGGCCAGGGGUCACU	GGAGCGGGGAGAAGGCCAGGGGTCACTCCAGGATTCCA
<b>NTS55</b>	<i>CYBB</i>	GCUGGAUACUGUGUGGUAGAGGG	CTAGCTGGATTACTGTGTGGTAGAGGGAGGTGATTAGC
<b>NTS58</b>	<i>AAVS1</i>	GUUUGCCUGGACACCCGUUCUCC	TTTCTTTGCTGGACACCCGTTCTCTGTGGATTTCGG
<b>NTS59</b>	<i>AAVS1</i>	GACCCACAGUGGGGCCACUAGGG	CTCCACCCACAGTGGGGCCACTAGGGACAGGATTGGT
<b>NTS63</b>	<i>AAVS1</i>	GAGUUAGAACUCAGGACCAACUUA	CCAAAGTTAGAACTCAGGACCAACTTATCTGATTTTG
<b>NTS64</b>	<i>Rosa26</i>	GGCAGAUACAGAGGGAAGAGGGGG	AGTTGCAGATCACAGGGAAGAGGGGGAAGGGATTCTC
DTS1-NmeCas9	<i>KL</i>	GCUGGCUGAGCCAUUUUCGGCUC	CTGGCTGGCTGAGCCATTTTCGGCTCTGGAGATTATC
<b>DTS3-NmeCas9</b>	<i>ARHGEF9</i>	GACUGAAGGCGAGGUCCGGGGCGG	ACTGACTGAAGGCGAGGTCCGGGGCGGAGGGGATTGGG
<b>DTS7-NmeCas9</b>	<i>LSP1</i>	GGCUGGCACCCUCAUGUACCCAG	GCCGGCTGGCACCCCTCATGTACCCAGGGGAGATTCCA
<b>DTS8-NmeCas9</b>	<i>PCDHB4</i>	GUGAGUGAAGAUUGAUUAGUUCU	CTGGTGAGTGAAGATATGATAAGTTCTGGGGGATTGTC
<b>SpyCas9 Site</b>	<b>Gene or locus</b>	<b>Spacer Sequence</b>	<b>Target site, with PAM</b>
STS60	<i>AAVS1</i>	GUUAAUGUGGCUCUGGUUCU	CCGGTTAATGTGGCTCTGGTTCTGGGTAC
STS61	<i>AAVS1</i>	GUCCCCUCCACCCACAGUC	TCTGTCCCTCCACCCACAGTGGGCCA

**Table 1** NmeCas9 or SpyCas9 guide and target sequences used in this study. NTS, NmeCas9 target site; STS, SpyCas9 target site. The sgRNA spacer sequences (5'→3') are shown with their canonical lengths, and with a 5'-terminal G residue; non-canonical lengths are described in the text and figures. Target site sequences are also 5'→3' and correspond to the DNA strand that is non-complementary to the sgRNA, with PAM sequences underlined. The names of sites that showed at least 3% editing with NmeCas9 are indicated in bold (Continued)

NmeCas9 Site	Gene or locus	Spacer Sequence	Target site, with PAM
STS62	<i>AAVS1</i>	GGGGCCACUAGGGACAGGAU	AGTGGGGCCACTAGGGACAGGATTGGTGA
DTS1–SpyCas9	<i>KL</i>	GCUGAGCCCAUUUUCGGCUC	CTGGCTGAGCCCATTTTCGGCTCTGGAGA
DTS3–SpyCas9	<i>ARHGEF9</i>	GAAGGCGAGGUCCGGGGCGG	ACTGAAGGCGAGGTCCGGGGCGGAGGGGA
DTS7–SpyCas9	<i>LSP1</i>	GGCACCUCUUAUGUACCCAG	GCTGGCACCTCCATGTACCCAGGGGAGA
DTS8–SpyCas9	<i>PCDHB4</i>	GUGAAGAUUAUGUAAGUUCU	TGAGTGAAGATATGATAAGTTCGGGGGA

true frequencies of indel formation [67]. Targeted deep sequencing of PCR amplicons, which is a more quantitative readout of editing efficiency, confirmed editing with indel efficiencies ranging from ~15 to 85% (Fig. 1e). These results show that NmeCas9 can induce, with variable efficiency, edits at many genomic target sites in human cells. Furthermore, we demonstrated NmeCas9 genome editing in multiple cell lines and via distinct delivery modes. Nucleofection of NmeCas9 ribonucleoprotein (RNP) (loaded with an in vitro-transcribed sgRNA) led to indel formation at three sites in K562 chronic myelogenous leukemia cells and in hTERT-immortalized human foreskin fibroblasts (gift from Dr. Job Dekker) (Fig. 1f). In addition, mouse embryonic stem cells (mESCs) and HEK293T cells were transduced with a lentivirus construct expressing NmeCas9. In these cells, transient transfection of plasmids expressing a sgRNA led to genome editing (Fig. 1f). Collectively, our results show that NmeCas9 can be used for genome editing in a range of human or mouse cell lines via plasmid transfection, RNP delivery, or lentiviral transduction.

#### PAM specificity of NmeCas9 in human cells

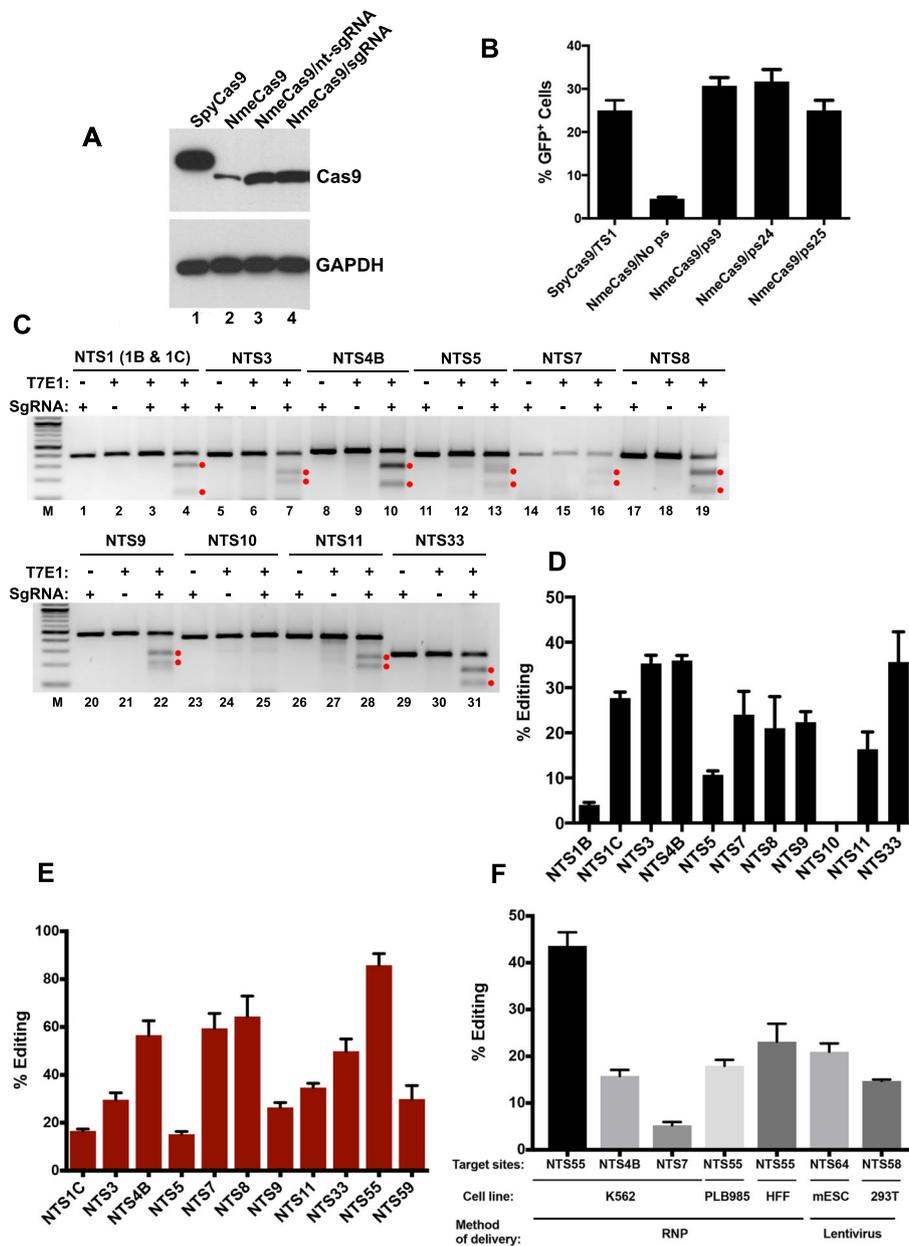
During native CRISPR interference in bacterial cells, considerable variation in the N<sub>4</sub>GATT PAM consensus is tolerated: although the G1 residue (N<sub>4</sub>GATT) is strictly required, virtually all other single mutations at A2 (N<sub>4</sub>GATT), T3 (N<sub>4</sub>GATT), and T4 (N<sub>4</sub>GATT) retain at least partial function in licensing bacterial interference [46, 52]. In contrast, fewer NmeCas9 PAM variants have been validated for genome editing in mammalian cells [53, 54]. To gain more insight into NmeCas9 PAM flexibility and specificity in mammalian cells, and in the context of an otherwise identical target site and an invariant sgRNA, we employed the split-GFP readout of cleavage activity (Additional file 1: Figure S1B). We introduced single-nucleotide mutations at every position of the PAM sequence of ps9, as well as all double-mutant combinations of the four most permissive single mutants, and then measured the ability of NmeCas9 to induce GFP fluorescence in transfected HEK293T cells. The results

are shown in Fig. 2a. As expected, mutation of the G1 residue to any other base reduced editing to background levels, as defined by the control reporter that lacks a protospacer [(no ps), see Fig. 3a]. As for mutations at the A2, T3, and T4 positions, four single mutants (N<sub>4</sub>GCTT, N<sub>4</sub>GTTT, N<sub>4</sub>GACT, and N<sub>4</sub>GATA) and two double mutants (N<sub>4</sub>GTCT and N<sub>4</sub>GACA) were edited with efficiencies approaching that observed with the N<sub>4</sub>GATT PAM. Two other single mutants (N<sub>4</sub>GAGT and N<sub>4</sub>GATG) and three double mutants (N<sub>4</sub>GCCT, N<sub>4</sub>GCTA, and N<sub>4</sub>GTTA) gave intermediate or low efficiencies, and the remaining mutants tested were at or near background levels. We note that some of the minimally functional or non-functional PAMs (e.g., N<sub>4</sub>GAAT and N<sub>4</sub>GATC) in this mammalian assay fit the functional consensus sequences defined previously in *E. coli* [46].

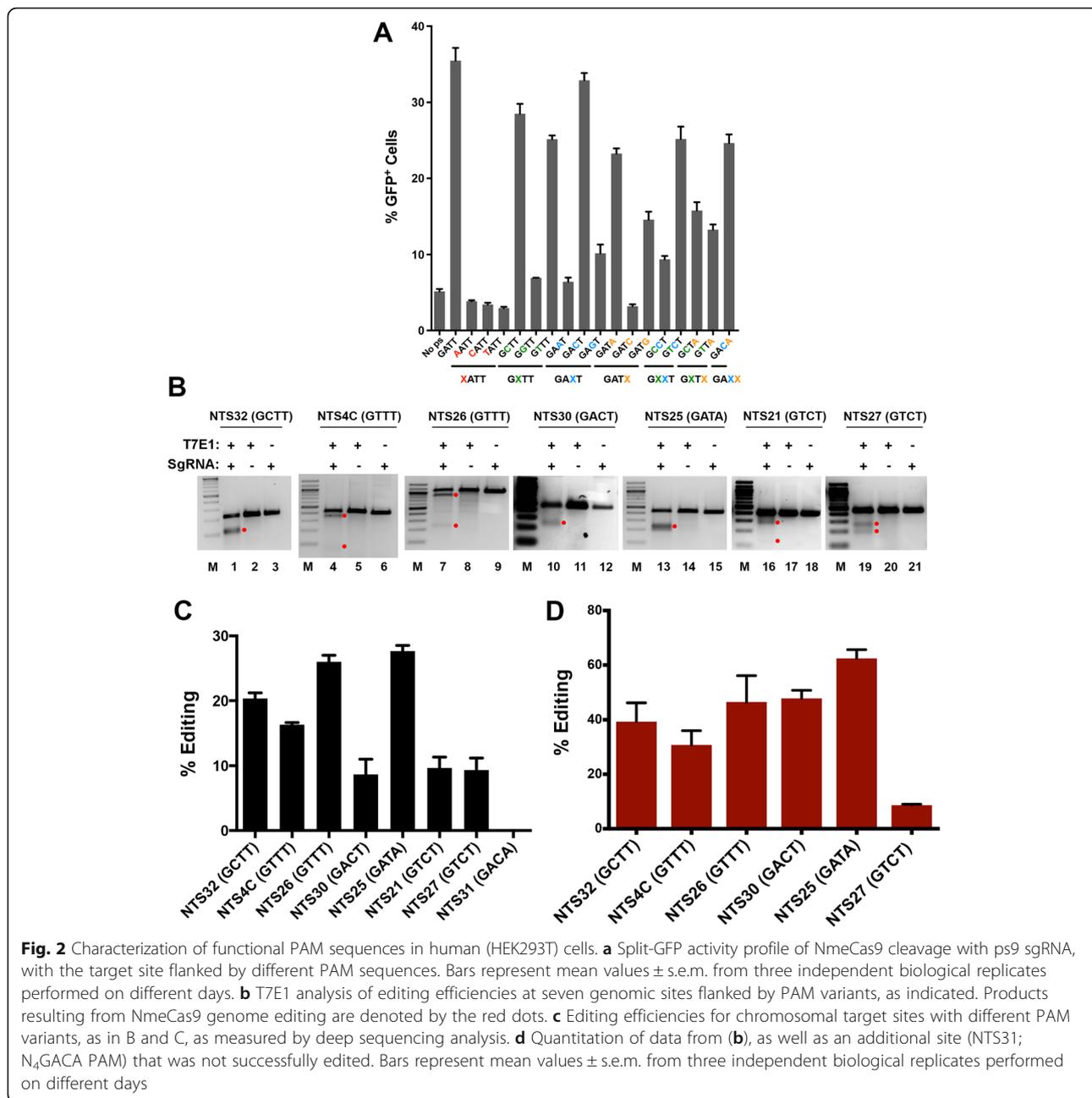
We then used T7E1 analysis to validate genome editing at eight native chromosomal sites associated with the most active PAM variants (N<sub>4</sub>GCTT, N<sub>4</sub>GTTT, N<sub>4</sub>GACT, N<sub>4</sub>GATA, N<sub>4</sub>GTCT, and N<sub>4</sub>GACA). Our results with this set of targets indicate that all of these PAM variants tested except N<sub>4</sub>GACA support chromosomal editing (Fig. 2b and c). Targeted deep sequencing confirmed editing with indel efficiencies ranging from ~8 to 60% (except for target site NTS21, which amplified poorly with Illumina-compatible primers) (Fig. 2d).

#### Apo NmeCas9 is not genotoxic to mammalian cells

NmeCas9 and several other type II-C Cas9 orthologs have been shown to possess an RNA-dependent ssDNA cleavage (DNase H) activity in vitro [52, 55]. R-loops (regions where an RNA strand invades a DNA duplex to form a DNA:RNA hybrid, with the other DNA strand displaced) occur naturally during transcription and other cellular processes [68]. Since DNase H activity is independent of the tracrRNA or the PAM sequence, it is theoretically possible that it could degrade naturally occurring R-loops in living cells. Global degradation of R-loops in cells could result in an increase in DNA damage detectable by increased  $\gamma$ H2AX staining [69]. To test whether the DNase H activity of NmeCas9 could

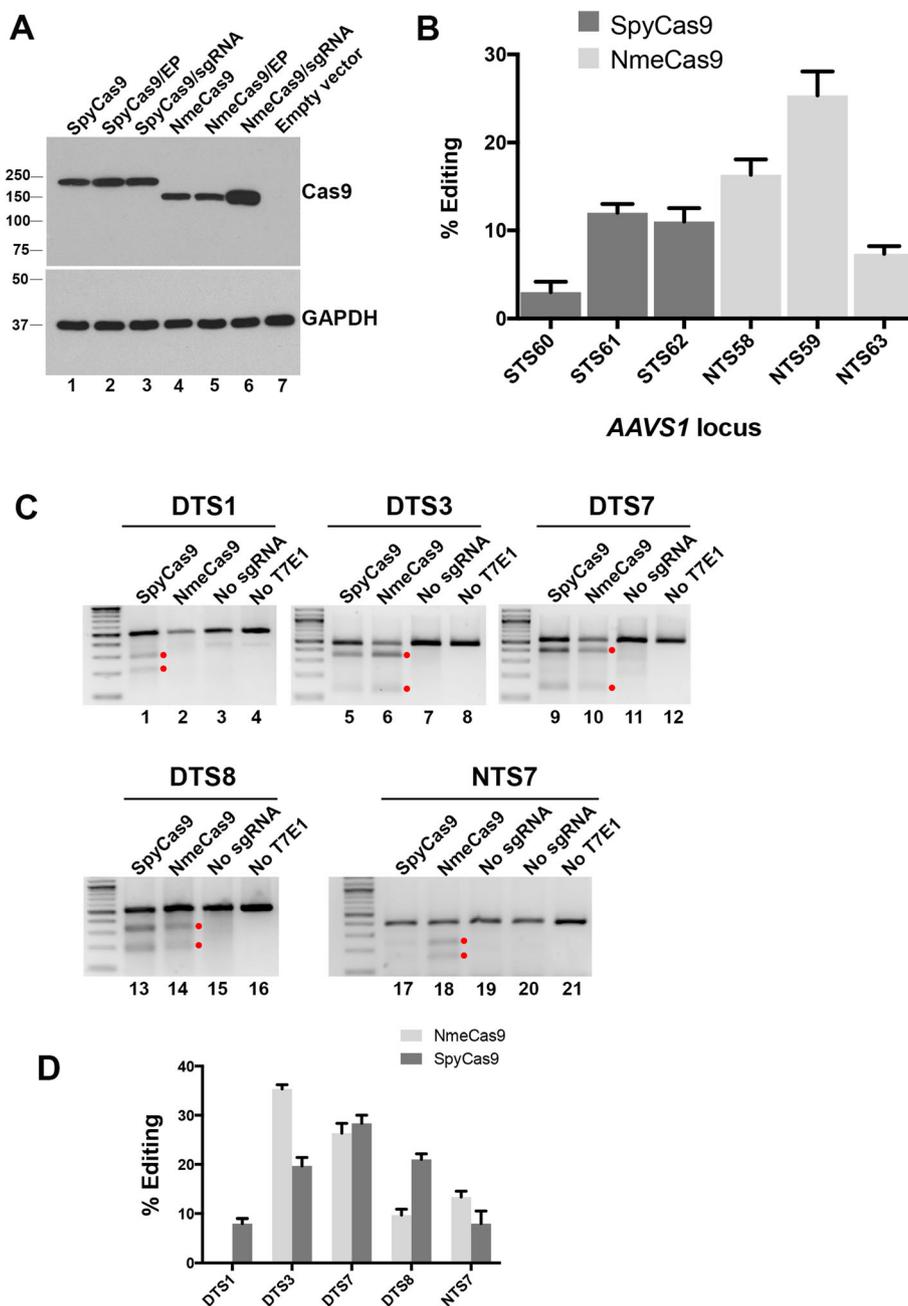


**Fig. 1** NmeCas9 expression and activity in human (HEK293T) cells. **a** Western blot detection of HA-tagged NmeCas9 in transiently transfected HEK293T cells. Lane 1: cells transfected with SpyCas9 plasmid under the control of the CMV promoter. Lane 2: cells transfected with NmeCas9 plasmid under the control of the elongation factor-1 $\alpha$  (EF1 $\alpha$ ) promoter. Lane 3: cells expressing NmeCas9 and a non-targeting sgRNA (nt-sgRNA), which lacks a complementary site in the human genome. Lane 4: cells expressing NmeCas9 and a sgRNA targeting chromosomal site NTS3. Upper panel: anti-HA western blot. Lower panel: anti-GAPDH western blot as a loading control. **b** NmeCas9 targeting co-transfected split-GFP reporter with ps9, ps24, and ps25 sites. Plasmid cleavage by SpyCas9 is used as a positive control, and a reporter without a guide-complementary site (No ps: no protospacer) is used as a negative control to define background levels of recombination leading to GFP<sup>+</sup> cells. **c** NmeCas9 programmed independently with different sgRNAs targeting eleven genomic sites flanked by an N<sub>4</sub>GATT PAM, detected by T7E1 analysis. Products resulting from NmeCas9 genome editing are denoted by the red dots. **d** Quantitation of editing efficiencies from three independent biological replicates performed on different days. Error bars indicate  $\pm$  standard error of the mean ( $\pm$  s.e.m.). **e** Editing efficiencies for chromosomal target sites as measured by PCR and high-throughput sequencing (Deep sequencing). Data are mean values  $\pm$  s.e.m. from three biological replicates performed on different days. **f** Genomic edits with NmeCas9 programmed independently with different guides in different cell lines and using different methods of delivery



lead to an increase in  $\gamma$ H2AX, we transduced mouse embryonic stem cells E14 (mESCs) with lentiviral plasmids expressing NmeCas9 (Addgene #120076) and dNmeCas9 (Addgene #120077) (which lacks DNase H activity [52]) to create stable cell lines expressing NmeCas9 or dNmeCas9, respectively. mESCs are ideal for this purpose as R-loops have been extensively studied in these cells and have been shown to be important for differentiation [70]. We performed  $\gamma$ H2AX staining of these two cell lines and compared them to wildtype E14 cells. As a positive control for  $\gamma$ H2AX induction, we exposed wildtype E14 cells to UV, a known stimulator of the global DNA damage response.

Immunofluorescence microscopy of cells expressing NmeCas9 or dNmeCas9 exhibited no increase in  $\gamma$ H2AX foci compared to wildtype E14, suggesting that sustained NmeCas9 expression is not genotoxic (Additional file 1: Figure S2A). In contrast, cells exposed to UV light showed a significant increase in  $\gamma$ H2AX levels. Flow cytometric measurements of  $\gamma$ H2AX immunostaining confirmed these results (Additional file 1: Figure S2B). We next tested whether continuous NmeCas9 expression in human HEK293T cells has genotoxic effects. We performed  $\gamma$ H2AX staining as above and found no difference between the wildtype cells and those expressing NmeCas9



**Fig. 3** NmeCas9 and SpyCas9 have comparable editing efficiencies in human (HEK293T) cells when targeting the same chromosomal sites. **a** Western blot analysis of NmeCas9 and SpyCas9. HEK293T cells were transfected with the indicated Cas9 ortholog cloned in the same plasmid backbone and fused to the same HA epitope tags and NLSs. Top panel: anti-HA western blot (EP, empty sgRNA plasmid). Bottom panel: anti-GAPDH western blot, used as a loading control. Mobilities of protein markers are indicated. **b** T7E1 analysis of three previously validated SpyCas9 guides targeting the AAVS1 locus, in comparison with NmeCas9 guides targeting nearby AAVS1 sites (mean  $\pm$  s.e.m.,  $n = 3$ ). **c** Representative T7E1 analyses comparing editing efficiencies at the dual target sites DTS1, DTS3, DTS7, DTS8, and NTS7, using the indicated Cas9/sgRNA combinations. Products resulting from Cas9 genome editing are denoted by the red dots. **d** Quantitation of data from (c) (mean  $\pm$  s.e.m.,  $n = 3$ ). Two-tailed paired Student's *T* test showed significant difference between NmeCas9 and SpyCas9 editing of DTS1, DTS3, and DTS8 ( $p < 0.05$ )

(Additional file 1: Figure S2C). These data indicate that NmeCas9 expression does not lead to a global DNA damage response in mESCs or human cells.

**Comparative analysis of NmeCas9 and SpyCas9**  
 SpyCas9 is by far the best-characterized Cas9 ortholog and is therefore the most informative benchmark when

defining the efficiency and accuracy of other Cas9s. To facilitate comparative experiments between NmeCas9 and SpyCas9, we developed a matched Cas9 + sgRNA expression system for the two orthologs. This serves to minimize the expression differences between the two Cas9s in our comparative experiments, beyond those differences dictated by the sequence variations between the orthologs themselves. To this end, we employed the separate pCSDest2-SpyCas9-NLS-3XHA-NLS (Addgene #69220) and pLKO.1-puro-U6sgRNA-BfuA1 (Addgene #52628) plasmids reported previously for the expression of SpyCas9 (driven by the CMV promoter) and its sgRNA (driven by the U6 promoter), respectively [58, 65]. We then replaced the bacterially derived SpyCas9 sequence (i.e., not including the terminal fusions) with that of NmeCas9 in the CMV-driven expression plasmid. This yielded an NmeCas9 expression vector (pEJS424) that is identical to that of the SpyCas9 expression vector in every way (backbone, promoters, UTRs, poly(A) signals, terminal fusions, etc.) except for the Cas9 sequence itself. Similarly, we replaced the SpyCas9 sgRNA cassette in pLKO.1-puro-U6sgRNA-BfuA1 with that of the NmeCas9 sgRNA [46, 53], yielding the NmeCas9 sgRNA expression plasmid pEJS333. This matched system facilitates direct comparisons of the two enzymes' accumulation and activity during editing experiments. To assess relative expression levels of the identically tagged Cas9 orthologs, the two plasmids were transiently transfected into HEK293T cells for 48 h, and the expression of the two proteins was monitored by anti-HA western blot (Fig. 3a). Consistent with our previous data (Fig. 1a), analyses of samples from identically transfected cells show that NmeCas9 accumulation is stronger when co-expressed with its cognate sgRNA (Fig. 3a, compare lane 6 to 4 and 5), whereas SpyCas9 is not affected by the presence of its sgRNA (lanes 1–3).

For an initial comparison of the cleavage efficiencies of the two Cas9s, we chose three previously validated SpyCas9 guides targeting the *AAVS1* “safe harbor” locus [20, 71] and used the CRISPRseek package [72] to design three NmeCas9 guides targeting the same locus within a region of ~700 base pairs (Additional file 1: Figure S3A). The matched Cas9/sgRNA expression systems described above were used for transient transfection of HEK293T cells. T7E1 analysis showed that the editing efficiencies were comparable, with the highest efficiency being observed when targeting the NTS59 site with NmeCas9 (Fig. 3b and Additional file 1: Figure S3B).

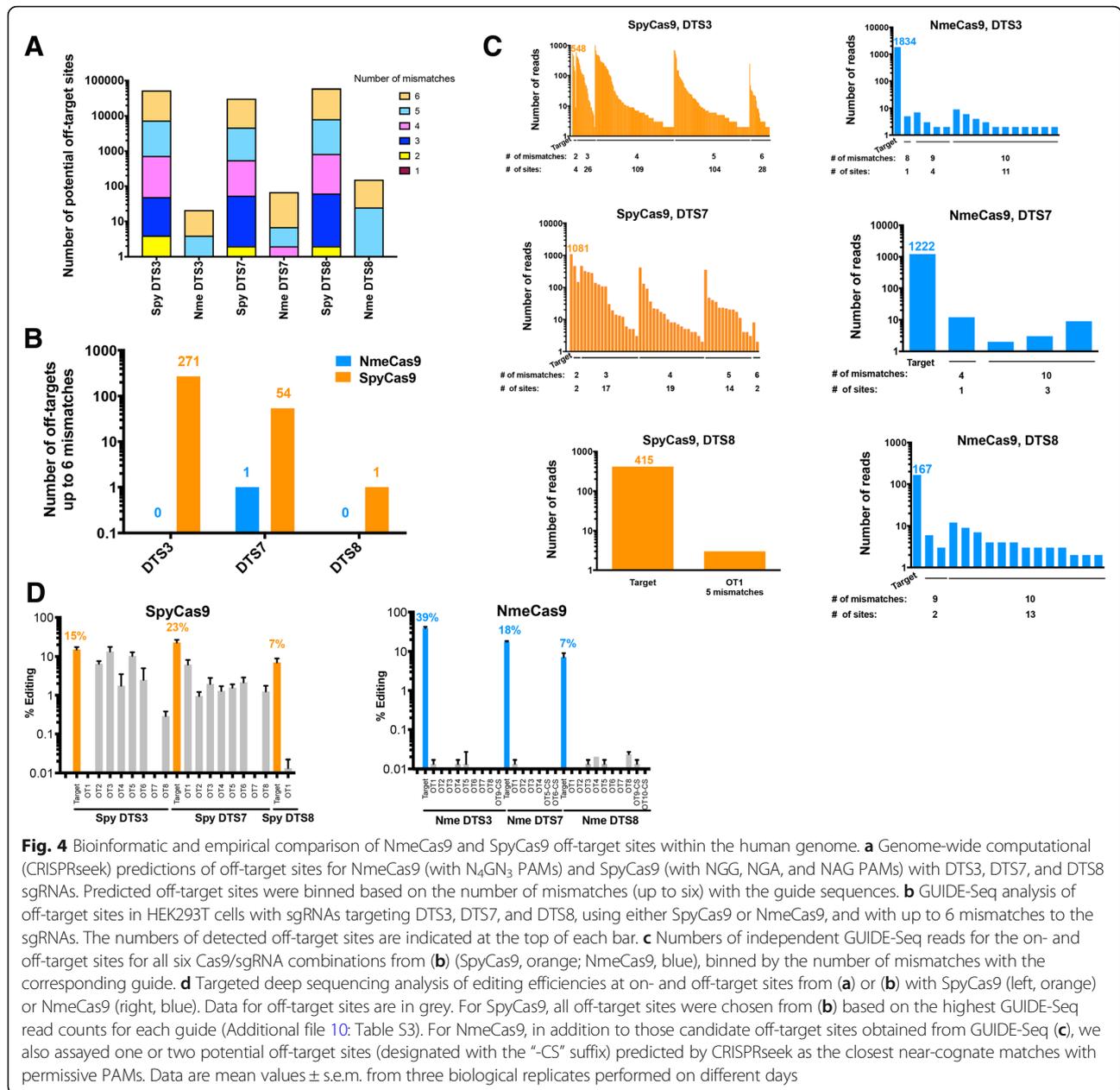
To provide a direct comparison of editing efficiency between the SpyCas9 and NmeCas9 systems, we took advantage of the non-overlapping PAMs of SpyCas9 and NmeCas9 (NGG and N<sub>4</sub>GATT, respectively). Because the optimal SpyCas9 and NmeCas9 PAMs are non-overlapping, it is simple to identify chromosomal target sites

that are compatible with both orthologs, i.e., that are dual target sites (DTSs) with a composite PAM sequence of NGGNGATT that is preferred by both nucleases. In this sequence context, both Cas9s will cleave the exact same internucleotide bond (NN/NNNGGNGATT; cleaved junction in bold, and PAM region underlined), and both Cas9s will have to contend with the exact same sequence and chromatin structural context. Furthermore, if the target site contains a G residue at position -24 of the sgRNA-noncomplementary strand (relative to the PAM) and another at position -20, then the U6 promoter can be used to express perfectly matched sgRNAs for both Cas9 orthologs. Four DTSs with these characteristics were used in this comparison (Additional file 1: Figure S4A). We had previously used NmeCas9 to target a site (NTS7) that happened also to match the SpyCas9 PAM consensus, so we included it in our comparative analysis as a fifth site, even though it has a predicted rG-dT wobble pair at position -24 for the NmeCas9 sgRNA (Additional file 1: Figure S4A).

We compared the editing activities of both Cas9 orthologs programmed to target the five chromosomal sites depicted in Additional file 1: Figure S4A, initially via T7E1 digestion. SpyCas9 was more efficient than NmeCas9 at generating edits at the DTS1 and DTS8 sites (Fig. 3c, lanes 1–2 and 13–14). In contrast, NmeCas9 was more efficient than SpyCas9 at the DTS3 and NTS7 sites (Fig. 3c, lanes 5–6 and 17–18). Editing at DTS7 was approximately equal with both orthologs (Fig. 3c, lanes 9–10). Data from three biological replicates of all five target sites are plotted in Fig. 3d. The remainder of our comparative studies focused on DTS3, DTS7, and DTS8, as they provided examples of target sites with NmeCas9 editing efficiencies that are greater than, equal to, or lower than those of SpyCas9, respectively. The editing efficiency of these three sites was confirmed by targeted deep sequencing (see below). At all three of these sites, the addition of an extra 5'-terminal G residue had little to no effect on editing by either SpyCas9 or NmeCas9 (Additional file 1: Figure S4B). Truncation of the three NmeCas9 guides down to 20 nucleotides (all perfectly matched) again had differential effects on editing efficiency from one site to the next, with no reduction in DTS7 editing, partial reduction in DTS3 editing, and complete loss of DTS8 editing (Additional file 1: Figure S4B). These results establish the guide/target context for deeper comparative analyses of SpyCas9 and NmeCas9 indel spectra and accuracy at shared chromosomal sites.

#### Indel spectrum at NmeCas9- and SpyCas9-edited sites

Our targeted deep sequencing data at the three dual target sites DTS3, DTS7 and DTS8 (Fig. 4d, Additional file 1: Figure S4A and Additional file 2: Table S5) enabled us to



analyze the spectrum of insertions and deletions generated by NmeCas9, in comparison with those of SpyCas9 when editing the exact same sites (Additional file 1: Figures S5B–S8). Although small deletions predominated at all three sites with both Cas9 orthologs, the frequency of insertions was lower for NmeCas9 than it was with SpyCas9 (Additional file 1: Figures S5B–S8). For both SpyCas9 and NmeCas9, the vast majority of insertions were only a single nucleotide (Additional file 1: Figure S7). The sizes of the deletions varied from one target site to the other for both Cas9 orthologs. Our data suggest that at Cas9 edits, deletions predominated over

insertions and the indel size varies considerably from site to site (Additional file 1: Figures S5B, S9 and S10).

**Assessing the genome-wide precision of NmeCas9 editing**

All Cas9 orthologs described to date have some propensity to edit off-target sites lacking perfect complementarity to the programmed guide RNA, and considerable effort has been devoted to developing strategies (mostly with SpyCas9) to increase editing specificity (reviewed in [31, 34, 35]). In comparison with SpyCas9, orthologs such as NmeCas9 that employ longer guide sequences and that require longer PAMs have the potential for

greater on-target specificity, possibly due in part to the lower density of near-cognate sequences. As an initial step in exploring this possibility, we used CRISPRseek [72] to perform a global analysis of potential NmeCas9 and SpyCas9 off-target sites with six or fewer mismatches in the human genome, using sgRNAs specific for DTS3, DTS7, and DTS8 (Fig. 4a) as representative queries. When allowing for permissive and semi-permissive PAMs (NGG, NGA, and NAG for SpyCas9; N<sub>4</sub>GHTT, N<sub>4</sub>GACT, N<sub>4</sub>GAYA, and N<sub>4</sub>GTCT for NmeCas9), potential off-target sites for NmeCas9 were predicted with two to three orders of magnitude lower frequency than for SpyCas9 (Table 2). These results hold true even when we relaxed the PAM requirement to the “slipped” N<sub>5</sub>GHTT, N<sub>5</sub>GACT, N<sub>5</sub>GAYA, and N<sub>5</sub>GTCT PAMs with variant spacing (analyzed because purified, recombinant NmeCas9 has been observed to catalyze DNA cleavage in vitro at such sites [52]). Furthermore, NmeCas9 off-target sites with fewer than five mismatches were rare (two sites with four mismatches) for DTS7, and non-existent for DTS3 and DTS8 (Table 2). Even when we relaxed the NmeCas9 PAM requirement to N<sub>4</sub>GN<sub>3</sub>, which includes some PAMs that enable only background levels of targeting (e.g., N<sub>4</sub>GATC (Fig. 2a)), the vast majority of predicted off-target sites (> 96%) for these three guides had five or more mismatches, and none had fewer than four mismatches among the 24 nucleotides of the spacer (Fig. 4a). In contrast, the SpyCas9 guides targeting DTS3, DTS7, and DTS8 had 49, 54, and 62 predicted off-target sites with three or fewer mismatches, respectively, among the 20 nucleotides of the spacer (Table 2). As speculated previously [53, 54], these bioinformatic predictions suggest the possibility that the NmeCas9 genome-editing system may induce very few undesired mutations, or perhaps none, even when targeting sites that induce substantial off-targeting with SpyCas9.

Although bioinformatic predictions of off-targeting can be useful, it is well established that off-target profiles must be defined experimentally in a prediction-independent fashion due to our limited understanding of target specificity determinants, and the corresponding inability of algorithms to predict all possible sites successfully [31, 34, 35]. The need for empirical off-target profiling is especially acute with Cas9 orthologs that are far less thoroughly characterized than SpyCas9. A previous report used PCR amplification and high-throughput sequencing to detect the frequencies of mutations at 15–20 predicted NmeCas9 off-target sites for each of three guides in human cells and found only background levels of indels in all cases, suggesting a very high degree of precision for NmeCas9 [54]. However, this report restricted its analysis to candidate sites with N<sub>4</sub>GNTT PAMs and three or fewer mismatches (or two mismatches combined with a 1 nucleotide bulge) in the PAM-proximal 19 nucleotides, leaving open the possibility that legitimate off-target sites that did not fit these specific criteria remained unexamined. Accordingly, empirical and minimally biased off-target profiles have never been generated for any NmeCas9/sgRNA combination, and the true off-target propensity of NmeCas9 therefore remains unknown. At the time we began this work, multiple methods for prediction-independent detection of off-target sites had been reported including GUIDE-seq, BLESS, Digenome-Seq, HTGTS, and IDLV capture, each with their own advantages and disadvantages (reviewed in [31, 34, 35]); additional methods (SITE-Seq [64], CIRCLE-seq [73], and BLISS [74]) have been reported more recently. Initially, we chose to apply GUIDE-seq [63], which takes advantage of oligonucleotide incorporation into double-strand break sites, for defining the off-target profiles of both SpyCas9 and NmeCas9 when each is programmed to edit the DTS3, DTS7, and DTS8 sites (Fig. 3c–d) in the human genome.

**Table 2** Number of predicted near-cognate sites in the human genome for the three dual target sites (DTS3, DTS7 and DTS8) analyzed in this study. These potential off-target sites differ from the on-target site by six or fewer mismatches, as listed on the left, and include the functional or semi-functional PAMs shown at the top

Number of mismatches	SpyCas9 sites (NGG, NGA, NAG PAMs)			NmeCas9 sites (N <sub>4</sub> GATT, N <sub>4</sub> GCTT, N <sub>4</sub> GTTT, N <sub>4</sub> GACT, N <sub>4</sub> GATA, N <sub>4</sub> GTCT, N <sub>4</sub> GACA PAMs)			NmeCas9 sites (N <sub>5</sub> GATT, N <sub>5</sub> GCTT, N <sub>5</sub> GTTT, N <sub>5</sub> GACT, N <sub>5</sub> GATA, N <sub>5</sub> GTCT, N <sub>5</sub> GACA “slipped” PAMs)		
	DTS3	DTS7	DTS8	DTS3	DTS7	DTS8	DTS3	DTS7	DTS8
1	0	0	0	0	0	0	0	0	0
2	4	2	2	0	0	0	0	0	0
3	45	52	60	0	0	0	0	0	0
4	680	500	772	0	2	0	0	2	1
5	6691	4116	7325	4	5	25	0	4	12
6	45,897	26,474	52,547	17	61	129	24	62	124
Total	53,317	31,144	60,706	21	68	154	24	68	137

After confirming that the co-transfected double-stranded oligodeoxynucleotide (dsODN) was incorporated efficiently at the DTS3, DTS7, and DTS8 sites during both NmeCas9 and SpyCas9 editing (Additional file 1: Figure S4C), we then prepared GUIDE-seq libraries for each of the six editing conditions, as well as for the negative control conditions (i.e., in the absence of any sgRNA) for both Cas9 orthologs. The GUIDE-seq libraries were then subjected to high-throughput sequencing, mapped, and analyzed as described [75] (Fig. 4b–c). On-target editing with these guides was readily detected by this method, with the number of independent reads ranging from a low of 167 (NmeCas9, DTS8) to a high of 1834 (NmeCas9, DTS3) (Fig. 4c and Additional file 3: Table S2).

For our initial analyses, we scored candidate sites as true off-targets if they yielded two or more independent reads and had six or fewer mismatches with the guide, with no constraints placed on the PAM match at that site. For SpyCas9, two of the sgRNAs (targeting DTS3 and DTS7) induced substantial numbers of off-target editing events (271 and 54 off-target sites, respectively (Fig. 4b)) under these criteria. The majority of these SpyCas9 off-target sites (88% and 77% for DTS3 and DTS7, respectively) were associated with a canonical NGG PAM. Reads were very abundant at many of these loci and at five off-target sites (all with the DTS3 sgRNA) even exceeded the number of on-target reads (Fig. 4c). SpyCas9 was much more precise with the DTS8 sgRNA: we detected a single off-target site with five mismatches and an NGG PAM, and it was associated with only three independent reads, far lower than the 415 reads that we detected at the on-target site (Fig. 4c and Additional file 3: Table S2). Overall, the range of editing accuracies that we measured empirically for SpyCas9—very high (e.g., DTS8), intermediate (e.g., DTS7), and poor (e.g., DTS3)—are consistent with the observations of other reports using distinct guides (reviewed in [31, 34, 35]).

In striking contrast, GUIDE-seq analyses with NmeCas9, programmed with sgRNAs targeting the exact same three sites, yielded off-target profiles that were exceptionally specific in all cases (Fig. 4b–c). For DTS3 and DTS8, we found no reads at any site with six or fewer guide mismatches; for DTS7, we found one off-target site with four mismatches (three of which were at the PAM-distal end; see Additional file 3: Table S2), and even at this site, there were only 12 independent reads,  $\sim 100\times$  fewer than the 1222 reads detected at DTS7 itself. This off-target site was also associated with a PAM (N<sub>4</sub>GGCT) that would be expected to be poorly functional, though it could also be considered a “slipped” PAM with a more optimal consensus but variant spacing (N<sub>5</sub>GCTT). To explore the off-targeting potential of NmeCas9 further, we decreased the stringency of our

mapping to allow detection of off-target sites with up to 10 mismatches. Even in these conditions, only four (DTS7), 15 (DTS8), and 16 (DTS3) candidate sites were identified, most of which had only four or fewer reads (Fig. 4c) and were associated with poorly functional PAMs (Additional file 3: Table S2). We consider it likely that most if not all of these low-probability candidate off-target sites represent background noise caused by spurious priming and other sources of experimental error.

As an additional test of off-targeting potential, we repeated the DTS7 GUIDE-seq experiments with both SpyCas9 and NmeCas9, but this time using a different transfection reagent (Lipofectamine3000 rather than Polyfect). These repeat experiments revealed that >96% (29 out of 30) of off-target sites with up to five mismatches were detected under both transfection conditions for SpyCas9 (Additional file 4: Table S1). However, the NmeCas9 GUIDE-seq data showed no overlap between the potential sites identified under the two conditions, again suggesting that the few off-target reads that we did observe are unlikely to represent legitimate off-target editing sites.

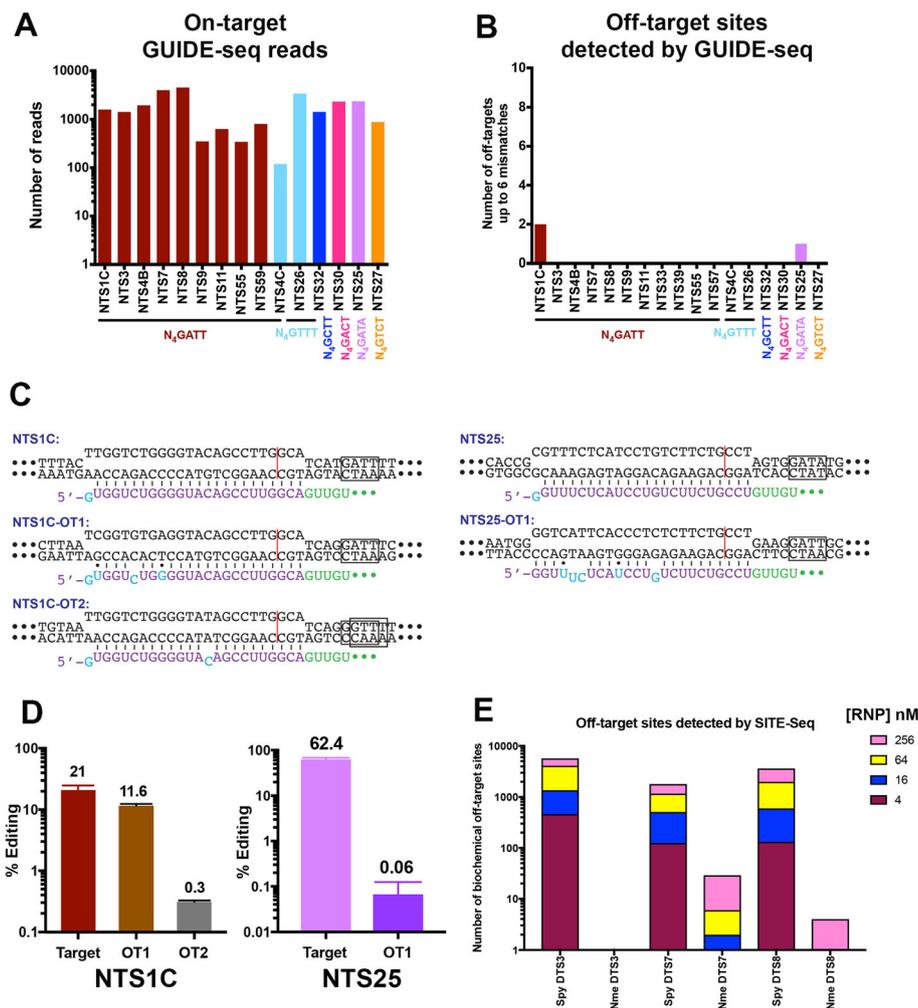
To confirm the validity of the off-target sites defined by GUIDE-seq, we designed primers flanking candidate off-target sites identified by GUIDE-seq, PCR-amplified those loci following standard genome editing (i.e., in the absence of co-transfected GUIDE-seq dsODN) (3 biological replicates), and then subjected the PCR products to high-throughput sequencing to detect the frequencies of Cas9-induced indels. For this analysis, we chose the top candidate off-target sites (as defined by GUIDE-seq read count) for each of the six cases (DTS3, DTS7, and DTS8, each edited by either SpyCas9 or NmeCas9). In addition, due to the low numbers of off-target sites and the low off-target read counts observed during the NmeCas9 GUIDE-seq experiments, we analyzed the top two predicted off-target sites for the three NmeCas9 sgRNAs, as identified by CRISPRseek (Fig. 4a and Table 2) [72]. On-target indel formation was detected in all cases, with editing efficiencies ranging from 7% (DTS8, with both SpyCas9 and NmeCas9) to 39% (DTS3 with NmeCas9) (Fig. 4d). At the off-target sites, our targeted deep-sequencing analyses largely confirmed our GUIDE-seq results: SpyCas9 readily induced indels at most of the tested off-target sites when paired with the DTS3 and DTS7 sgRNAs, and in some cases, the off-target editing efficiencies approached those observed at the on-target sites (Fig. 4d). Although some SpyCas9 off-targeting could also be detected with the DTS8 sgRNA, the frequencies were much lower (<0.1% in all cases). Off-target edits induced by NmeCas9 were far less frequent in all cases, even with the DTS3 sgRNA that was so efficient at on-target mutagenesis: many off-target sites exhibited

editing efficiencies that were indistinguishable from background sequencing error rates (Fig. 4d). These results, in combination with the GUIDE-seq analyses described above, reveal wildtype NmeCas9 to be an exceptionally precise genome-editing enzyme.

To explore NmeCas9 editing accuracy more deeply, we used 16 NmeCas9 target sites among the 24 sites across the genome we tested earlier, 10 with canonical N<sub>4</sub>GATT PAMs and six with variant functional PAMs (Additional file 5: Table S9). We then performed

GUIDE-seq analyses of NmeCas9 editing at these sites. GUIDE-seq analysis readily revealed editing at each of these sites, with on-target read counts ranging from ~100 to ~5000 reads (Fig. 5a) confirming the on-target editing shown earlier by T7E1 assay and deep sequencing analyses (Fig. 1c–e and Fig. 2b–d). Most notably, off-target reads were undetectable by GUIDE-seq with 14 out of the 16 sgRNAs (Fig. 5b).

The two guides with off-target activity (NTS1C and NTS25) had only two and one off-target sites, respectively



**Fig. 5** Off-target analyses for additional NmeCas9 sgRNAs, targeting sites with consensus and variant PAMs. **a** Number of GUIDE-Seq reads for the on-target sites, with the PAM sequences for each site indicated underneath. **b** Number of GUIDE-Seq-detected off-target sites using the Bioconductor package GUIDESeq version 1.1.17 [75] with default settings except that PAM.size = 8, PAM = “NNNNNGATT,” min.reads = 2, max.mismatch = 6, allowed.mismatch.PAM = 4, PAM.pattern = “NNNNNNNN\$,” BSgenomeName = Hsapiens, txdb = TxDb.Hsapiens.UCSC.hg19.knownGene, orgAnn = org.Hs.egSYMBOL gRNA.size was set to length of the gRNA used, and various number of 0's were added at the beginning of weights to make the length of weights equal to the gRNA size. For example, for gRNA with length 24, weights = c(0,0,0,0, 0, 0.014, 0, 0, 0.395, 0.317, 0, 0.389, 0.079, 0.445, 0.508, 0.613, 0.851, 0.732, 0.828, 0.615, 0.804, 0.685, 0.583) for all sixteen sgRNAs used in (a). **c** Schematic diagrams of NmeCas9 sgRNA/DNA R-loops for the NTS1C (left) and NTS25 (right) sgRNAs, at the GUIDE-Seq-detected on- and off-target sites. Black, DNA residues; boxed text, PAM; red line, NmeCas9 cleavage site; cyan and purple, mismatch/wobble and complementary nts (respectively) in the NmeCas9 sgRNA guide region; green, NmeCas9 sgRNA repeat nts. **d** NmeCas9 editing efficiencies at the NTS1C (left) and NTS25 (right) on-target sites, and at the off-target sites detected by GUIDE-Seq from (b), as measured by PCR and high-throughput sequencing. Data are mean values  $\pm$  s.e.m. from three biological replicates performed on different days. **e** Comparison of NmeCas9 and SpyCas9 biochemical off-target sites using SITE-Seq analysis

(Fig. 5b and Additional file 1: Figure S11). Off-target editing was confirmed by high-throughput sequencing and analysis of indels (Fig. 5d). Compared with the on-target site (perfectly matched at all positions other than the 5'-terminal guide nucleotide, and with an optimal N<sub>4</sub>GATT PAM), the efficiently targeted NTS1C-OT1 had two wobble pairs and one mismatch (all in the nine PAM-distal nucleotides), as well as a canonical N<sub>4</sub>GATT PAM (Fig. 5c and Additional file 3: Table S2). The weakly edited NTS1C-OT2 site had only a single mismatch (at the 11th nucleotide, counting in the PAM-distal direction), but was associated with a non-canonical N<sub>4</sub>GGTT (or a "slipped" N<sub>5</sub>GTTT) PAM (Fig. 5c and Additional file 3: Table S2). NTS25 with an N<sub>4</sub>GATA PAM was the other guide with a single off-target site (NTS25-OT1), where NmeCas9 edited up to ~1000× less efficiently than at the on-target site (Fig. 5d). This minimal amount of off-target editing arose despite the association of NTS25-OT1 with an optimal N<sub>4</sub>GATT PAM, unlike the variant N<sub>4</sub>GATA PAM that flanks the on-target site. Overall, our GUIDE-seq and sequencing-based analyses demonstrate that NmeCas9 genome editing is exceptionally accurate: we detected and confirmed cellular off-target editing with only two of the 19 guides tested, and even in those two cases, only one or two off-target sites could be found for each. Furthermore, of the three bona fide off-target sites that we identified, only one generated indels at substantial frequency (11.6%); indel frequencies were very modest (0.3% or lower) at the other two off-target sites.

We next sought to corroborate and expand on our GUIDE-seq results with a second prediction-independent method. We applied the SITE-Seq assay, a biochemical method that does not rely on cellular events such as DNA repair, thus potentially enabling a more thorough profiling of genome-wide specificity [64]. SITE-Seq libraries were prepared for the three dual target sites with both Cas9 orthologs as well as for 12 of the NmeCas9-only target sites. SITE-Seq was performed on HEK293T genomic DNA (gDNA) treated with a range of RNP concentrations (4–256 nM) previously shown to discriminate high- and low-probability cellular off-targets [64]. Finally, the resulting libraries were sequenced, aligned, and then analyzed as previously described [64].

Negative controls without RNP recovered zero sites across any concentrations, whereas SpyCas9 assembled with sgRNAs targeting DTS3, DTS7, or DTS8 recovered hundreds (at 4 nM RNP) to thousands (at 256 nM RNP) of biochemical off-target sites (Fig. 5e). In contrast, NmeCas9 assembled with sgRNAs targeting the same three sites recovered only their on-target sites at 4 nM RNP and at most 29 off-target sites at 256 nM RNP (Fig. 5e). Moreover, the 12 additional NmeCas9 target sites showed

similarly high specificity: eight samples recovered only the on-target sites at 4 nM RNP and six of those recovered no more than nine off-targets at 256 nM RNP (Additional file 1: Figure S5A). Across NmeCas9 RNPs, off-target sequence mismatches appeared enriched in the 5' end of the sgRNA target sequence (Additional file 6: Table S4). Finally, three of the NmeCas9 RNPs (NTS30, NTS4C, and NTS59) required elevated concentrations to retrieve their on-targets, potentially due to poor sgRNA transcription and/or RNP assembly. These RNPs were therefore excluded from further analysis.

We next performed cell-based validation experiments to investigate whether any of the biochemical off-targets were edited in cells. Since NmeCas9 recovered only ~100 biochemical off-targets across all RNPs and concentrations, we could examine each site for editing in cells. SpyCas9 generated >10,000 biochemical off-targets across all DTS samples, preventing comprehensive cellular profiling. Therefore, for each RNP, we randomly selected 95 high cleavage sensitivity SITE-Seq sites (i.e., recovered at all concentrations tested in SITE-Seq) for examination, as we predicted those were more likely to accumulate edits in cells [64] (Additional file 2: Table S5). Notably, only a subset of the sites validated from GUIDE-seq were contained within this list of sites (1/8 and 5/8 overlapping sites for DTS3 and DTS7, respectively). SITE-Seq and GUIDE-seq validations were performed on the same gDNA samples to facilitate comparisons between data sets.

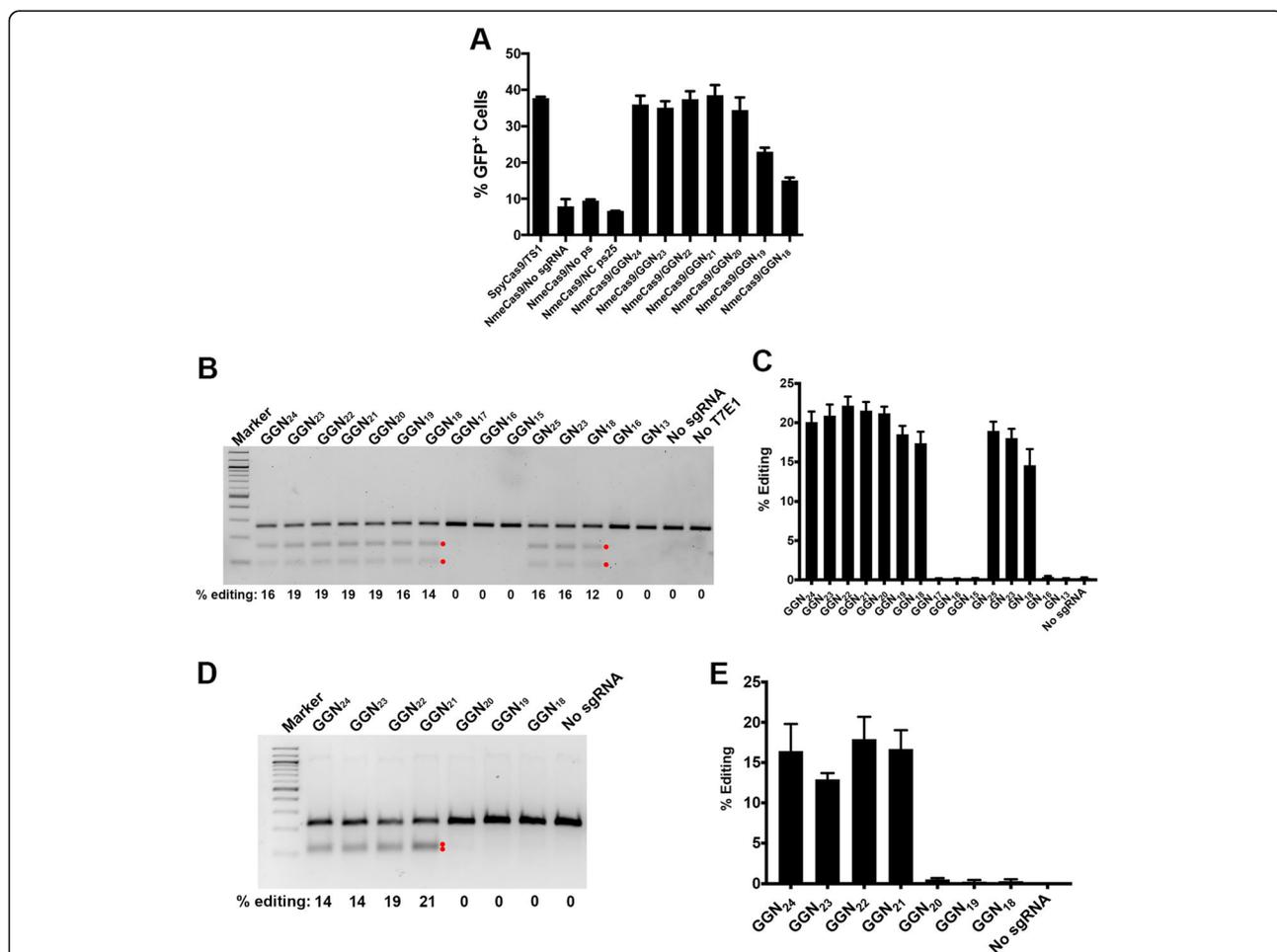
Across all NmeCas9 RNPs, only three cellular off-targets were observed. These three all belonged to the NTS1C RNP, and two of them had also been detected previously with GUIDE-seq. All high cleavage sensitivity SITE-Seq sites (i.e., all on-targets and the single prominent NTS1C off-target, NTS1C-OT1) showed editing in cells. Conversely, SITE-Seq sites with low cleavage sensitivity, defined as being recovered at only 64 nM and/or 256 nM RNP, were rarely found as edited (2/93 sites). Importantly, this suggests that we identified all or the clear majority of NmeCas9 cellular off-targets, albeit at our limit of detection. Across all SpyCas9 RNPs, 14 cellular off-targets were observed (8/70 sites for DTS3, 6/83 sites for DTS7, and 0/79 sites for DTS8) [Additional file 2: Table S5; not all 95 amplicons were included in the final analysis as some were filtered due to low read coverage or high variant calls in the untreated sample (see materials and methods for more details)]. Since our data set was only a subset of the total number of high cleavage sensitivity SITE-Seq sites and excluded many of the GUIDE-seq validated sites, we expect that sequencing all SITE-Seq sites may uncover additional cellular off-targets. Taken together, these data corroborate our GUIDE-seq results, suggesting that NmeCas9 can serve as a highly specific genome editing platform.

**Functionality of truncated sgRNAs with NmeCas9**

SpyCas9 can accommodate limited variation in the length of the guide region (normally 20 nucleotides) of its sgRNAs [76–79], and sgRNAs with modestly lengthened (22-nt) or shortened (17–18-nt) guide regions can even enhance editing specificity by reducing editing at off-target sites by a greater degree than they affect editing at the on-target site [76, 77]. To test the length dependence of the NmeCas9 guide sequence (normally 24 nucleotides [51]) during mammalian editing, we constructed a series of sgRNAs containing 18, 19, 20, 21, 22, 23, and 24 nucleotides of complementarity to ps9 cloned into the split-GFP reporter plasmid (Additional file 1: Figure S12A). All designed guides started with two guanine nucleotides (resulting in 1–2 positions of target

non-complementarity at the very 5' end of the guide) to facilitate transcription, analogous to the SpyCas9 “GGN<sub>20</sub>” sgRNAs [76]. We then measured the abilities of these sgRNAs to direct NmeCas9 cleavage of the reporter in human cells. sgRNAs that have 20–23 nucleotides of target complementarity showed activities comparable to the sgRNA with the natural 24 nucleotides of complementarity, whereas sgRNAs containing 18 or 19 nucleotides of complementarity show lower activity (Fig. 6a).

We next used a native chromosomal target site (NTS33 in *VEGFA*, as in Fig. 1c–e) to test the editing efficiency of NmeCas9 spacers of varying lengths (Additional file 1: Figure S12B). sgRNA constructs included one or two 5'-terminal guanine residues to enable transcription by the U6 promoter, sometimes resulting in 1–2 nucleotides



**Fig. 6** NmeCas9 guide length requirements in mammalian cells. **a** Split-GFP activity profile of NmeCas9 cleavage with ps9 sgRNAs bearing spacers of varying lengths (18–24 nt) along with 5'-terminal G residues to enable transcription. Bars represent mean values ± s.e.m. from three independent biological replicates performed on different days. **b** T7E1 analysis of editing efficiencies at the NTS33 genomic target site (with an N<sub>4</sub>GATT PAM) with sgRNAs bearing spacers of varying lengths (13–25 nt) with 1–2 5'-terminal G residues. Products resulting from NmeCas9 genome editing are denoted by the red dots. **c** Quantitation of editing efficiencies (of experiment in **b**) from three independent biological replicates performed on different days. Error bars indicate ± standard error of the mean (± s.e.m.). **d** As in (**b**), but targeting the NTS32 genomic site (with a N<sub>4</sub>GCTT PAM). **e** Quantitation of editing efficiencies (of experiment in **d**) from three independent biological replicates performed on different days. Error bars indicate ± standard error of the mean (± s.e.m.)

of target non-complementarity at the 5' end of the guide sequence. sgRNAs with 20, 21, or 22 nucleotides of target complementarity (GGN<sub>18</sub>, GGN<sub>19</sub>, and GGN<sub>20</sub>, respectively) performed comparably to the natural guide length (24 nucleotides of complementarity, GN<sub>23</sub>) at this site (Fig. 6b–c), and within this range, the addition of 1–2 unpaired G residues at the 5' end had no adverse effect. These results are consistent with the results obtained with the GFP reporter (Fig. 6a). sgRNAs with guide lengths of 19 nucleotides or shorter, along with a single mismatch in the first or second position (GGN<sub>17</sub>, GGN<sub>16</sub>, and GGN<sub>15</sub>), did not direct detectable editing, nor did a sgRNA with perfectly matched guide sequences of 17 or 14 nucleotides (GN<sub>16</sub> and GN<sub>13</sub>, respectively) (Fig. 6b–c). However, a 19-nt guide with no mismatches (GN<sub>18</sub>) successfully directed editing. These results indicate that 19–26-nt guides can be tolerated by NmeCas9, but that activity can be compromised by guide truncations from the natural length of 24 nucleotides down to 17–18 nucleotides and smaller, and that single mismatches (even at or near the 5'-terminus of the guide) can be discriminated against with a 19-nt guide.

The target sites tested in Fig. 6a–c are both associated with a canonical N<sub>4</sub>GATT PAM. To examine length dependence at a site with a variant PAM, we varied guide sequence length at the N<sub>4</sub>GCTT-associated NTS32 site (also in *VEGFA*). In this experiment, each of the guides had two 5'-terminal G residues, accompanied by 1–2 terminal mismatches with the target sequence (Additional file 1: Figure S12C). At the NTS32 site, sgRNAs with 21–24 nucleotides of complementarity (GGN<sub>24</sub>, GGN<sub>23</sub>, GGN<sub>22</sub>, and GGN<sub>21</sub>) supported editing, but shorter guides (GGN<sub>20</sub>, GGN<sub>19</sub>, and GGN<sub>18</sub>) did not (Fig. 6d–e). We conclude that sgRNAs with 20 nucleotides of complementarity can direct editing at some sites (Fig. 6b–c) but not all (Fig. 6d–e). It is possible that this minor variation in length dependence can be affected by the presence of mismatched 5'-terminal G residues in the sgRNA, the adherence of the target to the canonical N<sub>4</sub>GATT PAM consensus, or both, but the consistency of any such relationship will require functional tests at much larger numbers of sites. Nonetheless, NmeCas9 guide truncations of 1–3 nucleotides appear to be functional in most cases, in agreement with the results of others [54].

#### Truncated sgRNAs reduce off-target cleavage by NmeCas9

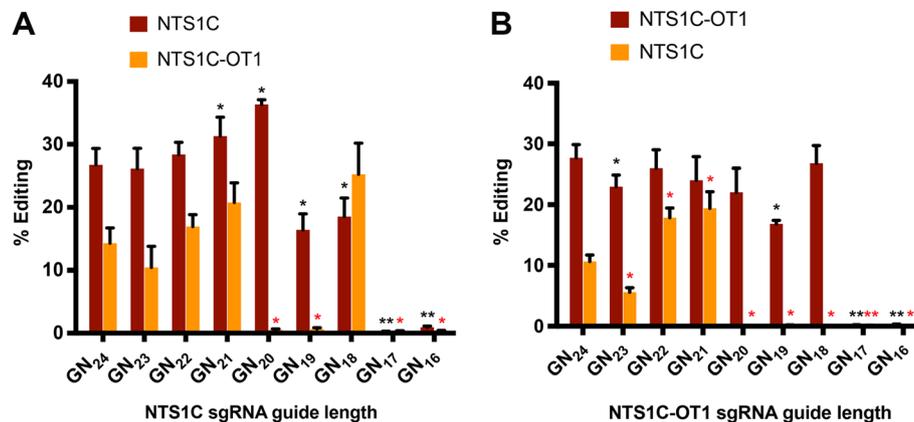
Although NmeCas9 exhibits very little propensity to edit off-target sites, for therapeutic applications, it may be desirable to suppress even the small amount of off-targeting that occurs (Fig. 5). Several strategies have been developed to suppress off-targeting by SpyCas9 [31, 34, 35], some of which could be readily applied to other orthologs. For

example, truncated sgRNAs (tru-sgRNAs) sometimes suppress off-target SpyCas9 editing more than they suppress on-target editing [77]. Because 5'-terminal truncations are compatible with NmeCas9 function (Fig. 6), we tested whether NmeCas9 tru-sgRNAs can have similar suppressive effects on off-target editing without sacrificing on-target editing efficiency.

First, we tested whether guide truncation can lead to NmeCas9 editing at novel off-target sites (i.e., at off-target sites not edited by full-length guides), as reported previously for SpyCas9 [77]. Our earlier tests of NmeCas9 on-target editing with tru-sgRNAs used guides targeting the NTS33 (Fig. 6b–c) and NTS32 (Fig. 6d–e) sites. We again used GUIDE-seq with a subset of the validated NTS32 and NTS33 tru-sgRNAs to determine whether NmeCas9 guide truncation leads to off-target editing at new sites, and found none (Additional file 1: Figure S13). Although we cannot rule out the possibility that other NmeCas9 guides could be identified that yield novel off-target events upon truncation, our results suggest that de novo off-targeting by NmeCas9 tru-sgRNAs is unlikely to be a pervasive problem.

The most efficiently edited off-target site from our previous analyses was NTS1C-OT1, providing us with our most stringent test of off-target suppression. When targeted by the NTS1C sgRNA, NTS1C-OT1 has one rG-dT wobble pair at position –16 (i.e., at the 16th base pair from the PAM-proximal end of the R-loop), one rC-dC mismatch at position –19, and one rU-dG wobble pair at position –23 (Fig. 5c). We generated a series of NTS1C-targeting sgRNAs with a single 5'-terminal G (for U6 promoter transcription) and spacer complementarities ranging from 24 to 15 nucleotides (GN<sub>24</sub> to GN<sub>15</sub>, Additional file 1: Figure S14A, top panel). Conversely, we designed a similar series of sgRNAs with perfect complementarity to NTS1C-OT1 (Additional file 1: Figure S14B, top panel). Consistent with our earlier results with other target sites (Fig. 6), T7E1 analyses revealed that both sets of guides enabled editing of the perfectly-matched on-target site with truncations down to 19 nucleotides (GN<sub>18</sub>), but that shorter guides were inactive. On-target editing efficiencies at both sites were comparable across the seven active guide lengths (GN<sub>24</sub> through GN<sub>18</sub>), with the exception of slightly lower efficiencies with the GN<sub>19</sub> guides (Additional file 1: Figure S14A & B, middle and bottom panels).

We then used targeted deep sequencing to test whether off-target editing is reduced with the truncated sgRNAs. With both sets of sgRNAs (perfectly complementary to either NTS1C or NTS1C-OT1), we found that off-targeting at the corresponding near-cognate site persisted with the four longest guides (GN<sub>24</sub>, GN<sub>23</sub>, GN<sub>22</sub>, GN<sub>21</sub>; Fig. 7). However, off-targeting was abolished with the GN<sub>20</sub> guide, without any significant



**Fig. 7** Guide truncation can suppress off-target editing by NmeCas9. **a** Editing efficiencies at the NTS1C (on-target, red) and NTS1C-OT1 (off-target, orange) genomic sites, after editing by NmeCas9 and NTS1C sgRNAs of varying lengths, as measured by PCR and high-throughput sequencing. Data are mean values  $\pm$  s.e.m. from three biological replicates performed on different days. **b** As in **(a)**, but using sgRNAs perfectly complementary to the NTS1C-OT1 genomic site. Two-tailed paired Student's *T* test showed significant difference in on- and off-target editing efficiency as a function of guide truncation. On-target editing (black asterisk) or off-target editing (red asterisk) is compared to the baseline condition GN<sub>24</sub>, respectively ( $p < 0.05$  is annotated with one asterisk;  $p < 0.01$  is annotated with two asterisks)

reduction in on-target editing efficiencies (Fig. 7). Off-targeting was also absent with the GN<sub>19</sub> guide, though on-target editing efficiency was compromised. These results, albeit from a limited data set, indicate that truncated sgRNAs (especially those with 20 or 19 base pairs of guide/target complementarity, 4–5 base pairs fewer than the natural length) can suppress even the limited degree of off-targeting that occurs with NmeCas9.

Unexpectedly, even though off-targeting at NTS1C-OT1 was abolished with the GN<sub>20</sub> and GN<sub>19</sub> truncated NTS1C sgRNAs, truncating by an additional nucleotide (to generate the GN<sub>18</sub> sgRNA) once again yielded NTS1C-OT1 edits (Fig. 7a). This could be explained by the extra G residue at the 5'-terminus of each sgRNA in the truncation series (Additional file 1: Figure S14). With the NTS1C GN<sub>19</sub> sgRNA, both the 5'-terminal G residue and the adjacent C residue are mismatched with the NTS1C-OT1 site. In contrast, with the GN<sub>18</sub> sgRNA, the 5'-terminal G is complementary to the off-target site. In other words, with the NTS1C GN<sub>19</sub> and GN<sub>18</sub> sgRNAs, the NTS1C-OT1 off-target interactions (which are identical in the PAM-proximal 17 nucleotides) include two additional nucleotides of non-complementarity or one additional nucleotide of complementarity, respectively. Thus, the more extensively truncated GN<sub>18</sub> sgRNA has *greater* complementarity with the NTS1C-OT1 site than the GN<sub>19</sub> sgRNA, explaining the re-emergence of off-target editing with the former. This observation highlights the fact that the inclusion of a 5'-terminal G residue that is mismatched with the on-target site, but that is complementary to a C residue at an off-target site, can limit the effectiveness of a truncated guide at suppressing off-target editing, necessitating care in truncated sgRNA

design when the sgRNA is generated by cellular transcription. This issue is not a concern with sgRNAs that are generated by other means (e.g., chemical synthesis) that do not require a 5'-terminal G. Overall, our results demonstrate that NmeCas9 genome editing is exceptionally precise, and even when rare off-target editing events occur, tru-sgRNAs can provide a simple and effective way to suppress them.

## Discussion

The ability to use type II and type V CRISPR-Cas systems as RNA-programmable DNA-cleaving systems [13, 14, 27] is revolutionizing many aspects of the life sciences and holds similar promise for biotechnological, agricultural, and clinical applications. Most applications reported thus far have used a single Cas9 ortholog (Spy-Cas9). Thousands of additional Cas9 orthologs have also been identified [28], but only a few have been characterized, validated for genome engineering applications, or both. The development of additional orthologs promises to increase the number of targetable sites (through new PAM specificities), extend multiplexing possibilities (for pairwise combinations of Cas9 orthologs with orthogonal guides), and improve deliverability (for the more compact Cas9 orthologs). In addition, some Cas9s may show mechanistic distinctions (such as staggered vs. blunt dsDNA breaks) [80], greater protein stability in vivo, improved control mechanisms (e.g., via multiple anti-CRISPRs that act at various stages of the DNA cleavage pathway) [57, 58, 60, 61, 81–83], and other enhancements. Finally, some may exhibit a greater natural propensity to distinguish between on- vs. off-target sites during genome editing applications, obviating the need

for extensive engineering (as was necessary with SpyCas9) to attain the accuracy needed for many applications, especially therapeutic development.

Here we have further defined the properties of NmeCas9 during editing in human cells, including validation and extension of previous analyses of guide length and PAM requirements [46, 53, 54]. Intriguingly, the tolerance to deviations from the N<sub>4</sub>G(A/C)TT natural PAM consensus [51] observed *in vitro* and in bacterial cells [46, 52] is considerably reduced in the mammalian context, i.e., fewer PAM variations are permitted during mammalian editing. The basis for this context-dependent difference is not clear but may be due in part to the ability to access targets within eukaryotic chromatin, or to the larger numbers of PAMs that must be scanned (due to the much larger genome sizes in mammalian cells), since lower SpyCas9/sgRNA concentrations (relative to potential DNA substrates) have been shown to improve accuracy [30, 84, 85]. We have also found that steady-state NmeCas9 levels in human cells are markedly increased in the presence of its cognate sgRNA, suggesting that sgRNA-loaded NmeCas9 is more stable than *apo* NmeCas9. An increased proteolytic sensitivity of *apo* Cas9 relative to the sgRNA-bound form has been noted previously for a different type II-C ortholog (CdiCas9 [55]).

A previous report indicated that NmeCas9 has high intrinsic accuracy, based on analyses of candidate off-target sites that were predicted bioinformatically [54]. However, the true genome-wide accuracy of NmeCas9 was not assessed empirically, as is necessary given well-established imperfections in bioinformatic predictions of off-targeting [31, 34, 35]. We have used GUIDE-seq [63] and SITE-Seq [64] to define the genome-wide accuracy of wildtype NmeCas9, including side-by-side comparisons with wildtype SpyCas9 during editing of identical on-target sites. We find that NmeCas9 is a consistently high-accuracy genome editor, with off-target editing undetectable above background with 17 out of 19 analyzed sgRNAs, and only one or three verified off-target edits with the remaining two guides. We observed this exquisite specificity by NmeCas9 even with sgRNAs that target sites (DTS3 and DTS7 (see Fig. 4)) that are highly prone to off-target editing when targeted with SpyCas9. Of the four off-target sites that we validated, three accumulated < 1% indels. Even with the one sgRNA that yielded a significant frequency of off-target editing (NTS1C, which induced indels at NTS1C-OT1 with approximately half the efficiency of on-target editing), the off-targeting with wildtype NmeCas9 could be easily suppressed with truncated sgRNAs. Our ability to detect NTS25-OT1 editing with GUIDE-seq, despite its very low (0.06%) editing efficiency based on high-throughput sequencing, indicates that our GUIDE-seq experiments can identify even very low-efficiency off-target editing sites.

Similar considerations apply to our SITE-Seq analyses. We observed high accuracy even when NmeCas9 is delivered by plasmid transfection, a delivery method that is associated with higher off-target editing than more transient delivery modes such as RNP delivery [86, 87]. Low off-target activity even with sustained effector expression could be particularly useful for AAV delivery, as recently validated for NmeCas9 [88].

The two type II-C Cas9 orthologs (NmeCas9 and CjeCas9) that have been validated for mammalian genome editing and assessed for genome-wide specificity [47, 54] (this work) and both have proven to be naturally hyper-accurate. Both use longer guide sequences than the 20-nucleotide guides employed by SpyCas9, and both also have longer and more restrictive PAM requirements. For both type II-C orthologs, it is not yet known whether the longer PAMs, longer guides, or both account for the limited off-target editing. Type II-C Cas9 orthologs generally cleave dsDNA more slowly than SpyCas9 [49, 55], and it has been noted that lowering  $k_{cat}$  can, in some circumstances, enhance specificity [89]. In keeping with the generally lower enzymatic activity, a recent report found NmeCas9 to exhibit consistently lower cellular editing activity than the two type II-A Cas9s tested (SpyCas9 and SauCas9), though most tests in that report were done with suboptimal guide lengths for NmeCas9 [90]. Whatever the mechanistic basis for the high intrinsic accuracy, it is noteworthy that it is a property of the native proteins, without a requirement for extensive engineering. This adds to the motivation to identify more Cas9 orthologs with human genome-editing activity, as it suggests that it may be unnecessary in many cases (perhaps especially among type II-C enzymes) to invest heavily in structural and mechanistic analyses and engineering efforts to attain sufficient accuracy for many applications and with many desired guides, as was done with (for example) SpyCas9 [32, 33, 37, 38, 65]. Although Cas9 orthologs with more restrictive PAM requirements (such as NmeCas9, CjeCas9, and GeoCas9) by definition will afford lower densities of potential target sites than SpyCas9 (which also usually affords the highest on-target editing efficiencies among established Cas9 orthologs), the combined targeting possibilities for multiple such Cas9s will increase the targeting options available within a desired sequence window, with little propensity for off-targeting. The continued exploration of natural Cas9 variation, especially for those orthologs with other advantages such as small size and anti-CRISPR off-switch control, therefore has great potential to advance the CRISPR genome editing revolution.

## Conclusions

NmeCas9 is an intrinsically high-accuracy genome-editing enzyme in mammalian cells, and the limited

off-target editing that occurs can (at least in some cases) be suppressed by guide truncation. Continued exploration of Cas9 orthologs could therefore yield additional enzymes that do not require extensive characterization and engineering to prevent off-target editing.

## Methods

### Plasmids

Two plasmids for the expression of NmeCas9 were used in this study. The first construct (used in Figs. 1, 2, 6 and 7) was derived from the plasmid pSimpleII where NmeCas9 was cloned under the control of the elongation factor-1 $\alpha$  promoter, as described previously [53]. The *Cas9* gene in this construct expresses a protein with two NLSs and an HA tag. To make an all-in-one expression plasmid, a fragment containing a *BsmBI*-crRNA cassette linked to the tracrRNA by six nucleotides, under the control of U6 RNA polymerase III promoter, was synthesized as a gene block (Integrated DNA Technologies) and inserted into pSimpleII, generating the pSimpleII-Cas9-sgRNA-*BsmBI* plasmid that includes all elements needed for editing (Addgene #115694). To insert specific spacer sequence into the crRNA cassette, synthetic oligonucleotides were annealed to generate a duplex with overhangs compatible with those generated by *BsmBI* digestion of the pSimpleII-Cas9-sgRNA-*BsmBI* plasmid. The insert was then ligated into the *BsmBI*-digested plasmid. For Figs. 3 and 4, NmeCas9 (Addgene #87448) and SpyCas9 (Addgene #69220) constructs were expressed from the pCS2-Dest Gateway plasmid under the control of the CMV IE94 promoter [91]. All sgRNAs used with pCS2-Dest-Cas9 were driven by the U6 promoter in pLKO.1-puro (Addgene #52628 and #86195) [62]. The M427 GFP reporter plasmid [66] was used as described [65].

### Cell culture, transfection, and transduction

HEK293T were cultured in DMEM with 10% FBS and 1% Penicillin/Streptomycin (Gibco) in a 37 °C incubator with 5% CO<sub>2</sub>. K562 cells were grown in the same conditions but using IMDM. HFF cells were grown in the same conditions but using DMEM with Glutamax and 20% FBS without antibiotics. mESCs were grown in DMEM supplemented with 10% FBS, glutamine beta-ME and LIF. For transient transfection, we used early to mid-passage cells (passage number 4–18). Approximately  $1.5 \times 10^5$  cells were transfected with 150 ng Cas9-expressing plasmid, 150 ng sgRNA-expressing plasmid and 10 ng mCherry plasmid using Polyfect transfection reagent (Qiagen) in a 24-well plate according to the manufacturer's protocol. For the GFP reporter assay, 100 ng M427 plasmid was included in the co-transfection mix. Transduction was done as described previously [92].

### Western blotting

Forty-eight hours after transfection, cells were harvested and lysed with 50  $\mu$ l of RIPA buffer. Protein concentration was determined with the BCA kit (Thermo Scientific), and 12  $\mu$ g of proteins were used for electrophoresis and blotting. The blots were probed with anti-HA (Sigma, H3663) and anti-GAPDH (Abcam, ab9485) as primary antibodies, and then with horseradish peroxidase-conjugated anti-mouse IgG (Thermoscientific, 62-6520) or anti-rabbit IgG (Biorad, 1706515) secondary antibodies, respectively. Blots were visualized using the Clarity Western ECL substrate (Biorad, 170-5060).

### Flow cytometry

The GFP reporter was used as described previously [65]. Briefly, cells were harvested 48 h after transfection and used for FACS analysis (BD Accuri 6C). To minimize the effects of differences in the efficiency of transfection among samples, cells were initially gated for mCherry-expression, and the percentage of GFP-expressing cells were quantified within mCherry positive cells. All experiments were performed in triplicate with data reported as mean values with error bars indicating the standard error of the mean (s.e.m.).

### Genome editing

Seventy-two hours after transfection, genomic DNA was extracted via the DNeasy Blood and Tissue kit (Qiagen), according to the manufacturer's protocol. Fifty-nanogram DNA was used for PCR amplification using primers specific for each genomic site (Additional file 5: Table S9) with High Fidelity 2X PCR Master Mix (New England Biolabs). For T7E1 analysis, 10  $\mu$ l of PCR product was hybridized and treated with 0.5  $\mu$ l T7 Endonuclease I (10 U/ $\mu$ l, New England Biolabs) in 1X NEB Buffer 2 for 1 h. Samples were run on a 2.5% agarose gel, stained with SYBR-safe (ThermoFisher Scientific), and quantified using the ImageMaster-TotalLab program. Indel percentages are calculated as previously described [93, 94]. Experiments for T7E1 analysis are performed in triplicate with data reported as mean  $\pm$  s.e.m. For indel analysis by TIDE, 20 ng of PCR product is purified and then sequenced by Sanger sequencing. The trace files were subjected to analysis using the TIDE web tool (<https://tide.deskgen.com>).

### Expression and purification of NmeCas9

NmeCas9 was cloned into the pMCSG7 vector containing a T7 promoter followed by a 6XHis tag and a tobacco etch virus (TEV) protease cleavage site. Two NLSs on the C-terminus of NmeCas9 and another NLS on the N-terminus were also incorporated (Addgene #120078). This construct was transformed into the

Rosetta 2 DE3 strain of *E. coli*. Expression of NmeCas9 was performed as previously described for SpyCas9 [14]. Briefly, a bacterial culture was grown at 37 °C until an OD600 of 0.6 was reached. At this point, the temperature was lowered to 18 °C followed by addition of 1 mM Isopropyl  $\beta$ -D-1-thiogalactopyranoside (IPTG) to induce protein expression. Cells were grown overnight and then harvested for purification. Purification of NmeCas9 was performed in three steps: Nickel affinity chromatography, cation exchange chromatography, and size exclusion chromatography. The detailed protocols for these can be found in [14].

#### RNP delivery of NmeCas9

RNP delivery of NmeCas9 was performed using the Neon transfection system (ThermoFisher). Approximately 40 picomoles of NmeCas9 and 50 picomoles of sgRNA (transcribed from PCR-assembled templates in vitro with T7 RNA polymerase) were mixed in buffer R and incubated at room temperature for 30 min. This preassembled complex was then mixed with 50,000–150,000 cells, and electroporated using 10  $\mu$ L Neon tips. After electroporation, cells were plated in prewarmed 24-well plates containing the appropriate culture media without antibiotics. The number of cells used and pulse parameters of electroporation were different for different cell types tested. The number of cells used was 50,000, 100,000, and 150,000 for PLB985 cells, HEK293T cells, and K562/HFF cells respectively. Electroporation parameters (voltage, width, number of pulses) were 1150 v, 20 ms, 2 pulses for HEK293T cells; 1000 v, 50 ms, 1 pulse for K562 cells; 1350 v, 35 ms, 1 pulse for PLB985 cells; and 1700 v, 20 ms, 1 pulse for HFF cells.

#### $\gamma$ H2AX immunofluorescence staining and flow cytometry

For immunofluorescence, mouse embryonic stem cells (mESCs) were crosslinked with 4% paraformaldehyde and stained with anti- $\gamma$ H2AX (LP BIO, AR-0149-200) as primary antibody and Alexa Fluor<sup>®</sup> 488 goat anti-rabbit IgG (Invitrogen, A11034) as secondary antibody. DNA was stained with DAPI. For a positive control, E14 cells were irradiated with 254 nm UV light (3 mJ/cm<sup>2</sup>). Images were taken by a Nikon Eclipse E400 and representative examples were chosen.

For flow cytometry, cells were fixed with 70% ethanol, primary and secondary antibody were as described above for immunofluorescence, and DNA was stained with propidium iodide. Cells were analyzed by BD FACSCalibur. The box plot was presented with the bottom line of the box representing the first quartile, the band inside box indicating the median, the top line being the third quartile, the bottom end of whisker denoting data of first quartile minus 1.5 times of interquartile range (no less than 0),

and the top end of the whisker indicating data of third quartile plus 1.5 times of interquartile. Outliers are not shown. All experiments were performed in duplicate. For genotoxicity assay in HEK293T cells, we treated positive control cells with 5 mJ/cm<sup>2</sup> UV irradiation. We used the same protocol described above for antibody staining. Cells were analyzed by MACS-Quant flow cytometer. Two independent experiments were performed on different days.

#### CRISPRseek analysis of potential off-target sites

Global off-target analyses for DTS3, DTS7, and DTS8 with NmeCas9 sgRNAs were performed using the Bioconductor package CRISPRseek 1.9.1 [72] with parameter settings tailored for NmeCas9. Specifically, all parameters are set as default except the following: gRNA.size = 24, PAM = “NNNNGATT,” PAM.size = 8, RNA.PAM.pattern = “NNNNGNNN\$” (or “NNNNGNNN\$” for slipped PAM prediction), weights = c(0, 0, 0, 0, 0, 0, 0.014, 0, 0, 0.395, 0.317, 0, 0.389, 0.079, 0.445, 0.508, 0.613, 0.851, 0.732, 0.828, 0.615, 0.804, 0.685, 0.583), max.mismatch = 6, allowed.mismatch.PAM = 7, topN = 10,000, min.score = 0. This setting means that all seven permissive PAM sequences (N<sub>4</sub>GATT, N<sub>4</sub>GCTT, N<sub>4</sub>GTTT, N<sub>4</sub>GACA, N<sub>4</sub>GACT, N<sub>4</sub>GATA, N<sub>4</sub>GTCT) were allowed and all off-targets with up to 6 mismatches were collected [the sgRNA length was changed from 20 to 24; four additional zeros were added to the beginning of the weights series to be consistent with the gRNA length of 24; and topN (the number of off-target sites displayed) and min.score (the minimum score of an off-target to be included in the output) were modified to enable identification of all off-target sites with up to 6 mismatches]. Predicted off-target sites for DTS3, DTS7, and DTS8 with SpyCas9 sgRNAs were obtained using CRISPRseek 1.9.1 default settings for SpyCas9 (with NGG, NAG, and NGA PAMs allowed). Batch scripts for high-performance computing running the IBM LSF scheduling software are included in the supplemental section. Off-target sites were binned according to the number of mismatches relative to the on-target sequence. The numbers of off-targets for each sgRNA were counted and plotted.

#### GUIDE-seq

We performed GUIDE-seq experiment with some modifications to the original protocol [63], as described [65]. Briefly, in 24-well format, HEK293T cells were transfected with 150 ng of Cas9, 150 ng of sgRNA, and 7.5 pmol of annealed GUIDE-seq oligonucleotide using Polyfect transfection reagent (Qiagen) for all six guides (DTS3, DTS7, and DTS8 for both the NmeCas9 and SpyCas9 systems). Experiments with DTS7 sgRNAs were repeated using Lipofectamine 3000 transfection reagent (Invitrogen) according to the manufacturer's protocol. Forty-eight hours after

transfection, genomic DNA was extracted with a DNeasy Blood and Tissue kit (Qiagen) according to the manufacturer's protocol. Library preparation (Additional file 7: Table S6), sequencing, and read analyses were done according to protocols described previously [63, 65]. Only sites that harbored a sequence with up to six or ten mismatches with the target site (for SpyCas9 or NmeCas9, respectively) were considered potential off-target sites. Data were analyzed using the Bioconductor package GUIDEseq version 1.1.17 (Zhu et al., 2017). For SpyCas9, default setting was used except that `min.reads = 2`, `max.mismatch = 6`, `allowed.mismatch.PAM = 2`, `PAM.pattern = "NNN$"`, `BSgenomeName = Hsapiens`, `txdb = TxDb.Hsapiens.UCSC.hg19.knownGene`, `orgAnn = org.Hs.egSYMBOL`. For NmeCas9, default setting was used except that `PAM.size = 8`, `PAM = "NNNNGATT"`, `min.reads = 2`, `allowed.mismatch.PAM = 4`, `PAM.pattern = "NNNNNNNN$"`, `BSgenomeName = Hsapiens`, `txdb = TxDb.Hsapiens.UCSC.hg19.knownGene`, `orgAnn = org.Hs.egSYMBOL`. NmeCas9 dataset was analyzed twice with `max.mismatch = 6` and `max.mismatch = 10` respectively. The `gRNA.size` was set to the length of the gRNA used, and various number of 0's was added at the beginning of weights to make the length of weights equal to the gRNA size. For example, for gRNA with length 24, `weights = c(0,0,0,0,0, 0, 0.014, 0, 0, 0.395, 0.317, 0, 0.389, 0.079, 0.445, 0.508, 0.613, 0.851, 0.732, 0.828, 0.615, 0.804, 0.685, 0.583)` (Zhu et al., 2017). These regions are reported in Additional file 3: Table S2.

#### SITE-Seq

We performed the SITE-Seq assay as described previously [64]. In 50-mL conical tubes, high molecular weight genomic DNA (gDNA) was extracted from HEK293T cells using the Blood and Cell Culture DNA Maxi Kit (Qiagen) according to the manufacturer's protocol. sgRNAs for both NmeCas9 and SpyCas9 RNP assembly were transcribed from PCR-assembled DNA templates containing T7 promoters. Oligo sequences used in DNA template assembly can be found in Additional file 8: Table S8. PCR reactions were performed using Q5 Hot Start High-Fidelity 2X Master Mix (New England Biolabs) with the following thermal cycling conditions: 98 °C for 2 min, 30 cycles of 20 s at 98 °C, 20 s at 52 °C, 15 s at 72 °C, and a final extension at 72 °C for 2 min. sgRNAs were in vitro transcribed using the HiScribe T7 High Yield RNA Synthesis Kit (New England Biolabs) according to the manufacturer's protocol. Transcription reactions were digested with 2 units RNase-free DNase I (New England Biolabs) at 37 °C for 10 min; the reaction was stopped by adding EDTA to a final concentration of 35 mM and incubating at 75 °C for 10 min. All guides were purified with RNA-Clean beads (Beckman Coulter) and quantified with the Quant-IT Ribogreen RNA Assay kit (ThermoFisher) according to the manufacturers' protocols.

Individual RNPs were prepared by incubating each sgRNA at 95 °C for 2 min, then allowed to slowly come to room temperature over 5 min. Each sgRNA was then combined with its respective Cas9 in a 3:1 sgRNA:Cas9 molar ratio and incubated at 37 °C for 10 min in cleavage reaction buffer (20 mM HEPES, pH 7.4, 150 mM KCl, 10 mM MgCl<sub>2</sub>, 5% glycerol). In 96-well format, 10 µg of gDNA was treated with 0.2 pmol, 0.8 pmol, 3.2 pmol, and 12.8 pmol of each RNP in 50 µL total volume in cleavage reaction buffer, in triplicate. Negative control reactions were assembled in parallel and did not include any RNP. gDNA was treated with RNPs for 4 h at 37 °C. Library preparation and sequencing were done according to protocols described previously [64] using the Illumina NextSeq platform, and ~3 million reads were obtained for each sample. Any SITE-Seq sites without off-target motifs located within 1 nt of the cut-site were considered false-positives and discarded.

#### Targeted deep sequencing analysis

To measure indel frequencies, targeted deep sequencing analyses were done as previously described [65]. Briefly, we used two-step PCR amplification to produce DNA fragments for each on-target and off-target site. In the first step, we used locus-specific primers bearing universal overhangs with complementary ends to the TruSeq adaptor sequences (Additional file 9: Table S7). DNA was amplified with Phusion High Fidelity DNA Polymerase (New England Biolabs) using annealing temperatures of 60 °C, 64 °C or 68 °C, depending on the primer pair. In the second step, the purified PCR products were amplified with a universal forward primer and an indexed reverse primer to reconstitute the TruSeq adaptors (Additional file 9: Table S7). Input DNA was PCR-amplified with Phusion High Fidelity DNA Polymerase (98 °C, 15 s; 61 °C, 25 s; 72 °C, 18 s; 9 cycles), and equal amounts of the products from each treatment group were mixed and run on a 2.5% agarose gel. Full-size products (~250 bp in length) were gel-extracted. The purified library was deep sequenced using a paired-end 150 bp MiSeq run.

MiSeq data analysis was performed using a suite of Unix-based software tools. First, the quality of paired-end sequencing reads (R1 and R2 fastq files) was assessed using FastQC (<http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>). Raw paired-end reads were combined using paired end read merger (PEAR) [95] to generate single merged high-quality full-length reads. Reads were then filtered by quality [using Filter FASTQ [96]] to remove those with a mean PHRED quality score under 30 and a minimum per base score under 24. Each group of reads was then aligned to a corresponding reference sequence using BWA (version 0.7.5) and SAMtools (version 0.1.19). To determine indel frequency, size, and distribution, all edited reads from each experimental replicate were

combined and aligned, as described above. Indel types and frequencies were then cataloged in a text output format at each base using bam-readcount (<https://github.com/genome/bam-readcount>). For each treatment group, the average background indel frequencies (based on indel type, position, and frequency) of the triplicate negative control group were subtracted to obtain the nuclease-dependent indel frequencies. Indels at each base were marked, summarized, and plotted using GraphPad Prism. Deep sequencing data and the results of statistical tests are reported in Additional file 10: Table S3.

SITE-Seq cell-based validation was performed as previously described with minor modifications [64]. In brief, SITE-Seq sites were amplified from ~1000 to 4000 template copies per replicate and sequencing data from Cas9-treated samples were combined to minimize any variability due to uneven coverage across replicates. Cas9 cleavage sites were registered from the SITE-Seq data, and mutant reads were defined as any non-reference variant calls within 20 bp of the cut site. Sites with low sequencing coverage (<1000 reads in the combined, Cas9-treated samples or <200 reads in the reference samples) or >2% variant calls in the reference samples were discarded. Sites were tallied as cellular off-targets if they accumulated >0.5% mutant reads in the combined, Cas9-treated samples. This threshold corresponded to sites that showed unambiguous editing when DNA repair patterns were visually inspected.

## Additional files

**Additional file 1: Figure S1.** (A) Sequences of sgRNAs associated with their respective DNA targets used in Fig. 1b. (B) Schematic representation of the split-GFP reporter system used in Figs. 1b, 2a, and 6a. **Figure S2.** NmeCas9 does not lead to an increase in  $\gamma$ H2AX levels in mouse ESC or human HEK293T cells. **Figure S3.** SpyCas9 and NmeCas9 exhibit similar editing activity when targeting the *AAVS1* locus. **Figure S4.** NmeCas9 and SpyCas9 editing efficiencies at dual target sites (DTSs) that can be cleaved by both enzymes. **Figure S5.** SITE-Seq read counts and heat map of the frequencies of indels by length at DTS3, DTS7, and DTS8 for SpyCas9 and NmeCas9. **Figure S6.** Positions of insertions and deletions at the DTS3, DTS7, and DTS8 sites. **Figure S7.** Sizes of insertions and deletions at the DTS3, DTS7, and DTS8 sites. **Figure S8.** Fractions of all mutations that are insertions or deletions at the DTS3, DTS7, and DTS8 sites for SpyCas9 and NmeCas9. **Figure S9.** As in **Figure S6**. **Figure S10.** As in **Figure S7**. **Figure S11.** GUIDE-seq read counts at NmeCas9 on- and off-target sites. **Figure S12.** Guide truncation series for Protospacer 9, NTS33 and NTS32. **Figure S13.** GUIDE-seq analysis to determine whether NmeCas9 sgRNA truncation leads to de novo off-target events. **Figure S14.** NmeCas9 guide length requirements at the NTS1C and NTS1C-OT1 editing sites. (PDF 5232 kb)

**Additional file 2: Table S5.** Targeted deep sequencing read count data for all sgRNAs used and for all off-targets detected by SITE-Seq. (XLSX 144 kb)

**Additional file 3: Table S2.** Summary of potential off-target sites detected by GUIDE-seq analysis using the Bioconductor package GUIDE-seq. The table includes sites for both NmeCas9 (with  $N_4GN_3$  PAM sequences) and SpyCas9 (with NGG, NGA, and NAG PAM sequences) for the DTS3, DTS7 and DTS8 sgRNAs, as well as for NmeCas9 with the other sgRNAs used in this study. (XLSX 105 kb)

**Additional file 4: Table S1.** List of off-target sites with canonical and variant SpyCas9 and NmeCas9 PAMs predicted by CRISPRseek, for all guides used in this study. (XLSX 42214 kb)

**Additional file 5: Table S9.** Plasmids used in this study (including sequences of synthetic oligonucleotides that were used for sgRNA guide sequence cloning), and primer sequences used for analysis of editing at the corresponding genomic sites by T7E1 and TIDE analyses. (XLSX 30 kb)

**Additional file 6: Table S4.** Summary of potential off-target sites detected by SITE-seq analysis. The table includes sites for both NmeCas9 and SpyCas9 for the DTS3, DTS7 and DTS8 sgRNAs, as well as for other sgRNAs used in this study. (XLSX 3308 kb)

**Additional file 7: Table S6.** GUIDE-seq primer index sequences used in this study. (XLSX 45 kb)

**Additional file 8: Table S8.** Primer sequences used for on- and off-target deep sequencing analysis for all genomic editing sites analyzed by SITE-Seq. (XLSX 48 kb)

**Additional file 9: Table S7.** Primer sequences used for on- and off-target deep sequencing analysis for all genomic editing sites analyzed by GUIDE-seq. (XLSX 28 kb)

**Additional file 10: Table S3.** Targeted deep sequencing read count data for all sgRNAs used and for all off-targets detected by GUIDE-seq. (XLSX 44 kb)

## Abbreviations

AAV: Adeno-associated virus; BLESS: Breaks labelling, enrichment on streptavidin and next-generation sequencing; BLISS: Breaks labeling in situ and sequencing; bp: Base pair; Cas: CRISPR-associated; CdiCas9: *Corynebacterium diphtheria* Cas9; Circle-seq: Circularization of in vitro reporting of cleavage effects by sequencing; CjeCas9: *Campylobacter jejuni* Cas9; CMV: Cytomegalovirus; CRISPR: Clustered, regularly interspaced, short palindromic repeats; crRNAs: CRISPR RNAs; dCas9: "Dead" Cas9; Digenome-seq: Digested genome sequencing; DSB: Double-strand break; dsODN: Double-stranded oligodeoxynucleotide; DTS: Dual target site; EF1 $\alpha$ : Elongation factor-1 $\alpha$ ; GeoCas9: *Geobacillus stearothermophilus* Cas9; GUIDE-seq: Genome-wide unbiased identification of double strand breaks enabled by sequencing; HDR: Homology-directed repair; HTGTS: High-throughput genome-wide translocation sequencing; IDLV: Integrase-defective lentiviral vector; mESC: Mouse embryonic stem cell; NHEJ: Non-homologous end joining; NLS: Nuclear localization signal; NmeCas9: *Neisseria meningitidis* (strain 8013) Cas9; NTS: NmeCas9 target site; PAM: Protospacer adjacent motif; RNP: Ribonucleoprotein; SauCas9: *Staphylococcus aureus* Cas9; sgRNA: Single-guide RNA; SITE-Seq: Selective enrichment and identification of tagged genomic DNA ends by sequencing; SpyCas9: *Streptococcus pyogenes* Cas9; T7E1: T7 Endonuclease 1; tracrRNA: *Trans*-acting CRISPR RNA; tru-sgRNAs: Truncated sgRNAs

## Acknowledgements

We thank Yin Guo for technical assistance, Wen Xue and members of the Sontheimer and Wolfe laboratories for insightful comments and discussions, and Phil Zamore for the use of his flow cytometer. We are also grateful to the UMMS Deep Sequencing and Molecular Biology Core Laboratories for providing outstanding technical support services for this research project.

## Funding

This work was supported by funds from Intellia Therapeutics to E.J.S., and by NIH grant R01 GM115911 to E.J.S. and S.A.W.

## Availability of data and materials

The deep sequencing data from this study have been submitted to the NCBI Sequence Read Archive (SRA; <http://www.ncbi.nlm.nih.gov/sra>) under accession number SRP150215 [97]. Plasmids are available via Addgene.

## Authors' contributions

NA, XDG, PL, AE, AM, RI, AG, KES, and TW carried out genome-editing experiments. TW, XDG, and TGF carried out genotoxicity experiments and analyses. NA, XDG, PL, AM, AG, LJZ, and SAW carried out GUIDE-seq experiments and analysis. PDD, AHS, AML, KM, KKF, and PC carried out SITE-Seq experiments and

analyses using NmeCas9 protein supplied by AM. XDG, PL, AM, LJZ, and SAW carried out additional bioinformatic and statistical analyses. All authors analyzed and interpreted data. NA, XDG, PC, and EJS wrote the manuscript, and all authors edited the manuscript. All authors read and approved the final manuscript.

#### Ethics approval and consent to participate

Not applicable.

#### Consent for publication

Not applicable.

#### Competing interests

EJS is a co-founder and scientific advisor of Intellia Therapeutics. P.D.D, A.H.S, A.M.L, K.M, C.K.F, and P.C are current or former employees of Caribou Biosciences, Inc., a company that develops and commercializes genome engineering technologies, and such individuals may own shares or stock options in Caribou Biosciences. SITE-Seq is a trademark of Caribou Biosciences, Inc. The other authors declare that they have no competing interests.

#### Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

#### Author details

<sup>1</sup>RNA Therapeutics Institute, University of Massachusetts Medical School, 368 Plantation Street, Worcester, MA 01605, USA. <sup>2</sup>Program in Molecular Medicine, University of Massachusetts Medical School, 368 Plantation Street, Worcester, MA 01605, USA. <sup>3</sup>Department of Molecular, Cell and Cancer Biology, University of Massachusetts Medical School, 368 Plantation Street, Worcester, MA 01605, USA. <sup>4</sup>Department of Biochemistry and Molecular Pharmacology, University of Massachusetts Medical School, 368 Plantation Street, Worcester, MA 01605, USA. <sup>5</sup>Program in Bioinformatics and Integrative Biology, University of Massachusetts Medical School, 368 Plantation Street, Worcester, MA 01605, USA. <sup>6</sup>Caribou Biosciences, Inc., 2929 7th Street, Suite 105, Berkeley, CA 94710, USA. <sup>7</sup>Present Address: Sangamo Therapeutics, Inc., Richmond, CA, USA. <sup>8</sup>Present Address: Bluebird bio, Cambridge, MA, USA. <sup>9</sup>Present Address: Molecular Pathology Unit, Massachusetts General Hospital, Charlestown, MA, USA. <sup>10</sup>Present Address: Memorial Sloan Kettering Cancer Center, New York, NY, USA.

Received: 9 May 2018 Accepted: 17 November 2018

Published online: 05 December 2018

#### References

- Marraffini LA. CRISPR-Cas immunity in prokaryotes. *Nature*. 2015;526:55–61.
- Mohanraju P, Makarova KS, Zetsche B, Zhang F, Koonin EV, van der Oost J. Diverse evolutionary roots and mechanistic variations of the CRISPR-Cas systems. *Science*. 2016;353:aad5147.
- Sontheimer EJ, Barrangou R. The bacterial origins of the CRISPR genome-editing revolution. *Hum Gene Ther*. 2015;26:413–24.
- Barrangou R, Fremaux C, Deveau H, Richards M, Boyaval P, Moineau S, Romero DA, Horvath P. CRISPR provides acquired resistance against viruses in prokaryotes. *Science*. 2007;315:1709–12.
- Brouns SJ, Jore MM, Lundgren M, Westra ER, Slijkhuis RJ, Snijders AP, Dickman MJ, Makarova KS, Koonin EV, van der Oost J. Small CRISPR RNAs guide antiviral defense in prokaryotes. *Science*. 2008;321:960–4.
- Marraffini LA, Sontheimer EJ. CRISPR interference limits horizontal gene transfer in staphylococci by targeting DNA. *Science*. 2008;322:1843–5.
- Makarova KS, Wolf YI, Alkhnbashi OS, Costa F, Shah SA, Saunders SJ, Barrangou R, Brouns SJ, Charpentier E, Haft DH, et al. An updated evolutionary classification of CRISPR-Cas systems. *Nat Rev Microbiol*. 2015;13:722–36.
- Deltcheva E, Chylinski K, Sharma CM, Gonzales K, Chao Y, Pirzada ZA, Eckert MR, Vogel J, Charpentier E. CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III. *Nature*. 2011;471:602–7.
- Garneau JE, Dupuis ME, Villion M, Romero DA, Barrangou R, Boyaval P, Fremaux C, Horvath P, Magadan AH, Moineau S. The CRISPR/Cas bacterial immune system cleaves bacteriophage and plasmid DNA. *Nature*. 2010;468:67–71.
- Sapranaukas R, Gasiunas G, Fremaux C, Barrangou R, Horvath P, Siksnys V. The *Streptococcus thermophilus* CRISPR/Cas system provides immunity in *Escherichia coli*. *Nucleic Acids Res*. 2011;39:9275–82.
- Deveau H, Barrangou R, Garneau JE, Labonte J, Fremaux C, Boyaval P, Romero DA, Horvath P, Moineau S. Phage response to CRISPR-encoded resistance in *Streptococcus thermophilus*. *J Bacteriol*. 2008;190:1390–400.
- Mojica FJ, Diez-Villasenor C, Garcia-Martinez J, Almendros C. Short motif sequences determine the targets of the prokaryotic CRISPR defence system. *Microbiol*. 2009;155:733–40.
- Gasiunas G, Barrangou R, Horvath P, Siksnys V. Cas9-crRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive immunity in bacteria. *Proc Natl Acad Sci U S A*. 2012;109:E2579–86.
- Jinek M, Chylinski K, Fonfara I, Hauer M, Doudna JA, Charpentier E. A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science*. 2012;337:816–21.
- Jiang W, Bikard D, Cox D, Zhang F, Marraffini LA. RNA-guided editing of bacterial genomes using CRISPR-Cas systems. *Nat Biotechnol*. 2013;31:233–9.
- Cho SW, Kim S, Kim JM, Kim JS. Targeted genome engineering in human cells with the Cas9 RNA-guided endonuclease. *Nat Biotechnol*. 2013;31:230–2.
- Cong L, Ran FA, Cox D, Lin S, Barretto R, Habib N, Hsu PD, Wu X, Jiang W, Marraffini LA, Zhang F. Multiplex genome engineering using CRISPR/Cas systems. *Science*. 2013;339:819–23.
- Hwang WY, Fu Y, Reyon D, Maeder ML, Tsai SQ, Sander JD, Peterson RT, Yeh JR, Joung JK. Efficient genome editing in zebrafish using a CRISPR-Cas system. *Nat Biotechnol*. 2013;31:227–9.
- Jinek M, East A, Cheng A, Lin S, Ma E, Doudna J. RNA-programmed genome editing in human cells. *eLife*. 2013;2:e00471.
- Mali P, Yang L, Esvelt KM, Aach J, Guell M, DiCarlo JE, Norville JE, Church GM. RNA-guided human genome engineering via Cas9. *Science*. 2013;339:823–6.
- Urnov FD. Genome Editing B.C. (before CRISPR): lasting lessons from the "old testament". *CRISPR J*. 2018;1:34–46.
- Hsu PD, Lander ES, Zhang F. Development and applications of CRISPR-Cas9 for genome engineering. *Cell*. 2014;157:1262–78.
- Komor AC, Badran AH, Liu DR. CRISPR-based technologies for the manipulation of eukaryotic genomes. *Cell*. 2017;168:20–36.
- Sternberg SH, Doudna JA. Expanding the biologist's toolkit with CRISPR-Cas9. *Mol Cell*. 2015;58:568–74.
- Dominguez AA, Lim WA, Qi LS. Beyond editing: repurposing CRISPR-Cas9 for precision genome regulation and interrogation. *Nat Rev Mol Cell Biol*. 2016;17:5–15.
- Wang H, La Russa M, Qi LS. CRISPR/Cas9 in genome editing and beyond. *Annu Rev Biochem*. 2016;85:227–64.
- Zetsche B, Gootenberg JS, Abudayyeh OO, Slaymaker IM, Makarova KS, Essletzbichler P, Volz SE, Joung J, van der Oost J, Regev A, et al. Cpf1 is a single RNA-guided endonuclease of a class 2 CRISPR-Cas system. *Cell*. 2015;163:759–71.
- Shmakov S, Smargon A, Scott D, Cox D, Pyzocha N, Yan W, Abudayyeh OO, Gootenberg JS, Makarova KS, Wolf YI, et al. Diversity and evolution of class 2 CRISPR-Cas systems. *Nat Rev Microbiol*. 2017;15:169–82.
- Fu Y, Foden JA, Khayter C, Maeder ML, Reyon D, Joung JK, Sander JD. High-frequency off-target mutagenesis induced by CRISPR-Cas nucleases in human cells. *Nat Biotechnol*. 2013;31:822–6.
- Hsu PD, Scott DA, Weinstein JA, Ran FA, Konermann S, Agarwala V, Li Y, Fine EJ, Wu X, Shalem O, et al. DNA targeting specificity of RNA-guided Cas9 nucleases. *Nat Biotechnol*. 2013;31:827–32.
- Bolukbasi MF, Gupta A, Wolfe SA. Creating and evaluating accurate CRISPR-Cas9 scalpels for genomic surgery. *Nat Methods*. 2015;13:41–50.
- Casini A, Olivieri M, Petris G, Montagna C, Reginato G, Maule G, Lorenzin F, Prandi D, Romanel A, Demichelis F, et al. A highly specific SpCas9 variant is identified by *in vivo* screening in yeast. *Nat Biotechnol*. 2018;36:265–71.
- Chen JS, Dagdas YS, Kleinstiver BP, Welch MM, Sousa AA, Harrington LB, Sternberg SH, Joung JK, Yildiz A, Doudna JA. Enhanced proofreading governs CRISPR-Cas9 targeting accuracy. *Nature*. 2017;550(7676):407–10.
- Tsai SQ, Joung JK. Defining and improving the genome-wide specificities of CRISPR-Cas9 nucleases. *Nat Rev Genet*. 2016;17:300–12.
- Tycko J, Myer VE, Hsu PD. Methods for optimizing CRISPR-Cas9 genome editing specificity. *Mol Cell*. 2016;63:355–70.
- Yin H, Song CQ, Suresh S, Kwan SY, Wu Q, Walsh S, Ding J, Bogorad RL, Zhu LJ, Wolfe SA, et al. Partial DNA-guided Cas9 enables genome editing with reduced off-target activity. *Nat Chem Biol*. 2018;14:311–6.

37. Kleinstiver BP, Pattanayak V, Prew MS, Tsai SQ, Nguyen NT, Zheng Z, Joung JK. High-fidelity CRISPR-Cas9 nucleases with no detectable genome-wide off-target effects. *Nature*. 2016;529:490–5.
38. Slaymaker IM, Gao L, Zetsche B, Scott DA, Yan WX, Zhang F. Rationally engineered Cas9 nucleases with improved specificity. *Science*. 2016;351:84–8.
39. Jiang F, Taylor DW, Chen JS, Kornfeld JE, Zhou K, Thompson AJ, Nogales E, Doudna JA. Structures of a CRISPR-Cas9 R-loop complex primed for DNA cleavage. *Science*. 2016;351:867–71.
40. Jiang F, Zhou K, Ma L, Gressel S, Doudna JA. A Cas9-guide RNA complex preorganized for target DNA recognition. *Science*. 2015;348:1477–81.
41. Jinek M, Jiang F, Taylor DW, Sternberg SH, Kaya E, Ma E, Anders C, Hauer M, Zhou K, Lin S, et al. Structures of Cas9 endonucleases reveal RNA-mediated conformational activation. *Science*. 2014;343:1247997.
42. Nishimasu H, Ran FA, Hsu PD, Konermann S, Shehata SI, Dohmae N, Ishitani R, Zhang F, Nureki O. Crystal structure of Cas9 in complex with guide RNA and target DNA. *Cell*. 2014;156:935–49.
43. Chylinski K, Makarova KS, Charpentier E, Koonin EV. Classification and evolution of type II CRISPR-Cas systems. *Nucleic Acids Res*. 2014;42:6091–105.
44. Fonfara I, Le Rhun A, Chylinski K, Makarova KS, Lecrivain AL, Bzdrenga J, Koonin EV, Charpentier E. Phylogeny of Cas9 determines functional exchangeability of dual-RNA and Cas9 among orthologous type II CRISPR-Cas systems. *Nucleic Acids Res*. 2014;42:2577–90.
45. Briner AE, Donohoue PD, Gomas AA, Selle K, Slorach EM, Nye CH, Haurwitz RE, Beisel CL, May AP, Barrangou R. Guide RNA functional modules direct Cas9 activity and orthogonality. *Mol Cell*. 2014;56:333–9.
46. Esvelt KM, Mali P, Braff JL, Moosburner M, Young SJ, Church GM. Orthogonal Cas9 proteins for RNA-guided gene regulation and editing. *Nat Methods*. 2013;10:1116–21.
47. Kim E, Koo T, Park SW, Kim D, Kim K, Cho HY, Song DW, Lee KJ, Jung MH, Kim S, et al. In vivo genome editing with a small Cas9 orthologue derived from *Campylobacter jejuni*. *Nat Commun*. 2017;8:14500.
48. Ran FA, Cong L, Yan WX, Scott DA, Gootenberg JS, Kriz AJ, Zetsche B, Shalem O, Wu X, Makarova KS, et al. In vivo genome editing using *Staphylococcus aureus* Cas9. *Nature*. 2015;520:186–91.
49. Mir A, Edraki A, Lee J, Sontheimer EJ. Type II-CRISPR-Cas9 biology, mechanism and application. *ACS Chem Biol*. 2018;13(2):357–65.
50. Zhang Y. The CRISPR-Cas9 system in *Neisseria* spp. *Pathog Dis*. 2017;75. <https://doi.org/10.1093/femspd/ftx036>.
51. Zhang Y, Heidrich N, Ampattu BJ, Gunderson CW, Seifert HS, Schoen C, Vogel J, Sontheimer EJ. Processing-independent CRISPR RNAs limit natural transformation in *Neisseria meningitidis*. *Mol Cell*. 2013;50:488–503.
52. Zhang Y, Rajan R, Seifert HS, Mondragón A, Sontheimer EJ. DNase H activity of *Neisseria meningitidis* Cas9. *Mol Cell*. 2015;60:242–55.
53. Hou Z, Zhang Y, Propson NE, Howden SE, Chu LF, Sontheimer EJ, Thomson JA. Efficient genome engineering in human pluripotent stem cells using Cas9 from *Neisseria meningitidis*. *Proc Natl Acad Sci U S A*. 2013;110:15644–9.
54. Lee CM, Cradick TJ, Bao G. The *Neisseria meningitidis* CRISPR-Cas9 system enables specific genome editing in mammalian cells. *Mol Ther*. 2016;24:645–54.
55. Ma E, Harrington LB, O'Connell MR, Zhou K, Doudna JA. Single-stranded DNA cleavage by divergent CRISPR-Cas9 enzymes. *Mol Cell*. 2015;60:398–407.
56. Rousseau BA, Hou Z, Gramelspacher MJ, Zhang Y. Programmable RNA cleavage and recognition by a natural CRISPR-Cas9 system from *Neisseria meningitidis*. *Mol Cell*. 2018;69(5):906–14.
57. Harrington LB, Doxzen KW, Ma E, Liu JJ, Knott GJ, Edraki A, Garcia B, Amrani N, Chen JS, Cofsky JC, et al. A broad-spectrum inhibitor of CRISPR-Cas9. *Cell*. 2017;170:1224–33 e1215.
58. Pawluk A, Amrani N, Zhang Y, Garcia B, Hidalgo-Reyes Y, Lee J, Edraki A, Shah M, Sontheimer EJ, Maxwell KL, Davidson AR. Naturally occurring off-switches for CRISPR-Cas9. *Cell*. 2016;167(7):1829–38.
59. Sontheimer EJ, Davidson AR. Inhibition of CRISPR-Cas systems by mobile genetic elements. *Curr Opin Microbiol*. 2017;37:120–7.
60. Rauch BJ, Silvis MR, Hultquist JF, Waters CS, McGregor MJ, Krogan NJ, Bondy-Denomy J. Inhibition of CRISPR-Cas9 with bacteriophage proteins. *Cell*. 2017;168:150–8 e110.
61. Hynes AP, Rousseau GM, Lemay M-L, Horvath P, Romero DA, Fremaux C, Moineau S. An anti-CRISPR from a virulent streptococcal phage inhibits *Streptococcus pyogenes* Cas9. *Nat Microbiol*. 2017;2:1374–80.
62. Kearns NA, Pham H, Tabak B, Genga RM, Silverstein NJ, Garber M, Maehr R. Functional annotation of native enhancers with a Cas9-histone demethylase fusion. *Nat Methods*. 2015;12:401–3.
63. Tsai SQ, Zheng Z, Nguyen NT, Liebers M, Topkar W, Thapar V, Wyvekens N, Khayter C, Iafrate AJ, Le LP, et al. GUIDE-seq enables genome-wide profiling of off-target cleavage by CRISPR-Cas nucleases. *Nat Biotechnol*. 2014;33:187–97.
64. Cameron P, Fuller CK, Donohoue PD, Jones BN, Thompson MS, Carter MM, Gradia S, Vidal B, Garner E, Slorach EM, et al. Mapping the genomic landscape of CRISPR-Cas9 cleavage. *Nat Methods*. 2017;14:600–6.
65. Bolukbasi MF, Gupta A, Oikemus S, Derr AG, Garber M, Brodsky MH, Zhu LJ, Wolfe SA. DNA-binding-domain fusions enhance the targeting range and precision of Cas9. *Nat Methods*. 2015;12:1150–6.
66. Wilson KA, McEwen AE, Pruett-Miller SM, Zhang J, Kildebeck EJ, Porteus MH. Expanding the repertoire of target sites for zinc finger nuclease-mediated genome modification. *Mol Ther Nucleic Acids*. 2013;2:e88.
67. Kim HJ, Lee HJ, Kim H, Cho SW, Kim JS. Targeted genome editing in human cells with zinc finger nucleases constructed via modular assembly. *Genome Res*. 2009;19:1279–88.
68. Santos-Pereira JM, Aguilera A. R loops: new modulators of genome dynamics and function. *Nat Rev Genet*. 2015;16:583–97.
69. Ciccía A, Elledge SJ. The DNA damage response: making it safe to play with knives. *Mol Cell*. 2010;40:179–204.
70. Chen PB, Chen HV, Acharya D, Rando OJ, Fazio TG. R loops regulate promoter-proximal chromatin architecture and cellular differentiation. *Nat Struct Mol Biol*. 2015;22:999–1007.
71. Aouida M, Eid A, Ali Z, Cradick T, Lee C, Deshmukh H, Atef A, AbuSamra D, Gadhouni SZ, Merzaban J, et al. Efficient fdCas9 synthetic endonuclease with improved specificity for precise genome engineering. *PLoS One*. 2015;10:e0133373.
72. Zhu LJ, Holmes BR, Aronin N, Brodsky MH. CRISPRseek: a bioconductor package to identify target-specific guide RNAs for CRISPR-Cas9 genome-editing systems. *PLoS One*. 2014;9:e108424.
73. Tsai SQ, Nguyen NT, Malagon-Lopez J, Topkar W, Aryee MJ, Joung JK. CIRCLE-seq: a highly sensitive in vitro screen for genome-wide CRISPR-Cas9 nuclease off-targets. *Nat Methods*. 2017;14:607–14.
74. Yan WX, Mirzazadeh R, Garnerone S, Scott D, Schneider MW, Kallas T, Custodio J, Wernersson E, Li Y, Gao L, et al. BLISS is a versatile and quantitative method for genome-wide profiling of DNA double-strand breaks. *Nat Commun*. 2017;8:15058.
75. Zhu LJ, Lawrence M, Gupta A, Pagés H, Cucukural A, Garber M, Wolfe SA. GUIDEseq: a bioconductor package to analyze GUIDE-Seq datasets for CRISPR-Cas nucleases. *BMC Genomics*. 2017;18:379.
76. Cho SW, Kim S, Kim Y, Kweon J, Kim HS, Bae S, Kim JS. Analysis of off-target effects of CRISPR/Cas-derived RNA-guided endonucleases and nickases. *Genome Res*. 2014;24:132–41.
77. Fu Y, Sander JD, Reyon D, Cascio VM, Joung JK. Improving CRISPR-Cas nuclease specificity using truncated guide RNAs. *Nat Biotechnol*. 2014;32:279–84.
78. Hwang WY, Fu Y, Reyon D, Maeder ML, Kaini P, Sander JD, Joung JK, Peterson RT, Yeh JR. Heritable and precise zebrafish genome editing using a CRISPR-Cas system. *PLoS One*. 2013;8:e68708.
79. Ran FA, Hsu PD, Lin CY, Gootenberg JS, Konermann S, Trevino AE, Scott DA, Inoue A, Matoba S, Zhang Y, Zhang F. Double nicking by RNA-guided CRISPR Cas9 for enhanced genome editing specificity. *Cell*. 2013;154:1380–9.
80. Chen F, Ding X, Feng Y, Seebeck T, Jiang Y, Davis GD. Targeted activation of diverse CRISPR-Cas systems for mammalian genome editing via proximal CRISPR targeting. *Nat Commun*. 2017;8:14958.
81. Dong, Guo M, Wang S, Zhu Y, Wang S, Xiong Z, Yang J, Xu Z, Huang Z. Structural basis of CRISPR-SpyCas9 inhibition by an anti-CRISPR protein. *Nature*. 2017;546:436–9.
82. Shin CS, Jiang F, Liu J-J, Bray NL, Rauch BJ, Baik SH, Nogales E, Bondy-Denomy J, Corn JE, Doudna JA. Disabling Cas9 by an anti-CRISPR DNA mimic. *Sci Adv*. 2017; in press.
83. Yang H, Patel DJ. Inhibition mechanism of an anti-CRISPR suppressor AcrIIA4 targeting SpyCas9. *Mol Cell*. 2017;67(1):117–27.
84. Fu BX, Hansen LL, Artilles KL, Nonet ML, Fire AZ. Landscape of target:guide homology effects on Cas9-mediated cleavage. *Nucleic Acids Res*. 2014;42:13778–87.
85. Pattanayak V, Lin S, Guillinger JP, Ma E, Doudna JA, Liu DR. High-throughput profiling of off-target DNA cleavage reveals RNA-programmed Cas9 nuclease specificity. *Nat Biotechnol*. 2013;31:839–43.
86. Kim S, Kim D, Cho SW, Kim J, Kim JS. Highly efficient RNA-guided genome editing in human cells via delivery of purified Cas9 ribonucleoproteins. *Genome Res*. 2014;24:1012–9.

87. Zuris JA, Thompson DB, Shu Y, Guilinger JP, Bessen JL, Hu JH, Maeder ML, Joung JK, Chen Z-Y, Liu DR. Cationic lipid-mediated delivery of proteins enables efficient protein-based genome editing *in vitro* and *in vivo*. *Nat Biotechnol.* 2015;33:73–80.
88. Ibraheim R, Song CQ, Mir A, Amrani N, Xue W, Sontheimer EJ. All-in-one adeno-associated virus delivery and genome editing by *Neisseria meningitidis* Cas9 *in vivo*. *Genome Biol.* 2018;19:137.
89. Bisaria N, Jarmoskaite I, Herschlag D. Lessons from enzyme kinetics reveal specificity principles for RNA-guided nucleases in RNA interference and CRISPR-based genome editing. *Cell Syst.* 2017;4:21–9.
90. Wang Y, Liu KJ, Sutrisnoh NB, Srinivasan H, Zhang J, Li J, Zhang F, Lalith CRJ, Xing H, Shanmugam R, et al. Systematic evaluation of CRISPR-Cas systems reveals design principles for genome editing in human cells. *Genome Biol.* 2018;19:62.
91. Villefranc JA, Amigo J, Lawson ND. Gateway compatible vectors for analysis of gene function in the zebrafish. *Dev Dyn.* 2007;236:3077–87.
92. Mou H, Smith JL, Peng L, Yin H, Moore J, Zhang XO, Song CQ, Sheel A, Wu Q, Ozata DM, et al. CRISPR/Cas9-mediated genome editing induces exon skipping by alternative splicing or exon deletion. *Genome Biol.* 2017;18:108.
93. Gupta A, Hall VL, Kok FO, Shin M, McNulty JC, Lawson ND, Wolfe SA. Targeted chromosomal deletions and inversions in zebrafish. *Genome Res.* 2013;23:1008–17.
94. Guschin DY, Waite AJ, Katibah GE, Miller JC, Holmes MC, Rebar EJ. A rapid and general assay for monitoring endogenous gene modification. *Methods Mol Biol.* 2010;649:247–56.
95. Zhang J, Kobert K, Flouri T, Stamatakis A. PEAR: a fast and accurate Illumina paired-end reAd mergeR. *Bioinformatics.* 2014;30:614–20.
96. Blankenberg D, Gordon A, Von Kuster G, Coraor N, Taylor J, Nekrutenko A, Galaxy T. Manipulation of FASTQ data with galaxy. *Bioinformatics.* 2010;26:1783–5.
97. Amrani N, Gao XD, Liu P, Edraki A, Mir A, Ibraheim R, et al. NmeCas9 is an intrinsically high-fidelity genome editing platform. *Sequence Read Archive (SRA) [Database];* 2018.

**Ready to submit your research? Choose BMC and benefit from:**

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

**At BMC, research is always in progress.**

Learn more [biomedcentral.com/submissions](https://biomedcentral.com/submissions)

