NmeCas9 is an intrinsically high-fidelity genome editing platform

Nadia Amrani[1,*], Xin D. Gao[1,*], Pengpeng Liu[3,4], Ankit Gupta[3,#], Alireza Edraki[1], Raed Ibraheim[1], Kanae E. Sasaki[1†], Lihua Julie Zhu[2,3,5], Scot A. Wolfe[3,4], and Erik J. Sontheimer[1,2,Δ]

[1]RNA Therapeutics Institute
[2]Program in Molecular Medicine
[3]Department of Molecular, Cell and Cancer Biology
[4]Department of Biochemistry and Molecular Pharmacology
[5]Program in Bioinformatics and Integrative Biology

University of Massachusetts Medical School
Worcester, MA 01605
U.S.A.

* These authors contributed equally to this manuscript
# Present address: bluebird bio, Cambridge, Massachusetts, USA
† Present address: Molecular Pathology Unit, Massachusetts General Hospital, Charlestown, Massachusetts, USA

Δ Correspondence: erik.sontheimer@umassmed.edu

Running title: Intrinsically precise genome editing by NmeCas9

Key Words: Cas9; CRISPR; sgRNA; *Neisseria meningitidis*

ABSTRACT

The development of CRISPR-Cas9 RNA-guided 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, 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 efforts to achieve single-site accuracy within large (e.g. mammalian) genomes. In addition to its off-targeting propensity, SpyCas9 is encoded by a relatively large (~4.2 kb) open reading frame (ORF), limiting its utility in applications that require size-restricted delivery strategies such as adeno-associated virus (AAV) vectors. In contrast, some genome-editing-validated Cas9 orthologs [e.g. *Staphylococcus aureus* Cas9 (SauCas9), *Campylobacter jejuni* Cas9 (CjeCas9), and *Neisseria meningitidis* Cas9 (NmeCas9)] are considerably smaller and therefore better suited for viral delivery. Here we show that wild-type NmeCas9, when programmed with guide sequences of natural length (24 nucleotides), exhibits a nearly complete absence of unintended targeting in human cells, even when targeting sites that are highly prone to off-target activity when employing SpyCas9. We also validate at least six variant protospacer adjacent motifs (PAMs), in addition to the preferred consensus PAM (5'-N_4GATT-3'), for NmeCas9 genome editing in human cells. 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 efforts.

INTRODUCTION

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 (Marraffini 2015; Sontheimer and Barrangou 2015; Mohanraju et al. 2016). 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 (Barrangou et al. 2007; Brouns et al. 2008) and other mobile genetic elements (Marraffini and Sontheimer 2008). 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 (Makarova et al. 2015). 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 tracrRNA (Garneau et al. 2010; Deltcheva et al. 2011; Sapranauskas et al. 2011). Successful interference also requires that the DNA target (the "protospacer") be highly complementary to the spacer portion of the crRNA, and that the target also matches a PAM consensus at neighboring base pairs (Deveau et al. 2008; Mojica et al. 2009).

Following the discovery that Type II interference occurs via double-strand breaks (DSBs) in the DNA target (Garneau et al. 2010), 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 (Sapranauskas et al. 2011). 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 (Gasiunas

et al. 2012; Jinek et al. 2012). *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) (Jinek et al. 2012). 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 (Jiang et al. 2013) and eukaryotic (Cho et al. 2013; Cong et al. 2013; Hwang et al. 2013b; Jinek et al. 2013; Mali et al. 2013). As with earlier DSB-inducing systems, 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 (Hsu et al. 2014; Sternberg and Doudna 2015; Komor et al. 2017). 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 (Dominguez et al. 2016; Wang et al. 2016). Similar RNA-guided tools for genome manipulations have since been developed from Type V CRISPR-Cas systems that use the Cas12a (formerly Cpf1) enzyme (Zetsche et al. 2015).

Type II CRISPR-Cas systems are currently grouped into three subtypes (II-A, II-B and II-C) (Makarova et al. 2015; Shmakov et al. 2017). 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 (Deltcheva et al. 2011; Jinek et al. 2012)]. 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 (Deltcheva et al. 2011; Jinek et al. 2012). 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 (Fu et al. 2013; Hsu et al. 2013), prompting the development of numerous approaches to limit editing activity at unwanted sites (Bolukbasi et al. 2015b; Tsai and Joung 2016; Tycko et al. 2016). Although these strategies have been shown to minimize off-

118 targeting to various degrees, they do not always abolish it, and they can also reduce on-target activity, at
119 least with some sgRNAs. Furthermore, each of these approaches has required extensive testing, validation,
120 and optimization, and in some cases (Kleinstiver et al. 2016; Slaymaker et al. 2016) depended heavily
121 upon prior high-resolution structural characterization (Jinek et al. 2014; Nishimasu et al. 2014; Jiang et al.
122 2015; Jiang et al. 2016).

123 Thousands of other Cas9 orthologs have been documented (Chylinski et al. 2014; Fonfara et al.
124 2014; Makarova et al. 2015; Shmakov et al. 2017), providing tremendous untapped potential for
125 additional genome editing capabilities beyond those offered by SpyCas9. Many Cas9 orthologs will
126 provide distinct PAM specificities, increasing the number of targetable sites in any given genome. Many
127 pair-wise Cas9 combinations also have orthogonal guides that load into one ortholog but not the other,
128 facilitating multiplexed applications (Esvelt et al. 2013; Briner et al. 2014; Fonfara et al. 2014). Finally,
129 some Cas9 orthologs (especially those from subtype II-C) are hundreds of amino acids smaller than the
130 1,368 amino acid SpyCas9 (Chylinski et al. 2014; Fonfara et al. 2014; Makarova et al. 2015), and are
131 therefore more amenable to combined Cas9/sgRNA delivery via a single size-restricted vector such as
132 AAV (Ran et al. 2015; Kim et al. 2017). Finally, there may be native Cas9 orthologs that exhibit
133 additional advantages such as greater efficiency, hyper-accuracy, distinct activities, reduced
134 immunogenicity, or novel means of control over editing. Deeper exploration of the Cas9 population could
135 therefore enable expanded or improved genome engineering capabilities.

136 We have used *N. meningitidis* (strain 8013) as a model system for the interference functions and
137 mechanisms of Type II-C CRISPR-Cas systems (Zhang et al. 2013; Zhang et al. 2015). In addition, we
138 and others previously reported that the Type II-C Cas9 ortholog from *N. meningitidis* (NmeCas9) can be
139 applied as a genome engineering platform (Esvelt et al. 2013; Hou et al. 2013; Lee et al. 2016). At 1,082
140 amino acids, NmeCas9 is 286 residues smaller than SpyCas9, making it nearly as compact as SauCas9
141 (1,053 amino acids) and well within range of all-in-one AAV delivery. Its spacer-derived guide sequences
142 are longer (24 nts) than those of most other Cas9 orthologs (Zhang et al. 2013), and like SpyCas9, it
143 cleaves both DNA strands between the third and fourth nts of the protospacer (counting from the PAM-

144 proximal end). NmeCas9 also has a longer PAM consensus (5'-N_4GATT-3', after the 3' end of the
145 protospacer's crRNA-noncomplementary strand) (Esvelt et al. 2013; Hou et al. 2013; Zhang et al. 2013;
146 Fonfara et al. 2014; Zhang et al. 2015; Lee et al. 2016), leading to a lower density of targetable sites
147 compared to SpyCas9. Considerable variation from this consensus is permitted during bacterial
148 interference (Esvelt et al. 2013; Zhang et al. 2015), and a smaller number of variant PAMs can also
149 support targeting in mammalian cells (Hou et al. 2013; Lee et al. 2016). Recently, natural Cas9 inhibitors
150 (encoded by bacterial mobile elements) have been identified and validated in *N. meningitidis* and other
151 bacteria with type II-C systems, providing for genetically encodable off-switches for NmeCas9 genome
152 editing (Pawluk et al. 2016). These "anti-CRISPR" (Acr) proteins enable temporal, spatial, or conditional
153 control over the NmeCas9 system. Natural inhibitors of Type II-A systems have also been discovered in
154 *Listeria monocytogenes*, some of which are effective at inhibiting SpyCas9 (Rauch et al. 2017).
155 The longer PAM consensus and longer guide sequence for NmeCas9 could result in a reduced
156 propensity for off-targeting, and targeted deep sequencing at bioinformatically predicted near-cognate
157 sites is consistent with this possibility (Lee et al. 2016). A high degree of genome-wide specificity has also
158 been noted for the dNmeCas9 platform (Kearns et al. 2015a). However, the true, unbiased accuracy of
159 NmeCas9 is not known, since empirical assessments of genome-wide off-target editing activity
160 (independent of bioinformatics prediction) have not been reported for this ortholog. Here we define and
161 confirm many of the parameters of NmeCas9 editing activity in mammalian cells including PAM
162 sequence preferences, guide length limitations, and off-target profiles. Most notably, we use an empirical
163 approach (GUIDE-seq) (Tsai et al. 2014) to define NmeCas9 off-target profiles and find that wild-type
164 NmeCas9 is a high-fidelity genome editing platform in mammalian cells, with far lower levels of off-
165 targeting than SpyCas9. These results further validate NmeCas9 as a genome engineering platform, and
166 suggest that continued exploration of Cas9 orthologs could identify additional RNA-guided nucleases that
167 exhibit favorable properties, even without the extensive engineering efforts that have been applied to
168 SpyCas9 (Bolukbasi et al. 2015b; Tsai and Joung 2016; Tycko et al. 2016).
169

RESULTS

Co-expressed sgRNA increases NmeCas9 accumulation in mammalian cells

Previously we demonstrated that NmeCas9 [derived from *N. meningitidis* strain 8013 (Zhang et al. 2013)] can efficiently target chromosomal loci in human stem cells using either dual RNAs (crRNA + tracrRNA) or an sgRNA (Hou et al. 2013). 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 an sgRNA in a single transfection vector, similar to our previous all-in-one dual-RNA plasmid (pSimple-Cas9-Tracr-crRNA; Addgene #47868) (Hou et al. 2013). The pEJS15 plasmid expresses NmeCas9 fused to a C-terminal single-HA epitope tag and NLS sequences at both N- and C-termini, under the control of the elongation factor-1α (EF1α) promoter. The sgRNA cassette (driven by the U6 promoter) includes two *BsmB*I 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) (Zhang et al. 2013; Zhang et al. 2015) to express sgRNAs that target protospacer (ps) 9, ps24 or ps25, respectively (Supplemental Fig. 1A). 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 N-TS3 (Table 1). Two of the resulting all-in-one plasmids (spacer9/sgRNA and N-TS3/sgRNA), as well as a plasmid lacking the sgRNA cassette, were transiently transfected into HEK293T cells for 48h, 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 CMV promoter) (Bolukbasi et al. 2015a) (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). Recently, a different Type II-C Cas9 (CdiCas9) was shown to be dramatically stabilized by its cognate sgRNA when subjected to proteolysis *in vitro* (Ma et al. 2015); if similar resistance to proteolysis occurs with NmeCas9 upon sgRNA binding, it could explain some or all

Site	Gene or locus	Spacer Sequence of sgRNA	Target site, with PAM
NTS1B	*SLC9A9*	GGGCAUCAUGAUUUUGAACUCCCU	CCTTGGCATCATGATTTTGAACTCCCTATGTGATTCTA
NTS1C	*SLC9A9*	GUGGUCUGGGGUACAGCCUUGGCA	TACTUGGTCTGGGGTACAGCCTTGGCATCATGATTTTG
NTS1C-OT1	*PHKG2*	GCGGUGUGAGGUACAGCCUUGGCA	TAATCGGTGTGAGGTACAGCCTTGGCATCAGGATTTCT
NTS3	*AL158836*	GAUGCUCAGAAAGAGGAAGCUGGU	GGGGATGCTCAGAAAGAGGAAGCTGGTTTATGATTGGA
NTS4B	*FLJ00328*	GGACAGGAGUCGCCAGAGGCCGGU	GCAGGACAGGAGTCGCCAGAGGCCGGTGGTGGATTTCC
NTS4C	*FLJ00328*	GGGGCUGGCUCCACGUCGCGCCGC	TGCGGGGCTGGCTCCACGTCGCGCCGCGGCGGTTTGGG
NTS5	*AF064860*	GAAACAGACUCGCAAGACUUCAGA	GACAAAACAGACTCGCAAGACTTCAGATACAGATTCCA
NTS7	*LOC100505797*	GAGGGAGAGAGGUGAGCGGAUGAA	GCAAAGGGAGAGAGGTGAGCGGATGAAGGGAGATTGGT
NTS8	*ESPN*	GGACGCAAUUCCAGAGGUGAUGGG	CGGCGACGCAATTCCAGAGGTGATGGGGAGTGATTGTC
NTS9	*ZNHIT2*	GGCGCUGUGUUUUCGCAAAGCUUC	CGGCGCGCTGTGTTTTCGCAAAGCTTCCGAGGATTCTC
NTS10	*HHLA1*	GCAGCCAAGUUUGAGAACUGCUGU	TGTGCAGCCAAGTTTGAGAACTGCTGTTACAGATTTCC
NTS11	*SMARCB1*	GUUCCAGUUGGGAAGGGCCAGUGC	TAGATTCCAGTTGGGAAGGGCCAGTGCCTCCGATTCCA
NTS21	*TNNC1*	GCCAGAGCUGCCGCCAGACAGUGA	CAGTCCAGAGCTGCCGCCAGACAGTGATGCTGTCTTGG
NTS25	*AC193513*	GGUUUCUCAUCCUGUCUUCUGCCU	CCGCGTTTCTCATCCTGTCTTCTGCCTAGTGGATATGT
NTS26	*LOC105378512*	GUUCAAAAGUAGCGGGCGCUAGGC	GTACTTCAAAAGTAGCGGGCGCTAGGCGGGTGTTTCTG
NTS27	*TIE1*	GUUCUCCAAGCCCUCGGACCUCGU	CGGCTTCTCCAAGCCCTCGGACCTCGTGGGCGTCTTCT
NTS30	*NEK8*	GGGGCUCCGGAGCCCACCCCAGGA	CGCGGGGCTCCGGAGCCCACCCCAGGACCAGGACTTAG
NTS31	*POC1A*	GUGGGAAGUGUAGCUCCACCUUCC	ATGTTGGGAAGTGTAGCTCCACCTTCCTCCAGACATAG
NTS32	*VEGFA*	GCCCCGGCUCUGGCUAAAGAGGGA	CACACCCCGGCTCTGGCTAAAGAGGGAATGGGCTTTGG
NTS33	*VEGFA*	GCGGGGAGAAGGCCAGGGGUCACU	GGAGCGGGGAGAAGGCCAGGGGTCACTCCAGGATTCCA
NTS55	*CYBB*	GCUGGAUUACUGUGUGGUAGAGGG	CTAGCTGGATTACTGTGTGGTAGAGGGAGGTGATTAGC
NTS58	*AAVS1*	GUUUGCCUGGACACCCCGUUCUCC	TTTGTTTGCCTGGACACCCCGTTCTCCTGTGGATTCCG
NTS59	*AAVS1*	GACCCCACAGUGGGGCCACUAGGG	CTCCACCCCACAGTGGGGCCACTAGGGACAGGATTGGT
STS60	*AAVS1*	GUUAAUGUGGCUCUGGUUCU	CCGGTTAATGTGGCTCTGGTTCTGGGTAC
STS61	*AAVS1*	GUCCCCUCCACCCCACAGUG	TCTGTCCCCTCCACCCCACAGTGGGGCCA
STS62	*AAVS1*	GGGGCCACUAGGGACAGGAU	AGTGGGGCCACTAGGGACAGGATTGGTGA
NTS63	*AAVS1*	GAGUUAGAACUCAGGACCAACUUA	CCAAAGTTAGAACTCAGGACCAACTTATTCTGATTTTG

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.

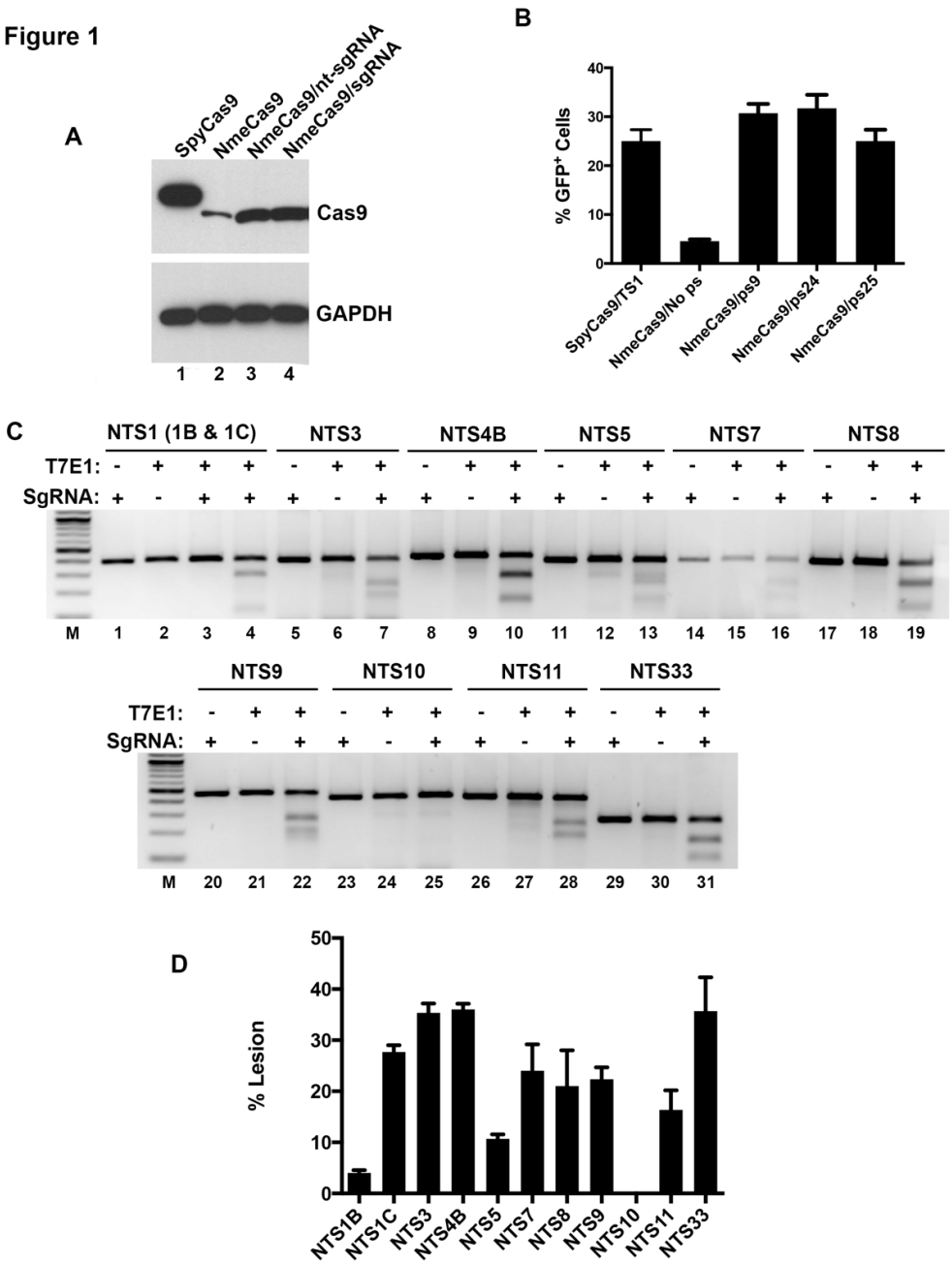


Figure 1. NmeCas9 expression and activity in human (HEK293T) cells. (*A*) Western blot detection of HA-tagged NmeCas9 in transiently transfected HEK293T cells. Lane 1: Cells transfected with SpyCas9 plasmid under the control of the CMV promoter. Lane 2: Cells transfected with NmeCas9 plasmid under the control of the elongation factor-1α (EF1α) 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 an 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+ cells. (*C*) NmeCas9 programmed independently with different sgRNAs targeting eleven genomic sites flanked by an N_4GATT PAM, detected by T7E1 analysis. (*D*) Quantitation of lesion efficiencies from three independent biological replicates performed on different days. Error bars indicate ± standard error of the mean (± s.e.m.).

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 (Wilson et al. 2013) into which we can insert target protospacers for NmeCas9. Cleavage promotes a single-strand-annealing-based repair pathway that generates an intact GFP ORF, leading to fluorescence (Wilson et al. 2013) that can be scored after 48 hours by flow cytometry. We generated reporters carrying three validated bacterial protospacers (ps9, ps24 and ps25, as described above) (Zhang et al. 2013; Zhang et al. 2015) 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 targeted 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 eleven human chromosomal sites with an N_4GATT PAM (Table 1). These sgRNAs induced 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 true frequencies of indel formation (Guan et al. 2004). These data confirm that NmeCas9 can induce, with variable efficiency, DNA lesions at many potential genomic target sites in human cells.

Functionality of truncated sgRNAs with NmeCas9

SpyCas9 can accommodate limited variation in the length of the guide region (normally 20 nt) of its sgRNAs (Hwang et al. 2013a; Ran et al. 2013; Cho et al. 2014; Fu et al. 2014b), 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 (Cho et

al. 2014; Fu et al. 2014b). To test the length dependence of the NmeCas9 guide sequence [normally 24 nt; (Zhang et al. 2013)] during mammalian editing, we constructed a series of sgRNAs containing 18, 19, 20, 21, 22, 23, and 24 nts of complementarity to ps9, cloned into the split-GFP reporter plasmid (Supplemental Fig. 1B). All designed guides started with two guanine nts (resulting in 1-2 positions of target noncomplementarity at the very 5' end of the guide) to facilitate transcription and to test the effects of extra 5'-terminal G residues, analogous to the SpyCas9 "GGN_{20}" sgRNAs (Cho et al., 2014). We then measured the abilities of these sgRNAs to direct NmeCas9 cleavage of the reporter in human cells. SgRNAs that have 20 to 23 nts of target complementarity showed activities comparable to the sgRNA with the natural 24 nts of complementarity, whereas sgRNAs containing 18 and 19 nts of complementarity show lower activity (Fig. 2A).

We next used a native chromosomal target site (NTS33 in *VEGFA*, as in Figs. 1C and 1D) to test the editing efficiency of NmeCas9 spacers of varying lengths (Supplemental Fig. 1C). SgRNA constructs included one or two 5'-terminal guanine residues to enable transcription by the U6 promoter, sometimes resulting in 1-2 nts of target non-complementarity at the 5' end of the guide sequence. SgRNAs with 20, 21, and 22 nts of target complementarity (GGN_{18}, GGN_{19}, and GGN_{20}, respectively) performed comparably to the natural guide length (24 nts of complementarity, GN_{23}) at this site (Fig. 2B), 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. 2A). SgRNAs with guide lengths of 19 nts or shorter, along with a single mismatch in the first or second position (GGN_{17}, GGN_{16}, and GGN_{15}), did not direct detectable editing, nor did an sgRNA with perfectly matched guide sequences of 17 or 14 nts (GN_{16} and GN_{13}, respectively) (Fig. 2B). However, a 19-nt guide with no mismatches (GN_{18}) successfully directed editing, albeit with slightly reduced efficiency. 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 nts down to 17-18 nts 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 Figs. 2A and 2B are both associated with a canonical N_4GATT PAM,

Figure 2

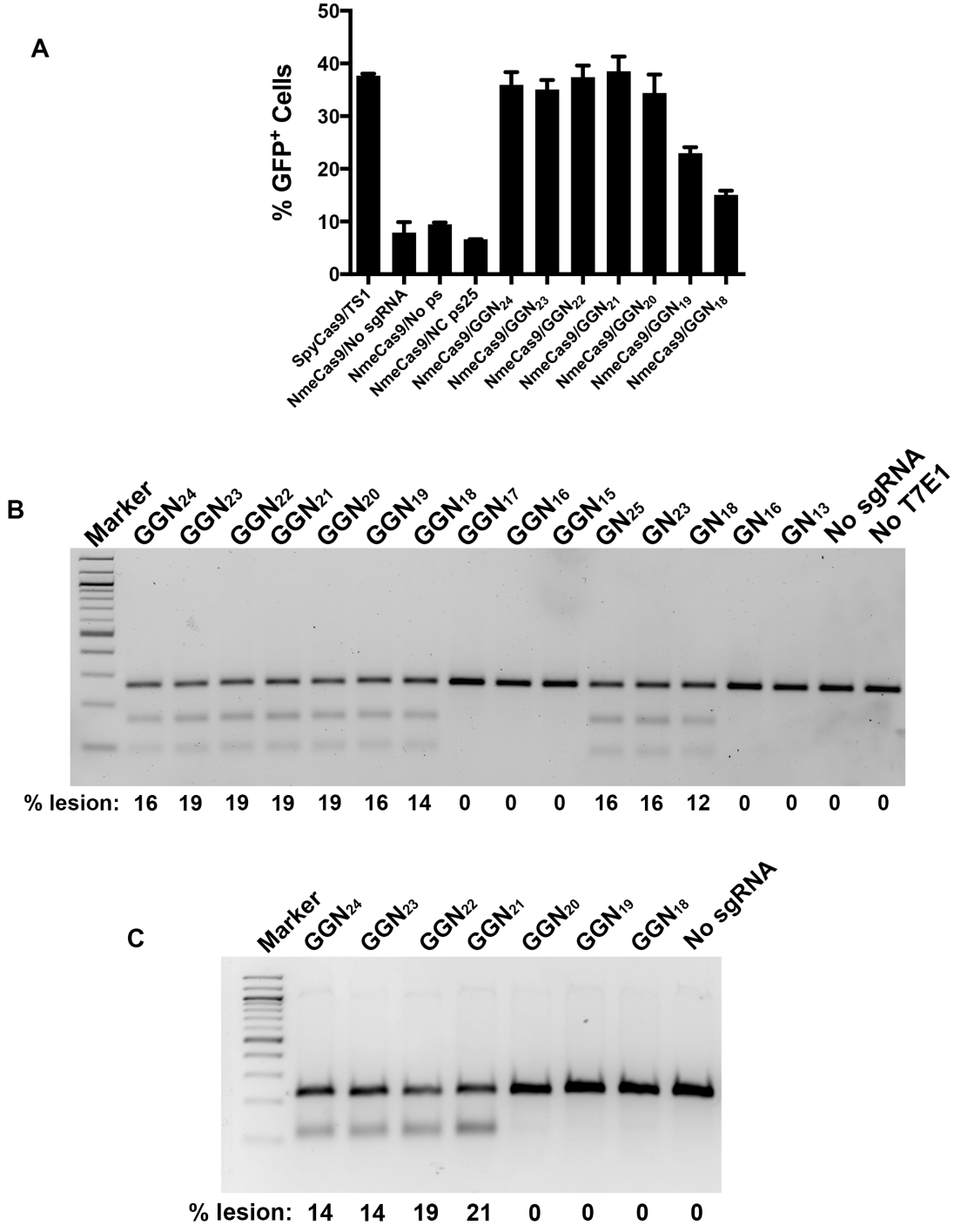


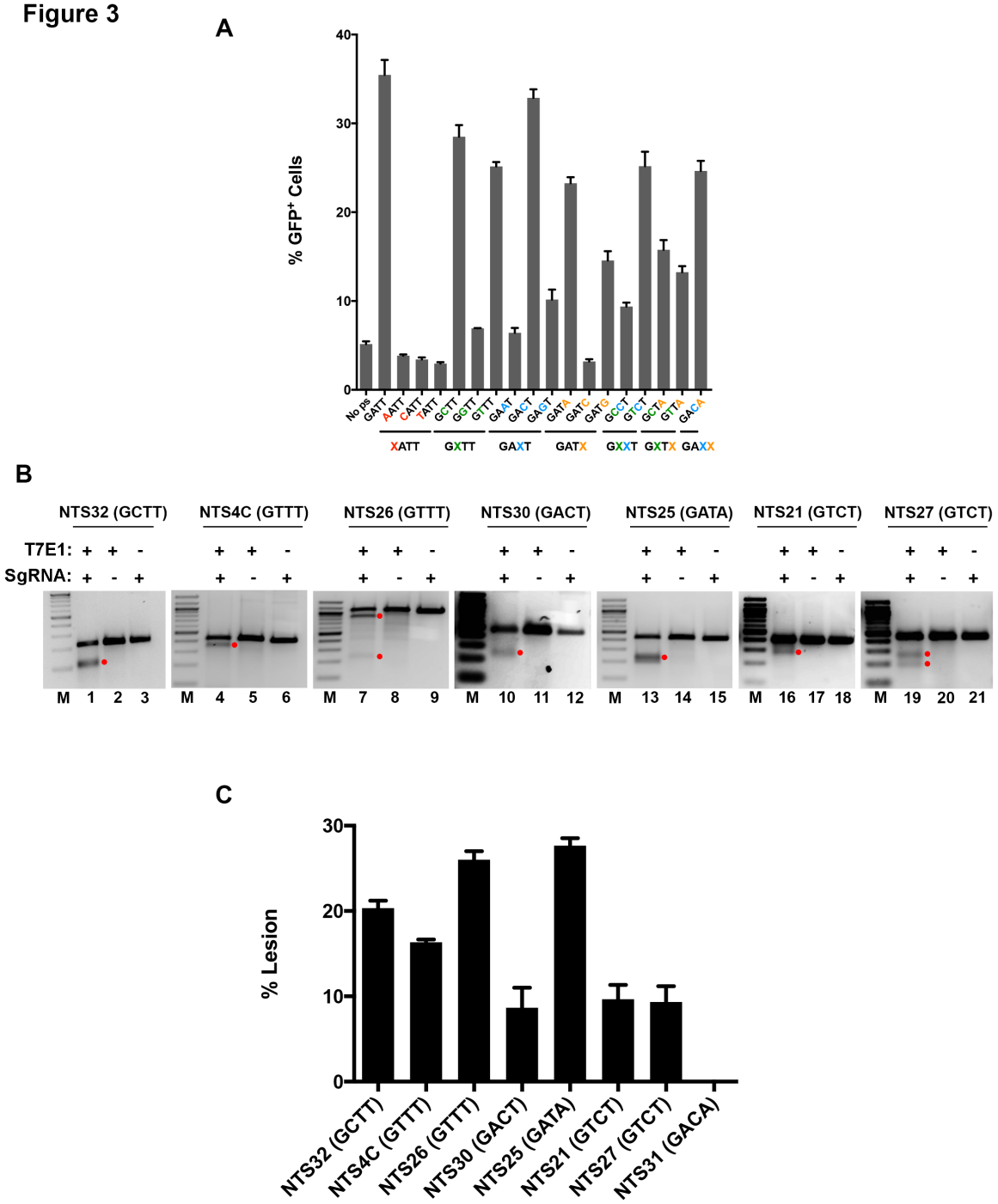
Figure 2. NmeCas9 guide length requirements in mammalian cells. (*A*) Split-GFP activity profile of NmeCas9 cleavage with ps9 sgRNAs bearing spacers of varying lengths (18-24 nts) along with 5'-terminal G residues to enable transcription. Bars represent mean values ± s.e.m. from three independent biological replicates performed on different days. (*B*) T7EI analysis of editing efficiencies at the NTS33 genomic target site (with an N_4GATT PAM) with sgRNAs bearing spacers of varying lengths (13-25 nts) with 1-2 5'-terminal G residues. (*C*) As in (*B*), but targeting the NTS32 genomic site (with an N_4GCTT PAM).

but efficient NmeCas9 editing at mammalian chromosomal sites associated with N_4GCTT (Hou et al. 2013) and other variant PAMs [(Lee et al. 2016); also see below] has also been reported. To examine length dependence at a site with a variant PAM, we varied guide sequence length at the N_4GCTT-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 (Supplemental Fig. 1D). At the NTS32 site, sgRNAs with 21-24 nts of complementarity (GGN_{24}, GGN_{23}, GGN_{22}, and GGN_{21}) supported editing, but shorter guides (GGN_{20}, GGN_{19}, and GGN_{18}) did not (Fig. 2C). We conclude that sgRNAs with 20 nts of complementarity can direct editing at some sites (Fig. 2B) but not all (Fig. 2C). 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_4GATT 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 nts appear to be functional in most cases, in agreement with the results of others (Lee et al. 2016).

PAM specificity of NmeCas9 in human cells

During native CRISPR interference in bacterial cells, considerable variation in the N_4GATT PAM consensus is tolerated: although the G1 residue (N_4<u>G</u>ATT) is strictly required, virtually all other single mutations at A2 (N_4G<u>A</u>TT), T3 (N_4GA<u>T</u>T), and T4 (N_4GAT<u>T</u>) retain at least partial function in licensing bacterial interference (Esvelt et al. 2013; Zhang et al. 2015). In contrast, fewer NmeCas9 PAM variants have been validated during genome editing in mammalian cells (Hou et al. 2013; Lee et al. 2016). 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 described above. We introduced single-nt 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 293T cells. The results are shown in Fig. 3A. As expected, mutation of the G1 residue to any other base reduced editing to

Figure 3



298 **Figure 3.** Characterization of functional PAM sequences in human (HEK293T) cells. (*A*) Split-GFP activity profile
299 of NmeCas9 cleavage with ps9 sgRNA, with the target site flanked by different PAM sequences. Bars represent
300 mean values ± s.e.m. from three independent biological replicates performed on different days. (*B*) T7E1 analysis of
301 editing efficiencies at seven genomic sites flanked by PAM variants, as indicated. Products resulting from NmeCas9
302 genome editing are denoted by the red dots. (*C*) Quantitation of data from (*B*), as well as an additional site (NTS31;
303 N_4GACA PAM) that was not successfully edited. Bars represent mean values ± s.e.m. from three independent
304 biological replicates performed on different days.

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_4GCTT, N_4GTTT, N_4GACT, and N_4GATA) and two double mutants (N_4GTCT and N_4GACA) were edited with efficiencies approaching that observed with the N_4GATT PAM. Two other single mutants (N_4GAGT and N_4GATG), and three double mutants (N_4GCCT, N_4GCTA, and N_4GTTA) 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_4GAAT and N_4GATC) in this mammalian assay fit the functional consensus sequences defined previously in *E. coli* (Esvelt et al. 2013).

We then used T7E1 analysis to validate genome editing at eight native chromosomal sites associated with the most active PAM variants (N_4GCTT, N_4GTTT, N_4GACT, N_4GATA, N_4GTCT, and N_4GACA). Our results with this set of targets indicate that all of these PAM variants tested except N_4GACA support chromosomal editing (Fig. 3B and C).

Comparative analysis of NmeCas9 and SpyCas9

SpyCas9 is by far the best-characterized Cas9 orthologue, 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 orthologues 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 (Bolukbasi et al. 2015a; Pawluk et al. 2016). 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 (Esvelt et al. 2013; Hou et al. 2013), 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 48h, and the expression of the identically tagged proteins was monitored by anti-HA western blot (Fig. 4A). 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. 4A, compare lane 6 to 4 and 5), whereas SpyCas9 is not affected by the presence of its sgRNA (lanes 1-3). Apo NmeCas9 (without sgRNA) accumulates less strongly than SpyCas9, but more strongly than SpyCas9 in the presence of cognate sgRNA.

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 (Mali et al. 2013; Aouida et al. 2015) and used the CRISPRseek package (Zhu et al. 2014) to design three NmeCas9 guides targeting the same locus within a region of ~700 bp (Supplemental Fig. 2A). 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. 4B).

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_4GATT, respectively). Because the optimal SpyCas9 and NmeCas9 PAMs are non-overlapping, it is simple to identify chromosomal target sites that are compatible with both orthologues, 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 (N**N/N**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

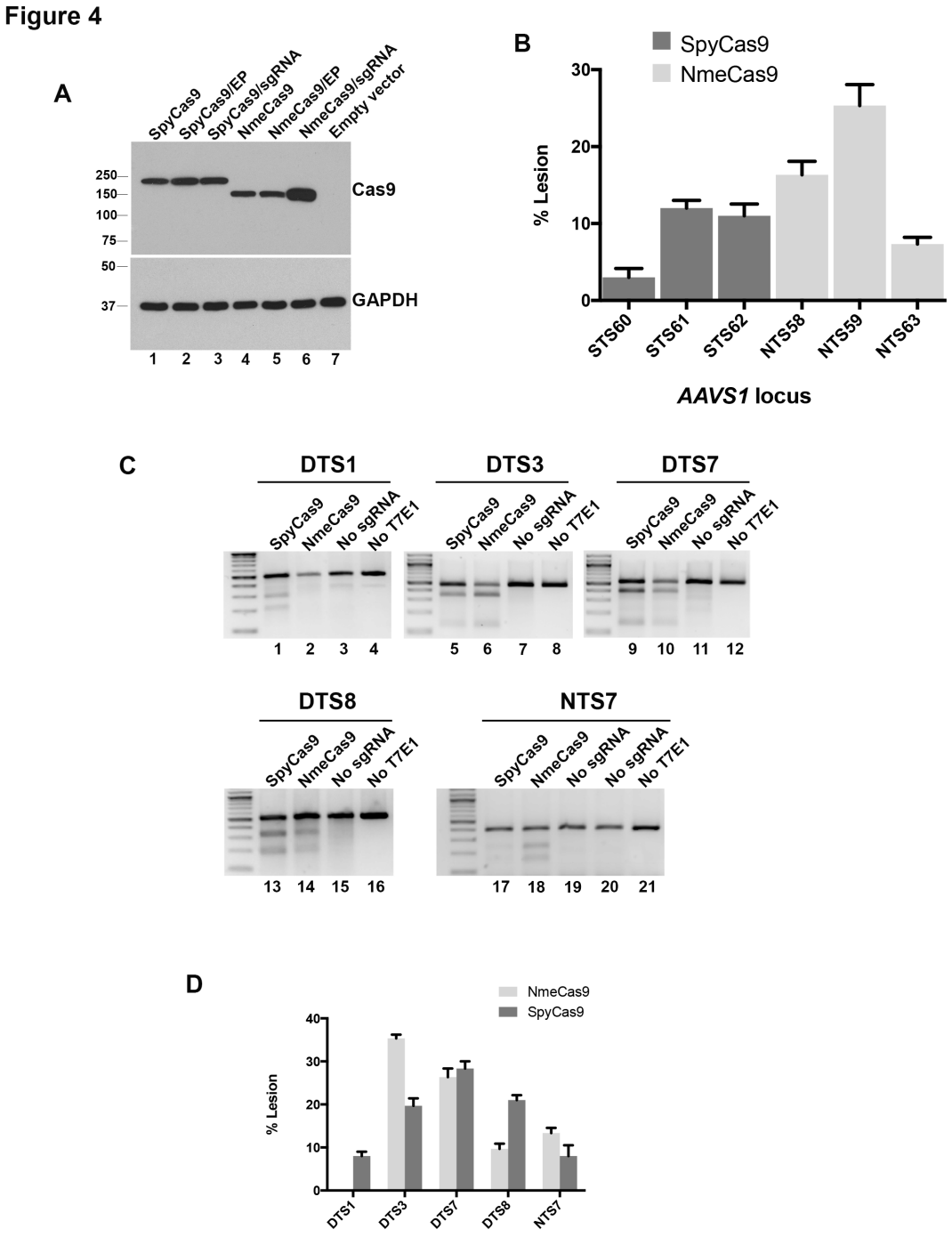


Figure 4. NmeCas9 and SpyCas9 have comparable editing efficiencies in human (HEK293T) cells when targeting the same chromosomal sites. (*A*) Western blot analysis of NmeCas9 and SpyCas9. HEK293T cells were transfected with the indicated Cas9 ortholog cloned in the same plasmid backbone, and fused to 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 ± s.e.m., $n = 3$). (*C*) Representative T7EI analyses comparing editing efficiencies at the dual target sites DTS1, DTS3, DTS7, DTS8, and NTS7, using the indicated Cas9/sgRNA combinations. (*D*) Quantitation of data from (*C*) (mean ± s.e.m., $n = 3$).

366 at position -24 of the sgRNA-noncomplementary strand (relative to the PAM) and another at position -20,
367 then the U6 promoter can be used to express perfectly-matched sgRNAs for both Cas9 orthologues. Four
368 DTSs with these characteristics were used in this comparison (Supplemental Fig. 3A). We had previously
369 used NmeCas9 to target a site (NTS7) that happened also to match the SpyCas9 PAM consensus, so we
370 included it in our comparative analysis as a fifth site, even though it has a predicted rG-dT wobble pair at
371 position -24 for the NmeCas9 sgRNA (Supplemental Fig. 3A).

372 We set out next to compare the editing activities of both Cas9 orthologs programmed to target the
373 five chromosomal sites depicted in Supplemental Fig. 3A, initially via T7E1 digestion. SpyCas9 was more
374 efficient than NmeCas9 at generating lesions at the DTS1 and DTS8 sites (Fig. 4C, lanes 1-2 and 13-14).
375 In contrast, NmeCas9 was more efficient than SpyCas9 at the DTS3 and NTS7 sites (Fig. 4C, lanes 5-6
376 and 17-18). Editing at DTS7 was approximately equal with both orthologs (Fig. 4C, lanes 9-10). Data
377 from three biological replicates of all five target sites are plotted in Fig. 4D. The remainder of our
378 comparative studies focused on DTS3, DTS7, and DTS8, as they provided examples of target sites with
379 NmeCas9 editing efficiencies that are greater than, equal to, or lower than those of SpyCas9, respectively.
380 At all three of these sites, the addition of an extra 5'-terminal G residue had little to no effect on editing by
381 either SpyCas9 or NmeCas9 (Supplemental Fig. 3B). Truncation of the three NmeCas9 guides down to
382 20 nt (all perfectly matched) again had differential effects on editing efficiency from one site to the next,
383 with no reduction in DTS7 editing, partial reduction in DTS3 editing, and complete loss of DTS8 editing
384 (Supplemental Fig. 3B).

385

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

387 All Cas9 orthologs described to date have some propensity to edit off-target sites lacking perfect
388 complementarity to the programmed guide RNA, and considerable effort has been devoted to developing
389 strategies (mostly with SpyCas9) to increase editing specificity (reviewed in (Bolukbasi et al. 2015b; Tsai
390 and Joung 2016; Tycko et al. 2016)). In comparison with SpyCas9, orthologs such as NmeCas9 that
391 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 (Zhu et al. 2014) 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. 5A) as representative queries. When allowing for permissive and semi-permissive PAMs [NGG, NGA, and NAG for SpyCas9; N_4GHTT, N_4GACT, N_4GAYA, and N_4GTCT for NmeCas9], potential off-target sites for NmeCas9 were predicted with two to three orders of magnitude lower frequency than for SpyCas9 (Table 2). Furthermore, NmeCas9 off-target sites with fewer

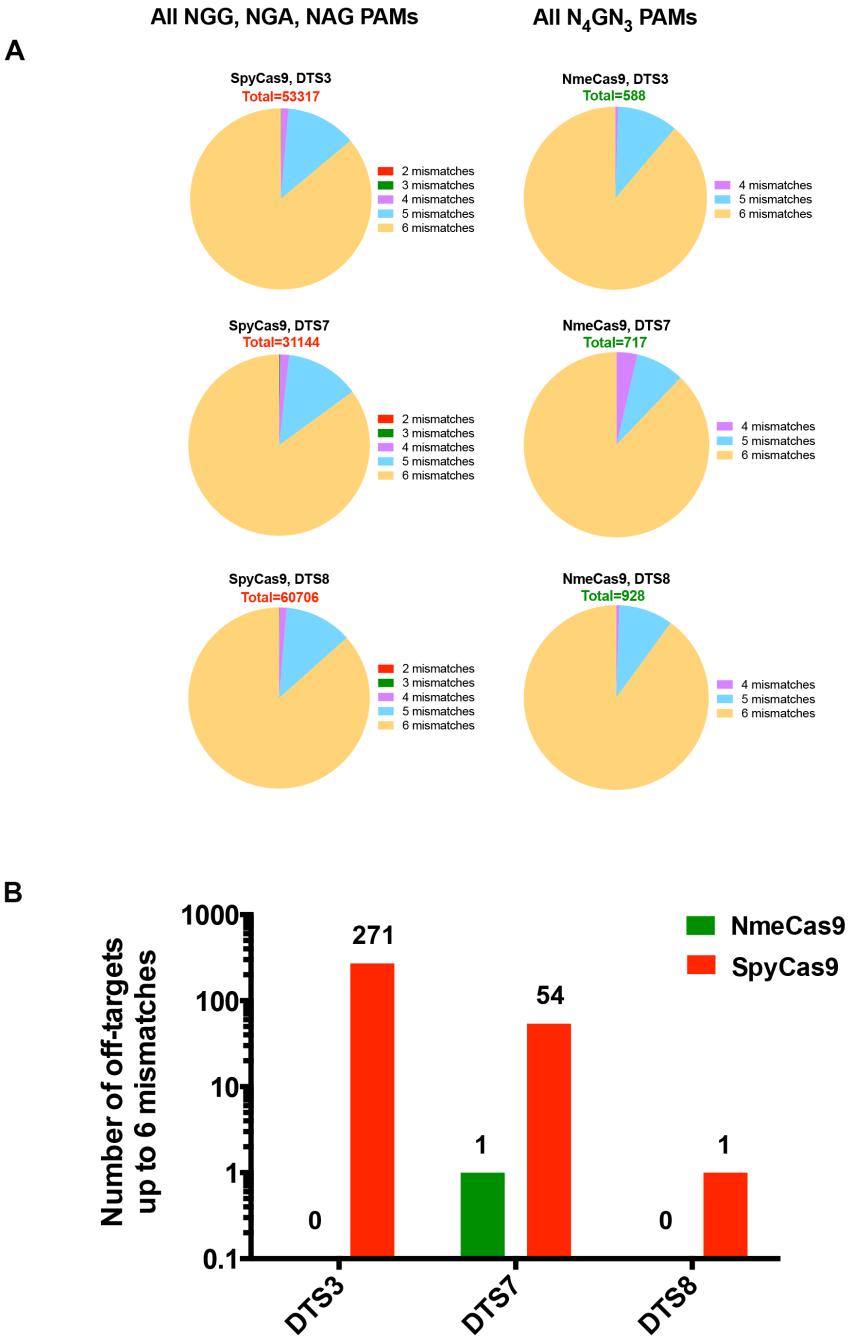
Number of mismatches	**SpyCas9 sites (NGG, NGA, NAG PAMs)**			**NmeCas9 sites (N_4GATT, N_4GCTT, N_4GTTT, N_4GACT, N_4GATA, N_4GTCT, N_4GACA PAMs)**		
	DTS3	**DTS7**	**DTS8**	**DTS3**	**DTS7**	**DTS8**
1	0	0	0	0	0	0
2	4	2	2	0	0	0
3	45	52	60	0	0	0
4	680	500	772	0	2	0
5	6,691	4,116	7,325	4	5	25
6	45,897	26,474	52,547	17	61	129
Total	**53,317**	**31,144**	**60,706**	**21**	**68**	**154**

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.

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_4GN_3, which includes some PAMs that enable only background levels of targeting [e.g. N_4GATC (Fig. 3A)], 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 (Fig. 5A). 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 (Table 2). As speculated previously (Hou et al. 2013; Lee et al. 2016), these bioinformatic predictions suggest the intriguing 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-

Figure 5



412

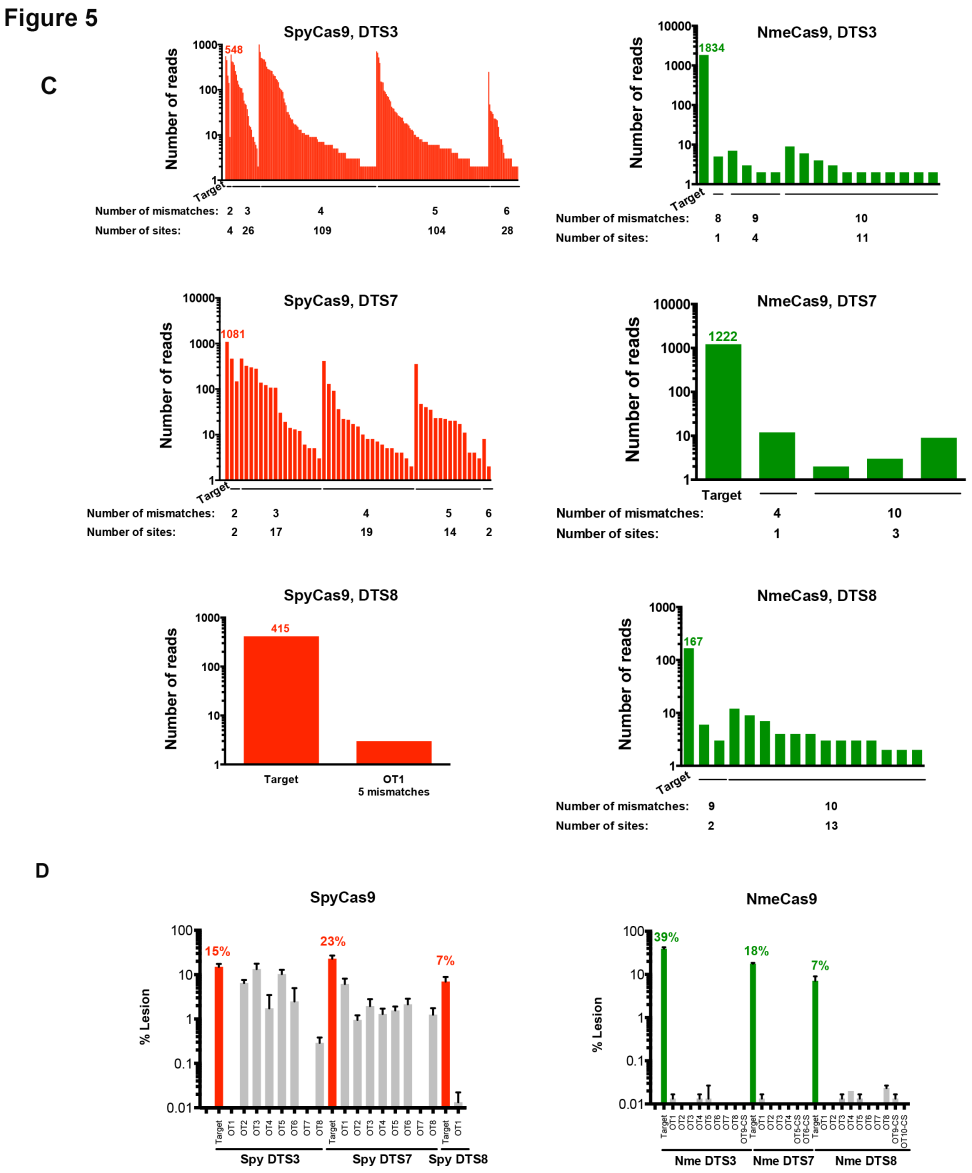


Figure 5. Bioinformatic and empirical comparison of NmeCas9 and SpyCas9 off-target sites within the human genome. (*A*) Genome-wide computational (CRISPRseek) predictions of off-target sites for NmeCas9 (with N_4GN_3 PAMs) and SpyCas9 (with NGG, NGA, and NAG PAMs) with DTS3, DTS7 and DTS8 sgRNAs. Predicted off-target sites were binned based on the number of mismatches (up to six) with the guide sequences. (*B*) GUIDE-seq analysis of off-target sites in HEK293T cells with sgRNAs targeting DTS3, DTS7 and DTS8, using either SpyCas9 or NmeCas9, and with up to 6 mismatches to the sgRNAs. The numbers of detected off-target sites are indicated at the top of each bar. (*C*) Numbers of independent GUIDE-seq reads for the on- and off-target sites for all six Cas9/sgRNA combinations from (*B*) (SpyCas9, red; NmeCas9, green), binned by the number of mismatches with the corresponding guide. (*D*) Targeted deep sequencing analysis of lesion efficiencies at on- and off-target sites from (*A*) or (*B*) with SpyCas9 (left, red) or NmeCas9 (right, green). Data for off-target sites are in grey. For SpyCas9, all off-target sites were chosen from (*B*) based on the highest GUIDE-seq read counts for each guide (Supplemental Table 3). For NmeCas9, in addition to those candidate off-target sites obtained from GUIDE-seq (*C*), we also assayed one or two potential off-target sites (designated with the "-CS" suffix) predicted by CRISPRseek as the closest near-cognate matches with permissive PAMs. Data are mean values ± s.e.m. from three biological replicates performed on different days.

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 (Bolukbasi et al. 2015b; Tsai and Joung 2016; Tycko et al. 2016). 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 lesions 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 (Lee et al. 2016). However, this report restricted its analysis to candidate sites with N_4GNTT PAMs and three or fewer mismatches (or two mismatches combined with a 1-nt bulge) in the PAM-proximal 19 nts, 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 (Bolukbasi et al. 2015b; Tsai and Joung 2016; Tycko et al. 2016)); additional methods [SITE-seq (Cameron et al. 2017), CIRCLE-seq (Tsai et al. 2017), and BLISS (Yan et al. 2017)] have been reported very recently. We chose to apply GUIDE-Seq (Tsai et al. 2014), 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. 4C-D) in the human genome.

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 (Supplemental Fig. 3C), 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

454 described (Zhu et al. 2017) (Fig. 5B-C). On-target editing with these guides was readily detected by this
455 method, with the number of independent reads ranging from a low of 167 (NmeCas9, DTS8) to a high of
456 1,834 (NmeCas9, DTS3) (Fig. 5C and Supplemental Table 2).

457 For our initial analyses, we scored candidate sites as true off-targets if they yielded two or more
458 independent reads and had six or fewer mismatches with the guide, with no constraints placed on the
459 PAM match at that site. For SpyCas9, two of the sgRNAs (targeting DTS3 and DTS7) induced
460 substantial numbers of off-target editing events (271 and 54 off-target sites, respectively (Fig. 5B)] under
461 these criteria. The majority of these SpyCas9 off-target sites (88% and 77% for DTS3 and DTS7,
462 respectively) were associated with a canonical NGG PAM. Reads were very abundant at many of these
463 loci, and at five off-target sites (all with the DTS3 sgRNA) even exceeded the number of on-target reads
464 (Fig. 5C). SpyCas9 was much more precise with the DTS8 sgRNA: we detected a single off-target site
465 with five mismatches and an NGG PAM, and it was associated with only three independent reads, far
466 lower than the 415 reads that we detected at the on-target site (Fig. 5C and Supplemental Table 2).
467 Overall, the range of editing accuracies that we measured empirically for SpyCas9 – very high (e.g.
468 DTS8), intermediate (e.g. DTS7), and poor (e.g. DTS3) – are consistent with the observations of other
469 reports using distinct guides (reviewed in (Bolukbasi et al. 2015b; Tsai and Joung 2016; Tycko et al.
470 2016)).

471 In striking contrast, GUIDE-Seq analyses with NmeCas9, programmed with sgRNAs targeting
472 the exact same three sites, yielded off-target profiles that were exceptionally clean in all cases (Fig. 5B-C).
473 For DTS3 and DTS8 we found no reads at any site with six or fewer guide mismatches; for DTS7 we
474 found one off-target site with four mismatches (three of which were at the PAM-distal end; see
475 Supplemental Table 2), and even at this site there were only twelve independent reads, ~100x fewer than
476 the 1,222 reads detected at DTS7 itself. This off-target site was also associated with a PAM (N_4GGCT)
477 that would be expected to be poorly functional, though it could also be consider a "slipped" PAM with a
478 more optimal consensus but variant spacing (N_5GCTT). Purified, recombinant NmeCas9 has been
479 observed to catalyze DNA cleavage *in vitro* at a site with a similarly slipped PAM (Zhang et al. 2015). To

480 explore the off-targeting potential of NmeCas9 further, we decreased the stringency of our mapping to
481 allow detection of off-target sites with up to 10 mismatches. Even in these conditions, only four (DTS7),
482 fifteen (DTS8), and sixteen (DTS3) candidate sites were identified, most of which had only four or fewer
483 reads (Fig. 5C) and were associated with poorly functional PAMs (Supplemental Table 2). We consider it
484 likely that most if not all of these low-probability candidate off-target sites represent background noise
485 caused by spurious priming and other sources of experimental error.

486 As an additional test of off-targeting potential, we repeated the DTS7 GUIDE-Seq experiments
487 with both SpyCas9 and NmeCas9, but this time using a different transfection reagent (Lipofectamine3000
488 rather than Polyfect). These repeat experiments revealed that >96% (29 out of 30) of off-target sites with
489 up to 5 mismatches were detected under both transfection conditions for SpyCas9 (Supplemental Table
490 1). However, the NmeCas9 GUIDE-Seq data showed no overlap between the potential sites identified
491 under the two conditions, again suggesting that the few off-target reads that we did observe are unlikely to
492 represent legitimate off-target editing sites.

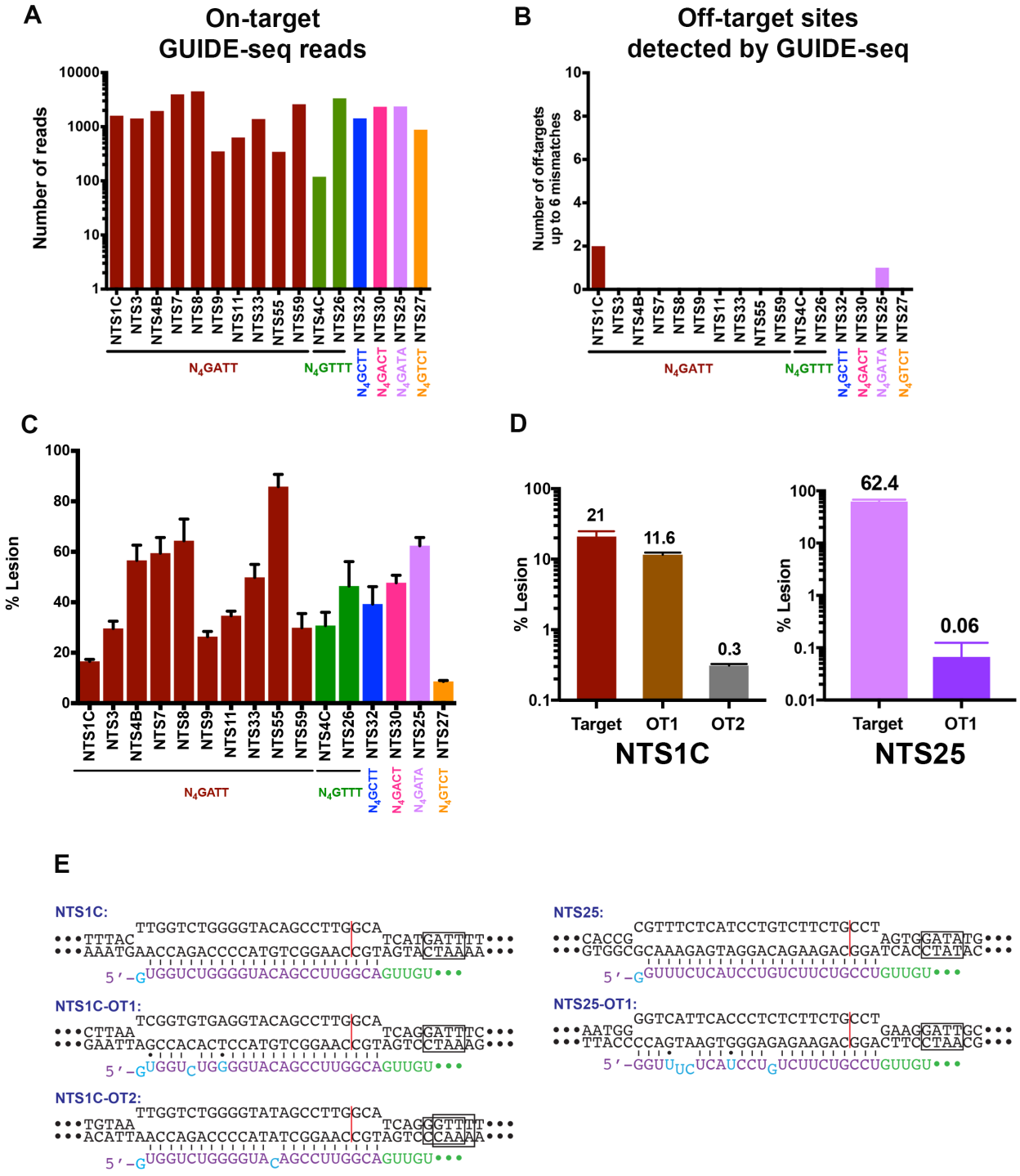
493 To confirm the validity of the off-target sites defined by GUIDE-seq, we designed primers
494 flanking candidate off-target sites identified by GUIDE-Seq, PCR-amplified those loci following standard
495 genome editing (i.e., in the absence of co-transfected GUIDE-Seq dsODN) (3 biological replicates), and
496 then subjected the PCR products to high-throughput sequencing to detect the frequencies of Cas9-
497 induced indels. For this lesion analysis we chose the top candidate off-target sites (as defined by GUIDE-
498 Seq read count) for each of the six cases (DTS3, DTS7 and DTS8, each edited by either SpyCas9 or
499 NmeCas9). In addition, due to the low numbers of off-target sites and the low off-target read counts
500 observed during the NmeCas9 GUIDE-Seq experiments, we analyzed the top two predicted off-target
501 sites for the three NmeCas9 sgRNAs, as identified by CRISPRseek (Fig. 5A and Table 2) (Zhu et al.
502 2014). On-target indel formation was detected in all cases, with lesion efficiencies ranging from 7%
503 (DTS8, with both SpyCas9 and NmeCas9) to 39% (DTS3 with NmeCas9) (Fig. 5D). At the off-target
504 sites, our targeted deep-sequencing analyses largely confirmed our GUIDE-Seq results: SpyCas9 readily
505 induced lesions at most of the tested off-target sites when paired with the DTS3 and DTS7 sgRNAs, and

in some cases the off-target lesion efficiencies approached those observed at the on-target sites (Fig. 5D). 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 lesions 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 lesion efficiencies that were indistinguishable from background, and never rose above ~0.02% (Fig. 5D). These results, in combination with the GUIDE-Seq analyses described above, reveal wild-type NmeCas9 to be an exceptionally precise genome editing enzyme.

To explore NmeCas9 editing accuracy more deeply, we chose 16 additional NmeCas9 target sites across the genome, ten with canonical N_4GATT PAMs and six with variant functional PAMs (Supplementary Table 5). We then performed GUIDE-Seq and lesion 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 ~5,000 reads (Fig. 6A). More notably, off-target reads were undetectable by GUIDE-seq with 14 out of the 16 sgRNAs (Fig. 6B). Targeted deep sequencing of PCR amplicons, which is a more quantitative readout of editing efficiency than either GUIDE-seq or T7E1 analysis, confirmed on-target editing in all cases, with indel efficiencies ranging from ~5-85% (Fig. 6C).

The two guides with off-target activity (NTS1C and NTS25) had only two and one off-target sites, respectively (Fig. 6B and Supplemental Fig. 4). Off-targeting was confirmed by high-throughput sequencing and analysis of indels (Fig. 6D). Compared with the on-target site (perfectly matched at all positions other than the 5'-terminal guide nt, and with an optimal N_4GATT PAM), the efficiently targeted NTS1C-OT1 had two wobble pairs and one mismatch (all in the nine PAM-distal nts), as well as a canonical N_4GATT PAM (Fig. 6E and Supplementary Table 2). The weakly edited NTS1C-OT2 site had only a single mismatch (at the 11th nt, counting in the PAM-distal direction), but was associated with a non-canonical N_4GGTT (or a "slipped" N_5GTTT) PAM (Fig. 6E and Supplementary Table 2). NTS25 with an N_4GATA PAM was the other guide with a single off-target site (NTS25-OT1), where NmeCas9 cleaved and edited ~1,000x less efficiently than at the on-target site (Fig. 6D). This minimal amount of off-targeting arose despite the association of NTS25-OT1 with an optimal N_4GATT PAM, unlike the

Figure 6



532
533 **Figure 6.** GUIDE-seq off-target analyses for sixteen additional NmeCas9 sgRNAs, targeting sites with consensus
534 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 (Zhu et al. 2017) with default settings except that PAM.size = 8, PAM = "NNNNGATT", 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, 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*) Lesion efficiencies for the on-target sites as measured by PCR and high-throughput sequencing. Data are mean values ± s.e.m. from three biological replicates performed on different days. (*D*) NmeCas9 lesion 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 ± s.e.m. from three biological replicates performed on different days. (*E*) 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 nts, 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.

variant N_4GATA PAM that flanks the on-target site. Overall, our GUIDE-Seq and lesion analyses demonstrate that NmeCas9 genome editing is exceptionally accurate: we detected and confirmed off-targeting with only two of the nineteen 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.

Indel spectrum at NmeCas9-edited sites

Our targeted deep sequencing data at the three dual target sites (Fig. 5D and Supplemental Fig. 3A) 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 (Supplemental Fig. 5). Although small deletions predominated at all three sites with both Cas9 orthologs, the frequency of insertions was even lower for NmeCas9 than it was with SpyCas9 (Supplemental Figs. 6 and 7). For both SpyCas9 and NmeCas9, the vast majority of insertions were only a single nucleotide (Supplemental Fig. 7). The sizes of the deletions varied from one target site to the other for both Cas9 orthologs. Our overall conclusions hold for the additional NmeCas9 target sites that we analyzed (Supplemental Figs. 8 and 9): deletions always predominated over insertions, and we observed considerable variations in indel size from one site to the next.

570

571 **Truncated sgRNAs reduce off-target cleavage by NmeCas9**

572 Although NmeCas9 exhibits very little propensity to edit off-target sites, for therapeutic
573 applications it may be desirable to suppress even the small amount of off-targeting that occurs (Fig. 6).
574 Several strategies have been developed to suppress off-targeting by SpyCas9 (Bolukbasi et al. 2015b; Tsai
575 and Joung 2016; Tycko et al. 2016), some of which could be readily applied to other orthologs. For
576 example, truncated sgRNAs (tru-sgRNAs) sometimes suppress off-target SpyCas9 editing more than they
577 suppress on-target editing (Fu et al. 2014b). Because 5'-terminal truncations are compatible with
578 NmeCas9 function (Fig. 2), we tested whether NmeCas9 tru-sgRNAs can have similar suppressive effects
579 on off-target editing.

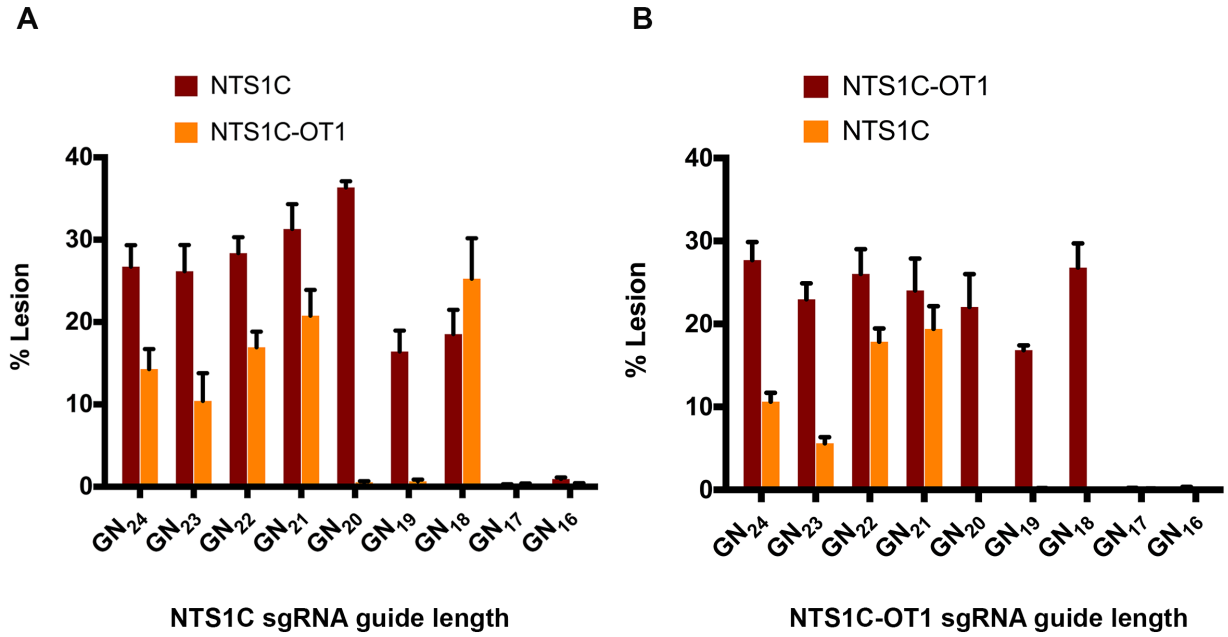
580 First, we tested whether guide truncation can lead to NmeCas9 editing at novel off-target sites (i.e.
581 at off-target sites not edited by full-length guides), as reported previously for SpyCas9 (Fu et al. 2014b).
582 Our earlier tests of NmeCas9 on-target editing with tru-sgRNAs used guides targeting the NTS33 (Fig.
583 2B) and NTS32 (Fig. 2C) sites. GUIDE-seq did not detect any NmeCas9 off-target sites during editing
584 with full-length NTS32 and NTS33 sgRNAs (Fig. 6). We again used GUIDE-seq with a subset of the
585 validated NTS32 and NTS33 tru-sgRNAs to determine whether NmeCas9 guide truncation induces new
586 off-target sites, and found none (Supplemental Fig. 10). Although we cannot rule out the possibility that
587 other NmeCas9 guides could be identified that yield novel off-target events upon truncation, our results
588 suggest that *de novo* off-targeting by NmeCas9 tru-sgRNAs is unlikely to be a pervasive problem.

589 The most efficiently edited off-target site from our previous analyses was NTS1C-OT1, providing
590 us with our most stringent test of off-target suppression. When targeted by the NTS1C sgRNA, NTS1C-
591 OT1 has one rG-dT wobble pair at position -16 (i.e., at the 16th base pair from the PAM-proximal end of
592 the R-loop), one rC-dC mismatch at position -19, and one rU-dG wobble pair at position -23 (Fig. 6E).
593 We generated a series of NTS1C-targeting sgRNAs with a single 5'-terminal G (for U6 promoter
594 transcription) and spacer complementarities ranging from 24 to 15 nts (GN_{24} to GN_{15}, Supplemental Fig.
595 11A). Conversely, we designed a similar series of sgRNAs with perfect complementarity to NTS1C-OT1

596 (Supplemental Fig. 11B). Consistent with our earlier results with other target sites (Fig. 2), T7E1 analyses
597 revealed that both sets of guides enabled editing of the cognate, perfectly-matched site with truncations
598 down to 19 nt (GN_{18}), but that shorter guides were inactive. On-target editing efficiencies at both sites
599 were comparable across the seven active guide lengths (GN_{24} through GN_{18}), with the exception of slightly
600 lower efficiencies with the GN_{19} guides (Supplemental Fig. 11A & B).

601 We then used targeted deep sequencing to test whether off-target editing is reduced with the
602 truncated sgRNAs. With both sets of sgRNAs (perfectly complementary to either NTS1C or NTS1C-
603 OT1), we found that off-targeting at the corresponding near-cognate site persisted with the four longest
604 guides (GN_{24}, GN_{23}, GN_{22}, GN_{21}; Fig. 7). However, off-targeting was abolished with the GN_{20} guide,

Figure 7



605 **Figure 7.** Guide truncation can suppress off-target editing by NmeCas9. (*A*) Lesion efficiencies at the NTS1C (on-
606 target, red) and NTS1C-OT1 (off-target, orange) genomic sites, after editing by NmeCas9 and NTS1C sgRNAs of
607 varying lengths, as measured by PCR and high-throughput sequencing. Data are mean values ± s.e.m. from three
608 biological replicates performed on different days. (*B*) As in (*A*), but using sgRNAs perfectly complementary to the
609 NTS1C-OT1 genomic site.
610

611 without any significant reduction in on-target lesion efficiencies (Fig. 7). Off-targeting was also absent with
612 the GN_{19} guide, though on-target editing efficiency was compromised. These results indicate that
613 truncated sgRNAs (especially those with 20 or 19 bp of guide/target complementarity, 4-5 bp 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_{20} and GN_{19} truncated NTS1C sgRNAs, truncating by an additional nt (to generate the GN_{18} sgRNA) once again yielded NTS1C-OT1 lesions (Fig. 7A). This could be explained by the extra G residue at the 5'-terminus of each sgRNA in the truncation series (Supplemental Fig. 11). With the NTS1C GN_{19} sgRNA, both the 5'-terminal G residue and the adjacent C residue are mismatched with the NTS1C-OT1 site. In contrast, with the GN_{18} sgRNA, the 5'-terminal G is complementary to the off-target site. In other words, with the NTS1C GN_{19} and GN_{18} sgRNAs, the NTS1C-OT1 off-target interactions (which are identical in the PAM-proximal 17 nts) include two additional nts of non-complementarity or one addition nt of complementarity, respectively. Thus, the more extensively truncated GN_{18} sgRNA has *greater* complementarity with the NTS1C-OT1 site than the GN_{19} 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 (Gasiunas et al. 2012; Jinek et al. 2012; Zetsche et al. 2015) 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 (SpyCas9). Thousands of additional Cas9 orthologs have also been identified (Shmakov et al. 2017), but only a few have been characterized, validated for genome engineering applications, or both. Adding 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) (Chen et al. 2017), greater protein stability *in vivo*, improved control mechanisms (e.g. via multiple anti-CRISPRs that act at various stages of the DNA cleavage pathway) (Pawluk et al. 2016; Dong et al. 2017; Rauch et al. 2017; Shin et al. 2017; Yang and Patel 2017), 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 (Esvelt et al. 2013; Hou et al. 2013; Lee et al. 2016). Intriguingly, the tolerance to deviations from the $N_4G(A/C)TT$ natural PAM consensus (Zhang et al. 2013) observed *in vitro* and in bacterial cells (Esvelt et al. 2013; Zhang et al. 2015) 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 decreased expression levels relative to potential DNA substrates, since lower SpyCas9/sgRNA concentrations have been shown to improve accuracy (Hsu et al. 2013; Pattanayak et al. 2013; Fu et al. 2014a). Also related to Cas9 accumulation, we

659 have found that steady-state NmeCas9 levels in human cells are markedly improved in the presence of its
660 cognate sgRNA, suggesting that sgRNA-loaded NmeCas9 is more stable than *apo* NmeCas9. An increased
661 proteolytic sensitivity of *apo* Cas9 relative to the sgRNA-bound form has been noted previously for a
662 different Type II-C ortholog [*Corynebacterium diphtheria* Cas9 (CdiCas9) (Ma et al. 2015)].

663 A previous report indicated that NmeCas9 has high intrinsic accuracy, based on analyses of
664 candidate off-target sites that were predicted bioinformatically (Lee et al. 2016). However, the true
665 genome-wide accuracy of NmeCas9 was not assessed empirically, as is necessary given well-established
666 imperfections in bioinformatic predictions of off-targeting (Bolukbasi et al. 2015b; Tsai and Joung 2016;
667 Tycko et al. 2016). We have use GUIDE-seq (Tsai et al. 2014) to define the genome-wide accuracy of
668 NmeCas9, including side-by-side comparisons with SpyCas9 during editing of identical on-target sites.
669 We find that wild-type NmeCas9 is a consistently high-accuracy genome editor, with off-targeting being
670 undetectable above background with seventeen out of nineteen analyzed sgRNAs, and only one or two
671 verified off-targets with the remaining two guides. We observed this exquisite specificity by NmeCas9 even
672 with sgRNAs that target sites [DTS3 and DTS7 (see Fig. 5)] that are highly prone to off-targeting when
673 edited with SpyCas9. Of the three off-target sites that we validated, two were edited with ~15-fold
674 (NTS1C-OT2) or ~1,000-fold (NTS25-OT1) lower efficiencies than at the corresponding on-target site.
675 Even with the one sgRNA that yielded a significant frequency of off-target editing (NTS1C, which induces
676 lesions at NTS1C-OT1 with approximately half the efficiency of on-target editing), the off-targeting with
677 wild-type NmeCas9 could be easily suppressed with truncated sgRNAs. Our ability to detect NTS25-OT1
678 editing with GUIDE-seq, despite its very low (0.06%) editing efficiency based on high-throughput
679 sequencing, indicates that our GUIDE-seq experiments can identify even very low-efficiency off-target
680 editing sites. As more off-target profiling strategies are developed that have ever-increasing sensitivities
681 (Cameron et al. 2017; Tsai et al. 2017; Yan et al. 2017), it will be useful to test whether NmeCas9 off-
682 targeting remains nearly always undetectable even with improved detection limits.

683 The two Type II-C Cas9 orthologs (NmeCas9 and CjeCas9) that have been validated for
684 mammalian genome editing and assessed for genome-wide specificity (Lee et al. 2016; Kim et al. 2017)

(this work) have both proven to be naturally hyper-accurate. Both use longer guide sequences than the 20-nt 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-targeting. 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 (Bolukbasi et al. 2015b; Tsai and Joung 2016; Tycko et al. 2016). Although Cas9 orthologs with more restrictive PAM requirements (such as NmeCas9 and CjeCas9) by definition will afford lower densities of potential target sites than SpyCas9, 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.

Methods

Plasmids

Two plasmids for the expression of NmeCas9 were used in this study. The first construct (used in Figs. 1 and 2) was derived from the plasmid pSimpleII where NmeCas9 was cloned under the control of the elongation factor-1α promoter, as described previously (Hou et al. 2013). 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 *BsmB*I-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-*BsmB*I plasmid that includes all elements needed for editing. To insert specific spacer sequence into the crRNA cassette, synthetic oligonucleotides were annealed to generate a duplex with overhangs compatible with those generated by *BsmB*I digestion of the pSimpleII-Cas9-sgRNA-*BsmB*I plasmid. The insert was then ligated into the *BsmB*I-digested plasmid. For Figs. 3-7, NmeCas9 and SpyCas9 constructs were expressed from the pCS2-Dest Gateway plasmid under the control of the CMV IE94 promoter (Villefranc et al. 2007). All sgRNAs used with pCS2-Dest-Cas9 were driven by the U6 promoter in pLKO.1-puro (Kearns et al. 2015b). The M427 GFP reporter plasmid (Wilson et al. 2013) was used as described (Bolukbasi et al. 2015a).

Cell culture and transfection

HEK293T cells were cultured in DMEM with 10% FBS and 1% Penicillin/Streptomycin (Gibco) in a 37°C incubator with 5% CO2. For transient transfection, we used early to mid-passage cells (passage number 4-18). Approximately 1.5 x 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.

727 **Western blotting**

728 48 h after transfection, cells were harvested and lysed with 50 µl of RIPA buffer. Protein concentration
729 was determined with the BCA kit (Thermo Scientific) and 12 µg of proteins were used for electrophoresis
730 and blotting. The blots were probed with anti-HA (Sigma, H3663) and anti-GAPDH (Abcam, ab9485) as
731 primary antibodies, and then with horseradish peroxidase–conjugated anti-mouse IgG (Thermoscientific,
732 62-6520) or anti-rabbit IgG (Biorad, 1706515) secondary antibodies, respectively. Blots were visualized
733 using the Clarity Western ECL substrate (Biorad, 170-5060).

734 **Flow cytometry**

735 The GFP reporter was used as described previously (Bolukbasi et al. 2015a). Briefly, cells were harvested
736 48 hours after transfection and used for FACS analysis (BD Accuri 6C). To minimize the effects of
737 differences in the efficiency of transfection among samples, cells were initially gated for mCherry-
738 expression, and the percentage of GFP-expressing cells were quantified within mCherry positive cells. All
739 experiments were performed in triplicate with data reported as mean values with error bars indicating the
740 standard error of the mean (s.e.m.).

741 **Genome editing**

742 72 hours after transfection, genomic DNA was extracted via the DNeasy Blood and Tissue kit (Qiagen),
743 according to the manufacturer's protocol. 50 ng DNA was used for PCR-amplification using primers
744 specific for each genomic site (Supplementary Table 6) with High Fidelity 2X PCR Master Mix (New
745 England Biolabs). For T7E1 analysis, 10 µl of PCR product was hybridized and treated with 0.5 µl T7
746 Endonuclease I (10 U/µl, New England Biolabs) in 1X NEB Buffer 2 for 1 hour. Samples were run on a
747 2.5% agarose gel, stained with SYBR-safe (ThermoFisher Scientific), and quantified using the
748 ImageMaster-TotalLab program. Indel percentages are calculated as previously described (Guschin et al.
749 2010; Gupta et al. 2013). Experiments for T7E1 analysis are performed in triplicate with data reported as
750 mean ± s.e.m.
751

752 **CRISPRseek analysis of potential off-target sites**

753 Global off-target analyses for DTS3, DTS7, and DTS8 with NmeCas9 sgRNAs were performed using the
754 Bioconductor package CRISPRseek 1.9.1 (Zhu et al. 2014) with parameter settings tailored for NmeCas9.
755 Specifically, all parameters are set as default except the following: gRNA.size = 24, PAM =
756 "NNNNGATT", PAM.size = 8, RNA.PAM.pattern = "NNNNGNNN$", weights = c(0, 0, 0, 0, 0, 0,
757 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,
758 0.583), max.mismatch = 6, allowed.mismatch.PAM = 7, topN = 10000, min.score = 0. This setting
759 means that all seven permissive PAM sequences (N_4GATT, N_4GCTT, N_4GTTT, N_4GACA, N_4GACT,
760 N_4GATA, N_4GTCT) were allowed and all off-targets with up to 6 mismatches were collected [the sgRNA
761 length was changed from 20 to 24; four additional zeros were added to the beginning of the weights series
762 to be consistent with the gRNA length of 24; and topN (the number of off-target sites displayed) and
763 min.score (the minimum score of an off-target to be included in the output) were modified to enable
764 identification of all off-target sites with up to 6 mismatches]. Predicted off-target sites for DTS3, DTS7,
765 and DTS8 with SpyCas9 sgRNAs were obtained using CRISPRseek 1.9.1 default settings for SpyCas9
766 (with NGG, NAG, and NGA PAMs allowed). Batch scripts for high-performance computing running the
767 IBM LSF scheduling software are included in the supplemental section. Off-target sites were binned
768 according to the number of mismatches relative to the on-target sequence. The numbers of off-targets for
769 each sgRNA were counted and plotted as pie charts.
770
771 **GUIDE-Seq**
772 We performed GUIDE-seq experiment with some modifications to the original protocol (Tsai et al. 2014),
773 as described (Bolukbasi et al. 2015a). Briefly, in 24-well format, HEK293T cells were transfected with 150
774 ng of Cas9, 150 ng of sgRNA, and 7.5 pmol of annealed GUIDE-seq oligonucleotide using Polyfect
775 transfection reagent (Qiagen) for all six guides (DTS3, DTS7 and DTS8 for both the NmeCas9 and
776 SpyCas9 systems). Experiments with DTS7 sgRNAs were repeated using Lipofectamine 3000 transfection
777 reagent (Invitrogen) according to the manufacturer's protocol. 48 h after transfection, genomic DNA was

778 extracted with a DNeasy Blood and Tissue kit (Qiagen) according to the manufacturer protocol. Library
779 preparation, sequencing, and read analyses were done according to protocols described previously (Tsai et
780 al. 2014; Bolukbasi et al. 2015a). Only sites that harbored a sequence with up to six or ten mismatches
781 with the target site (for SpyCas9 or NmeCas9, respectively) were considered potential off-target sites. Data
782 were analyzed using the Bioconductor package GUIDEseq version 1.1.17 (Zhu et al., 2017). For
783 SpyCas9, default setting was used except that min.reads = 2, max.mismatch = 6, allowed.mismatch.PAM
784 = 2, PAM.pattern = "NNN$", BSgenomeName = Hsapiens, txdb =
785 TxDb.Hsapiens.UCSC.hg19.knownGene, orgAnn = org.Hs.egSYMBOL For NmeCas9, default setting
786 was used except that PAM.size = 8, PAM = "NNNNGATT", min.reads = 2, allowed.mismatch.PAM =
787 4, PAM.pattern = "NNNNNNNN$", BSgenomeName = Hsapiens, txdb =
788 TxDb.Hsapiens.UCSC.hg19.knownGene, orgAnn = org.Hs.egSYMBOL. NmeCas9 dataset was
789 analyzed twice with max.mismatch = 6 and max.mismatch = 10 respectively. The gRNA.size was set to
790 the length of the gRNA used, and various number of 0's was added at the beginning of weights to make
791 the length of weights equal to the gRNA size. For example, for gRNA with length 24, weights =
792 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,
793 0.804, 0.685, 0.583) (Zhu et al., 2017). These regions are reported in Supplemental Table 2.

794 **Targeted deep sequencing analysis**

795 To measure indel frequencies, targeted deep sequencing analyses were done as previously described
796 (Bolukbasi et al. 2015a). Briefly, we used two-step PCR amplification to produce DNA fragments for each
797 on-target and off-target site. In the first step, we used locus-specific primers bearing universal overhangs
798 with complementary ends to the TruSeq adaptor sequences (Supplemental Table 5). DNA was amplified
799 with Phusion High Fidelity DNA Polymerase (New England Biolabs) using annealing temperatures of
800 60°C, 64°C or 68°C, depending on the primer pair. In the second step, the purified PCR products were
801 amplified with a universal forward primer and an indexed reverse primer to reconstitute the TruSeq
802 adaptors (Supplemental Table 5). Input DNA was PCR-amplified with Phusion High Fidelity DNA

Polymerase (98°C, 15s; 61°C, 25s; 72°C, 18s; 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 (~250bp in length) were gel-extracted. The purified library was deep sequenced using a paired-end 150bp 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) (Zhang et al. 2014) to generate single merged high-quality full-length reads. Reads were then filtered by quality [using Filter FASTQ (Blankenberg et al. 2010)] 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 Supplemental Table 3.

Data access

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 XXXXXX.

Acknowledgments

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. 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. E.J.S. is a co-founder and scientific advisor of Intellia Therapeutics.

References

Aouida M, Eid A, Ali Z, Cradick T, Lee C, Deshmukh H, Atef A, AbuSamra D, Gadhoum SZ, Merzaban J et al. 2015. Efficient fdCas9 Synthetic Endonuclease with Improved Specificity for Precise Genome Engineering. *PLoS One* **10**: e0133373.

Barrangou R, Fremaux C, Deveau H, Richards M, Boyaval P, Moineau S, Romero DA, Horvath P. 2007. CRISPR provides acquired resistance against viruses in prokaryotes. *Science* **315**: 1709-1712.

Blankenberg D, Gordon A, Von Kuster G, Coraor N, Taylor J, Nekrutenko A, Galaxy T. 2010. Manipulation of FASTQ data with Galaxy. *Bioinformatics* **26**: 1783-1785.

Bolukbasi MF, Gupta A, Oikemus S, Derr AG, Garber M, Brodsky MH, Zhu LJ, Wolfe SA. 2015a. DNA-binding-domain fusions enhance the targeting range and precision of Cas9. *Nat Methods* **12**: 1150-1156.

Bolukbasi MF, Gupta A, Wolfe SA. 2015b. Creating and evaluating accurate CRISPR-Cas9 scalpels for genomic surgery. *Nat Methods* **13**: 41-50.

Briner AE, Donohoue PD, Gomaa AA, Selle K, Slorach EM, Nye CH, Haurwitz RE, Beisel CL, May AP, Barrangou R. 2014. Guide RNA functional modules direct Cas9 activity and orthogonality. *Mol Cell* **56**: 333-339.

Brouns SJ, Jore MM, Lundgren M, Westra ER, Slijkhuis RJ, Snijders AP, Dickman MJ, Makarova KS, Koonin EV, van der Oost J. 2008. Small CRISPR RNAs guide antiviral defense in prokaryotes. *Science* **321**: 960-964.

852 Cameron P, Fuller CK, Donohoue PD, Jones BN, Thompson MS, Carter MM, Gradia S, Vidal B,
853 Garner E, Slorach EM et al. 2017. Mapping the genomic landscape of CRISPR-Cas9 cleavage. *Nat*
854 *Methods* **14**: 600-606.

855 Chen F, Ding X, Feng Y, Seebeck T, Jiang Y, Davis GD. 2017. Targeted activation of diverse CRISPR-
856 Cas systems for mammalian genome editing via proximal CRISPR targeting. *Nat Commun* **8**: 14958.

857 Cho SW, Kim S, Kim JM, Kim JS. 2013. Targeted genome engineering in human cells with the Cas9
858 RNA-guided endonuclease. *Nat Biotechnol* **31**: 230-232.

859 Cho SW, Kim S, Kim Y, Kweon J, Kim HS, Bae S, Kim JS. 2014. Analysis of off-target effects of
860 CRISPR/Cas-derived RNA-guided endonucleases and nickases. *Genome Res* **24**: 132-141.

861 Chylinski K, Makarova KS, Charpentier E, Koonin EV. 2014. Classification and evolution of type II
862 CRISPR-Cas systems. *Nucleic Acids Res* **42**: 6091-6105.

863 Cong L, Ran FA, Cox D, Lin S, Barretto R, Habib N, Hsu PD, Wu X, Jiang W, Marraffini LA et al.
864 2013. Multiplex genome engineering using CRISPR/Cas systems. *Science* **339**: 819-823.

865 Deltcheva E, Chylinski K, Sharma CM, Gonzales K, Chao Y, Pirzada ZA, Eckert MR, Vogel J,
866 Charpentier E. 2011. CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III.
867 *Nature* **471**: 602-607.

868 Deveau H, Barrangou R, Garneau JE, Labonte J, Fremaux C, Boyaval P, Romero DA, Horvath P,
869 Moineau S. 2008. Phage response to CRISPR-encoded resistance in Streptococcus thermophilus. *J*
870 *Bacteriol* **190**: 1390-1400.

871 Dominguez AA, Lim WA, Qi LS. 2016. Beyond editing: repurposing CRISPR-Cas9 for precision
872 genome regulation and interrogation. *Nat Rev Mol Cell Biol* **17**: 5-15.

873 Dong, Guo M, Wang S, Zhu Y, Wang S, Xiong Z, Yang J, Xu Z, Huang Z. 2017. Structural basis of
874 CRISPR-SpyCas9 inhibition by an anti-CRISPR protein. *Nature* **546**: 436-439.

875 Esvelt KM, Mali P, Braff JL, Moosburner M, Yaung SJ, Church GM. 2013. Orthogonal Cas9 proteins
876 for RNA-guided gene regulation and editing. *Nat Methods* **10**: 1116-1121.

877 Fonfara I, Le Rhun A, Chylinski K, Makarova KS, Lecrivain AL, Bzdrenga J, Koonin EV, Charpentier
878 E. 2014. Phylogeny of Cas9 determines functional exchangeability of dual-RNA and Cas9 among
879 orthologous type II CRISPR-Cas systems. *Nucleic Acids Res* **42**: 2577-2590.

880 Fu BX, Hansen LL, Artiles KL, Nonet ML, Fire AZ. 2014a. Landscape of target:guide homology effects
881 on Cas9-mediated cleavage. *Nucleic Acids Res* **42**: 13778-13787.

882 Fu Y, Foden JA, Khayter C, Maeder ML, Reyon D, Joung JK, Sander JD. 2013. High-frequency off-
883 target mutagenesis induced by CRISPR-Cas nucleases in human cells. *Nat Biotechnol* **31**: 822-826.

884 Fu Y, Sander JD, Reyon D, Cascio VM, Joung JK. 2014b. Improving CRISPR-Cas nuclease specificity
885 using truncated guide RNAs. *Nat Biotechnol* **32**: 279-284.

886 Garneau JE, Dupuis ME, Villion M, Romero DA, Barrangou R, Boyaval P, Fremaux C, Horvath P,
887 Magadan AH, Moineau S. 2010. The CRISPR/Cas bacterial immune system cleaves bacteriophage and
888 plasmid DNA. *Nature* **468**: 67-71.

889 Gasiunas G, Barrangou R, Horvath P, Siksnys V. 2012. Cas9-crRNA ribonucleoprotein complex
890 mediates specific DNA cleavage for adaptive immunity in bacteria. *Proc Natl Acad Sci USA* **109**: E2579-
891 2586.

892 Guan C, Kumar S, Kucera R, Ewel A. 2004. Changing the enzymatic activity of T7 endonuclease by
893 mutations at the beta-bridge site: alteration of substrate specificity profile and metal ion requirements by
894 mutation distant from the catalytic domain. *Biochemistry* **43**: 4313-4322.

Gupta A, Hall VL, Kok FO, Shin M, McNulty JC, Lawson ND, Wolfe SA. 2013. Targeted chromosomal deletions and inversions in zebrafish. *Genome Res* **23**: 1008-1017.

Guschin DY, Waite AJ, Katibah GE, Miller JC, Holmes MC, Rebar EJ. 2010. A rapid and general assay for monitoring endogenous gene modification. *Methods Mol Biol* **649**: 247-256.

Hou Z, Zhang Y, Propson NE, Howden SE, Chu LF, Sontheimer EJ, Thomson JA. 2013. Efficient genome engineering in human pluripotent stem cells using Cas9 from *Neisseria meningitidis*. *Proc Natl Acad Sci USA* **110**: 15644-15649.

Hsu PD, Lander ES, Zhang F. 2014. Development and applications of CRISPR-Cas9 for genome engineering. *Cell* **157**: 1262-1278.

Hsu PD, Scott DA, Weinstein JA, Ran FA, Konermann S, Agarwala V, Li Y, Fine EJ, Wu X, Shalem O et al. 2013. DNA targeting specificity of RNA-guided Cas9 nucleases. *Nat Biotechnol* **31**: 827-832.

Hwang WY, Fu Y, Reyon D, Maeder ML, Kaini P, Sander JD, Joung JK, Peterson RT, Yeh JR. 2013a. Heritable and precise zebrafish genome editing using a CRISPR-Cas system. *PLoS One* **8**: e68708.

Hwang WY, Fu Y, Reyon D, Maeder ML, Tsai SQ, Sander JD, Peterson RT, Yeh JR, Joung JK. 2013b. Efficient genome editing in zebrafish using a CRISPR-Cas system. *Nat Biotechnol* **31**: 227-229.

Jiang F, Taylor DW, Chen JS, Kornfeld JE, Zhou K, Thompson AJ, Nogales E, Doudna JA. 2016. Structures of a CRISPR-Cas9 R-loop complex primed for DNA cleavage. *Science* **351**: 867-871.

Jiang F, Zhou K, Ma L, Gressel S, Doudna JA. 2015. A Cas9-guide RNA complex preorganized for target DNA recognition. *Science* **348**: 1477-1481.

Jiang W, Bikard D, Cox D, Zhang F, Marraffini LA. 2013. RNA-guided editing of bacterial genomes using CRISPR-Cas systems. *Nat Biotechnol* **31**: 233-239.

916 Jinek M, Chylinski K, Fonfara I, Hauer M, Doudna JA, Charpentier E. 2012. A programmable dual-
917 RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* **337**: 816-821.

918 Jinek M, East A, Cheng A, Lin S, Ma E, Doudna J. 2013. RNA-programmed genome editing in human
919 cells. *eLife* **2**: e00471.

920 Jinek M, Jiang F, Taylor DW, Sternberg SH, Kaya E, Ma E, Anders C, Hauer M, Zhou K, Lin S et al.
921 2014. Structures of Cas9 endonucleases reveal RNA-mediated conformational activation. *Science* **343**:
922 1247997.

923 Kearns NA, Pham H, Tabak B, Genga RM, Silverstein NJ, Garber M, Maehr R. 2015a. Functional
924 annotation of native enhancers with a Cas9-histone demethylase fusion. *Nat Methods* **12**: 401-403.

925 Kearns NA, Pham H, Tabak B, Genga RM, Silverstein NJ, Garber M, Maehr R. 2015b. Functional
926 annotation of native enhancers with a Cas9-histone demethylase fusion. *Nat Methods* **12**: 401-403.

927 Kim E, Koo T, Park SW, Kim D, Kim K, Cho HY, Song DW, Lee KJ, Jung MH, Kim S et al. 2017. In
928 vivo genome editing with a small Cas9 orthologue derived from Campylobacter jejuni. *Nat Commun* **8**:
929 14500.

930 Kleinstiver BP, Pattanayak V, Prew MS, Tsai SQ, Nguyen NT, Zheng Z, Joung JK. 2016. High-fidelity
931 CRISPR-Cas9 nucleases with no detectable genome-wide off-target effects. *Nature* **529**: 490-495.

932 Komor AC, Badran AH, Liu DR. 2017. CRISPR-Based Technologies for the Manipulation of
933 Eukaryotic Genomes. *Cell* **168**: 20-36.

934 Lee CM, Cradick TJ, Bao G. 2016. The *Neisseria meningitidis* CRISPR-Cas9 system enables specific
935 genome editing in mammalian cells. *Mol Ther* **24**: 645-654.

936 Ma E, Harrington LB, O'Connell MR, Zhou K, Doudna JA. 2015. Single-Stranded DNA Cleavage by
937 Divergent CRISPR-Cas9 Enzymes. *Mol Cell* **60**: 398-407.

938 Makarova KS, Wolf YI, Alkhnbashi OS, Costa F, Shah SA, Saunders SJ, Barrangou R, Brouns SJ,
939 Charpentier E, Haft DH et al. 2015. An updated evolutionary classification of CRISPR-Cas systems. *Nat*
940 *Rev Microbiol* **13**: 722-736.

941 Mali P, Yang L, Esvelt KM, Aach J, Guell M, DiCarlo JE, Norville JE, Church GM. 2013. RNA-guided
942 human genome engineering via Cas9. *Science* **339**: 823-826.

943 Marraffini LA. 2015. CRISPR-Cas immunity in prokaryotes. *Nature* **526**: 55-61.

944 Marraffini LA, Sontheimer EJ. 2008. CRISPR interference limits horizontal gene transfer in
945 staphylococci by targeting DNA. *Science* **322**: 1843-1845.

946 Mohanraju P, Makarova KS, Zetsche B, Zhang F, Koonin EV, van der Oost J. 2016. Diverse
947 evolutionary roots and mechanistic variations of the CRISPR-Cas systems. *Science* **353**: aad5147.

948 Mojica FJ, Diez-Villasenor C, Garcia-Martinez J, Almendros C. 2009. Short motif sequences determine
949 the targets of the prokaryotic CRISPR defence system. *Microbiol* **155**: 733-740.

950 Nishimasu H, Ran FA, Hsu PD, Konermann S, Shehata SI, Dohmae N, Ishitani R, Zhang F, Nureki O.
951 2014. Crystal structure of Cas9 in complex with guide RNA and target DNA. *Cell* **156**: 935-949.

952 Pattanayak V, Lin S, Guilinger JP, Ma E, Doudna JA, Liu DR. 2013. High-throughput profiling of off-
953 target DNA cleavage reveals RNA-programmed Cas9 nuclease specificity. *Nat Biotechnol* **31**: 839-843.

954 Pawluk A, Amrani N, Zhang Y, Garcia B, Hidalgo-Reyes Y, Lee J, Edraki A, Shah M, Sontheimer EJ,
955 Maxwell KL et al. 2016. Naturally occurring off-switches for CRISPR-Cas9. *Cell*.

Ran FA, Cong L, Yan WX, Scott DA, Gootenberg JS, Kriz AJ, Zetsche B, Shalem O, Wu X, Makarova KS et al. 2015. In vivo genome editing using *Staphylococcus aureus* Cas9. *Nature* **520**: 186-191.

Ran FA, Hsu PD, Lin CY, Gootenberg JS, Konermann S, Trevino AE, Scott DA, Inoue A, Matoba S, Zhang Y et al. 2013. Double nicking by RNA-guided CRISPR Cas9 for enhanced genome editing specificity. *Cell* **154**: 1380-1389.

Rauch BJ, Silvis MR, Hultquist JF, Waters CS, McGregor MJ, Krogan NJ, Bondy-Denomy J. 2017. Inhibition of CRISPR-Cas9 with Bacteriophage Proteins. *Cell* **168**: 150-158 e110.

Sapranauskas R, Gasiunas G, Fremaux C, Barrangou R, Horvath P, Siksnys V. 2011. The *Streptococcus thermophilus* CRISPR/Cas system provides immunity in *Escherichia coli*. *Nucleic Acids Res* **39**: 9275-9282.

Shin CS, Jiang F, Liu J-J, Bray NL, Rauch BJ, Baik SH, Nogales E, Bondy-Denomy J, Corn JE, Doudna JA. 2017. Disabling Cas9 by an anti-CRISPR DNA mimic. *Science Advances*: , in press.

Shmakov S, Smargon A, Scott D, Cox D, Pyzocha N, Yan W, Abudayyeh OO, Gootenberg JS, Makarova KS, Wolf YI et al. 2017. Diversity and evolution of class 2 CRISPR-Cas systems. *Nat Rev Microbiol* **15**: 169-182.

Slaymaker IM, Gao L, Zetsche B, Scott DA, Yan WX, Zhang F. 2016. Rationally engineered Cas9 nucleases with improved specificity. *Science* **351**: 84-88.

Sontheimer EJ, Barrangou R. 2015. The bacterial origins of the CRISPR genome-editing revolution. *Hum Gene Ther* **26**: 413-424.

Sternberg SH, Doudna JA. 2015. Expanding the Biologist's Toolkit with CRISPR-Cas9. *Mol Cell* **58**: 568-574.

976 Tsai SQ, Joung JK. 2016. Defining and improving the genome-wide specificities of CRISPR-Cas9
977 nucleases. *Nat Rev Genet* **17**: 300-312.

978 Tsai SQ, Nguyen NT, Malagon-Lopez J, Topkar VV, Aryee MJ, Joung JK. 2017. CIRCLE-seq: a highly
979 sensitive in vitro screen for genome-wide CRISPR-Cas9 nuclease off-targets. *Nat Methods* **14**: 607-614.

980 Tsai SQ, Zheng Z, Nguyen NT, Liebers M, Topkar VV, Thapar V, Wyvekens N, Khayter C, Iafrate AJ,
981 Le LP et al. 2014. GUIDE-seq enables genome-wide profiling of off-target cleavage by CRISPR-Cas
982 nucleases. *Nat Biotechnol* **33**: 187-197.

983 Tycko J, Myer VE, Hsu PD. 2016. Methods for optimizing CRISPR-Cas9 genome editing specificity. *Mol*
984 *Cell* **63**: 355-370.

985 Villefranc JA, Amigo J, Lawson ND. 2007. Gateway compatible vectors for analysis of gene function in
986 the zebrafish. *Dev Dyn* **236**: 3077-3087.

987 Wang H, La Russa M, Qi LS. 2016. CRISPR/Cas9 in Genome Editing and Beyond. *Annu Rev Biochem*
988 **85**: 227-264.

989 Wilson KA, McEwen AE, Pruett-Miller SM, Zhang J, Kildebeck EJ, Porteus MH. 2013. Expanding the
990 Repertoire of Target Sites for Zinc Finger Nuclease-mediated Genome Modification. *Mol Ther Nucleic*
991 *Acids* **2**: e88.

992 Yan WX, Mirzazadeh R, Garnerone S, Scott D, Schneider MW, Kallas T, Custodio J, Wernersson E, Li
993 Y, Gao L et al. 2017. BLISS is a versatile and quantitative method for genome-wide profiling of DNA
994 double-strand breaks. *Nat Commun* **8**: 15058.

995 Yang H, Patel DJ. 2017. Inhibition Mechanism of an Anti-CRISPR Suppressor AcrIIA4 Targeting
996 SpyCas9. *Mol Cell* doi:10.1016/j.molcel.2017.05.024.

997 Zetsche B, Gootenberg JS, Abudayyeh OO, Slaymaker IM, Makarova KS, Essletzbichler P, Volz SE,
998 Joung J, van der Oost J, Regev A et al. 2015. Cpf1 is a single RNA-guided endonuclease of a class 2
999 CRISPR-Cas system. *Cell* **163**: 759-771.

1000 Zhang J, Kobert K, Flouri T, Stamatakis A. 2014. PEAR: a fast and accurate Illumina Paired-End reAd
1001 mergeR. *Bioinformatics* **30**: 614-620.

1002 Zhang Y, Heidrich N, Ampattu BJ, Gunderson CW, Seifert HS, Schoen C, Vogel J, Sontheimer EJ.
1003 2013. Processing-independent CRISPR RNAs limit natural transformation in Neisseria meningitidis. *Mol*
1004 *Cell* **50**: 488-503.

1005 Zhang Y, Rajan R, Seifert HS, Mondragón A, Sontheimer EJ. 2015. DNase H activity of *Neisseria*
1006 *meningitidis* Cas9. *Mol Cell* **60**: 242-255.

1007 Zhu LJ, Holmes BR, Aronin N, Brodsky MH. 2014. CRISPRseek: a bioconductor package to identify
1008 target-specific guide RNAs for CRISPR-Cas9 genome-editing systems. *PLoS One* **9**: e108424.

1009 Zhu LJ, Lawrence M, Gupta A, Pagés H, Kucukural A, Garber M, Wolfe SA. 2017. GUIDEseq: a
1010 bioconductor package to analyze GUIDE-Seq datasets for CRISPR-Cas nucleases. *BMC Genomics* **18**:
1011 379.
1012