ARTICLE

doi:10.1038/nature14299

In vivo genome editing using *Staphylococcus aureus* Cas9

F. Ann Ran[1,2]*, Le Cong[1,3]*, Winston X. Yan[1,4,5]*, David A. Scott[1,6,7], Jonathan S. Gootenberg[1,8], Andrea J. Kriz[3], Bernd Zetsche[1], Ophir Shalem[1], Xuebing Wu[9,10], Kira S. Makarova[11], Eugene V. Koonin[11], Phillip A. Sharp[3,9] & Feng Zhang[1,6,7,12]

The RNA-guided endonuclease Cas9 has emerged as a versatile genome-editing platform. However, the size of the commonly used Cas9 from *Streptococcus pyogenes* (SpCas9) limits its utility for basic research and therapeutic applications that use the highly versatile adeno-associated virus (AAV) delivery vehicle. Here, we characterize six smaller Cas9 orthologues and show that Cas9 from *Staphylococcus aureus* (SaCas9) can edit the genome with efficiencies similar to those of SpCas9, while being more than 1 kilobase shorter. We packaged SaCas9 and its single guide RNA expression cassette into a single AAV vector and targeted the cholesterol regulatory gene *Pcsk9* in the mouse liver. Within one week of injection, we observed >40% gene modification, accompanied by significant reductions in serum Pcsk9 and total cholesterol levels. We further assess the genome-wide targeting specificity of SaCas9 and SpCas9 using BLESS, and demonstrate that SaCas9-mediated *in vivo* genome editing has the potential to be efficient and specific.

Cas9, an RNA-guided endonuclease derived from the type II CRISPR-Cas bacterial adaptive immune system[1–7], has been harnessed for genome editing[8,9] and holds tremendous promise for biomedical research. Genome editing of somatic tissue in postnatal animals, however, has been limited in part by the challenge of delivering Cas9 *in vivo*. For this purpose, adeno-associated virus (AAV) vectors are attractive vehicles[10] because of their low immunogenic potential, reduced oncogenic risk from host-genome integration[11], and broad range of serotype specificity[12–15]. Nevertheless, the restrictive cargo size (~4.5 kb, excluding the inverted terminal repeats) of AAV presents an obstacle for packaging the commonly used *Streptococcus pyogenes* Cas9 (SpCas9, ~4.2 kb) and its single guide RNA (sgRNA) in a single vector; although technically feasible[17], this approach leaves little room for customized expression and control elements[16].

In search of smaller Cas9 enzymes for efficient *in vivo* delivery by AAV, we have previously described a short Cas9 from the CRISPR1 locus of *Streptococcus thermophilus* LMD-9 (St1Cas9, ~3.3 kb)[8] as well as a rationally-designed truncated form of SpCas9 (ref. 18) for genome editing in human cells. However, both systems have important practical drawbacks: the former requires a complex protospacer-associated motif (PAM) sequence (NNAGAAW)[3], which restricts the range of accessible targets, whereas the latter exhibits reduced activity. Given the substantial diversity of CRISPR-Cas systems present in sequenced microbial genomes[19], we therefore sought to interrogate and discover additional Cas9 enzymes that are small, efficient and broadly targeting.

In vitro cleavage by small Cas9 enzymes

Type II CRISPR-Cas systems require only two main components for eukaryotic genome editing: a Cas9 enzyme, and a chimaeric sgRNA[6] derived from the CRISPR RNA (crRNA) and the noncoding trans-activating crRNA (tracrRNA)[4,20]. Analysis of over 600 Cas9 orthologues shows that these enzymes are clustered into two length groups with characteristic protein sizes of approximately 1,350 and 1,000 amino acid residues, respectively[19,21] (Extended Data Fig. 1a), with shorter Cas9 enzymes having significantly truncated REC domains (Fig. 1a). From these shorter Cas9 enzymes, which belong to Type IIA and IIC subtypes, we selected six candidates for profiling (Fig. 1a and Extended Data Fig. 1b). To determine the cognate crRNA and tracrRNA for each Cas9, we computationally identified regularly interspaced repeat sequences (direct repeats) within a 2-kb window flanking the CRISPR locus. We then predicted the tracrRNA by detecting sequences with strong complementarity to the direct repeat sequence (an anti-repeat region), at least two predicted stem-loop structures, and a Rho-independent transcriptional termination signal up to 150 nucleotides downstream of the anti-repeat region. Although a truncated tracrRNA can support robust DNA cleavage *in vitro*[6], previous reports show that the secondary structures of the tracrRNA are important for Cas9 activity in mammalian cells[8,9,18,22]. Therefore, we designed sgRNA scaffolds for each orthologue by fusing the 3′ end of a truncated direct repeat with the 5′ end of the corresponding tracrRNA, including the full-length tail, via a 4-nucleotide linker[6] (Extended Data Fig. 1b and Supplementary Table 1). To identify the PAM sequence for each Cas9, we first constructed a library of plasmid DNA containing a constant 20-bp target followed by a degenerate 7-bp sequence (5′-NNNNNNN). We then incubated cell lysate from human embryonic kidney 293FT (293FT) cells expressing the Cas9 orthologue with its *in vitro*-transcribed sgRNA and the plasmid library. By generating a consensus from the 7-bp sequence found on successfully cleaved DNA plasmids (Fig. 1b), we determined putative PAMs for each Cas9 (Fig. 1c). We observed that, similar to SpCas9, most Cas9 orthologues cleaved targets 3-bp upstream of the PAM (Extended Data Fig. 2). To validate each putative PAM from the library, we then incubated a DNA template bearing the consensus PAM with cell lysate

[1]Broad Institute of MIT and Harvard, Cambridge, Massachusetts 02142, USA. [2]Society of Fellows, Harvard University, Cambridge, Massachusetts 02138, USA. [3]Department of Biology, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA. [4]Graduate Program in Biophysics, Harvard Medical School, Boston, Massachusetts 02115, USA. [5]Harvard-MIT Division of Health Sciences and Technology, Harvard Medical School, Boston, Massachusetts 02115, USA. [6]McGovern Institute for Brain Research, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA. [7]Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA. [8]Department of Systems Biology, Harvard Medical School, Boston, Massachusetts 02115, USA. [9]David H. Koch Institute for Integrative Cancer Research, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA. [10]Computational and Systems Biology Graduate Program, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA. [11]National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health, Bethesda, Maryland 20894, USA. [12]Department of Biological Engineering, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA.
*These authors contributed equally to this work.

©2015 Macmillan Publishers Limited. All rights reserved

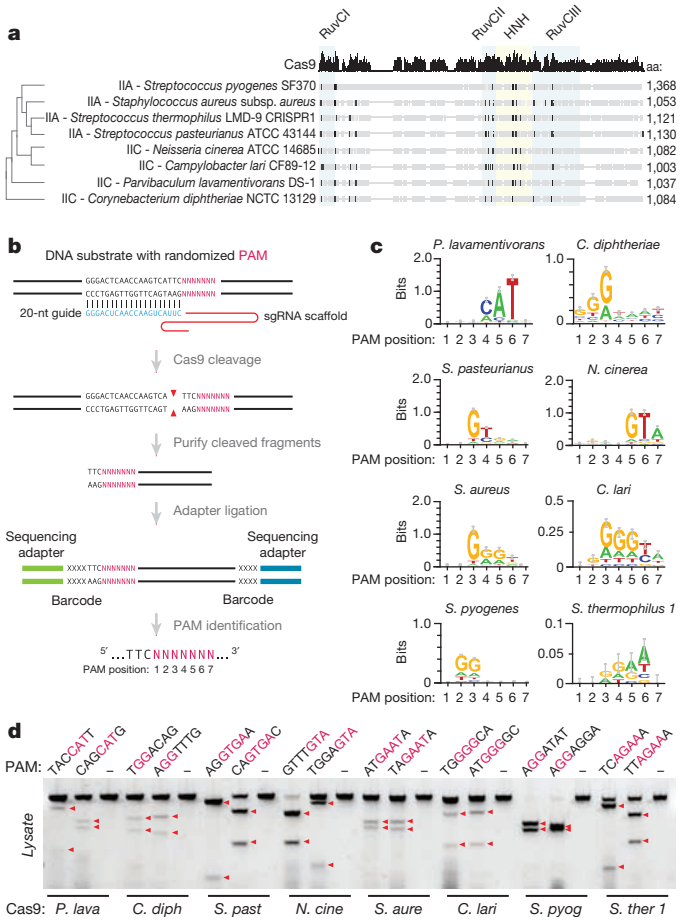


Figure 1 | Biochemical screen for small Cas9 orthologues. **a**, Phylogenetic tree of selected Cas9 orthologues. Subfamily and sizes (amino acids) are indicated, with nuclease domains highlighted in coloured boxes, and conserved sequences in black. **b**, Schematic illustration of the *in vitro* cleavage-based method used to identify the first seven positions (5′-NNNNNNN) of protospacer adjacent motifs (PAMs). **c**, Consensus PAMs for eight Cas9 orthologues from sequencing of cleaved fragments. Error bars are Bayesian 95% confidence interval[45]. **d**, Cleavage using different orthologues and sgRNAs targeting loci bearing the putative PAMs (consensus shown in red). Red triangles indicate cleavage fragments.

and the corresponding sgRNA. We found that the Cas9 orthologues, in combination with the sgRNA designs, successfully cleaved the appropriate targets (Fig. 1d and Supplementary Table 2).

To test whether each Cas9 orthologue can facilitate genome editing in mammalian cells, we co-transfected 293FT cells with individual Cas9 enzymes and their respective sgRNAs targeting human endogenous loci containing the appropriate PAMs. Of the six Cas9 orthologues tested, only the one from *Staphylococcus aureus* (SaCas9) produced indels with efficiencies comparable to those of SpCas9 (Extended Data Fig. 3a, b and Supplementary Table 3), suggesting that DNA-cleavage activity in cell-free assays does not necessarily predict activity in mammalian cells. These observations prompted us to focus on harnessing SaCas9 and its sgRNA for *in vivo* applications.

SaCas9 sgRNA design and PAM discovery

Although mature crRNAs in *S. pyogenes* are processed to contain 20-nucleotide spacers (guides) and 19- to 22-nucleotide direct repeats[4], RNA sequencing of crRNAs from other organisms reveals that the spacer and direct repeat sequence lengths can vary[4,20,23]. We therefore tested sgRNAs for SaCas9 with variable guide lengths and repeat:anti-repeat duplexes. We found that SaCas9 achieves the highest editing efficiency in mammalian cells with guides between 21 and 23 nucleotides long and can accommodate a range of lengths for the direct repeat:anti-repeat region (Fig. 2a, b, Extended Data Fig. 4). This notably contrasts with SpCas9, where the natural 20-nucleotide guide length can be truncated to 17 nucleotides without significantly compromising nuclease activity, while increasing specificity[24]. Additionally, replacing the first base of the guide with guanine further improved SaCas9 activity (Extended Data Fig. 3c).

To fully characterize the SaCas9 PAM and the seed region within its guide sequence[25], we performed chromatin immunoprecipitation (ChIP) using catalytically mutant forms of SaCas9 (dSaCas9, D10A and N580A mutations, based on homology to SpCas9) or SpCas9 (dSpCas9, D10A and H840A mutations) and their corresponding sgRNAs. We targeted two loci in the human *EMX1* gene with composite NGGRRT PAMs, which allow targeting by both dCas9s. A search for motifs containing both the guide region and PAM within 50 nucleotides of the ChIP peak summits revealed seed sequences of 7–8 nucleotides for dSaCas9 (Fig. 2c). In addition, NNGRRT and NGG PAMs were found adjacent to the seed sequences for dSaCas9 and dSpCas9, respectively (Extended Data Fig. 5). Although the 6th position of the PAM is predominantly thymine, we did observe low levels of degeneracy in both the biochemical and ChIP-based PAM discovery assays (Fig. 1c and Extended Data Fig. 5a). We therefore tested the base preference for this position and determined that, although SaCas9 cleaves genomic targets most efficiently with NNGRRT, all NNGRR PAMs can be cleaved and should be considered as potential targets, especially in the context of off-target evaluations (Fig. 2d, Extended Data Fig. 6 and Supplementary Table 4).

Unbiased profiling of Cas9 specificity

As advances in Cas9 technology promise to enable a broad range of *in vivo* and therapeutic applications, accurate, genome-wide identification of off-target nuclease activity has become increasingly important. Although a number of studies have employed sequence similarity-based off-target search[22,26–30] or dCas9-ChIP[31,32] to predict off-target sites for Cas9, such approaches cannot assess the nuclease activity of Cas9 in a comprehensive and unbiased manner. To measure the genome-wide

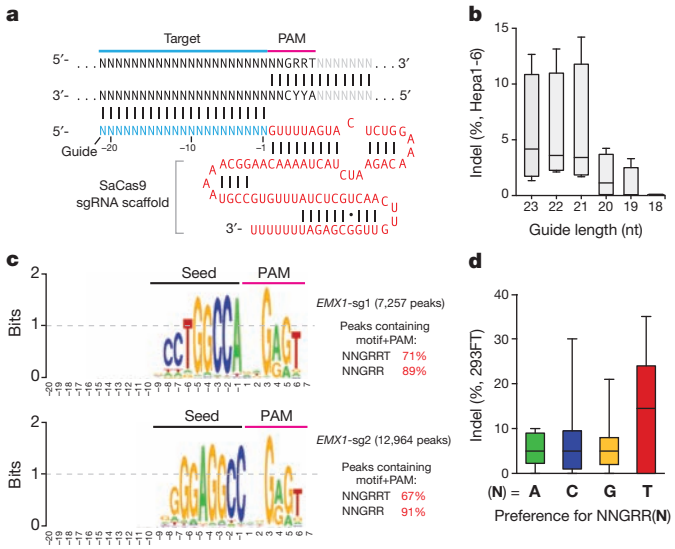


Figure 2 | Characterization of *Staphylococcus aureus* Cas9 (SaCas9) in 293FT cells. **a**, SaCas9 sgRNA scaffold (red) and guide (blue) base-pairing at target locus (black) immediately 5′ of PAM. **b**, Box-whisker plot showing indel levels vary depending on the length of the guide sequence ($n = 4$). **c**, dSaCas9-ChIP reveals peaks associated with seed + PAM. Text to the right indicates the total number of peaks and percentage containing significant (false discovery rate < 0.1) match to the guide motif followed by NNGRRT or NNGRR PAMs. **d**, Pooled indel values for NNGRR(A), (C), (G), or (T) PAM combinations ($n = 12$, 21, 39 and 44, respectively).

©2015 Macmillan Publishers Limited. All rights reserved

cleavage activity of SaCas9 and SpCas9 directly, we applied BLESS (direct *in situ* breaks labelling, enrichment on streptavidin and next-generation sequencing)[33] to capture a snapshot of Cas9-induced DNA double-stranded breaks (DSBs) in cells. We transfected 293FT cells with SaCas9 or SpCas9 and the same *EMX1* targeting guides used in the previous ChIP experiment, or pUC19 as a negative control. After cells are fixed, free genomic DNA ends from DSBs are captured using biotinylated adaptors and analysed by deep sequencing (Fig. 3a). To identify candidate Cas9-induced DSB sites genome-wide, we established a three-step analysis pipeline following alignment of the sequenced BLESS reads to the genome (Extended Data Fig. 7a, Supplementary Discussion). First, we applied nearest-neighbour clustering on the aligned reads to identify groups of DSBs (DSB clusters) across the genome. Second, we sought to separate potential Cas9-induced DSB clusters from background DSB clusters resulting from low frequency biological processes and technical artefacts, as well as high-frequency telomeric and centromeric DSB hotspots[33]. From the on-target and a small subset of verified off-target sites (predicted by sequence similarity using a previously established method[22] and sequenced to detect indels), we found that reads in Cas9-induced DSB clusters mapped to characteristic, well-defined genomic positions compared to the more diffuse alignment pattern at background DSB clusters. To distinguish between the two types of DSB clusters, we calculated in each cluster the distance between all possible pairs of forward and reverse-oriented reads (corresponding to 3′ and 5′ ends of DSBs), and filtered out the background DSB clusters based on the distinctive pairwise-distance distribution of these clusters (Extended Data Fig. 7b, c). Third, the DSB score for a given locus was calculated by comparing the count of DSBs in the experimental and negative control samples using a maximum-likelihood estimate[22] (Supplementary Discussion). This analysis identified the on-target loci for both SaCas9 and SpCas9 guides as the top scoring sites, and revealed additional sites with high DSB scores (Fig. 3b–d).

Next, we sought to assess whether DSB scores correlated with indel formation. We used targeted deep sequencing to detect indel formation

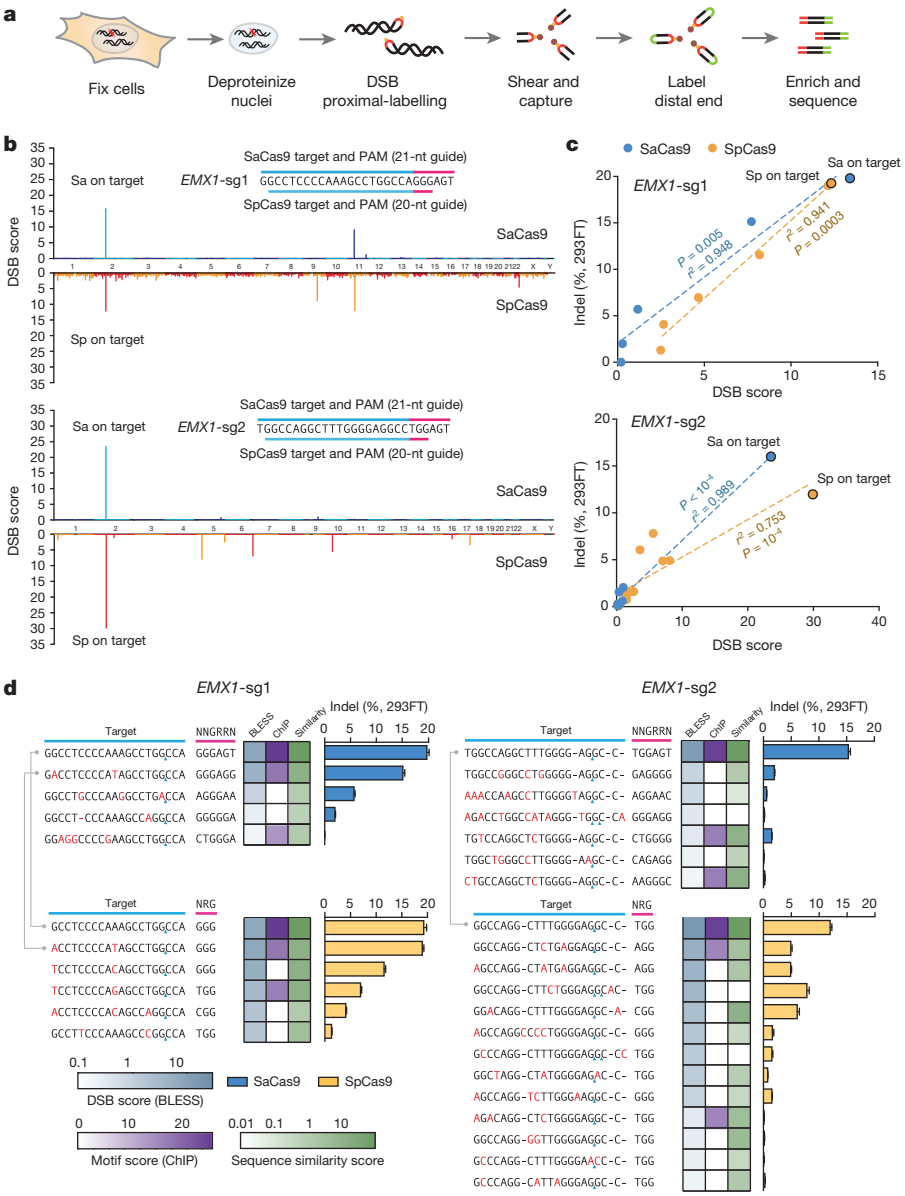


Figure 3 | Characterization of genome-wide nuclease activity of SaCas9 and SpCas9. **a**, Schematic of BLESS processing steps. **b**, Manhattan plots of genome-wide DSB clusters generated by each Cas9 and sgRNA pair, with on-target loci shown above (see Supplementary Discussion). **c**, Correlation between DSB scores and indel levels for top-scoring DSB clusters. Trend lines, r^2 and P values are calculated using ordinary least squares method. **d**, Off-target loci from BLESS with detectable indels through targeted deep sequencing ($n = 3$) are shown. Heat maps indicate DSB score (blue), motif score from ChIP (purple), or sequence similarity score (green) for each locus. Blue triangles indicate peak positions of BLESS signal.

©2015 Macmillan Publishers Limited. All rights reserved

on the ~30 top-ranking off-target sites identified by BLESS for each Cas9 and sgRNA combination. We found that only those sites that contained PAM and homology to the guide sequence exhibited indels (Extended Data Fig. 8). We observed a strong linear correlation between DSB scores and indel levels for each Cas9 and sgRNA pairing ($r^2 = 0.948$ and 0.989 for the two *EMX1* targets with SaCas9 and $r^2 = 0.941$ and 0.753 for those with SpCas9) (Fig. 3c, Extended Data Fig. 9b–d). Furthermore, BLESS identified additional off-target sites not previously predicted by sequence similarity to target or ChIP (Extended Data Figs 7 and 9, Supplementary Tables 5 and 6). These new off-target sites include not only those containing Watson–Crick base-pairing mismatches to the guide, but also the recently reported insertion and deletion mismatches in the guide:target heteroduplex (Fig. 3d)[29,30]. Together, these results highlight the need for more precise understanding of rules governing Cas9 nuclease activity, a requisite step towards improving the predictive power of computational guide design programs.

In vivo genome editing using SaCas9

Following *in vitro* characterization, we incorporated SaCas9 and its sgRNA into an AAV vector to test its efficacy and specificity *in vivo*. The small size of SaCas9 enables packaging of both a U6-driven sgRNA and a cytomegalovirus (CMV)- or thyroxine-binding globulin (TBG)-driven SaCas9 expression cassette into a single AAV vector within the 4.5-kb packaging limit. Using hepatocyte-tropic AAV serotype 8, we targeted the mouse apolipoprotein (*Apob*) gene (Extended Data Fig. 10a). One week after intravenous administration of virus into C57BL/6 mice, we observed ~5% indel formation in liver tissue; after four weeks, the liver tissue showed characteristic hepatic lipid accumulation from Apob knockdown following histology analysis using oil red staining[34–37] (Extended Data Fig. 10b, c).

We next targeted proprotein convertase subtilisin/kexin type 9 (*Pcsk9*), a therapeutically relevant gene involved in cholesterol homeostasis[38]. Inhibitors of the human convertase PCSK9 have emerged as a promising new class of cardioprotective drugs, after human genetic studies revealed that loss of *PCSK9* is associated with a reduced risk of cardiovascular disease and lower levels of low-density lipoprotein (LDL) cholesterol[39–41]. We designed two *Pcsk9*-targeting sgRNAs (20-nucleotide guides with additional 5′ guanine) and validated their activity *in vitro*. Each sgRNA was packaged into AAV-SaCas9 and injected into mice (2×10^{11} total genome copies) (Fig. 4a). One week after administration, we observed greater than 40% indel formation at either locus in whole liver tissue, with similar levels two and four weeks post-injection (Fig. 4b). To determine the effect of *Pcsk9*-targeting AAV-SaCas9 dosage on serum Pcsk9 and total cholesterol levels, we administered a range of AAV titres from 0.5×10^{11} to 4×10^{11} total genome copies. With all titres, we observed a ~95% decrease in serum Pcsk9 and a ~40% decrease in total cholesterol one week after administration, both of which were sustained throughout the course of four weeks (Fig. 4c, d).

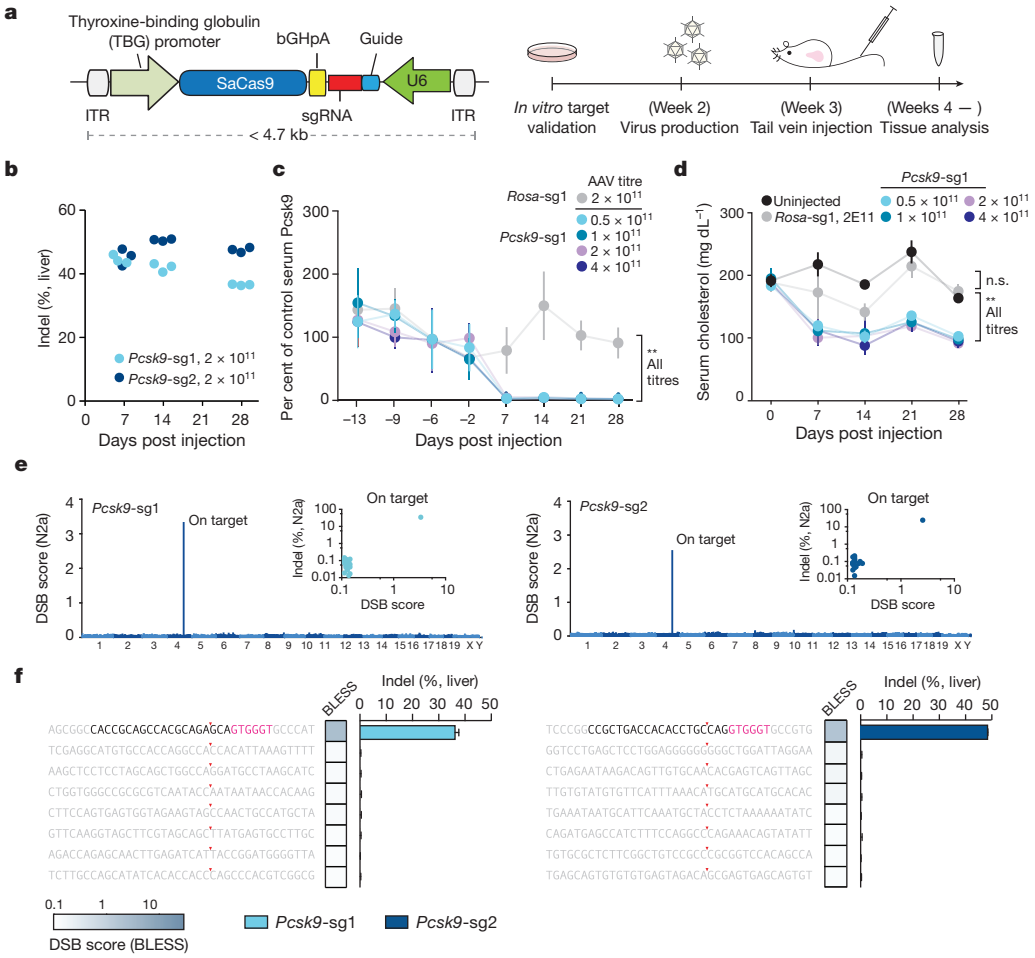


Figure 4 | AAV-delivery of SaCas9 for *in vivo* genome editing. **a**, Single-vector AAV system and experimental timeline. **b**, Indels at *Pcsk9* targets in liver tissue following injection of AAV at 2×10^{11} total genome copies ($n = 3$ animals). **c**, **d**, Time course of serum Pcsk9 (**c**) and total cholesterol in animals (**d**; $n = 3$ for all titres and time points, error bars show s.e.m.). **e**, Manhattan plots of BLESS-identified DSB clusters in N2a cells. Inset indicates indel levels at top DSB scoring loci. **f**, Indels in liver tissue ($n = 3$ animals, error bars indicate Wilson intervals) at BLESS-identified off-target loci from N2a cells. Heat map indicates DSB scores.

©2015 Macmillan Publishers Limited. All rights reserved

We next considered SaCas9 off-target modifications in the liver tissue samples. To first identify candidate off-target cleavage sites for the two *Pcsk9*-targeting guides, we transiently transfected an AAV-CMV::SaCas9 vector into mouse Neuroblastoma-2a (N2a) cells and applied BLESS to detect Cas9-induced DSBs in the genome. For both guides, we found very low levels of DSB signal across the genome except at the on-target loci (Fig. 4e). Targeted deep sequencing of the candidate off-target sites identified by BLESS in N2a cells did not reveal appreciable levels of indels in either N2a cells or liver tissue (4 weeks post injection of 2×10^{11} total genome copies) (Fig. 4e, f and Supplementary Table 8). We additionally sequenced off-target sites predicted by target sequence similarity, and likewise did not detect indel formations (Supplementary Table 9).

Finally, we examined the titre-matched *Pcsk9*-targeting and enhanced green fluorescent protein-conjugated (EGFP) TBG–EGFP cohorts as well as naive animals for signs of toxicity or acute immune response. At 1 week post-injection, necropsy and gross examination of liver tissue of the cohorts revealed no abnormalities; further histological examination of the liver by haematoxylin and eosin (H&E) staining showed no signs of inflammation, such as aggregates of lymphocytes or macrophages (Fig. 5a). Throughout the time course of the experiment, there were no elevated levels of serum alanine aminotransferase (ALT), albumin, and total bilirubin in any of the cohorts. We observed a slight trend in aspartate transaminase (AST) increase across all cohorts at four weeks, including the uninjected animals. The elevated levels did not exceed the upper limit of normal and is not indicative of hepatocellular injury in animals (Fig. 5b). However, a larger cohort study should be conducted to further evaluate the potential side-effects of Cas9-mediated *in vivo* genome editing. In addition, the differences between mouse and human immune responses need to be better elucidated before considering this approach for therapeutic applications.

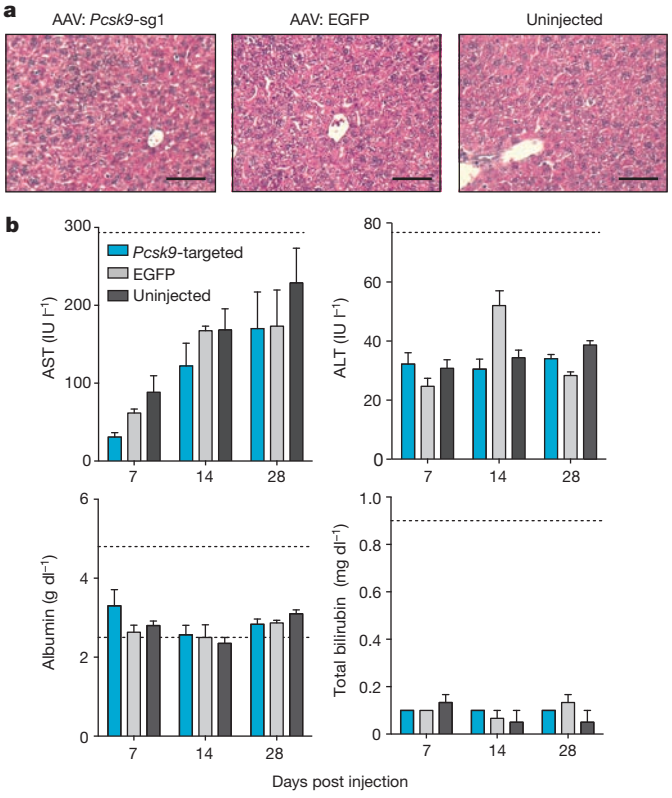


Figure 5 | Liver function tests and toxicity examination in injected animals. **a**, Histological analysis of the liver at 1 week post-injection by haematoxylin and eosin stain. Scale bars, 10 μm. **b**, Liver function tests in Pcsk9-targeted (both Pcsk9-sg1 and Pcsk9-sg2; 2×10^{11} total genome copies, $n \geq 4$), TBG–EGFP-injected (2×10^{11} total genome copies, $n = 3$), and uninjected ($n = 5$) animals. Dashed lines show the upper and lower ranges of normal value in mice where applicable.

Discussion

Here, we develop a small and efficient Cas9 from *S. aureus* for *in vivo* genome editing[17]. The results of these experiments highlight the power of using comparative genomic analysis[19,42] in expanding the CRISPR-Cas9 toolbox. Identification of new Cas9 orthologues[19,42], in addition to structure-guided engineering, could yield a repertoire of Cas9 variants with expanded capabilities and minimized molecular weight, for nucleic acid manipulation to further advance genome and epigenome engineering.

The AAV-SaCas9 system is able to mediate efficient and rapid editing of *Pcsk9* in the mouse liver, resulting in reductions of serum Pcsk9 and total cholesterol levels. To assess the specificity of SaCas9, we used an unbiased DSB detection method, BLESS, to identify a list of candidate off-target cleavage sites in a mouse cell line. We examined these sites in liver tissue transduced by AAV-SaCas9 and did not observe any indel formation within the detection limits of *in vitro* BLESS and targeted deep sequencing. Importantly, the off-target sites identified *in vitro* might differ from those *in vivo*, which need to be further evaluated by the applications of BLESS or other unbiased techniques such as those published during the revision of this work[43,44]. Finally, we did not observe any overt signs of acute toxicity in mice at one to four weeks after virus administration. Although more studies are needed to further improve the SaCas9 system for *in vivo* genome editing, such as assessing the long-term impact of Cas9 and sgRNA expression, these findings suggest that *in vivo* genome editing using SaCas9 has the potential to be highly efficient and specific.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 17 February 2014; accepted 5 February 2015.
Published online 1 April 2015.

1. Bolotin, A., Quinquis, B., Sorokin, A. & Ehrlich, S. D. Clustered regularly interspaced short palindrome repeats (CRISPRs) have spacers of extrachromosomal origin. *Microbiology* **151,** 2551–2561 (2005).
2. Barrangou, R. *et al.* CRISPR provides acquired resistance against viruses in prokaryotes. *Science* **315,** 1709–1712 (2007).
3. Garneau, J. E. *et al.* The CRISPR/Cas bacterial immune system cleaves bacteriophage and plasmid DNA. *Nature* **468,** 67–71 (2010).
4. Deltcheva, E. *et al.* CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III. *Nature* **471,** 602–607 (2011).
5. Sapranauskas, R. *et al.* The *Streptococcus thermophilus* CRISPR/Cas system provides immunity in *Escherichia coli*. *Nucleic Acids Res.* **39,** 9275–9282 (2011).
6. Jinek, M. *et al.* A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* **337,** 816–821 (2012).
7. Gasiunas, G., Barrangou, R., Horvath, P. & Siksnys, V. Cas9-crRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive immunity in bacteria. *Proc. Natl Acad. Sci. USA* **109,** E2579–E2586 (2012).
8. Cong, L. *et al.* Multiplex genome engineering using CRISPR/Cas systems. *Science* **339,** 819–823 (2013).
9. Mali, P. *et al.* RNA-guided human genome engineering via Cas9. *Science* **339,** 823–826 (2013).
10. Gaudet, D. *et al.* Review of the clinical development of alipogene tiparvovec gene therapy for lipoprotein lipase deficiency. *Atheroscler. Suppl.* **11,** 55–60 (2010).
11. Vasileva, A. & Jessberger, R. Precise hit: adeno-associated virus in gene targeting. *Nature Rev. Microbiol.* **3,** 837–847 (2005).
12. Mingozzi, F. & High, K. A. Therapeutic *in vivo* gene transfer for genetic disease using AAV: progress and challenges. *Nature Rev. Genet.* **12,** 341–355 (2011).
13. Gao, G., Vandenberghe, L. H. & Wilson, J. M. New recombinant serotypes of AAV vectors. *Curr. Gene Ther.* **5,** 285–297 (2005).
14. Kay, M. A. State-of-the-art gene-based therapies: the road ahead. *Nature Rev. Genet.* **12,** 316–328 (2011).
15. Zincarelli, C., Soltys, S., Rengo, G. & Rabinowitz, J. E. Analysis of AAV serotypes 1–9 mediated gene expression and tropism in mice after systemic injection. *Mol. Ther.* **16,** 1073–1080 (2008).
16. Swiech, L. *et al.* *In vivo* interrogation of gene function in the mammalian brain using CRISPR-Cas9. *Nature Biotechnol.* **33,** 102–106 (2015).
17. Senís, E. *et al.* CRISPR/Cas9-mediated genome engineering: an adeno-associated viral (AAV) vector toolbox. *Biotechnol. J.* **9,** 1402–1412 (2014).
18. Nishimasu, H. *et al.* Crystal structure of Cas9 in complex with guide RNA and target DNA. *Cell* **156,** 935–949 (2014).

©2015 Macmillan Publishers Limited. All rights reserved

19. Chylinski, K., Makarova, K. S., Charpentier, E. & Koonin, E. V. Classification and evolution of type II CRISPR-Cas systems. *Nucleic Acids Res.* **42,** 6091–6105 (2014).
20. Chylinski, K., Le Rhun, A. & Charpentier, E. The tracrRNA and Cas9 families of type II CRISPR-Cas immunity systems. *RNA Biol.* **10,** 726–737 (2013).
21. Hsu, P. D., Lander, E. S. & Zhang, F. Development and applications of CRISPR-Cas9 for genome engineering. *Cell* **157,** 1262–1278 (2014).
22. Hsu, P. D. *et al.* DNA targeting specificity of RNA-guided Cas9 nucleases. *Nature Biotechnol.* **31,** 827–832 (2013).
23. Hou, Z. *et al.* Efficient genome engineering in human pluripotent stem cells using Cas9 from *Neisseria meningitidis*. *Proc. Natl Acad. Sci. USA* **110,** 15644–15649 (2013).
24. Fu, Y., Sander, J. D., Reyon, D., Cascio, V. M. & Joung, J. K. Improving CRISPR-Cas nuclease specificity using truncated guide RNAs. *Nature Biotechnol.* **32,** 279–284 (2014).
25. Semenova, E. *et al.* Interference by clustered regularly interspaced short palindromic repeat (CRISPR) RNA is governed by a seed sequence. *Proc. Natl Acad. Sci. USA* **108,** 10098–10103 (2011).
26. Fu, Y. *et al.* High-frequency off-target mutagenesis induced by CRISPR-Cas nucleases in human cells. *Nature Biotechnol.* **31,** 822–826 (2013).
27. Mali, P. *et al.* CAS9 transcriptional activators for target specificity screening and paired nickases for cooperative genome engineering. *Nature Biotechnol.* **31,** 833–838 (2013).
28. Pattanayak, V. *et al.* High-throughput profiling of off-target DNA cleavage reveals RNA-programmed Cas9 nuclease specificity. *Nature Biotechnol.* **31,** 839–843 (2013).
29. Lin, Y. *et al.* CRISPR/Cas9 systems have off-target activity with insertions or deletions between target DNA and guide RNA sequences. *Nucleic Acids Res.* **42,** 7473–7485 (2014).
30. Bae, S., Park, J. & Kim, J.-S. Cas-OFFinder: a fast and versatile algorithm that searches for potential off-target sites of Cas9 RNA-guided endonucleases. *Bioinformatics* **30,** 1473–1475 (2014).
31. Wu, X. *et al.* Genome-wide binding of the CRISPR endonuclease Cas9 in mammalian cells. *Nature Biotechnol.* **32,** 670–676 (2014).
32. Kuscu, C., Arslan, S., Singh, R., Thorpe, J. & Adli, M. Genome-wide analysis reveals characteristics of off-target sites bound by the Cas9 endonuclease. *Nature Biotechnol.* **32,** 677–683 (2014).
33. Crosetto, N. *et al.* Nucleotide-resolution DNA double-strand break mapping by next-generation sequencing. *Nature Methods* **10,** 361–365 (2013).
34. Young, S. G. Recent progress in understanding apolipoprotein B. *Circulation* **82,** 1574–1594 (1990).
35. Soutschek, J. *et al.* Therapeutic silencing of an endogenous gene by systemic administration of modified siRNAs. *Nature* **432,** 173–178 (2004).
36. Rozema, D. B. *et al.* Dynamic PolyConjugates for targeted *in vivo* delivery of siRNA to hepatocytes. *Proc. Natl Acad. Sci. USA* **104,** 12982–12987 (2007).
37. Wolfrum, C. *et al.* Mechanisms and optimization of *in vivo* delivery of lipophilic siRNAs. *Nature Biotechnol.* **25,** 1149–1157 (2007).
38. Fitzgerald, K. *et al.* Effect of an RNA interference drug on the synthesis of proprotein convertase subtilisin/kexin type 9 (PCSK9) and the concentration of serum LDL cholesterol in healthy volunteers: a randomised, single-blind, placebo-controlled, phase 1 trial. *Lancet* **383,** 60–68 (2014).
39. Abifadel, M. *et al.* Mutations in *PCSK9* cause autosomal dominant hypercholesterolemia. *Nature Genet.* **34,** 154–156 (2003).
40. Cohen, J. *et al.* Low LDL cholesterol in individuals of African descent resulting from frequent nonsense mutations in *PCSK9*. *Nature Genet.* **37,** 161–165 (2005).
41. Horton, J. D., Cohen, J. C. & Hobbs, H. H. Molecular biology of PCSK9: its role in LDL metabolism. *Trends Biochem. Sci.* **32,** 71–77 (2007).
42. Briner, A. E. *et al.* Guide RNA functional modules direct Cas9 activity and orthogonality. *Mol. Cell* **56,** 333–339 (2014).
43. Tsai, S. Q. *et al.* GUIDE-seq enables genome-wide profiling of off-target cleavage by CRISPR-Cas nucleases. *Nature Biotechnol.* **33,** 187–197 (2015).
44. Frock, R. L. *et al.* Genome-wide detection of DNA double-stranded breaks induced by engineered nucleases. *Nature Biotechnol.* **33,** 179–186 (2015).
45. Crooks, G. E., Hon, G., Chandonia, J.-M. & Brenner, S. E. WebLogo: a sequence logo generator. *Genome Res.* **14,** 1188–1190 (2004).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank E. Charpentier, I. Fonfara and K. Chylinski for discussions; A. Scherer-Hoock, B. Clear and the MIT Division of Comparative Medicine for assistance with animal experiments; Boston Children's Hospital Viral Core and R. Xiao for assistance with AAV production; N. Crosetto for advice on BLESS; C.-Y. Lin and I. Slaymaker for experimental assistance; and the entire Zhang laboratory for support and advice. F.A.R. is a Junior Fellow at the Harvard Society of Fellows. W.X.Y. is supported by T32GM007753 from the National Institute of General Medical Sciences and a Paul and Daisy Soros Fellowship. J.S.G. is supported by a US Department of Energy Computational Science Graduate Fellowship. X.W. is a Howard Hughes Medical Institute International Student Research Fellow. P.A.S. is supported by United States Public Health Service grants RO1-GM34277, R01-CA133404 from the National Institutes of Health, and PO1-CA42063 from the National Cancer Institute, and partially by Cancer Center Support (core) grant P30-CA14051 from the National Cancer Institute. F.Z. is supported by the National Institutes of Health through NIMH (5DP1-MH100706) and NIDDK (5R01DK097768-03), a Waterman Award from the National Science Foundation, the Keck, New York Stem Cell, Damon Runyon, Searle Scholars, Merkin, and Vallee Foundations, and B. Metcalfe. F.Z. is a New York Stem Cell Foundation Robertson Investigator. The Children's Hospital virus core is supported by an NIH core grant (5P30EY012196-17). The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institute of General Medical Sciences or the National Institutes of Health. CRISPR reagents are available to the academic community through Addgene, and information about the protocols, plasmids, and reagents can be found at the Zhang laboratory website http://www.genome-engineering.org.

Author Contributions F.A.R. and F.Z. conceived this study. F.A.R., L.C., W.X.Y. and F.Z. designed and performed the experiments with help from all authors. F.A.R., J.S.G., O.S., K.S.M., E.V.K. and F.Z. contributed to analysis of Cas9 orthologues, crRNA and tracrRNA, and PAM. A.J.K., F.A.R., X.W., and P.A.S. led ChIP and computational analysis and validation. F.A.R., W.X.Y. and L.C. performed BLESS and targeted sequencing of BLESS-identified off-target sites, and D.A.S. contributed computational analysis of BLESS data. W.X.Y., F.A.R., L.C. and B.Z. contributed animal data. W.X.Y., F.A.R., L.C., J.S.G., and F.Z. wrote the manuscript with help from all authors.

Author Information All reagents described in this manuscript have been deposited with Addgene (plasmid IDs 61591, 61592 and 61593). Source data are available online and deep sequencing data are available at Sequence Read Archive under BioProject accession number PRJNA274149. Reprints and permissions information is available at www.nature.com/reprints. The authors declare competing financial interests: details are available in the online version of the paper. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to F.Z. (zhang@broadinstitute.org).

©2015 **Macmillan Publishers Limited. All rights reserved**

METHODS

No statistical methods were used to predetermine sample size.

***In vitro* transcription and cleavage assay.** Cas9 orthologues were human codon-optimized and synthesized by GenScript, and transfected into 293FT cells as described below. Whole-cell lysates from 293FT cells were prepared with lysis buffer (20 mM HEPES, 100 mM KCl, 5 mM MgCl2, 1 mM DTT, 5% glycerol, 0.1% Triton X-100) supplemented with Protease Inhibitor Cocktail (Roche). T7-driven sgRNA was transcribed *in vitro* using custom oligonucleotides (Supplementary Information) and HiScribe T7 *In vitro* Transcription Kit (NEB), following the manufacturer's recommended protocol. The *in vitro* cleavage assay was carried out as follows: for a 20 µl cleavage reaction, 10 µl of cell lysate was incubated with 2 µl cleavage buffer (100 mM HEPES, 500 mM KCl, 25 mM $MgCl_2$, 5 mM DTT, 25% glycerol), 1 µg *in vitro* transcribed RNA and 200 ng EcoRI-linearized pUC19 plasmid DNA or 200 ng purified PCR amplicons from mammalian genomic DNA containing target sequence. After 30 min incubation, cleavage reactions were purified using QIAquick Spin Columns and treated with RNase A at final concentration of 80 ng μl^{-1} for 30 min and analysed on a 1% agarose E-Gel (Life Technologies).

***In vitro* PAM screen.** Rho-independent transcriptional termination was predicted using the ARNold terminator search tool[46,47]. For the PAM library, a degenerate 7-bp sequence was cloned into a pUC19 vector. For each orthologue, the *in vitro* cleavage assay was carried out as above with 1 µg T7-transcribed sgRNA and 400 ng pUC19 with degenerate PAM. Cleaved plasmids were linearized by NheI, gel extracted, and ligated with Illumina sequencing adaptors. Barcoded and purified DNA libraries were quantified by Quant-iT PicoGreen dsDNA Assay Kit or Qubit 2.0 Fluorometer (Life Technologies) and pooled in an equimolar ratio for sequencing using the Illumina MiSeq Personal Sequencer (Life Technologies). MiSeq reads were filtered by requiring an average Phred quality (Q score) of at least 23, as well as perfect sequence matches to barcodes. For reads corresponding to each orthologue, the degenerate region was extracted. All extracted regions were then grouped and analysed with Weblogo[45].

Cell culture and transfection. Human embryonic kidney 293FT (Life Technologies), Neuro-2a (N2a), and Hepa1-6 (ATCC) cell lines were maintained in Dulbecco's modified Eagle's medium (DMEM) supplemented with 10% FBS (HyClone), 2 mM GlutaMAX (Life Technologies), 100 U ml^{-1} penicillin, and 100 µg ml^{-1} streptomycin at 37 °C with 5% CO_2 incubation.

Cells were seeded into 24-well plates (Corning) one day before transfection at a density of 240,000 cells per well, and transfected at 70–80% confluency using Lipofectamine 2000 (Life Technologies) following the manufacturer's recommended protocol. For each well of a 24-well plate, a total of 500 ng DNA was used. For ChIP and BLESS, a total of 4.5 million cells are seeded the day before transfection into a 100-mm plate, and a total of 20 µg DNA was used.

DNA isolation from cells and tissue. Genomic DNA was extracted using the QuickExtract DNA Extraction Solution (Epicentre). Briefly, pelleted cells were resuspended in QuickExtract solution and incubated at 65 °C for 15 min, 68 °C for 15 min, and 98 °C for 10 min (ref. 8). Genomic liver DNA was extracted from bulk tissue fragments using a microtube bead mill homogenizer (Beadbug, Denville Scientific) by homogenizing approximately 30–50 mg of tissue in 600 µl of DPBS (Gibco). The homogenate was then centrifuged at 2,000 to 3,000*g* for 5 min at 4 °C and the pellet was resuspended in 300–600 µl QuickExtract DNA Extraction Solution (Epicentre) and incubated as above.

Indel analysis and guide:target base-pairing mismatch search. Indel analyses by SURVEYOR assay and targeted deep sequencing were carried out and analysed as previously described[8,22]. The methods for identification of potential off-target sites for SpCas9 based on Watson–Crick base-pairing mismatch between guide RNA and target DNA has been previously described[22], and adapted for SaCas9 by considering NNGRR for possible off-target PAMs. Alignment was manually adjusted to allow for insertion and deletion mismatches in the guide:target heteroduplex[29,30].

Chromatin immunoprecipitation and analysis. Cells were passaged at 24 h post-transfection into a 150-mm dish, and fixed for ChIP processing at 48 h post-transfection. For each condition, 10 million cells are used for ChIP input, following experimental protocols and analyses as previously described[31] with the following modifications: instead of pairwise peak-calling, ChIP peaks were only required to be enriched over both 'empty' controls (dSpCas9 only, dSaCas9 only) as well as the other Cas9/other sgRNA sample (for example, SpCas9/EMX-sg2 peaks must be enriched over SaCas9/EMX-sg1 peaks in addition to the empty controls). This was done to avoid filtering out of real peaks present in two related samples as much as possible.

To identify off-targets ranked by motif or sequence similarity to guide, motif scores for ChIP peaks were calculated as follows: for a given ChIP peak, the 100-nucleotide interval around the peak summit, the target sequence, and a given sgRNA guide region of length L, the query, an alignment score is calculated for every subsequence of L in the target. The subsequence with the highest score is reported as the best match to the query. For each subsequence alignment, the score calculation begins at the 5′ end of the query. For each position in the alignment, 1 is added or subtracted for match or mismatch between the query and target, respectively. If the score becomes negative, it is set to 0 and the calculation continued for the remainder of the alignment. The score at the 3′ end of the query is reported as the final score for the alignment. MACS scores $= -10\log(P$ value relative to the empty control) are determined as previously described[48]. For unbiased determination of PAM from ChIP peaks, the peaks were analysed for the best match by motif score to the guide region only within 50 nucleotides of the peak summit; the alignment was extended for 10 nucleotides at the 3′ end and visualized using Weblogo[45].

To calculate the motif score threshold at which false discovery rate < 0.1 for each sample, 100-nucleotide sequences centred around peak summits were shuffled while preserving dinucleotide frequency. The best match by motif score to the guide+PAM (NGG for SpCas9, NNGRRT for SaCas9) in these shuffled sequences was then found. The score threshold for false discovery rate < 0.1 was defined as the score such that less than 10% of shuffled peaks had a motif score above that score threshold.

BLESS for DSB detection. Cells are harvested at 24 h post-transfection, then processed as previously described[33] with the following alterations: a total of 10 million cells are fixed for nuclei isolation and permeabilization, and treated with Proteinase K for 4 min at 37 °C before inactivation with PMSF. All deproteinized nuclei are used for DSB labelling with 100 mM of annealed proximal linkers overnight. After Proteinase K digestion of labelled nuclei, chromatin was mechanically sheared with a 26G needle before sonication (BioRuptor, 20 min on high, 50% duty cycle). 20 µg of sheared chromatin are captured on streptavidin beads, washed, and ligated to 200 mM of distal linker. Linker hairpins are then cleaved off with I-SceI digestion for 1 h at 37 °C, and products PCR-enriched for 18 cycles before proceeding to library preparation with TruSeq Nano LT Kit (Illumina). For the negative control, cells mock transfected with Lipofectamine 2000 and pUC19 DNA were parallel processed through the assay.

BLESS analysis. Fastq files were demultiplexed, and 30-bp genomic sequences were separated from the BLESS ligation handles for alignment. Bowtie was used to map the genomic sequences to hg19 or mm9, allowing for a maximum of 2 mismatches. Following alignment, reads from all bio-replicates for an individual sample were first pooled, and then nearest neighbour clustering was performed with a 30-bp moving window to identify regions of enrichment across the genome. Within each cluster, the pairwise distance was calculated between all forward and reverse read strand mappings (Extended Data Fig. 7b, c). Pairwise distance distributions were used to filter out wide and poorly defined DSB clusters from the well-defined DSB clusters characteristically found at Cas9-induced cleavage sites (see Supplementary Information). Finally, we adjusted the count of predicted Cas9-induced DSBs at a given locus by using a binomial model to calculate the maximum-likelihood estimate of peak enrichment in the Cas9-sgRNA treated sgRNAs given BLESS measurements from an untreated negative control. After the maximum-likelihood estimate calculation, a list of loci ranked by their DSB scores could be obtained and plotted (Fig. 3b, Extended Data Fig. 8). Additional descriptions can be found in Supplementary Information.

The top-ranking ~30 sites from the list of Cas9 induced DSB clusters were sequenced for indel formation (Extended Data Fig. 8; validated targets in Fig. 3d). Within these loci, PAMs and regions of target homology were identified by first searching all PAM sites within a ± 50 bp window around the DSB cluster, then selecting the adjacent sequence with fewest mismatches to the target sequence.

Code availability. BLESS analysis code is available at https://github.com/fengzhanglab/BLESS.

Virus production and titration. For in-house viral production, 293FT cells (Life Technologies) were maintained as described above in 150 mm plates. For each transfection, 8 µg of pAAV8 serotype packaging plasmid, 10 µg of pDF6 helper plasmid, and 6 µg of AAV2 plasmid carrying the construct of interest were added to 1 mL of serum-free DMEM. 125 µl of PEI "Max" solution (1 mg ml^{-1}, pH = 7.1) was then added to the mixture and incubated at room temperature for 5 to 10 s. After incubation, the mixture was added to 20 ml of warm maintenance media and applied to each dish to replace the old growth media. Cells were harvested between 48 h and 72 h post transfection by scraping and pelleting by centrifugation. The AAV2/8 (AAV2 inverted terminal repeat (ITR) vectors pseudo-typed with AAV8 capsid) viral particles were then purified from the pellet according to a previously published protocol[49].

High titre and purity viruses were also produced by vector core facilities at Children's Hospital Boston and Massachusetts Eye and Ear Infirmary (MEEI). These AAV vectors were then titred by real-time qPCR using a customized TaqMan probe against the transgene, and all viral preparations were titre-matched across different batches and production facilities before experiments. The purity of AAV vector was further verified by SDS–PAGE.

©2015 Macmillan Publishers Limited. All rights reserved

Animal injection and processing. All mice cohorts were maintained at animal facility with standard diet and housing following IRB-approved protocols. AAV vector was delivered to 5–6 week old male C57/BL6 mice intravenously via lateral tail vein injection. All dosages of AAV were adjusted to 100 µl or 200 µl with sterile phosphate buffered saline (PBS), pH 7.4 (Gibco) before the injection. Animals were not immunosuppressed or otherwise handled differently before injection or during the course of the experiment except the pre-bleed fasting as noted below. The animals were randomized to the different experimental conditions, with the investigator not blinded to the assignments.

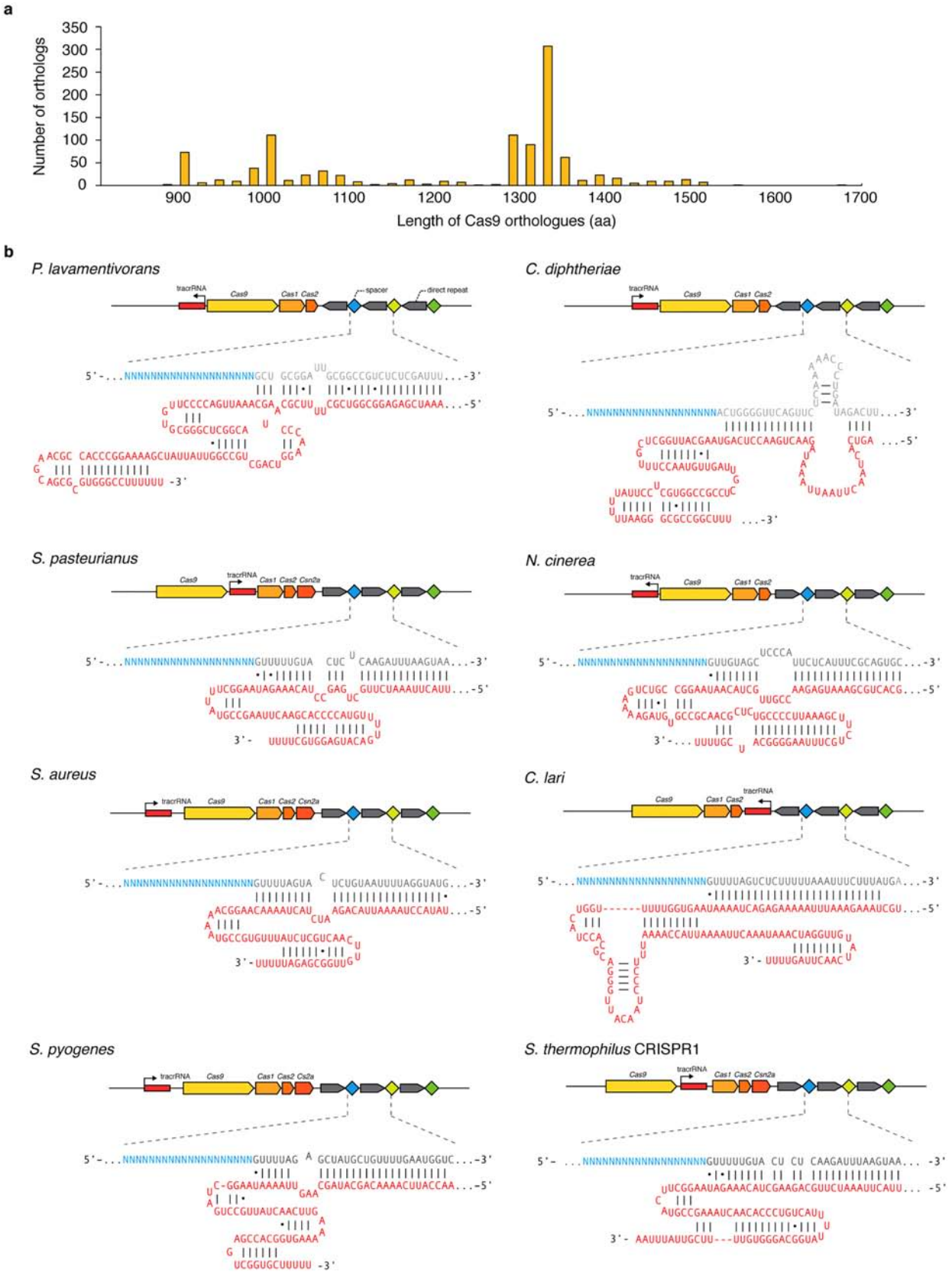
To track the serum levels of Pcsk9 and total cholesterol, animals were fasted overnight for 12 h before blood collection by saphenous vein bleeds (no more than 100 µl or 10% of total blood volume per week). Multiple bleeds were made before tail vein delivery of AAV vector or control to collect pre-injection samples and to habituate the animals to handling during the procedure. After the blood was allowed to clot at room temperature, the serum was separated by centrifugation and stored at −20 °C for subsequent analysis. For terminal procedures to collect liver tissue and larger serum volumes for chemistry panels, mice were euthanized by carbon dioxide inhalation. Subsequently, blood was collected via cardiac puncture. Transcardial perfusion with 30 ml PBS removed the remaining blood, after which liver samples were collected. The median lobe of liver was removed and fixed in 10% neutral buffered formalin for histological analysis, while the remaining lobes were sliced in small blocks of size less than $1 \times 1 \times 3\ mm^3$ and frozen for subsequent DNA or protein extraction.

Histology and serum analysis. Following tissue harvesting as described above, flash-frozen mouse liver samples were embedded in OCT compound (Tissue Tek, Cat # 4583), snap-frozen, and stored at −80 °C before processing. Frozen tissues were cryosectioned at 4 µm in thickness and stained with Oil Red O following manufacturer's recommended protocol. Liver histology was assessed by H&E staining sections of 10% neutral buffer formalin fixed liver sections.

Serum levels of Pcsk9 were determined by ELISA using the Mouse Proprotein Convertase 9/PCSK9 Quantikine ELISA Kit (MPC-900, R&D Systems), following the manufacturer's instructions. Total cholesterol levels were measured using the Infinity Cholesterol Reagent (Thermo Fisher) per the manufacturer's instructions. Serum ALT, AST, albumin and total bilirubin were measured by an Olympus AU5400 (IDEXX Memphis, TN).

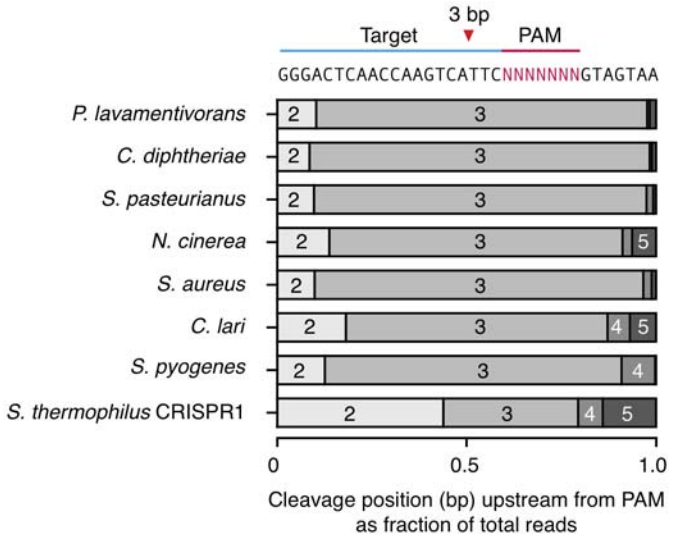
46. Gautheret, D. & Lambert, A. Direct RNA motif definition and identification from multiple sequence alignments using secondary structure profiles. *J. Mol. Biol.* **313,** 1003–1011 (2001).
47. Macke, T. J. *et al.* RNAMotif, an RNA secondary structure definition and search algorithm. *Nucleic Acids Res.* **29,** 4724–4735 (2001).
48. Zhang, Y. *et al.* Model-based analysis of ChIP-seq (MACS). *Genome Biol.* **9,** R137 (2008).
49. Veldwijk, M. R. *et al.* Development and optimization of a real-time quantitative PCR-based method for the titration of AAV-2 vector stocks. *Mol. Ther.* **6,** 272–278 (2002).
50. Zuker, M. Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res.* **31,** 3406–3415 (2003).

©2015 **Macmillan Publishers Limited. All rights reserved**



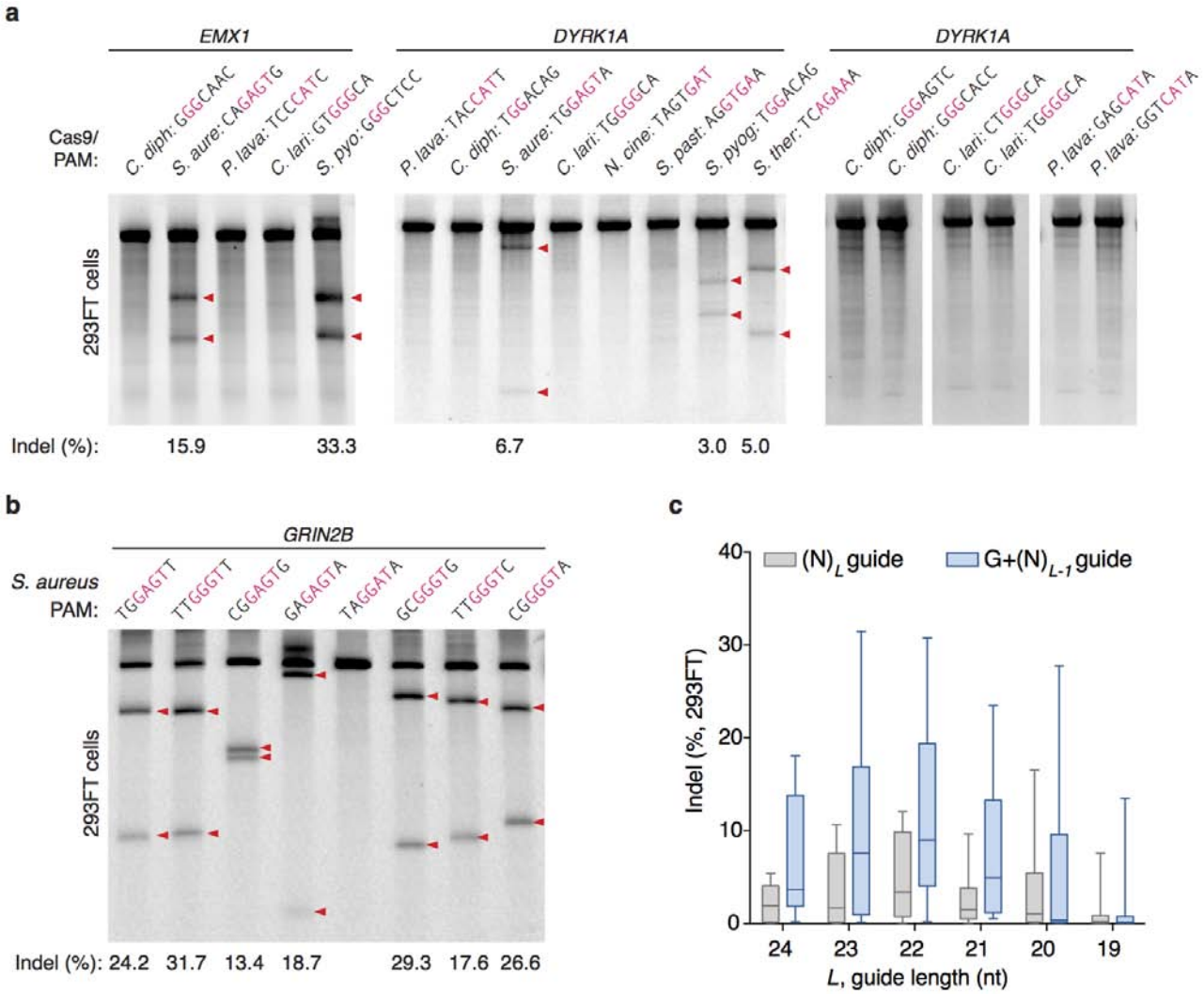
Extended Data Figure 1 | Selection of Type II CRISPR-Cas loci from eight bacterial species. a, Distribution of lengths for Cas9 > 600 Cas9 orthologues[19]. **b**, Schematic of Type II CRISPR-Cas loci and sgRNA from eight bacterial species. Spacer or 'guide' sequences are shown in blue, followed by direct repeats (grey). Predicted tracrRNAs are shown in red, and folded based on the Constraint Generation RNA folding model[50].

©2015 Macmillan Publishers Limited. All rights reserved



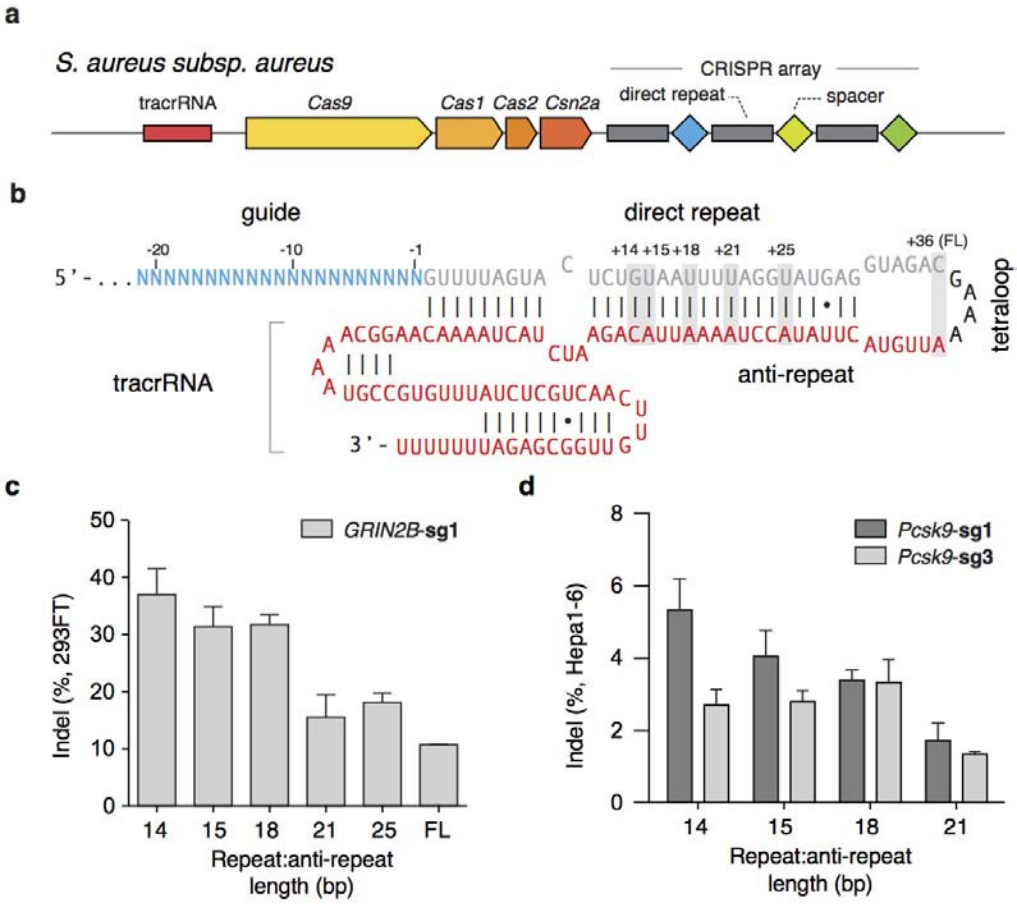
Extended Data Figure 2 | Cas9 orthologue cleavage pattern *in vitro*. Stacked bar graph indicates the fraction of targets cleaved at 2, 3, 4, or 5 bp upstream of PAM for each Cas9 orthologue; most Cas9 enzymes cleave stereotypically at 3 bp upstream of PAM (red triangle).

©2015 Macmillan Publishers Limited. All rights reserved



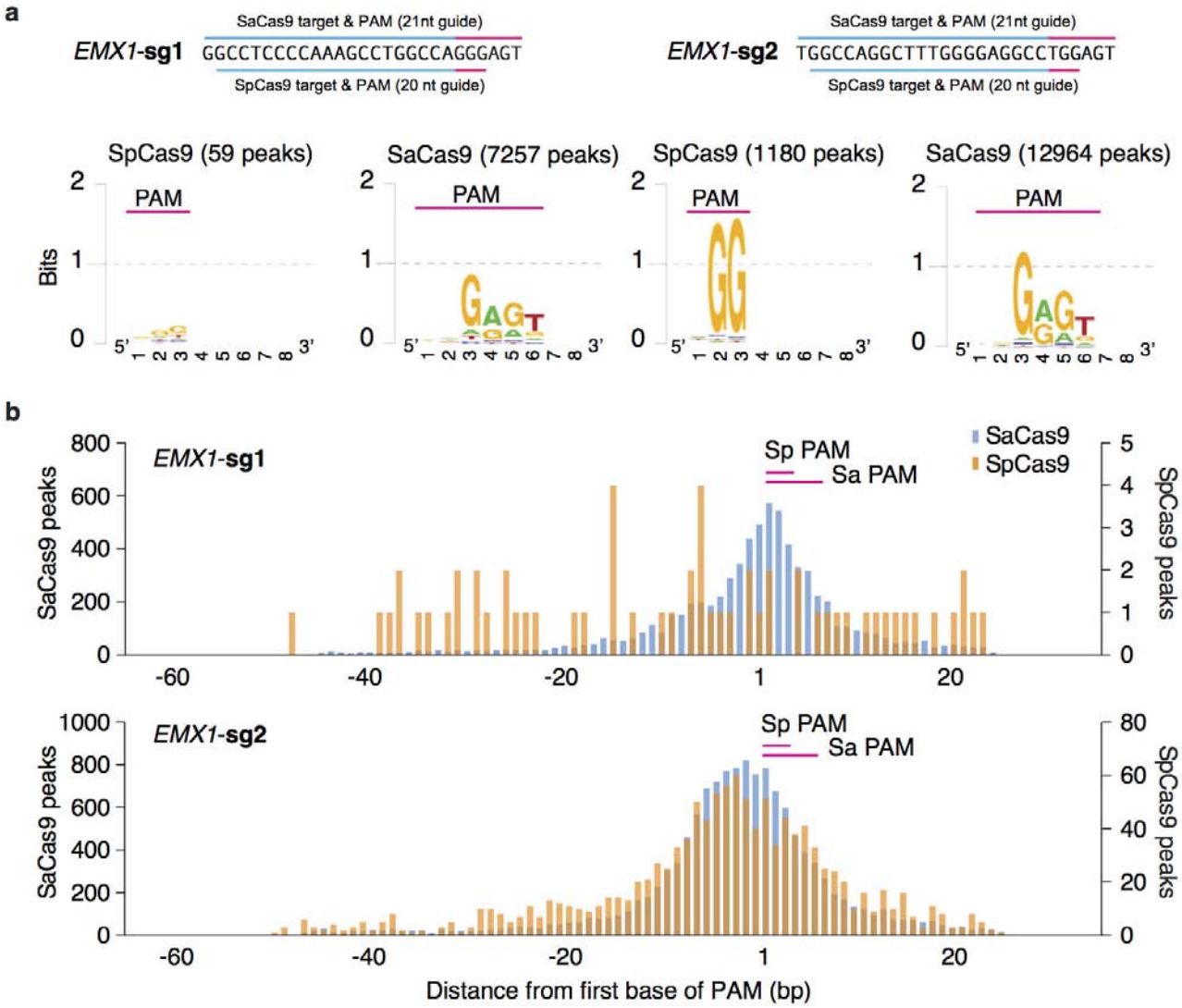
Extended Data Figure 3 | Test of Cas9 orthologue activity in 293FT cells. **a**, SURVEYOR assays showing indel formation at human endogenous loci from co-transfection of Cas9 orthologues and sgRNA. PAM sequences for individual targets are shown above each lane, with the consensus region for each PAM highlighted in red. Red triangles indicate cleaved fragments. **b**, SaCas9 generates indels efficiently for a multiple targets. **c**, Box-whisker plot of indel formation as a function of SaCas9 guide length *L*, with unaltered guides (perfect match of *L* nucleotides, grey bars) or replacement of the 5′-most base of guide with guanine (G + *L* − *1* nucleotides, blue bars) ($n = 8$ guides).

©2015 Macmillan Publishers Limited. All rights reserved



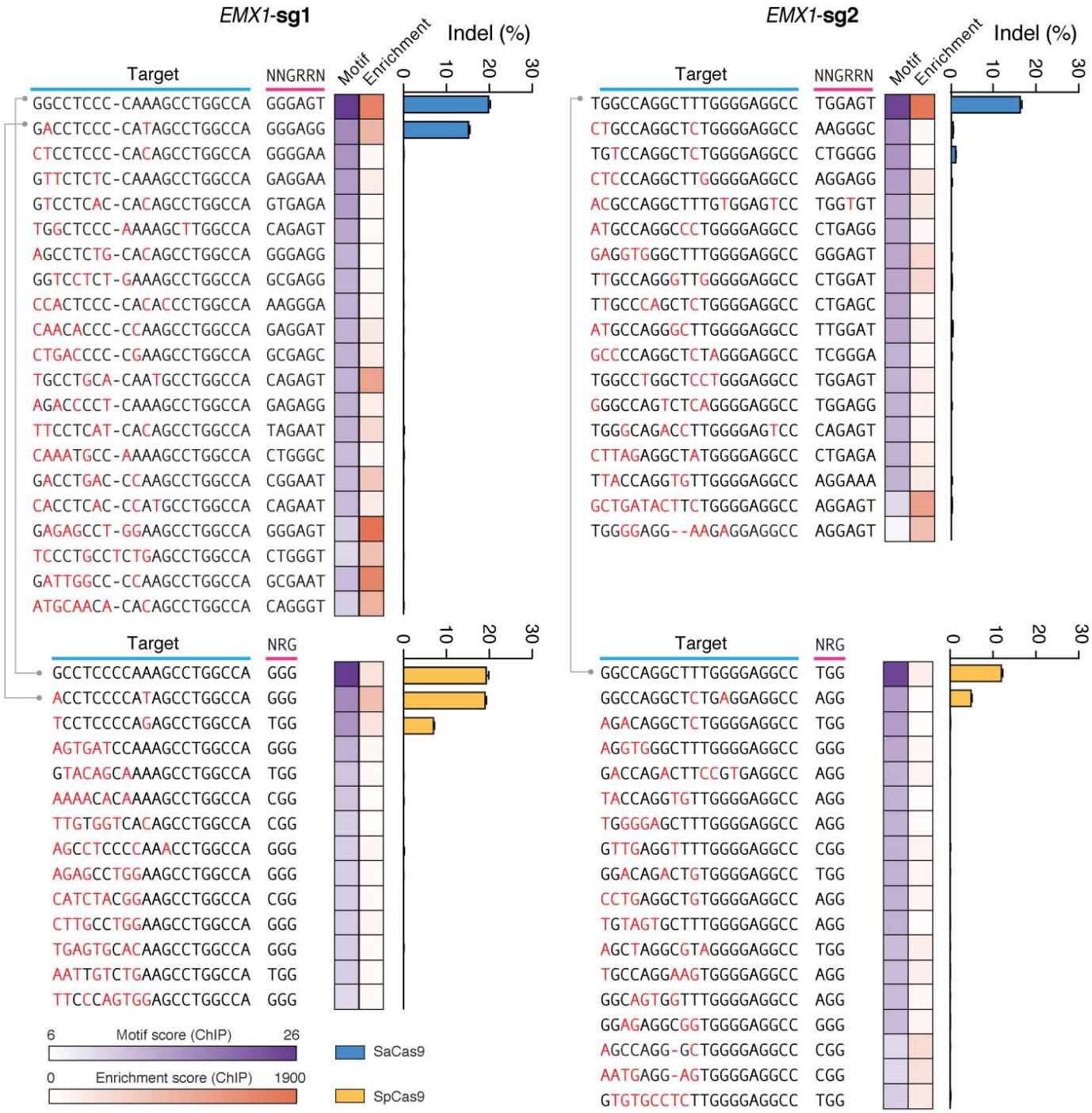
Extended Data Figure 4 | Optimization of SaCas9 sgRNA scaffold in mammalian cells. **a**, Schematic of the *Staphylococcus aureus subspecies aureus* CRISPR locus. **b**, Schematic of SaCas9 sgRNA with 21-nucleotide guide, crRNA repeat (grey), tetraloop (black) and tracrRNA (red). The number of crRNA repeat to tracrRNA anti-repeat base-pairing is indicated above the grey boxes. SaCas9 cleaves targets with varying repeat:anti-repeat lengths in **c**, HEK 293FT and **d**, Hepa1-6 cell lines. ($n = 3$, error bars show s.e.m.)

©2015 Macmillan Publishers Limited. All rights reserved



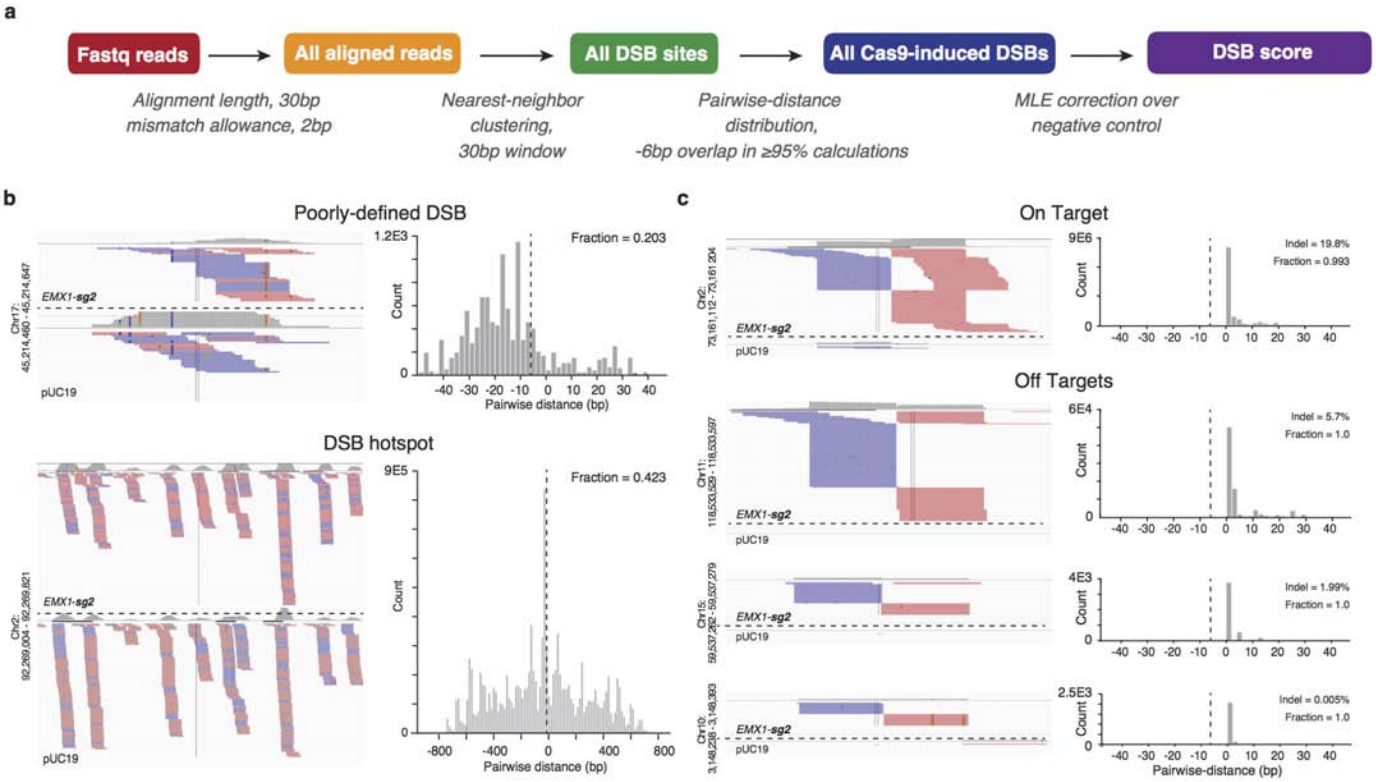
Extended Data Figure 5 | Genome-wide binding by Cas9-chromatin immunoprecipitation (dCas9-ChIP). **a**, Unbiased identification of PAM motif for dSaCas9 and dSpCas9. Peaks were analysed for the best match by motif score to the guide region only within 50 nucleotides of the peak summit. The alignment was extended for 10 nucleotides at the 3′ end and visualized using Weblogo. Numbers in parentheses indicate the number of called peaks. **b**, Histograms show the distribution of the peak summit relative to motif for dSaCas9 and dSpCas9. Position 1 on *x* axis indicates the first base of PAM.

©2015 Macmillan Publishers Limited. All rights reserved



Extended Data Figure 6 | Indel measurements at candidate off-target sites based on ChIP. Indels at top off-target sites predicted by dCas9-ChIP for each Cas9 and sgRNA pair, based on ChIP peaks ranked by sequence similarity of the genomic loci to the guide motif (heat map in purple), or *P* value of ChIP enrichment over control (heat map in red). Lines connect the common targets (*EMX1*) and off-targets between the two Cas9 enzymes.

©2015 Macmillan Publishers Limited. All rights reserved



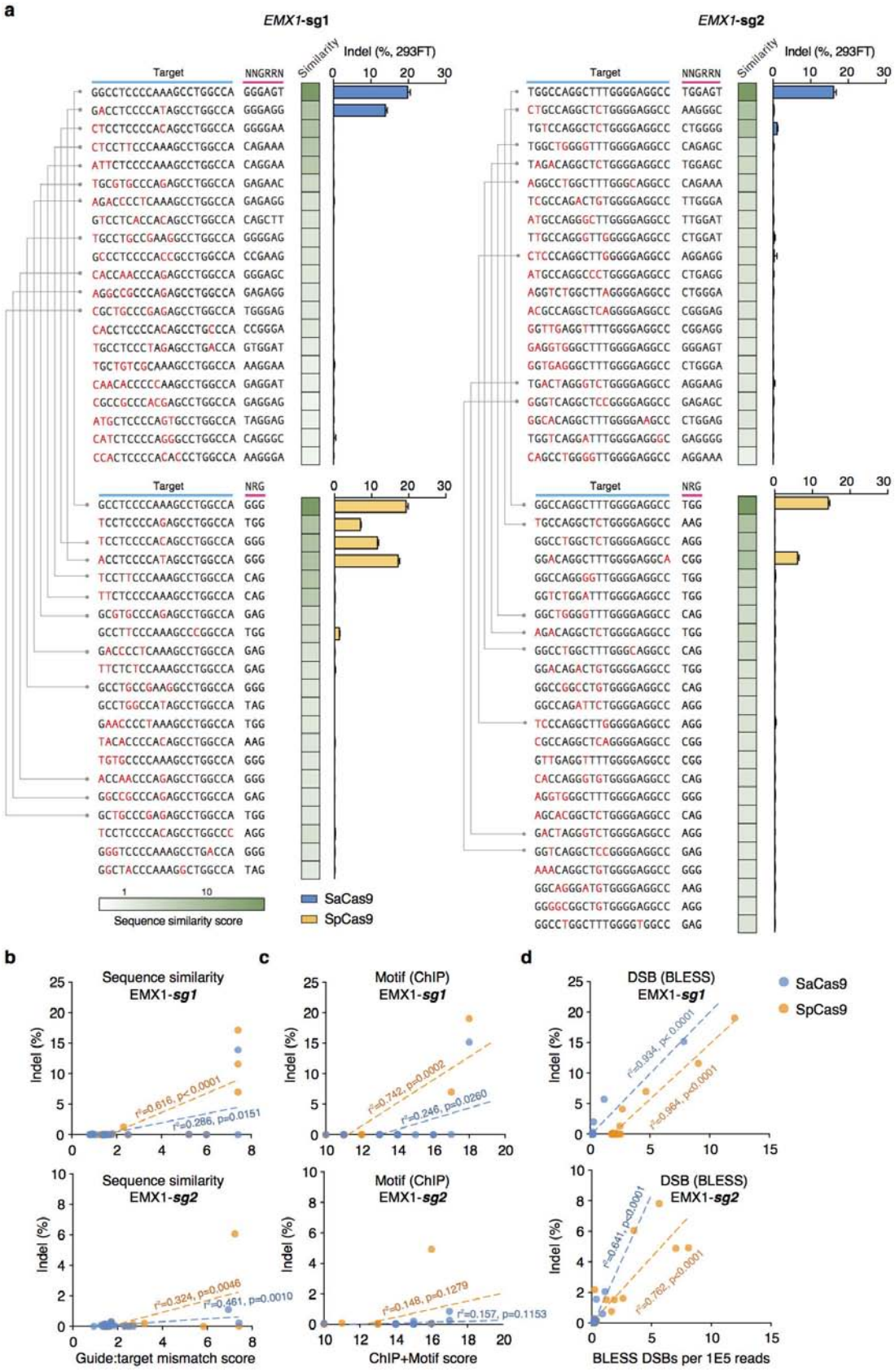
Extended Data Figure 7 | Analysis pipeline of sequencing data from BLESS. **a**, Overview of the data analysis pipeline starting from the raw sequencing reads. Representative sequencing read mappings and corresponding histograms of the pairwise distances between all the forward orientation (red) reads and reverse orientation (blue) reads, displayed for representative **b**, DSB hotspots and poorly defined DSB sites and **c**, Cas9-induced DSBs with detectable indels. Fraction of pairwise distances between reads overlapping by no more than 6 bp (dashed vertical line) are indicated over histogram plots.

©2015 Macmillan Publishers Limited. All rights reserved



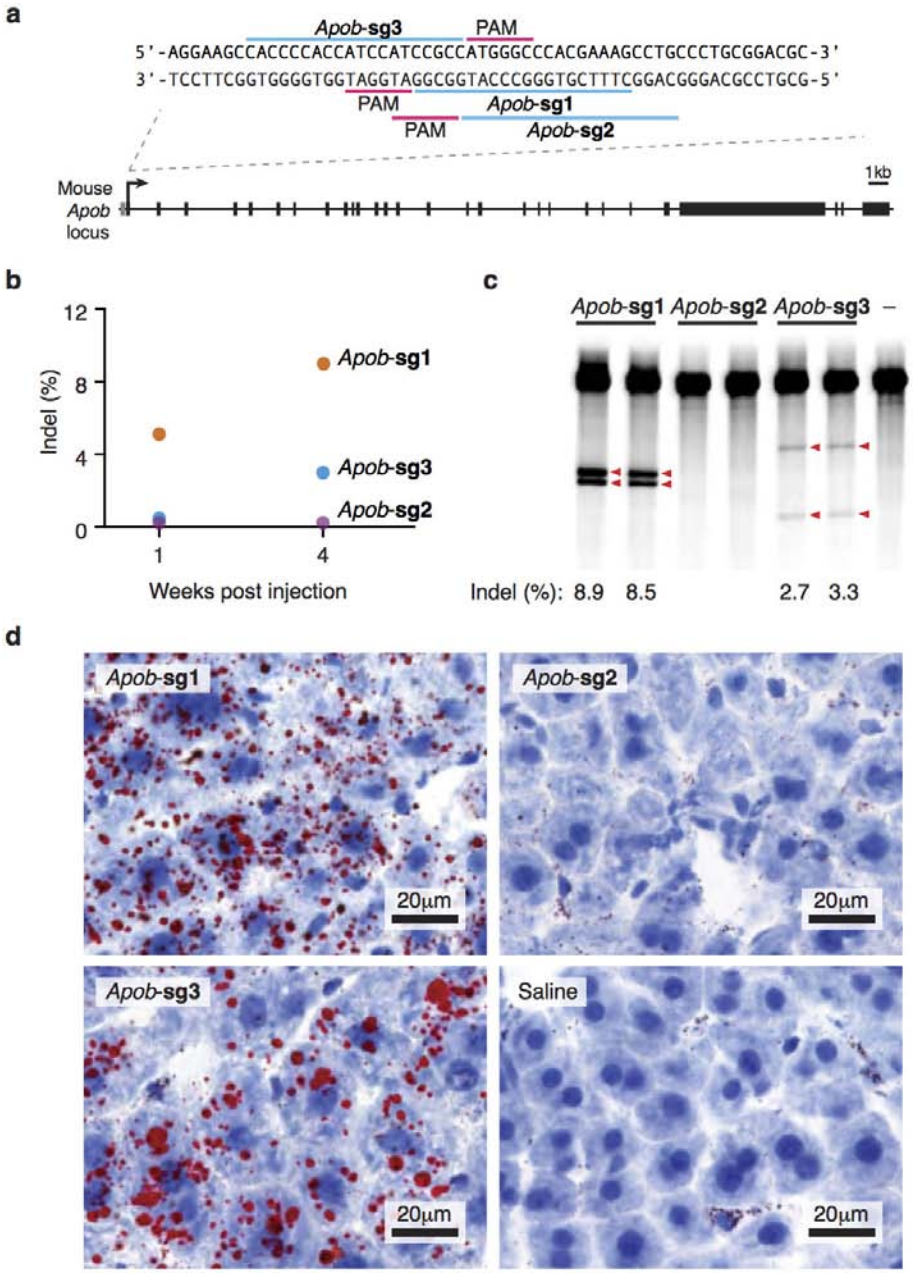
Extended Data Figure 8 | Indel measurements at off-target sites based on DSB scores. List of top off-target sites ranked by DSB scores for each Cas9 and sgRNA pair. Indel levels are determined by targeted deep sequencing. Blue triangles indicate positions of peak BLESS signal, and where present, PAMs and targets with sequence homology to the guide are highlighted. Lines connect the common on-targets (*EMX1*) and off-targets between the two Cas9 enzymes. N.D., not determined.

©2015 Macmillan Publishers Limited. All rights reserved



Extended Data Figure 9 | Indel measurements of top candidate off-target sites based on sequence similarity score. Off-targets are predicted based on sequence similarity to on-target, accounting for number and position of Watson–Crick base-pairing mismatches as previously described[22]. NNGRR and NRG are used as potential PAMs for SaCas9 and SpCas9, respectively. Lines connect the common targets (*EMX1*) and off-targets between the two Cas9 enzymes. Correlation plots between indel percentages and **b**, prediction based on sequence similarity, **c**, ChIP peaks ranked by motif similarity, or **d**, DSB scores for top ranking off-target loci. Trendlines, r^2, and P values are calculated using ordinary least squares.

©2015 Macmillan Publishers Limited. All rights reserved



Extended Data Figure 10 | SaCas9 targeting *Apob* locus in the mouse liver. **a**, Schematics illustrating the mouse *Apob* gene locus and the positions of the three guides tested. **b**, Experimental time course and **c**, SURVEYOR assay showing indel formation at target loci after intravenous injection of AAV2/8 carrying thyroxine-binding globulin (TBG) promoter-driven SaCas9 and U6-driven guide at 2×10^{11} total genome copies ($n = 1$ animal each). **d**, Oil-red staining of liver tissue from AAV- or saline-injected animals. Male C56BL/6 mice were injected at 8 weeks of age and analysed 4 weeks post injection.

©2015 Macmillan Publishers Limited. All rights reserved