LETTER

https://doi.org/10.1038/s41586-018-0500-9

In vivo CRISPR editing with no detectable genome-wide off-target mutations

Pinar Akcakaya[1,13], Maggie L. Bobbin[2,3,4,13], Jimmy A. Guo[2,3], Jose Malagon-Lopez[2,3,4], Kendell Clement[2,3,4], Sara P. Garcia[2], Mick D. Fellows[5], Michelle J. Porritt[1], Mike A. Firth[6], Alba Carreras[1,9], Tania Baccega[1,10], Frank Seeliger[7], Mikael Bjursell[1], Shengdar Q. Tsai[2,3,4,11], Nhu T. Nguyen[2,3], Roberto Nitsch[8], Lorenz M. Mayr[1,12], Luca Pinello[2,4], Mohammad Bohlooly-Y[1], Martin J. Aryee[2,4], Marcello Maresca[1]* & J. Keith Joung[2,3,4]*

CRISPR–Cas genome-editing nucleases hold substantial promise for developing human therapeutic applications[1–6] but identifying unwanted off-target mutations is important for clinical translation[7]. A well-validated method that can reliably identify off-targets in vivo has not been described to date, which means it is currently unclear whether and how frequently these mutations occur. Here we describe 'verification of in vivo off-targets' (VIVO), a highly sensitive strategy that can robustly identify the genome-wide off-target effects of CRISPR–Cas nucleases in vivo. We use VIVO and a guide RNA deliberately designed to be promiscuous to show that CRISPR–Cas nucleases can induce substantial off-target mutations in mouse livers in vivo. More importantly, we also use VIVO to show that appropriately designed guide RNAs can direct efficient in vivo editing in mouse livers with no detectable off-target mutations. VIVO provides a general strategy for defining and quantifying the off-target effects of gene-editing nucleases in whole organisms, thereby providing a blueprint to foster the development of therapeutic strategies that use in vivo gene editing.

VIVO consists of two steps (Fig. 1a). In an initial in vitro 'discovery' step, a superset of potential off-target cleavage sites for a nuclease is identified using circularization for in vitro reporting of cleavage effects by sequencing (CIRCLE-seq)[8]. This method is highly sensitive, avoids potential confounding effects associated with cell-based assays[8] and can successfully identify supersets of sites that include bona fide off-targets in cultured human cells[8]. In a second in vivo 'confirmation' step, sites identified by CIRCLE-seq are examined for indel mutations in target tissues that have been treated with the nuclease.

To test VIVO, we designed a *Streptococcus pyogenes* Cas9 (hereafter, Cas9) guide RNA (gRNA) targeted to the mouse *Pcsk9* gene that we expected would have a high likelihood of inducing multiple off-target mutations in the mouse genome (Extended Data Fig. 1a, Extended Data Table 1 and Methods). To deliver this 'promiscuous' gRNA (gP) and Cas9 to mouse livers in vivo, we infected cohorts of mice with an adenoviral vector that encodes these components or with a negative control vector that encodes GFP and Cas9. Adenoviral vectors have a known biodistribution, which includes efficient delivery to the liver (Extended Data Fig. 2). Two strains of mice were infected: a wild-type C57BL/6N strain (hereafter, wild-type mice) and littermates that contain a single copy of a human *PCSK9* open reading frame knocked into the *Rosa26* locus (hereafter, knock-in mice; A.C. et al., manuscript submitted; Extended Data Fig. 3 and Methods). gP–Cas9 induced stable and efficient mutation of the on-target *Pcsk9* site and reductions in Pcsk9 protein levels in plasma, in both wild-type and knock-in mice (Extended Data Fig. 1b, c).

Having established the efficacy of gP–Cas9 for on-target *Pcsk9* modification in vivo, we conducted the first screening step of VIVO by performing CIRCLE-seq with gP–Cas9 on liver genomic DNA from wild-type and knock-in mice (Fig. 1a and Methods). We identified many off-target cleavage sites in vitro: 3,107 and 2,663 sites with wild-type and knock-in mice genomes, respectively (Extended Data Fig. 4 and Supplementary Table 1). These sites represent only a small percentage of all sites in the genome that have seven or fewer mismatches relative to the on-target site (Extended Data Fig. 5). There were 2,368 sites that were identified in both mouse genomes and that showed strong concordance ($r^2 = 0.902$) in their CIRCLE-seq read counts (Extended Data Fig. 4), which semi-quantitatively reflect cleavage efficiency[8]. Sites identified in only one or the other genome generally had among the lowest CIRCLE-seq read counts (Extended Data Fig. 4). This is consistent with the possibility that these sites may or may not be detected by chance alone because they lie at the assay limit of detection, as has previously been observed with analogous experiments in human cells in culture[8]. Single nucleotide polymorphisms or indels might also have a role in accounting for a very small subset of these sites (Extended Data Table 2). The 20 sites with the highest CIRCLE-seq read counts all had 3 or fewer mismatches in the spacer sequence (Extended Data Fig. 4 and Supplementary Table 1). Many off-target sites contained protospacer adjacent motif (PAM) mismatches, with NAG as the most prevalent (Extended Data Fig. 4 and Supplementary Table 1)—probably because gP has an unusually high relative number of closely matched sites in the mouse genome with a NAG PAM (Extended Data Table 3), a known alternative PAM for Cas9[8,9].

To perform the second step of VIVO, we assessed whether the off-target cleavage sites of gP–Cas9 identified by CIRCLE-seq showed evidence of indels in vivo in the livers of wild-type and knock-in mice treated with gP–Cas9 and control GFP–Cas9 adenoviral vectors. Because of the very large number of sites identified by CIRCLE-seq, we performed targeted amplicon sequencing using liver genomic DNA from mice euthanized on day four and on week three after infection on only the following subsets of sites: the *Pcsk9* on-target site, 11 'class I' sites with the highest CIRCLE-seq read counts (containing 1–3 mismatches relative to the on-target site), 17 'class II' sites with moderate CIRCLE-seq read counts (containing 2–4 mismatches) and 17 'class III' sites with lower CIRCLE-seq read counts (containing 1–6 mismatches) (Fig. 1b and Extended Data Fig. 4). The *Pcsk9* on-target site was efficiently mutagenized and remained stably mutated in both wild-type and knock-in mice (mean indel frequencies ranging from about 23 to 30%) (Fig. 1b and Supplementary Table 2). Of the

[1]Discovery Biology, Discovery Sciences, IMED Biotech Unit, AstraZeneca, Gothenburg, Sweden. [2]Molecular Pathology Unit and Center for Cancer Research, Massachusetts General Hospital, Charlestown, MA, USA. [3]Center for Computational and Integrative Biology, Massachusetts General Hospital, Charlestown, MA, USA. [4]Department of Pathology, Harvard Medical School, Boston, MA, USA. [5]Advanced Medicines Safety, Drug Safety and Metabolism, IMED Biotech Unit, AstraZeneca, Cambridge, UK. [6]Quantitative Biology, Discovery Sciences, IMED Biotech Unit, AstraZeneca, Cambridge, UK. [7]Pathology Science, Drug Safety and Metabolism, IMED Biotech Unit, AstraZeneca, Gothenburg, Sweden. [8]Advanced Medicines Safety, Drug Safety and Metabolism, IMED Biotech Unit, AstraZeneca, Gothenburg, Sweden. [9]Present address: Wallenberg Laboratory and Sahlgrenska Center for Cardiovascular and Metabolic Research, Department of Molecular and Clinical Medicine, University of Gothenburg, Gothenburg, Sweden. [10]Present address: San Raffaele Telethon Institute for Gene Therapy, IRCCS San Raffaele Scientific Institute, Milan, Italy. [11]Present address: Department of Hematology, St. Jude Children's Research Hospital, Memphis, TN, USA. [12]Present address: GE Healthcare Life Sciences, The Grove Centre, Amersham, UK. [13]These authors contributed equally: Pinar Akcakaya, Maggie L. Bobbin. *e-mail: marcello.maresca@astrazeneca.com; jjoung@mgh.harvard.edu

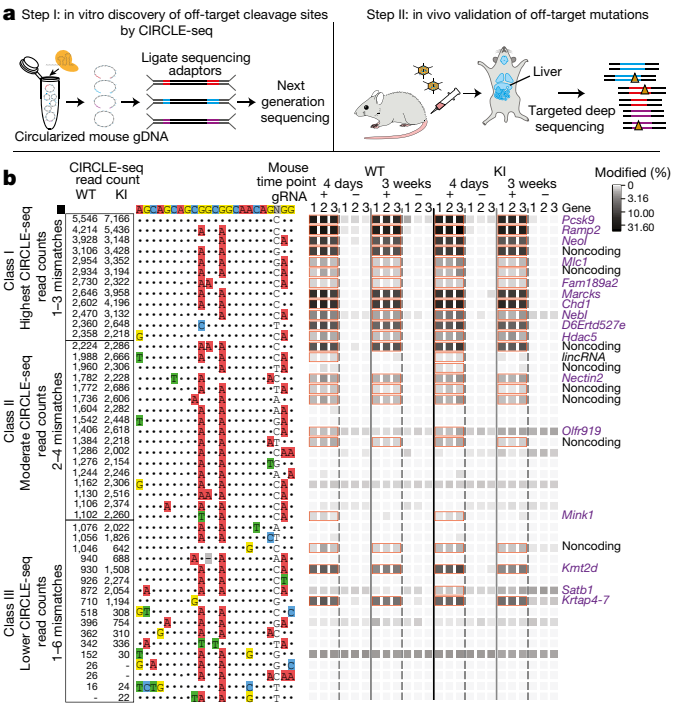


Fig. 1 | Overview and validation of VIVO. a, Schematic illustrating the two-step VIVO method. In step I, CIRCLE-seq identifies off-target sites cleaved in vitro. In step II, the sites identified in step I are assessed in vivo for indel mutations by targeted amplicon sequencing performed with genomic DNA isolated from the livers of nuclease-treated mice. **b**, Assessment of in vivo off-target indels induced by gP–Cas9. Indel frequencies as determined by targeted amplicon sequencing are presented as heat maps for the gP–Cas9 on-target site (black square) and the class I, class II and class III off-target sites (identified from CIRCLE-seq experiments). Each locus was assayed in $n = 3$ biologically independent wild-type (WT) and knock-in (KI) mice (labelled as 1, 2 and 3) using genomic DNA isolated from the liver of mice treated with experimental adenoviral vector that encodes gP–Cas9 (gRNA +) or control adenoviral vector that encodes GFP–Cas9 (gRNA −). Mismatches relative to the on-target site are shown with coloured boxes. CIRCLE-seq read-count numbers for each site are shown. Sites that showed a significant difference between the experimental (gRNA +) and control (gRNA −) samples are outlined with orange boxes and labelled by genomic locus, with coding regions shown in purple text. *P* values and significance were obtained by fitting a negative binomial generalized linear model (for source data and *P* values, see Supplementary Table 2).

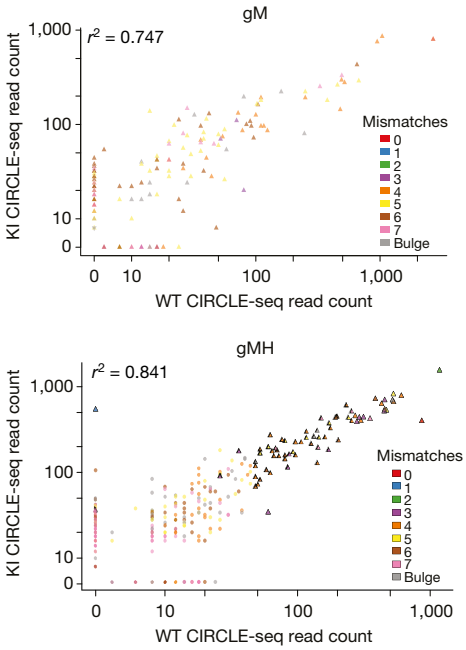


Fig. 2 | Characterization of *Pcsk9*-targeted gRNAs designed to be orthogonal to the mouse genome by CIRCLE-seq. Scatter plots of CIRCLE-seq read counts for off-target cleavage sites identified in vitro with gM–Cas9 and gMH–Cas9 on genomic DNA from wild-type ($n = 1$) and knock-in ($n = 1$) mice (for source data, see Supplementary Tables 3, 4). Each site is colour-coded for the number of mismatches it has relative to the on-target site. Sites represented as triangles were chosen for targeted amplicon sequencing. Correlation r^2 values shown were obtained by linear regression performed using all values in each scatter plot.

45 sites we examined, 19 showed significant evidence of indels (mean frequencies ranging from 41.9% to 0.13%) in wild-type and knock-in mouse livers at day four and at week three after infection (Fig. 1b and Supplementary Table 2). Notably, higher CIRCLE-seq read counts generally correlated well with the likelihood of finding indels in vivo: 11 of the 11 class I off-target sites, 5 of the 17 class II sites and 3 of the 17 class III sites contained indels (Fig. 1b). All 19 of these sites had 3 or fewer mismatches relative to the on-target site, with the majority of the sites located within gene-coding sequences (Fig. 1b). Three additional sites—including two sites each containing three mismatches in the spacer and one mismatch in the PAM—showed significant evidence of indel mutations at the four-day time point, but which was no longer significant by three weeks (Supplementary Table 2); the mutation frequencies observed at these sites were around 0.13%, which is close to the limit of detection (0.1%) for next-generation sequencing[10]. These data show that Cas9 with a promiscuous gRNA can generate stable, and sometimes high-frequency, off-target mutations in vivo and that VIVO can identify such mutations even with frequencies as low as 0.13%.

Because potential therapeutic applications would not use a promiscuous gRNA with so many closely matched genomic sites, we sought to assess the in vivo off-target profiles of Cas9 gRNAs designed to be more orthogonal to the mouse genome. We constructed two additional gRNAs (abbreviated as gM[11] and gMH) that are targeted to mouse *Pcsk9* (Extended Data Fig. 6a) but have relatively few closely matched sites in the C57BL6/N mouse genome (Extended Data Table 1). gMH also targets a site (with one mismatch) in the human *PCSK9* open reading frame, which is integrated in the knock-in mouse genome (Extended Data Fig. 6a). Delivery of gM–Cas9 and gMH–Cas9 by adenoviral vectors induced expected genetic alterations in the mouse *Pcsk9* gene and human *PCSK9* transgene as well as corresponding decreases in the levels of Pcsk9 and PCSK9 in plasma (Extended Data Fig. 6b, c).

We conducted the first in vitro CIRCLE-seq step of VIVO with gM–Cas9 and gMH–Cas9 on wild-type and knock-in mouse genomic DNA. The on-target mouse *Pcsk9* was identified in all four experiments and the human *PCSK9* transgene site only in the experiment with gMH on knock-in mouse DNA (Fig. 2). We identified fewer off-target sites with gM–Cas9 and gMH–Cas9 (Fig. 2 and Supplementary Tables 3, 4) than we had with gP–Cas9. When using gM, we found 182 off-target sites: 129 with wild-type mouse DNA, 145 with knock-in mouse DNA and 92 sites that were common to both. When using gMH, we found 529 off-target sites: 333 with wild-type mouse DNA, 394 with knock-in mouse DNA and 198 sites that were common to both. All but 2 of the 711 off-target sites that were identified with these gRNAs had 3 or more mismatches, consistent with the higher orthogonality of these gRNAs relative to the mouse genome. For both gRNAs, there were good concordances in CIRCLE-seq read counts between the wild-type and knock-in mice when considering all off-target sites that were identified ($r^2 = 0.747$ and 0.841 for gM and gMH, respectively) (Fig. 2). As when using gP, sites that were found in only one mouse genome generally had low CIRCLE-seq read counts (Fig. 2), and analysis of gM CIRCLE-seq and targeted amplicon sequencing data suggests that these differences are not due to single nucleotide polymorphisms (Extended Data Table 2 and Supplementary Table 5).

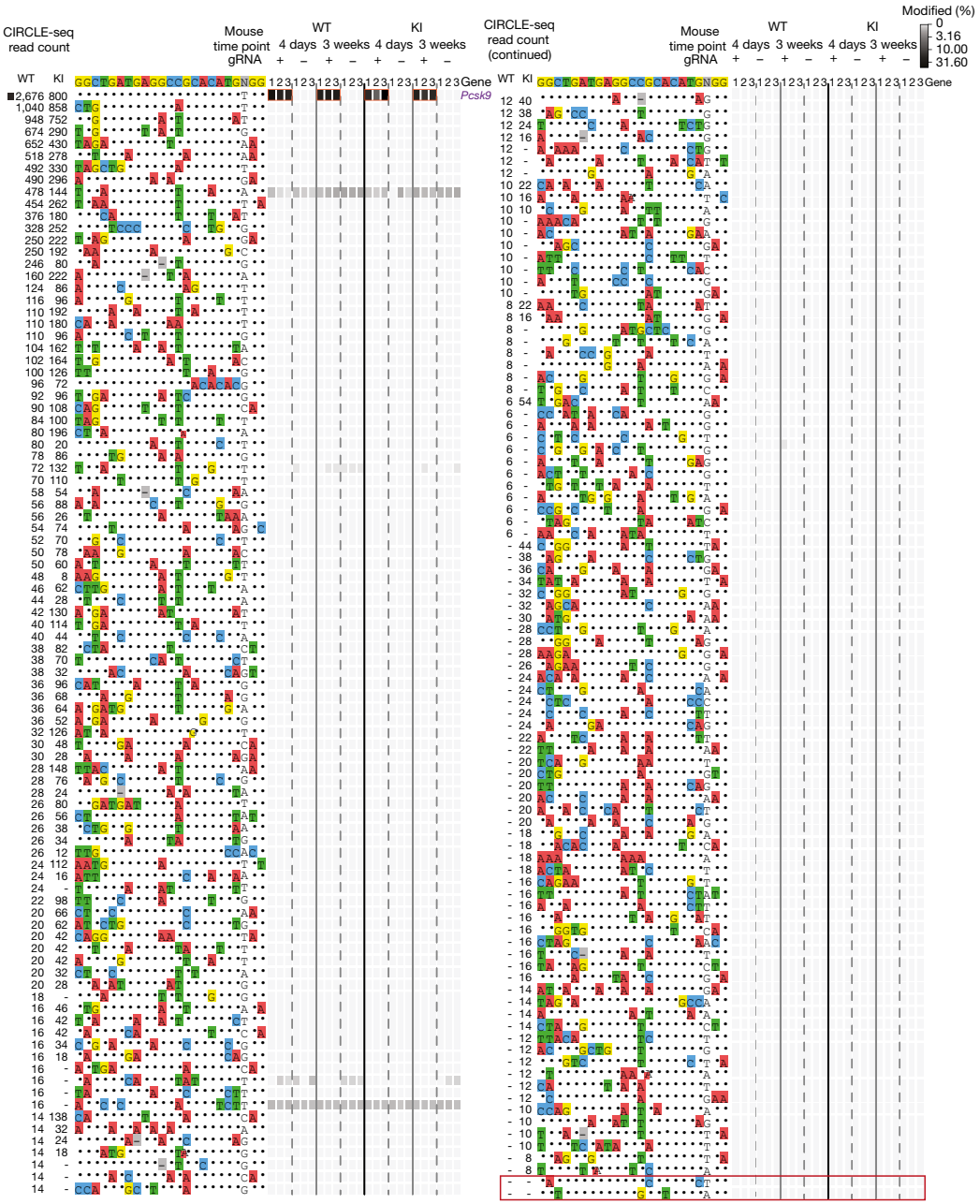


Fig. 3 | Assessment of in vivo off-target indels induced by gM–SpCas9. Indel frequencies determined by targeted amplicon sequencing for the gM–Cas9 on-target site (black square) and 181 off-target sites identified by CIRCLE-seq. Each condition shown was assayed in $n = 3$ biologically independent mice (labelled 1, 2 and 3) using genomic DNA isolated from the liver of mice treated with experimental adenoviral vector that encodes gM–Cas9 (gRNA +) or control adenoviral vector that encodes GFP–Cas9 (gRNA −). Data are presented as in Fig. 1b. The single site (the on-target site) that was significantly different between the experimental (gRNA +) and control (gRNA –) samples is highlighted with orange boxes. Additional closely matched sites in the mouse genome (not identified from the CIRCLE-seq experiments) examined for indel mutations are boxed in red. For source data and *P* values (negative binomial), see Supplementary Table 6.

We conducted the second in vivo step of VIVO for gM and gMH by performing targeted amplicon sequencing of sites found by CIRCLE-seq in liver genomic DNA of adenovirus-treated wild-type and knock-in mice. For gM, we comprehensively examined 181 of the 182 off-target CIRCLE-seq cleavage sites (one site could be amplified but not sequenced; see Methods) and the on-target site from liver DNA collected at day 4 or week 3 after adenovirus infection. Notably, only the gM on-target site showed significant evidence of indels (ranging from 12.6% to 18.5%) (Fig. 3 and Supplementary Table 6); no significant off-target indels were identified at any of the 181 off-target sites in either mouse at either time point. Because CIRCLE-seq identified a large number of potential off-target sites (529 in total) when using gMH, we examined the on-target site and a subset of the off-target sites (comprising 69 sites) that had the highest CIRCLE-seq read counts (and up to 6 mismatches). These 69 sites encompassed all but 1 of the CIRCLE-seq sites that had up to 3 mismatches (1 site could be amplified but not sequenced; see Methods) and the human *PCSK9* transgene site (with 1 mismatch) (Fig. 2). This choice was guided by our finding that none of the 19 stable in vivo off-target sites we found for gP–Cas9 had 4 or more mismatches relative to the on-target site (Fig. 1b). Among the 69 sites, we found significant indel mutations at only the on-target mouse gMH site (27.4–43.6% indels) and the human *PCSK9* transgene site bearing one mismatch (20.4–21.7% indels) (Extended Data Fig. 7 and Supplementary Table 7).

To further exclude the possibility that CIRCLE-seq missed any bona fide off-target sites, we also performed targeted amplicon sequencing

of the most closely matched sites in the mouse C57BL/6N genome (containing up to three mismatches in the spacer) that were not identified in our CIRCLE-seq experiments (four sites for gM and ten sites for gMH) (Extended Data Table 1). Three of the gM sites could not be individually selectively amplified and so these were assessed together as a pool (Methods). We did not observe significant indels at any of these sites in all treated mice at both time points (Fig. 3, Extended Data Fig. 7 and Supplementary Tables 6, 7).

To our knowledge, our report provides the first demonstration that CRISPR–Cas nucleases can robustly induce off-target mutations in vivo. Previous in vivo studies have reported no or very few off-target mutations, but used the cell-based 'genome-wide unbiased identification of double-strand breaks enabled by sequencing' (GUIDE-seq) method[12–14] or other in silico approaches that have not been validated to effectively identify these sites in vivo (see Supplementary Discussion). By contrast, VIVO enabled the robust and sensitive identification of off-target sites in vivo, even those with mutation frequencies as low as about 0.13%. The high sensitivity of CIRCLE-seq is most probably what enabled the identification of a superset of all potential off-target sites, including those actually mutated in vivo (Supplementary Discussion). The detection limit of VIVO—as with all existing methods—is bounded by the current error rate of next-generation sequencing for indels (approximately 0.1%). In addition, we did not attempt to detect large-scale chromosomal rearrangements (translocations, inversions or large deletions) but we expect the frequencies of such alterations to be no higher than that of any off-target indels. Other methods[15] might be used in future studies to test for these alterations.

VIVO defines a pathway for assessing the in vivo genome-wide specificities of CRISPR–Cas nucleases. Our work suggests that gRNAs should be designed to have the lowest possible number of closely matched genomic sites (with three or fewer mismatches), which can be done using existing in silico tools[16]. Such gRNAs can be assessed in step 1 of VIVO to identify those that have a reasonable number (<100–200) of in vitro off-target cleavage sites and then these sites can be comprehensively examined for indels in vivo using targeted amplicon sequencing in step 2 of VIVO. We also recommend examination of any other closely matched genomic sites (fewer than four mismatches) that are not identified by CIRCLE-seq. Persistent off-target mutations might be further reduced using various strategies for improving the genome-wide specificities of CRISPR–Cas nucleases (Supplementary Discussion).

We believe VIVO sets an important standard for defining in vivo off-target effects of gene-editing nucleases. The approach should be generalizable to non-CRISPR gene-editing nucleases, to non-mammalian organisms and to other delivery methods (Supplementary Discussion). VIVO should also be useful for characterizing the in vivo specificities of engineered CRISPR–Cas9 variants and other CRISPR–Cas orthologues (Supplementary Discussion) and for assessing the effectiveness of methods for reducing the off-target effects of these nucleases. We expect our findings will advance research and methods that will further spur the clinical translation of in vivo genome-editing therapeutic strategies.

Note added in proof: a recent study[17] that investigated large-scale chromosome alterations resulting from CRISPR–Cas9 editing found that the frequencies of these alterations were indeed no higher than those of the off-target indels we report here.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, statements of data availability and associated accession codes are available at https://doi.org/10.1038/s41586-018-0500-9.

Received: 27 February 2018; Accepted: 23 July 2018;
Published online: 12 September 2018

1. Musunuru, K. The hope and hype of CRISPR-Cas9 genome editing: a review. *JAMA Cardiol.* **2**, 914–919 (2017).
2. Fellmann, C., Gowen, B. G., Lin, P. C., Doudna, J. A. & Corn, J. E. Cornerstones of CRISPR–Cas in drug discovery and therapy. *Nat. Rev. Drug Discov.* **16**, 89–100 (2017).
3. Komor, A. C., Badran, A. H. & Liu, D. R. CRISPR-based technologies for the manipulation of eukaryotic genomes. *Cell* **168**, 20–36 (2017).
4. Koo, T. & Kim, J. S. Therapeutic applications of CRISPR RNA-guided genome editing. *Brief. Funct. Genomics* **16**, 38–45 (2017).
5. Cornu, T. I., Mussolino, C. & Cathomen, T. Refining strategies to translate genome editing to the clinic. *Nat. Med.* **23**, 415–423 (2017).
6. Dunbar, C. E. et al. Gene therapy comes of age. *Science* **359**, eaan4672 (2018).
7. Tsai, S. Q. & Joung, J. K. Defining and improving the genome-wide specificities of CRISPR–Cas9 nucleases. *Nat. Rev. Genet.* **17**, 300–312 (2016).
8. Tsai, S. Q. et al. CIRCLE-seq: a highly sensitive *in vitro* screen for genome-wide CRISPR–Cas9 nuclease off-targets. *Nat. Methods* **14**, 607–614 (2017).
9. Hsu, P. D. et al. DNA targeting specificity of RNA-guided Cas9 nucleases. *Nat. Biotechnol.* **31**, 827–832 (2013).
10. Tsai, S. Q. et al. GUIDE-seq enables genome-wide profiling of off-target cleavage by CRISPR–Cas nucleases. *Nat. Biotechnol.* **33**, 187–197 (2015).
11. Ding, Q. et al. Permanent alteration of PCSK9 with in vivo CRISPR-Cas9 genome editing. *Circ. Res.* **115**, 488–492 (2014).
12. Yin, H. et al. Structure-guided chemical modification of guide RNA enables potent non-viral *in vivo* genome editing. *Nat. Biotechnol.* **35**, 1179–1187 (2017).
13. Yin, H. et al. Therapeutic genome editing by combined viral and non-viral delivery of CRISPR system components *in vivo*. *Nat. Biotechnol.* **34**, 328–333 (2016).
14. Gao, X. et al. Treatment of autosomal dominant hearing loss by *in vivo* delivery of genome editing agents. *Nature* **553**, 217–221 (2018).
15. Giannoukos, G. et al. UDiTaS™, a genome editing detection method for indels and genome rearrangements. *BMC Genomics* **19**, 212 (2018).
16. 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).
17. Kosicki, M., Tomberg, K. & Bradley, A. Repair of double-strand breaks induced by CRISPR–Cas9 leads to large deletions and complex rearrangements. *Nat. Biotechnol.* **36**, 765–771 (2018).

Acknowledgements J.K.J. is supported by the Desmond and Ann Heathwood MGH Research Scholar Award. J.K.J., M.L.B. and J.A.G. were supported by a sponsored research agreement with AstraZeneca. L.P. is supported by a National Human Genome Research Institute (NHGRI) Career Development Award (R00HG008399). J.K.J., M.J.A. and J.M.-L. are supported by a National Institutes of Health Maximizing Investigators' Research Award (MIRA) (R35 GM118158). J.K.J., L.P. and K.C. are supported by the Defense Advanced Research Projects Agency (HR0011-17-2-0042). We thank M. Snowden, S. Platz and S. Rees for resource allocation from AstraZeneca Research Funds. We thank J. Y. Hsu for discussions and input.

Reviewer information *Nature* thanks F. Urnov and the other anonymous reviewer(s) for their contribution to the peer review of this work.

Author contributions P.A., M.J.P., T.B. and M.B. executed intravenous tail vein injections. P.A. and T.B. coordinated vena saphena blood sampling and performed plasma extractions. P.A., M.J.P., A.C., T.B., M.B., M.M. and R.N. performed in vivo terminations and organ collection. P.A. performed Surveyor assays for genomic DNAs of mouse livers. P.A. and T.B. performed ELISA for Pcsk9 protein detection in plasma samples. M.L.B, J.A.G., S.Q.T. and N.T.N. performed the CIRCLE-seq and targeted amplicon sequencing experiments. J.M.-L., K.C., S.P.G., L.P. and M.J.A. performed bioinformatic and computational analysis of the off-target experiments. M.A.F. generated AstraZeneca proprietary software for gRNA identification. P.A. designed gRNAs with help of M.A.F., and validated their functional activity. F.S. performed mouse phenotypic characterization. P.A., M.L.B., S.Q.T., M.M., M.B.-Y, A.C., R.N., M.D.F., L.M.M. and J.K.J. conceived of and designed the study. P.A., M.L.B., M.M. and J.K.J. organized and supervised experiments. P.A., M.L.B., J.A.G., M.M. and J.K.J. prepared the manuscript with input from all authors.

Competing interests J.K.J. has financial interests in Beam Therapeutics, Blink Therapeutics, Editas Medicine, Endcadia, Monitor Biotechnologies (formerly known as Beacon Genomics), Pairwise Plants, Poseida Therapeutics and Transposagen Biopharmaceuticals. M.J.A. and S.Q.T. have financial interests in Monitor Biotechnologies. J.K.J.'s and M.J.A.'s interests were reviewed and are managed by Massachusetts General Hospital and Partners HealthCare in accordance with their conflict of interest policies. S.Q.T. and J.K.J. are co-inventors on a patent describing the CIRCLE-seq method. P.A., M.D.F., M.J.P., M.A.F., F.S., M.B., R.N., M.B.Y. and M.M. are employees and shareholders of AstraZeneca. L.M.M. is an employee and shareholder of GE Healthcare and a shareholder of AstraZeneca.

Additional information
Extended data is available for this paper at https://doi.org/10.1038/s41586-018-0500-9.
Supplementary information is available for this paper at https://doi.org/10.1038/s41586-018-0500-9.
Reprints and permissions information is available at http://www.nature.com/reprints.
Correspondence and requests for materials should be addressed to M.M. or J.K.J.
Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

METHODS

gRNA design. gP was identified by searching for an on-target sequence within mouse *Pcsk9* (ENSMUSG000044254) exons one to three that showed a high number of closely matched sites (two or fewer mismatches to the on-target site) in the mouse genome. gMH was designed by searching for a gRNA that can cleave both mouse *Pcsk9* and human *PCSK9*, and that showed a perfect alignment to mouse *Pcsk9* and up to two nucleotides mismatch to human *PCSK9* (ENSG00000169174) at least eight nucleotides distal from the PAM. For both gRNA designs, AstraZeneca proprietary software was used as an in silico tool, which was developed based on the codebase of the Wellcome Trust Sanger Institute (WGE: http://www.sanger.ac.uk/htgt/wge/)[18] with the addition of the NAG PAM motif and NGG in the alignments for potential off-targets[9]. GRCm38/mm10 and GRCh38/hg38 genomes were used as reference for alignments. The gM targeted to mouse *Pcsk9* has previously been described[11].

Adenoviral constructs. Adenoviruses that express Cas9 and gRNAs (Ad-Cas9-gM, Ad-Cas9-gMH and Ad-Cas9-gP) were generated by Vector Biolabs (Malvern). Cas9 and gRNAs were expressed from chicken β-actin hybrid (CBh) and U6 promoters, respectively, in a replication-deficient adenoviral-serotype 5 (dE1/E3) backbone. A negative control adenovirus (Ad-Cas9-GFP) that expresses Cas9 and GFP from the CBh and CMV promoters, respectively, but no gRNA was also generated.

Animal studies. All animal experiments were approved by the AstraZeneca internal committee for animal studies as well as the Gothenburg Ethics Committee for Experimental Animals (license number: 162-2015+), compliant with EU directives on the protection of animals used for scientific purposes.

C57BL/6N mice (Charles River) were individually housed in a temperature (21 ± 2 °C) and humidity (55 ± 15%) controlled room with a 12:12-h light:dark cycle. R3 diet (Lactamin AB) and tap water were provided ad libitum. Cage bedding and enrichments include spen chips, shredded paper, gnaw sticks and a plastic house.

A humanized hypercholesterolaemia mouse model was generated by liver-specific overexpression of human PCSK9 in C57BL/6N mice (A.C. et al., manuscript submitted). In brief, the knock-in mouse was generated by cloning the human *PCSK9* open reading frame downstream of the mouse albumin promoter from albumin–Cre mice, in a vector designed to target the mouse *Rosa26*. Founder C57BL/6N hPCSK9 heterozygous males were crossed with C57BL/6N females to generate experimental animals, which are littermates with two genotypes: C57BL/6N hPCSK9KI$^{+/-}$ (referred to as knock-in) and wild-type C57BL/6N hPCSK9KI$^{-/-}$ (referred to as wild-type) (Extended Data Fig. 3).

For in vivo *Pcsk9* gene editing, nine- to eleven-week-old male mice received a tail vein injection with a dose of 1×10^9 infection units (IFU) of adenovirus (Ad-Cas9-gM, Ad-Cas9-gMH, Ad-Cas9-gP or Ad-Cas9-GFP) in 200 μl diluted with phosphate-buffered saline. Peripheral blood was sampled before virus administration (baseline), a week after virus administration and at termination (four days or three weeks after virus administration). Animals were euthanized by cardiac puncture under isoflurane anaesthesia at the experimental endpoint. The organs—including liver, spleen, lungs, kidney, muscle, brain and testes—were dissected, snap-frozen in liquid nitrogen and stored at −80 °C until further analyses.

Ten milligrams of frozen liver tissue was lysed to obtain 30–50 μg genomic DNA using the Gentra Puregene Tissue kit (Qiagen). In vivo gene-editing efficiency was evaluated using Surveyor mismatch cleavage assay (Integrated DNA Technologies, BVBA) (using primers: for gP, GAGGCCGAAACCTGATCCTT and CTTAGAGACCACCAGACGGC; for gM, GGAGGACACGTTTTCTGCAT and CTGCTGCTGTTGCTGCTAC; for gMH mouse locus, GACTTTGTGA AGGCTGGGGA and TGCATGGAGCAATGCAGAGA; for gMH human transgene, TAGCCTTGCGTTCCGAGGAG and CATTCTCGAAGTCGGTGACCA; with amplicon sizes 619, 415, 349, 495 base pairs, respectively) and targeted deep sequencing (primers listed in Supplementary Tables 2, 6 and 7).

Animals were randomized based on their weights measured before the experiments. The investigators were blinded to group allocation during data collection and analysis. No sample size calculation was performed. Sample size was determined to generate triple independent samples for comparisons between groups sufficient to perform statistical tests.

Assessment of human PCSK9 and/or mouse Pcsk9 protein levels in plasma. Peripheral blood was collected in EDTA-coated capillary tubes from *vena saphena* during the course of the study and by cardiac puncture at the time of termination. Samples were kept on ice for up to 2 h before extraction of plasma by centrifugation at 10,000 rpm for 20 min at 4 °C. Plasma was stored at −80 °C until the samples were analysed. Levels of human PCSK9 and mouse Pcsk9 in plasma were determined with a standard ELISA kit (DPC900 and MPC900; R&D Systems) according to the manufacturer's instructions. Prior to the assay, plasma samples were diluted 1:800 and 1:1,000 for human PCSK9 and mouse Pcsk9, respectively.

Reference genome for CIRCLE-seq, CRISPResso and Cas-OFFinder. Build 38 of the C57BL/6NJ genome, sequenced by the Sanger Mouse Genomes Project (http://csbio.unc.edu/CCstatus/index.py?run=PseudoOld), was used as the reference genome for the experiments of this report. The human *PCSK9* gene DNA sequence was inserted into the mouse genome as an extra chromosome in the reference and was named as 'chrPCSK9KI'.

CIRCLE-seq. CIRCLE-seq was performed experimentally as previously described[8]. Data were processed using v.1.1 of the CIRCLE-Seq analysis pipeline[8] (https://github.com/tsailabSJ/circleseq) with parameters: 'window_size: 3; mapq_threshold: 50; start_threshold: 1; gap_threshold: 3; mismatch_threshold: 7; merged_analysis: False, variant_analysis: True'. The gP off-target sites were adjusted for mapping artefacts in highly repetitive regions by consolidating contiguous sites whose mapping positions differed by 5 base pairs or less. The off-target sequences reported for these consolidated sites were obtained by performing the local alignment used in the CIRCLE-seq pipeline.

Targeted amplicon deep sequencing of off-targets. Genomic DNA from liver tissue of adenovirus-injected mice was extracted at day 4 and at week 3 post-treatment for indel analysis. As detailed in the text, we validated off-targets identified by CIRCLE-seq by selecting sites with read counts above 50% of the on-target and a variety of lower-ranked sites (containing up to 6 mismatches relative to the on-target) for targeted deep sequencing. In addition, we ruled out the possibility that CIRCLE-seq was missing potential off-target sites identified using in silico tools by sequencing all sites containing up to 3 mismatches, identified by Cas-OFFinder[16] for gM and gMH. All sites we analysed were amplified from 150 ng of input genomic DNA (approximately 5×10^4 genomes) with Phusion Hot Start Flex DNA polymerase (New England Biolabs). PCR products were purified using magnetic beads made as previously described[19], quantified using a QuantiFlor dsDNA System kit (Promega), normalized to 10 ng/μl per amplicon and pooled. Pooled samples were end-repaired and A-tailed using an End Prep Enzyme Mix and reaction buffer from NEBNext Ultra II DNA Library Prep Kit for Illumina, and ligated to Illumina TruSeq adapters using a ligation master mix and ligation enhancer from the same kit. Library samples were then purified with magnetic beads made as previously described[19], size-selected using PEG/NaCl SPRI solution (KAPA Biosystems), quantified using droplet digital PCR (BioRad) and loaded onto an Illumina MiSeq for deep sequencing. To analyse amplicon sequencing of potential on- and off-targets, we used CRISPResso software[20] v1.0.11 (https://github.com/lucapinello/crispresso) with the following parameters: '-q 30 --ignore_substitutions --hide_mutations_outside_window_NHEJ'.

For each of the 45 gP off-target sites we examined, we obtained 10,000 or more sequencing reads in at least 2 samples for treated and control samples at all time points (Supplementary Table 2). One potential gM off-target site (chr15:98037617-98037640) and one potential gMH off-target site (chr15:4878177-4878200) were amplified but could not be successfully sequenced. The problematic gM site was amplified with two different sets of primers and both amplicons failed to sequence. The gMH site is in a highly repetitive area with low complexity and we were unable to differentiate this site from other sites in the genome; therefore, the site was removed from analysis. For all of the gM off-target sites we were able to sequence, we obtained 10,000 or more sequencing reads in at least 2 samples for treated and control samples at all time points (Supplementary Table 6). For all but 1 of the gMH off-target sites we were able to sequence, we obtained 10,000 or more sequencing reads in at least 2 samples for treated and control samples at all time points (Supplementary Table 7). One of the gMH off-target sites we sequenced (chr17:33501685-33501708) did not reach the 10,000 read threshold for any samples or time points but read counts ranged from 2,509 to 9,149. We were unable to individually and selectively amplify three sites that were identified in silico as being highly similar to the gM on-target site but that were not identified by CIRCLE-seq, owing to their sequence similarities: these sites were chr14:25878231-25878254, chr14:26018001-26018024 and chr14:26157615-26157638. Therefore, for these three sites the read counts were pooled into one amplicon that encompasses all locations and that is labelled as 'chr14:pooled' in Supplementary Table 6.

Targeted amplicon deep sequencing of wild-type and knock-in untreated mice. Genomic DNA from liver tissue of untreated mice—the same mice upon which CIRCLE-seq was performed—was amplified at sites that contained no reads in either the knock-in or wild-type CIRCLE-seq for gM and deep sequenced to look for single nucleotide polymorphisms (Supplementary Table 5). Primers used are the same as in Supplementary Table 6. Sites were analysed with CRISPRessoPooled with the following parameters: '--cleavage_offset -7 --window_around_sgrna 13', to perform a focused variant analysis in the spacer.

Cas-OFFinder. Identification of potential off-targets by Cas-OFFinder[16] (https://github.com/snugel/cas-offinder, v. 2.4) was done using the off-line version, allowing up to 7 mismatches and non-canonical PAMs. We then restricted the output to the sites with at most 6 mismatches in the spacer and at most 1 mismatch in the PAM.

Non-reference genetic variation. samtools (mpileup and bcftools, v.1.3.1[21]) was used to discover non-reference genetic variation at the off-target sites identified by CIRCLE-seq. Positions with a genotype-quality score greater than 5 and depth

of at least 3 were considered as potential variants if they did not fall adjacent to the cleavage site or at the edge of the reads, and were not located in a highly repetitive region with poor mapping quality.

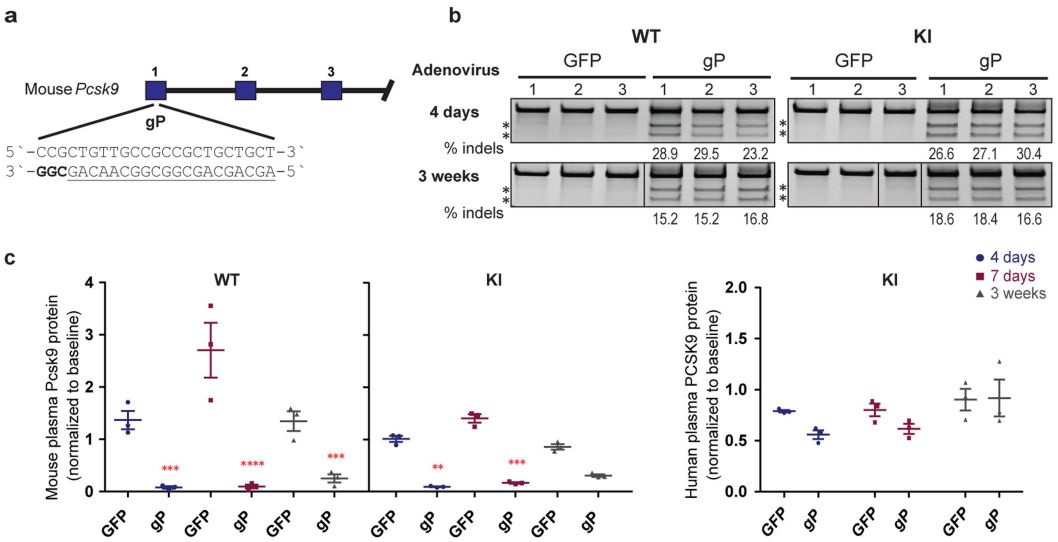
Statistical analysis of levels of protein in plasma. Data visualization and statistical analyses for plasma protein measurements were performed using GraphPad Prism 7.02. Protein levels after the adenoviral administration were normalized to baseline levels, and values for gRNA treatment groups were compared with the control treatment group. Comparisons between groups were performed using two-way ANOVA test followed by Sidak's or Dunnett's two-sided adjusted multiple comparisons test, depending on the number of comparison groups. $P < 0.05$ was considered to be statistically significant. The level of significance in all graphs is represented as follows: $*P < 0.05$, $**P < 0.01$, $***P < 0.001$ and $****P < 0.0001$. Exact P values, confidence intervals and effect size are presented in the Source Data.

Statistical analysis of targeted amplicon deep-sequencing data. P values were obtained by fitting a negative binomial generalized linear model (function MASS:glm.nb in R version 3.4.2 with parameter init.theta = 1 and with the logarithm of the total number of reads as the offset) to the control and nuclease-treated samples for each evaluated site with at least one non-zero indel count among the nuclease-treated samples. To avoid convergence issues, we added 1 to all the indel counts and confirmed that rerunning the models without the addition of the 1 did not result in any additional significant off-target sites. We adjusted for multiple comparisons using the Benjamini and Hochberg method (function p.adjust in R version 3.4.2). Multiple testing adjustment was performed within strata defined by gRNA, mouse background and time point. We considered the indel percentage in the gRNA–Cas9-treated replicates to be significantly greater than the indel percentage in the GFP–Cas9-treated controls if the adjusted P value was less than 0.1, the nuclease-treatment coefficient was greater than zero and the median indel frequency of the treated replicates was greater than 0.1%.

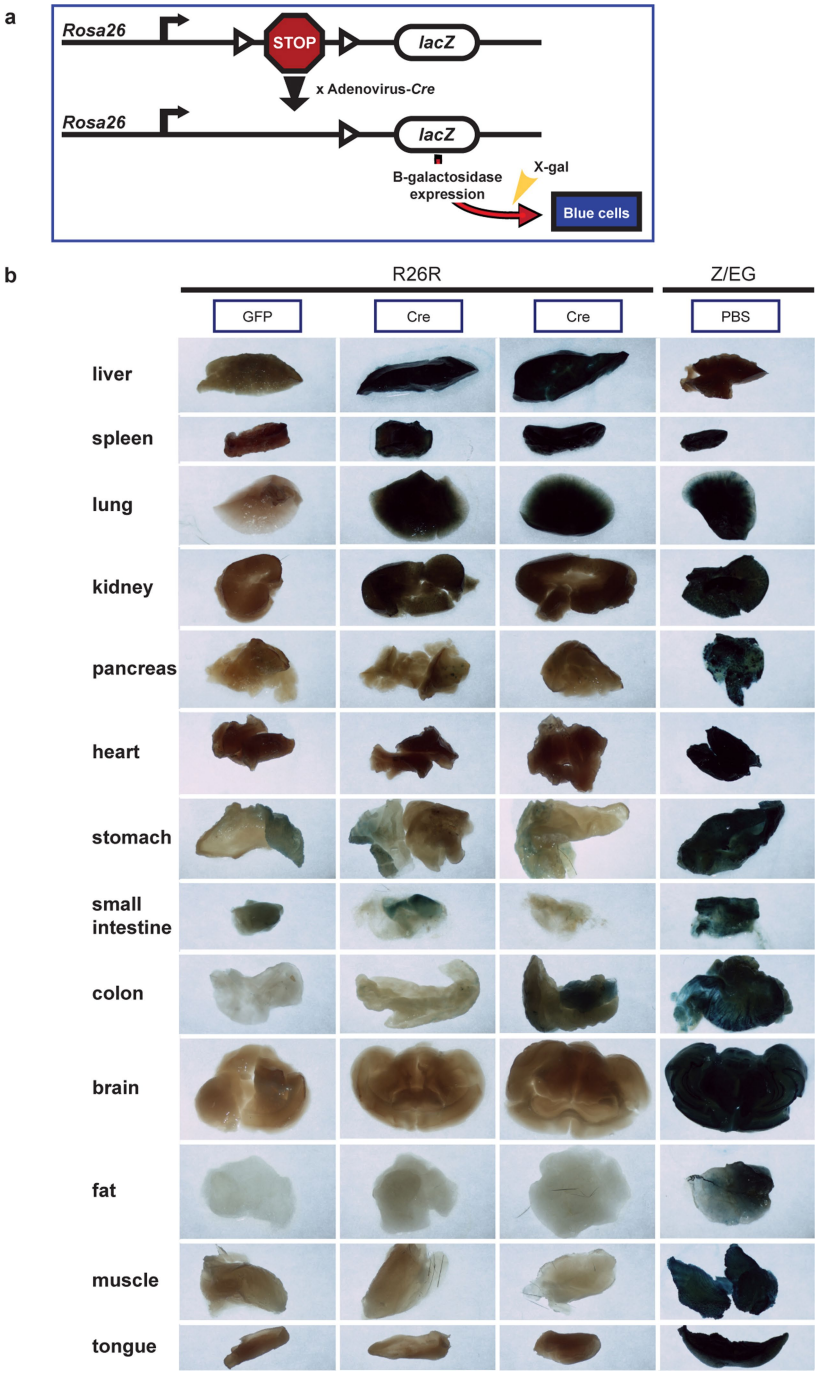
Reporting summary. Further information on research design is available in the Nature Research Reporting Summary linked to this paper.

Data availability. Sequence data that support the findings of this study have been deposited with SRA accession number SRP151131. All other data that support the findings of this study are available from the corresponding authors upon reasonable request.

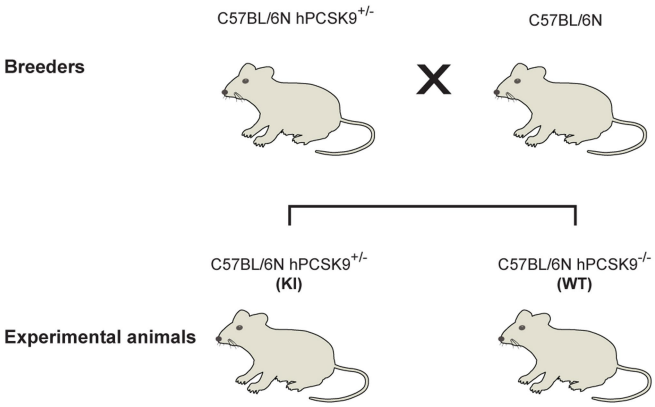
18. Hodgkins, A. et al. WGE: a CRISPR database for genome engineering. *Bioinformatics* **31**, 3078–3080 (2015).
19. Rohland, N. & Reich, D. Cost-effective, high-throughput DNA sequencing libraries for multiplexed target capture. *Genome Res.* **22**, 939–946 (2012).
20. Pinello, L. et al. Analyzing CRISPR genome-editing experiments with CRISPResso. *Nat. Biotechnol.* **34**, 695–697 (2016).
21. Li, H. et al. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).



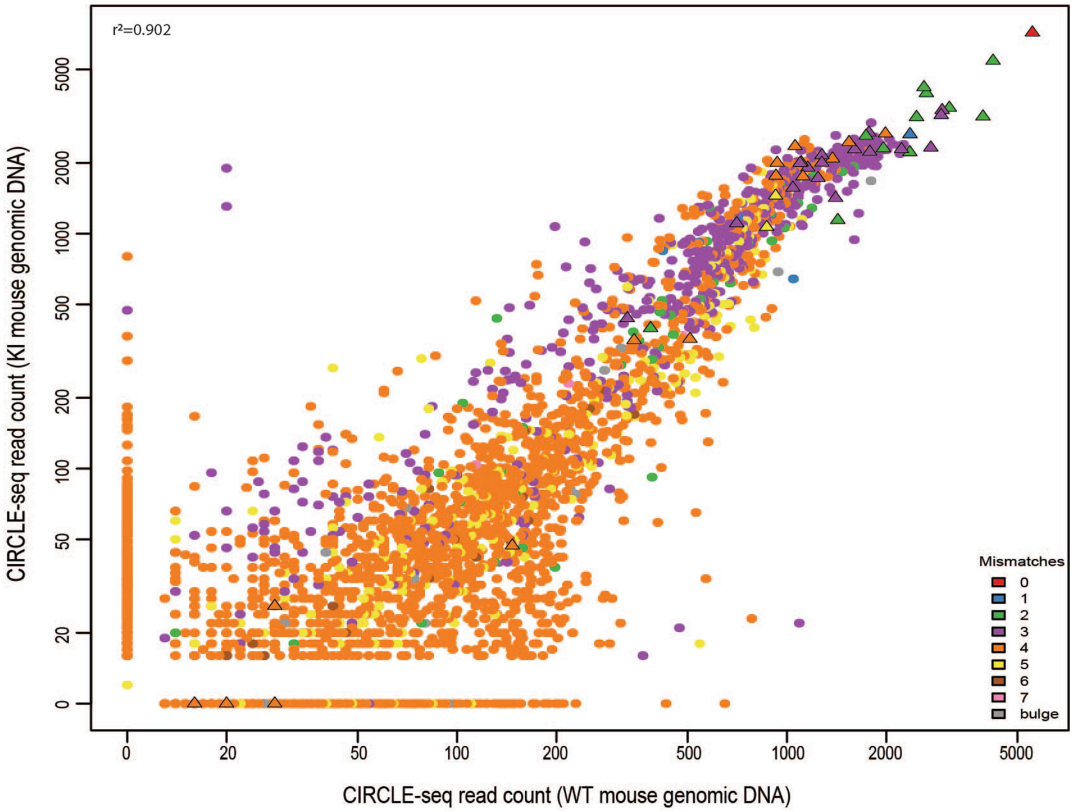
Extended Data Fig. 1 | gP–Cas9 efficiently mutates the mouse *Pcsk9* gene and reduces levels of Pcsk9 protein in plasma in vivo. a, gP was designed to target a sequence within exon 1 of the mouse *Pcsk9* gene that has many closely related genomic sites (that is, those with 1–3 mismatches relative to the on-target site; Extended Data Table 1). Blue bars indicate exons for the mouse genomic region. **b**, Surveyor assay and next-generation DNA sequencing data demonstrate efficient in vivo modification of the on-target mouse *Pcsk9* gene site in mouse liver by gP–Cas9. Assays were performed on day 4 and on week 3 after the administration of adenoviral vectors that encode gP–Cas9 (gP) or negative control GFP–Cas9 (GFP). For each time point, the assays used genomic DNA isolated from livers of $n = 3$ biologically independent wild-type C57BL/6N (WT) mice or C57BL/6N-derived mice containing a single copy of the human *PCSK9* open reading frame under albumin promoter, knocked into the *Rosa26* locus (KI). Asterisks indicate the cleaved PCR products expected after treatment with Surveyor nuclease. Percentages show the frequencies of indel mutations determined by targeted amplicon sequencing using next-generation sequencing; these are the same values shown for the on-target site in Fig. 1b. Lines divide lanes taken from different locations on the same gel. For source data for Surveyor assays and targeted amplicon sequencing, see Supplementary Fig. 1 and Supplementary Table 2, respectively. **c**, Mouse Pcsk9 protein levels in plasma measured in $n = 3$ biologically independent wild-type and knock-in mice and human PCSK9 protein levels in plasma measured in $n = 3$ biologically independent knock-in mice, after nuclease treatment. Protein levels were assessed on day 4, 7 and week 3 after administration of gP or control GFP adenoviral vectors and normalized to baseline levels. Significant differences between experimental and control groups were determined using two-way ANOVA and Sidak's two-sided adjusted multiple comparisons test; $^{**}P < 0.01$, $^{***}P < 0.001$, $^{****}P < 0.0001$. See Source Data for Extended Data Fig. 4b for exact adjusted *P* values. All values are presented as group means, and error bars represent standard error of the mean. The enhanced reduction of levels of Pcsk9 in plasma relative to the frequency of observed *Pcsk9* genetic alteration is consistent with previously published studies (Supplementary Discussion).



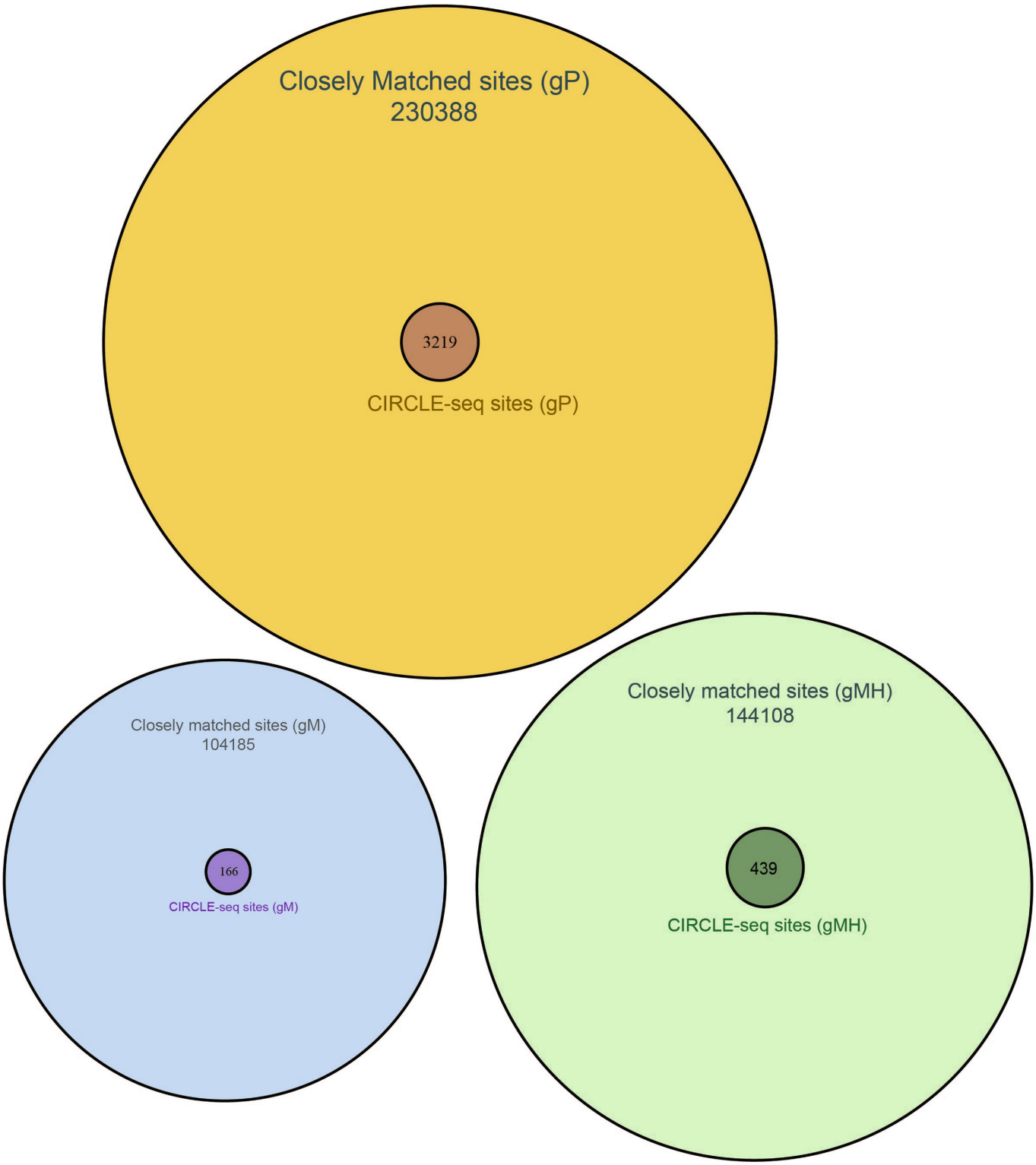
Extended Data Fig. 2 | Bio-distribution studies of adenovirus-serotype 5 in mice. a, Schematic of integrated reporter construct in R26R mice used to assess delivery of Cre recombinase using adenovirus serotype 5 vector. Cre-mediated excision of a *lox*P-flanked transcriptional stop signal upstream of a *lacZ* gene results in expression of β-galactosidase enzyme. β-Galactosidase expression can be quantified by staining dissected tissues with X-gal, a compound that turns blue when cleaved by this enzyme. **b**, Quantification of β-galactosidase expression in sections of various dissected organs from $n = 2$ biologically independent R26R mice intravenously injected with adenovirus serotype 5 vector encoding Cre. Matched organs sections from a R26R mouse intravenously injected with an adenovirus serotype 5 vector encoding GFP were used to determine background staining levels and serve as a negative control. Matched organ sections from Z/EG mice that constitutively express lacZ (β-galactosidase) and intravenously injected with PBS (rather than adenovirus) were used to provide positive staining controls. All mice were evaluated one week after adenovirus or PBS injection. The experiment was performed once.



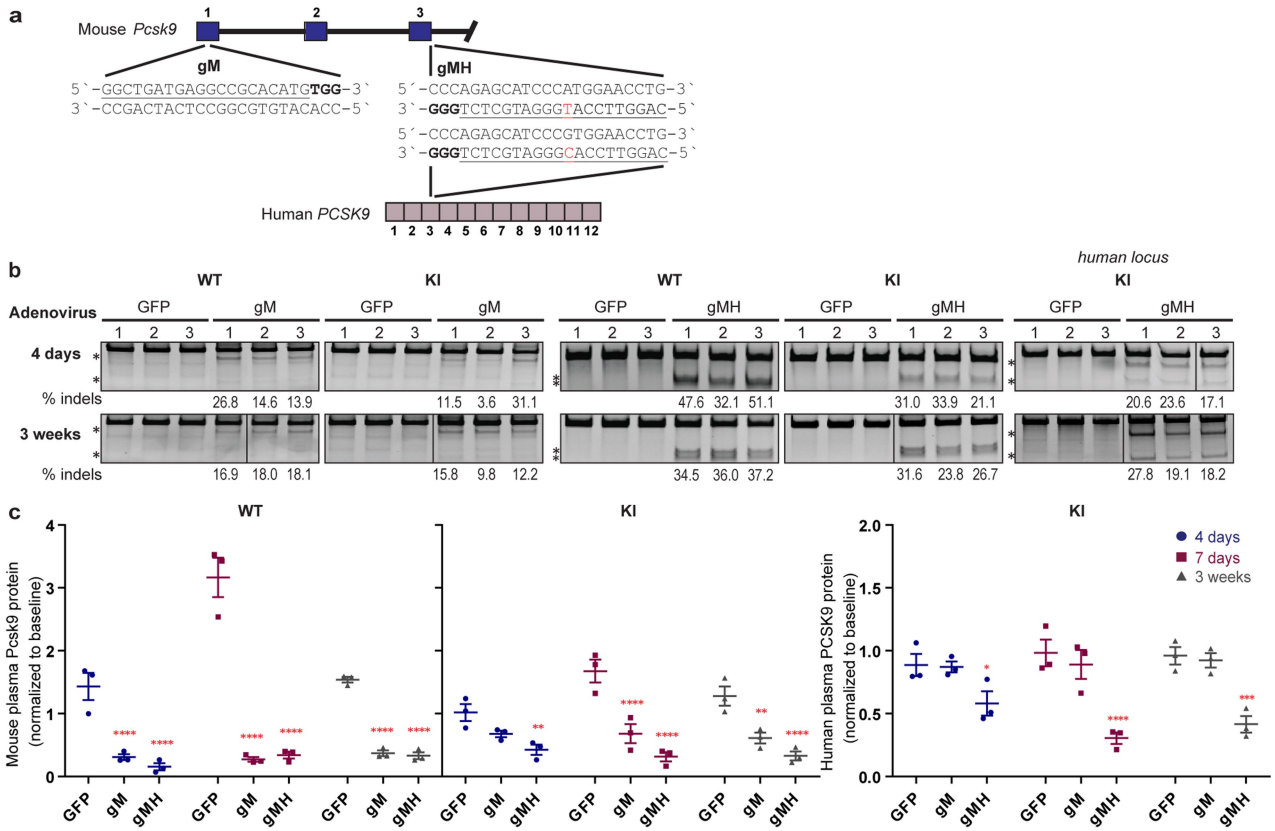
Extended Data Fig. 3 | Breeding strategy for generating experimental mice containing human *PCSK9* open reading frame knocked into the *Rosa26* locus. C57BL/6N-derived mouse line containing a single copy of the human *PCSK9* open reading frame knocked into the *Rosa26* locus (C57BL/6N hPCSK9KI$^{+/-}$) are used for breeding with C57BL/6N mice. Offspring yielded experimental animals that are C57BL/6N hPCSK9KI$^{+/-}$ (referred to as knock-in) and C57BL/6N hPCSK9KI$^{-/-}$ (referred to as wild type) males.



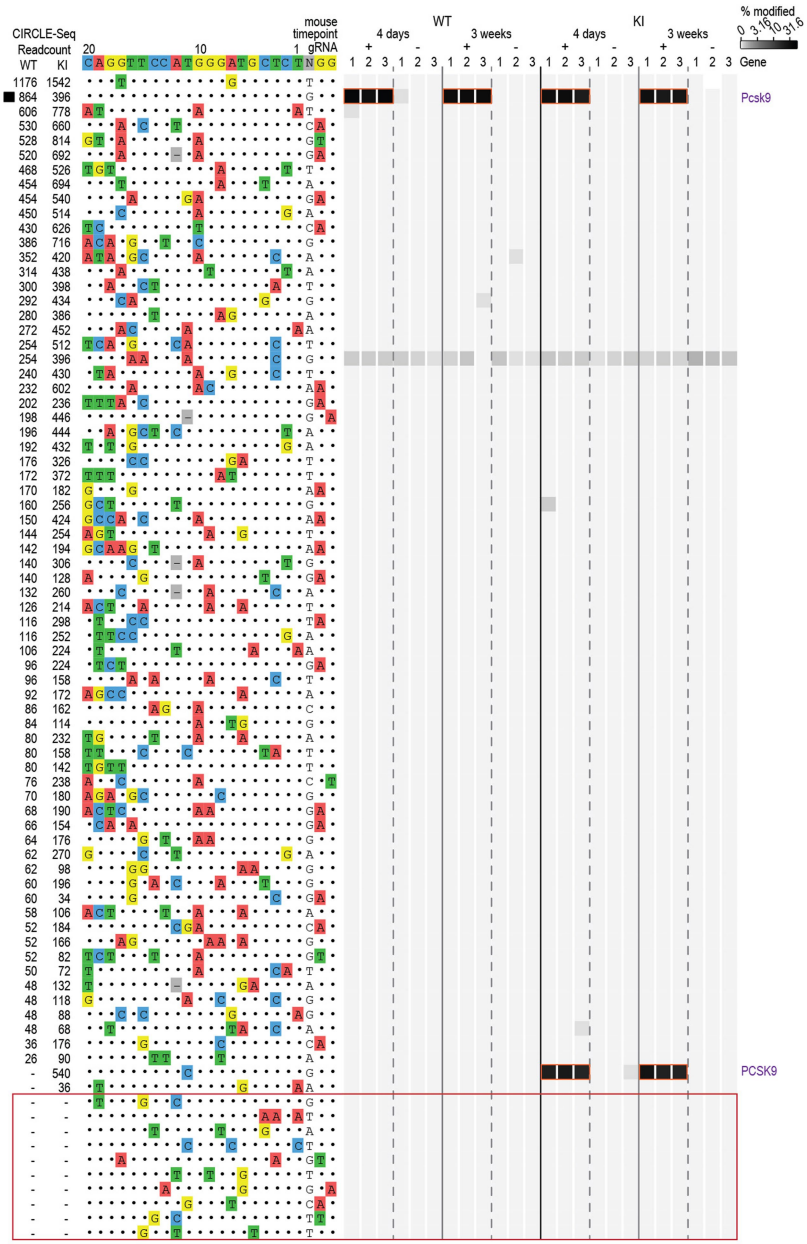
Extended Data Fig. 4 | Scatter plot of CIRCLE-seq read counts for sites identified with gP–Cas9 on genomic DNA from $n = 1$ wild-type and $n = 1$ knock-in mice. Read counts are shown on a logarithmic scale and colours indicate the number of mismatches in each off-target site relative to the on-target site. Sites shown as triangles were chosen for targeted amplicon sequencing. The correlation r^2 value obtained using all values in the scatter plot is shown in the upper left-hand corner and was obtained using a linear regression.



Extended Data Fig. 5 | Comparison of closely matched sites identified in silico and off-target cleavage sites identified by CIRCLE-seq. Venn diagrams comparing off-target cleavage sites in mouse genomic DNA identified by CIRCLE-seq experiments with closely matched sites (up to six mismatches relative to the on-target site) in the mouse genome identified in silico by Cas-OFFinder are shown for the Cas9 gRNAs gP, gM and gMH.



Extended Data Fig. 6 | Genetic and phenotypic alterations induced by delivery of gM–Cas9 and gMH–Cas9 in vivo. a, Sequence and location of the Cas9 gM (mouse) and gMH (mouse and human) target sites in the endogenous mouse *Pcsk9* gene and human *PCSK9* transgene inserted at the mouse *Rosa26* locus. The single base position that differs between the gMH target sites in the mouse *Pcsk9* gene and the human *PCSK9* transgene is highlighted in red. Blue bars indicate exons for the mouse genomic region and purple bars represent exons for the human genomic locus; the PAM sequence for the sites is in bold and the spacer sequence is underlined. **b**, Surveyor assay and next-generation DNA sequencing data demonstrate efficient in vivo modification of the on-target endogenous mouse *Pcsk9* site and human *PCSK9* transgene in mouse liver. Assays were performed at day 4 and at week 3 after administration of adenoviral vectors that encode gM and Cas9 (gM), gMH and Cas9 (gMH) or GFP and Cas9 (GFP) using genomic DNA isolated from livers of $n = 3$ biologically independent wild-type and knock-in mice. Asterisks indicate the cleaved PCR products expected following treatment with Surveyor nuclease. Percentages show the frequencies of indel mutations determined by targeted amplicon sequencing using next-generation sequencing; these are the same values shown for the on-target sites in Fig. 3 and Extended Data Fig. 7. Lines divide lanes taken from different locations on the same gel. For source data for Surveyor assays, see Supplementary Fig. 1. For source data for targeted amplicon sequencing, see Supplementary Tables 6 and 7 for gM and gMH, respectively. **b**, Mouse Pcsk9 protein levels measured in plasma in $n = 3$ biologically independent wild-type and knock-in mice, and human PCSK9 protein levels measured in plasma in $n = 3$ biologically independent knock-in mice after CRISPR–Cas nuclease treatment. Protein levels in plasma were assessed at day 4, 7 and week 3 after the administration of gM, gMH or control GFP adenoviral vectors and normalized to baseline levels at each time point. Significant differences between groups were determined using two-way ANOVA and Dunnett's two-sided adjusted multiple comparisons test; $^{*}P < 0.05$, $^{**}P < 0.01$, $^{***}P < 0.001$, $^{****}P < 0.0001$. See Source Data for exact adjusted P values. Values are presented as group means, error bars represent standard errors of the mean.



Extended Data Fig. 7 | Assessment of in vivo off-target indel mutations induced by gMH–Cas9. Indel mutation frequencies determined by targeted amplicon sequencing (using high-throughput sequencing) are presented as heat maps for the gMH–Cas9 on-target site (black square) and 63 off-target sites identified from CIRCLE-seq experiments. Each locus was assayed in $n = 3$ biologically independent mice (labelled 1, 2 and 3) using genomic DNA isolated from the liver of wild-type and knock-in mice treated with experimental adenoviral vector that encodes gMH–Cas9 (gRNA +) or control adenoviral vector GFP–Cas9 (gRNA −). For each site, mismatches relative to the on-target site are shown with coloured boxes and bases in the spacer sequence and are numbered from 1 (most proximal to the PAM) to 20 (most distal from the PAM). The number of read counts found for each site from the CIRCLE-seq experiments on wild-type and knock-in mouse genomic DNA are shown in the left columns (ranked from highest to lowest based on counts in the wild-type genomic DNA CIRCLE-seq experiment). Each box in the heat map represents a single sequencing experiment. Sites that were significantly different between the experimental (gRNA +) and control (gRNA −) samples are highlighted with an orange outline around the boxes. Additional closely matched sites in the mouse genome (not identified from the CIRCLE-seq experiments) that were examined for indel mutations are boxed in red at the bottom of the figure. See Supplementary Table 7 for source data and *P* values (negative binomial).

Extended Data Table 1 | Numbers of off-target sites for gP, gM and gMH identified by Cas-OFFinder (in silico) and CIRCLE-seq (experimental)

		Sites with canonical NGG PAM							
		Number of spacer mismatches							
gRNA	Method	0	1	2	3	4	5	6	7
gP	Cas-OFFinder	1	5	41	355	1073	3347	21900	94051
	CIRCLE-seq (total WT&KI)	1	5	38	231	121	44	15	4
	CIRCLE-seq (WT mouse)	1	5	38	226	117	43	13	2
	CIRCLE-seq (KI mouse)	1	5	38	218	93	25	11	4
gM	Cas-OFFinder	1	0	0	8	77	780	8315	55093
	CIRCLE-seq (total WT&KI)	1	0	0	4	18	36	32	25
	CIRCLE-seq (WT mouse)	1	0	0	4	17	26	23	17
	CIRCLE-seq (KI mouse)	1	0	0	3	17	30	23	14
gMH	Cas-OFFinder	1	1	1	15	178	1609	10992	55363
	CIRCLE-seq (total WT&KI)	1	1	1	9	52	65	82	159
	CIRCLE-seq (WT mouse)	1	0	1	8	46	46	45	81
	CIRCLE-seq (KI mouse)	1	1	1	9	43	53	64	102

		Sites with PAM harboring single mismatch						
		Number of spacer mismatches						
gRNA	Method	0	1	2	3	4	5	6
gP	Cas-OFFinder	0	16	370	12028	9828	20097	70495
	CIRCLE-seq (total WT&KI)	0	14	279	2160	271	31	6
	CIRCLE-seq (WT mouse)	0	14	272	1648	164	13	2
	CIRCLE-seq (KI mouse)	0	14	271	1904	262	31	4
gM	Cas-OFFinder	0	0	1	20	423	4911	34722
	CIRCLE-seq (total WT&KI)	0	0	1	3	18	18	10
	CIRCLE-seq (WT mouse)	0	0	0	3	15	16	8
	CIRCLE-seq (KI mouse)	0	0	1	3	14	8	4
gMH	Cas-OFFinder	0	0	7	80	907	8285	67109
	CIRCLE-seq (total WT&KI)	0	0	3	19	21	9	17
	CIRCLE-seq (WT mouse)	0	0	3	18	18	7	9
	CIRCLE-seq (KI mouse)	0	0	3	16	9	5	13

For sites identified by CIRCLE-seq, the total number of sites found in the wild-type and knock-in mice are listed and immediately below that are the number of sites found in each of the two mice.

Extended Data Table 2 | Off-target sites identified by CIRCLE-seq for gP, gM and gMH that exhibit single nucleotide polymorphisms or indels based on CIRCLE-seq data

Location of site	gRNA	Mouse	gRNA on-target site	Off-target sequence	Number of Mismatches	SNPs confirmed by targeted amplicon-sequencing
chr9:78832014-78832037	gM	KI	GGCTGATGAGGCCGCACATGNGG	atCaGATaAaaCCaCACATGGaG	8	No
chr4:129226173-129226196	gMH	KI	CAGGTTCCATGGGATGCTCTNGG	agGGcTCacctGGATGCTCTGtG	8	N.D.
chr9:68916733-68916756	gMH	KI	CAGGTTCCATGGGATGCTCTNGG	tAGGgagagaGGGATGCTCTGaG	8	N.D.
chr3:19461541-19461564	gMH	KI	CAGGTTCCATGGGATGCTCTNGG	CAtGTaCCAaGGGATGtTCTAcG	5	N.D.
chr13:37109353-37109376	gP	KI	AGCAGCAGCGGCGGCAACAGNGG	taCAGCAGCaGCaGCAACAaCGa	6	N.D.
chr6:112201818-112201841	gP	WT	AGCAGCAGCGGCGGCAACAGNGG	AGCAaCAGCaGCaGCAgCAGTaG	5	N.D.

Mismatches relative to the on-target site are shown as lower-case letters. Single nucleotide polymorphisms or indels that differ from the C57BL/6N mouse strain are shown as red-coloured letters. Targeted amplicon sequencing of sites for gM (first row) was performed with the same genomic DNA from wild-type and knock-in mice used for CIRCLE-seq experiments (data in Supplementary Table 7). N.D. = not done.

Extended Data Table 3 | Numbers of closely matched sites in the mouse genome with canonical NGG, alternate NAG and other alternate non-NGG or non-NAG PAMs for gP, gM and gMH

Guide RNA	Method for finding off-target sites	Canonical NGG PAM	NAG PAM	Non-canonical Non-NAG	Total sites
gP	Found in mouse genome by Cas-OFFinder	120773	64538	78925	264236
gP	Found by CIRCLE-seq	472	2668	263	3403
gM	Found in mouse genome by Cas-OFFinder	64274	10285	37421	111980
gM	Found by CIRCLE-seq	130	25	28	183
gMH	Found in mouse genome by Cas-OFFinder	68160	16404	79049	163613
gMH	Found by CIRCLE-seq	448	52	30	530

Sites with NAG PAM

Number of mismatches	gP sites	gMH sites	gM sites
1	0	0	0
2	14	0	0
3	331	4	0
4	11349	26	8
5	7817	220	122
6	12416	1858	1676
7	32611	14296	8479

Top, the total numbers of these sites identified by Cas-OFFinder and CIRCLE-seq.
Bottom, sites with alternate NAG PAMs identified by Cas-OFFinder are shown by the number of mismatches present in the spacer region for each gRNA.

nature research

Corresponding author(s): J. Keith Joung and Marcello Maresca

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see Authors & Referees and the Editorial Policy Checklist.

Statistical parameters

When statistical analyses are reported, confirm that the following items are present in the relevant location (e.g. figure legend, table legend, main text, or Methods section).

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒	An indication of whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☐	☒	A description of all covariates tested
☐	☒	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒	A full description of the statistics including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒	For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☐	☒	Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated
☐	☒	Clearly defined error bars *State explicitly what error bars represent (e.g. SD, SE, CI)*

Our web collection on statistics for biologists may be useful.

Software and code

Policy information about availability of computer code

Data collection	For guide RNA design we used AstraZeneca proprietary software as an in silico tool, which was developed based on Wellcome Trust Sanger Institute`s codebase (WGE: http://www.sanger.ac.uk/htgt/wge/) with the addition of the NAG PAM motif as well as NGG in the alignments for potential off-targets. In silico mismatch data was obtained with Cas Offinder software (https://github.com/snugel/cas-offinder).
Data analysis	Data visualization and statistical analyses for plasma protein measurements were performed using GraphPad Prism 7.02. For circle-seq (data analyzed with CIRCLE-seq software: https://github.com/tsailabSJ/circleseq) and amplicon sequencing data (analyzed with CRISPResso: https://github.com/lucapinello/CRISPResso) was visualized and statistically tested with R, version 3.4.3.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers upon request. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

Sequence data that support the findings of this study have been deposited with SRA accession number SRP151131 and SUB4194113. Raw data for Extended Data Figures 1 and 6 available in Supplementary Figure 1.

Field-specific reporting

Please select the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/authors/policies/ReportingSummary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	No sample size calculation was performed. Sample size was determined as generation of triple independent samples for comparisons between groups that is sufficient to perform statistical tests.
Data exclusions	No data were excluded.
Replication	The experimental findings can be reliably reproducible. Some experiments producing the key data were repeated by different co-authors.
Randomization	Animals were randomized based on their weights measured prior to the experiments.
Blinding	The investigators were blinded to group allocation during data collection and analysis, except Surveyor assays - due to the need of visual organization of the groups to generate easily readable images.

Reporting for specific materials, systems and methods

Materials & experimental systems

n/a	Involved in the study
☐	☒ Unique biological materials
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology
☐	☒ Animals and other organisms
☒	☐ Human research participants

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging

Unique biological materials

Policy information about availability of materials

Obtaining unique materials	The original research article describing the humanized PCSK9 knock-in mouse model is currently under review for publication. AstraZeneca can share the animal model under a signed material transport agreement after that article is accepted for publication.

Animals and other organisms

Policy information about studies involving animals; ARRIVE guidelines recommended for reporting animal research

Laboratory animals	C57BL/6N mice (Charles River, Sulzfeld, Germany), male, 9-11 weeks old. Humanized PCSK9 knock-in mouse model is derived from C57BL/6N mice, male, 9-11 weeks old.
Wild animals	N/A
Field-collected samples	N/A