Unexpected mutations after CRISPR–Cas9 editing *in vivo*

To the Editor: CRISPR–Cas9 editing shows promise for correcting disease-causing mutations. For example, in a recent study we used CRISPR-Cas9 for sight restoration in blind *rd1* mice by correcting a mutation in the *Pde6b* gene[1]. However, concerns persist regarding secondary mutations in regions not targeted by the single guide RNA (sgRNA)[2]. Algorithms generate likely off-target sites for a given gRNA, but these algorithms may miss mutations. Whole-genome sequencing (WGS) has been used to assess the presence of small insertions and deletions (indels)[3] but not to probe for single-nucleotide variants (SNVs) in a whole organism. We performed WGS on a CRISPR–Cas9-edited mouse to identify all off-target mutations and found an unexpectedly high number of SNVs compared with the widely accepted assumption that CRISPR causes mostly indels at regions homologous to the sgRNA.

We tested four sgRNAs in cells then chose the sgRNA with the highest activity for *in vivo* targeting. DNA was isolated from two CRISPR-repaired mice (F03 and F05) and one uncorrected control[1]. CRISPR–Cas9-treated mice were sequenced at an average depth of 50×, and the control was sequenced at 30×. Variant calls were confirmed by at least 23× sequencing coverage (**Supplementary Tables 1** and **2**). Multiple variant-calling software pipelines identified indels and SNVs (**Fig. 1** and **Supplementary Methods**).

In the CRISPR-treated mice, targeted alleles were repaired[1]. Off-target mutations were identified as those present in the CRISPR-treated animals but absent in the uncorrected control. All pipelines showed that F03 harbored 164 indels and 1,736 SNVs (63 and 885 of these, respectively, associated with known genes). F05 harbored 128 indels and 1,696 SNVs (51 and 865 of these, respectively, associated with known genes) (**Fig. 1**). The same 117 indels and 1,397 SNVs were detected in both of the CRISPR-treated mice, which indicated nonrandom targeting. SNVs appeared to slightly favor transitions over transversions (**Supplementary Fig. 1**). The mutation rate detected in CRISPR-treated mice was substantially higher than that generated by spontaneous germline mutations (3 to 4 indels and 90 to 100 SNVs, *de novo*, per generation)[4,5].

As additional controls, each of the variants was compared with the FVB/NJ genome in the mouse dbSNP database (v138), and each of the SNVs was also compared with all 36 strains in the Mouse Genome Project (v3). None of the CRISPR-generated off-target mutations were found in any of these strains, which further confirmed that these WGS-identified SNVs were the result of CRISPR–Cas9 off targeting.

All pipelines identified 6 and 3 indels and 60 and 51 SNVs in F03 and F05 mice, respectively, in exonic regions only (**Fig. 1**); 5 indels and 24 SNVs caused nonsynonymous mutations in protein-coding sequences (**Supplementary Tables 3** and **4**). Of these, all five indels and one SNV (introducing a premature stop codon) were expected to be deleterious. Several mutated protein-coding genes were associated with a human and/or mouse phenotype (**Supplementary Tables 3** and **4**). Of the 29 coding-sequence variants, 7 variants were mutated identically in both mice. 24 CRISPR-associated variants were selected, and all were confirmed by Sanger sequencing (**Supplementary Fig. 2** and **Supplementary Methods**). Among the top-fifty sequences predicted for off targeting, none were mutated. Additionally, there was poor sequence homology between the sgRNA and sequences near the actual off-target coding and noncoding variants (**Supplementary Fig. 3**). Our results suggest current *in silico* modeling cannot predict bona fide off-target sites.

Together, these results indicate that at least certain sgRNAs may target loci independently of their target *in vivo*. The unpredictable generation of these variants is of concern. The impact of the numerous mutations occurring in noncoding RNAs or other regulatory intragenic regions could be detrimental to key cellular processes (**Supplementary Fig. 4** and **Supplementary Table 5**)[6]. Although our CRISPR-treated mice did not display obvious extraocular phenotypes, it is possible the mice may reveal phenotypes in time, when they are challenged or bred to homozygosity.

The present study demonstrates WGS analysis of both indels and SNVs as the most thorough method for identifying off-target mutations and shows a significantly higher number of potentially deleterious CRISPR–Cas9-induced mutations than have been previously reported[3]. It is not clear whether improved sgRNA
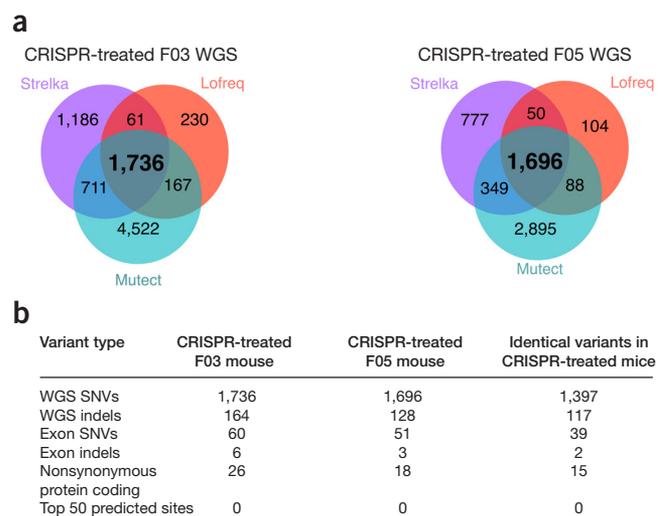
b

Variant type	CRISPR-treated F03 mouse	CRISPR-treated F05 mouse	Identical variants in CRISPR-treated mice
WGS SNVs	1,736	1,696	1,397
WGS indels	164	128	117
Exon SNVs	60	51	39
Exon indels	6	3	2
Nonsynonymous protein coding	26	18	15
Top 50 predicted sites	0	0	0

Figure 1 | CRISPR gene correction introduces an unexpectedly high number of mutations in a mouse model of gene therapy. (**a**) Venn diagrams display SNVs detected in WGS data by the indicated bioinformatics tools. (**b**) Summary of variant types found in mouse F03, mouse F05, and shared between both mice.

design or use of high-fidelity Cas9 may reduce off-target mutations, or whether *in vivo* off targets are a general problem of any sgRNA. Our study places the onus on researchers to carefully assay their specific gRNA and Cas9 for off-target mutations. More work may be needed to increase the fidelity of CRISPR–Cas9 with regard to off-target mutation generation before the CRISPR platform can be used without risk, especially in the clinical setting. Future studies employing new CRISPR methods and reagents should consider using WGS to determine the presence of off-target mutations *in vivo*.

Data availability statement. Experimental design information is available at Bioproject (accession PRJNA382177). Sequencing data available at SRA: FVB mouse (accession SRR5450998), F03 mouse (accession SRR5450997), F05 mouse (accession SRR5450996).

Note: Any Supplementary Information and Source Data files are available in the online version of the paper.

Kellie A Schaefer[1,2,8], Wen-Hsuan Wu[3,4,8], Diana F Colgan[1], Stephen H Tsang[3,4], Alexander G Bassuk[5,6] & Vinit B Mahajan[1,7]

[1]Omics Laboratory, Stanford University, Palo Alto, California, USA. [2]Interdisciplinary Graduate Program in Genetics, University of Iowa, Iowa City, Iowa, USA. [3]Jonas Children's Vision Care and Bernard & Shirlee Brown Glaucoma Laboratory, Departments of Ophthalmology, Pathology & Cell Biology, Institute of Human Nutrition, College of Physicians and Surgeons, Columbia University, New York, New York, USA. [4]Edward S. Harkness Eye Institute, New York-Presbyterian Hospital, New York, New York, USA. [5]Department of Pediatrics, University of Iowa, Iowa City, Iowa, USA. [6]Department of Neurology, University of Iowa, Iowa City, Iowa, USA. [7]Byers Eye Institute, Stanford University, Palo Alto, CA, USA [8]These authors contributed equally to this work. Correspondence should be addressed to V.B.M. (vinit.mahajan@stanford.edu) or A.G.B. (alexander-bassuk@uiowa.edu).

AUTHOR CONTRIBUTIONS
K.A.S., W.-H.W., S.H.T., A.G.B., and V.B.M. designed experiments. S.H.T., A.G.B., and V.B.M. managed this study. W.-H.W. carried out mouse experiments. K.A.S., D.F.C., S.H.T., A.G.B., and V.B.M. analyzed sequencing data. K.A.S., W.H.W., S.H.T., D.F.C., A.G.B., and V.B.M. wrote the manuscript. All authors participated in discussion and revision of the manuscript.

COMPETING FINANCIAL INTERESTS
The authors declare no competing financial interests.

1. Wu, W.H. *et al.* CRISPR repair reveals causative mutation in a preclinical model of retinitis pigmentosa. *Mol. Ther.* **24**, 1388–1394 (2016).
2. Koo, T., Lee, J. & Kim, J.S. Measuring and reducing off-target activities of programmable nucleases including CRISPR–Cas9. *Mol. Cells* **38**, 475–481 (2015).
3. Iyer, V. *et al.* Off-target mutations are rare in Cas9-modified mice. *Nat. Methods* **12**, 479 (2015).
4. Nachman, M.W. & Crowell, S.L. Estimate of the mutation rate per nucleotide in humans. *Genetics* **156**, 297–304 (2000).
5. Roach, J.C. *et al.* Analysis of genetic inheritance in a family quartet by whole-genome sequencing. *Science* **328**, 636–639 (2010).
6. Mattick, J.S. & Makunin, I.V. Non-coding RNA. *Hum. Mol. Genet.* **15**, R17–R29 (2006).

Digenome-seq web tool for profiling CRISPR specificity

To the Editor: We recently reported Digenome-seq (digested genome sequencing), a method for *in vitro* identification of potential off-target sites, and we evaluated the specificity of CRISPR–Cas9 (refs. 1,2) and CRISPR–Cpf1(ref. 3) endonuclease by whole-genome sequencing. Digenome-seq pinpoints the exact location of double-strand break (DSB) sites by recognizing specific patterns of aligned reads. However, the analysis pipeline presented in our previous report required extensive manual interaction and produced several large intermediate files, resulting in a long running time. Here, we present a redesigned analysis tool for Digenome-seq data that runs on web browsers. The core algorithm of the tool is written in C++ and compiled to asm.js (http://asmjs.org/), a preoptimized subset of JavaScript. Users can instantly perform the complete analysis in an ordinary web browser (**Supplementary Note 1**) with fast execution speed without uploading any data to a server and without local tool installation. In our benchmark, the full analysis for 100 GB of BAM file took 3 h for whole analysis on Intel i5 3570k central processing unit in a single thread.

With the new tool, a user submits both the mock- and nuclease-treated BAM files (**Fig. 1a**) together with target sequences and required parameters (**Supplementary Note 2**). The result page is then displayed and updated in real time. Next, the tool finds patterns of aligned reads that characterize DSB sites. 80-bp flanking sequences around each site are retrieved from the Ensembl database[4] and aligned to the target sequence using a semiglobal alignment algorithm for user convenience. Each site is reported along with the alignment, the number of DSB sites per chromosome (**Fig. 1b**) and in an interactive plot (**Fig. 1c**). The web tool is freely available at http://www.rgenome.net/digenome-js.

Note: Any Supplementary Information and Source Data files are available in the online version of the paper.
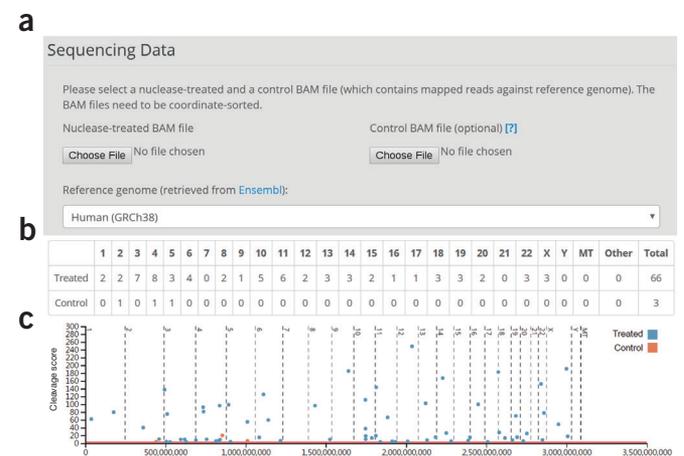
Figure 1 | Digenome-seq overview. (**a**) Main input form. (**b**) Summary table showing the number of candidate sites on each chromosome and mitochondrion (MT). (**c**) Interactive summary plot showing the genomic position (the x-axis) and DNA cleavage score (the y-axis) for each candidate. The minimum cleavage score cutoff is depicted with a red line.