prefer, a valid alibi—for the arrangement of faster authorization of the numerous future GMOs on which the European Food Safety Agency will, in all likelihood, express a positive opinion in terms of their health and environmental impact. In fact, the EU's official list of GMOs authorized for import is not so short: 58 items have been imported for years, plus 19 cleared on April 24, 2015 (ref. 13), and some 40 requests are still pending.

COMPETING FINANCIAL INTERESTS
The author declares no competing financial interests.

Giovanni Tagliabue

Carugo, Como, Italy.
e-mail: giovanni.tagliabue@uniedi.com

1. Ammann, K. *N. Biotechnol.* **31**, 1–17 (2014).
2. Tagliabue, G. *Nat. Biotechnol.* **33**, 907–908 (2015).
3. Wager, R. & McHughen, A. *EMBO Rep.* **11**, 258–262 (2010).
4. European Union. Directive 2001/18/EC of the European Parliament and of the Council of 12 March 2001 on the deliberate release into the environment of genetically modified organisms and repealing Council Directive 90/220/EEC. http://europa.eu/legislation_summaries/agriculture/food/l28130_en.htm (EU, 2001).
5. European Union. *Regulation (EC) No* 1829/2003 *of the European Parliament and of the Council of 22 September 2003 on genetically modified food and feed.* http://eur-lex.europa.eu/legal-content/EN/TXT/?qid=1427098789359&uri=URISERV:l21154 (EU, 2003).
6. European Union. Directive (EU) 2015/412 of the European Parliament and of the Council of 11 March 2015 amending Directive 2001/18/EC as regards the possibility for the Member States to restrict or prohibit the cultivation of genetically modified organisms (GMOs) in their territory. http://eur-lex.europa.eu/legal-content/EN/TXT/?qid=1426590211658&uri=OJ:JOL_2015_068_R_0001 (EU, 2015).
7. Rabesandratana, T. *Science* **350**, 18–19 (2015).
8. European Commission. Proposal for a Regulation of the European Parliament and of the Council amending Regulation (EC) No 1829/2003 as regards the possibility for the Member States to restrict or prohibit the use of genetically modified food and feed on their territory, COM/2015/0177 final - 2015/0093 (COD). http://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:52015PC0177 (EU, 2015).
9. European Parliament. GMO imports: MEPs object to draft law allowing national bans, call for plan B. http://www.europarl.europa.eu/news/en/news-room/content/20150713IPR80661/html/GMO-imports-MEPs-object-to-draft-law-allowing-national-bans-call-for-plan-B (EP, 2015).
10. Baulcombe, D. *et al. GM science update. A report to the Council for Science and Technology.* http://www.gov.uk/government/publications/genetic-modification-gm-technologies (UK Government, London, March 2014)
11. European Parliament. Parliament rejects national GMO bans proposal. http://www.europarl.europa.eu/news/en/news-room/content/20151022IPR98805/html/Parliament-rejects-national-GMO-bans-proposal (EP; 2015).
12. European Commission. *Economic Impact of Unapproved GMOs on EU Feed Imports and Livestock Production.* http://ec.europa.eu/agriculture/envir/gmo/economic_impactGMOs_en.pdf (EC; Brussels, 2007).
13. European Commission. Commission authorises 17 GMOs for food/feed uses and 2 GM carnations. http://europa.eu/rapid/press-release_IP-15-4843_en.htm (EC; Brussels, 2015).

Analyzing CRISPR genome-editing experiments with CRISPResso

To the Editor:

Recent progress in genome editing technologies, in particular the CRISPR-Cas9 nuclease system, has provided new opportunities to investigate the biological functions of genomic sequences by targeted mutagenesis[1–4]. Double-strand breaks (DSBs) arising from site-specific Cas9 cleavage can be resolved by non-homologous end joining (NHEJ) or homology-directed repair (HDR), which result in a spectrum of diverse outcomes, including insertions, deletions, nucleotide substitutions and, in the case of HDR, recombination of extrachromosomal donor sequences[1–3,5,6]. Deep sequencing of amplified genomic regions or whole genome sequencing (WGS) allows quantitative and sensitive detection of targeted mutations. However, to date, no standard analytic tool has been developed to systematically enumerate and visualize these events, resulting in inconsistencies among different experiments and across laboratories. Challenging issues for the interpretation of CRISPR-Cas9-edited sequences include amplification or sequencing errors, experimental variation in sequence quality, ambiguous alignment of variable length indels, deconvolution of mixed HDR–NHEJ outcomes, and analytical complexities resulting from large WGS data sets and pooled experiments where many different target sites are present in a single sequencing library. To both solve these issues and attempt to standardize data analysis, we developed CRISPResso, a robust and easy-to-use computational pipeline (**Supplementary Note 1** and **Supplementary Fig. 1**). CRISPResso enables accurate quantification and visualization of CRISPR-Cas9 outcomes, as well as comprehensive evaluation of effects on coding sequences, non-coding elements and selected off-target sites.

CRISPResso is a suite of computational tools to qualitatively and quantitatively evaluate the outcomes of genome-editing experiments in which target loci are subject to deep sequencing. It provides an integrated, user-friendly interface that can be operated by biologists and bioinformaticians alike (**Supplementary Figs. 1** and **2**). Compared with existing tools[7], CRISPResso offers several notable features, including the following: batch sample analysis by command line interface; integration with other pipelines; tunable parameters of sequence quality and alignment fidelity; discrete measurement of insertions, deletions and nucleotide substitutions (which are ignored by other methods); tunable windows around the cleavage site to minimize false-positive classification; quantification of frameshift versus in-frame coding mutations; and distinction between NHEJ, HDR and mixed mutation events. CRISPResso automates the following steps: first, filtering low-quality reads; second, trimming adapters; third, aligning the reads to a reference amplicon; fourth, quantifying the proportion of HDR and NHEJ outcomes; and fifth, determining the proportion of frameshift and in-frame mutations as well as detecting potential splice-site mutations. A graphical report is generated to visualize mutagenesis profiles (**Fig. 1** and **Supplementary Figs. 3–5**), and plain text output files are also produced for further integrative analyses (**Supplementary Note 2**). This pipeline can be used for assessment of on-target editing efficacy as well as of off-target editing at selected loci[8,9].

We initially assessed the performance and limitations of CRISPResso by performing simulations with various genome-editing outcomes, with and without sequencing errors included (**Supplementary Note 3** and **Supplementary Figs. 6–9**). We found that CRISPResso, even in the presence of sequencing errors, robustly and accurately recovered editing events with a negligible false-positive rate (<0.1%). Then we applied CRISPResso to experimental paired-end deep-sequencing data either from cells expressing Cas9 and single guide RNA (sgRNA)-1 targeted to the *HBB* coding sequence (experiment 1) or from cells expressing Cas9, an extrachromosomal homologous donor template and either sgRNA-2 (experiment 2) or sgRNA-1 (experiment 3), with the intent of targeted introduction of four nucleotide substitutions at *HBB* (**Supplementary Note 4** and **Supplementary Figs. 3–5,10**). For experiment 1, CRISPResso provides a quantification of the proportion of NHEJ occurrences, mutated allele size distribution and precise mutation localization with respect to the reference amplicon (**Fig. 1a–c**). When coding sequences were provided as an optional input, the software quantified

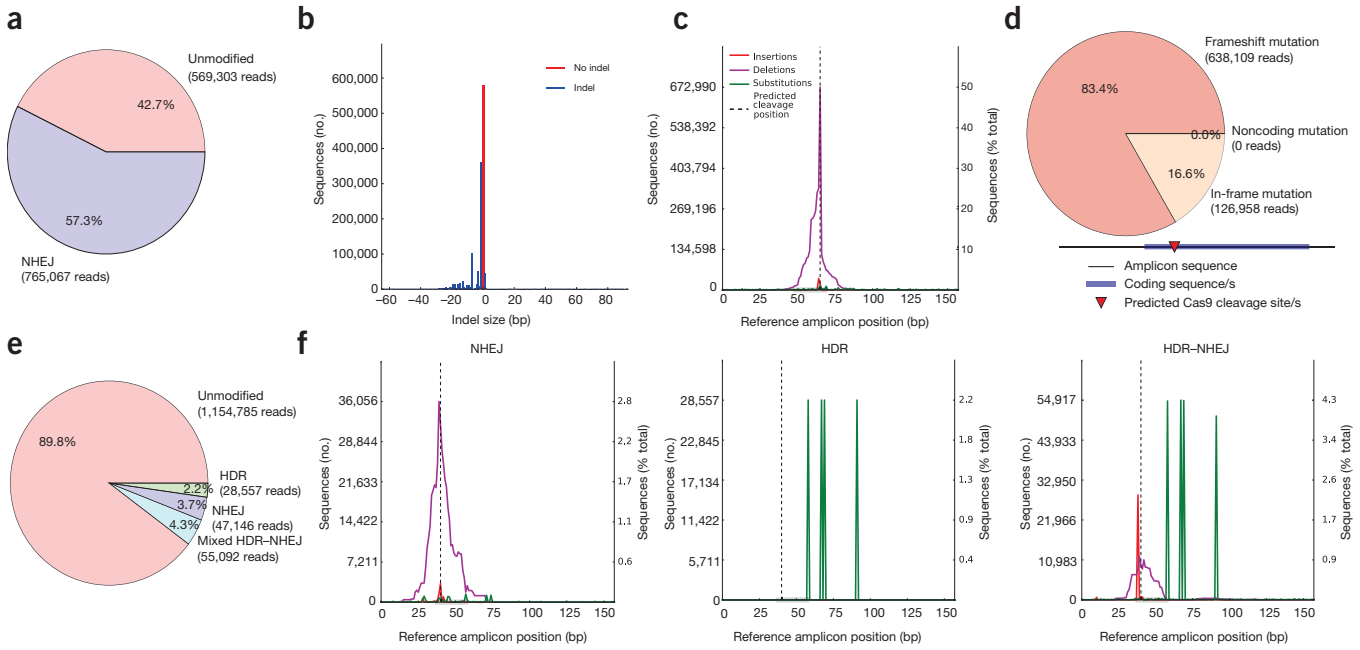

Figure 1 Quantification and visualization of NHEJ and HDR mutagenesis profiles. (**a–d**) An example of NHEJ-mediated disruption of a coding sequence by CRISPR-Cas9 (experiment 1). (**a**) Quantification of editing frequency as determined by the percentage and number of sequence reads showing unmodified and modified alleles. When no donor sequence is provided, CRISPResso classifies any mutation overlapping a window around the expected cleavage site/s as an NHEJ event. (**b**) Frequency distribution of alleles with indels and without indels. Length-conserving substitutions are not classified as indels in this plot. In this example, the indels are dominated by small deletions, consistent with the anticipated CRISPR-Cas9 effect. (**c**) NHEJ reads with insertions, deletions and substitutions mapped to reference amplicon. For insertions, the positions immediately adjacent to the insertion are indicated. In this example, the mutations cluster around the predicted cleavage position, consistent with the anticipated CRISPR-Cas9 effect. A low level of substitutions apparent throughout the amplicon suggests low-level technical error, although these errors do not contribute to the quantification of the NHEJ. (**d**) Frameshift analysis of coding sequence reads affected by modifications. Frameshift and in-frame mutations include any mutations that partially or fully overlap coding sequences as input by the user, with any non-overlapping mutations classified as noncoding (see also **Supplementary Fig. 11**). (**e**,**f**) HDR-mediated recombination of an extrachromosomal donor sequence resulting in four substitutions relative to the reference amplicon (experiment 2). (**e**) When an expected HDR amplicon is provided, CRISPResso classifies sequence reads as HDR if they preferentially align to the expected HDR amplicon sequence and NHEJ (or unmodified) if they preferentially align to the reference amplicon. An alignment threshold may be provided to distinguish HDR alleles from those showing evidence of mixed HDR–NHEJ repair. (**f**) Mapping of mutation position to reference amplicon of reads classified as NHEJ (left), HDR (center) and mixed HDR–NHEJ (right). In this example with the alignment threshold set to 100% sequence identity, the HDR alleles show only the four expected substitutions, whereas the mixed HDR–NHEJ alleles show additional indels at the predicted cleavage position, consistent with sequential cleavages initially repaired by HDR and subsequently by NHEJ.

frameshift and in-frame mutations and predicted splice site mutations (**Fig. 1d** and **Supplementary Fig. 11**). When an expected HDR amplicon sequence was provided (experiments 2 and 3), CRISPResso was able to deconvolve and characterize unmodified, NHEJ-modified and HDR-modified alleles as distinct outcomes (**Fig. 1e**,**f**, **Supplementary Note 1** and **Supplementary Figs. 3–5**). In addition, it identified mixed alleles that may result from sequential cleavages initially resulting in HDR and later NHEJ repair (**Fig. 1f**). In a case when the donor sequence disrupted the guide RNA seed sequence or the protospacer-adjacent motif (PAM), the relative fraction of mixed events appeared substantially reduced, consistent with the effect of these HDR alleles on resisting subsequent cleavage (**Supplementary Figs. 12** and **13**). By specifying the sequence identity required to classify an event as HDR, the user can control the specificity of HDR and the sensitivity of mixed HDR–NHEJ allele detection (**Supplementary Figs. 12** and **13**).

In addition, the CRISPResso suite accommodates single- or pooled-amplicon deep sequencing and WGS data sets and allows the direct comparison of individual experiments. In fact, four additional utilities are provided (**Supplementary Fig. 1**).

First, CRISPRessoPooled is a tool for the analysis of pooled amplicon experiments that first preprocesses the input data to highlight and remove PCR amplification or trimming artifacts. This tool is recommended to be run in a mixed mode with alignment to both a reference genome and a list of amplicons to help resolve alignment artifacts or contamination (although it may also be run in individual modes). The outputs of CRISPRessoPooled include individual CRISPResso reports with detailed mapping statistics for each region as well as a summary table for all target regions (**Supplementary Note 5**, **Supplementary Figs. 14**, **15** and **Supplementary Tables 1–6**). Second, CRISPRessoWGS is a tool for the analysis of WGS data that provides detailed CRISPResso reports for any set of sites throughout the genome (e.g., potential off-target sites) and separate .bam files (for discrete visualization in a genome browser) (**Supplementary Note 6** and **Supplementary Fig. 16**). Third, CRISPRessoCompare is a tool for the comparison of two CRISPResso analyses, useful, for example, to compare treated and untreated samples or to compare different experimental conditions (**Supplementary Note 7** and **Supplementary Fig. 17**). And finally, CRISPRessoPooledWGSCompare is a tool to compare experiments involving several regions analyzed by either CRISPRessoPooled or CRISPRessoWGS (**Supplementary Note 8** and **Supplementary Table 7**).

All the analysis presented in this manuscript and the source code of

CRISPResso can be downloaded from **Supplementary Data 1–10** with the most updated version available at github (http://github.com/lucapinello/CRISPResso) and the experimental data produced deposited in the GEO database (GSE78729). CRISPResso can be run either as a stand-alone command line utility (http://github.com/lucapinello/CRISPResso) or as a web application (http://www.crispresso.rocks; **Supplementary Note 2**).

In summary, the CRISPResso suite offers flexible and powerful tools not only to evaluate and quantify genome editing outcomes from sequencing experiments, but also to standardize and streamline analyses that currently require development of custom in-house algorithms.

Accession codes. GEO: GSE78729.

Editor's note: This article has been peer reviewed.

Note: Any Supplementary Information and Source Data files are available in the online version of the paper (doi:10.1038/nbt.3583).

ACKNOWLEDGMENTS

This work was supported by NHGRI Career Development Award K99HG008399 to L.P. M.C.C. is supported by F30DK103359-01A1. S.H.O. is supported by P01HL032262 and P30DK049216 (Center of Excellence in Molecular Hematology). D.E.B. is supported by NIDDK Career Development Award K08DK093705, Doris Duke Charitable Foundation, Charles H. Hood Foundation, Cooley's Anemia Foundation, American Society of Hematology and Burroughs Wellcome Fund. G.-C.Y. is supported by NIH R01HL119099 and R01HG005085.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

Luca Pinello[1,2], Matthew C Canver [3–5], Megan D Hoban [6], Stuart H Orkin [3–5,7], Donald B Kohn[6], Daniel E Bauer [3–5] & Guo-Cheng Yuan[1,2]

[1]Department of Biostatistics and Computational Biology, Dana-Farber Cancer Institute, Boston, Massachusetts, USA. [2]Department of Biostatistics, Harvard TH Chan School of Public Health, Boston, Massachusetts, USA. [3]Division of Hematology/Oncology, Boston Children's Hospital, Boston, Massachusetts, USA. [4]Department of Pediatric Oncology, Dana-Farber Cancer Institute, Boston, Massachusetts, USA. [5]Department of Pediatrics, Harvard Medical School and Harvard Stem Cell Institute, Harvard University, Boston, Massachusetts, USA. [6]Department of Microbiology, Immunology, and Molecular Genetics, Eli & Edythe Broad Center of Regenerative Medicine & Stem Cell Research, University of California, Los Angeles, Los Angeles, California, USA. [7]Howard Hughes Medical Institute, Boston, Massachusetts, USA.
e-mail: lpinello@jimmy.harvard.edu or daniel.bauer@childrens.harvard.edu; or gcyuan@jimmy.harvard.edu

1. Cong, L. *et al. Science* **339**, 819–823 (2013).
2. Mali, P. *et al. Science* **339**, 823–826 (2013).
3. Ran, F.A. *et al. Nat. Protoc.* **8**, 2281–2308 (2013).
4. Sander, J.D. & Joung, J.K. *Nat. Biotechnol.* **32**, 347–355 (2014).
5. Canver, M.C. *et al. J. Biol. Chem.* **289**, 21312–21324 (2014).
6. Ran, F.A. *et al. Cell* **154**, 1380–1389 (2013).
7. Güell, M., Yang, L. & Church, G.M. *Bioinformatics* **30**, 2968–2970 (2014).
8. Hsu, P.D. *et al. Nat. Biotechnol.* **31**, 827–832 (2013).
9. Fu, Y. *et al. Nat. Biotechnol.* **31**, 822–826 (2013).

Data mining differential clinical outcomes associated with drug regimens using adverse event reporting data

To the Editor:

The US Food and Drug Administration (FDA) Adverse Event Reporting System (FAERS) contains millions of clinical records that detail drug exposure(s), disease indications and clinical outcomes. However, a variety of challenges limit the ability of these data to be exploited for the extraction of clinically meaningful information beyond single-agent safety signal associations. Here, we report an open access web data mining tool, AERS*Mine*, that enables data exploration and hypothesis generation among alternatively medicated cohorts to discover patterns of differential outcomes. Cohort groups can be formed as a function of demographics, underlying disorders, drug classes and other user-driven constraints across the entire range of human diseases and approved drugs.

Understanding short-term and long-term clinical outcomes associated with drug therapies is challenging. Virtually all drugs can cause unwanted side effects, therapeutic efficacy can vary widely between individuals, and long-term outcomes of chronic regimens are highly confounded by indication-associated risks. In some cases, severe drug-associated adverse effects have become apparent only after the onset of treatment, and have led to withdrawal or restriction of drugs by the FDA. For example, troglitazone was withdrawn by the FDA in 2000 owing to increased risk of hepatotoxicity, whereas cerivastatin was withdrawn in 2001 owing to increased risk of rhabdomyolysis[1]. Furthermore, our ability to identify differential patterns of drug responses and clinical outcomes among population subgroups has been limited.

To address these challenges, the FDA and the World Health Organization (WHO; Geneva) conduct pharmacovigilance and monitor safety standards of approved drugs on the market. The FDA maintains the FAERS (http://www.fda.gov/Drugs/GuidanceComplianceRegulatoryInformation/Surveillance/AdverseDrugEffects/default.htm), which stores manually reviewed adverse event (AE) reports received by the FDA from healthcare professionals, drug manufacturers and consumers from around the world. Although the FAERS provides patient demographic details such as age, gender, clinical indications, drugs, AEs and outcomes that can be used to identify latent risks of approved therapeutics and their combinations[2–6], accurate mining of this information remains difficult.

AERS*Mine* (https://research.cchmc.org/aers) is a tool that effectively mines the FAERS data through systematic normalization, unification and ontological aggregation of the drugs, clinical indications and AEs (Online Methods, **Supplementary Tables 1** and **2** and **Supplementary Figs. 1–5**). This allows analysis of large clinical cohorts and comparison of differential long-term outcomes between treatment regimens. AERS*Mine* facilitates aggregation, subcategorization and simultaneous comparison of multiple patient groups, based on explicit combinations of demographics, clinical indication(s), and exposure to different drug(s) or drug classes. For instance, AERS*Mine* allows users to create mutually exclusive treatment cohorts (e.g., drugs *a* AND *b*, OR *c*, NOT *x*, *y*, *z*) with normalized row values of clinical events (AEs) and generate data matrices for pattern-based analyses of differential AE risks and incidence rates (**Fig. 1**). These high-resolution analyses can enable the detection of differential effects between varied drug classes and within specific, defined patient subgroups that would be otherwise confounded by mixing population subgroups that differ in their relative risks of specific clinical events or outcomes.