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


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 mutagenesis1–3. 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 sequences1,3–5. 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 tools1, 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 ampiclon; 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 loci8,9.

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 ampiclon (Fig. 1a–c).

When coding sequences were provided as an optional input, the software quantified
frameshift and in-frame mutations and predicted splice site mutations (Fig. 1d and Supplementary Fig. 11). When an expected HDR ampiclon 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. 1c, 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 frequency 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, CRISPRessoWGSCompare 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
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, AERSMine, 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.

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, accurate mining of this information remains difficult.

AERSMine (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. AERSMine 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, AERSMine 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.