Evaluation of CRISPR gene-editing tools in zebrafish

José M. Uribe-Salazar¹,², Gulhan Kaya¹, Aadithya Sekar¹, KaeChandra Weyenberg¹, Cole Ingamells¹ and Megan Y. Dennis¹,²*

Abstract

Background: Zebrafish have practical features that make them a useful model for higher-throughput tests of gene function using CRISPR/Cas9 editing to create ‘knockout’ models. In particular, the use of G₀ mosaic mutants has potential to increase throughput of functional studies significantly but may suffer from transient effects of introducing Cas9 via microinjection. Further, a large number of computational and empirical tools exist to design CRISPR assays but often produce varied predictions across methods leaving uncertainty in choosing an optimal approach for zebrafish studies.

Methods: To systematically assess accuracy of tool predictions of on- and off-target gene editing, we subjected zebrafish embryos to CRISPR/Cas9 with 50 different guide RNAs (gRNAs) targeting 14 genes. We also investigate potential confounders of G₀-based CRISPR screens by assaying control embryos for spurious mutations and altered gene expression.

Results: We compared our experimental in vivo editing efficiencies in mosaic G₀ embryos with those predicted by eight commonly used gRNA design tools and found large discrepancies between methods. Assessing off-target mutations (predicted in silico and in vitro) found that the majority of tested loci had low in vivo frequencies (< 1%). To characterize if commonly used ‘mock’ CRISPR controls (larvae injected with Cas9 enzyme or mRNA with no gRNA) exhibited spurious molecular features that might exacerbate studies of G₀ mosaic CRISPR knockout fish, we generated an RNA-seq dataset of various control larvae at 5 days post fertilization. While we found no evidence of spontaneous somatic mutations of injected larvae, we did identify several hundred differentially-expressed genes with high variability between injection types. Network analyses of shared differentially-expressed genes in the ‘mock’ injected larvae implicated a number of key regulators of common metabolic pathways, and gene-ontology analysis revealed connections with response to wounding and cytoskeleton organization, highlighting a potentially lasting effect from the microinjection process that requires further investigation.

Conclusion: Overall, our results provide a valuable resource for the zebrafish community for the design and execution of CRISPR/Cas9 experiments.

Keywords: Danio rerio, Zebrafish, CRISPR, Cas9, Gene knockout, CIRCLE-seq, RNA-seq

*Correspondence: mydennis@ucdavis.edu

¹ Integrative Genetics and Genomics Graduate Group, University of California, Davis, Davis, CA, USA
Full list of author information is available at the end of the article

© The Author(s) 2021. Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/. The Creative Commons Public Domain Dedication waiver (http://creativecommons.org/publicdomain/zero/1.0/) applies to the data made available in this article, unless otherwise stated in a credit line to the data.
transmitted de novo single-nucleotide mutations due to CRISPR-editing using exome sequencing and a stringent bioinformatic pipeline [30] in a similar approach used to identify off-target mutations in mouse trios [31, 32]. Other studies have observed off-target mutation rates ranging from 0.07 to 3.17% in zebrafish by sequencing the top three to four predicted off-target regions based on sequence homology [11, 12, 18]. Although off-target mutations should not significantly impact studies of stable mutants, since unwanted mutations can be outcrossed out of studied lines relatively easily [14, 21], they may be problematic in rapid genetic screens using $G_0$ mosaics that quickly test gene functions in a single generation.

The increasing number of tools available for the design and execution of CRISPR screens provide an important resource to the zebrafish community. Here, we assayed different available CRISPR on- and off-target prediction methods using empirical data from Cas9-edited zebrafish embryos. We quantified CRISPR cleavage efficiencies in vivo employing a variety of experimental approaches and used these results to compare the accuracy of in silico and in vitro tools for predicting Cas9 on- and/or off-target activity. Finally, to examine potential confounders that may arise from microinjection of Cas9 into embryos on resulting phenotypes, we assayed $G_0$ ‘mock’ negative control embryos injected with a buffer containing either Cas9 enzyme or mRNA in the absence of gRNAs by performing RNA-seq and obtained a list of genes with significant differential expression versus un.injected wild-type siblings. In all, these results will serve as a useful resource to the research community as larger-scale $G_0$ CRISPR screens become more common in assaying gene functions in zebrafish.

Results

Identification of CRISPR-induced indels in zebrafish

We generated a dataset of experimentally confirmed indels within 14 protein-coding genes from injected NHGRI-1 wild-type zebrafish larvae targeted by 50 gRNAs (2–4 different gRNAs/target gene, assembled through the annealing of crRNA:tracrRNA) (Fig. 1A, Supplementary Tables 1 and 2). These 50 gRNAs were designed using CRISPRScan [26] and include a range of predicted editing efficiencies (mean 57.6, range 23–83). To obtain experimental in vivo editing efficiency values for each gRNA, DNA extracted from a pool of 20 $G_0$ mutant embryos—generated via microinjections of individual gRNAs at the one-cell stage and harvested at 5 days post-fertilization (dpf)—and ~200 bp regions surrounding predicted cut sites for all gRNAs were amplified and Illumina sequenced. To extract the proportion of reads carrying indel alleles, we used CrisprVariants [33] with uninject ed batch siblings DNA...
as reference (Supplementary Table 2 for all scores obtained). From this, we inferred an in vivo ‘efficiency score’, calculated as the percentage of DNA from injected embryos harboring indels compared to un.injected batch siblings (Fig. 1B).

To compare our efficiency scores with those produced from Sanger-based tools, we also amplified and sequenced ~500bp fragments surrounding the targeted sites from the same DNA. We extracted the percentage of indels using two different tools that deconvolve major mutations
and their frequencies within Sanger traces—Tracking of Indels by DEcomposition (TIDE) [34] and Inference from CRISPR Edits (ICE) [35] (Fig. 1B). Briefly, these tools use the gRNA sequence to predict the cutting site in the control sequence trace, map the sample trace to this reference, identify indels by deconvolving all base reads at each position, and provide a frequency of the indel spectrum [34, 35]. As previously reported [35], both tools provided positively correlated in vivo scores across all gRNAs (Spearman $\rho = 0.87$, $p = 6.78 \times 10^{-15}$) with an average score difference of 8.8±12.1 between tools (Fig. 1B). We noted a higher correlation between tools in scores below the median (Spearman $\rho = 0.96$, $p = 1.23 \times 10^{-15}$) than above the median (Spearman $\rho = 0.65$, $p = 0.00072$; Fig. 1B), suggesting that the deconvolution process of both tools is more accurate when fewer molecules from the pool carry indels. Both ICE and TIDE efficiency scores were correlated with our Illumina-based editing scores (ICE: Spearman $\rho = 0.88$, $p = 9.14 \times 10^{-16}$; TIDE: Spearman $\rho = 0.59$, $p = 7.33 \times 10^{-6}$; Fig. 1C), though they significantly underestimated editing efficiencies with, for example, Illumina estimates 19.4±16.3 higher than ICE estimated scores (Fig. 1C). Based on its higher correlation, we reported Sanger-based ICE in vivo scores for the rest of this study.

To ascertain consistency of editing efficiencies across embryos, we also repeated microinjections for four gRNAs targeting a single gene (srgap2) and assessed in vivo efficiencies scores of 20 individual larvae using the ICE tool. This resulted in low variance across injections and relative parity of efficiencies versus results from our pooled-larvae DNA preparations (e.g., low efficiency gRNA targeting exon 2; average±standard error ICE score in individual larvae 18.9±3.3 versus an ICE score of 13 in pooled larvae). We next compared the accuracy of CRISPR on-target predictions by in silico methods. We compared our in vivo results (Fig. 2A) to scores from four other in silico tools (E-CRISP, CRISP-GE, CCTop, CRISPRon) and observed a generalized underestimation of editing efficiency (Fig. 2A). Strikingly, only scores predicted by CRISPRScan exhibited significant, albeit weak, correlation with in vivo scores. Further, the correlation with Illumina-based scores was significant only when using our NHGRI-1 strain (Spearman $\rho = 0.31$, $p = 0.028$; Fig. 2B), though CRISPRScan scores were highly concordant with those obtained using the Tubingen-derived reference (Spearman $\rho = 0.88$, $p = 5.02 \times 10^{-17}$; Fig. 2B), with an average difference between scores of 4.2±4.6 (range 0–31) (Supplementary Table 2). CHOPCHOP values exhibited correlations with scores from four other in silico tools (E-CRISP, CRISPRon, DeepSpCas9, IDT) but none were concordant with our in vivo results (Fig. 2B). Additionally, two tools that utilize deep learning methods (CRISPRon and DeepSpCas9) were significantly correlated with each other but failed to predict in vivo editing efficiencies in our assay (Fig. 2B). Thus, despite the research community broadly adapting all methods for designing gRNAs, there is little consensus in

(See figure on next page.)

**Fig. 2** Correlation of on-target efficiencies calculated using different methods. A Heatmap of the efficiency scores obtained from the design tool (CRISPRScan), in silico prediction tools, and cutting cleavages obtained in vivo using Illumina sequencing and a deconvolution tool from Sanger sequencing [35] for 50 gRNAs. Each box represents a gRNA and the efficiency scores range from 0 (blue) to 100 (red). B Spearman correlations between all efficiency scores from in silico predictions, an in vitro protocol [28], and in vivo cutting assays. Each box includes the correlation result with the $p$-value in parenthesis. The color of the boxes represent the correlation values, ranging between −1 (blue) and 1 (red). CHOPCHOP scores were obtained using two different scoring methods, CHOPCHOP (D) (based on [39]) and CHOPCHOP (X) (based on [40]).
predicting activity of a particular gRNA among these tools. Based on our results, we recommend using CRISPRScan for choosing gRNAs in zebrafish experiments.

**Accuracy of CRISPR on-target predictions by an in vitro method**

Next, we evaluated the possibility of using the in vitro protocol CIRCLE-seq [28], an approach designed to identify target sites of a given gRNA by subjecting naked genomic DNA to Cas9 enzyme/gRNA cleavage followed by Illumina sequencing, to obtain an editing efficiency score. It is important to emphasize that such in vitro assays are not designed for predicting on-target editing efficiencies. Nevertheless, we sought to understand if such an approach could be used for this purpose. We tested individually the 50 gRNAs described above using the CIRCLE-seq protocol [47], following the standard recommendations, and computed a log enrichment score normalized by the sequence library size, termed reads per million normalized (RPMN) (see Methods). We found that in vitro-obtained enrichment scores were not correlated with in vivo efficiencies (Illumina: Spearman \( \rho = 0.10, \ p = 0.494 \); ICE: Spearman \( \rho = -0.02, \ p = 0.911 \), Fig. 2B) or with in silico predictions, with the exception of CHOPCHOP using the scoring method by [40] (Fig. 2). This indicates that the CIRCLE-seq assay does not necessarily predict on-target CRISPR cleavage activity, at least quantitatively. Previous work from in vivo CRISPR studies of zebrafish suggests that increased GC-content predicts increased activity of gRNAs [26]. Examining GC content of our tested gRNAs, ranging from 31.8 to 77.3%, we observed a positive correlation with CRISPRScan in silico scores (linear model: \( \beta = 68.18, \ p = 0.003, \text{adjusted-}R^2 = 0.16 \)) and CIRCLE-seq in vitro RPMN scores (linear model: \( \beta = 6.4, \ p = 0.006, \text{adjusted-}R^2 = 0.14 \)) (Supplementary Fig. 1A and B); however, our experimentally determined in vivo scores were not correlated with GC content (Illumina linear model: \( \beta = -18.57, \ p = 0.689, \text{adjusted-}R^2 = -0.02 \); ICE linear model: \( \beta = 12.36, \ p = 0.817, \text{adjusted-}R^2 = -0.02 \); Supplementary Fig. 1C), suggesting that additional variables should also be considered (e.g., depletion of A nucleotide bases, nucleosome positioning or DNA accessibility [26, 48]).

**CRISPR off-target mutation prediction methods**

To avoid spurious phenotypes, off-target mutations should be minimized when choosing gRNAs in CRISPR experiments. To characterize off-target mutations for our set of 50 gRNAs, we queried predictions from in silico (CRISPRScan) and in vitro (CIRCLE-seq) methods. CRISPRScan provides a list of predicted off-target sites (between 55 and 1350, median 206.5; Supplementary Table 3) for each gRNA within the zebrafish NHGRIz reference genome (GRCh11/danRer11) based on a cutting frequency determination (CFD) score that primarily takes into account sequence similarity, location, and type of sequence mismatches [26, 38]. The CIRCLE-seq empirical approach also produced variable numbers of sites (between 18 and 874, median 113.5; Supplementary Table 3) per gRNA (defined as ‘CIRCLE-seq sites’) relative to the control library digested solely with Cas9 enzyme. The number of off-target sites predicted by CRISPRScan exhibited a significant, albeit weak, correlation with the number of CIRCLE-seq sites per gRNA (Spearman \( \rho = 0.33, \ p = 0.022 \), Fig. 3A). Focusing on putatively impactful off-target predictions, an average of 20±13% CRISPRScan-predicted and 64±7% CIRCLE-seq sites per gRNA intersected at least one gene (Supplementary Table 3). The sites predicted in silico or in vitro intersecting genes predominantly did not overlap, with an average of 1.6±1.8 (range 0–7) genes per gRNA overlapping between the two approaches for the same gRNA.

To verify if predicted off-target sites were subjected to in vivo Cas9 cleavage, we performed Sanger sequencing of sites within genes identified in silico (\( n = 17 \)) and in vitro (\( n = 20 \)) for eight gRNAs, an average of six regions per gRNA (see Supplementary Table 1 for description of sites). Using the ICE tool, we found mosaic mutations at frequencies between 0 and 11%, with 23 out of the 37 sites evidencing indel frequencies below 1% (Fig. 3B), and no differences observed between off-target sites predicted by CRISPRScan or CIRCLE-seq (Mann-Whitney U = 175.5, \( \ p = 0.873 \); Fig. 3B). To validate the accuracy of ICE at these low indel frequencies, we again performed Illumina sequencing of predicted off-target sites for six of the eight evaluated gRNAs (see Supplementary Table 1 for description) and found significant concordance in results (0.29–7.62% of mosaicism; Spearman \( \rho = 0.83, \ p = 0.039 \); Fig. 3C). The average difference in mosaicism between ICE and Illumina was low (1.6±2.0), with ICE tending to slightly underestimate indel frequencies, highlighting its utility to quickly and economically assess predicted off-targets regions.

We also tested if sites predicted with higher likelihoods of off-target cutting events resulted in higher mutation rates by comparing the indel frequencies among the different levels of prediction (top 1, 2, or 3 prediction scores by CRISPRScan or CIRCLE-seq). No differences were found between prediction groups (Kruskal-Wallis: \( H_{(2)} = 2.26, \ p = 0.320 \); Fig. 3B), suggesting that the information used by the tools to assign probabilities of off-target activity (e.g., CFD scores in CRISPRScan or normalized read counts in CIRCLE-seq) do not necessarily predict the efficiency of cutting at off-target sites in vivo. Thus, off-target cutting mutations at the assessed
sites exhibited low frequencies with no clear method performing best. Moreover, none of the on-target scores previously obtained (in silico, in vitro, or in vivo) correlated with the number of predicted off-target sites per gRNA (using either CRISPRScan or CIRCLE-seq), nor the frequency of indels at validated off-target sites (Spearman \( \rho = 0.27, p = 0.111 \), Fig. 3D), suggesting that higher on-target efficiencies do not necessarily translate into increased frequencies of spurious off-target mutations.

**Evaluating CRISPR Cas9-injection controls**

A commonly used ‘mock’ injection control for phenotypic screens of CRISPR-generated G_0 mosaic lines are embryos injected with buffer and Cas9 in the absence of a gRNA. We sought to determine if such control treatments could significantly impact the genome or transcriptome of our zebrafish larvae. To characterize its impact on genes, we performed RNA-seq of wild-type NHGRI-1 embryos injected with either Cas9 enzyme or Cas9 mRNA (three biological replicates of a pool of five injected larvae), uninjected batch siblings (two biological replicates of a pool of five larvae), and uninjected siblings from another batch (three biological replicates with a pool of five larvae) as controls.

**Potential genomic mutations in controls**

Recently, Sundaresan and colleagues [49] found that Cas9 in the presence of Mn^{2+} ions can result in double-strand cleavage of genomic DNA in the absence of a gRNA. Although their study did not show this same off-target cleavage activity in the presence of Mg^{2+}, we hypothesized that aberrant genomic mutations could be incurred by Cas9 due to the presence of MgCl_2 in our injection buffer since Mg^{2+} has been shown to compete with Mn^{2+} in activating common enzymes [50]. Using our RNA-seq data, we used an optimized pipeline [51] to identify somatic mosaic mutations with uninjected wild-type controls as a reference for common polymorphisms. Focusing only on high-confidence variants (minimum sequence read depth of 20), we filtered already-reported variants in the NHGRI-1 zebrafish line [46], and used the Variant Effect Predictor tool from ENSEMBL to obtain a list of frameshift mutations in protein-coding genes present in our
Cas9-injected larvae. A total of 48 and 38 genes were identified with frameshifting variants in larvae injected with Cas9-enzyme and Cas9-mRNA, respectively, with 14 of these genes shared across both injection types (Fig. 4A, Supplementary Table 4). On average, each pool of larvae injected with Cas9 enzyme or mRNA carried frameshift variants in 18.7 ± 3.1 genes. All identified frameshift variants evidenced low allelic frequencies (Cas9-enzyme: average 0.043, range 0.0036-0.142; Cas9-mRNA: average 0.059, range 0.002-0.316) and high read depth (Cas9-enzyme: average 386.5, range 22-2076; Cas9-mRNA: average 343.8, range 20-3453) (Supplementary Table 4). Additionally, frameshift variants were positioned closer to a potential Cas9 PAM site (NGG) than by random chance (4 bp median observed distance to closest PAM site; empirical p = 0.0016 using the whole-genome and p = 0.006 using protein-coding regions only, from 10,000 permutations). Therefore, we

![Diagram A]

**Fig. 4** Evaluation of spurious genomic mutations in CRISPR-injection controls. **A** The abundance of protein-coding genes carrying frameshift variants for each Cas9-injected treatment are depicted in a Venn diagram, with mutated genes identified in both treatments listed. **B** Genomic DNA from zebrafish larvae injected with Cas9 enzyme, Cas9 mRNA, catalytically dead Cas9 (dCas9), a scrambled gRNA, uninjected batch siblings, and a fin clip from their parents was used to perform targeted Illumina sequencing of 21 genes to quantify indel mosaicism with average ± standard deviation values listed in the table (see Supplementary Tables 1 and 7 for the description of the genes).
decided to evaluate if indels would consistently arise in these genes in an additional set of microinjections. We performed a new set of microinjections in NHGRI-1 larvae using these same controls (Cas9 enzyme and Cas9 mRNA) and two additional ones commonly used in CRISPR experiments (catalytically dead Cas9 (dCas9) enzyme and a scrambled gRNA coupled with Cas9 enzyme, sequence published in [19]) and evaluated the presence of mutations in 21 genes, including 14 genes with identified frameshift mutations in our RNA-seq data and seven controls with no mutations observed (Fig. 4B, see Supplementary Tables 1 and 5 for the description of all sites). Briefly, genomic DNA was harvested from (1) three pools of five larvae from each group injected at the one-cell stage (Cas9 enzyme, Cas9 mRNA, dCas9, scrambled gRNA); (2) three pools of five uninjected batch siblings larvae; and (3) finclips of the crossing parents as controls. Subsequently, ~200 bp regions surrounding the closest Cas9 PAM site to the previously RNA-seq-identified variants were Illumina sequenced and the alleles extracted using CrispRVariants [33]. We did not observe evidence of inflation of indels in any of the injected groups relative to the uninjected batch siblings or the parental fish, with an overall average mosaicism of 3.1±0.8% per site (below the expected 10% allele ratio for a heterozygous variant in a single individual from a pool of five; Fig. 4B, Supplementary Table 5). Our NHGRI-1 zebrafish carried common single nucleotide variants in the targeted regions, particularly in gene stich1073-110a20 where two variants were present in close to 50 and 20% of the reads (Supplementary Fig. 2). Interestingly, we did observe a subtly higher mosaicism in the genes previously detected with variants in our RNA-seq data relative to the regions used as controls (Mann-Whitney U = 2251.5, p = 0.00074, median mosaicism in tested genes 3.4%, median mosaicism in control genes 2.88%; Fig. 4B, Supplementary Table 5). Thus, it is possible that the genes we identified with variants in our RNA-seq data may be naturally prone to carry variants. In summary, these results suggest that currently used CRISPR controls do not suffer systematic DNA cleavages in the absence of a gRNA.

**Differential gene expression in controls**

We also characterized the impact of injecting Cas9 enzyme or mRNA on the transcriptomes of our zebrafish larvae. Comparisons of transcripts abundances show significant variance across biological replicates when quantifying in both Cas9 treatments, particularly evident in samples injected with the Cas9 enzyme, versus wild-type uninjected larvae (Fig. 5A). This suggests that considerable stochasticity may exist regarding the effects of Cas9 injections in these controls. Examining the genes impacted, we identified hundreds of differentially-expressed (DE) genes in our Cas9-injected versus uninjected controls, with a greater number of upregulated genes than downregulated genes (Fig. 5B, Supplementary Table 6). Specifically, Cas9-enzyme injections resulted in a total of 1100 DE genes (3.6% of the genes assayed), with 756 genes (68.7%) upregulated (fold change > 1) and 344 (31.3%) downregulated (fold change < −1). Cas9-mRNA injected larvae exhibited 548 DE genes (1.8% of the genes assayed), 376 (68.6%) of these upregulated and 172 (31.4%) downregulated (Fig. 5B). We observed 248 (197 upregulated and 51 downregulated) common DE genes between the two treatments (Fig. 5C), which could be part of a common response to the microinjection process. Network analyses identified commonalities in the shared DE genes enriched in key regulators of different KEGG pathways [52], including spliceosome and ribosome (including genes eif4g1a, hnrnpd, magoh, hnrnpa0a), hedgehog signaling (shha), glutathione metabolism (gsto2, gsr), GnRH signaling (disp6), aminoacyl-tRNA biosynthesis (yars), cell cycle (ki67), glycolysis (aloda), and cellular senescence (ppp3ccca) (Supplementary Fig. 3). Furthermore, while we observed no enrichment in gene ontology terms for downregulated genes, common upregulated genes from both treatments were related to response to wounding (GO:0009611, adjusted p-value = 0.009) and cytoskeleton organization (GO:0045104, adjusted p-value = 0.009) (Supplementary Table 7), revealing molecular consequences of the microinjection process that were still detectable five days later.

**Discussion**

Our study presents a comprehensive evaluation of empirical and predictive tools currently used for CRISPR editing in zebrafish. Cleavage scores obtained by an in vivo assessment of 50 gRNAs via Sanger sequencing and deconvolution tools (ICE and TIDE) were concordant with Illumina sequencing, the gold standard in predicting efficiencies, as previously reported [35]. Both tools underestimated the presence of non-edited alleles by ~20%, contrary to previous comparisons of TIDE and Illumina sequencing in cell lines, where TIDE showed a ~10–20% overestimation of non-edited alleles [53]. For sites with lower indel frequencies, as we observed for predicted off-target mutations, ICE scores were more concordant with Illumina results (~1–2% difference, again mostly underestimates). Therefore, we suggest that Sanger sequencing deconvolution tools are valuable for establishing relative gRNAs efficiencies but do not necessarily accurately predict absolute cleavage efficiencies in zebrafish in vivo, except at sites with low indel frequencies. In addition, we formalized an empirical ‘intensity ratio’ score from the commonly-used PAGE approach
to assay CRISPR indels and verified its utility in approximating cleavage efficiencies, making it a more affordable and rapid approach to assay editing efficiencies versus sequencing.

On-target efficiency prediction tools showed large differences using the same set of gRNAs sequences, highlighting the importance of understanding features accounted for by each tool. A recent review [25] provides a comprehensive overview of different design tools available and the source of experimental data used to train each one. CRISPRScan [26] was the only tool that could predict on-target efficiency in our set of gRNAs, while no other method provided scores that were correlated with cleavage activities observed in vivo. One limitation of our study was the skew in higher efficiency gRNAs (mean predicted CRISPRScan score of 57.6), which could feasibly impact correlations. Notably, we did obtain more accurate CRISPRScan predictions when we utilized our NHGRIzed reference [46] compared to the current Tübingen-derived reference [4], highlighting the importance of accounting for known genetic variation when designing suitable gRNAs [54, 55]. Considering CRISPRScan was the only tool that incorporated empirical data from zebrafish, with most methods tested using in vitro-derived data, our results emphasize the importance of utilizing a tool trained using in vivo experimental data specific to the study’s target species.

An in silico (CRISPRScan) and in vitro (CIRCLE-seq) method predicted ~ 20 and 65% potential off-target regions impacting genes, respectively. Notably, we did not evaluate if other predicted sites included cis-regulatory elements that could also potentially alter gene expression. Future assessments should include tests targeting a diversity of loci for a more thorough understanding of the potential off-target indels caused by unwanted CRISPR cleavage sites. We observed low off-target mutation frequencies (most < 1%), similar to those previously reported from using single [11, 12] or multiple gRNAs [18], although did observe off-target indel frequencies as high as 11% for certain gRNAs. Notably, neither predictive method (CRISPRScan or CIRCLE-seq) nor their likelihood score (using CFD or
normalized read count) could accurately predict indel frequencies at off-target sites. Typically, such low mutation frequencies should not be of high impact when generating stable knockout zebrafish lines as these could be easily outcrossed. However, such mutations could have significant impacts on phenotypic outcomes when injected G0 mosaic populations are analyzed directly.

The adequate selection of controls is a fundamental process in evaluating gene function using G0 knockout crisptant zebrafish, as these larvae serve as baselines from which inferences will be made from. Currently, no consensus exists for preferred controls used in high-throughput CRISPR workflows of zebrafish larvae, which can include targeting a known gene as a positive control (e.g., tyr) [14], uninjected larvae [17, 18], sham injections with a Cas9:tracrRNA complex [15], and injections of a scrambled gRNA [16, 19], among others. Our RNA-seq assay identified several genes carrying frameshift mutations using uninjected clutch siblings as reference. A follow-up analysis of a second set of controls showed existence of mosaic variants in all injected controls (e.g., Cas9 mRNA, enzyme, and scrambled gRNA), in addition to uninjected siblings and crossed parents at low allelic frequencies (~3%). Nevertheless, even though we were limited to our targeted regions, we did observe a higher mosaicism in genes identified as carrying frameshift mutations from our RNA-seq assay compared to control genes, suggesting that these genes could be naturally prone to exhibit mutations in the NHGRI-1 zebrafish line. We also observed high variability in gene expression in larvae solely injected with Cas9 enzyme or mRNA, with several of these DE genes involved in response to wounding processes. Notably, these DE genes were retrieved from 5 dpf larvae suggesting that damage incurred during the microinjection process has a lasting effect. These results suggest that caution should be taken in using G0 mosaic mutants in investigating phenotypes related to pathways found to be significantly skewed in injection controls, including those involving the function of the spliceosome, ribosomes, and cytoskeleton dynamics.

Conclusions
Overall, we performed a simultaneous assessment of gRNA activities predicted by several commonly used in silico and in vitro methods with those determined experimentally in vivo in injected zebrafish embryos. These results provide valuable information that can be incorporated into the design and execution of CRISPR/Cas9 assays in zebrafish using available workflows [8, 13, 14, 17, 18]. Namely, we make the following conclusions and recommendations:

- Sanger-based efficiency estimates (TIDE and ICE) tend to underestimate indel mosaicism in zebrafish, though they are more accurate when lower mutational mosaicism exists (such as those observed at off-target sites).
- Quantifying heterodimers via PAGE gels represents an affordable method to qualitatively assay CRISPR cutting efficiencies.
- Of the existing tools, we recommend CRISPRScan for predicting gRNA on-target efficiency, preferably matched to the zebrafish strain being used.
- Off-target mutations occur at relatively low rates with neither in silico nor in vitro prediction methods performing significantly better.
- Microinjection of Cas9 (enzymes or mRNA) into embryos does not result in spurious genomic mutations but does impact expression of certain genes and pathways. Caution should be exercised if studying phenotypes related to these genes when performing G0 mosaic zebrafish screens.

Our aim was to provide information to aid in the decision-making process for future projects using affordable and reliable gene-editing tools in zebrafish. As high-throughput methods continue to be developed for assaying multiple genes simultaneously, it will be important to use optimal tools for predicting and assessing on- and off-target activity in zebrafish larvae for accurate interpretation of phenotypic outcomes.

Methods
Zebrafish husbandry
NHGRI-1 wild-type zebrafish [46] were maintained through standard protocols [56] and their use was approved by the Institutional Animal Care and Use Committee (IACUC) from the Office of Animal Welfare Assurance, University of California, Davis. Animals were kept in a temperature (28 ± 0.5°C) and light (10 h dark/14 h light cycle) controlled modular system with UV-sterilized filtered water (Aquaneering, San Diego, CA), with a density of 25 adult fish per tank. Feeding and general monitoring of all zebrafish was performed twice a day (9 am and 4 pm). Food included rotifers (Rotigrow Nanno, Reed Mariculture, Campbell, CA), brine shrimp (Artemia Brine Shrimp 90% hatch, Aquaneering, San Diego, CA), and flakes (Zebrafish Select Diet, Aquaneering, San Diego, CA). For all experimental procedures, eggs were collected via natural spawning of randomly selected adult NHGRI-1 zebrafish in 11 crossing tanks (Aquaneering, San Diego, CA), using a minimum of five breeding pairs (1 male, 1 female) unless otherwise specified. Embryos were grown in standard Petri dishes with E3 media (0.03% Instant Ocean salt in deionized water)
and incubated at 28 ± 0.5 °C, using a dissecting microscope (Leica, Buffalo Grove, IL) for developmental staging and daily monitoring until their use for molecular procedures.

**Design and in silico predictions for gRNAs**

Fifty gRNAs targeting exons of 14 genes were designed using CRISPRScan [26] (scores ranging between 24 and 83 with a mean value of 57.6) with zebrafish genome version GRCz11/danRer11 as the reference (see description of gRNAs in Supplementary Tables 1 and 2). All targeted genes were protein coding. For each designed gRNA, we obtained the efficiency scores predicted by CRISPRScan [26], CHOPCHOP [36–38] using the scoring method from [39] and [40], E-CRISP [41], CRISPR-GE [42], CCTop [43], CRISPRon [44], DeepSpCas9 [45], and the IDT design tool (www.idtdna.com). From CRISPRScan, we also gathered the predicted off-target sites for each gRNA defined by the CFD score [26]. Additionally, we utilized bedtools [57] to determine the GC percentage for each gRNA. To incorporate NHGRI-1 variants into the zebrafish reference, we used the FastaAlternateReferenceMaker function from GATK [58] with the reported high-confidence variants for the NHGRI-1 zebrafish strain [46].

**Microinjections to generate CRISPR G₀ mosaic mutants**

All gRNAs were individually injected into NHGRI-1 embryos to estimate the frequency of indels. gRNAs were prepared following the manufacturer’s protocol (IDT). Briefly, 2.5 μl of 100 μM crRNA, 2.5 μl of 100 μM tracrRNA, and 5 μl of Nuclease-free Duplex Buffer using an annealing program consisting of 5 min at 95 °C, a ramp from 95 °C to 50 °C with a −0.1 °C/s change, 10 min (min) at 50 °C, and a ramp from 50 °C to 4 °C with a −1 °C/s change. Ribonucleaseprotein injection mix was prepared with 1.30 μl of Cas9 enzyme (20 μM, New England Biolabs), 1.60 μl of prepared gRNAs, 2.5 μl of 4x Injection Buffer (containing 0.2% phenol red, 800 mM KCl, 4 mM MgCl₂, 4 mM TCEP, 120 mM HEPES, pH 7.0), and 4.6 μl of Nuclease-free water. Microinjections directly into the yolks of NHGRI-1 embryos at the one-cell stage were performed as described previously [59], using needles from a micropipette puller (Model P-97, Sutter Instruments) and an air injector (Pneumatic MPPI-2 Pressure Injector). Embryos were collected and ~1 nl of ribonucleoprotein mix was injected per embryo, after previous calibration with a microruler. Twenty injected embryos per Petri dish were grown up to 5 dpf at 28 °C.

**Illumina and Sanger amplicon sequencing**

DNA extractions were performed on 20 pooled embryos by adding 100 μl of 50 mM NaOH, incubation at 95 °C for 20 min, ramp from 95 °C to 4 °C at a 0.7 °C/s decrease, followed by an addition of 10 μl of 1 M Tris-HCl and a 15 min spin at 4680 rpm. We amplified a ~200 bp region surrounding the targeted site of each gRNA (see Supplementary Table 1 for description of primers). PCR amplifications were performed using 12.5 μl of 2X DreamTaq Green PCR Master Mix (Thermo Fisher), 9.5 μl of Nuclease-Free water, 1 μl of 10 μM primers, and 1 μl extracted DNA. Thermocycler program included 3 min at 95 °C, followed by 35 cycles of 15 s at 95 °C, 30 s at 60 °C, and 20 s at 72 °C, and a final 5 min incubation at 72 °C. Reactions were purified using Ampure XP magnetic beads (Beckman Coulter) and Illumina sequenced (Genewiz, San Diego, CA). To obtain percent mosaicity of mutants by mapping paired-end fastq reads to the zebrafish reference genome (GRCz11/danRer11) using bwa [60] and the R package CrispRVariants [33]. Additionally, we amplified a ~500 bp region surrounding the targeted site of each gRNA from the same extracted DNA for six gRNAs and performed and performed Sanger sequencing (Genewiz, San Diego, CA). Raw trace files were used in the TIDE [34] and ICE [35] tools to predict the percentage of indels, which we used as our in vivo editing score for each gRNA. For both Sanger and Illumina sequencing, we used uninjected batch-sibling embryos as a control reference.

**PAGE and intensity-ratio estimation**

An empirical cleavage analysis from each gRNA was performed using PAGE. Briefly, we amplified a ~200 bp region in DNA around the targeted site from gRNA-injected and uninjected embryos, as described above. Reactions of the uninjected and injected samples from the same amplicon were run on a 7.5% polyacrylamide gel together for 75 min at 110 V and revealed using GelRed (VWR International). Gel images were processed in the software Fiji [61]. For each sample, we defined areas A and B as follows:

For each gRNA, the mean-intensity value was obtained for the A and B areas in both the injected and uninjected samples. The A and B areas were exactly the same size...
between samples. The intensity ratio was calculated as: [injected B / injected A] / [uninjected B / un.injected A]. Log-normalized intensity ratios followed a normal distribution (Shapiro-Wilk test: W = 0.96, p = 0.167) with an average value of 1.21 ± 0.70.

**CIRCLE-seq**

CIRCLE-seq libraries were prepared for each gRNA (IDT) using genomic DNA extracted from NHGRI-1 (DNA Blood & Tissue kit, Qiagen) following the described protocol [47]. Libraries were sequenced using one HiSeq XTen lane (Novogene, Sacramento, CA), providing an average of 7.3 million reads (range: 4.0 - 13.3 million reads) and >Q30 for 92% of reads per gRNA library. Raw reads were processed using the bioinformatic pipeline described [47] (mapping rate >99% in all samples) to identify regions with cutting events relative to a control sample (treated with Cas9 enzyme and no gRNA). In an attempt to obtain an on-target efficiency estimation from in vitro digestions, we calculated the reads per million normalized (RP MN). For this purpose, we used samtools [61] to extract read coverage from aligned bam files. For each gRNA, coverage was obtained for the third and fourth base upstream of the PAM site as it is the region expected to be cut by Cas9 [62]. RPMN for each gRNA was calculated as the sum of coverage at these two sites divided by the total mapped reads per sample and multiplied by 1 million to scale the values. RPMN scores ranged from 4.42 to 881 (median 99.3) so we decided to use a log normalization to reduce this range.

**RNA-seq**

We performed RNA-seq of Cas9 injected NHGRI-1 larvae. One-cell stage NHGRI-1 embryos were injected with either Cas9 enzyme or Cas9 mRNA. Injection mix for Cas9 enzyme included Cas9 enzyme (20 μM, New England Biolabs), 2.5 μl of 4x Injection Buffer (0.2% phenol red, 800 mM KCl, 4 mM MgCl₂, 4 mM TCEP, 120 mM HEPES, pH 7.0), and Nuclease-free water. Cas9 mRNA was obtained from plasmid pT3TS-nCas9n (Addgene, plasmid #46757) [5], using the MEGAlgorithms T3 transcription kit (Thermo Fisher) following manufacturer’s guidelines of 3.5h 56°C incubation with T3. mRNA was purified with the MEGAClear transcription clean-up kit (Thermo Fisher) and concentration of mRNA obtained using a NanoDrop (Thermo Fisher). The injection mix of Cas9 mRNA contained 100ng/μl of mRNA, 4x Injection Buffer (0.2% phenol red, 800 mM KCl, 4 mM MgCl₂, 4 mM TCEP, 120 mM HEPES, pH 7.0), and Nuclease-free water. Additionally, uninjected batch-siblings and uninjected siblings from an additional batch were used as controls. All embryos were grown at 28 °C in a density of <50 embryos per dish. At 5 dpf, three pools of five larvae were collected for each group (Cas9 enzyme, Cas9 mRNA, and uninjected) for RNA extraction using the RNaseasy kit (Qiagen) with genomic DNA eliminator columns for DNA removal. Whole RNA samples were subjected to RNA-seq using the poly-A selection method (Genewiz, San Diego, CA).

**Variant identification from RNA-seq data**

We followed a previously described pipeline to identify somatic variants from RNA-seq data [51]. Briefly, we mapped reads with STAR [63] using the 2-pass mode and a genomic reference created with GRCh31/danRer11 assembly and gtf files (release version 100). Variant calling was performed with MuTect2 as part of GATK [58] using the tumor versus normal mode. ‘Normal’ was defined by the two uninjected samples to identify all somatic mutations in our Cas9 injected embryos. Variants were annotated using the Variant Effect Predictor tool [64]. High confidence variants (minimum sequencing depth of 20) previously reported for the NHGRI-1 line [46] were removed. Only frameshift loss-of-function variants with a minimum read depth of 20 in canonical protein-coding genes were considered. We extracted the median distance between the identified variants and the nearest Cas9 PAM site (NGG sequence) using the coordinates in the CRISPRScan UCSC track. This median observed distance was compared to the result of median distances of 10,000 permutations of random sampling across the genome and their nearest PAM site. One-tailed empirical p values from this comparison were calculated as (M + N)/(N + 1), where M is the number of iterations with a median distance below the observed value and N is the total number of iterations. We orthogonally investigated the presence of variants in 23 genes via Illumina sequencing of a ~200bp region surrounding the identified variant location and the R package CrispRVariants [33] (Supplementary Table 1 for primers description). For this purpose, we extracted DNA from three pools of five embryos injected with Cas9 enzyme, Cas9 mRNA, dCas9 (Alt-R S. p. dCas9 protein V3 from IDT), a scrambled gRNA (see Supplementary Table 1 for sequence description), or uninjected. In addition, we extracted DNA from a finclip of the crossing parents of the embryos used for the injections (both female and male). In all of these groups, we quantified the percentage of mutations as all alleles different from the reference.

**Differential gene expression analysis from RNA-seq data**

Raw reads were processed using the elvers [https://github.com/dib-lab/elvers; version 0.1], release DOI: https://doi.org/10.5281/zenodo.3345045 bioinformatic pipeline that utilizes fastqc [65], trimmomatic [66], and salmon [67]
to obtain the transcripts per kilobase million (TPM) for each gene. DESeq2 [68] was used to extract differentially-expressed genes in the Cas9 enzyme or Cas9 mRNA injected samples relative to the uninjected larvae. R package clusterProfiler [69] was used to perform enrichment tests of differentially-expressed genes in biological pathways. Network analyses of the common differential expressed genes was performed using the NetworkAnalyst online tool (www.networkanalyst.ca) [70, 71].

**Statistical analyses**

All analyses were performed in R version 4.0.2 [72]. Normality of variables was checked using the Shapiro-Wilk test and parametric or nonparametric comparisons made accordingly. Spearman correlation tests (denoted as ρ) and linear regression models were used to determine the relationship between variables. All analyses compared across different experimental batches included batch as a factor in the model to prevent biases caused by inter-batch differences. Averages include the standard deviation unless otherwise specified. Alpha to determine significance across the different tests was set at 0.05 unless otherwise specified. Additional R packages used for making figures included eulerr [73] and pheatmap [74].

**Abbreviations**

CFD: Cutting frequency determination; CRISPR: Clustered regularly interspaced short palindromic repeats; gRNA: Guide RNA; indels: Insertions or deletions; PAGE: Polyacrylamide gel electrophoresis; RPMN: Reads per million normalized.

**Supplementary Information**

The online version contains supplementary material available at https://doi.org/10.1186/s12864-021-08238-1.

**Additional files**

**Additional file 1:** Supplementary Figure 1. Editing efficiencies across 50 gRNAs with different GC content. **Supplementary Figure 2.** Mosaicism predicted using Illumina sequencing of C9_ mutant larvae. **Supplementary Figure 3.** Network analyses of the shared differentially expressed genes in Cas9 enzyme and Cas9 mRNA injected larvae. **Supplementary Table 1.** Oligonucleotides utilized for all assays performed throughout the project. **Supplementary Table 2.** On-target efficiency scores for the 50 gRNAs obtained from different in silico prediction tools, empirical cleavage scores in vivo using Sanger sequencing and sequence deconvolution tools (TIDE and ICE), from the in vitro assay CIRCLE-seq (RPMN), and Illumina sequencing using the CrispRVariants software. **Supplementary Table 3.** Description of the off-targets predicted by CRISPRScan and CIRCLE-seq for the 50 evaluated gRNAs. The percentage of sites intersecting genes from the CRISPRScan tool was obtained from the top 30 predicted off-targets since the tool only provides detailed information from these. **Supplementary Table 4.** List of the frameshift variants identified in protein-coding genes in samples injected with either Cas9 enzyme or Cas9 mRNA using un.injected batch-siblings as reference. **Supplementary Table 5.** Illumina mosaicism percentages observed in larvae injected with Cas9 (enzyme or mRNA), catalytically dead Cas9 (dCas9), scrambled gRNA, uninjected batch siblings, and a finclip from their crossing parents. **Supplementary Table 6.** List of DE genes observed in Cas9-injected samples relative to uninjected batch-siblings and siblings from another batch. **Supplementary Table 7.** List of gene ontology terms enriched in upregulated genes found in both injection treatments (Cas9 enzyme and Cas9 mRNA).

**Acknowledgements**

We thank our team of UC Davis undergraduate students that maintain husbandry to keep our zebrafish healthy and happy. Thank you to Dr. Li-En Jiao for kindly providing the Cas9 mRNA plasmid and comments on the manuscript, and Daniela C. Soto for bioinformatic support. We are grateful to Dr. Charles Vejnar and Dr. Antonio Giráldez for the incorporation of our NHGrized zebrafish reference to the CRISPRScan online tool. Finally, we thank the anonymous reviewers of this manuscript for their helpful comments and suggestions that have significantly improved our study.

**Authors’ contributions**

JMUUS: conceptualization, data collection, analysis, writing of the article. AS, GK, KW, and CI: data collection. MYD: conceptualization, funding acquisition, analysis, writing of the article. All authors have read and approved the manuscript.

**Funding**

This work was supported, in part, by the U.S. National Institutes of Health (NIH) grants from the National Institute of Neurological Disorder and Stroke (R00NS083627 to M.Y.D.), the Office of the Director and National Institute of Mental Health (DP2 OD023824 to M.Y.D.), UC Davis MIND Institute Intellectual and Developmental Disabilities Research Center pilot grant (U54 HD079125 to M.Y.D.), and a UC Davis Graduate Research Award (J.M.U.S.). M.Y.D. is also supported by a Sloan Research Fellowship (FG-2016-6814). The funding sources did not play a role in the research or publication process.

**Availability of data and materials**

Original fastq files from the CIRCLE-seq and RNA-seq assays are deposited in the European Nucleotide Archive repository under project PRJEB39643.

**Declarations**

**Ethics approval and consent to participate**

Animal use was approved by the University of California, Davis IACUC from the Office of Animal Welfare Assurance accredited by the Association for Assessment and Accreditation of Laboratory Animal Care (Assurance Number A3433-01 on file with the Office of Laboratory Animal Welfare). The IACUC is constituted in accordance with U.S. Public Health Service Animal Welfare Policy and includes a member of the public and a non-scientist. The study and all methods were carried out in accordance with relevant guidelines and regulations and in compliance with the ARRIVE guidelines [75].

**Consent for publication**

Not applicable.

**Competing interests**

The authors declare that they have no competing interests.

**Author details**

1 Genome Center, MIND Institute, and Department of Biochemistry & Molecular Medicine, School of Medicine, University of California, Davis, Davis, CA, USA.
2 Integrative Genetics and Genomics Graduate Group, University of California, Davis, Davis, CA, USA.

**Received: 17 May 2021   Accepted: 4 December 2021**

**Published online:** 06 January 2022

**References**

1. Meyers JR. Zebrafish: Development of a Vertebrate Model Organism: Zebrafish: Development of a Vertebrate Model Organism. Current Protocols Essential Laboratory Techniques. 2018;16.e19.
1. Holtzman NG, Lovene MK, Liang JO, Morris J. Learning to fish with genetics: a primer on the vertebrate model Danio rerio. Genetics. 2016;203:1069–89.

4. Howe K, Clark MD, Torroja CF, Torrence J, Berthollet C, Muffato M, et al. The zebrafish reference genome sequence and its relationship to the human genome. Nature. 2013;496:498–503.

7. Jiao L-E, Wente SR, Chen W. Efficient multiplex biallelic zebrafish genome editing using a CRISPR/Cas9 nuclease system. Proc Natl Acad Sci U S A. 2013;110:1904–9.

10. Hwang WY, Fu Y, Reyon D, Maeder ML, Tsai SQ, Sander JD, et al. Efficient genome editing in zebrafish using a CRISPR-Cas9 system. Nat Biotechnol. 2013;31:227–9.

13. Iton U, Krauss J, Nüsslein-Volhard C. Precise and efficient genome editing in zebrafish using the CRISPR/Cas9 system. Development. 2014;141:4827–30.

16. Holtzman NG, Iovine MK, Liang JO, Morris J. Learning to fish with RNA-guided Cas9 nuclease in zebrafish embryos. Cell Res. 2013;23:465–72.

19. Irion U, Krauss J, Nüsslein-Volhard C. Precise and efficient genome editing in zebrafish using the CRISPR/Cas9 system. Nat Biotechnol. 2013;31:227–9.
49. Sundaresan R, Parameshwaran HP, Yogesha SD, Keilbarth MW, Rajan R. RNA-independent DNA cleavage activities of Cas9 and Cas12a. Cell Rep. 2017;21:3728–39.

50. Winhust JM, Manchester KL. Comparison of ability of Mg and Mn to activate the key enzymes of glycolysis. FEBS Lett. 1972;72:321–6.

51. Coudray A, Battenhouse AM, Bucher P, Iyer VR. Detection and benchmarking of somatic mutations in cancer genomes using RNA-seq data. PeerJ. 2018;6:e5362.

52. Kaneshia M, Goto S. KEGG: kyoto encyclopedia of genes and genomes. Nucleic Acids Res. 2000;28:27–30.

53. Sentmanat MF, Peters ST, Florian CP, Connelly JP, Pruett-Miller SM. A survey of validation strategies for CRISPR-Cas9 editing. Sci Rep. 2018;8:8888.

54. Suurväli J, Whiteley AR, Zheng Y, Gharbi K, Leptin M, Wiehe T. The laboratory domestication of zebrafish: from diverse populations to inbred substrains. Mol Biol Evol. 2020;37:1056–69.

55. Coe TS, Hamilton PB, Griffiths AM, Hodgson DJ, Wahab MA, Tyler CR. Genetic variation in strains of zebrafish (Danio rerio) and the implications for ecotoxicology studies. Ecotoxicology. 2009;18:44–50.

56. Westerfield M. The zebrafish book: a guide for the laboratory use of zebrafish (Danio Renio); 2007.

57. Quinlan AR. BEDTools: the Swiss-army tool for genome feature analysis. Curr Protoc Bioinformatics. 2014;47:11.12.1–34.

58. Wu X, Kriz AJ, Sharp PA. Target specificity of the CRISPR-Cas9 system. Quant Biol. 2014;2:59–70. https://doi.org/10.1007/s40484-014-0030-x.

59. Yu G, Wang L-G, Han Y, He Q-Y. clusterProfiler: an R package for comparing biological themes among gene clusters. OMICS. 2012;16:284–7.

60. Zhou G, Soufan O, Ewald J, Hancock REW, Basu N, Xia J. NetworkAnalyst 3.0: a visual analytics platform for comprehensive gene expression profiling and meta-analysis. Nucleic Acids Res. 2019;47:W234–41.

61. Team RC. R: a language and environment for statistical computing. 2017.

Publisher’s Note
Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.