



Analysis of Single Nucleotide Variants in CRISPR-Cas9 Edited Zebrafish Exomes Shows No Evidence of Off-Target Inflation

Marie R. Mooney¹, Erica E. Davis^{1,2,3} and Nicholas Katsanis^{1,2,3*}

¹ Center for Human Disease Modeling, Duke University Medical Center, Durham, NC, United States, ² Advanced Center for Translational and Genetic Medicine (ACT-GeM), Stanley Manne Children's Research Institute, Department of Pediatrics, Ann & Robert H. Lurie Children's Hospital of Chicago, Chicago, IL, United States, ³ Departments of Pediatrics and Cell and Developmental Biology, Feinberg School of Medicine, Northwestern University, Chicago, IL, United States

OPEN ACCESS

Edited by:

Emmanouil Dermizakis,
University of Geneva, Switzerland

Reviewed by:

Rui Chen,
Baylor College of Medicine,
United States
Michael R. Crowley,
University of Alabama at
Birmingham, United States

*Correspondence:

Nicholas Katsanis
nkatsanis@luriechildrens.org

Specialty section:

This article was submitted to
Genomic Assay Technology,
a section of the journal
Frontiers in Genetics

Received: 12 May 2019

Accepted: 05 September 2019

Published: 11 October 2019

Citation:

Mooney MR, Davis EE and
Katsanis N (2019) Analysis of Single
Nucleotide Variants in CRISPR-Cas9
Edited Zebrafish Exomes Shows No
Evidence of Off-Target Inflation.
Front. Genet. 10:949.
doi: 10.3389/fgene.2019.00949

Therapeutic applications of CRISPR-Cas9 gene editing have spurred innovation in Cas9 enzyme engineering and single guide RNA (sgRNA) design algorithms to minimize potential off-target events. While recent work in rodents outlines favorable conditions for specific editing and uses a trio design (mother, father, offspring) to control for the contribution of natural genome variation, the potential for CRISPR-Cas9 to induce *de novo* mutations *in vivo* remains a topic of interest. In zebrafish, we performed whole exome sequencing (WES) on two generations of offspring derived from the same founding pair: 54 exomes from control and CRISPR-Cas9 edited embryos in the first generation (F0), and 16 exomes from the progeny of inbred F0 pairs in the second generation (F1). We did not observe an increase in the number of transmissible variants in edited individuals in F1, nor in F0 edited mosaic individuals, arguing that *in vivo* editing does not precipitate an inflation of deleterious point mutations.

Keywords: CRISPR-Cas9, zebrafish, exome, *de novo* mutation, off-target effect

INTRODUCTION

CRISPR-Cas9 gene editing technology has offered powerful investigative tools and opened new potential avenues for the treatment of genetic disorders. Nonetheless, like preceding technologies, the *in vivo* implementation of CRISPR-Cas9 editing faces potential barriers. These include restricted control over the delivery and activity of the system; immune responses to the system components; and permanent alteration of unintended genomic targets (Ho et al., 2018). In cell culture systems, the alteration of off-target regions decreases precipitously with the use of stringently designed sgRNA sequences and Cas9 enzymes engineered for high specificity (Fu et al., 2013; Doench et al., 2016; Hu et al., 2018), though recent work demonstrates that precise control over the nature of editing even at on-target sites remains challenging (Kosicki et al., 2018). In rodents, these same factors influence the efficiency and specificity of CRISPR-Cas9 editing (Anderson et al., 2018). However, examination of atypical CRISPR-Cas9 influence on organisms remains limited; it is often focused primarily on predicted off-target assessment and is not always agnostic (Varshney et al., 2015).

Here, we evaluated the incidence and transmission of off-target effects in a cohort of CRISPR-Cas9 edited zebrafish embryos derived from the same founding pair. Using 52 zebrafish embryos from the same clutch targeted with sgRNAs with variable on-target efficiency, we whole-exome sequenced

DNA from the entire cohort and their genetic parents and we measured the transmission of variants to the next generation.

METHODS

CRISPR-Cas9 Gene Editing in Zebrafish Embryos

We used CHOPCHOP (Labun et al., 2016) to identify sgRNAs targeting a sequence within the coding regions of the target genes and sgRNAs were *in vitro* transcribed using the GeneArt precision gRNA synthesis kit (Thermo Fisher, Waltham, MA) according to the manufacturer's instructions. See **Supplemental Figure S1, Table S1**, and references (Shaw et al., 2017; Hall et al., 2018; Tsai et al., 2018) for details on targeting sequences/locations and sgRNA efficiency. Zebrafish embryos from a single clutch from a natural mating of a ZDR background founder pair were either uninjected or injected into the cell at the 1-cell stage with a 1 nl cocktail of 100 pg/nl sgRNA, 200 pg/nl Cas9 protein (PNA Bio, Newbury Park, CA), or a combination of both reagents. We extracted genomic DNA (gDNA) from tail clips of parental zebrafish or whole zebrafish embryos at 4 dpf. All zebrafish experiments were approved by the Duke University Institutional Care and Use Committee (Protocol A154-18-06).

Sample Selection for Sequencing

The ZDR strain in our laboratory gives consistently robust clutch sizes of ~100 embryos. To preserve enough individuals to generate an F1 generation, we anticipated that we would have approximately 50 individuals available for exome sequencing. Using the CFD score cut-off of 0.2 as a threshold for the likelihood of inducing transmissible off-target mutations, we expected that we would need at least 5-6 embryos per condition to observe one of these events. Thus, we selected six independent embryos per gRNA plus Cas9 condition for comparison with controls while maintaining the experiment within a single clutch to control for inherited variation.

Heteroduplex Editing Efficiency by PAGE

For each sgRNA plus Cas9 condition we PCR-amplified gDNA from 12 embryos per batch using site-specific primers and screened for heteroduplex formation as described (Zhu et al., 2014). Five samples with evidence of heteroduplex formation were gel purified alongside a control sample, 'A' overhangs were added to the PCR products, and the products were cloned into a TOPO4 vector (Thermo Fisher). We picked 12 colonies per embryo to estimate targeting efficiency by Sanger sequencing.

Whole Exome Sequencing

We used the manufacturer protocol for the Agilent SureSelect Capture kit for non-human exomes with 200 ng gDNA per individual (75 Mb capture designed on the zv9 version of the zebrafish genome; Agilent SSXT Zebrafish All Exon kit; Agilent Technologies, Santa Clara, CA). Samples were multiplexed and run across two lanes of the Illumina HiSeq 4000 as paired-end 150 bp reads. Sequence data were demultiplexed and Fastq files were generated using Bcl2Fastq conversion software (Illumina, San Diego, CA).

Variant Calling

Sequencing reads were processed using the TrimGalore toolkit (Krueger, 2017) which employs Cutadapt to trim low quality bases and Illumina sequencing adapters from the 3' end of the reads. Only reads that were 20 nt or longer after trimming were kept for further analysis. Using the BWA (v. 0.7.15) MEM algorithm (Li, 2013), reads were mapped to the Zv9 version of the zebrafish genome. Picard tools (Picard, 2017) (v. 2.14.1) were used to remove PCR duplicates and to calculate sequencing metrics. The Genome Analysis Toolkit (McKenna et al., 2010) (GATK, v. 3.8-0) MuTect2 caller was used to call variants between each experimental condition and the adult male and adult female samples separately. Independently, aligned reads were locally realigned with the GATK IndelRealigner and then processed with Samtools mpileup (Li, 2011) for variant calling with VarScan2 trio (Koboldt et al., 2013). VarScan2 variant call sets were generated with the minimum coverage specified at 30x.

Variant Analysis

We used BEDOPS (Neph et al., 2012) and Bedtools (Quinlan and Hall, 2010) intersect, window, and merge commands to exclude variants with support in either parent, variants reported to occur in wild-type zebrafish strains ensembl dbSNP version 79, variants in repeat regions or regions of predicted segmental duplication in the genome (Khaja et al., 2006), variants reported in both control individuals and CRISPR-edited individuals, and variants reported at the on-target locations for CRISPR-editing. The potential for variants to occur due to off-target CRISPR-mediated editing was assessed by comparing variant counts between groups with either a Wilcoxon rank test for two groups, or a Kruskal-Wallis rank test for more than two groups and assessing the p-value against a Bonferroni critical value to correct for multiple testing. In addition, variants from samples were compared with locations of predicted off-target regions (formatted into a bed file) from three algorithms: CRISPOR (Concordet and Haeussler, 2018), the CRISPRdirect engine with 12-mer to 20-mer hits, or Cas9-OFFinder allowing 3-mismatches and 1-bulge in either DNA or RNA. Hypergeometric p-values calculated with the Rothstein lab hypergeometric calculator, use the capture space (74691693 bp) as the population size, and a reasonable high vs low sequencing error rate for our Illumina platform (.24% vs .1%) (Pfeiffer et al., 2018) to calculate the expected number of population variants called by chance at a position covered at the F0 average read depth (4 or more errant reads at the position; AF > .05).

RESULTS

Generating and Sequencing CRISPR-Cas9-Edited F0 and F1 Individuals

We focused on three different genes (*anln*, *kmt2d*, and *smchd1*) for which a) we have substantial experience in this model organism and b) give reproducible, quantitative defects in kidney morphogenesis (Hall et al., 2018), mandibular and neuronal development (Tsai et al., 2018), and craniofacial morphogenesis (Shaw et al., 2017). For each locus, we used sgRNAs that had the following three characteristics. First, for each of the three genes, we selected an sgRNA with demonstrated high efficiency (100%) and an sgRNA

with low efficiency (~30%), as determined by heteroduplex analysis and Sanger sequencing of cloned PCR products (Shaw et al., 2017; Hall et al., 2018; Tsai et al., 2018) (**Supplementary Figure 1**). Second, we mandated that all sgRNAs have a high specificity score (MIT specificity score 79–99 for each sgRNA; **Supplementary Table S1**). Finally, we required that each sgRNA was predicted to generate few off-target effects. We used CRISPOR to assess the cutting frequency determination (CFD) scores of the sgRNAs and observed few predicted off-target loci at high risk (CFD > 0.2) genome-wide (mean = 0.17, range = 0–0.73; **Supplementary Figure 2**). In the exome, CRISPOR predicts 0–3 high risk loci per sgRNA.

Next, we co-injected each sgRNA and Cas9 protein into wild-type zebrafish embryos from the same clutch at the 1-cell stage. For each sgRNA, we harvested DNA from six edited individuals to serve as technical replicates. In addition, we collected DNA from two individuals for each of the following conditions: uninjected, sgRNA alone, or Cas9 alone (**Figure 1A**). Finally, to assess the potential transmission of *de novo* variants to the next generation, we raised the F0 cohort for the *smchd1* high efficiency sgRNA and intercrossed adults to obtain the F1 generation. We did not observe defects in fecundity or the expression of inconsistent phenotypes within the cohort. In total, we performed whole exome sequencing (WES) on two parents, 52 F0 individuals and 16 F1 individuals (**Figure 1A**). WES resulted in 76x average target coverage in F0 samples and 115x average target coverage in F1 individuals (**Figures 1B, C**). The F0 sequencing data covered 83% of the exome at $\geq 30\times$ and 65% at $\geq 50\times$. The F1 sequencing data covered 88% of the exome at $\geq 30\times$ and 78% of the exome at $\geq 50\times$.

De Novo Mutation Counts Are Not Inflated in F1 Exomes

Low-level mosaicism remains challenging to detect in WES data and it is prone to high false-positive and false-negative rates (Sandmann et al., 2017). For this reason, we first focused on transmitted events. If CRISPR-Cas9 editing does induce off-target *de novo* mutations, we should observe an increase above baseline in the number of heterozygous variants fixed in the CRISPR-edited F1 generation that were absent from the grandparents.

Given the estimated 0.01% gene level baseline mutation rate in zebrafish (Mullins et al., 1994), we expect approximately 2–3 exonic changes per generation. To measure the observed rates, we applied a trio sequencing workflow aligned with best practices for the Genome Analysis Toolkit (GATK) and we called both single nucleotide variants and indels with two established variant callers: VarScan2 or Mutect2 (**Figure 1D**). Starting with all calls, we performed multiple data filtering steps. First, we removed variants present in either of the grandparental exomes. Second, since a small number of variants might have appeared *de novo* because of missing data from either grandparent, we also excluded alleles reported in the zebrafish ensembl dbSNP database. Third, we removed variants from the on-target genome locations (**Supplementary Figure 3**). Together, these three filters removed 79% of the MuTect2 and 99% of the VarScan2 calls. As an additional data filtering step, we removed repetitive elements, regions of potential segmental duplication in zebrafish, and indel variants containing homo-, dinucleotide, and trinucleotide repeats. This step improved the transition-transversion ratio from 0.91 to 1.09 which approaches a previously reported ratio

of 1.2 for zebrafish (Stickney et al., 2002) (**Supplementary Figure 4**). Finally, we removed cross-noise variants found in two or more samples that likely represent systematic technical error or uncalled low-level mosaics from the grandparents.

Using this dataset (**Supplementary Table S2**), we then applied a filter for allele frequency (AF) above 0.3 to capture the fixed heterozygous variants and we compared the variant count differences between F1 embryos derived from edited and unedited F0 adults. VarScan2 reports candidate variant counts closer to the expected natural accumulation of *de novo* mutations in F1 than MuTect2 (average 20 vs 66, respectively; **Supplementary Table S3**). We calculated the critical p-value threshold Bonferroni correction for three groups ($p < 0.012$), and neither calling method reports a significant difference between progeny of edited and control adults ($p > 0.11$; Wilcoxon rank test, **Supplementary Table S4**).

Next, we focused on the VarScan2 results. Based on the >5-fold inflation of observed versus expected variant calls across the cohort (mean of 20 vs 2–3, respectively; **Figure 2A**) we hypothesized that these agnostically filtered calls still included false positives. Therefore, we reviewed the variant calls in the Integrative Genomics Viewer (IGV). We found two sources of false positives (**Supplementary Figure 5**). First, a subset of read alignments filled into small deletions observed in the grandparents rather than extend a gap (83% of calls). Second, local realignments involving small deletions misalign in the progeny, even though an alternative placement of the deletion results in a grandparental genotype (10% of calls).

Of the remaining calls, half were deemed unlikely to be bona fide variants for other reasons. These included complex regions with many error prone reads; abundance of mis-mapped read pairs; and remaining low level mosaicism in grandparents. The other half were unambiguous *de novo* heterozygous variants (**Figure 2B**). Notably, most of the unambiguous variants were also called by MuTect2 (10 of 11; **Figure 2C**). For this population of alleles, we observed no difference between control and edited groups called by both callers (**Supplementary Table S5**). Crucially, we confirmed all of the variants detected by both callers in F1 animals derived from CRISPR/Cas edited individuals by Sanger sequencing. While we were encouraged by these results, the two agnostic filtering criteria removing dbSNP calls and cross-noise variants within the same guide may have artificially reduced our candidate variant pool and caused us to overlook potential CRISPR-induced editing. We performed a re-analysis of these filters by: 1) removing the dbSNP filter entirely (**Supplementary Table S6**) and evaluating new variant calls (**Supplementary Figure 6**) and 2) evaluating the subset of variants called in more than one individual (**Supplementary Table S7**). Taken together, we found that, regardless of whether we consider agnostic or manually reviewed variant numbers, there is no predilection toward inflated variant counts in F1 offspring derived from edited versus control groups. Further, the observed number of *de novo* variants in F1s does not exceed the expected rate of 2–3 per exome, per generation.

De Novo Mutation Counts Are Not Inflated Across the Multigenerational Cohort

We then returned to the F0 cohort to investigate whether variant burden outside of the targeted locus differed among individuals injected with sgRNA in the presence or absence of Cas9. Importantly,

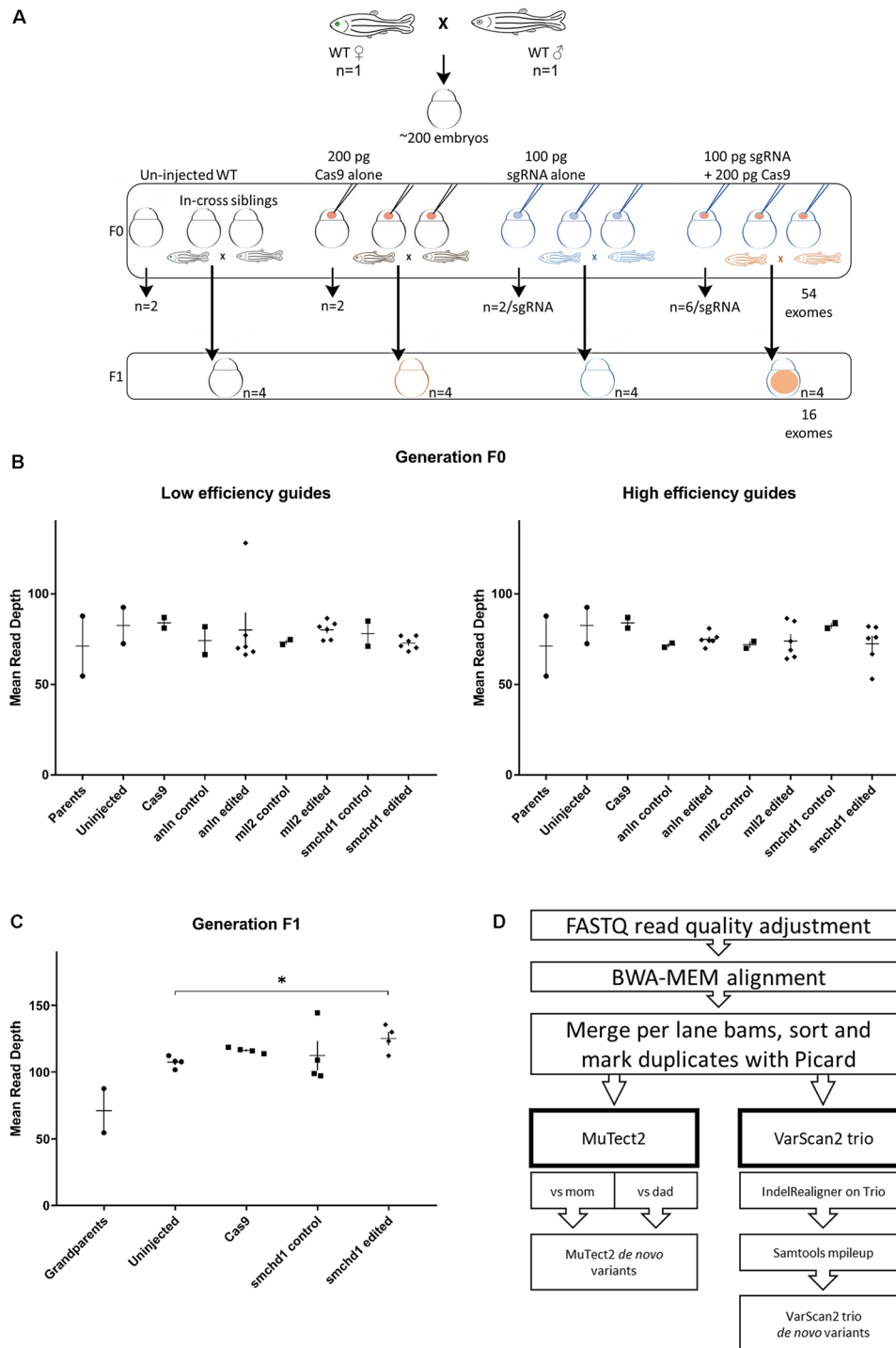
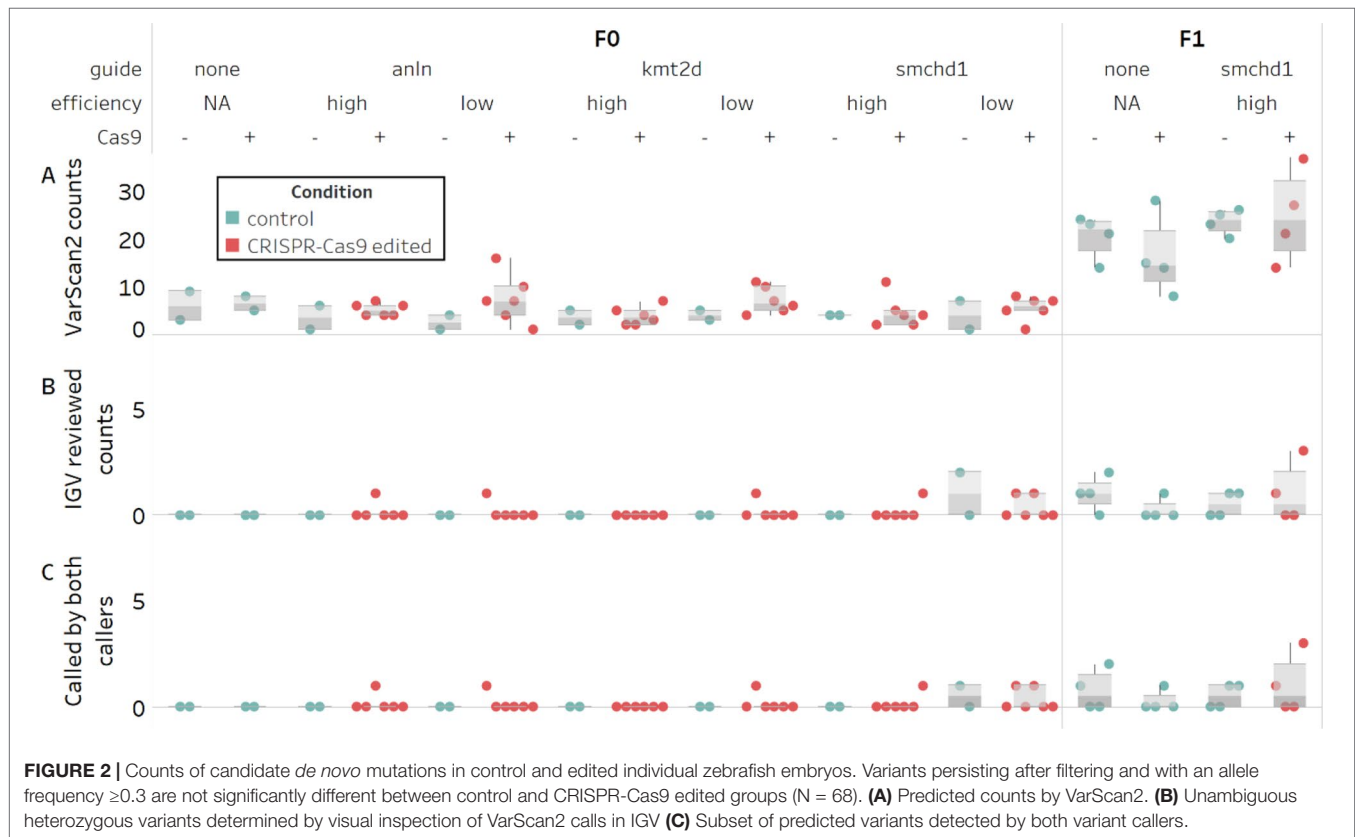


FIGURE 1 | Whole exome sequencing in two generations of CRISPR-Cas9 edited zebrafish. **(A)** The experimental design generates a single clutch of ~200 embryos from a founder pair of parents from the ZDR laboratory strain of wild-type zebrafish. The embryos were randomly assigned to four experimental arms: uninjected controls, Cas9 injected controls, sgRNA injected controls, and Cas9 + sgRNA gene edited samples. A total of 52 embryos were sampled for DNA extraction and sequencing at 4 dpf in the F0 generation (2 uninjected, 2 Cas9 injected, 2 sgRNA injected across 6 different sgRNAs targeting 3 genes for a total of 12 embryos, and 6 CRISPR-Cas9 embryos per sgRNA guide for a total of 36 edited individuals). Additional embryos for each condition were injected concurrently, but raised to adulthood. The F0 in-cross from pairs edited with the *smchd1* high efficiency guide generated F1 progeny for further sequencing: We sampled offspring from 4 uninjected, 4 Cas9 injected, 4 sgRNA injected, and 4 CRISPR-Cas9 injected embryos for a total of 16 F1 exomes. **(B)** The first round of exome sequencing (F0 and parents) generated a consistent read depth averaging 76x coverage. **(C)** The second round of exome sequencing (F1) generated a consistently higher read depth averaging 115x coverage. The *smchd1* edited individuals are also sequenced to a higher depth than the uninjected controls ($p < 0.05$). **(D)** After sequencing quality control and alignment, variant calling was performed with both somatic and germline callers to identify candidate *de novo* mutations.



the expected allelic series of variants are reported robustly at the on-target locations of the sgRNAs against two of the target genes, *anln* on chromosome 19 and *kmt2d* on chromosome 23 (**Supplementary Figure 3A**) (Hall et al., 2018; Tsai et al., 2018). No on-target variants are observed for the *smchd1* locus because our exome capture did not include baits for this locus in the *Zv9* assembly of the zebrafish genome. However, we demonstrated experimentally the on-target CRISPR-editing capability of the two *smchd1* sgRNAs and the transmission of on-target variants produced by the high-efficiency sgRNA to the F1 generation via Sanger sequencing (**Supplementary Figure 3B**), as described (Shaw et al., 2017).

We first considered the agnostic off-target VarScan2 variants called in the mosaic F0 generation (**Supplementary Table S8**). Initially, we applied the same arbitrary 0.3 AF threshold that we used with the F1 calls, reasoning that editing occurs at the one-to-two cell stage and would likely manifest as an off-target inflation at high allele frequencies. We determined the Bonferroni correction threshold for four groups ($p < 0.012$), and again, we did not observe a significant inflation in *de novo* variant counts between control and F0 edited groups, in either the algorithmically predicted counts or the manually reviewed counts ($p > 0.15$; Wilcoxon rank test; **Figures 2A, B; Supplementary Table S9**). We then repeated the analysis on the agnostic MuTect2 call set, and consistent with the filtered VarScan2 data, we did not observe an inflation in *de novo* mutation counts between control and edited groups ($p > 0.04$; **Supplementary Table S7**). Finally, because a 0.3 AF may fail to detect inefficient targeting events or lower mosaicism levels, we tested lower cutoff frequencies. We expected that as we

lowered the AF threshold beyond 10%, the sensitivity of the caller would decrease (Xu, 2018). However, at either an arbitrary 0.1 AF threshold, or without applying a threshold, we still observe no significant differences ($p > 0.08$; **Supplementary Table S9**).

For the VarScan2 dataset generated from F0 exomes, the variant count exceeded the expected 2-3 *de novo* changes per exome in at least one individual in half of the edited conditions (**Figure 2A**). To exclude the possibility that these could be false positive calls, similar to what we observed in the F1 cohort, we inspected all variants exceeding the 0.3 AF cutoff using IGV. We found that this dataset also was subject to similar technical artifacts as observed for F1; exclusion of these variants brought the *de novo* mutation call number within the expected range (**Figure 2B**). Using the same Bonferroni correction for four groups ($p < 0.012$), we were unable to detect a difference between control versus edited groups ($p > 0.38$; **Supplementary Table S10**). Since we had observed that variants detected by both callers represented an unbiased way to assess high confidence calls in F1, we also asked whether we could detect a difference in variant counts in this subset of calls in F0 (7 of 8 unambiguous calls; **Figure 2C**). Again, we observed no significant differences between controls and edited groups ($p > 0.78$; **Supplementary Table S11**).

De Novo Mutations Are Not Observed At Predicted Off-Target Sites

To examine the potential incidence of off-target mutations more sensitively, we removed the filters on the variant calls and searched predicted off target sites across our multigenerational

cohort using three algorithms: the MIT CRISPR design site, the CRISPR-direct engine, and CAS-OFFinder, for any variants occurring within 100 bp flanking a predicted off-target site. Consistent with previous reports (Hruscha et al., 2013; Varshney et al., 2015), we found no support for single nucleotide variants or small indels occurring at predicted off-target locations in the F1 generation, and sporadic low allele frequency calls near predicted off-target regions in F0s. The number of reported variants in the F0 samples are not significantly different than expected by chance ($p > 0.08$; **Supplementary Table S12**).

We reviewed the 15 reported variant calls near predicted off-target sites in F0s, and found that none are supported by both variant callers (**Supplementary Table S13**). Seven are also reported in siblings subjected to editing with alternative guides or control conditions, making them unlikely to be induced by Cas9-mediated genome editing. Another four were not supported by reads on both strands. Of the four remaining variants, one was only reported in a control condition, making it unlikely to be a result of editing. The other three occur at a 5% alternate allele frequency, near the limit of detection for the variant callers, increasing the likelihood that they may be artifacts. We do note that one variant has features consistent with an expected off-target cut. This is a small deletion reported directly at a predicted off-target cut site detected by two prediction engines. Notably, this small deletion occurs in an exonic region, has a high CFD risk score (CFD score = 0.52), and is observed at the predicted locus in a few reads from the VarScan2 call set as well, even though it is not called by that algorithm. Together, our analysis of reported variants near predicted off-target sites detects one potential off-target variant at low allele frequency in a single individual and does not demonstrate an inflated or transmissible mutation burden conjoint with expected on-target deletions.

DISCUSSION

Trio sequencing designs enable off-target analyses to distinguish gene editing effects from natural and inherited genetic variation. In our study, the bulk of variant calls in zebrafish exomes are filtered out due to their existence in the parental strain. Our ability to recover transmissible on-target deletions and Sanger-validated *de novo* mutations outside of predicted off-target regions and in quantities indistinguishable from natural variation suggests that off-target CRISPR events occur infrequently.

Our results are consistent with previous results in zebrafish demonstrating limited off-target activity at select predicted regions (Hruscha et al., 2013; Varshney et al., 2015) and with recent work in mice that found limited support for off-target effects genome-wide (Iyer et al., 2018). Indeed, limited assessments in several organisms including dog (Zou et al., 2015), goat (Li et al., 2018), and pig (Carey et al., 2019) have suggested few off-target effects. An advantage to our approach is the ability to generate and evaluate many individuals, and we have observed neither unexpected phenotypes nor additional off-target events. While our unbiased assessment is limited to detecting potential off-target variation within the exon-capture space of the genome, this analysis expands the search space considerably beyond the few algorithmically predicted sites investigated in preceding studies in zebrafish and other organisms.

Though off-target editing in non-coding regions of the genome will need to be assessed as well, the interpretation of such changes and their influence on gene expression will become more powerful as the genomic annotation of variation in these regions in unedited individuals becomes more widely available. Several large-scale projects in the zebrafish community are currently seeking to fill this need, including the DANIO-CODE project (<https://danio-code.zfin.org/>), and we look forward to having the community resources to better address these questions in the future. Furthermore, we did not assess large structural variants or long deletions at the on-target site. In addition, we occasionally observed trends toward variant inflation in the predicted variant call sets that were related to sequencing depth and did not survive visual inspection or cross-validation with a secondary variant caller. This observation suggests that even with trio designs and other precautionary measures, care should be exercised in interpreting variant predictions agnostically and that sequencing even more individuals per condition may be required to expose subtle differences in off-target effects.

In response to initial reports that CRISPR-Cas9 edited mammalian cells harbored off-target variants (Fu et al., 2013; Zhang et al., 2015), many iterative improvements in technology and experimental design have outlined conditions for achieving CRISPR-Cas9 gene editing while limiting off-target events. Our experimental and sgRNA design incorporated such advancements (high on-target MIT ranking, low off-target CFD scores, high cutting efficiency, and short Cas9 exposure), minimizing the chance of inducing off-target events to the extent possible within a typical experimental design for generating loss-of-function genetic models *in vivo*. Complementary approaches like DIG-Seq have recently shown empirically that the *in vivo* context itself further reduces the incidence of off-targeting events (Kim and Kim, 2018). However, unexpected nuances of the CRISPR-Cas9 editing system continue to emerge. Varied biological responses to CRISPR-Cas9, such as DNA damage repair (Haapaniemi et al., 2018), enzymatic immunity (Crudele and Chamberlain, 2018), and alternative templating (Ma et al., 2017) exemplify our still nascent understanding of DNA and RNA editing. While the reversibility of RNA editing provides an enticing possibility for reducing the risk of off-target events, the off-target rates, effects, and subsequent engineering advances to RNA editing systems like Cas13 and adenosine deaminase acting on RNA (ADAR) are still emerging as well (Cox et al., 2017; Katrekar et al., 2019). Furthermore, natural human genetic variation has been shown to influence both the efficaciousness of on-target DNA editing and the frequency of off-target events (Lessard et al., 2017); an observation that may extend to RNA editing technologies as well. Under these circumstances, use of emergent computational, laboratory, and animal modeling tools and unbiased genome-wide off-target assessments will facilitate the foundational knowledge required to reduce unnecessary risk in practice.

DATA AVAILABILITY STATEMENT

The dataset generated for this study was submitted to the Sequence Read Archive (SRA) and can be accessed by searching the BioProject ID PRJNA525401 on the NCBI website (<https://www.ncbi.nlm.nih.gov/bioproject/PRJNA525401>)

ETHICS STATEMENT

Zebrafish experiments were approved by the Duke University Institutional Care and Use Committee (Protocol A154-18-06).

AUTHOR CONTRIBUTIONS

NK conceived and designed the study. MM and ED processed the biological samples. MM performed the informatics and statistical analysis and drafted the manuscript. All authors contributed to manuscript revision and editing, and read and approved the submitted version.

FUNDING

This work was supported by a fellowship from U.S. National Institutes of Health Grant 5T32HG008955-02 (MM).

REFERENCES

- Anderson, K. R., Haeussler, M., Watanabe, C., Janakiraman, V., Lund, J., Modrusan, Z., et al. (2018). CRISPR off-target analysis in genetically engineered rats and mice. *Nat. Methods* 15, 512–514. doi: 10.1038/s41592-018-0011-5
- Carey, K., Ryu, J., Uh, K., Lengi, A. J., Clark-Deener, S., Corl, B. A., et al. (2019). Frequency of off-targeting in genome edited pigs produced via direct injection of the CRISPR/Cas9 system into developing embryos. *BMC Biotechnol.* 19, 25. doi: 10.1186/s12896-019-0517-7
- Concordet, J.-P., and Haeussler, M. (2018). CRISPOR: intuitive guide selection for CRISPR/Cas9 genome editing experiments and screens. *Nucleic Acids Res.* 46, W242–W245. doi: 10.1093/nar/gky354
- Crudele, J. M., and Chamberlain, J. S. (2018). Cas9 immunity creates challenges for CRISPR gene editing therapies. *Nat. Commun.* 9, 3497. doi: 10.1038/s41467-018-05843-9
- Cox, D. B. T., Gootenberg, J. S., Abudayyeh, O. O., Franklin, B., Kellner, M. J., Joung, J., et al. (2017). RNA editing with CRISPR-Cas13. *Science* 358, 1019–1027. doi: 10.1126/science.aag0180
- Doench, J. G., Fusi, N., Sullender, M., Hegde, M., Vaimberg, E. W., Donovan, K. F., et al. (2016). Optimized sgRNA design to maximize activity and minimize off-target effects of CRISPR-Cas9. *Nat. Biotechnol.* 34, 184–191. doi: 10.1038/nbt.3437
- Fu, Y., Foden, J. A., Khayter, C., Maeder, M. L., Reyon, D., Joung, J. K., et al. (2013). High-frequency off-target mutagenesis induced by CRISPR-Cas nucleases in human cells. *Nat. Biotechnol.* 31, 822–826. doi: 10.1038/nbt.2623
- Haapaniemi, E., Botla, S., Persson, J., Schmierer, B., and Taipale, J. (2018). CRISPR-Cas9 genome editing induces a p53-mediated DNA damage response. *Nat. Med.* 24, 927–930. doi: 10.1038/s41591-018-0049-z
- Hall, G., Lane, B. M., Khan, K., Padiaditakis, I., Xiao, J., Wu, G., et al. (2018). The human FSGS-causing ANLN R431C mutation induces dysregulated PI3K/AKT/mTOR/Rac1 signaling in podocytes. *J. Am. Soc. Nephrol.* 29, 2110–2122. doi: 10.1681/ASN.2017121338
- Ho, B. X., Loh, S. J. H., Chan, W. K., and Soh, B. S. (2018). *In vivo* genome editing as a therapeutic approach. *Int. J. Mol. Sci.* 19, 2721. doi: 10.3390/ijms19092721
- Hruscha, A., Krawitz, P., Rechenberg, A., Heinrich, V., Hecht, J., Haass, C., et al. (2013). Efficient CRISPR/Cas9 genome editing with low off-target effects in zebrafish. *Development* 140, 4982–4987. doi: 10.1242/dev.099085
- Hu, J. H., Miller, S. M., Geurts, M. H., Tang, W., Chen, L., Sun, N., et al. (2018). Evolved Cas9 variants with broad PAM compatibility and high DNA specificity. *Nature* 556, 57–63. doi: 10.1038/nature26155
- Iyer, V., Boroviak, K., Thomas, M., Doe, B., Riva, L., Ryder, E., et al. (2018). No unexpected CRISPR-Cas9 off-target activity revealed by trio sequencing of gene-edited mice. *PLoS Genet.* 14, e1007503. doi: 10.1371/journal.pgen.1007503
- Katrekar, D., Chen, G., Meluzzi, D., Ganesh, A., Worlikar, A., Shih, Y.-R., et al. (2019). *In vivo* RNA editing of point mutations via RNA-guided adenosine deaminases. *Nat. Methods* 16, 239. doi: 10.1038/s41592-019-0323-0
- Khaja, R., MacDonald, J. R., Zhang, J., and Scherer, S. W. (2006). Methods for identifying and mapping recent segmental and gene duplications in eukaryotic genomes. *Methods Mol. Biol. Clifton NJ* 338, 9–20. doi: 10.1385/1-59745-097-9:9
- Kim, D., and Kim, J.-S. (2018). DIG-seq: a genome-wide CRISPR off-target profiling method using chromatin DNA. *Genome Res.* 28, 1894–1900. doi: 10.1101/gr.236620.118
- Koboldt, D. C., Larson, D. E., and Wilson, R. K. (2013). Using VarScan 2 for germline variant calling and somatic mutation detection. *Curr. Protoc. Bioinforma. Ed. Board Andreas Baxeavanis Al* 44, 15.4.1–15.4.17. doi: 10.1002/0471250953.bi1504s44
- Kosicki, M., Tomberg, K., and Bradley, A. (2018). Repair of double-strand breaks induced by CRISPR-Cas9 leads to large deletions and complex rearrangements. *Nat. Biotechnol.* 36, 765–771. doi: 10.1038/nbt.4192
- Krueger, F. (2017). Trim galore! (version 0.4.3) Babraham Bioinformatics. Available from <https://github.com/FelixKrueger/TrimGalore>.
- Labun, K., Montague, T. G., Gagnon, J. A., Thyme, S. B., and Valen, E. (2016). CHOPCHOP v2: a web tool for the next generation of CRISPR genome engineering. *Nucleic Acids Res.* 44, W272–W276. doi: 10.1093/nar/gkw398
- Lessard, S., Francioli, L., Alfoldi, J., Tardif, J.-C., Ellinor, P. T., MacArthur, D. G., et al. (2017). Human genetic variation alters CRISPR-Cas9 on- and off-targeting specificity at therapeutically implicated loci. *Proc. Natl. Acad. Sci.* 114, E11257–E11266. doi: 10.1073/pnas.1714640114
- Li, C., Zhou, S., Li, Y., Li, G., Ding, Y., Li, L., et al. (2018). Trio-based deep sequencing reveals a low incidence of off-target mutations in the offspring of genetically edited goats. *Front. Genet.* 9, 449. doi: 10.3389/fgene.2018.00449
- Li, H. (2013). BWA-MEM (version 0.7.15). Available from <https://github.com/lh3/bwa>.
- Li, H. (2011). A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics* 27, 2987–2993. doi: 10.1093/bioinformatics/btr509
- Ma, H., Marti-Gutierrez, N., Park, S.-W., Wu, J., Lee, Y., Suzuki, K., et al. (2017). Correction of a pathogenic gene mutation in human embryos. *Nature* 548, 413–419. doi: 10.1038/nature23305
- McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytzky, A., et al. (2010). The genome analysis toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* 20, 1297–1303. doi: 10.1101/gr.107524.110
- Mooney, M., Davis, E. E., and Katsanis, N. (2019). Analysis of single nucleotide variants in CRISPR-Cas9 edited zebrafish embryos shows no evidence of off-target inflation. *bioRxiv*. <https://www.biorxiv.org/content/10.1101/568642v1>. doi: 10.1101/568642

ACKNOWLEDGMENTS

We are grateful to I-Chun Tsai, Maria Kousi, Zachary Kupchinsky and Igor Padiaditakis for technical assistance. We thank Nicolas Devos (Duke Sequencing and Genomic Technologies Shared Resource) and David Corcoran (Duke Genomic Analysis and Bioinformatics Shared Resource) for sequencing and informatics support, respectively. Some analyses were carried out using resources from the Duke Compute Cluster. NK is a Distinguished Jean and George Brumley Professor. This manuscript has been released as a Pre-Print at bioRxiv: (Mooney et al., 2019 <https://www.biorxiv.org/content/10.1101/568642v1>)

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fgene.2019.00949/full#supplementary-material>.

- Mullins, M. C., Hammerschmidt, M., Haffter, P., and Nüsslein-Volhard, C. (1994). Large-scale mutagenesis in the zebrafish: in search of genes controlling development in a vertebrate. *Curr. Biol.* 4, 189–202. doi: 10.1016/S0960-9822(00)00048-8
- Neph, S., Kuehn, M. S., Reynolds, A. P., Haugen, E., Thurman, R. E., Johnson, A. K., et al. (2012). BEDOPS: high-performance genomic feature operations. *Bioinformatics* 28, 1919–1920. doi: 10.1093/bioinformatics/bts277
- Pfeiffer, F., Gröber, C., Blank, M., Händler, K., Beyer, M., Schultze, J. L., et al. (2018). Systematic evaluation of error rates and causes in short samples in next-generation sequencing. *Sci. Rep.* 8, 10950. doi: 10.1038/s41598-018-29325-6
- Picard (2017). Broad Institute.
- Quinlan, A. R., and Hall, I. M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* 26, 841–842. doi: 10.1093/bioinformatics/btq033
- Sandmann, S., de Graaf, A. O., Karimi, M., van der Reijden, B. A., Hellström-Lindberg, E., Jansen, J. H., et al. (2017). Evaluating variant calling tools for non-matched next-generation sequencing data. *Sci. Rep.* 7, 43169. doi: 10.1038/srep43169
- Shaw, N. D., Brand, H., Kupchinsky, Z. A., Bengani, H., Plummer, L., Jones, T. I., et al. (2017). *SMCHD1* mutations associated with a rare muscular dystrophy can also cause isolated arhinia and Bosma arhinia microphthalmia syndrome. *Nat. Genet.* 49, 238–248. doi: 10.1038/ng.3743
- Stickney, H. L., Schmutz, J., Woods, I. G., Holtzer, C. C., Dickson, M. C., Kelly, P. D., et al. (2002). Rapid mapping of zebrafish mutations with SNPs and oligonucleotide microarrays. *Genome Res.* 12, 1929–1934. doi: 10.1101/gr.777302
- Tsai, I.-C., McKnight, K., McKinstry, S. U., Maynard, A. T., Tan, P. L., Golzio, C., et al. (2018). Small molecule inhibition of RAS/MAPK signaling ameliorates developmental pathologies of Kabuki Syndrome. *Sci. Rep.* 8, 10779. doi: 10.1038/s41598-018-28709-y
- Varshney, G. K., Pei, W., LaFave, M. C., Idol, J., Xu, L., Gallardo, V., et al. (2015). High-throughput gene targeting and phenotyping in zebrafish using CRISPR/Cas9. *Genome Res.* 25, 1030–1042. doi: 10.1101/gr.186379.114
- Xu, C. (2018). A review of somatic single nucleotide variant calling algorithms for next-generation sequencing data. *Comput. Struct. Biotechnol. J.* 16, 15–24. doi: 10.1016/j.csbj.2018.01.003
- Zhang, X.-H., Tee, L. Y., Wang, X.-G., Huang, Q.-S., and Yang, S.-H. (2015). Off-target effects in CRISPR/Cas9-mediated genome engineering. *Mol. Ther. Nucleic Acids* 4, e264. doi: 10.1038/mtna.2015.37
- Zhu, X., Xu, Y., Yu, S., Lu, L., Ding, M., Cheng, J., et al. (2014). An efficient genotyping method for genome-modified animals and human cells generated with CRISPR/Cas9 system. *Sci. Rep.* 4, 6420. doi: 10.1038/srep06420
- Zou, Q., Wang, X., Liu, Y., Ouyang, Z., Long, H., Wei, S., et al. (2015). Generation of gene-target dogs using CRISPR/Cas9 system. *J. Mol. Cell Biol.* 7, 580–583. doi: 10.1093/jmcb/mjv061

Conflict of Interest: NK is a paid consultant for and holds significant stock of Rescindo Therapeutics, Inc. The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2019 Mooney, Davis and Katsanis. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.