

Boosting targeted genome editing using the hei-tag

Thomas Thumberger^{1,6}, Tinatini Tavhelidse-Suck^{1,2,6}, Jose Arturo Gutierrez-Triana^{1,3,6}, Alex Cornean^{1,2}, Rebekka Medert^{2,4,5}, Bettina Welz^{1,2,5}, Marc Freichel^{4,5}, and Joachim Wittbrodt^{1,5*}

¹Centre for Organismal Studies (COS), Heidelberg University, Im Neuenheimer Feld 230, D-69120 Heidelberg, Germany

²Heidelberg Biosciences International Graduate School (HBIGS), Heidelberg, Germany

³present address: Escuela de Microbiología, Facultad de Salud, Universidad Industrial, Santander, Colombia

⁴Institute of Pharmacology, Heidelberg University, Heidelberg, Germany

⁵DZHK (German Centre for Cardiovascular Research), partner site Heidelberg/Mannheim, Heidelberg, Germany

⁶These authors contributed equally to this work: Thomas Thumberger, Tinatini Tavhelidse-Suck, Jose Arturo Gutierrez-Triana.

thomas.thumberger@cos.uni-heidelberg.de, tinatini.tavhelidse@cos.uni-heidelberg.de, jagutri@uis.edu.co, alex.cornean@cos.uni-heidelberg.de, rebekka.medert@pharma.uni-heidelberg.de, bettina.welz@cos.uni-heidelberg.de, marc.freichel@pharma.uni-heidelberg.de, jochen.wittbrodt@cos.uni-heidelberg.de

*Corresponding author:

Joachim Wittbrodt

Developmental Biology and Physiology

Centre for Organismal Studies (COS), Heidelberg University

Im Neuenheimer Feld 230, D-69120 Heidelberg, Germany

Phone +49 (6221) 54-6499

Fax +49 (6221) 54-5639

Abstract

Precise, targeted genome editing by CRISPR/Cas9 is key for basic research and translational approaches in model and non-model systems. While active in all species tested so far, editing efficiencies still leave room for improvement. The bacterial Cas9 needs to be efficiently shuttled into the nucleus as attempted by fusion with nuclear localization signals (NLSs). Additional peptide tags such as FLAG- or myc-tags are usually added for immediate detection or straight-forward purification. Immediate activity is usually granted by administration of pre-assembled protein/RNA complexes. We present the 'hei-tag (high efficiency-tag)' which boosts the activity of CRISPR/Cas genome editing tools already when supplied as mRNA. The addition of the hei-tag, a myc tag coupled to an optimized NLS via a flexible linker, to Cas9 or a C-to-T base editor dramatically enhances the respective targeting efficiency. This results in an increase in bi-allelic editing, yet reduction of allele variance, indicating an immediate activity even at early developmental stages. The hei-tag boost is active in model systems ranging from fish to mammals, including tissue culture applications. The simple addition of the hei-tag allows to instantly upgrade existing and potentially highly adapted systems as well as to establish novel highly efficient tools immediately applicable at the mRNA level.

Introduction

In the last decade, the CRISPR/Cas9 system and its derivatives facilitated and revolutionized genome editing across all phyla (Nidhi et al., 2021). The efficiency of editing crucially depends on the on-site activity of the particular Cas9 enzymes used (usually *Streptococcus pyogenes* Cas9, SpCas9) in the nucleus. State-of-the-art Cas9 variants differ by peptide tags added to the N- and C-termini of the respective endonuclease resulting in reported different activities (Liu et al., 2021; Zhang et al., 2014). Employed tags usually comprise diverse nuclear localization signals (NLSs) and epitope tags (e.g. FLAG, Myc, HA) for potential protein purification or visualization. To achieve nuclear localization of the Cas9 enzyme, the monopartite NLS originating from the SV40 large T-antigen (Kalderon et al., 1984) or a bipartite NLS discovered in *Xenopus* nucleoplasmin are routinely employed (Dingwall et al., 1988). However, the nuclear localization activity of commonly used NLSs is tightly controlled during early development (Poon and Jans, 2005) and is first detectable during gastrulation. In fish embryos, an optimized artificial NLS (Inoue et al., 2016) (oNLS) facilitates prominent nuclear localization already immediately after fertilization, while the SV40 NLS acts most prominently much later and facilitates nuclear localization approximately at the 1000 cell stage. For high targeting efficiency with low mosaicism, a peak activity should be achieved in the zygote or at early cleavage stages. Here we present the hei-tag, a short bi-partite tag composed of a myc-tag and optimized NLSs at the N- and C-termini that boosts Cas9 or C-T base editor-mediated targeted genome editing *in organismo* and cell culture.

Results

Assessing the genome editing efficiency requires a reliable and quantitative readout based on an apparent phenotype. We established a quantitative assay for loss-of-eye-pigmentation to address the activity of different Cas9 variants in two teleost model systems, medaka (*Oryzias latipes*) and zebrafish (*Danio rerio*) covering a wide evolutionary distance of 200 million years (Furutani-Seiki and Wittbrodt, 2004). Our assay on retinal pigmentation provides a highly reproducible quantitative readout for the loss of the conserved transporter protein *oculocutaneous albinism type 2* (*oca2*), required for melanin biosynthesis (Fig. 1a). Only its bi-allelic inactivation results in the

loss of pigmentation of eyes and skin (Lischik et al., 2019). A prominent knock-out phenotype thus can either result from a single to few early events, or from many events at subsequent developmental stages. Although phenotypically indifferent, the allele variance (genetic mosaicism) reflects the time point of action.

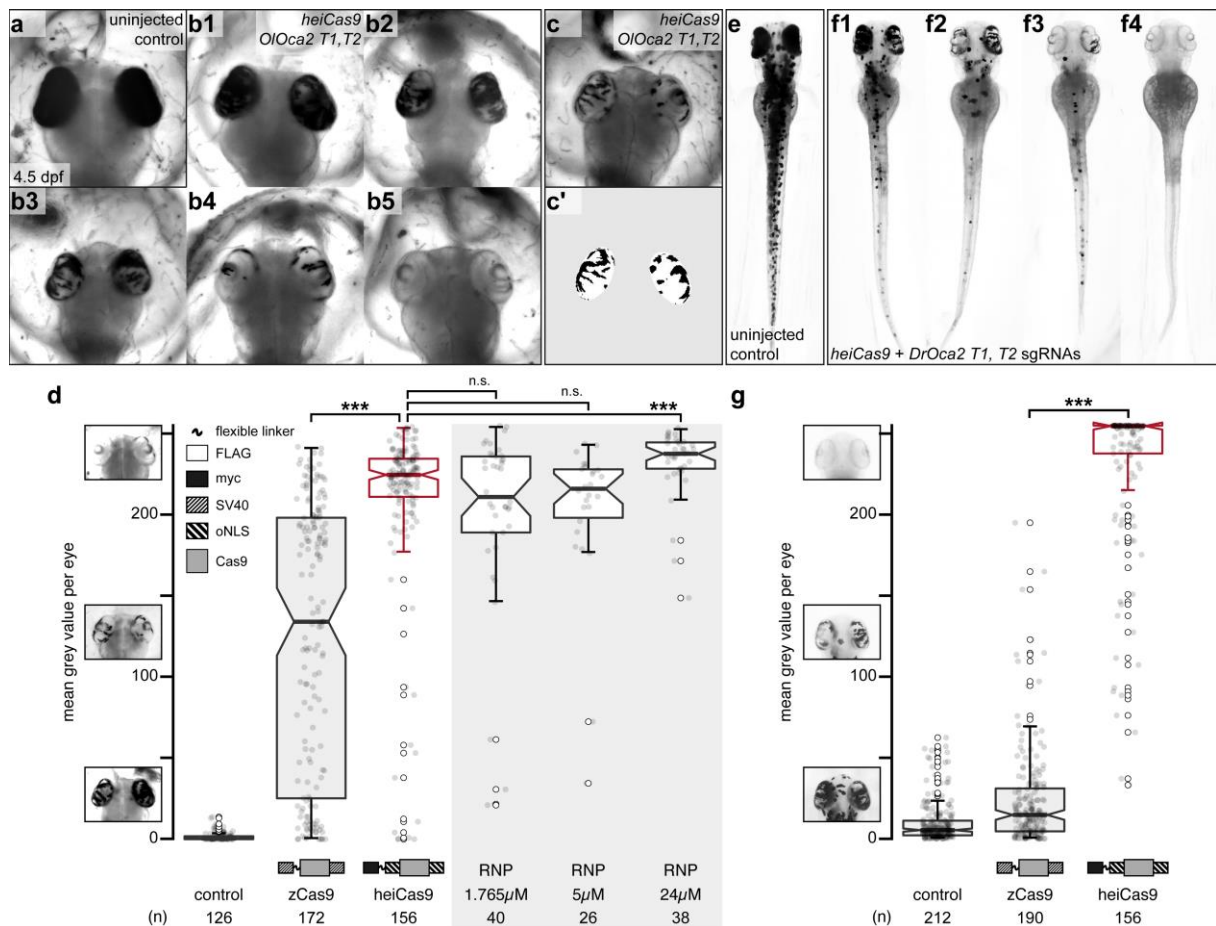


Fig. 1: heiCas9 exhibits outstanding bi-allelic targeting activity in fish.

Phenotypic range and quantification of *OIOca2* T1, T2 and *DrOca2* T1, T2 sgRNAs/Cas9 variant and sgRNA/Cas9 protein complex (RNP) mediated loss of pigmentation in medaka (a-d) and zebrafish (e-g) at high concentrations. (a) Fully pigmented eyes in un-injected control medaka embryo at 4.5 dpf. (b1-b5) Range of typically observed loss-of-pigmentation phenotypes upon injection with 150 ng/ μ l *heiCas9* mRNA and 30 ng/ μ l *OIOca2* T1, T2 sgRNAs. The observed phenotypes range from almost full pigmentation (b1) to completely unpigmented eyes (b5). (c) Minimum intensity projection of a medaka embryo at 4.5 days after injection with 150 ng/ μ l *heiCas9* and 30 ng/ μ l *OIOca2* T1, T2 sgRNAs. (c') Locally thresholded pigmentation on elliptical selection per eye (same embryo as in c). (d) Quantification

of mean grey values (0 = fully pigmented, 255 = completely unpigmented) of individual eyes from *Oca2* knock-out medaka crispants co-injected with 30 ng/μl *O/Oca2 T1, T2* sgRNAs and 150 ng/μl mRNAs of *zCas9* and *heiCas9* (red) compared to RNP injections (concentrations indicated). Medians: uninjected control = 0.4; *zCas9* = 134.5; *heiCas9* = 225.3; 1.765 μM RNP = 211.1; 5 μM RNP = 216.2; 24 μM RNP = 237.8. Note: highly significant pigment loss (70% increase) in *heiCas9* versus *zCas9* crispants ($p = 1.1e-25$); *heiCas9* reaches the same knock-out efficiency compared to RNP injections with only significant differences at highest RNP concentrations (24μM). (e) Fully pigmented uninjected control zebrafish embryo at 2.5 dpf. (f1-f4) Range of typically observed loss-of-pigmentation phenotypes upon injection with 150 ng/μl *heiCas9* mRNA and 30 ng/μl *DrOca2 T1, T2* sgRNAs. The observed phenotypes range from almost full pigmentation (f1) to completely unpigmented eyes and body (f4). (g) Quantification of mean grey values of individual eyes from *oca2* knock-out zebrafish embryos co-injected with 30 ng/μl *DrOca2 T1, T2* sgRNAs and 150 ng/μl mRNAs of *zCas9* and *heiCas9* (red) respectively. Medians: uninjected control = 5.3; *zCas9* = 14.7; *heiCas9* = 254.6. Note the very highly significant pigment loss (17-fold increase) in *heiCas9* versus *zCas9* crispants ($p = 2.1e-56$). dpf, days post fertilization; mean grey values ranged from 0, i.e. fully pigmented eye to 255, i.e. complete loss of pigmentation; n, number of eyes analyzed. Bold line, median. Statistical analysis performed in R, pairwise Wilcoxon rank sum test, Bonferroni corrected.

State-of-the-art protocols employ high concentrations of Cas9 and respective sgRNAs to ensure efficient on-site editing. To facilitate uniform Cas9 action, we followed our successful mRNA injection protocol (Gutierrez-Triana et al., 2018). One-cell stage medaka embryos were co-injected with sgRNAs targeting the *oca2* gene (*O/Oca2 T1, T2*) together with mRNA encoding a Cas9 endonuclease and mRNA encoding the Green Fluorescent Protein (GFP) as injection tracer. Injected embryos were fixed at 4.5 days post fertilization (Iwamatsu, 2004) well after the onset of pigmentation in control injections and subjected to image analysis (Fig. 1b). In brief, the eyes were segmented, (residual) pigmentation was thresholded (Fig. 1c-c') and

quantified according to mean grey values (0, i.e. fully pigmented, 255, i.e. completely unpigmented, Fig. 1d).

We first established the base activity level for the assay at standard conditions with high molar excess (150 ng/μl concentration) and determined the activity of a Cas9 variant codon optimized for zebrafish, i.e. a Cas9 carrying a SV40 NLS at the N- and C-terminus (nls-zCas9-nls, hereinafter: zCas9, Plasmid #47929 Addgene, (Jao et al., 2013)). The analysis of medaka *oca2* knock-out embryos injected with zCas9 revealed bi-allelic inactivation events of the *oca2* gene, yet with a strong overall variability as apparent by patchy unpigmented domains in the eyes (median of mean grey values = 134.5 compared to un-injected controls, median = 0.4; Fig. 1d). This patchy distribution of small, unpigmented areas indicated that bi-allelic targeting occurred only in few cells at later stages of development. To address whether different peptide domains (NLSs, Myc-tag, amino acid linkers) flanking the Cas9 enzyme enhance the targeting efficiency, we performed a permutation screen with Cas9 variants carrying these domains at different positions, which resulted in the identification of the „hei-tag“ (Figure 1-figure supplement 1). The hei-tag comprises a myc-tag connected via a flexible linker to an oNLS at the N-terminus complemented by a second oNLS fused to the C-terminus of a mammalian codon optimized Cas9 (see Supplementary File 1 for sequence) and in this conformation displayed highest editing activity. Any alteration of those domains in relative order or sequence negatively impacted on editing efficiency compared to the hei-tag (Supplementary File 2).

When assessing the activity of the resulting heiCas9 at high molar excess (standard conditions, 150 ng/μl), heiCas9 displayed a 70% increase in bi-allelic targeting efficiency versus the reference zCas9 (median zCas9 = 134.5, heiCas9 = 225.3; Fig. 1d) in medaka. Embryos co-injected with *heiCas9* mRNA and sgRNAs against *oca2* essentially lost pigmentation. The observed absence of pigmentation argues for an early time point of action due to high activity and efficient nuclear translocation of the tagged heiCas9 variant already at the earliest cleavage stages. In developing organisms, the time point of genome editing essentially impacts on the allele variance, i.e. the number of alleles established by the targeting attempt. To immediately provide a functional editing machinery, preassembled ribonucleoproteins (RNPs) containing Cas9 protein and guide RNA are popular, employing high molar excess/high concentrations of Cas9 (Kroll et al., 2021; Wu et al., 2018). Strikingly,

the editing efficiency of injected *heiCas9* mRNA was fully comparable to such RNP approaches (Fig. 1d, Figure 1-figure supplement 2).

To address whether the enhancement by *hei*-tag fusion to Cas9 is applicable to different models, we next compared the activities of the zCas9 and *heiCas9* in a second, evolutionarily distant fish species *Danio rerio* (zebrafish) targeting the orthologous *oca2* gene (sgRNAs *DrOca2 T1, T2* (Hammouda et al., 2019)). Injected and control embryos were fixed well after the onset of pigmentation at 2.5 dpf (Kimmel et al., 1995) (Fig. 1e-f) and subjected to the quantitative assay for eye pigmentation described above. Taking the activity of zCas9 as base level (median = 14.7), *heiCas9* delivered an outstanding targeting efficiency (median = 254.6), reflecting a 17-fold increase ($p = 2.1e-56$) (Fig. 1g, Figure 1-figure supplement 2). Similar to the results in medaka, yet even more pronounced, nearly unpigmented embryos were obtained with the *heiCas9*, arguing for highly efficient, early targeting. Taken together, addition of the *hei*-tag to a mammalian codon optimized Cas9 resulted in the highly efficient *heiCas9*, which boosted the targeting efficiency 17-fold, even when used at saturating concentrations. It prominently inactivated both alleles of the targeted *oca2* locus, with a putatively early onset of action upon injection of *heiCas9* mRNA and the respective sgRNAs at the one-cell stage.

To address whether the high targeting efficiency of *heiCas9* was conveyed by the high molar excess employed or was possibly restricted to the *oca2* locus, we turned to a multiplexing regime at ten-fold reduced concentrations of the Cas9 variants employed. We targeted four different genomic loci with four different sgRNAs: exonic targeting of *oca2* (*O/Oca2 T2*), targeting of the start codon of the *retina-specific transcription factor 2* (*rx2*; (Stemmer et al., 2015)) and the *crystallin alpha a* (*cryaa*; (Stemmer et al., 2015)) as well as intronic targeting of *rx3* (Zilova et al., 2021). Medaka one-cell stage embryos were co-injected with a mix of 12.5 ng/μl per sgRNA, the ten-fold reduced (15 ng/μl) zCas9 or *heiCas9* mRNA and 20 ng/μl *mCherry* mRNA as injection tracer.

For each multiplexing experiment, the genomic DNA of three pools each containing eight randomly picked crispants was extracted at 4 dpf and subjected to allele specific genotyping via Illumina sequencing. In the multiplexing approaches a total of 823,898 reads for the zCas9 and 824,817 reads for the *heiCas9*, compared to 711,739 control reads were analyzed (Supplementary File 3, Figure 2-figure

supplement 1). In all cases heiCas9 performed dramatically better than the reference zCas9 (Fig. 2a; mean percentage of modified alleles zCas9 (black dots) versus heiCas9 (red dots): *O/Oca2*: 3.38% vs. 54.59%, $p = 0.026$; *O/Rx2*: 20.82% vs. 95.85%, $p = 3.2e-06$; *O/Rx3*: 16.61% vs. 49.36%, $p = 0.0041$; *O/Cryaa*: 83.50% vs. 98.44%, $p = 0.039$). Strikingly, although the overall targeting efficiency was consistently higher as reflected by the high percentage of edited alleles (Fig. 2a), at the same time the allele variance was reduced in all cases when using heiCas9 (Fig. 2b; mean percentage of allele variance: zCas9 (black hollow dots) versus heiCas9 (red hollow dots): *O/Oca2*: 20.71% vs. 12.87%, $p = 0.025$; *O/Rx2*: 15.63% vs. 7.86%, $p = 7.6e-06$; *O/Rx3*: 17.91% vs. 12.75%, $p = 0.00021$; *O/Cryaa*: 10.17% vs. 8.74%, $p = 0.22$). This reduced allele variance for all multiplexed loci indicates an early editing by heiCas9. Given this and the overall higher targeting efficiency in all loci analyzed in the multiplexing approach, heiCas9 outperformed zCas9. It resulted in a massive performance boost, which was partially masked at saturating conditions, and now became fully apparent. The high efficiency of heiCas9 thus allows efficient editing at low concentrations with the potential to reduce off-target effects. Whether this putative reduction of off-targets is (over-)compensated by the efficient nuclear localization needs to be assessed by whole genome approaches in the future.

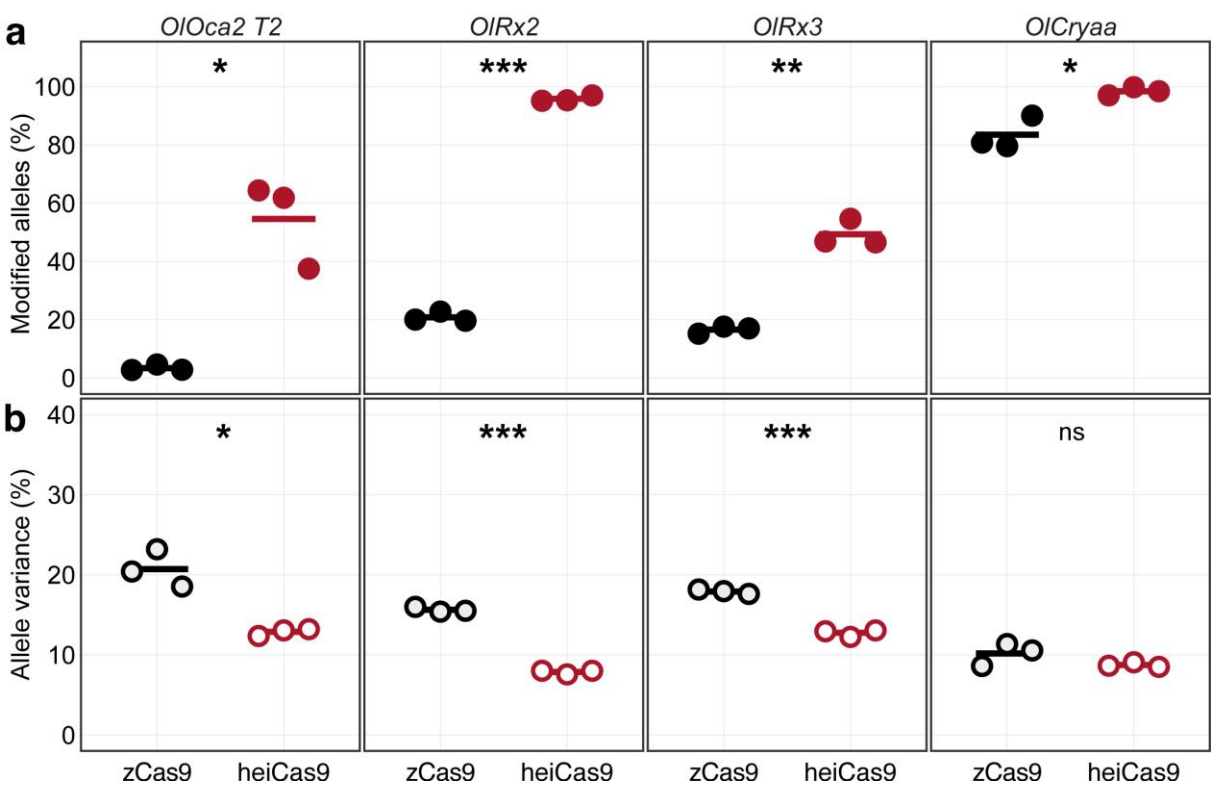


Figure 2 Increased knock-out activity and reduced allele variance in heiCas9 crispants.

Multiplexed injections with 15 ng/μl mRNA of zCas9 or heiCas9 (red) mRNA and 12.5 ng/μl per sgRNA targeting exonic sequences in *oculocutaneous albinism type 2* (*oca2*; *OIOca2 T2*), the start-codons of the *retina-specific homeobox transcription factor 2* (*rx2*; *OIRx2*) and of the *alpha a crystallin* (*cryaa*; *OICryaa*), as well as an intronic sequence in *rx3* (*OIRx3*). Illumina sequencing performed on three biological replicates (8 embryos each) per targeted locus. (a) Increased knock-out efficiency in heiCas9 crispants as shown by proportion of modified over all Illumina sequencing reads per replicate and locus. (b) Reduced allele variance in heiCas9 crispants as shown by abundance of specific allele divided by all modified alleles per replicate and locus.

Bold line, mean values of zCas9 (black) and heiCas9 (red). Total aligned Illumina-reads analyzed: *OIOca2*: zCas9 = 194,931, heiCas9 = 180,222; *OIRx2*: zCas9 = 224,146, heiCas9 = 269,103; *OIRx3*: zCas9 = 195,248, heiCas9 = 175,044; *OICryaa*: zCas9 = 209,573, heiCas9 = 200,448. Statistical analysis performed in R, Student's t-test.

While the early onset of action is required for uniform editing in developing organisms, cell culture approaches demand efficient translocation of the sgRNA/Cas9 complex in a large number of cells. To validate the range of action on the one hand and to address the relevance of the hei-tag in a mammalian setting, we expanded the scope of the analysis to mammalian cell culture. We focused on mRNA-based assays and compared the activity of heiCas9 to state-of-the-art Cas9 variants, i.e. the commercially available *GeneArt® CRISPR nuclease* as well as a mammalian codon optimized Cas9 (*JDS246-Cas9*, Addgene #43861) in mouse SW10 cells. We assessed the respective genome editing efficiencies by independent and complementary tools, the Tracking of Indels by Decomposition (TIDE) analysis (Brinkman et al., 2014) as well as by Inference of CRISPR Editing (ICE) (Hsiau et al., 2019). Both approaches decompose the mixed Sanger reads of PCR products spanning the CRISPR target site and compute an efficiency score as well as the distribution of expected indels. To target the murine *Periaxin* (*Prx*) locus, mouse SW10 cells were co-transfected with *MmPrx* crRNA / ATTO-550-linked tracrRNA and the mRNAs of either *JDS246-Cas9*, *GeneArt® CRISPR nuclease* or *heiCas9*. The *Prx* locus was PCR amplified and sequenced. Similar to targeting *in organismo*, *heiCas9* also exhibited the highest genome editing efficiency when compared to *JDS246-Cas9* (TIDE: 123.6%, ICE: 113%) and *GeneArt® CRISPR nuclease* (TIDE: 123.1%, ICE: 111%) in mammalian cell culture (Fig. 3, Figure 3-figure supplement 1, $R^2 > 0.9$ (TIDE) and > 0.9 (ICE) for all mRNAs tested). Notably, the KO-score efficiencies (ICE) amounted to 173% compared to *JDS246-Cas9* and to 167% compared to *GeneArt® CRISPR nuclease*, indicating higher abundance of frameshifts (Hsiau et al., 2019) at this genomic locus.

Remarkably, *heiCas9* transfected cells showed a highly increased number of mutant alleles with an increased abundance of a 26 nt deletion when compared to *GeneArt® CRISPR nuclease* and *JDS246-Cas9* (Figure 3-figure supplement 1).

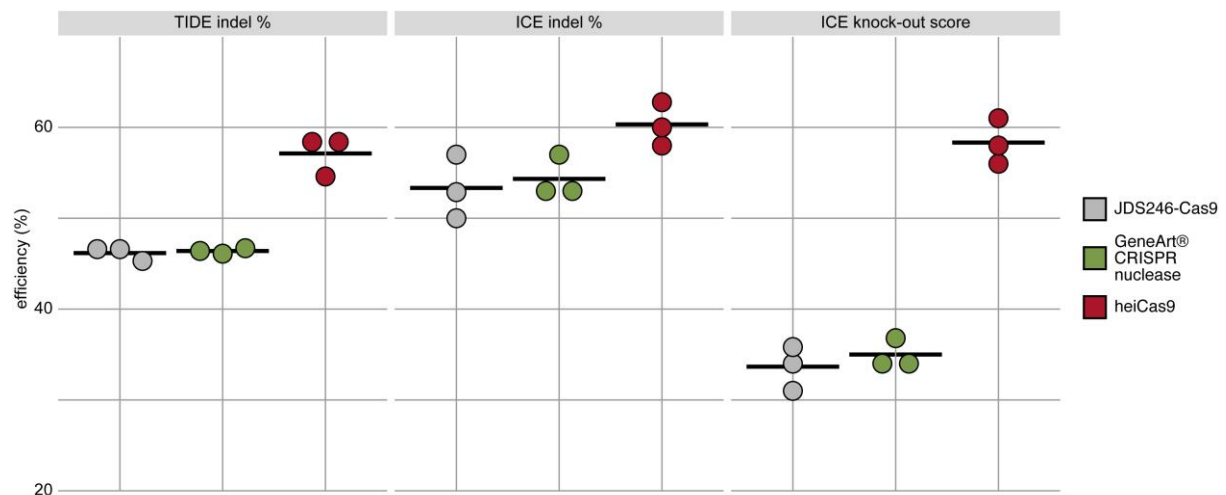


Figure 3: heiCas9 consistently exhibits high genome editing efficiency in mammalian cells. Mouse SW10 cells were co-transfected with *MmPrx* crRNA and mRNAs of *JDS246-Cas9*, *GeneArt® CRISPR nuclease* and *heiCas9* respectively. Genome editing efficiency was assessed by TIDE and ICE tools. ICE knock-out score represents proportion of indels that indicate a frameshift or ≥ 21 bp deletion. Data points represent three biological replicates, black line indicates respective mean: TIDE indel %: *JDS246-Cas9* = 46.2; *GeneArt® CRISPR nuclease* = 46.4, *heiCas9* = 57.1; ICE indel %: *JDS246-Cas9* = 53.3; *GeneArt® CRISPR nuclease* = 54.3, *heiCas9* = 60.3; ICE knock-out score %: *JDS246-Cas9* = 33.7; *GeneArt® CRISPR nuclease* = 35.0, *heiCas9* = 58.3. $R^2 > 0.9$ (TIDE) and > 0.9 (ICE) for all mRNAs tested. For representative indel spectrum for each mRNA see Figure 3-figure supplement 1.

Given the observed boosting of Cas9 activity by the simple addition of the hei-tag, we next tested if the hei-tag also improves further Cas9-based techniques. Base-editing is an increasingly applied method with a potential for therapeutics (Antoniou et al., 2021). Base editors (BEs) are composed of a modified Cas9 that only nicks one DNA strand and does not introduce a double-strand break (Cas9 nickase or Cas9n) and a nucleotide deaminase for precisely targeted nucleotide editing (Anzalone et al., 2020). To increase the efficiency of BEs, several iterative rounds of optimization of the employed de-aminases and linkers have been undertaken, yielding optimal performance with the newest variants (Carrington et al., 2020; Cornean et al., 2021; Rosello et al., 2021; Zhao et al., 2020). To investigate, if the addition of the hei-tag

provides an easy and straight forward alternative route for increasing the activity of a nuclear protein of interest, we selected a cytosine to thymine (C-to-T) base editor version with intermediate efficiency (BE4-Gam (Komor et al., 2017)) to introduce non-sense or severe miss-sense mutations into the pigmentation gene *Oca2*. We employed our tool ACEofBASEs (Cornean et al., 2021) to design and evaluate sgRNA target sites that introduce non-synonymous codon mutations and/or premature STOP codons upon editing of the respective open reading frame (ORF). We compared three different sgRNAs (*O/Oca2 T1*, *T3* and *T4*) employing the original BE4-Gam and the hei-tag fused variant (heiBE4-Gam). In the *oca2* ORF, the transition of cytosines 766, 922 and 997 to thymine all convert the respective codon to a pre-mature STOP (*O/Oca2 T3*: C766T, leading to Q256*; *O/Oca2 T4*: C922T, leading to Q308*; *O/Oca2 T1*: C995-997T, leading to T332I and Q333*). Again, the loss of pigmentation was used as proxy for bi-allelic targeting efficiency following medaka one-cell stage injections with either one of the three sgRNAs (*O/Oca2 T1*, *T3* or *T4*, 30 ng/μl) as well as 150 ng/μl mRNA of either *BE4-Gam* or *heiBE4-Gam*. Screening and analysis was performed at 4.5 dpf as described above. For each sgRNA employed, heiBE4-Gam resulted in more pronounced loss of pigmentation in comparison to BE4-Gam (Fig. 4a; control median = 0.0; medians BE4-Gam versus heiBE4-Gam: *O/Oca2 T1*, 0.6 vs. 28.0, $p = 1.737e-20$; *O/Oca2 T3*, 0.0 vs. 0.8, $p = 0.0471$; *O/Oca2 T4*, 93.8 vs. 170.1, $p = 5.215e-12$). Quantification of Sanger sequencing reads confirmed an increase of all C-to-T transitions at the *O/Oca2 T1* target site when heiBE4-Gam was used ($74.1\% \pm 8.9\%$ for heiBE4-Gam vs $44.2\% \pm 6.8\%$ for BE4-Gam; Figure 4-figure supplement 1, three replicates containing 5 randomly picked embryos each). In particular the C997T transition introducing a pre-mature STOP codon was increased 1.7-fold (i.e. 68% in heiBE4-Gam vs. 41% in BE4-Gam) in case of *heiBE4-Gam* (Fig. 4b, c).

In conclusion, using the hei-tag to extend the ORFs of a mammalian codon optimized *SpCas9* or a C-to-T base editor (BE4-Gam) severely enhanced the respective genome targeting efficiency.

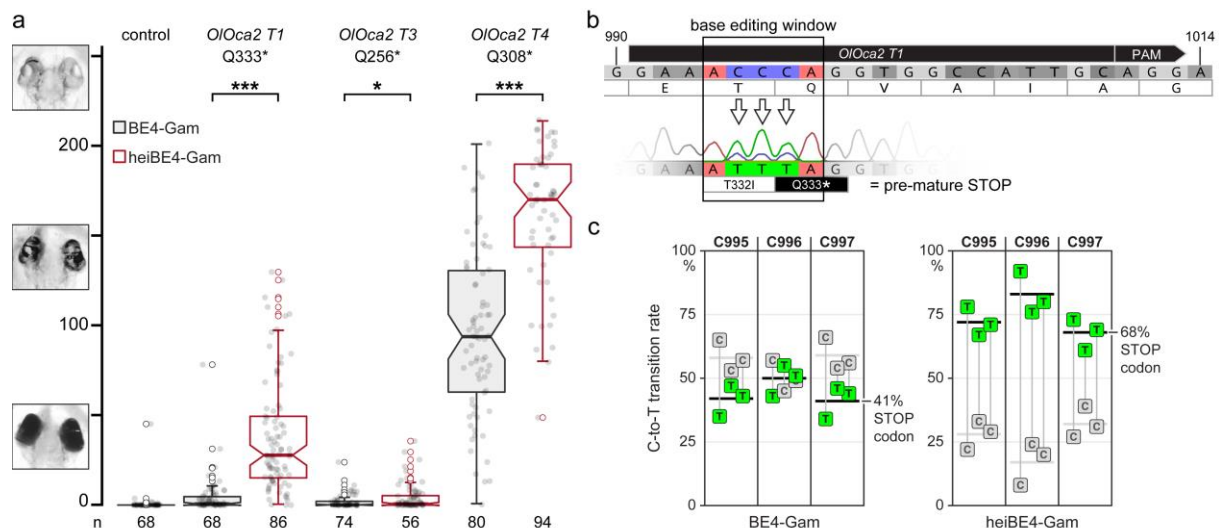


Figure 4: heiBE4-Gam mediates highly efficient cytosine to thymine transitions in medaka embryos. Phenotypic range and quantification of heiBE4-Gam mediated cytosine to thymine transitions in medaka embryos. (a) Categories of typically observed loss-of-pigmentation phenotypes in *oca2* editants. The observed pigmentation phenotypes range from (almost) unpigmented eyes, i.e. a very strong knock-out (top panel) over intermediate (central panel) to no loss-of-pigmentation (bottom panel). Quantification of phenotype resulting from injections with either *BE4-Gam* or *heiBE4-Gam* (red) mRNA and *OIOca2 T1*, *T3*, or *T4* sgRNAs. Note: dramatic increase of bi-allelic knock-out rate when using *heiBE4-Gam*. n, number of eyes analyzed. Control median = 0.0; medians *BE4-Gam* versus *heiBE4-Gam*: *OIOca2 T1*, 0.6 vs. 28.0, $p = 1.737$; *OIOca2 T3*, 0.0 vs. 0.8, $p = 0.0471$; *OIOca2 T4*, 93.8 vs. 170.1, $p = 5.215e-12$. Bold lines, median values. Statistical analysis performed in R, pairwise Wilcoxon rank sum test. (b) Schematic representation of base editing window in *OIOca2 T1* target site (PAM, protospacer adjacent motif). C-to-T transition of C995 and C996 edits the threonine (T) codon to isoleucine (I) (T332I); C997T creates a pre-mature STOP codon (Q333*). Nucleotide positions refer to the *oca2* open reading frame. (c) Quantification of Sanger sequencing reads at nucleotides C995, C996, C997 inside the base editing window of three injected embryo pools (five embryos each) reveals overall dramatic increase of C-to-T base transition when using *heiBE4-Gam*. Note 1.7-fold increase of C997T transition, i.e. efficient introduction of a pre-mature STOP codon. Mean values indicated by bold horizontal lines, cf. Figure 4-figure supplement 1.

Discussion

While the use of the optimized NLS in the hei-tag explains the earlier and better performance of the hei-tagged versions of Cas9 and base editors in developing organisms, the impact of the specific topology of domains contained in the hei-tag, remains elusive. It is speculated that the addition of certain peptide tags influences the efficacy and specificity of the fused protein of interest, due to their different isoelectric points and charge distributions (Zhang et al., 2014). Interestingly, our permutation screen demonstrated that although comprising the exact same peptides (for instance, compare MFO-Cas9-O (heiCas9) vs. OMF-Cas9-O and MSF-Cas9-S vs. SMF-Cas9-S in Figure 1-figure supplement 1), position of the particular tags relative to each other conveyed different genome editing efficiencies.

The hei-tag renders the resulting heiCas9 into a highly efficient endonuclease with broad applicability overcoming the limitations of current *SpCas9* variants by dramatically increasing the efficiency of targeted genome editing *in organismo*, as demonstrated in two evolutionarily distant fish models, as well as in mouse cell culture. In those systems heiCas9 introduces a lower number of mutant alleles, important for testing specific hypotheses or introducing site specific modifications by homology directed repair (Gutierrez-Triana et al., 2018). Conversely, Cas9 variants without the hei-tag are better suited for targeted screening approaches since they introduce a large number of mutant alleles. heiCas9 markedly increased the (bi-allelic) targeting rate alongside a decrease in allele variance, indicating a high targeting efficiency already at the earliest stages of development. Precedentially such early targeting in developing organisms was most of all reported using RNPs (Kroll et al., 2021; Wu et al., 2018), yet mRNA injection of heiCas9 is fully comparable to these protein approaches. The benefits of using mRNA over protein are apparent: new Cas9 variants can easily be generated and produced cost-efficiently by highly reproducible *in vitro* transcription, a standard method in molecular biology labs.

In light of the ever expanding CRISPR tool kit, the addition of the hei-tag provides the means to boost current specialized and future variants, as the simple addition of the hei-tag sequence also potentiated the activity of a cytosine base editor, with heiBE4-Gam resulting in an overall increase of about 30% of C-to-T transition rates (Fig. 4 and Figure 4-figure supplement 1). Taken together, the boosting activity of the hei-tag is neither limited by the species nor the approach, making it a powerful tweak to

swiftly upgrade any specifically adapted Cas-based genome editing approach (Anzalone et al., 2020).

Materials and Methods

Fish maintenance

Zebrafish (*Danio rerio*) and medaka (*Oryzias latipes*) fish were bred and maintained as previously described (Koster et al., 1997; Westerfield, 2000). The animal strains used in the present study were zebrafish AB/back and medaka Cab. All experimental procedures were performed according to the guidelines of the German animal welfare law and approved by the local government (Tierschutzgesetz §11, Abs. 1, Nr. 1, husbandry permit number 35–9185.64/BH Wittbrodt).

Cloning of Cas9 variants

The mammalian codon-optimized (Geneious 8.1.9, <https://www.geneious.com>) Cas9 sequence was gene-synthesized (GeneArt, ThermoFisher Scientific) as template for cloning the permuted peptide-tag Cas9 fusions (Supplemental File 2) using primers (Table 1) containing the sequences coding for a myc-tag (EQKLISEEDL), flexible or internal linkers and a SV40 (PKKKRKV) or optimized oNLS (PPPKRPRLD) (Inoue et al., 2016) (Figure 1-figure supplement 1). Cloning into the pCS2+ plasmid (Rupp et al., 1994) (multiple cloning site extended for *AgeI* site downstream of *BamHI* site) was performed using *AgeI* and *XbaI* restriction sites included in the 5' region of the forward or reverse primers, respectively. See Supplemental File 1 for full sequence of *heiCas9*. For consistent mRNA synthesis, the published *myc-Cas9* (Zhang et al., 2014) (MSI-Cas9-XI) was re-established with the pX330-U6-Chimeric_BB-CBh-hSpCas9 vector as template, primer based exchange of the N-terminal FLAG tag with the myc-tag sequence and brought into pCS2+ (Rupp et al., 1994) using *AgeI* and *XbaI* restriction sites included in the 5' region of the respective primers as well. pX330-U6-Chimeric_BB-CBh-hSpCas9 was a gift from Feng Zhang (Addgene plasmid #42230) (Cong et al., 2013).

Table 1 Primer sequences used for Cas9 variant cloning.

Restriction enzyme sites used for cloning are indicated in italics (*AgeI* in the forward primer, *XbaI* in the reverse primer), underscored sequence, binding to Cas9 ORF. F, flexible linker; I, internal linker; M, cMyc-tag; O, optimized NLS (Inoue et al., 2016); S, SV40 NLS (Kalderon et al., 1984); XI, bipartite *Xenopus laevis* nucleoplasmin NLS (Dingwall et al., 1988). For instance to establish the heiCas9 ORF, primers MFO-Cas9_fwd and Cas9-O_rev were used.

primer name	primer sequences in 5'-3'
MFO-Cas9_fwd	AATTTACCGGTTTACCATGGAGCAGAAGCTGATCAGCGAGGAGGACCTGGG AGGAAGCGGACCACCTCCCAAGAGGCCAGGCTGGACCTCGAGGATAAAAA <u>GTATTCTATTGGTTTAG</u>
MIS-Cas9_fwd	AATTTACCGGTTTACCATGGAGCAGAAGCTGATCAGCGAGGAGGACCTGGG TATCCACGGAGTCCCAGCAGCCGCTCCAAAGAAGAAGCGTAAGGTAGATAA <u>AAAGTATTCTATTGGTTTAG</u>
MSF-Cas9_fwd	AATTTACCGGTTTACCATGGAGCAGAAGCTGATCAGCGAGGAGGACCTGAT GGCTCCAAAGAAGAAGCGTAAGGTAGGAGGAAGCGGAGATAAAAAGTATTC <u>TATTGGTTTAG</u>
OMF-Cas9_fwd	AATTTACCGGTTTACCATGCCACCTCCCAAGAGGCCAGGCTGGACCTCGA GGAGCAGAAGCTGATCAGCGAGGAGGACCTGGGAGGAAGCGGAGATAAAAA <u>GTATTCTATTGGTTTAG</u>
SMF-Cas9_fwd	AATTTACCGGTTTACCATGGCTCCAAAGAAGAAGCGTAAGGTACTCGAGGA GCAGAAGCTGATCAGCGAGGAGGACCTGGGAGGAAGCGGAGATAAAAAGTA <u>TTCTATTGGTTTAG</u>
Cas9-O_rev	AATTTCTAGATTAGTCCAGCCTGGGCCTCTTGGGAGGAGGGGATCCGTC <u>CCCCCAAGCTGTGAC</u>
Cas9-S_rev	AATTTCTAGATTAATCTACCTTACGCTTCTTCTTTGGAGCAGCGGATCCG <u>TCACCCCCAAGCTGTGACA</u>
myc-Cas9_fwd	AATTTACCGGTCAAACATGGAGCAGAAGCTGATCAGCGAGGAGGACCTGAT GGCCCCAAAGAAGAAGCGGAAGGTC
myc-Cas9_rev	AATTTCTAGATTACTTTTTTCTTTTTTGCCTGGCCGGC

Cloning of BE4-Gam and heiBE4-Gam

BE4-Gam was subcloned from pCMV(BE4-Gam) (Addgene plasmid #100806, a gift from David Liu) (Komor et al., 2017) in a two step-process, first into pJET1.2 (Thermo Scientific), then into pGGEV4 (Kirchmaier et al., 2013) (Addgene plasmid #49284), by *Bam*HI, *Eco*RV and *Kpn*I restriction sites to create pGGEV4(BE4-Gam). heiBE4-Gam was assembled into pCS2+ (Rupp et al., 1994) by NEBuilder® HiFi DNA Assembly (NEB) with four inserts using Q5 polymerase PCR products (NEB): pCS2+ backbone, hei-tag fragment, Gam Mu-APOBEC1-Cas9n fragment, Cas9n-UGI fragment, 2xUGI-oNLS (see Table 2 for primers used).

Table 2 Primer sequences used for BE4-Gam and heiBE4-Gam cloning.

primer name	primer sequences in 5'-3'
pCS2+_backbone_fwd	GCCTCTAGAACTATAGTGAGTCG
pCS2+_backbone_rev	ATGGGATCCTGCAAAAAGAACAAG
hei-tag_fragment_fwd	CTTGTTCTTTTTGCAGGATCCCATTACCATGGAGCAGAAGCTG
hei-tag_fragment_rev	GCTGGTTTAGCCTCGAGGTCCAGCCTGG
Gam_Mu-APOBEC1-Cas9n_fragment_fwd	GACCTCGAGGCTAAACCAGCAAAACGTATCAAG
Gam_Mu-APOBEC1-Cas9n_fragment_rev	CTAGGGCCTTGAGAAGTGTC
Cas9n-UGI_fragment_fwd	GACACTTCTCAAGGCCCTAG
Cas9n-UGI_fragment_rev	CAGAGTCACCCCCAAGCTG
2xUGI-oNLS_fwd	CAGCTTGGGGGTGACTCTG
2xUGI-oNLS_rev	CGACTCACTATAGTTCTAGAGGCTTAGTCCAGCCTGGGCCTCTTGGGAGGG GGAGAACCACCAGAGAGC

sgRNA design

All sgRNAs for medaka (*O/Oca2*, *Rx2*, *Rx3*, *Cryaa*) and zebrafish (*DrOca2*) were designed using the CCTop target predictor with standard parameters (Stemmer et al., 2015). The sgRNAs used for base editing (*O/Oca2* T1, T3, T4) were designed or evaluated using ACEofBASEs (Cornean et al., 2021) and selected for introducing a pre-mature STOP codon. The following target sites were used [PAM in brackets]:

<i>O/Oca2</i> T1 (GAAACCCAGGTGGCCATTGC[AGG]),	<i>O/Oca2</i> T2
(TTGCAGGAATCATTCTGTGT[GGG]),	<i>O/Oca2</i> T3
(GATCCAAGTGGAGCAGACTG[AGG]),	<i>O/Oca2</i> T4
(CACAATCCAGGCCTTCCTGC[AGG])	<i>DrOca2</i> T1
(GTACAGCGACTGGTTAGTCA[TGG]),	<i>DrOca2</i> T2

(TAAGCACGTAGACTCCTGCC[AGG]), *Rx2* (GCATTTGTCAATGGATACCC[TGG]),
Cryaa (GGGAGAAGTGCTTGACATCC[AGG]), *Rx3*
(AGCAGAGCGCGCAAAGAACC[AGG]). *OIOca2 T1*, *OIOca2 T2* and *DrOca2 T1*
were the same as in (Hammouda et al., 2019), *OIOca2 T3* was the same as in
(Lischik et al., 2019) (OCA2_4), *OIRx2* and *OICryaa* are from (Stemmer et al., 2015),
and *OIRx3* is the same used in (Zilova et al., 2021). Cloning of sgRNA templates was
performed as described (Stemmer et al., 2015). Plasmid DR274 was a gift from Keith
Joung (Addgene plasmid #42250) (Hwang et al., 2013).

***In vitro* transcription of mRNA**

pCS2+ constructs in this work were linearized using NotI-HF (NEB) except for zCas9
– linearized with HpaI (NEB). The pGGEV4(BE4-Gam) was linearized using SpeI-HF
(NEB). mRNA was transcribed *in vitro* using the mMESSAGE mMACHINE SP6
transcription kit (ThermoFisher Scientific, AM1340). pCS2-nCas9n (zCas9) was a gift
from Wenbiao Chen (Addgene plasmid #47929) (Jao et al., 2013). The *JDS246-Cas9*
was linearized with MssI FD (ThermoFisher Scientific) and transcribed *in vitro* using
the mMESSAGE mMACHINE T7 Ultra Transcription Kit (ThermoFisher Scientific,
AM1345). *JDS246-Cas9* was a gift from Keith Joung (Addgene plasmid #43861).
sgRNAs were synthesized using the MEGAscript T7 transcription kit (Thermo Fisher
Scientific, AM1334) after plasmid digestion with DraI FD (ThermoFisher Scientific).

Microinjection

All microinjections were performed at the one cell stage. At standard concentrations,
zebrafish and medaka zygotes were injected with 150 ng/μl Cas9 (variant) mRNA,
Oca2 sgRNAs at 30 ng/μl and *H2B-GFP* mRNA at 10 ng/μl as injection tracer. The
multiplexing injection mixes contained 12.5 ng/μl per sgRNA (*OIOca2 T2*, *Rx2*, *Rx3*,
Cryaa) and 15 ng/μl of either zCas9 or *heiCas9* mRNA as well as 20 ng/μl mCherry
mRNA as injection tracer. For the protein injections, 24 μM RNP mix (Kroll et al.,
2021) was assembled in Cas9 buffer (20 mM Tris-HCl, 600 mM KCl, 20 %
glycerol;(Wu et al., 2018)) by mixing 61 μM Alt-R S.p. Cas9 Nuclease V3 (IDT) with
5710 ng of each sgRNA *OIOca2 T1* and *T2*. 285.6 ng GFP mRNA were added as
injection tracer. The mix was incubated for 5 min at 37 °C and stored on ice until
further dilution and injection. To obtain 5 μM RNPs (Wu et al., 2018), the 24 μM RNP

mix was diluted in a 1:1 mixture of Cas9 buffer and nuclease free water. 5 μ M RNP solution was further diluted in a 1:1 mixture of Cas9 buffer and nuclease free water to obtain 1.765 μ M RNPs.

For the base editing experiments, medaka zygotes were injected with *BE4-Gam* or *heiBE4-Gam* mRNA at 150 ng/ μ l, respective *Oca2* sgRNA at 30 ng/ μ l and *GFP* mRNA at 20 ng/ μ l as injection tracer. All injected embryos were maintained at 28°C in zebrafish medium (Westerfield, 2000) or medaka embryo rearing medium (ERM, 17 mM NaCl; 40 mM KCl; 0.27 mM $\text{CaCl}_2 \cdot 2\text{H}_2\text{O}$; 0.66 mM $\text{MgSO}_4 \cdot 7\text{H}_2\text{O}$, 17 mM Hepes).

Embryos were screened for *GFP* or *mCherry* expression four to seven hours or one day after injection using a Nikon SMZ18 stereomicroscope, and uninjected specimens were discarded.

Image acquisition and phenotype analysis

Medaka 4.5 days post fertilization (dpf) embryos (Iwamatsu, 2004) and zebrafish 2.5 dpf (Kimmel et al., 1995) embryos were fixed with 4% paraformaldehyde in 2xPTW (2x PBS pH 7.3, 0.1% Tween 20). Images of medaka embryos were acquired with the high content screening ACQUIFER Imaging Machine (DITABIS AG, Pforzheim, Germany). Images of zebrafish embryos were acquired with a Nikon digital sight DS-Ri1 camera mounted onto a Nikon Microscope SMZ18 and the Nikon Software NIS-Elements F version 4.0. Only properly developed embryos were included in the following analysis. Image analysis was performed with Fiji (Schindelin et al., 2012), i.e. mean grey values were obtained on minimum intensity projections and locally thresholded (Phansalkar algorithm with parameters $r = 20$, $p = 0.4$, $k = 0.4$) pictures and elliptical selections for each individual eye. The mean grey value per eye was used for the boxplot and statistical analysis (pairwise comparisons using Wilcoxon rank sum test, Bonferroni corrected) in RStudio (Team, 2020).

Targeted amplicon sequencing via Illumina

The multiplex approach was genotyped on DNA extractions of pools with each replicate containing 8 randomly picked crispants per *zCas9* or *heiCas9* injection or 6 control specimens. DNA was prepared by grinding and lysis in DNA extraction buffer (0.4 M Tris/HCl pH 8.0, 0.15 M NaCl, 0.1% SDS, 5 mM EDTA pH 8.0, 1 mg/ml proteinase K) at 60°C overnight. Proteinase K was inactivated at 95°C for 10 min and

the solution was diluted 1:2 with nuclease free water. For each DNA extraction, small libraries were constructed by PCR amplifying the four regions of interest (295-362 bp, *OIOca2*, *rx2*, *rx3*, *cryaa*) using locus-specific primers with 5' partial illumina adapter sequences (Table 3) and Q5® Hot Start High-Fidelity DNA Polymerase (New England Biolabs). PCR products were run on a 1% agarose gel, respective bands were excised and cleaned up using the Monarch® DNA Gel Extraction Kit (New England Biolabs). PCR products from the same genomic DNA source were pooled to equimolarity at 25 ng/µl and submitted to GeneWiz (Azenta Life Sciences) for sequencing (Amplicon-EZ: Illumina MiSeq, 2x250 bp sequencing, paired-end) obtaining between 48018 and 96899 reads per sample.

Table 3 Locus specific primers with 5' partial illumina adapter sequences
Locus specific primers with Illumina adapter sequence underscored

primer name	primer sequences in 5'-3'
oca2_F	<u>ACACTCTTTCCCTACACGACGCTCTTCCGATCT</u> CGTTAGAGTGGTATGGAGAACTGT
oca2_R	GACTGGAGTTCAGACGTGTGCTCTTCCGATCT <u>ATGGTCCTCACATCAGCAGC</u>
cryaa_F	<u>ACACTCTTTCCCTACACGACGCTCTTCCGATCT</u> CGCCATTTGCTTGTGTGTCA
cryaa_R	GACTGGAGTTCAGACGTGTGCTCTTCCGATCTAGTCTAGGAGGATGGGGCAG
rx2_F	<u>ACACTCTTTCCCTACACGACGCTCTTCCGATCT</u> AGAGGCACAAGAACTATTTGTTGATC
rx2_R	GACTGGAGTTCAGACGTGTGCTCTTCCGATCTAGGGCTCCGTTAACTTTGGG
rx3_F	<u>ACACTCTTTCCCTACACGACGCTCTTCCGATCT</u> ATGCAAACCAAGAAAGCGCC
rx3_R	GACTGGAGTTCAGACGTGTGCTCTTCCGATCTTGGGATTTCTCAAAGGCCCG

Analysis and plotting of next-generation sequencing data

Amplicon sequencing data was analyzed with CRISPResso2 v.2.1.2 (Clement et al., 2019) using the default -n nhej parameters. Demultiplexing was achieved by mapping to the four different wild-type loci, respectively. Downstream analysis was conducted using R v.3.6.3 in R studio (Team, 2020) (package: ggplot2 (Wickham 2016)), with data sourced from 'CRISPResso_quantification_of_editing_frequency.txt' output table. To determine the average read count per modified allele the 'Alleles_frequency_table.txt' output table was used. The number of modified alleles was determined by filtering > 'Read_status' > modified. Average read count per modified allele = modified reads/N modified alleles.

Genotyping of editants

Genotyping was performed on DNA extractions (see above) of three replicates containing five randomly picked editants each of BE4-Gam and heiBE4-Gam injections. Q5 polymerase (NEB), primers fwd 5'-GTTAAAACAGTTTCTTAAAAAGAACAGGA-3' and rev 5'-AGCAGAAGAAATGACTCAACATTTTG-3' (annealing at 62°C) were used on 1 µl of diluted DNA sample according to the manufacturer's instructions with 30x PCR cycles. PCR products were analyzed on a 1% agarose gel, bands excised, DNA-extraction performed using innuPREP Gel Extraction Kit (Analytik Jena) according to the manufacturer's instructions and subjected to Sanger sequencing (see below).

Cell lines

Mouse SW10 cells (ATCC, CRL-2766TM, Lot number 4117643) were cultured in DMEM (Gibco) supplemented with 1 g/ml glucose containing 10% FCS (Sigma), 1% penicillin (10,000 units/ml; Gibco) and 1% streptomycin (10 mg/ml; Gibco) and maintained at 33°C and 5% CO₂ and regularly tested negative for mycoplasma infections. Cells were passaged at 80-90% confluency. 24 h before transfection cells were seeded in a density of 85,000 cells per 12-well.

CRISPR Transfection

crRNA targeting exon 6 (TCGTATCCAGACACCGTCCC[GGG], PAM in brackets) of the mouse *Periaxin* (*MmPrx*) gene was selected from the IDT (crRNA XT) predesign crRNA database. crRNA (50 µM) and Alt-R® CRISPR-Cas9 tracrRNA, ATTO-550 (50 µM; IDT, 1075927) were diluted in nuclease-free duplex buffer (IDT) to a final concentration of 1 µM and incubated at 95°C for 5 minutes. 1 µg of the corresponding Cas9 mRNA (*GeneArt® CRISPR nuclease* Invitrogen, A29378; *JDS246-Cas9* or *heiCas9*) and 15 µl of tracrRNA+crRNA Mix (1 µM) were diluted in 34 µl Opti-MEM I (Gibco) and mixed with 3 µl Lipofectamine RNAiMAX (ThermoFisher) diluted in 47 µl Opti-MEM I. The tracrRNA+crRNA lipofection mix was incubated for 20 min at RT. Cell culture medium was exchanged with 900 µl Opti-MEM I and the tracrRNA+crRNA lipofection mix was added dropwise to the well. After 48h, genomic DNA was extracted using the DNeasy Blood and Tissue Kit (Qiagen, 69506) following the manufacturer's protocol. Q5-PCR was carried out using primers flanking the targeted exon 6 (fwd 5'-GAGACACTCACTCCAGACCC-3';

rev 5'-ACTCAGTAACCCAACAGCCA-3') and 30 cycles. PCR amplicons were purified using the Monarch DNA Gel Extraction Kit (NEB, T1020S) and subjected to sequencing.

Sanger sequencing

Sanger sequencing was performed by Eurofins Genomics using fwd 5'-GTTAAAACAGTTTCTTAAAAAGAACAGGA-3' to evaluate base editing at *OIOca2 T1* target site and using fwd 5'-GAGACACTCACTCCAGACCC-3' and rev 5'-ACTCAGTAACCCAACAGCCA-3' to evaluate genome editing of the *Prx* locus in SW10 cells. Quantification of base editing from Sanger sequencing reads was performed with EditR (Kluesner et al., 2018). Genome editing efficiency was assessed by sequence analysis using the TIDE web tool (Brinkman et al., 2014) and by ICE (Hsiau et al., 2019) using default parameters and indel size range up to 30 bp.

Data visualization

Data visualization and figure assembly was performed using Fiji (Schindelin et al., 2012), ggplot2 (Wickham, 2016) in RStudio (Team, 2020), Geneious Prime 2019.2.1 and Adobe® Illustrator® CS6.

Acknowledgements

This research was supported by grants of the European Research Council (ERC-SyG H2020, 81072), the Excellence Cluster '3D Matter Made to Order' (3DMM2O, EXC 2082/1 Wittbrodt C3) funded through the German Excellence Strategy via Deutsche Forschungsgemeinschaft (DFG), CRC873, project A3 and FOR2509 project 10 (WI 1824/9-1) to J.W. and CRC1118, project S03 to M.F.

A.C., B.W., R.M. and T.T.-S. are members/alumni of HBIGS, the Heidelberg Biosciences International Graduate School. B.W. was supported by the Deutsches Zentrum für Herz-Kreislauf-Forschung (DZHK B20-024 SE). We thank T. Kellner for sgRNA, base editor and Cas9 mRNA synthesis. We are thankful to M. Majewski, E. Leist and A. Saraceno for fish husbandry. We thank all members of the Wittbrodt lab for their critical, constructive feedback on the procedure and the manuscript.

Competing Interests statement

The authors declare the following competing interests: T.T., T.T.-S., J.A.G.T. and J.W. have a patent application pending (EP21166099.8) related to the findings described.

References

- Antoniou, P., Miccio, A., and Brusson, M. (2021). Base and Prime Editing Technologies for Blood Disorders. *Frontiers Genome Ed* 3, 618406.
- Anzalone, A.V., Koblan, L.W., and Liu, D.R. (2020). Genome editing with CRISPR–Cas nucleases, base editors, transposases and prime editors. *Nat Biotechnol* 38, 824–844.
- Brinkman, E.K., Chen, T., Amendola, M., and van Steensel, B. (2014). Easy quantitative assessment of genome editing by sequence trace decomposition. *Nucleic Acids Res* 42, e168–e168.
- Carrington, B., Weinstein, R.N., and Sood, R. (2020). BE4max and AncBE4max Are Efficient in Germline Conversion of C:G to T:A Base Pairs in Zebrafish. *Cells* 9, 1690.
- Clement, K., Rees, H., Canver, M.C., Gehrke, J.M., Farouni, R., Hsu, J.Y., Cole, M.A., Liu, D.R., Joung, J.K., Bauer, D.E., et al. (2019). CRISPResso2 provides accurate and rapid genome editing sequence analysis. *Nat Biotechnol* 37, 224–226.
- Cong, L., Ran, F.A., Cox, D., Lin, S., Barretto, R., Habib, N., Hsu, P.D., Wu, X., Jiang, W., Marraffini, L.A., et al. (2013). Multiplex Genome Engineering Using CRISPR/Cas Systems. *Science* 339, 819–823.
- Cornean, A., Gierten, J., Welz, B., Mateo, J.L., Thumberger, T., and Wittbrodt, J. (2021). Precise in vivo functional analysis of DNA variants with base editing using ACEofBASEs target prediction. *Biorxiv* 2021.07.26.453883.
- Dingwall, C., Robbins, J., Dilworth, S.M., Roberts, B., and Richardson, W.D. (1988). The nucleoplasmin nuclear location sequence is larger and more complex than that of SV-40 large T antigen. *J Cell Biology* 107, 841–849.
- Furutani-Seiki, M., and Wittbrodt, J. (2004). Medaka and zebrafish, an evolutionary twin study. *Mech Develop* 121, 629–637.
- Gutierrez-Triana, J.A., Tavhelidse, T., Thumberger, T., Thomas, I., Wittbrodt, B., Kellner, T., Anlas, K., Tsingos, E., and Wittbrodt, J. (2018). Efficient single-copy HDR by 5' modified long dsDNA donors. *Elife* 7, e39468.
- Hammouda, O.T., Böttger, F., Wittbrodt, J., and Thumberger, T. (2019). Swift Large-scale Examination of Directed Genome Editing. *Plos One* 14, e0213317.
- Hsiao, T., Conant, D., Rossi, N., Maures, T., Waite, K., Yang, J., Joshi, S., Kelso, R., Holden, K., Enzmann, B.L., et al. (2019). Inference of CRISPR Edits from Sanger Trace Data. *Biorxiv* 251082.
- Hwang, W.Y., Fu, Y., Reyon, D., Maeder, M.L., Tsai, S.Q., Sander, J.D., Peterson, R.T., Yeh, J.-R.J., and Joung, J.K. (2013). Efficient In Vivo Genome Editing Using RNA-Guided Nucleases. *Nat Biotechnol* 31, 227–229.
- Inoue, T., Iida, A., Maegawa, S., Sehara-Fujisawa, A., and Kinoshita, M. (2016). Generation of a transgenic medaka (*Oryzias latipes*) strain for visualization of nuclear dynamics in early developmental stages. *Dev Growth Differ* 58, 679–687.
- Iwamatsu, T. (2004). Stages of normal development in the medaka *Oryzias latipes*. *Mech Develop* 121, 605–618.

668 Jao, L.-E., Wente, S.R., and Chen, W. (2013). Efficient multiplex biallelic zebrafish genome editing
669 using a CRISPR nuclease system. *Proc National Acad Sci* 110, 13904–13909.

670 Kalderon, D., Roberts, B.L., Richardson, W.D., and Smith, A.E. (1984). A short amino acid sequence
671 able to specify nuclear location. *Cell* 39, 499–509.

672 Kimmel, C.B., Ballard, W.W., Kimmel, S.R., Ullmann, B., and Schilling, T.F. (1995). Stages of
673 embryonic development of the zebrafish. *Dev Dynam* 203, 253–310.

674 Kirchmaier, S., Lust, K., and Wittbrodt, J. (2013). Golden GATEway Cloning – A Combinatorial
675 Approach to Generate Fusion and Recombination Constructs. *Plos One* 8, e76117.

676 Kluesner, M.G., Nedveck, D.A., Lahr, W.S., Garbe, J.R., Abrahante, J.E., Webber, B.R., and
677 Moriarity, B.S. (2018). EditR: A Method to Quantify Base Editing from Sanger Sequencing. *Crispr J*
678 1, 239–250.

679 Komor, A.C., Zhao, K.T., Packer, M.S., Gaudelli, N.M., Waterbury, A.L., Koblan, L.W., Kim, Y.B.,
680 Badran, A.H., and Liu, D.R. (2017). Improved base excision repair inhibition and bacteriophage Mu
681 Gam protein yields C:G-to-T:A base editors with higher efficiency and product purity. *Sci Adv* 3,
682 eaao4774.

683 Koster, R., Stick, R., Loosli, F., and Wittbrodt, J. (1997). Medaka spalt acts as a target gene of
684 hedgehog signaling. *Dev Camb Engl* 124, 3147–3156.

685 Kroll, F., Powell, G.T., Ghosh, M., Gestri, G., Antinucci, P., Hearn, T.J., Tunbak, H., Lim, S., Dennis,
686 H.W., Fernandez, J.M., et al. (2021). A simple and effective F0 knockout method for rapid screening
687 of behaviour and other complex phenotypes. *Elife* 10, e59683.

688 Lischik, C.Q., Adelmann, L., and Wittbrodt, J. (2019). Enhanced in vivo-imaging in medaka by
689 optimized anaesthesia, fluorescent protein selection and removal of pigmentation. *Plos One* 14,
690 e0212956.

691 Liu, P., Liang, S.-Q., Zheng, C., Mintzer, E., Zhao, Y.G., Ponninselvan, K., Mir, A., Sontheimer,
692 E.J., Gao, G., Flotte, T.R., et al. (2021). Improved prime editors enable pathogenic allele correction
693 and cancer modelling in adult mice. *Nat Commun* 12, 2121.

694 Nidhi, S., Anand, U., Oleksak, P., Tripathi, P., Lal, J.A., Thomas, G., Kuca, K., and Tripathi, V.
695 (2021). Novel CRISPR–Cas Systems: An Updated Review of the Current Achievements,
696 Applications, and Future Research Perspectives. *Int J Mol Sci* 22, 3327.

697 Poon, I.K.H., and Jans, D.A. (2005). Regulation of Nuclear Transport: Central Role in Development
698 and Transformation? *Traffic* 6, 173–186.

699 Rosello, M., Vougny, J., Czarny, F., Mione, M.C., Concordet, J.-P., Albadri, S., and Bene, F.D.
700 (2021). Precise base editing for the in vivo study of developmental signaling and human pathologies in
701 zebrafish. *Elife* 10, e65552.

702 Rupp, R.A., Snider, L., and Weintraub, H. (1994). Xenopus embryos regulate the nuclear localization
703 of XMyoD. *Gene Dev* 8, 1311–1323.

704 Schindelin, J., Arganda-Carreras, I., Frise, E., Kaynig, V., Longair, M., Pietzsch, T., Preibisch, S.,
705 Rueden, C., Saalfeld, S., Schmid, B., et al. (2012). Fiji: an open-source platform for biological-image
706 analysis. *Nat Methods* 9, 676–682.

707 Stemmer, M., Thumberger, T., Keyer, M. del S., Wittbrodt, J., and Mateo, J.L. (2015). CCTop: An
708 Intuitive, Flexible and Reliable CRISPR/Cas9 Target Prediction Tool. *Plos One* *10*, e0124633.

709 Team, Rs. (2020). RStudio: Integrated Development Environment for R (Boston, MA).

710 Westerfield, M. (2000). The zebrafish book. A guide for the laboratory use of zebrafish (*Danio rerio*).
711 4th ed. (5274 University of Oregon, Eugene, OR 97403 USA: Univ. of Oregon Press, Eugene).

712 Wickham, H. (2016). ggplot2, Elegant Graphics for Data Analysis. R.

713 Wu, R.S., Lam, I.I., Clay, H., Duong, D.N., Deo, R.C., and Coughlin, S.R. (2018). A Rapid Method
714 for Directed Gene Knockout for Screening in G0 Zebrafish. *Dev Cell* *46*, 112-125.e4.

715 Zhang, J.-H., Pandey, M., Kahler, J.F., Loshakov, A., Harris, B., Dagur, P.K., Mo, Y.-Y., and
716 Simonds, W.F. (2014). Improving the specificity and efficacy of CRISPR/CAS9 and gRNA through
717 target specific DNA reporter. *J Biotechnol* *189*, 1–8.

718 Zhao, Y., Shang, D., Ying, R., Cheng, H., and Zhou, R. (2020). An optimized base editor with
719 efficient C-to-T base editing in zebrafish. *Bmc Biol* *18*, 190.

720 Zilova, L., Weinhardt, V., Tavhelidse, T., Schlagheck, C., Thumberger, T., and Wittbrodt, J. (2021).
721 Fish primary embryonic pluripotent cells assemble into retinal tissue mirroring in vivo early eye
722 development. *Elife* *10*, e66998.

723
724
725
726

Supplementary Figure and File Legends

Figure 1-figure supplement 1 identification of the hei-tag

Comparison of *OIOca2* knock-out efficiency (quantification of eye-pigmentation) using a permutation screen of peptide domains (NLSs, Myc-tag, amino acid linkers) flanking a mammalian codon optimized Cas9 enzyme. Injection mix contained 30 ng/μl *OIOca2* T1, T2, 150 ng/μl tagged Cas9 variant mRNA, 20 ng/μl GFP mRNA injection marker. Particular peptides and relative positions indicated by schematics. Constellation of peptides sorted by knock-out efficiency. The “hei-tag” myc-flexible-linker-oNLS-Cas9-oNLS (heiCas9) variant was identified being most efficient. JDS246-Cas9 (Addgene #43861), MSI-Cas9-XI (myc-Cas9) was cloned following Zhang et al., 2014 (Zhang et al., 2014).

Peptides used: FLAG, FLAG tag; F, flexible linker; I, internal linker; M, cMyc-tag; O, optimized NLS (Inoue et al., 2016); S, SV40 NLS (Kalderon et al., 1984); XI, bipartite *Xenopus laevis* nucleoplasmin NLS (Dingwall et al., 1988). For sequences see Supplementary Files 1-2.

0 = fully pigmented, 255 = completely unpigmented; n, number of eyes analyzed.

Figure 1-figure supplement 2 survival and abnormality rate of Cas9 mRNA and RNP injections

Percentage of dead, abnormal (e.g. delayed development or malformation) and properly developed injected embryos. Only properly developed embryos were included for analysis.

n, total number of injected embryos.

Figure 2-figure supplement 1 mode of editing of all modified alleles

Relative abundance of Illumina reads categorized by mode of editing among all modified alleles per replicate, locus (*OIOca2*, *OIRx2*, *OIRx3*, *OICryaa*) and Cas9 mRNA employed (zCas9, heiCas9). Categories: only insertions, only deletions, only substitutions, insertions and substitutions, deletions and substitutions.

n, total number of aligned modified illumina reads per replicate

Figure 3-figure supplement 1 Representative indel spectrum for each Cas9 mRNA used in the cell culture assay.

Indel spectrum diagram obtained from TIDE (red bargraphs) and ICE (blue bargraphs) analyses following *JDS246-Cas9* mRNA, *GeneArt® CRISPR nuclease* mRNA and *heiCas9* mRNA and *Prx* tracrRNA/crRNA transfections. Note decreased number of wildtype alleles (grey dashed line in TIDE analysis) in *heiCas9* transfected cells and increased abundance of 26 nt deletion (black arrowhead in ICE analysis).

Figure 4-figure supplement 1 Increased cytosine to thymine transition in medaka embryo pools injected with *heiBE4-Gam*.

(a) Schematic representation of base editing window in *OIOca2 T1* target site.
(b-c) Sanger sequencing quantifications (EditR (Kluesner et al., 2018)) of pools of five randomly picked embryos injected with sgRNA *OIOca2 T1* and either *BE4-Gam* (b) or *heiBE4-Gam* (c). Note: in *heiBE4-Gam* injections, for each cytosine, the C-to-T transition rate was higher than 60%, a level never observed in *BE4-Gam* injected embryos. C997T is highlighted with white frame.

Supplementary File 1

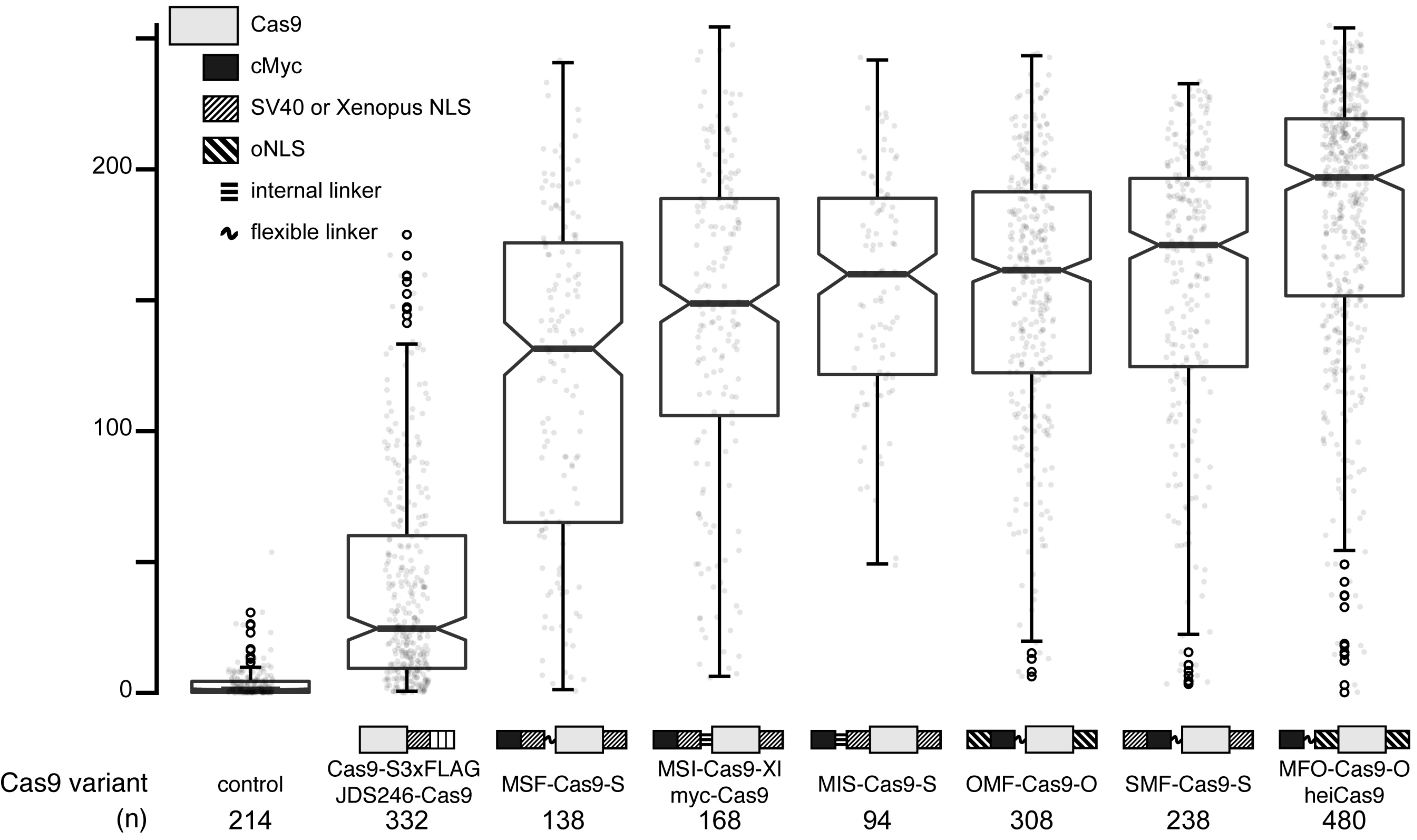
Nucleotide and translated amino acid sequence of *heiCas9*.

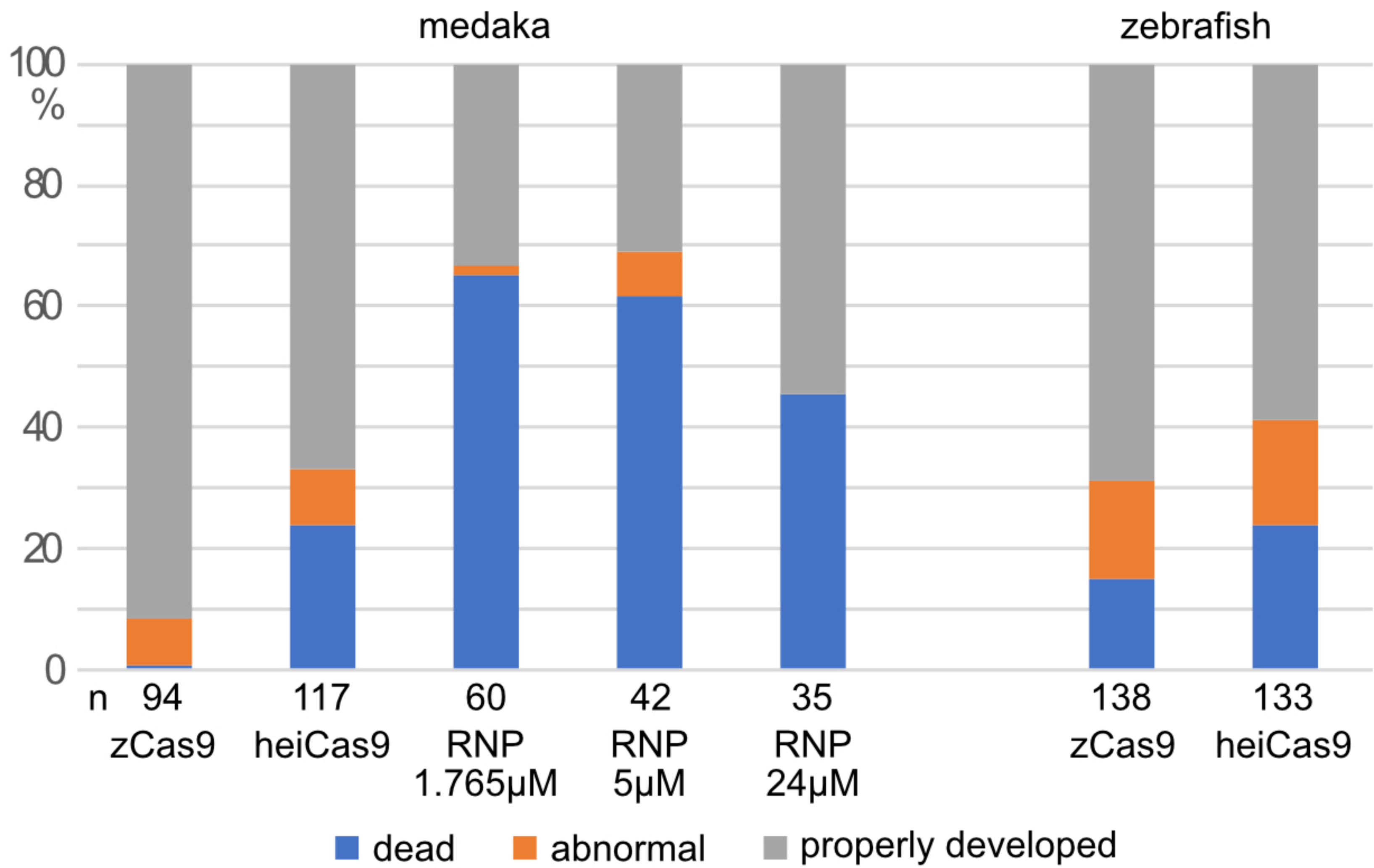
Supplementary File 2

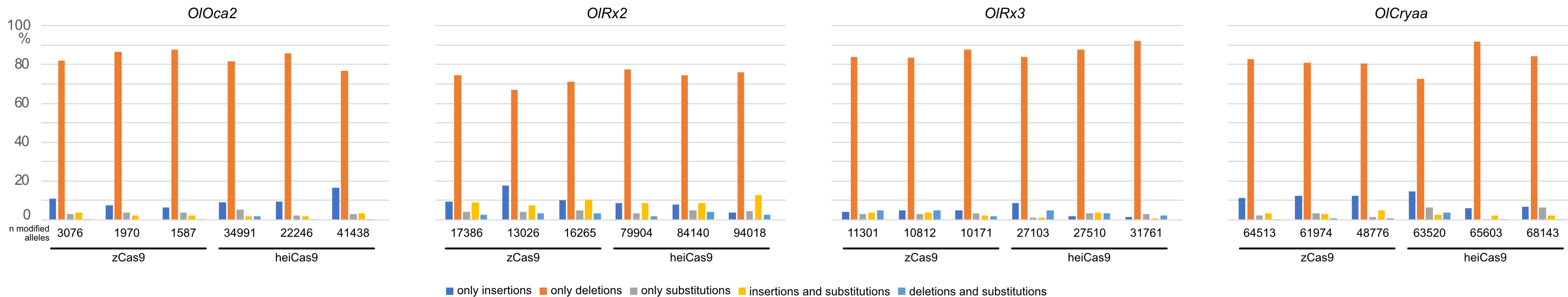
Sequences of peptide tags fused to mammalian SpCas9 (cf. Figure 1-supplementary file 1). M, cMyc-tag; O, optimized NLS (Inoue et al., 2016); S, SV40 NLS (Kalderon et al., 1984); XI, bipartite *Xenopus laevis* nucleoplasmin NLS (Dingwall et al., 1988).

Supplementary File 3

Allele variants and abundance in *OIOca2*, *rx2*, *rx3* and *cryaa*. Sequences and abundance of all locus-mapped reads per replicate (pool) and locus (*OIOca2*, *rx2*, *rx3*, *cryaa*) of multiplexing with either *zCas9* or *heiCas9* mRNA injections (cf. Figure 2). Sequences of allele variants (with more than 100 reads) displayed.

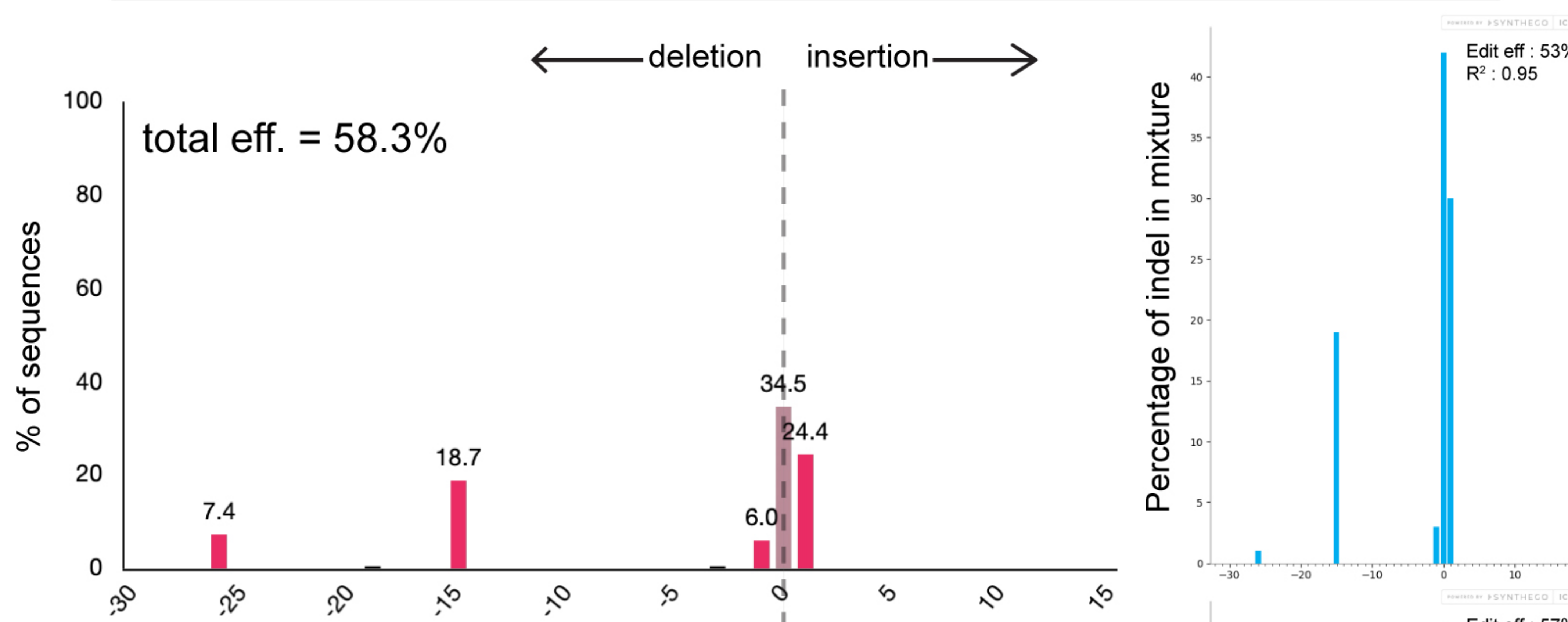




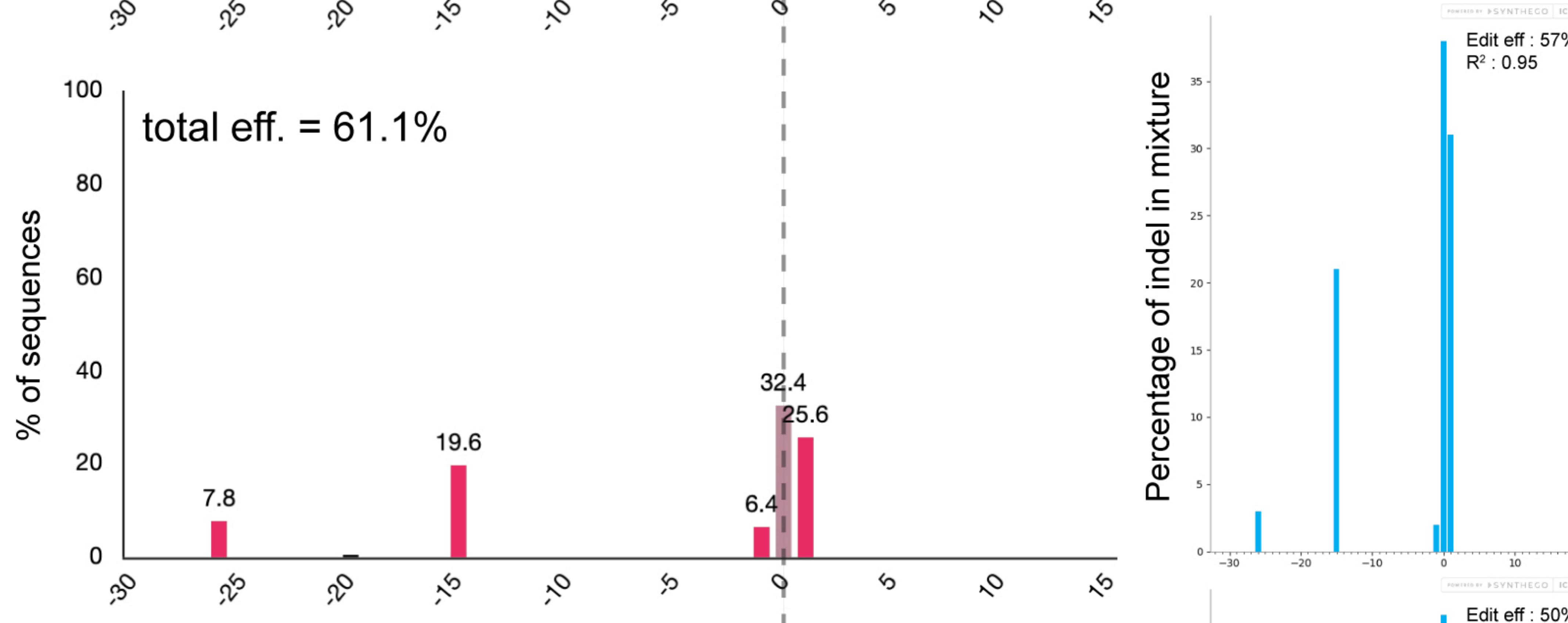


JDS246-Cas9

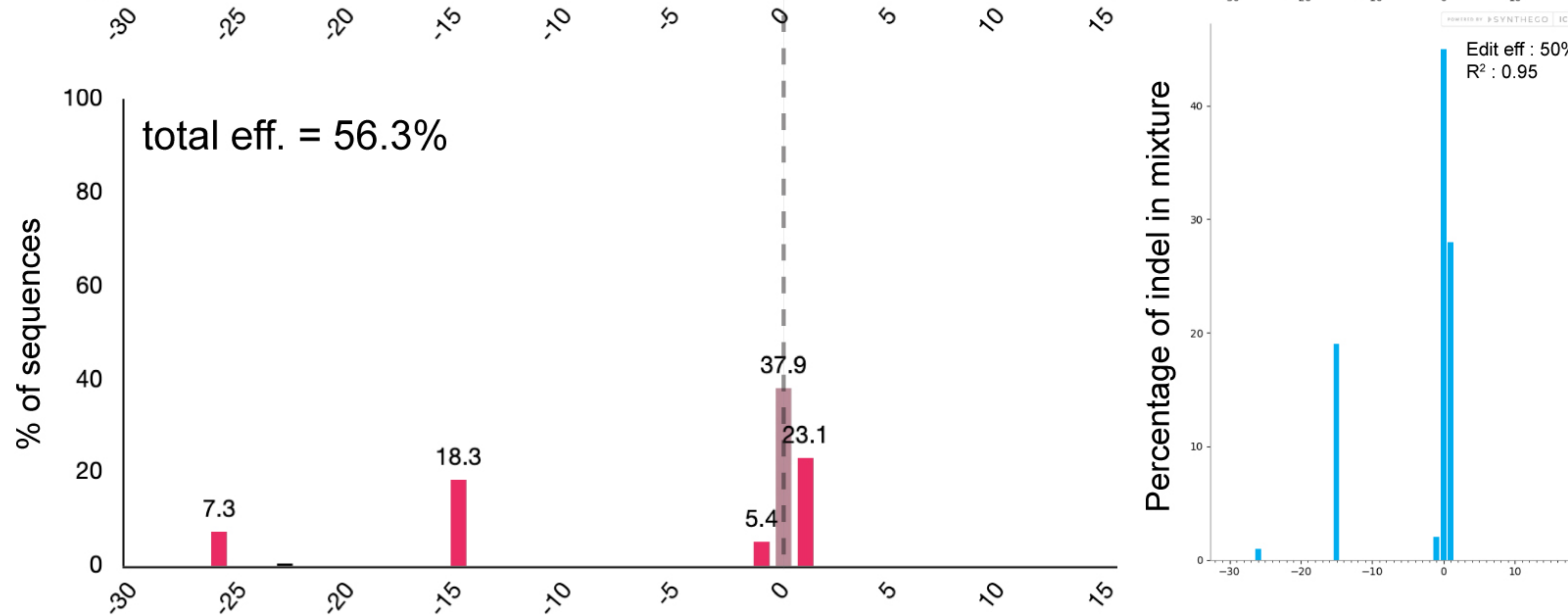
pool 1



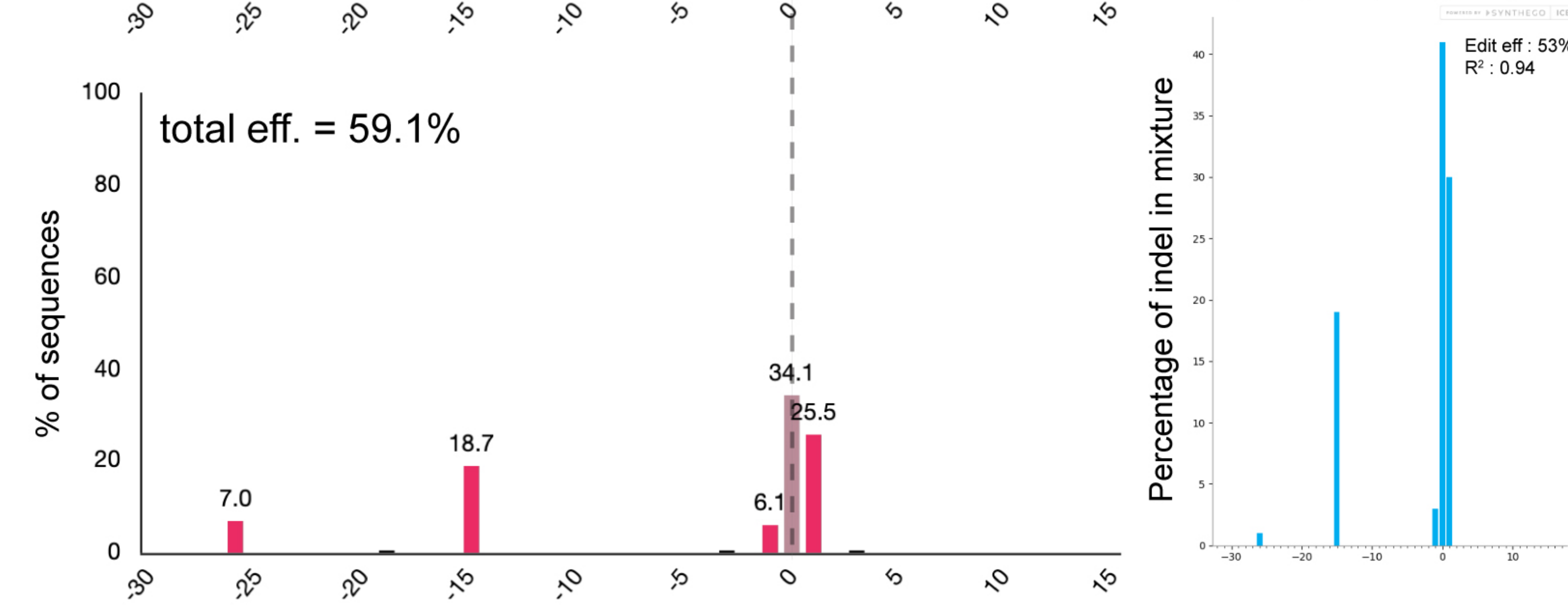
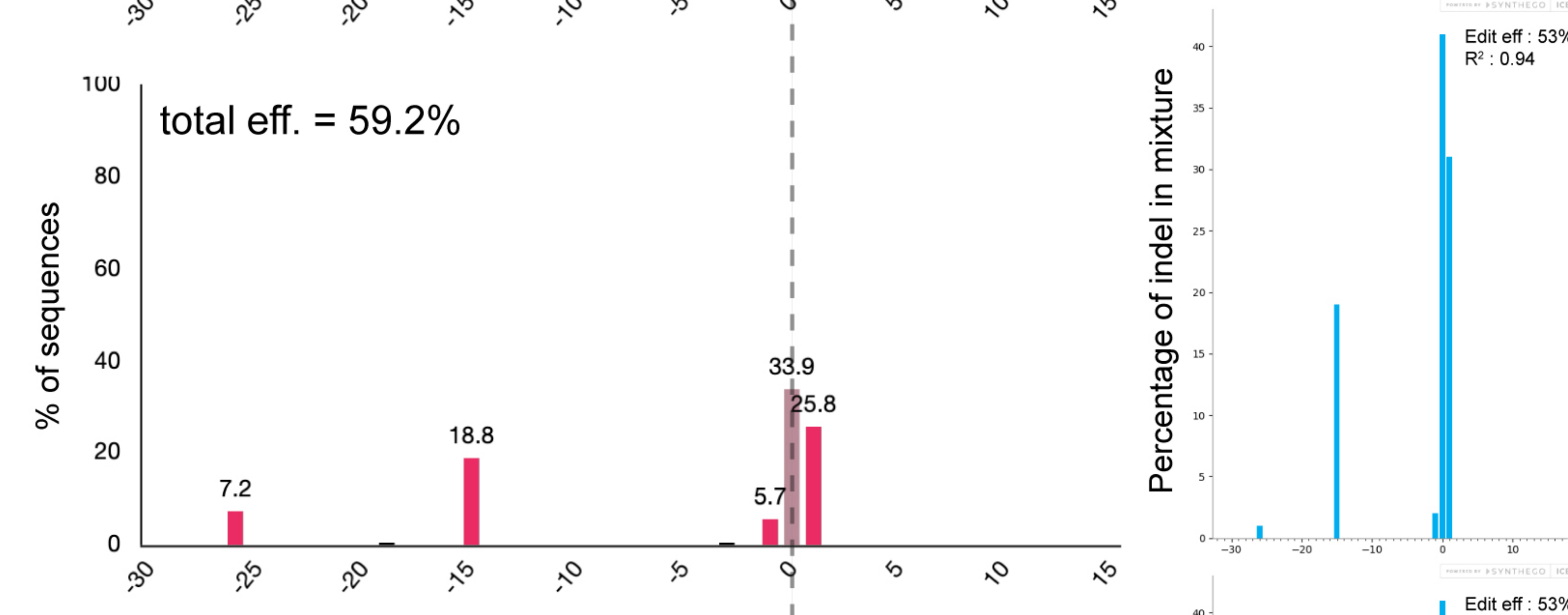
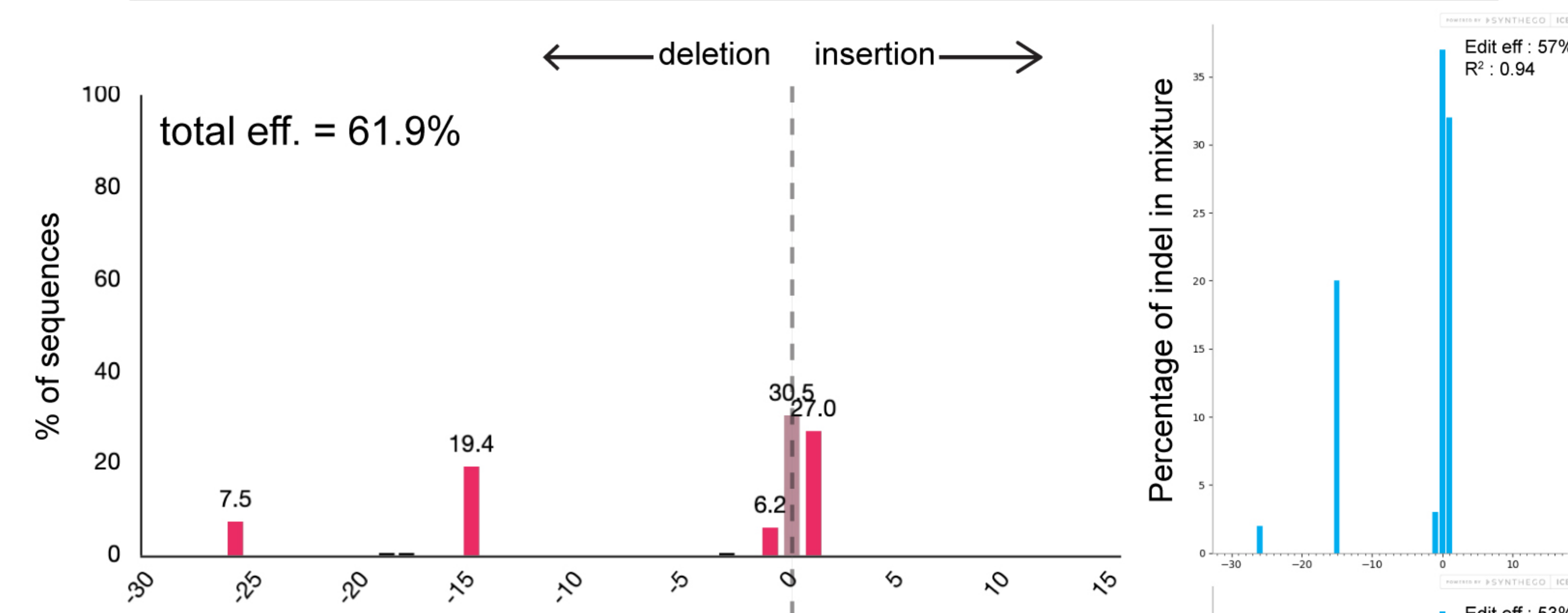
pool 2



pool 3



GeneArt® CRISPR nuclease



heiCas9

