1 **Full Title: Toward Precision Molecular Surgery: Robust, Selective**

2 **Induction of Microhomology-mediated End Joining *in vivo***

3 Short Title: MMEJ-based approach for precision genome engineering

4 *in vivo*

5 **Author Names and Affiliation**

6 **Hirotaka Ata[1,2,3,4], Thomas L. Ekstrom[1,4], Gabriel Martínez-Gálvez [2,4], Carla M.**

7 **Mann[4,5], Alexey V. Dvornikov[6], Kyle J. Schaefbauer[6], Alvin C. Ma[7], Drena Dobbs[4,5],**

8 **Karl J. Clark[4,6], and Stephen C. Ekker[1,4,6]**

9 [1]Mayo Clinic, Center for Clinical and Translational Science. Rochester, MN, USA

10 [2]Mayo Clinic, Graduate School of Biomedical Sciences. Rochester, MN, USA

11 [3]Mayo Clinic, Medical Scientist Training Program, Rochester, MN, USA

12 [4]Iowa State/Mayo Clinic Alliance for Genome Engineering

13 [5]Iowa State University, Genetics, Development, and Cell Biology, Ames, IA, USA

14 [6]Mayo Clinic, Biochemistry and Molecular Biology. Rochester, MN, USA

15 [7]The University of Hong Kong, Department of Paediatrics and Adolescent Medicine,

16 Hong Kong, Hong Kong

17

18 **Author Contribution**

19 **HA** contributed in Conceptualization, Data Curation, Formal Analysis, Investigation,

20 Funding Acquisition, Methodology, Validation, Visualization, Writing – Original draft

21 preparation, and Writing – Review and Editing. **TLE** contributed in Data Curation,

22 Investigation, Writing – Original draft preparation, and Writing – Review and Editing.

23 **GMG** contributed in Software, Validation, and Writing**. CMM** contributed in Software

24   Validation, and Writing. **AVD** contributed in Investigation, Methodology, Validation, and

25   Writing – Review and Editing. **KJS** contributed in Investigation and Writing – Review

26   and Editing. **ACM** contributed in Conceptualization, Data Curation, Investigation, and

27   Writing – Review and Editing. **DD** contributed in Funding Acquisition, Resources, and

28   Writing – Review and Editing. **KJC** contributed in Conceptualization, Funding

29   Acquisition, Resources, Supervision, and Writing – Review and Editing. **SCE**

30   contributed in Conceptualization, Funding Acquisition, Project Administration,

31   Resources, Supervision, Writing – Review and Editing.

32

33   **Abstract:**

34   One key problem in precision genome editing is the resultant unpredictable

35   plurality of sequence outcomes at the site of targeted DNA double-strand breaks

36   (DSBs). This is due to the typical activation of the versatile Non-homologous End

37   Joining (NHEJ) pathway. Such unpredictability limits the utility of somatic gene editing

38   for applications including gene therapy and functional genomics. For germline editing

39   work, the accurate reproduction of identical alleles using NHEJ is a labor intensive

40   process. In this study, we propose inducing Microhomology-mediated End Joining

41   (MMEJ) as a viable solution for improving somatic sequence homogeneity *in vivo*,

42   capable of generating a single predictable allele at high rates (56% ~ 86% of the entire

43   mutant allele pool). Using a combined dataset from zebrafish (*Danio rerio*) *in vivo* and

44   human HeLa cell *in vitro* as a training dataset, we identified specific contextual

45   sequence determinants surrounding genomic DSBs for robust MMEJ pathway

46   activation. We then applied our observation and prospectively designed MMEJ-inducing

47   sgRNAs against a variety of proof-of-principle genes and demonstrated a high level of

48   mutant allele homogeneity at these loci. F0 mutant zebrafish embryos and larvae

49   generated with these gRNAs faithfully recapitulated previously reported, recessive loss-

50   of-function phenotypes. We also provide a novel algorithm MENTHU

51   (http://genesculpt.org/menthu/) for improved prediction of candidate MMEJ loci, suitable

52   for both targeted and genome-wide applications. We believe that this MMEJ-centric

53   approach will have a broad impact on genome engineering and its applications. For

54   example, whereas somatic mosaicism hinders efficient recreation of a knockout mutant

55   allele at base pair resolution via the standard NHEJ-based approach, we demonstrate

56   that F0 founders transmitted the identical MMEJ allele of interest at high rates. Most

57   importantly, the ability to directly dictate the reading frame of an endogenous target will

58   have important implications for gene therapy applications in human genetic diseases.

59 **Author Summary:**

60   New gene editing tools precisely break DNA at pre-defined genomic locations,

61   but cells repair these lesions using diverse pathways that often lead to unpredictable

62   outcomes in the resulting DNA sequences. This sequence diversity in gene editing

63   outcomes represents an important obstacle to the application of this technology for

64   human therapies. Using a vertebrate animal as a model system, we provide strong

65   evidence that we can overcome this obstacle by selectively directing DNA repair of

66   double-stranded breaks through a lesser-described pathway termed Microhomology-

67   mediated End Joining (MMEJ). Unlike other, better-understood pathways, MMEJ uses

68   recurring short sequence patterns surrounding the site of DNA breakage. This enables

69   the prediction of repair outcomes with improved accuracy. Importantly, we also show

70   that preferential activation of MMEJ is compatible with effective gene editing. Finally, we

71   provide a simple algorithm and software for designing DNA-breaking reagents that have

72   high chance of activating the MMEJ pathway. We believe that the MMEJ-centric

73   approach to be broadly applicable for a variety of gene editing applications both within

74   the laboratory and for human therapies.

Introduction:

75

76   Programmable nucleases such as TALEN (Transcription Activator-like Effector

77   Nuclease) and CRISPR (Clustered Regularly Interspaced Short Palindromic Repeats)

78   systems have enabled a new era of scientific research(1, 2). Instead of relying on

79   knock-down models or expensively outsourced knockout lines, laboratories across the

80   world now have tools with which to generate indels (Insertions and deletions) of varying

81   sizes in the gene(s) of interest. However, DNA Double-strand Break (DSB) repairs

82   largely result in diverse sequence outcomes owing to the unpredictable nature of the

83   most commonly used Non-homologous End Joining (NHEJ) pathway(3, 4) (**Fig 1**). This

84   significantly confounds experimental readouts because knock-out cell lines often harbor

85   more than just one desired frameshift mutation. In the case of model organisms such as

86   zebrafish (*Danio rerio*), the F0 founders are genetically mosaic, warranting a complex

87   and time-consuming series of outcrossing to establish molecularly defined lines before

88   any biological questions can be addressed(5, 6).

89   In contrast to NHEJ, the MMEJ (Microhomology-mediated End Joining) DNA

90   repair pathway utilizes a pair of locally available direct sequence repeats on both sides

91   of a DSB that are apposed, annealed and extended(7-10) (**Fig 1**). As such, DSB repair

92   outcomes are highly stereotyped, resulting in deletion of the intervening sequence as

93   well as one of the repeats. Consequentially, there is an increasing interest in utilizing

94   MMEJ for precision genome engineering applications(11-14). To date, however,

95   effective harnessing of this pathway remains challenging due to the paucity of genetic

96   and mechanistic understanding(8).

97   Bae *et al.*(14) developed a sequence-based scoring system to estimate the

98   frequency of MMEJ-associated deletions induced by DSBs in human cells. While this

99   improved the predictability of MMEJ activation, the DSB repair outcomes tended to

100   consist of a heterogeneous population of multiple MMEJ alleles. In this study, we sought

101   to improve upon the existing algorithm with the goal of developing tools to more reliably

102   predict target loci that would be predisposed to generate a more homogeneous mutant

103   allele population through MMEJ. We demonstrate the feasibility and utility of such

104   reagent design on the molecular level (i.e., DNA repair outcomes) and on the

105   physiological level (i.e., F0 phenotype). We believe our approach can inform and benefit

106   applications such as rapid phenotype-genotype correlation in F0 animals, with an eye

107   toward applications in human gene therapy and facilitation of resource sharing and

108   recreation of various cell and animal lines on a global scale.

109 **Results:**

110 **MMEJ is an Active Repair Pathway in the Genetically Unaltered Zebrafish Embryo**

111 Prior work examining MMEJ activation in vertebrate organisms primarily focused on *in*

112 *vitro* models (8-10, 14-18). Initial analyses using a targeted knockin strategy suggested

113 that MMEJ was operational in the zebrafish embryo, though the efficiency of these

114 MMEJ outcomes was rather modest(13). One study reported incidental identification of

115 MMEJ inductions at two zebrafish genomic loci using programmable nucleases (19).

116 However, no consortium – small or large – of genomic loci that repair primarily through

117 NHEJ vs MMEJ has been compiled. To this end, we examined the repair outcomes of

118 previously designed TALEN and CRISPR-Cas9 genomic reagents (**S1 Table**). The

119 plurality of custom enzymes induced diverse sequence outcomes, consistent with NHEJ

120 being used as the primary DNA repair pathway. However, a few reagents induced

121 sequence outcomes satisfying the following criteria, suggesting that MMEJ was the

122 preferred repair pathway: 1) most predominant allele is the top predicted allele by the

123 Bae *et al.* algorithm(14); 2) most predominant allele comprises ≥ 50 % of the total

124 mutant allele population; and 3) mutagenic efficiency > 20%. For the purpose of this

125 study, a programmable nuclease satisfying all these criteria is referred to as a "Winner-

126 Take-All" reagent. Three TALEN (*chrd*, *mitfa* #4 & *surf1*) and two CRISPR-Cas9 (*surf1*

127 & *tyr* #2) reagents fell into this category (**S1 Table**, **Fig 2A**, **Fig 3A**).

128 Injecting the *chrd* TALEN pair (37.5 pg/arm) resulted in characteristic *chrd* loss of

129 function phenotypes: Intermediate-Cell-Mass (ICM) expansion and a smaller head by 1

130 day post-fertilization(20) (1 dpf; **Fig 2B**). Median penetrance for Moderate and Severe

131 phenotypes was 15.8% and 20.0%, respectively (**Fig 2B**, **S2 Table**). Strong MMEJ

132   activation by this TALEN pair was confirmed by subcloning results (**Fig 2A**) – 16/32

133   recovered mutant reads corresponded to the top predicted 7bp deletion allele. Similarly,

134   perturbing *tyr* gene with a CRISPR-Cas9 reagent recapitulated a previously reported

135   loss of melanin production phenotype, observable by 2 dpf(21) (**Fig 3B**).

136   Ribonucleoprotein (RNP) delivery at the dose of 300 pg *tyr* #2 sgRNA and 660 pg Cas9

137   resulted in Moderate and Severe loss of pigmentation phenotypes in 22.7% and 50.0%

138   of embryos, respectively (**Fig 3B**, **S2 Table**). Subcloning analysis showed 21/24 (88%;

139   **Fig 3A**) of resulting alleles contained a 4bp deletion consistent, with strong MMEJ

140   activation by this CRISPR-Cas9. Together with the *chrd* TALEN results, these data

141   support that MMEJ can be an effective repair pathway in F0 embryos at some genomic

142   loci.

143

144   **Many Bae *et al.* Predicted MMEJ Loci Are Preferentially Repaired by NHEJ**

145   A subset of the zebrafish reagents described above was prospectively designed using

146   the Bae *et al.* algorithm(14) (**S1 Table**). This algorithm calculates the strength of each

147   pair of microhomology arms according to the length and GC content of each pair, as

148   well as the length of the intervening sequence (i.e., *Pattern Score*). The additive sum of

149   all the possible *Pattern Scores* is then returned as *Microhomology Score*. This latter

150   score was found to have positive correlation with the rate of MMEJ activation in HeLa

151   cells(14). All fourteen prospectively designed reagents had a *Microhomology Score* of at

152   least 4000 – a median score found on human *BRCA1* gene. However, only four of these

153   reagents induced majority MMEJ outcomes as judged by the *Microhomology Fraction*

154   (**S1 Table**, **S1 Note**). We therefore retrospectively analyzed the repair outcomes of

155   these reagents to identify additional factor(s) that could enhance predictability of MMEJ

156   induction.

157

Rate of Pattern Score Change as a Discrimination Factor for MMEJ Induction *in vivo* and *in vitro*

160   Intriguingly, when the pattern score values clustered closely to one another (i.e.,

161   a flatter *Slope Value* as calculated according to **S2 Note**), this was indicative of an

162   unfavorable target for MMEJ activation in zebrafish embryos. Conversely, loci at which

163   pattern scores dropped precipitously (i.e., a steeper *Slope Value*) were good candidates

164   for MMEJ activation *in vivo* ($p = 0.0048$; **S1 Figure**). Based on these observations, we

165   hypothesized that locally available microhomology pairs are in direct competition with

166   one another, such that overabundance of these pairs is a negative predictor of MMEJ

167   activation. In other words, MMEJ activation is more favorable at loci with only one or two

168   predominant microhomology pair(s) (Low Competition loci) rather than many strong

169   microhomology pairs (High Competition loci).

170   To determine whether the zebrafish-based hypothesis was generalizable to

171   human cells (HeLa), we re-analyzed the deep sequencing dataset used to generate the

172   Bae *et a.l* algorithm(14). Available results from 90 genomic loci were sorted

173   alphabetically by the names of target genes. Outcomes from the first 50 targets showed

174   a correlation similar to that observed in zebrafish; higher *Microhomology Fractions*

175   generally correlated with low *Slope Values* from the first 50, alphabetically sorted

176   targets ($p = 0.00001$; **S2 Figure A**). This correlation was lost when microhomology arms

177   of 2 bp were included in the analysis ($p = 0.2644$; **S2 Figure B**). Accordingly,

178   microhomology arms of less than 3 bp were excluded from subsequent analyses. The

179   remaining 40 targets were then binned into High, Medium and Low Competition groups

180   based on quartiles (**S2 Figure C**) – the median *Microhomology Fraction* was

181   significantly higher in the High Competition group than in the Low Competition group

182   (0.300 vs 0.105, p = 0.011; **S2 Figure D**).

183

184   **Competition Hypothesis Predicts New Winner-Takes-All Reagents**

185   Based on this Competition Hypothesis, we designed and analyzed the DSB repair

186   outcomes of 20 Low Competition sgRNA targets across 9 genes (**S3 Table**). *Slope*

187   *Values* smaller than -40 was used as the cut-off for Low Competition, as 3 out of 4

188   previously designed zebrafish targets produced majority MMEJ outcomes in this range

189   (**S1 Table and S1 Figure**). For initial assessments, we used TIDE (Tracking Indels by

190   DEcomposition) analysis – a chromatogram analyzing tool that estimates proportions of

191   length varying mutant alleles present in a pool of mixed alleles (22) – which revealed

192   that 5 of these sgRNAs against 3 genes: *mtg1,* Mitochondrial GTPase 1; *tdgf1,*

193   Teratocarcinoma-Derived Growth Factor 1; and *ttn.2,* titin (*ttn.2 #1*, *ttn.2 #2*, *ttn.2* N2B

194   #1) were in the "Winner-Take-All" class. These results were subsequently confirmed

195   by subcloning analysis (**S3 Table**). Perturbation of *tdgf1* (alternatively known as *One-*

196   *eyed Pinhead*) causes aberrant "fused eyes" morphology and cyclopia, as judged by

197   reduced forebrain protrusion by 1 dpf(23) (**Fig 4B**). Aberrant head morphology alone

198   was classified as Weak, whereas that in combination with varying degrees of forebrain

199   protrusion was classified as Moderate or Strong phenotypes. RNP injections of

200   CRISPR-Cas9 at the dose of 300 pg sgRNA and 660 pg Cas9 resulted in median

201   penetrance for Moderate and Severe morphology at 21.8% and 11.4% (**Fig 4B**, **S2**

202   **Table**). The subcloning results were consistent with the noted phenotypic outcomes

203   from this common 4bp deletion (**Fig 4A**).

204   We next explored whether these "Winner-Take-All" reagents could be useful for

205   recapitulating a more subtle phenotype than the aberrant gross morphologies observed

206   in the *tdgf1* mutants. Splice blockade at the N2B exon of *ttn.2* gene by a synthetic

207   morpholino oligonucleotide was previously reported to reduce cardiac contractility by

208   ~70% on 2 dpf(24), phenocopying the *pickwick*^{m171} mutation(25). RNP delivery at the

209   dose of 300 pg *ttn.2* N2B #2 sgRNA + 660 pg Cas9 resulted in reduction of the

210   shortening fraction to a comparable degree (**Fig 5B**). Importantly, injecting RNP with *tyr*

211   *#2* sgRNA or sgRNA and Cas9 independently, at the same doses, did not affect the

212   shortening fraction. Due to the high editing efficiency (**Fig 5A**), animals injected with

213   these doses of CRISPR-Cas9 were not viable in post larval phases. For this reason,

214   animals injected at the lower dose of 75 pg sgRNA + 165 pg Cas9 protein were raised

215   to adulthood. Two F0 founders were successfully outcrossed to wild type zebrafish.

216   Heterozygous offspring were identified using the dsDNA heteroduplex-cleaving

217   Surveyor assay(26), and the transmission of the top predicted 5bp deletion allele was

218   confirmed from both founders by subcloning analyses (**S3 Figure**).

219   We also designed an sgRNA against exon 13 of *ttn.2* (*ttn.2* #2 sgRNA), expected

220   to produce a 12 bp deletion allele as a proof-of-principle for in-frame gene correction

221   (**Fig 6A**). RNP delivery at the dose of 300 pg sgRNA + 660 pg Cas9 resulted in the

222   induction of this 12 bp deletion allele at 72.7% of the resultant clones. While the injected

223   animals presented with mild cardiac edema evident by 2 dpf (median rate: 50.0%; **Fig**

224   **6B**, **S2 Table**), unlike the N2B sgRNA #1 CRISPR-Cas9 injected animals, these were

225   viable to adult age.

226

227   **Low Competition Plus Proximity of Microhomology Arms Strongly Predicts**

228   **Winner-Take-All Reagents: V2**

229   These data implicate the utility of "Winner-Take-All" class reagents for various

230   applications that require precision gene editing. However, sgRNA design based on the

231   Competition Hypothesis yielded only 5 Winner-Take-All reagents out of 20 that were

232   tested (**S3 Table**, **S3 Note**). Although this represents an improvement over the initial

233   approach relying solely on the *Microhomology Score*, we sought to further fine-tune the

234   predictability for the Winner-Take-All targets. To this end, we pooled the results from all

235   the programmable nucleases described above (**S1 and S3 Tables**) and seven Medium

236   ~ High Competition sgRNAs designed as controls based on the Competition Hypothesis

237   (**S4 Table**). In so doing, we noted that Winner-Take-All outcomes were observed only if

238   the two arms of the microhomology of the top predicted MMEJ allele for a locus were

239   separated by no more than 5 bp of intervening sequence. Thus, we identified a second

240   parameter: high ratio (≥ 1.5) of the *Pattern Scores* between the top and second

241   predicted MMEJ alleles for a given locus (**Fig 7**). Seven out of eight reagents that

242   satisfied both of these parameters were Winner-Take-All. Of the nine reagents that

243   satisfied the first parameter but not the second, 2 were Winner-Take-All. All the other 30

244   reagents that failed to meet the first parameter failed to induce the top predicted MMEJ

245   allele strongly. Most importantly, all the failed cases, i.e., incorrect predictions according

246   to the original Competition Hypothesis, can be explained using our revised approach

247   (Competition Hypothesis V2; **Fig 7C**). Version 2 also captured 3 Winner-Take-All

248   reagents that would have been missed by the original Competition Hypothesis alone,

249   and 1 Winner-Take-All reagent that would have been missed by the *Microhomology*

250   *Score* alone. Finally, a similar trend was observed with the HeLa cell dataset (**S4**

251   **Figure**). While the effect is not as dramatic as in zebrafish, this points to a possibly

252   conserved mechanism for MMEJ activation in vertebrate organisms.

253

254   **Accessing the Winner-Take-All Algorithm via MENTHU (MMEJ kNockout Target**

255   **Heuristic Utility)**

256   The broad potential utility of this updated "Winner-Take-All" Algorithm for MMEJ

257   prediction led us to develop a web-based automated analysis tool called MENTHU

258   (http://genesculpt.org/menthu/). The software can also be downloaded and installed on

259   a local computer (www.github.com/Dobbs-Lab/menthu/). MENTHU accepts a user-

260   specified DNA sequence and targeting scheme as input, and outputs recommended

261   CRISPR sgRNA target sites that are predicted to result in Winner-Take-All type

262   outcomes. We validated the accuracy and functionality of MENTHU against select

263   gRNA sites used in this study using whole exonic sequences as inputs (**S5 Table**).

264   Importantly, the software identified novel Winner-Take-All candidate loci against *surf1*

265   and *tdgf1,* where only Group 3 gRNA loci had been found by previous methods.

266

267   **Discussion:**

268   To date, precision genome engineering is limited by the ability to predictably,

269   efficiently, and reproducibly induce the identical sequence alterations in each and every

270   cell. Here, we demonstrate the feasibility and utility of creating allelic consistency by an

271   MMEJ-centric approach for designing programmable nucleases. While the precise

272   cellular components of the molecular machinery involved in MMEJ remain incompletely

273   understood(8), we provide evidence that we can enrich for MMEJ events by strictly

274   sequence-based queries.

275   Importantly, we demonstrate that MMEJ predominant repairs do not operate at

276   the cost of overall mutagenic efficiency; median edit efficiency for Winner-Take-All

277   reagents was 91.4%. As genetically unaltered wild type zebrafish were used throughout

278   the study, we have no reason to believe that NHEJ should have failed at any tested loci.

279   This is in contrast to the current perception that MMEJ is a back-up pathway to NHEJ(7,

280   8, 16, 17, 27).

281   Based on the data presented here, we speculate that there is a reaction-limiting

282   factor for MMEJ that is involved in identifying compatible microhomology pairs on both

283   sides of the DNA double stranded break. In the case of abundantly available local

284   microhomology pairs, sometimes this factor fails to localize to a single suitable pair, thus

285   rejecting the MMEJ activation. As end-resection is required for MMEJ and not for NHEJ

286   (9, 17, 18), this yet identified factor may be the deciding factor for committing DSB

287   repair through one End Joining pathway to another. This view is similar to a recent

288   report wherein CtIP/Artemis dependent limited end resection was a key trigger for a

289   slow-kinetic Lig1/3 independent NHEJ event that frequently utilized Microhomology to

290   repair a reporter plasmid(28). In our analysis, the primary driver of this decision making

291   process is the proximity of 2 microhomology arms. Our present findings break down the

292   key triggers for MMEJ activation into a simple 2-component math system with the aims

293   of making the MMEJ-centric approach to gene editing accessible. In addition, this study

294   should inform future studies on the MMEJ pathway by enabling the identification of

295   strongly MMEJ and strongly NHEJ loci under the same genetic context for in-depth

296   comparative molecular analyses.

297   Successful deployment of the Winner-Take-All reagents makes it possible to

298   directly dictate the reading frame or to do in-frame gene manipulations on endogenous

299   targets. Even assuming a somewhat modest outcome of 50% edit efficiency in which

300   50% of the mutant allele pool is of the desirable allele, more than 10% of the cell

301   population will be homozygous for this desired allele. Conversely, many real-life gene

302   editing applications would require only one of the diploid copies to be corrected. In

303   these settings under the same assumptions, just 11 viable cells are needed to achieve

304   95% confidence for establishing the right clone, bringing the idea of this kind of

305   molecular surgery closer to reality.

306   Our present study expands upon the current state-of-art for MMEJ activation and

307   demonstrates the ability to prospectively design robustly active Winner-Take--All

308   reagents *in-vivo*. We also provide evidence that this 2-component system to identify the

309   Winner-Take-All loci may be broadly applicable beyond zebrafish. To make this

310   algorithm more accessible, we developed the web-based server, MENTHU;

311   http://genesculpt.org/menthu/). MENTHU should enable testing the hypothesis that our

312   present findings are generalizable in other biological systems. In addition, MMEJ-based

313   loci are inherently restricted to genomic locations that leverage endogenous sequence

314   contexts. Consequently, the requirement for a GG dinucleotide for SpCas9 restricts the

315   potential MMEJ sites. However, active investigations are underway to accommodate

316   alternative or more lax PAM requirements. One such example is a recent variant of

317   Cas9 (xCas) that may function efficiently on an NG PAM(29). However, MENTHU

318   allows users to flexibly define any PAM sequence and the cut site (in nts from PAM) to

319   accommodate potential future variants of the CRISPR system.

320   Finally, we provide strong evidence to support the utility of the MMEJ-centric

321   approach beyond correlating phenotype-genotype in F0 animals. We envision this

322   approach to be useful for: 1) studying the effects of homozygous gene knock-out in

323   culture cells (as opposed to more common, compound heterozygous loss-of-function

324   cell lines), 2) rapid small molecule screening in F0 animals as a complimentary

325   approach to studying in germline mutant animals, 3) globally sharing and reproducing

326   gene knock-out cell and animal lines, and finally, 4) human gene therapy.

327 **Materials and Methods:**

328 <u>Microhomology arms</u>: For the purpose of this study, microhomology is defined as any

329 endogenous direct sequence repeats of ≥ 3 bp surrounding a DSB site. ≤ 2bp direct

330 sequence repeats were not considered sufficient substrates of MMEJ activation based

331 on our initial analyses of the DSB repair outcomes by previously designed

332 programmable nucleases. Correlation for *Microhomology Fraction* vs the *Slope Value*

333 was tangentially stronger when only ≥ 3 bp arms were considered (r^2 = 0.382 vs r^2 =

334 0.353; **S1 Figure**) in zebrafish, whereas the correlation was lost when 2bp arms were

335 considered in HeLa cells (r^2 = 0.339 vs r^2 = 0.034; **S2 Figure**).

336

337 *<u>Zebrafish Husbandry</u>*: All zebrafish (*Danio rerio*) were maintained in accordance with

338 protocols approved by the Institutional Animal Care and Use Committee at Mayo Clinic.

339 Zebrafish pairwise breeding was set up one day before microinjections and dividers

340 were removed the following morning. Following microinjections, the fertilized eggs were

341 transferred to Petri dishes with E3 media [5mM NaCl, 0.17mM KCL, 0.33mM CaCl, and

342 0.33mM $MgSO_4$ at pH 7.4] and incubated at 28.5°C. All subsequent assays were

343 conducted on fish less than 3 dpf, with the exception of assessing for germline

344 transmission. In this case, injected founders were raised to adulthood per the standard

345 zebrafish husbandry protocol.

346

347 *<u>DNA Oligonucleotide Preparation</u>*: All of the oligonucleotides used for this study were

348 purchased from IDT (San Jose, CA). Upon arrival, they were reconstituted into 100μM

349 suspensions in 1x TE and stored at -20°C until use.

350

351   *sgRNA Expression Vector synthesis*: pT7-gRNA was a gift from Wenbiao Chen

352   (Addgene plasmid # 46759). Given that the minimum requirement for the T7 promoter is

353   a single 5' G, the GG start on this vector was mutagenized to accommodate GA, GC,

354   GT starts, using Forward and Reverse primers given (**S5 Table**). Platinum Pfx DNA

355   Polymerase (Invitrogen 11708013. Carlsbad, CA) was used for 20 cycles of PCR

356   amplification with the Tm of 60 °C and extension time of 3 minutes. DpnI (NEB R0176.

357   Ipswich, MA) was subsequently added to reaction prior to transforming DH5α cells. The

358   target sequence was cloned in as previously described, with the exception of

359   conducting oligo annealing and T4 ligation (NEB M0202. Ipswich, MA) in 2 separate

360   steps. In each case, transformed cells were cultured with Carbenicillin, and plasmids

361   were purified with Plasmid Mini Kit (Qiagen 12123. Hilden, Germany).

362

363   *TALEN synthesis*: TALEN constructs were generated using the FusX kit (Addgene #

364   1000000063) as previously described(30). In short, RCIscript-GoldyTALEN was

365   linearized with BsmBI (NEB R0580. Ipswich, MA) along with 6 triplet RVD (Repeat-

366   Variable Diresidue) plasmids. Subsequently, they were ligated together in one reaction

367   by a modified Golden-Gate Assembly. Blue-White colony screening with X-Gal/IPTG,

368   colony PCR and finally pDNA sequencing were done to ascertain the correct assembly.

369

370   *In-vitro Transcription and RNA preparation*: pT3TS-nCas9n (a gift from Wenbiao Chen:

371   Addgene plasmid # 46757) was linearized with XbaI (NEB R0145. Ipswich, MA),

372   whereas TALEN constructs were linearized with SacI-HF (NEB R3156) and sgRNA

373   vector with BamHI-HF (NEB R3136. Ipswich, MA). Tyr sgRNA #2 – a construct made in

374   the Essner Lab – was linearized with HindIII (NEB R0104. Ipswich, MA). RNA was

375   made using T3 mMessage mMachine kit (Ambion AM1348. Foster City, CA) or HiScribe

376   T7 High Yield RNA synthesis kit (NEB E2040. Ipswich, MA) according to manufacturer's

377   protocols with the addition of RNA Secure to the reaction (Ambion AM7010. Foster City,

378   CA). To purify RNA, phenol-chloroform extraction was performed using Acid Phenol,

379   Chloroform, and MaXtract High Density Tubes (Qiagen 129046. Hilden, Germany). RNA

380   was then precipitated with Isopropanol at -20 °C, pelleted, air dried and resuspended

381   into nuclease free water. The quality and quantity of RNA were ascertained by using a

382   Nanodrop spectrophotometer and running aliquot on agarose gel. Each batch of RNA

383   was aliquoted into small single use tubes and stored at -80 °C until the morning of

384   microinjections.

385

386   _CRISPR-Cas9 RNP preparation for microinjections_: sgRNA was thawed on ice in the

387   morning of microinjections. This was then diluted to the concentration of 300 ng/µL in

388   Duplex Buffer [100 mM KCH_3COO, 30 mM HEPES at pH 7.5]. Appropriate folding of

389   sgRNA was induced by heating it to 95 °C for 5 minutes and gradually cooling the

390   solution to room temperature. Equal volumes of sgRNA and 0.66 mg/mL Alt-R S.p.

391   Cas9 Nuclease 3NLS (IDT 1074181. San Jose, CA) in Cas9 Working Buffer [20 mM

392   HEPES, 100 mM NaCl, 5 mM $MgCl_2$, 0.1 mM EDTA at pH 6.5] were mixed and

393   incubated at 37 °C for 10 minutes. RNP solutions were subsequently kept on ice until

394   immediately before use.

395

396   *TALEN and CRISPR-Cas9 RNA preparation for microinjections*: RNA was thawed on

397   ice in the morning of microinjections. TALEN mRNA was diluted to working

398   concentrations in the range of 12.5 ng/μL to 100 ng/μL in Danieau solution [58 mM

399   NaCl, 0.7mM KCl, 0.4 mM MgSO$_4$, 0.6 mM Ca(NO$_3$)$_2$, 5.0 mM HEPES at pH 7.6].

400   sgRNA and nCas9n mRNA were mixed and diluted to the final concentrations of 150

401   ng/μL and 100 ng/μL, respectively, in Danieau solution. These were all kept on wet ice

402   until immediately before use.

403

404   *Microinjections*: Microinjections were carried out as previously described(31). In short,

405   1-cell stage fertilized embryos were harvested and aligned on an agarose plate with E3

406   media. In the case of CRISPR-Cas9 reagents, either 1 or 2 nL was delivered to the cell.

407   In the case of TALEN reagents, 1 ~ 3 nL was delivered to the yolk mass. They were

408   then transferred to Petri dishes in E3 media for incubation at 28.5 °C. Dead and/or

409   nonviable embryos were counted and removed each subsequent morning.

410

411   *Phenotype Scoring*: Each experiment was conducted in at least a technical triplicate

412   and a biological duplicate. Detailed outcomes are provided in **S4 Table**. Gross

413   phenotypes were scored visually on either 1 dpf or 2 dpf using a standard dissecting

414   microscope. Subsequently, representative pictures were taken with Lightsheet Z.1

415   (Zeiss 2583000135. Oberkochen, Germany). Shortening Fractions were scored as

416   previously reported. In short, live 2 dpf larvae were immobilized and positioned in 3%

417   methylcellulose. An Amscope camera (MU1403. Irvine, CA) mounted on a Leica

418   Microscope (M165. Wetzlar, Germany) was used to capture a 15 second clip of the

419   beating heart at 66 fps. These clips were subsequently used to measure the distance of

420   the long axis along the ventricle at maximum dilation and maximum contraction using

421   ImageJ software(32). Shortening Fraction was calculated as below:

$$\text{Shortening Fraction} = 100 * (1 - \frac{Distance\ at\ Maximal\ Shortening}{Distance\ at\ Maximum\ Dilation})$$

422   Shortening Fractions from 5 cycles were averaged for each animal.

423

424   _DNA extraction and assessing mutagenic outcomes_: Typically, 8 uninjected wildtype

425   fish and 8 injected fish were randomly collected without prior screening for phenotype.

426   Chorion was predigested with 1 mg/mL Pronase at room temperature as needed. 1 ~ 3

427   dpf animals were then sacrificed for individual DNA extractions in 100mM NaOH for 15

428   minutes at 95 °C. Equal volumes of DNA from the same condition were then mixed and

429   used as templates for PCR with either MyTaq (Bioline BIO-21108. London, UK),

430   Phusion (NEB M0530. Ipswich, MA), or KOD (EMD Millipore 71085. Burlington, MA)

431   polymerases per manufacturer's protocols. The PCR amplicon was resolved on agarose

432   gel, gel extracted with either Monarch DNA Gel Extraction Kit (NEB T1020. Ipswich,

433   MA) or QiaEx II Gel Extraction Kit (Qiagen 20021. Hilden, Germany), and subsequently

434   sent out for sequencing. The chromatograms from both uninjected and injected

435   amplicons were used for TIDE analysis(22). Alternatively, purified amplicons were used

436   for subcloning analysis with either Topo-TA Cloning Kit (Thermo Fisher Scientific

437   451641. Waltham, MA) or StrataClone PCR Cloning Kit (Agilent 240205. Santa Clara,

438   CA) per manufacturer's protocols. Resultant white to pale blue colonies by Blue-White

439   screening were subjected to colony PCR with M13F and R primers, using MyTaq

440   polymerase. Once successful amplification was confirmed on agarose gel, these

441   amplicons were sent out for sequencing either with M13F, M13R or endogenous gene

442   target primers.

443

444   *Germline Transmission for 5bp deletion generated by N2B sgRNA #1*: RNP containing

445   N2B sgRNA #1 was prepared at 4x diluted dose as described above. Following

446   microinjections, viable fish were raised to sexual maturity. Both F0 founders we

447   attempted to outcross successfully mated and produced viable embryos. DNA was

448   extracted from all viable embryos on 1 dpf, and individual DNA was used as template

449   for PCR amplification using MyTaq Polymerase. Once the thermocycling ran to

450   completion, the amplicons were melted by heating to 95 °C and re-annealed by a

451   gradual step-wise cooling. Surveyor assay(26) was conducted per the manufacturer's

452   protocol (IDT 706025. San Jose, CA), and the results were analyzed by resolving the

453   post-digest amplicons on agarose gel. Amplicons from 4 heterozygous offspring each

454   were subcloned, and 5 colonies each were sent for Sanger Sequencing to confirm

455   successful transmission of the 5bp deletion allele.

456

457   *MENTHU:*

458   We developed a software tool, MENTHU (MMEJ kNockout Target Heuristic Utility), to

459   automate calculations required to implement the 2-component Winner-Take-All strategy:

460   1) identification of top predicted microhomology arms separated by ≤ 5 bp of intervening

461   sequence; 2) identification of "low competition" target sites (i.e., with a #1-ranked to #2-

462   ranked Pattern Score ratio ≥ 1.5. We designed MENTHU to first compute two of the

463   same sequence-based parameters (*Pattern Score and Microhomology Score*) used in

464   the algorithm of Bae *et al.*, (which are computed online by the RGEN online tool,

465   http://www.rgenome.net), by re-implementing and modifying the original Python source

466   code provided in Supplemental Figure 3 (14) in R. The MENTHU webserver operates

467   under R(33) version 3.4.1 and RShiny (34) v1.0.5. The MENTHU code was built through

468   RStudio(35) v1.1.442. Details regarding specific R package versions, complete

469   documentation and a full downloadable version of MENTHU for local installation are

470   provided at www.github.com/Dobbs-Lab/menthu/. MENTHU v2.0 can be freely

471   accessed online at http://genesculpt.org/menthu/.

472

473   *Statistical Analyses*: All of the statistical analyses were carried out using JMP software

474   (SAS Institute. Cary, NC). In all instances, p-values were calculated assuming non-

475   Gaussian Distributions. Wilcoxon Each Pair calculation was used for multiple group

476   comparisons with adjusted p-values.

477

492

493   References Cited (PLOS format – ICMJE or Vancouver style)

494   1.   Campbell JM, Hartjes KA, Nelson TJ, Xu X, Ekker SC. New and TALENted genome engineering
495   toolbox. Circ Res. 2013;113(5):571-87.
496   2.   Doudna JA, Charpentier E. Genome editing. The new frontier of genome engineering with
497   CRISPR-Cas9. Science. 2014;346(6213):1258096.
498   3.   Lieber MR. The mechanism of double-strand DNA break repair by the nonhomologous DNA end-
499   joining pathway. Annu Rev Biochem. 2010;79:181-211.
500   4.   Carroll D. Genome engineering with targetable nucleases. Annu Rev Biochem. 2014;83:409-39.
501   5.   Jao LE, Wente SR, Chen W. Efficient multiplex biallelic zebrafish genome editing using a CRISPR
502   nuclease system. Proc Natl Acad Sci U S A. 2013;110(34):13904-9.
503   6.   Ata H, Clark KJ, Ekker SC. The zebrafish genome editing toolkit. Methods Cell Biol. 2016;135:149-
504   70.
505   7.   Boulton SJ, Jackson SP. Identification of a Saccharomyces cerevisiae Ku80 homologue: roles in
506   DNA double strand break rejoining and in telomeric maintenance. Nucleic Acids Res. 1996;24(23):4639-
507   48.
508   8.   Seol JH, Shim EY, Lee SE. Microhomology-mediated end joining: Good, bad and ugly. Mutat Res.
509   2017.
510   9.   Sharma S, Javadekar SM, Pandey M, Srivastava M, Kumari R, Raghavan SC. Homology and
511   enzymatic requirements of microhomology-dependent alternative end joining. Cell Death Dis.
512   2015;6:e1697.
513   10.   Kent T, Chandramouly G, McDevitt SM, Ozdemir AY, Pomerantz RT. Mechanism of
514   microhomology-mediated end-joining promoted by human DNA polymerase theta. Nat Struct Mol Biol.
515   2015;22(3):230-7.
516   11.   Nakade S, Tsubota T, Sakane Y, Kume S, Sakamoto N, Obara M, et al. Microhomology-mediated
517   end-joining-dependent integration of donor DNA in cells and animals using TALENs and CRISPR/Cas9.
518   Nat Commun. 2014;5:5560.
519   12.   Yao X, Wang X, Liu J, Hu X, Shi L, Shen X, et al. CRISPR/Cas9 - Mediated Precise Targeted
520   Integration In Vivo Using a Double Cut Donor with Short Homology Arms. EBioMedicine. 2017;20:19-26.
521   13.   Hisano Y, Sakuma T, Nakade S, Ohga R, Ota S, Okamoto H, et al. Precise in-frame integration of
522   exogenous DNA mediated by CRISPR/Cas9 system in zebrafish. Sci Rep. 2015;5:8841.
523   14.   Bae S, Kweon J, Kim HS, Kim JS. Microhomology-based choice of Cas9 nuclease target sites. Nat
524   Methods. 2014;11(7):705-6.
525   15.   Qi Z, Redding S, Lee JY, Gibb B, Kwon Y, Niu H, et al. DNA sequence alignment by microhomology
526   sampling during homologous recombination. Cell. 2015;160(5):856-69.
527   16.   Corneo B, Wendland RL, Deriano L, Cui X, Klein IA, Wong SY, et al. Rag mutations reveal robust
528   alternative end joining. Nature. 2007;449(7161):483-6.
529   17.   Truong LN, Li Y, Shi LZ, Hwang PY, He J, Wang H, et al. Microhomology-mediated End Joining and
530   Homologous Recombination share the initial end resection step to repair DNA double-strand breaks in
531   mammalian cells. Proc Natl Acad Sci U S A. 2013;110(19):7720-5.
532   18.   Zha S, Boboila C, Alt FW. Mre11: roles in DNA repair beyond homologous recombination. Nat
533   Struct Mol Biol. 2009;16(8):798-800.
534   19.   He MD, Zhang FH, Wang HL, Wang HP, Zhu ZY, Sun YH. Efficient ligase 3-dependent
535   microhomology-mediated end joining repair of DNA double-strand breaks in zebrafish embryos. Mutat
536   Res. 2015;780:86-96.
537   20.   Schulte-Merker S, Lee KJ, McMahon AP, Hammerschmidt M. The zebrafish organizer requires
538   chordino. Nature. 1997;387(6636):862-3.

539   21.   Page-McCaw PS, Chung SC, Muto A, Roeser T, Staub W, Finger-Baier KC, et al. Retinal network
540   adaptation to bright light requires tyrosinase. Nat Neurosci. 2004;7(12):1329-36.
541   22.   Brinkman EK, Chen T, Amendola M, van Steensel B. Easy quantitative assessment of genome
542   editing by sequence trace decomposition. Nucleic Acids Res. 2014;42(22):e168.
543   23.   Zhang J, Talbot WS, Schier AF. Positional cloning identifies zebrafish one-eyed pinhead as a
544   permissive EGF-related ligand required during gastrulation. Cell. 1998;92(2):241-51.
545   24.   Seeley M, Huang W, Chen Z, Wolff WO, Lin X, Xu X. Depletion of zebrafish titin reduces cardiac
546   contractility by disrupting the assembly of Z-discs and A-bands. Circ Res. 2007;100(2):238-45.
547   25.   Xu X, Meiler SE, Zhong TP, Mohideen M, Crossley DA, Burggren WW, et al. Cardiomyopathy in
548   zebrafish due to mutation in an alternatively spliced exon of titin. Nat Genet. 2002;30(2):205-9.
549   26.   Vouillot L, Thelie A, Pollet N. Comparison of T7E1 and surveyor mismatch cleavage assays to
550   detect mutations triggered by engineered nucleases. G3 (Bethesda). 2015;5(3):407-15.
551   27.   Decottignies A. Microhomology-mediated end joining in fission yeast is repressed by pku70 and
552   relies on genes involved in homologous recombination. Genetics. 2007;176(3):1403-15.
553   28.   Biehs R, Steinlage M, Barton O, Juhasz S, Kunzel J, Spies J, et al. DNA Double-Strand Break
554   Resection Occurs during Non-homologous End Joining in G1 but Is Distinct from Resection during
555   Homologous Recombination. Mol Cell. 2017;65(4):671-84 e5.
556   29.   Hu JH, Miller SM, Geurts MH, Tang W, Chen L, Sun N, et al. Evolved Cas9 variants with broad
557   PAM compatibility and high DNA specificity. Nature. 2018.
558   30.   Ma AC, McNulty MS, Poshusta TL, Campbell JM, Martinez-Galvez G, Argue DP, et al. FusX: A
559   Rapid One-Step Transcription Activator-Like Effector Assembly System for Genome Science. Hum Gene
560   Ther. 2016;27(6):451-63.
561   31.   Bill BR, Petzold AM, Clark KJ, Schimmenti LA, Ekker SC. A primer for morpholino use in zebrafish.
562   Zebrafish. 2009;6(1):69-77.
563   32.   Schneider CA, Rasband WS, Eliceiri KW. NIH Image to ImageJ: 25 years of image analysis. Nat
564   Methods. 2012;9(7):671-5.
565   33.   R Core Team. R: A language and environment for statistical computing.: R Foundation for
566   Statistical Computing; 2016.
567   34.   Chang W CJ, Allaire JJ, Xie Y, and McPherson J. shiny: Web Application Framework for R. R
568   package version 1.0.5. 2017.
569   35.   R Studio Team. RStudio: Integrated Development for R. 2016.

570
571   **Figure Captions:**

572   **Fig 1 – MMEJ is a unique DSB repair pathway that results in highly efficient and**

573   **highly stereotyped mutagenesis.**

574   DSB repair generally begins with end resections to varying degrees. Once the ends are

575   processed, end joining in NHEJ often occurs between incompatible ends, producing

576   unpredictable and heterogeneous mutant alleles. In contrast, MMEJ uses region of

577   sequence microhomology flanking a DSB to temporarily appose the two strands. A

578   polymerase will then elongate DNA from the homology arms in a templated fashion,

579 resulting in predictable mutagenesis. Although both MMEJ and NHEJ can lead to high

580 efficiency, biallelic mutagenesis of a target gene, HDR is usually a low-efficiency,

581 monoallelic process. This is because DSB is repaired by recruiting homologous DNA as

582 a template for repair. Of the HDR pathways, HR results in a high-fidelity repair, usually

583 confined to late S ~ G2 phase because of the proximity requirement of the two

584 chromosomes. Rectangular boxes represent homology arms of varying lengths

585 indicated above them.

586

587 **Fig 2 Winner-Take-All TALEN reagent can be used to recapitulate previously**

588 **reported loss-of-*chrd*-function phenotype in 1 dpf F0, injected larvae**

589 **A** *Top* – Wildtype *chrd* sequence with TALEN binding sites annotated in teal. The dotted

590 red boxes are MH arms predicted to be used most frequently. Raw sequence alignment

591 of the whole PCR amplicon demonstrates that the majority of reads are the expected

592 7bp deletion allele. *Bottom* – summary data from subcloning analyses. 50% of the

593 mutant allele recovered were of the predicted MH allele. **B** Previously reported *chrd*

594 loss-of-function phenotype was successfully recapitulated using this TALEN pair.

595 Phenotype severity was graded by the degree of ICM expansion in the tail and by the

596 reduced head size by 1 dpf. Box plot demonstrating phenotypic penetrance is provided.

597 N = 3 biological and technical replicates. At least 29 injected animals were scored in

598 each experiment.

599

600 **Fig 3 Winner-Take-All sgRNA against *tyr* can be used to recapitulate loss-of-**

601 **melanophore phenotype in 2 dpf, injected F0 larvae injected**

602   **A** *Top* – Wildtype *tyr* sequence with the #2 sgRNA target site annotated in green. The

603   dotted red boxes are MH arms predicted to be used most frequently. Raw sequence

604   alignment of the whole PCR amplicon demonstrates that the majority of reads are the

605   expected 4bp deletion allele. *Bottom* – summary data from subcloning analyses. 88% of

606   the mutant allele recovered were of the predicted MH allele. **B** Previously reported *tyr*

607   loss-of-function phenotype was successfully recapitulated using this CRISPR-Cas9.

608   Phenotype severity was graded by the loss of retinal pigmentation. Partial loss of retinal

609   pigmentation was considered a Weak phenotype, whereas complete loss of

610   pigmentation in one or both eyes were considered Moderate and Strong phenotypes,

611   respectively. Box plot demonstrating phenotypic penetrance is provided. N = 3 biological

612   and technical replicates. At least 12 injected animals were scored in each experiment.

613

614   **Fig 4 Prospectively designed Winner-Take-All reagent against *tdgf1* can be used**

615   **to reproduce gross developmental defect in 1 dpf, injected F0 larvae**

616   **A** *Top* – Wildtype *tdgf1* sequence with sgRNA target site annotated in orange. The

617   dotted red boxes are MH arms predicted to be used most frequently. Raw sequence

618   alignment of the whole PCR amplicon demonstrates that the majority of reads are the

619   expected 4bp deletion allele. *Bottom* – summary data from subcloning analyses. 72% of

620   the mutant allele recovered were of the predicted MH allele. **B** Previously reported *tdgf1*

621   loss-of-function phenotype was successfully recapitulated using this CRISPR-Cas9.

622   Phenotype severity was graded by the "pinhead" morphology and cyclopia. Pinhead

623   morphology alone was classified as Weak, whereas Moderate and Strong phenotypes

624   also presented with varying degrees of cyclopia judged by the distance of forebrain

625 protrusion. In the Strong class, the forebrain does not separate the eyes, and they are

626 fused together. Box plot demonstrating phenotypic penetrance is provided. N = 4 with 3

627 biological and 4 technical replicates. At least 42 injected animals were scored in each

628 experiment.

629

630 **Fig 5 Winner-Take-All reagent against *ttn.2* N2B results in specific reduction of**

631 **shortening fraction in 2 dpf F0 zebrafish**

632 **A** *Top* – Wildtype *ttn.2* sequence at the N2B exon with sgRNA target site annotated in

633 red. The dotted red boxes are MH arms predicted to be used most frequently. Raw

634 sequence alignment of the whole PCR amplicon demonstrates that the majority of reads

635 are the expected 5bp deletion allele. *Bottom* – summary data from subcloning analyses.

636 86% of the mutant allele recovered were of the predicted MH allele. **B** Previously

637 reported *pickwick* phenotype was successfully recapitulated using this CRISPR-Cas9. 2

638 dpf zebrafish were immobilized in 3% methylcellulose for live recording of cardiac

639 functions. Whereas injections with Cas9 only (660pg), N2B #1 sgRNA only (300pg), or

640 *tyr* #2 sgRNA RNP (300pg sgRNA + 660pg Cas9) did not result in changes in

641 shortening fraction at this age, RNP injection containing N2B #1 sgRNA (300pg sgRNA

642 + 660pg Cas9) resulted in a specific reduction in shortening fraction by 65%. N > 3

643 biological and technical replicates. At least 5 injected animals were scored in each

644 experiment. P-values calculated using the Wilcoxon Each Pair Calculation (adjusted for

645 multiple comparisons)

646

647 **Fig 6 Winner-Take-All reagent can be used for in-frame gene alteration**

648 **A** *Top* – Wildtype *ttn.2* sequence with sgRNA target site annotated in red. The dotted

649 red boxes are MH arms predicted to be used most frequently. Raw sequence alignment

650 of the whole PCR amplicon demonstrates that the majority of reads are the expected

651 12bp deletion allele. *Bottom* – summary data from subcloning analyses. 73% of the

652 mutant allele recovered were of the predicted MH allele. **B** 2 dpf zebrafish larvae

653 injected with *ttn.2 #2* sgRNA RNP (300pg sgRNA + 660pg Cas9) grossly appear normal

654 with the exception of mild cardiac edema. Median penetrance was 50%. N = 3 biological

655 and technical replicates. At least 9 injected animals were scored in each experiment.

656

657 **Fig 7 Competition Hypothesis Version 2**

658 **A** Outlier plot summarizing repair outcomes from 47 genomic targets using TALEN and

659 CRISPR-Cas9. Close proximity of 2 MH arms (Groups 3 and 4) appears to be the

660 primary determinant for generating Winner-Take-All type outcomes as no target from

661 Groups 1 and 2 had Top MH Fraction exceeding 0.5. When the top predicted allele had

662 at least 50% higher Pattern Score than the second predicted allele (Groups 2 and 4), it

663 was a strong indicator for inducing MMEJ-class repairs. **B** *Top* Definition for each of the

664 4 groups used in Panel **A**. Each and every zebrafish genomic locus was segmented into

665 these categories. Pattern scores were derived using RGEN online tool. *Bottom* P-

666 values calculated using the Wilcoxon Each Pair Calculation (adjusted for multiple

667 comparisons). **C** Graphical representation of each group detailed in Panel A. Groups 1

668 and 2 are prone to activate NHEJ-type outcomes, presumably because the yet-

669 unidentified MMEJ factor fails to localize to suitable microhomology arm pairs, limited by

670 how far apart the arms are. Group 4 is most suitable for strong MMEJ activation

671   because it satisfies the proximity requirement AND the relative strength requirement.

672   The latter may aid in the kinetics of the yet-unidentified MMEJ factor binding to the

673   microhomology arms. Our data suggest that Group 3 is an intermediate group in terms

674   of MMEJ activation. Perhaps extragenetic factors, such as cell cycle and epigenetic

675   status may determine how favorable the loci are for MMEJ inductions.

676

677   **S1 Figure Overabundance of Microhomology arms is a negative predictor of**

678   **MMEJ activation in zebrafish**

679   **A** Box plot showing the distribution of *Slope Values* across 19 zebrafish genomic

680   targets. **B** Scatter plot of 3bp MH Fraction against *Slope Value*. Linear fit with 95%

681   Confidence Interval (shade) is shown. r^2 = 0.382, p = 0.0048. **C** Scatter plot of 2bp MH

682   Fraction against Slope Value. Linear fit with 95% Confidence Interval (shade) is shown.

683   r^2 = 0.353, p = 0.0073 Shade is 95% CI. *Pattern Scores* and *Microhomology Scores*

684   were derived using RGEN online tool (http://www.rgenome.net).

685

686   **S2 Figure Overabundance of Microhomology arms is a negative preidictor of**

687   **MMEJ activation in HeLa cell**

688   **A** Scatter plot of 3bp MH Fraction against *Slope Value* from the first 50, alphabetically

689   sorted HeLa cell targets. Linear fit with 95% Confidence Interval (shade) is shown. r^2 =

690   0.339, p = 0.0001. **B** Scatter plot of 2bp MH Fraction against *Slope Value* from the first

691   50, alphabetically sorted HeLa cell targets. Linear fit with 95% Confidence Interval

692   (shade) is shown. r^2 = 0.034, p = 0.2644. **C** Box plot showing the distribution of Slope

693   Values across the first, alphabetically sorted HeLa cell targets. **D** Box plot showing the

694 3bp MH Fractions for High and Low competition sites amongst the remaining 40 HeLa

695 cell targets. p = 0.011. Targets with < 20% overall edit efficiency were excluded in all

696 panels. *Pattern Scores* and *Microhomology Scores* were derived using RGEN online

697 tool (http://www.rgenome.net).

698

699 **S3 Figure Microhomology allele generated by *ttn.2* N2B sgRNA #1 is germline**

700 **transmitted**

701 Agarose gel showing PCR amplicon post Surveyor digest. 752bp band is the whole

702 amplicon. The expected bands due to mutations at the CRISPR site are denoted by

703 yellow arrowheads. The red asterisk denotes positive digest band due to a background

704 T -> A SNP at position 389 from the 5' end of the amplicon. Heterozygous animals are

705 bolded and underlined. Genotypes of the first 4 heterozygous progenies from each

706 founder were ascertained by subcloning analyses.

707

708 **S4 Fig Fitting Competition Hypothesis Version using HeLa cell dataset**

709 Outlier plot summarizing repair outcomes from 90 genomic targets using CRISPR-Cas9.

710 Similar to the findings in zebrafish, close proximity of 2 MH arms (Groups 3 and 4)

711 appears to be the primary determinant for utilizing this MH pair efficiently. When the top

712 predicted allele had at least 50% higher *Pattern Score* than the second predicted allele

713 (Groups 2 and 4), median Top MH Fractions trended higher compared to Group 1 and

714 3, respectively. P-values calculated using the Wilcoxon Each Pair Calculation (adjusted

715 for multiple comparisons). Targets with < 20% overall edit efficiency were excluded from

716   analysis. *Pattern Scores* and *Microhomology Scores* were derived using RGEN online

717   tool (http://www.rgenome.net).

718

719   **S1 Table**

720   List and summary mutagenic outcomes of TALEN and CRISPR-Cas9 reagents that

721   were designed primarily using the Bae *et al.* algorithm(14). Underlined & italicized bases

722   in sgRNA sequence denote mismatched bases due to the promoter requirement.

723   *Pattern Scores* and *Microhomology Scores* were derived using RGEN online tool

724   (http://www.rgenome.net).

725   MH: Microhomology, SC: Subcloning

726   *Reagents prospectively designed according to Bae *et al.* algorithm(14).

727   †No raw sequencing data were available. However, the outcome had been compiled

728   into a table prior to conception of this study.

729   ‡Injected with sgRNA and Cas9 mRNA (150 pg and 100 pg, respectively)

730   ^gift from Wenbiao Chen (Addene # 46761).

731

732   **S2 Table**

733   Summary gross phenotyping outcomes from Winner-Take-All reagent injections. For

734   *tdgf1*, Experiments 1a and 1b correspond to technical replicates using WT 1 as

735   reference, uninjected control. *chrd* and *tdgf1* phenotypes were scored on 1 dpf, whereas

736   *tyr, ttn.2* N2B, *ttn.2* phenotypes were scored on 2 dpf.

737

738   **S3 Table**

739   List and summary sequence outcomes of Low Competition sgRNA that were designed

740   around the Competition Hypothesis. Underlined & Italicized bases in gRNA sequence

741   denote mismatched bases due to the promoter requirement. *Pattern Scores* and

742   *Microhomology Scores* were derived using RGEN online tool (http://www.rgenome.net).

743   MH: Microhomology, SC: Subcloning, TIDE: Tracking Indels by DEcomposition.

744   [†] injected RNP at the dose of 115 pg sgRNA and 245 pg Cas9 due to poor viability at

745   higher doses

746

747   **S4 Table**

748   List and summary sequence outcomes of Medium ~ High Competition sgRNA that were

749   designed around the Competition Hypothesis. Underlined & Italicized bases in sgRNA

750   sequence denote mismatched bases due to the promoter requirement. *Pattern Scores*

751   and *Microhomology Scores* were derived using RGEN online tool

752   (http://www.rgenome.net).

753   MH: Microhomology, SC: Subcloning, TIDE: Tracking Indels by Decomposition.

754   [†]injected RNP at the dose of *115 pg sgRNA and 245 pg Cas9* due to poor viability at

755   higher doses

756

757   **S5 Table**

758   Sample MENTHU output from select CRISPR-Cas9 targets used in this study. The

759   output was obtained by using the entire target exon sequence with 40bp intronic

760   sequence each on both 5' and 3' ends. The MENTHU output provides a 3' NGG PAM

761   sequence for each gRNA targets (italicized and underlined). MENTHU gRNA outputs

762   that matched the target sequences used in this study are bolded. *Criteria 1* and *2* refer

763   to 1) if top predicted microhomology arm is separated by 5bp or less, and 2) if the ratio

764   of top to second predicted pattern scores is at least 1.5. MENTHU is programmed to

765   terminate calculations if the target site is negative for Criterion 1. As a result, no gRNA

766   sequence output is obtained for *chrd* #1 and *mitfa* #2. Importantly, in two instances

767   (*surf1* and *tgdf1*) where we only had Group 3 reagents, novel candidate Winner-Take-

768   All sites were identified.

769

770   **S6 Table**

771   List of primers used in this study. All the primer sequences are provided in 5' → 3'

772   order. For *urod* Reverse primer, M13F primer sequence was added at the 5' end of the

773   endogenous target sequence (bolded and italicized). For SDM primers, intended point

774   mutation is indicated by bold and italic.

775

A
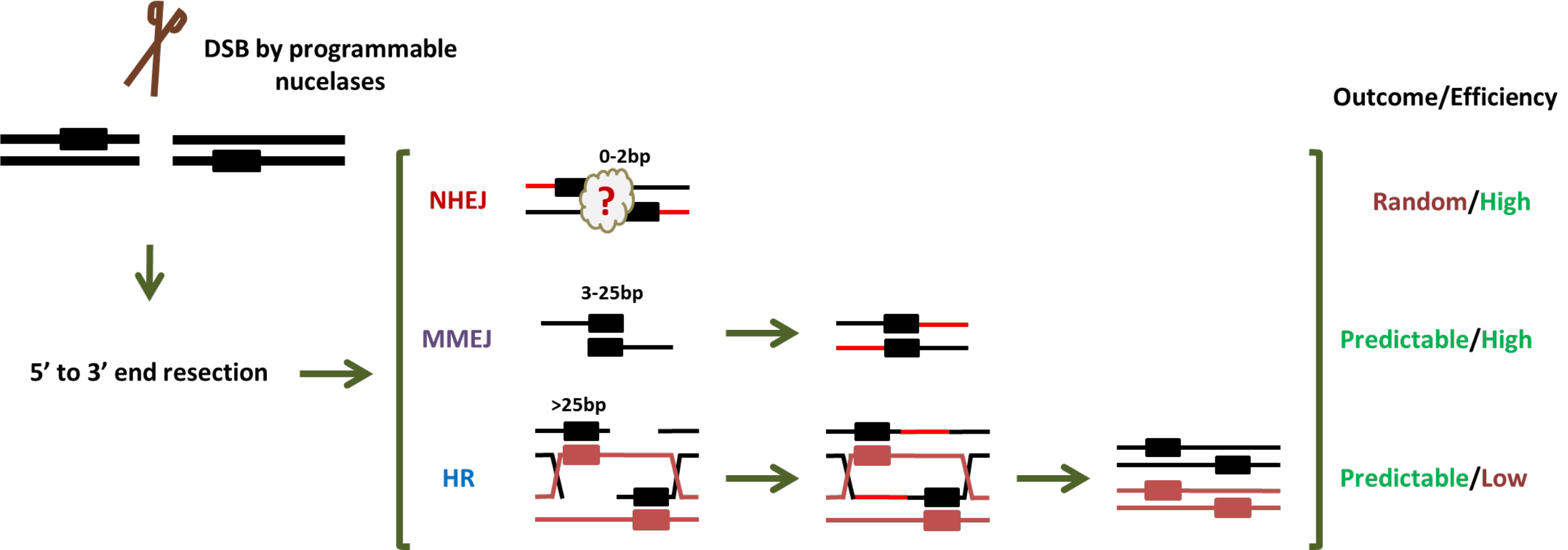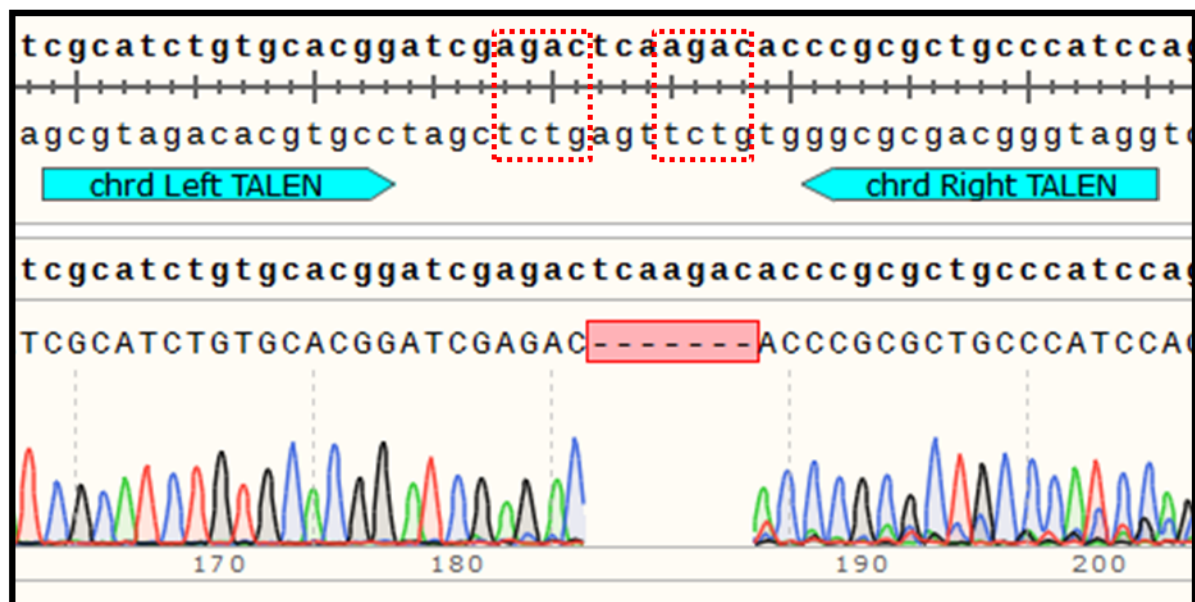
	Mutation	MH used	Number of Colonies
WT	–	–	3
Mut #1	7bp del	AGAC	16
Mut #2	7bp del	N/A	2
Mut #3	11bp del/19 ins	N/A	2
Others	–	–	12

B
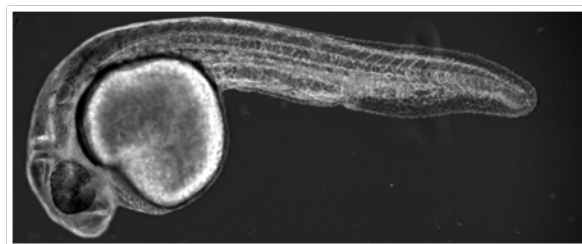
Penetrance

A

	Mutation	MH used	Number of Colonies
WT	–	–	0
Mut #1	4bp del	CTCA	21
Others	–	–	3

B
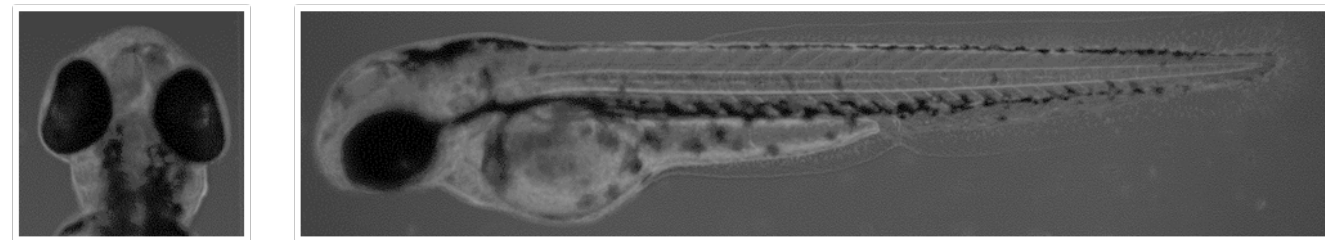
Penetrance

A

	Mutation	MH used	Number of Colonies
WT	–	–	0
Mut #1	4bp del	CGTT	28
Mut #2	13bp del	–	4
Mut #3	2bp del/4bp ins	–	3
Mut #4	13bp del	–	2
Others	–	–	2

B

Penetrance

A
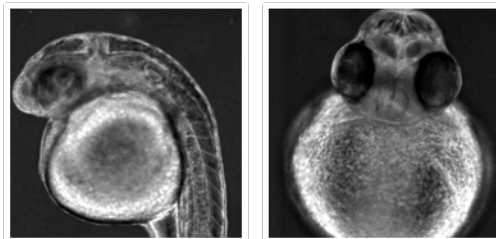
	Mutation	MH used	Number of Colonies
WT	–	–	1
Mut #1	5bp del	CTGTT	18
Mut #2	8bp del	TGT	2
Others	–	–	1

B
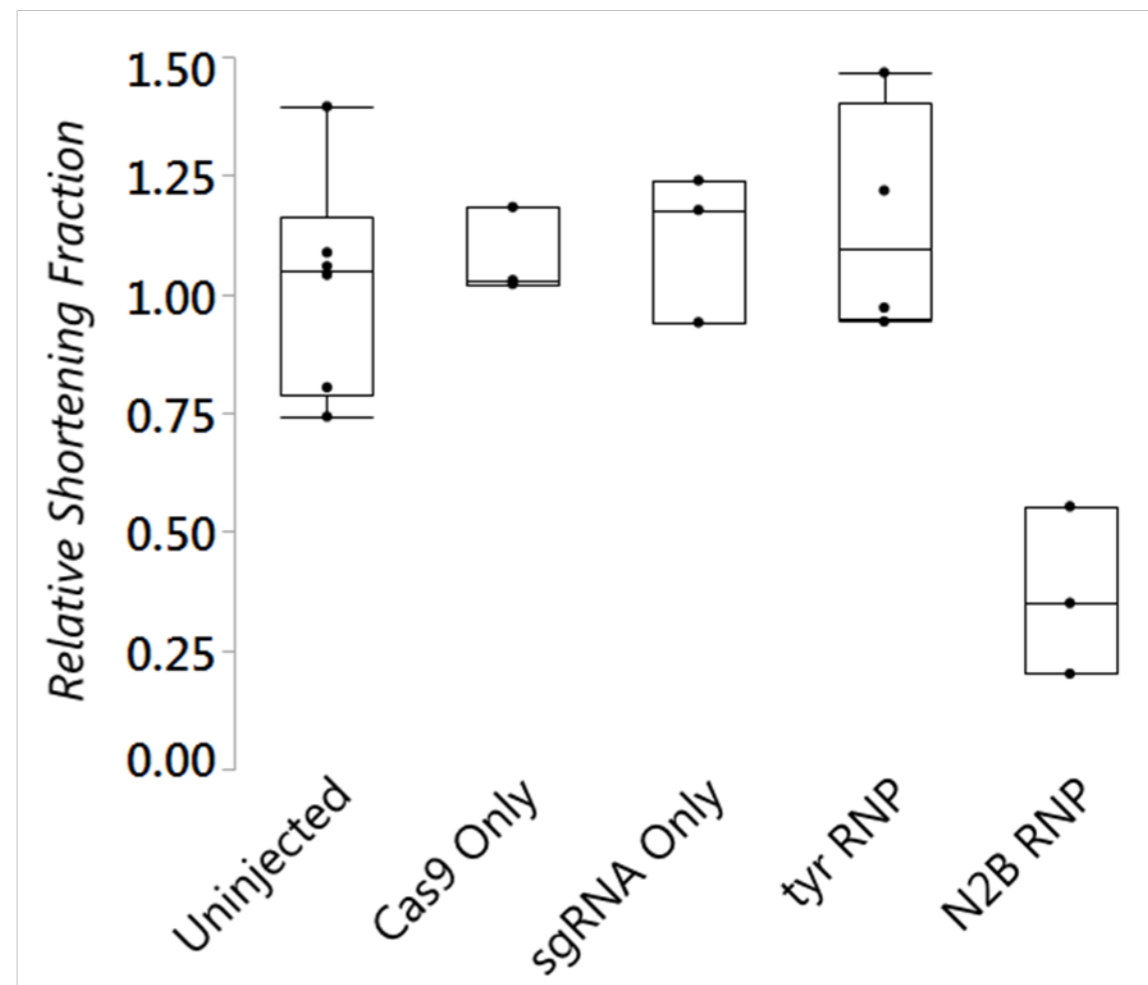
p-Values	Cas9	sgRNA	tyrRNP	N2B RNP
Uninj	1.0000	0.5186	0.5940	0.0282
Cas9		1.0000	1.0000	0.0809
sgRNA			0.8597	0.0809
tyr RNP				0.0518

A

	Mutation	**MH used**	**Number of Colonies**
WT	–	–	0
Mut #1	12bp del	GACCCCA	16
Others	–	–	6

B

Uninjected
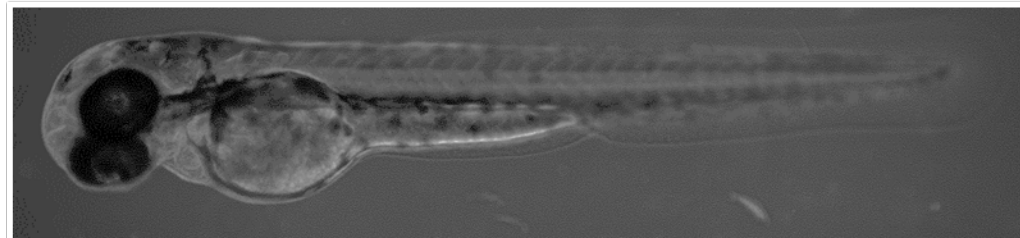
Injected — WT
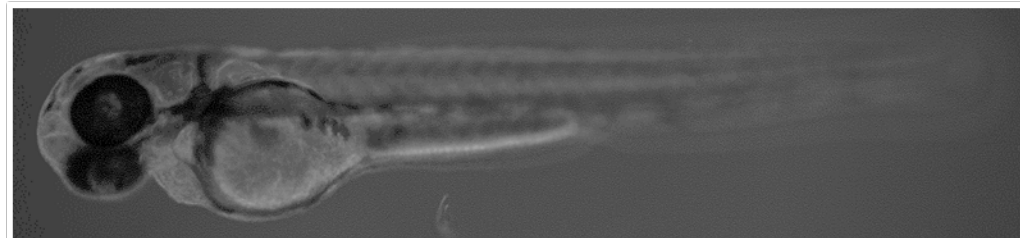
Injected — Cardiac Edema

A
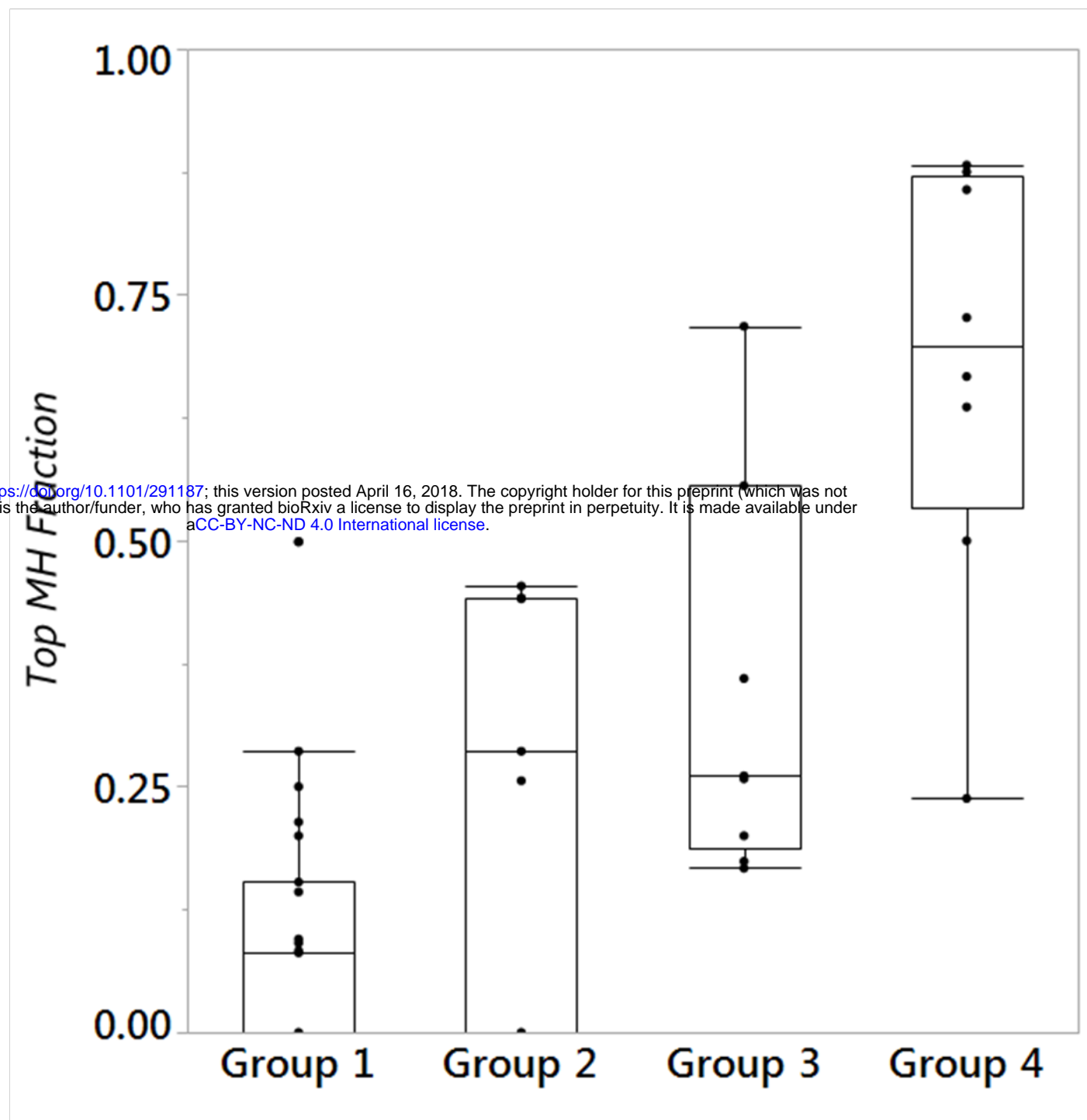
		Distance between the top MH arm pair	
		> 5bp	≤ 5bp
Top/Second Pattern Score ratio	< 1.5	Group 1	Group 3
	≥ 1.5	Group 2	Group 4

p-Values	Group 1	Group 2	Group 3	Group 4
Group 1		0.0489	0.0004	< 0.0001
Group 2			0.5961	0.0092
Group 3				0.0183
Group 4				

C