Deep learning image recognition enables efficient genome editing in zebrafish by automated injections

Maria Lorena Cordero-Maldonado[1][¶][,*], Simon Perathoner[1][¶][,#a], Kees-Jan van der Kolk[2][¶],

Ralf Boland[3], Ursula Heins-Marroquin[1], Herman P. Spaink[3], Annemarie H. Meijer[3],

Alexander D. Crawford[1][,#b], Jan de Sonneville[2][,*]

[1] Luxembourg Centre for Systems Biomedicine, University of Luxembourg, Belvaux, Luxembourg

[2] Life Science Methods BV, Leiden, the Netherlands

[3] Institute of Biology, Leiden University, Leiden, the Netherlands

[#a] Current address: Max Planck Institute for Heart and Lung Research, Bad Nauheim, Germany

[#b] Current address: Faculty of Veterinary Medicine, Norwegian University of Life Sciences, Oslo, Norway

* Corresponding authors

E-mail: marialorena.corderomaldonado@uni.lu (MLCM),

jan@lifesciencemethods.com (JdS)

¶These authors contributed equally to this work

Abstract

One of the most popular techniques in zebrafish research is microinjection, as it is a rapid and efficient way to genetically manipulate early developing embryos, and to introduce microbes or tracers at larval stages.

Here we demonstrate the development of a machine learning software that allows for microinjection at a trained target site in zebrafish eggs at unprecedented speed. The software is based on the open-source deep-learning library Inception v3.

In a first step, the software distinguishes wells containing embryos at one-cell stage from wells to be skipped with an accuracy of 93%. A second step was developed to pinpoint the injection site. Deep learning allows to predict this location on average within 42 µm to manually annotated sites. Using a Graphics Processing Unit (GPU), both steps together take less than 100 milliseconds. We first tested our system by injecting a morpholino into the middle of the yolk and found that the automated injection efficiency is as efficient as manual injection (~ 80%). Next, we tested both CRISPR/Cas9 and DNA construct injections into the zygote and obtained a comparable efficiency to that of an experienced experimentalist. Combined with a higher throughput, this results in a higher yield. Hence, the automated injection of CRISPR/Cas9 will allow high-throughput applications to knock out and knock in relevant genes to study their mechanisms or pathways of interest in diverse areas of biomedical research.

Introduction

Microinjection is one of the most powerful techniques used in zebrafish (*Danio rerio*), as it allows to follow cell fate [1], evaluate pathogenesis of bacteria [2], produce

2

56   chimeric individuals [3], study tumour progression [4,5], manipulate protein levels [6,7]

57   and create genetically altered lines [8]. The intrinsic biological properties of zebrafish

58   make it particularly amenable to this technique, since these cyprinids are highly

59   fecund, a spawning pair typically producing more than 400 eggs at a time. Moreover,

60   fertilization is external and spawning is confined to a brief period at dawn (natural or

61   artificial), allowing for timing of the experiments. Furthermore, the chorion of zebrafish

62   eggs is supple and easy to pierce.

63   Classically, injection of tracer dyes is used to identify single cell populations [9,10], to

64   follow cell lineages and to build fate maps in zebrafish [1,11]. The development of

65   molecular methods for the zebrafish model enabled functional studies by manipulating

66   the expression of specific genes. Injection of messenger RNA (mRNA) can be used to

67   overexpress and misexpress a specific protein [12], while morpholino antisense

68   oligonucleotides (MOs) can be employed to knock down a given target gene [13].

69   Morpholinos are synthetic oligonucleotides designed to be complementary to a

70   specific RNA target. Typically they target the translation start site blocking translation

71   initiation. Alternatively, MOs can also be designed to mask splicing sites, hindering the

72   proper processing of precursor mRNA. In zebrafish mRNA and MO injections are

73   simply performed by introducing a fine-tipped needle into the yolk of one-cell stage

74   eggs and delivering nanoliter volumes of the injection material into it [14]. As

75   cytoplasmic streaming will move the mRNA or MOs into the cytoplasm, it is not

76   necessary that the injection targets the cell. While injection into the yolk requires some

77   skill, it can usually be learned within a few weeks. Nevertheless, injections of mRNAs

78   and MOs have their drawbacks. First of all, the effect is only transient, *i.e.* the injected

79   molecules will be degraded and/or diluted with time. Moreover, in the case of mRNA

80   injection, tissue-specific upregulation is not possible and a given mRNA will be

81   expressed in all tissues indiscriminately. Also, the specificity of MO antisense

82   technology has recently been questioned. Indeed, MOs can sometimes lead to

83   misleading results, in many cases due to toxicity and off-target effects, but also due to

84   the difficulties in estimating their efficacy and controlling their dosage [15]. In a recent

85   study [16], loss-of-function mutations for ten different genes previously thought to have

86   an essential role in development failed to recapitulate the corresponding morpholino-

87   induced phenotypes. In several cases, the discrepancy between mutant and morphant

88   phenotypes, could be explained by genetic compensation mechanisms that occur in

89   mutants [17], however, undoubtedly rigorous controls are required to ascertain the

90   reliability of MO-induced phenotypes [15,18,19].

91   In the last years, with the implementation of targeted nuclease techniques in the

92   zebrafish, the demand for genetic evidence to define gene function has greatly

93   increased. Fortunately, after a somewhat slow start using zinc-finger nucleases

94   (ZFNs) [20] and transcription activator-like effector nucleases (TALENs) [21], the

95   adaptation of the prokaryotic CRISPR/Cas9 (clustered regularly interspaced short

96   palindromic repeats/CRISPR associated protein 9) defence system to engineer

97   genomes [22] has revolutionized reverse genetics in zebrafish.

98   The CRISPR/Cas9 method builds on the type II CRISPR system of the adaptive

99   immunity of certain archaea and bacteria [23]. When infected by a bacteriophage,

100   these prokaryotes respond by integrating fragments of the viral genome into clustered

101   regularly interspaced short palindromic repeat (CRISPR) loci. These loci are

102   subsequently transcribed and processed to CRISPR RNAs (crRNAs). crRNAs form a

103   base-paired structure with a transactivating CRISPR RNA (tracrRNA) to recruit

104   CRISPR associated protein 9 (Cas9) [22]. This nuclease is guided to the viral genome

105   by the RNA complex and the target sequence is subsequently cleaved, neutralizing

106   the invading entity. The cleavage site of the Cas9 nuclease is determined by a short

107   sequence adjacent to the target DNA and by sequence complementarity of the crRNA

108   with the target site itself.

109   Recently the system was adapted and optimised to engineer genomes. crRNA and

110   tracrRNA can be replaced by a single synthetic guide RNA (gRNA) that directs Cas9-

111   mediated cleavage of target DNA [22,24], and the method was implemented in multiple

112   systems including zebrafish [25,26], finally paving the road for knock-ins in this model

113   [27]. Along with *Tol2* mediated transgenesis, a transposon system based on the *Tol2*

114   element of medaka (*Oryzias latipes*) widely used in zebrafish to create transgenic lines

115   [8], the CRISPR/Cas9 system has become an essential tool for genome editing in

116   zebrafish. In this context microinjection is an essential technique. For the creation of

117   genetically altered lines in zebrafish, be it through *Tol2* transgenesis or by means of

118   zinc finger nucleases, TALEN or Cas9 nucleases, it is critical to inject the solution

119   directly into the blastomere at the one-cell stage or at least at the interface between

120   blastomere and yolk [28-31]. Contrary to RNA or MOs, DNA appears not to be

121   transported into the blastomere via cytoplasmic streaming. Moreover, efficiency of all

122   these genome editing techniques is much lower compared to mRNA or MOs injections.

123   Therefore, in order to create genetically altered zebrafish lines it is essential to master

124   microinjections into the cell. This can be challenging as this type of injection requires

125   long training and excellent technical skills.

Automated microinjection system

128   The first version of our automated microinjection system featured half-spherical wells,

129   moulded in agarose gel, which allowed for high-throughput microinjection into the yolk

130   of zebrafish eggs. This was used for microinjection of bacteria, morpholinos [32] and

131   cancer cells [33]. The great advantage over other systems was the higher batch size

5

132   and speed of the injections. As the initial cell division steps in zebrafish embryos occur

133   in intervals of 20-40 minutes, speed is crucial for the accuracy, reproducibility and

134   number of experiments.

135   In our experience, it is apparent that injections into the middle of the yolk are less

136   suitable for DNA injections. Therefore as a first step, the program "click-to-inject" was

137   developed to test the efficiency of injections closer to the first cell [33]. With this, we

138   noticed that we could achieve a great increase in efficiency, similar to manual

139   injections done into the first cell. Therefore, we set out to automate this procedure.

140   In this study we demonstrate the results of autonomous site selection and injection for

141   CRISPR/Cas9 and DNA manipulation of the zebrafish genome.

142   # Materials and methods

143

144   ## Animals

145

146   At the Luxembourg Centre for Systems Biomedicine, wild type adult zebrafish (AB

147   strain) are maintained in the Aquatic Facility according to standard protocols [34].

148   Zebrafish eggs were obtained by natural spawning on the day of each experiment,

149   kept in 0.3X Danieau's solution (14 mM NaCl, 2 mM KCl, 0.12 mM $MgSO_4$, 1.8 mM

150    $Ca(NO_3)_2$, 1.5 mM HEPES pH 7.5 and 0.03 M methylene blue), and staged by

151   morphology (one-cell stage) for the injections. After each series of injections, the eggs

152   were incubated at 28 °C (±0.5) and evaluated daily until 5 days post-fertilization (dpf).

153   At the Institute of Biology, Leiden University, wild type adult zebrafish (strain AB/TL)

154   are maintained and handled according to standard protocols [34]. Zebrafish eggs were

155   obtained by natural spawning on the day of each experiment, and after each series of

156   injections, the eggs were incubated at 28.5 °C in egg water (60 µg/ml sea salt, Sera

157   Marin, Heinsberg, Germany) and evaluated at 6 hours post-fertilization (hpf), 1 dpf and

158   5 dpf. Anaesthesia of larvae used for live imaging and COPAS [32,35] analysis was

159   done with 0.02% buffered Tricaine (3-aminobenzoic acid ethyl ester, Sigma-Aldrich)

160   in egg water.

Ethics statement

163   All procedures with zebrafish were performed in accordance with European laws,

164   guidelines and policies for animal experimentation, housing and care (European

165   Directive 2010/63/EU on the protection of animals used for scientific purposes). The

166   operation of the Aquatic Facility at the Luxembourg Centre for Systems Biomedicine

167   is allowed by the relevant agencies of the Government of Luxembourg (authorization

168   through the Grand-Ducal decree of 20 January 2016). Maintaining and handling of

169   adult zebrafish at the Institute of Biology, Leiden University, is in compliance with the

170   animal welfare directives of the local animal welfare committee.

Morpholino antisense oligonucleotide

173   The translation blocking morpholino for *slc45a2* (solute carrier family 45 member 2)

174   was obtained from Gene Tools according to Dooley *et al*., 2012 [36] with the following

175   sequence: 5'-GCTGGTCCTCAGTAAGAAGAGTCAT-3'. In addition, a 3' fluorescein

176   modification was included, which allowed fluorescent differentiation of injected eggs.

177   A standard MO with sequence 5'-CCTCTTACCTCAGTTACAATTTATA-3' was used

178   as an injection control. In both cases, stock solutions (1 mM ~ 8 ng/nL) were prepared

179   according to the specifications of the provider and titrated working solutions were

180   freshly prepared for each experiment.

Preparation of Cas9 mRNA and *slc45a2* sgRNA

183   Both *slc45a2* sgRNA and Cas9 mRNA were prepared according to Gagnon *et al*.,

184   2014 [30]. Briefly, the *slc45a2* DNA template was synthetized with T4 DNA polymerase

185   (New England BioLabs) using the oligonucleotides: *slc45a2*-specific

186   (taatacgactcactataGGTTTGGGAACCGGTCTGATgttttagagctagaaatagcaag) and

187   constant (AAAAGCACCGACTCGGTGCCACTTTTTCAAGTTGATAACGGACTAGCC

188   TTATTTTAACTTGCTATTTCTAGCTCTAAAAC). The sgRNA was synthetized using

189   T7 RNA polymerase (Ambion MEGAscript) and then diluted to 400 ng/µl. Cas9 mRNA

190   was synthetized using the pCS2-Cas9 plasmid [37], transcribed using the SP6

191   mMessage mMachine kit (Ambion) and finally diluted to 600 ng/µl.

192   Manual microinjections of *slc45a2*-MO and *slc45a2*

193   sgRNA/Cas9

194

195   Manual microinjections of zebrafish embryos were done as described by Rosen *et al*.

196   2009 [14]. Briefly, gene knockdown was achieved through manual microinjection of 2

197   nL of *slc45a2*-MO into the center of the yolk of wild-type zebrafish embryos at one- to

198   two-cell stage using an Eppendorf FemtoJet 4X® microinjector. We titrated the amount

199   of the *slc45a2*-MO to 3.7 ng per injection, and the same amount was used for the

200   control MO. On the other hand, gene editing was achieved through manual

201   microinjection of 4 nL of *slc45a2*-sgRNA, Cas9 and phenol red in a single mix into the

202   cell cytoplasm of wild-type zebrafish embryos at one-cell stage using an Eppendorf

203   FemtoJet 4X® microinjector. After each series of microinjections, the eggs were

204   incubated at 28 °C (±0.5) in 0.3X Danieau's medium and evaluated daily until 5 dpf to

205   record non-viable embryos, (*i.e.* non-fertilized, fluorescent negative at 6 hpf for MO-

206   injections, dead and dysmorphic embryos/larvae from 1 to 5 dpf), and the efficiency of

207   injection displayed by the albino phenotype.

Manual microinjection of DNA

Manual microinjection of zebrafish embryos was done as described previously using standard methods [38] using an Eppendorf FemtoJet microinjector. Briefly, a DNA construct containing a GFP fusion gene expressed under a constitutive promoter (*actb*:-NLSmCherry-IRES-GFP) in the Gateway *Tol2* vector was co-injected together with *Tol2* transposase RNA in a volume of 1 nL at a concentration of 25 ng/µL for both the DNA construct and the *Tol2* transposase RNA. Injections were performed at the one-cell stage at the blastomere/yolk boundary. After each series of injections, the eggs were incubated at 28.5 °C in egg water and evaluated at 6 hpf and 1 dpf to record non-viable embryos, and at 5 dpf for fluorescent signal.

Automated microinjection of *slc45a2*-MO

Automated microinjection of zebrafish embryos using the robotic injector (Life Science Methods BV) was done following guidelines described in Spaink *et al*. 2013 [33]. Briefly, gene knockdown was achieved through fully automated microinjection of 2 nL of *slc45a2*-MO (3.7 ng per injection) into the yolk of wild-type zebrafish embryos at one- to two-cell stage. A 1% agarose covered grid (9 blocks x 100 wells) was used to carefully arrange the zebrafish embryos, and then placed in the motorized stage coupled to a controlled and motorized micro-manipulator. The needle that was first loaded with the sample and calibrated for a microinjection volume of 2 nL was then placed into the micro-manipulator (Eppendorf FemtoJet). After the robotic injector was properly set (position of grid and needle) automated injection occurred at a speed of approximately 100 embryos per 160 seconds (1 block in the grid). All components of the robotic injector are connected to a controlling computer that is equipped with a software control program written in Python. After each series of microinjections, the

9

234 eggs were incubated at 28 °C (±0.5) in 0.3X Danieau's medium and evaluated daily

235 until 5 dpf to record non-viable embryos, (*i.e.* non-fertilized and fluorescent negative

236 at 6 hpf, dead and dysmorphic embryos/larvae from 1 to 5 dpf), and the efficiency of

237 injection displayed by the albino phenotype.

Automated microinjection of *slc45a2* sgRNA/Cas9

240 Automated microinjection of zebrafish embryos using the robotic injector (Life Science

241 Methods BV) was done following guidelines described in Spaink *et al*. 2013 [33].

242 Briefly, microinjection of 4 nL of *slc45a2*-sgRNA, Cas9 and phenol red in a single mix

243 was done in zebrafish embryos at one-cell stage. A 1% agarose covered grid (9 blocks

244 x 100 wells) was used to carefully arrange and orient one by one the zebrafish

245 embryos so that the first cell is visible. The grid was then placed in the motorized stage

246 coupled to a controlled and motorized micro-manipulator. The needle that was first

247 loaded with the sample and calibrated for a microinjection volume of 4 nL was then

248 placed into the micro-manipulator (Eppendorf FemtoJet). After the robotic injector was

249 properly set (position of grid and needle) the automated cell injections took place at a

250 speed of approximately 100 embryos per 2 minutes and 30 seconds (1 block in the

251 grid). All components of the robotic injector are connected to a controlling computer

252 that is equipped with a software control program written in Python. This allowed to

253 have the total count of the type of microinjections performed depending on the image

254 classification (see section below), *i.e.* "Close-to-cell" #72, "Center" #13, "Empty" #1,

255 "Two-Cell" #14, and "Sick" #0, resulting in 100 injected eggs. After each series of

256 microinjections, the eggs were incubated at 28 °C (±0.5) in 0.3X Danieau's medium

257 and evaluated daily until 5 dpf to record non-viable embryos, (*i.e.* non-fertilized at 6

258 hpf, dead and dysmorphic embryos/larvae from 1 to 5 dpf), and the efficiency of

259 injection displayed by the albino phenotype.

Deep learning algorithm for image classification

As a first step we used the Inception v3 network to learn to distinguish between five different categories: "Empty", "No-Cell", "First-Cell", "Two-Cell", "Sick" (this term is used to refer to non-viable eggs). We used a total of 11,000 annotated images. To prevent overfitting, we artificially increased the number of training samples by performing four types of image transformation: *1)* rotations about the center of the image; *2)* zooming by a factor 0.9-1.1; *3)* shifting by 28 pixels orthogonally in the +/- x and y direction, and *4)* flipping the image horizontally. The neural network architecture consisted of: *1)* the top part of the Inception v3 network (containing all inception blocks); *2)* a 2D global spatial average pooling layer; *3)* a fully connected layer of 1024 nodes with ReLU activation function, and *4)* a fully connected layer of 5 nodes, with softmax activation function. Training of the classification step was done using the Adam stochastic optimizer [39], with a learning rate of 10^{-4}. For a more in-depth description see the supporting information (S1).

Deep learning algorithm for finding the injection site

For the injection point determination, we translated the (x y) coordinates to a vector in a triangular mesh using a barycentric coordinate system. We let the outputs of the neural net correspond to vertices in the mesh. In our case, we used 160 vertices.

The neural network architecture consists of: *1)* the top part of the Inception v3 network (containing all inception blocks); *2)* a 2D global spatial average pooling layer; *3)* a fully connected layer of 1024 nodes with ReLU activation function, and *4)* a fully connected layer of 160 nodes, with softmax activation function. We used 2724 images for training and 674 images for validation (these are the same images as used for label "first-cell" in the classification step). Training of the injection point determination step was done

286   using the Adam optimizer, with a learning rate varying from 10^{-3} to 10^{-5}. More details

287   can be found in the supporting information (S1).

Software and hardware

290   For deep learning and robot control we used a Shuttle SZ170R8 equipped with an Intel

291   Core i3 6100 CPU, 16 GB kit Kingston DDR4 2133Mhz, ECC memory and an NVidia

292   GeForce GTX 1070 GPU. Installed software are: Keras 1.2.2, Theano 0.9.0, NumPy

293   1.11.0, SciPy 0.17.0. For the analysis of the data, raw data for all the series of

294   microinjections was processed in excel. Statistical analysis was done using excel and

295   GraphPad Prism 6 followed by unpaired *t*-test with Welch's correction for single

296   comparisons (when applicable). The criterion for statistical significance was $P < 0.05$.

297   Graphs were plotted using GraphPad Prism 6 and error bars on all graphs represent

298   standard deviation.

Microscopy and fluorescent analysis

301   At the Institute of Biology, Leiden University, representative pictures were taken using

302   a Leica M205 FA stereo fluorescence microscope equipped with a DFC345 FX

303   monochrome camera. Fluorescent signal was quantified using a Complex Object

304   Parameter Analyzer and Sorter (COPAS, Union Biometrica). At the Luxembourg

305   Centre for Systems Biomedicine, fluorescent sorting of fluorescein positive embryos

306   (for *slc45a2*-MO injections) was done using a Nikon SMZ25 stereomicroscope.

307   Representative pictures of control larvae and injected larvae displaying an albino

308   phenotype were taken using the Nikon SMZ25 stereomicroscope equipped with a

309   Nikon Digital Sight DS-Ri1 camera.

Results and discussion

Manual and automated injections of *slc45a2*-MO

In order to test MO efficiency of manual and automated injections we employed a translation-blocking MO against *slc45a2* (solute carrier family 45 member 2). Downregulation of this gene induces albino and/or hypo-pigmented morphants, as the melanophores are unable to produce melanin [36]. Manual and automated yolk microinjections were performed in parallel, and in both cases the induced albino phenotype was assessed in larvae at 3 dpf. The results obtained with both microinjection approaches are comparable and show that downregulation of *slc45a2* is highly efficient using morpholino antisense technology (Fig 1A). Additionally, the manual injections were performed by two different experimentalists (Fig 1B) and this shows that efficiency and variation of efficiency obtained by manual morpholino injections differs from person to person and, surprisingly, the variation of the efficiency of the automated injections is slightly larger.

Fig 1. **Morpholino knockdown efficiency with manual and automated injections**
A. The survival and knockdown efficiency of *slc45a2*-MO manual and automated (auto) microinjections were measured as the number of larvae displaying an albino phenotype at 3 days post-fertilization (dpf). Control-MO injected larvae and uninjected larvae were processed in parallel and the resulting pigmented (wild-type) larvae were also counted at 3 dpf. "n=" indicates the number of eggs used to obtain this cumulative result. **B.** Efficiency comparison between the automated injection into the yolk and manual injections performed by two independent experimentalists (not statistically significant). "n=" indicates the number of experiments used to calculate the average and standard deviation.

Semi-automated "click-to-inject"

After demonstrating that automated injection into the yolk is an efficient way to

339 generate morphants, we sought to apply the robotic injector for generating

340 CRISPR/Cas9 mutants for *slc45a2*. To investigate the dependence on the injection

341 location we used the "click-to-inject" program [33] to test the efficiency of injections

342 closer to the first cell. In the "click-to-inject" program the injection depth is set, but the

343 (x y) position is chosen by the operator. To inject, the operator moves the mouse

344 pointer to a specific site (*e.g.* the first cell) and clicks to trigger an injection and a

345 subsequent movement to the next egg. Based on this, next we set out to develop an

346 automated image recognition to more precisely identify the first cell and to automate

347 CRISPR/Cas9 injections.

348 Imaging conditions

349

350 In manual microinjection setups, as well as in standard microscopy, near-perfect

351 imaging conditions are applied with lighting from the bottom and imaging from the top,

352 or vice versa. As the zebrafish egg is very transparent, epi illumination from below is

353 not suitable; most contrast and edges are then lost. As the egg is spherical, a ring-

354 light displays a very bright circle on top of the egg. Therefore, to obtain better and

355 more reproducible imaging conditions in different locations, we placed a large (L x B

356 = 60 x 80 cm^2) diffuse light source above the robotic injector. Five different classes

357 were used to annotate the images (Fig 2). In the "Inject" class the ideal injection

358 position for automated microinjection is also annotated. Instead of injecting directly

359 into the zygote, we have chosen to inject in the yolk, close to the visible zygote. The

360 reason is that injections in a thin zygote (less ideal orientation, or very early stage)

361 would often cause a rotation of the egg, and bounce the needle off. Injections into the

362 yolk-blastomere boundary almost never show this problem, and thus gave a higher

363 yield.

364

14

365 **Fig 2. Imaging classification for injection**

366 Representative digital images measured from below of an agarose grid ("Empty") that supports

367 zebrafish eggs with the first cell visible ("Inject") or not visible ("No cell"), eggs in a two- or

368 higher cell stage ("Two cell") or non-viable eggs ("Sick"). In the "Inject" image an injection

369 location is indicated by a black dot with (x y) coordinates.

Machine learning

370

371

372 Initially, we tried a classic approach of machine vision on these images. The Hough

373 circle transform [40] allowed us to detect the yolk with an above 90% accuracy (data

374 not shown). However, the next step to find the first cell was problematic. In cases

375 where the shadow of the micromanipulator overlapped with the first cell, the edge

376 detection algorithm failed. As an alternative to edge detection, we annotated a

377 database of images with injection positions. We used a Fast Fourier Transform (FFT)

378 algorithm to find a closest matching egg in this database and used that image to infer

379 an injection position. This worked reasonably well with a peak error (distance between

380 calculated position and annotated position) of 20 μm. However, when a good match

381 could not be found, the error was quite high, and as a result the tail of the error was

382 quite large (data not shown). An explanation for the large variation in results is that

383 there is also a large variation in first cell shapes, especially when looking from an

384 arbitrary angle. It can be an early very thin line up to about a third of the yolk depending

385 on the developmental stage and orientation of the egg. To overcome this variation, we

386 could make the annotated database larger, to increase the chance of a close match.

387 Nevertheless, the downside of this solution is that more images have to be compared,

388 and this takes more processing time during injection. Thus, we sought to apply a better

389 approach based on deep learning.

390

391 Using a database of annotated images as input, one can also train a deep learning

392 network. Instead of comparing images during runtime, one trains an algorithm that is

15

393 afterwards used to interpret new images. The execution time of this algorithm is

394 independent of the size of the training image set. Thus, roughly speaking, the larger

395 the number of annotated images, the higher the accuracy of the algorithm. We used

396 the Inception v3 open source deep learning software [41]. This software has been built

397 and tested to categorize images, based on a large training database of images, initially

398 for the annual ImageNet Large Scale Visual Recognition Challenge (ILSVRC;

399 www.image-net.org). The Inception v3 architecture uses a neural network that takes

400 the pixels of images as input and extracts features. Many features are subsequently

401 built on top of features, in different layers of neurons, in higher and higher levels of

402 abstraction, until the neurons reach an output of defined categories [42]. One

403 advantage of the Inception v3 software is that one can reuse the first layers of feature

404 extraction for a different set of images. This is built on the idea that the basic features,

405 *e.g.* lines and simple patterns, can be used in all higher-level features that are used to

406 train new categories with new sets of images. Within eight hours of training time we

407 reached a 93% accuracy, with an execution time in the order of tens of milliseconds.

408

409 After finding the images with a visible first cell, the next step was to determine the

410 injection location. To enable the use of deep learning for this problem, we had to

411 modify the output from categories into an ideal location. When just the pixel (x y)

412 coordinate is used as output, only one pixel of the whole image is correct. With this

413 output the neuronal network cannot easily distinguish between locations closer to the

414 annotated location and further away, and this makes learning impossible. Therefore,

415 we translated the (x y) coordinates to a barycentric coordinate system [43]. The Greek

416 word "*barys*" means heavy and refers to the centre of gravity. In a barycentric

417 coordinate system a grid of triangles is used, with a weight assigned to each vertex.

418 This is used as follows. A chosen grid of triangles is placed on top of each image. The

16

419   annotated injection position will fall within one triangle; then the weights of these

420   triangle vertices are given a value according to the location within that triangle. These

421   weights sum up to one, whereas the other vertices in the grid are all zero. This vertices

422   output vector then represents the ideal outcome. The advantage is now that a small

423   deviation from this ideal output vector can be scored gradually instead of binary. This

424   then allows for efficient training. A more detailed explanation is available in the

425   supporting information (S1). After eight hours of training we created a table of (x y)

426   coordinates using validation images. We calculated the distance between the

427   annotated injection position and the position as predicted by the deep learning network

428   (Fig 3). The average distance is 42 µm, as depicted in Fig 3B, and for 83% of the

429   images this distance is smaller than 60 µm.

430

431   **Fig 3. Distance between annotated and computed injection location**
432   **A.** Bar graph depicting the frequency of the distance between annotated and computed
433   injection position (prediction). **B.** Digital image with a circle around an annotated injection point
434   to illustrate the average distance between annotation and prediction.

435   # Automated injection of *slc45a2* gRNA/Cas9
436
437   Trial and error in many laboratories have led to a best practice of injecting into the first

438   cell for the application of the CRISPR/Cas9 editing technique. In our robotic

439   microinjection system, injecting in the middle of the yolk gives the highest speed.

440   Image recognition used to customise an injection location takes time but can increase

441   the injection efficiency. To balance efficiency and speed, and to be able to monitor

442   improvements of our image recognition model, we started by measuring efficiency of

443   CRISPR/Cas9 injections performed in the yolk. Both manual and automated yolk

444   injections gave a very low efficiency of 12% (Fig 4A). Then, with the "click-to-inject"

445   program, resulting in injections closer to the first cell we could generate albino larvae

17

446   at an almost three times higher efficiency than with the injections in the middle of the

447   yolk (Fig 4A, Semi-auto). Next, using deep learning, we could automate this procedure

448   and with this we reached a slightly lower efficiency when comparing it the "click-to-

449   inject" injections but a higher efficiency than the one obtained with automated and

450   manual injections in the middle of the yolk (Fig 4A, Auto). Still, manual injections into

451   the first cell reached the highest efficiency of 43% (Fig 4A, Manual). Fig 4B shows that

452   both the efficiency and the variation between experiments differs considerably

453   depending on the experimentalist (displayed as P1, P2 and P3). In contrast, here, the

454   automated injections show relatively little variation, also when compare them to the

455   "click-to-inject" injections (Fig 4B, Semi-auto). Also, it can be seen that the efficiency

456   is quickly surpassed by humans given enough experience (P1 and P2). This lower

457   efficiency achieved with the robot can be explained by the injection location – close-

458   to-cell instead of into the zygote – and by the fact that not all the eggs are oriented

459   with a cell visible on the side, despite the fact that they are oriented in the agar grid.

460   Hence, the automated injections tend to be a mixture of injections into the middle of

461   the yolk, and close to the first cell, when the first cell was detected. With this we

462   obtained an efficiency of 24% on average (Fig 4B).

463

464   **Figure 4 Automated injections of CRISPR/Cas9**

465   **A.** Survival and average efficiency of *slc45a2* gRNA/Cas9 manual, click-to-inject (semi-auto)

466   and automated (auto) microinjections both in the yolk and in the cell were measured as the

467   number of larvae displaying an albino phenotype at 3 days post-fertilization (dpf). Uninjected

468   larvae were processed as controls and the resulting pigmented (wild-type) larvae were also

469   counted at 3 dpf. "n=" indicates the number of eggs that were used to obtain the cumulative

470   results. **B.** Comparison of the average efficiency and standard deviation between the

471   automated (auto), click-to-inject (semi-auto) and manual injections performed by three

472   independent experimentalists (P1, P2, P3). "n=" indicates the number of experiments that

473   were used to calculate the average and standard deviation. * P<0.05.

474

Automated injection of DNA

The injections with DNA were performed with an identical robotic microinjection system, but at a different laboratory. Here, we used a COPAS (Complex Object Parameter Analyzer and Sorter) system to measure the efficiency of the injections (Fig 5A). For this, we first measured the highest red fluorescence signal of the uninjected control larvae and took the highest signal as a threshold at 5 dpf. Then we measured the DNA-injected larvae and counted the larvae that passed this threshold. The survival was measured at 1 dpf to focus on differences as a result of the injection. Prior to placing the larvae into the COPAS system, larvae with visible developmental defects were removed. Both the manual and automated injected eggs had a similar relative number of malformed embryos (4% on average, results not shown).

Fig 5. Automated injections of DNA
A. Average survival and efficiency of DNA automated (auto) and manual injections as measured by the COPAS system. "n=" indicates the number of eggs that were used to obtain the cumulative results. **B.** Comparison of the average efficiency and standard deviation between the automated and manual cell injections. P4 indicates a different experimentalist and "n=" indicates the number of experiments that were used to calculate the average and standard deviation.

These results show that DNA injections are less demanding in terms of injection location. Injections into the middle of the yolk reached an average efficiency of 32%. This can be improved by injecting close to the first cell, when possible, to reach an efficiency of 39%. Surprisingly, here manual injections close to the first cell (personal preference) had a lower efficiency than could be obtained by automated injections and gave on average the same efficiency as injections into the middle of the yolk.

Microinjection throughput

To calculate the microinjection throughput, we divided the average injection time by the average efficiency. This results in the average time needed for one successfully injected larva. We measured and compared the throughput for the different genetic modifications and experimental setups described in this article, *i.e.* automated and manual injections for gene knockdown by morpholino antisense, gene knockout by CRISPR/Cas9 and transgenesis by *Tol2* (Fig 6).

Fig 6. Average injection time required to obtain one positive genetically modified larva. Abbreviations: MO, *slc45a2* morpholino; gRNA, *slc45a2* gRNA/Cas9; DNA, *Tol2* construct; Auto, automated injections; P1-4, four different experimentalists.

In the case of the manual injections, the throughput differs greatly depending on the experimentalist, as experience can lead to a higher throughput by increasing both the efficiency and speed of the injection process. It can also be seen that the robot is on par with fast human performance in case of the morpholino injections, but 1.5 times as fast as average human performance.

With deep learning, a robot can outperform humans on the more complex cell injections. With CRISPR/Cas9 the robot needs 6 seconds of injection time to obtain a positive larva, and humans need 8 up to 43 seconds. On average, the robot is more than three times (3.6x) faster. Manual injections of DNA constructs close to the cell are faster to perform than injections into the cell (2.5 seconds *vs* 6.8 seconds). However, this also reduces the manual efficiency, resulting in a 1.5 times higher throughput of the robot. A movie showing the robotic injection process in real-time is available in the supporting information (S2). The movie shows that the time between capturing the image and placing the cross (demonstrating the calculated injection

20

529   location) is only about 100 milliseconds.

Efficiency dependence on the injection location

532   Contrary to what might be expected, the efficiency of injections into the middle of the

533   yolk to alter the genome were not negligible as the efficiency was 12% for

534   CRISPR/Cas9, 32% for DNA injections and 80% for morpholino injections. Using the

535   measured efficiencies and statistics of image classification we can calculate the

536   efficiencies of injections close to the first cell. During the injections of CRISPR/Cas9,

537   on average 65% of the eggs were oriented with a first cell visible, and 35% were

538   injected into the middle of the yolk. The increase in efficiency, 24%, was caused by

539   65% of the eggs being injected with efficiency much higher than 12%. Using the

540   efficiency of the yolk injections we can predict the efficiency of injections close to the

541   first cell. Solving the equation $0.65*X+0.35*0.12 = 0.24$ for X results in an efficiency of

542   around 30% for injections close to the first cell. For DNA injections we have chosen to

543   not orient the eggs after placing them in a grid, and therefore less eggs, 46%, were

544   injected close to the first cell. Solving the equation $0.46*X + 0.54*0.32 = 0.39$ for X

545   results in a predicted efficiency of 47% for injections close to the first cell. Surprisingly,

546   this is much higher than what was obtained by manual injections close to the first cell.

547   These measured and calculated efficiencies can also be used to make a prediction of

548   positive embryos, directly after the injection.

Conclusion and perspectives

551   In this study we have demonstrated how we improved an automated injection robot to

552   inject close to the first cell using image recognition in order to enable efficient genome

553   editing in zebrafish embryos. This was accomplished using a modified open-source

21

554 deep-learning software and annotation of thousands of images. A step-by-step

555 approach of first testing an annotation strategy and efficiency helped to predict the

556 increase in efficiency that can be obtained. Initially we tested the efficiency with a semi-

557 automated click-to-inject program. This click-to-inject approach is also suitable as a

558 first step for other microinjection applications, such as injections into older zebrafish

559 larvae or different organisms.

560 Because of its transparency, rapid development and easy genetic manipulation,

561 zebrafish have become a key vertebrate model organism for the elucidation of

562 developmental processes. With the advent of CRISPR/Cas9 technology, zebrafish are

563 becoming an even more powerful tool for the study of diverse human disorders. The

564 CRISPR/Cas9 system achieves mutagenesis rate of around 80% for generation of

565 knockout lines [25], and has proven to have fewer side effects than other genome

566 editing technologies. However, generation of specific hereditable mutations or epitope

567 tagging of chromosomal genes in zebrafish is still challenging. Unfortunately, genome

568 editing in zebrafish is unpredictable and efficiency sometimes drops to 3.5% [44].

569 Therefore, higher number of eggs should be injected for the generation of the expected

570 mutation. Creation of zebrafish mutant lines using CRISPR/Cas9 requires precise

571 injections into or close to the zygote. These types of injections take time to master and

572 are tedious if many batches of hundreds of eggs have to be injected, particularly for

573 the generation of knock-in lines. Our results have showed that efficiency and

574 reproducibility of manual cell injections highly depend on the training stage of the

575 person performing the experiment, making it more difficult to have this technique as a

576 routine procedure in the laboratory. Here, we show the establishment of automated

577 injections as a reliable tool for the generation of CRISPR/Cas9 mutants. This method

578 could also be used for high-throughput gene overexpression studies by microinjection

22

579   of mRNA. Automated microinjections are simple to learn and allow the cell injection of

580   100 embryos in 2.5 minutes with comparable efficiency to manual cell injections.

581

The need for high-throughput genome manipulation

582

583

584   To date there have been almost 9,000 morpholinos used in zebrafish research. In

585   addition, the adaption of CRISPR/Cas9 editing technology is progressing faster than

586   any other gene silencing method, and even faster than the adoption of morpholino

587   knockdown technology (statistics on zfin.org). However, injections of mRNAs or DNA

588   must be more precise and are more time consuming. Therefore injection can be a

589   limiting step for high-throughput genetic studies. For the moment, there are about

590   30,000 known gene loci that could be interesting to manipulate in order to investigate

591   their function in development, disease or expressed phenotype (zfin.org). Multiplied

592   with 300 injections that are typically used to obtain a mutant, and multiple mutants per

593   gene, this brings us to tens of millions of injections. Much time and efforts would be

594   saved if this tedious but needed task can be performed mostly by robotic systems.

595

Acknowledgements

596

597

References

1. Kimmel CB, Warga RM, Schilling TF. Origin and organization of the zebrafish fate map. Development. 1990; 108(4): 581-594.

2. Takaki K, Davis JM, Winglee K, Ramakrishnan L. Evaluation of the pathogenesis and treatment of Mycobacterium marinum infection in zebrafish. Nature protocols. 2013; 8(6): 1114-1124.

3. Lin S, Long W, Chen J, Hopkins N. Production of germ-line chimeras in zebrafish by cell transplants from genetically pigmented to albino embryos. Proceedings of the National Academy of Sciences. 1992; 89(10): 4519-4523.

4. Jung DW, Oh ES, Park SH, Chang YT, Kim CH, Choi SY, et al. A novel zebrafish human tumor xenograft model validated for anti-cancer drug screening. Molecular BioSystems. 2012; 8(7): 1930-1939.

5. Nicoli S, Presta M. The zebrafish/tumor xenograft angiogenesis assay. Nature protocols. 2007; 2(11): 2918-2923.

6. Nasevicius A, Ekker SC. Effective targeted gene 'knockdown' in zebrafish. Nature genetics. 2000; 26(2): 216- 20.

7. Toyama R, O'Connell ML, Wright CV, Kuehn MR, Dawid IB. Nodal induces ectopic goosecoid and lim1 expression and axis duplication in zebrafish. Development. 1995; 121(2): 383-391.

8. Suster ML, Kikuta H, Urasaki A, Asakawa K, Kawakami K. Transgenesis in zebrafish with the tol2 transposon system. Transgenesis Techniques: Principles and Protocols. 2009: 41-63.

9. Kaethner RJ, Stürmer C. Dynamics of process formation during differentiation of tectal neurons in embryonic zebrafish. Journal of Neurobiology. 1997; 32(6): 627-639.

10. Van Asselt E, De Graaf F, Smit-Onel MJ, Van Raamsdonk W. Spinal neurons in the zebrafish labeled with fluoro-gold and wheat-germ agglutinin. Neuroscience. 1991; 43(2): 611-622.

11. Stainier DY, Lee RK, Fishman MC. Cardiovascular development in the zebrafish. I. Myocardial fate map and heart tube formation. Development. 1993; 119(1): 31-40.

12. Kim CH, Bae YK, Yamanaka Y, Yamashita S, Shimizu T, Fujii R, et al. Overexpression of neurogenin induces ectopic expression of HuC in zebrafish. Neuroscience letters. 1997; 239(2): 113-116.

13. Gore AV, Maegawa S, Cheong A, Gilligan PC, Weinberg ES, Sampath K. The zebrafish dorsal axis is apparent at the four-cell stage. Nature. 2005; 438(7070): 1030-1035.

14. Rosen JN, Sweeney MF, Mably JD. Microinjection of zebrafish embryos to analyze gene function. Journal of visualized experiments: JoVE. 2009 (25).

15. Eisen JS, Smith JC. Controlling morpholino experiments: don't stop making antisense. Development. 2008; 135(10): 1735-1743.

24

16. Kok FO, Shin M, Ni CW, Gupta A, Grosse AS, van Impel A, et al. Reverse genetic screening reveals poor correlation between morpholino-induced and mutant phenotypes in zebrafish. Developmental cell. 2015; 32(1): 97-108.

17. Rossi A, Kontarakis Z, Gerri C, Nolte H, Hölper S, Krüger M, et al. Genetic compensation induced by deleterious mutations but not gene knockdowns. Nature. 2015; 524(7564): 230-233.

18. Stainier DY, Kontarakis Z, Rossi A. Making sense of anti-sense data. Developmental cell. 2015; 32(1): 7-8.

19. Stainier DY, Raz E, Lawson ND, Ekker SC, Burdine RD, Eisen JS, et al. Guidelines for morpholino use in zebrafish. PLOS Genetics. 2017; 13(10): e1007000. doi: 10.1371/journal.pgen.1007000

20. Urnov FD, Rebar EJ, Holmes MC, Zhang HS, Gregory PD. Genome editing with engineered zinc finger nucleases. Nature Reviews Genetics. 2010; 11(9): 636-646.

21. Hwang WY, Peterson RT, Yeh JR. Methods for targeted mutagenesis in zebrafish using TALENs. Methods. 2014; 69(1): 76-84.

22. Jinek M, Chylinski K, Fonfara I, Hauer M, Doudna JA, Charpentier E. A programmable dual-RNA–guided DNA endonuclease in adaptive bacterial immunity. Science. 2012; 337(6096): 816-821.

23. Wiedenheft B, Sternberg SH, Doudna JA. RNA-guided genetic silencing systems in bacteria and archaea. Nature. 2012; 482(7385): 331-338.

24. Mali P, Yang L, Esvelt KM, Aach J, Guell M, DiCarlo JE, et al. RNA-guided human genome engineering via Cas9. Science. 2013; 339(6121): 823-826.

25. Hwang WY, Fu Y, Reyon D, Maeder ML, Tsai SQ, Sander JD, et al. Efficient genome editing in zebrafish using a CRISPR-Cas system. Nature biotechnology. 2013; 31(3): 227-229.

26. Sander JD, Joung JK. CRISPR-Cas systems for editing, regulating and targeting genomes. Nature biotechnology. 2014; 32(4): 347-355.

27. Auer TO, Del Bene F. CRISPR/Cas9 and TALEN-mediated knock-in approaches in zebrafish. Methods. 2014; 69(2): 142-150.

28. Clark KJ, Voytas DF, Ekker SC. A TALE of two nucleases: gene targeting for the masses? Zebrafish. 2011; 8(3): 147-149.

29. Foley JE, Yeh JR, Maeder ML, Reyon D, Sander JD, Peterson RT, et al. Rapid mutation of endogenous zebrafish genes using zinc finger nucleases made by Oligomerized Pool ENgineering (OPEN). PloS one. 2009; 4(2):e4348. doi: 10.1371/journal.pone.0004348

30. Gagnon JA, Valen E, Thyme SB, Huang P, Ahkmetova L, Pauli A, et al. Efficient mutagenesis by Cas9 protein-mediated oligonucleotide insertion and large-scale assessment of single-guide RNAs. PloS one. 2014; 9(5):e98186. doi: 10.1371/journal.pone.0098186

31. Gupta A, Hall VL, Kok FO, Shin M, McNulty JC, Lawson ND, et al. Targeted chromosomal deletions and inversions in zebrafish. Genome research. 2013; 23(6): 1008-1017.

32. Carvalho R, de Sonneville J, Stockhammer OW, Savage ND, Veneman WJ, Ottenhoff TH, et al. A high-throughput screen for tuberculosis progression. PLoS One. 2011; 6(2):e16779. doi: 10.1371/journal.pone.0016779.

33. Spaink HP, Cui C, Wiweger MI, Jansen HJ, Veneman WJ, Marín-Juez R, et al. Robotic injection of zebrafish embryos for high-throughput screening in disease models. Methods. 2013; 62(3): 246-254.

34. Westerfield, M. The zebrafish book. A guide for the laboratory use of zebrafish (Danio rerio). 4th ed., Univ. of Oregon Press, Eugene. 2000

35. Dove A. Screening for content—the evolution of high throughput. Nature biotechnology. 2003; 21(8): 859-864.

36. Dooley CM, Schwarz H, Mueller KP, Mongera A, Konantz M, Neuhauss S, et al. Slc45a2 and C-ATPase are regulators of melanosomal pH homeostasis in zebrafish, providing a mechanism for human pigment evolution and disease. Pigment Cell Melanoma Res. 2012; 26: 205-217.

37. Irion U, Krauss J, Nüsslein-Volhard C. Precise and efficient genome editing in zebrafish using the CRISPR/Cas9 system. Development. 2014; 141, 4827-4830.

38. Lieschke G. Zebrafish: Methods and Protocols. Springer protocols, Heidelberg. 2008.

39. Kingma D, Ba J. Adam: A method for stochastic optimization. arXiv:1412.6980. 2014.

40. Wang W, Liu X, Gelinas D, Ciruna B, Sun Y. A fully automated robotic system for microinjection of zebrafish embryos. PloS one. 2007; 2(9):e862.

41. Szegedy C, Vanhoucke V, Ioffe S, Shlens J, Wojna Z. Rethinking the inception architecture for computer vision. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition 2016; 2818-2826.

42. LeCun Y, Bengio Y, Hinton G. Deep learning. Nature. 2015; 521(7553): 436-444.

43. Ungar AA. Barycentric calculus in Euclidean and hyperbolic geometry: A comparative introduction. 2010.

44. Hruscha A, Krawitz P, Rechenberg A, Heinrich V, Hecht J, Haass C, et al. Efficient CRISPR/Cas9 genome editing with low off-target effects in zebrafish. Development. 2013; 140(24): 4982-4987.

Supporting information

S1: Deep learning supplement

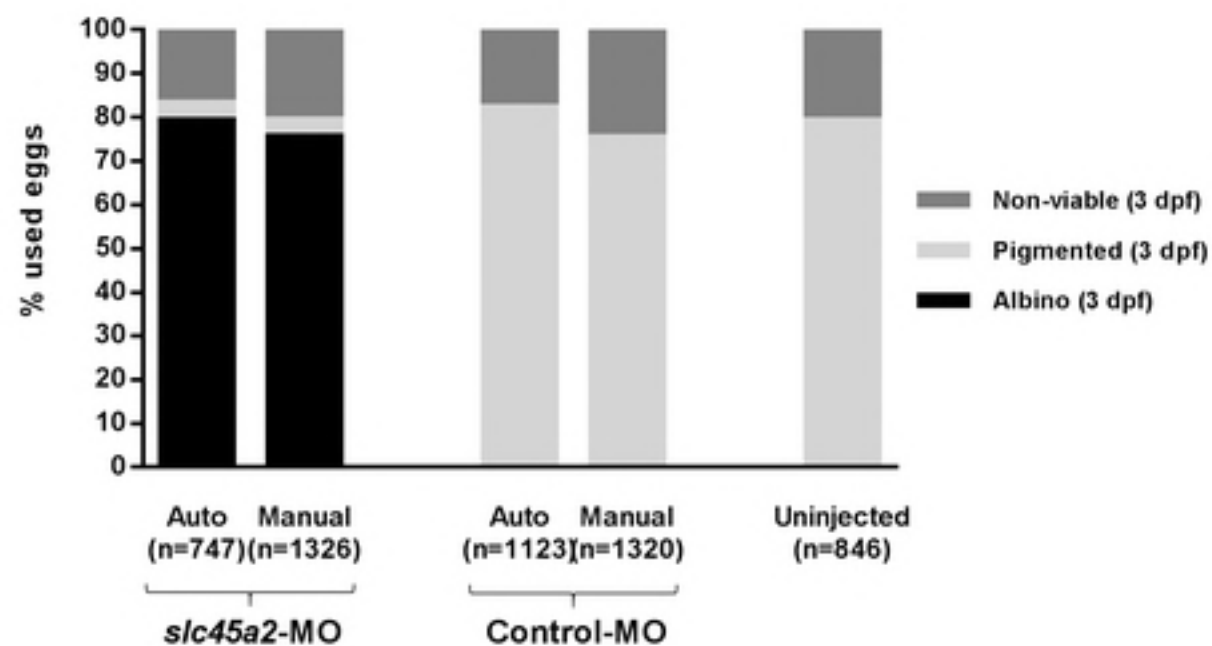S2: Movie demonstrating robotic injections with deep learning

S3: Representative images of *slc45a2*-MO and *slc45a2* gRNA/Cas9

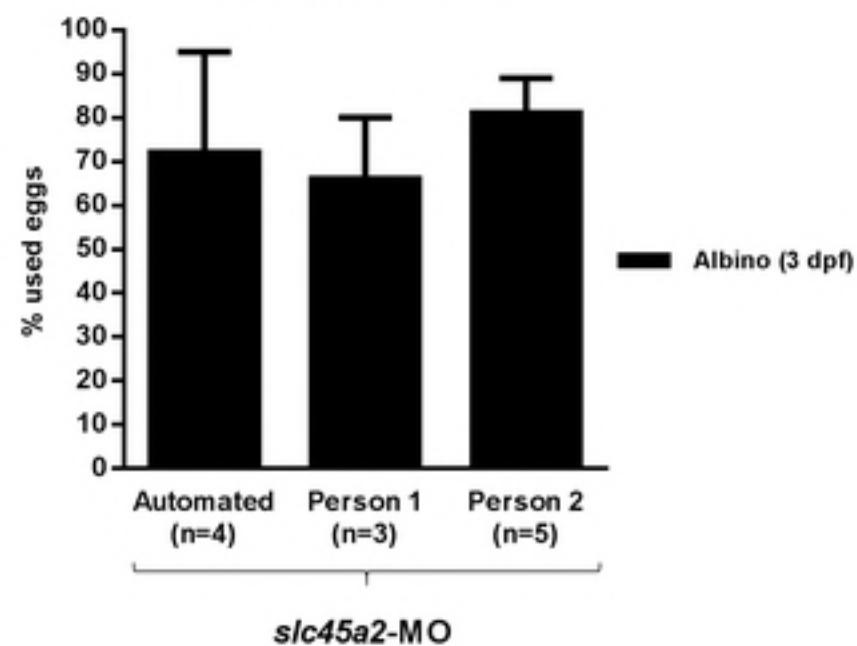S4: Representative images of DNA injected larvae

Fig 1

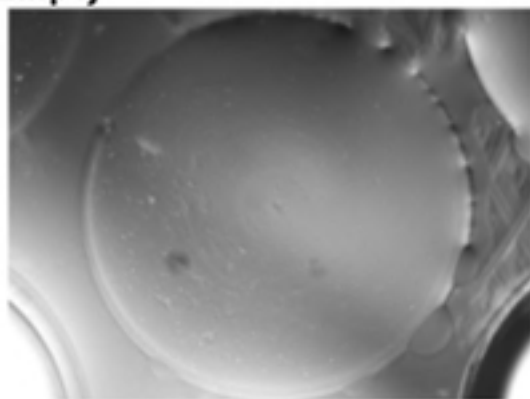**A** Manual *vs* automated yolk injections

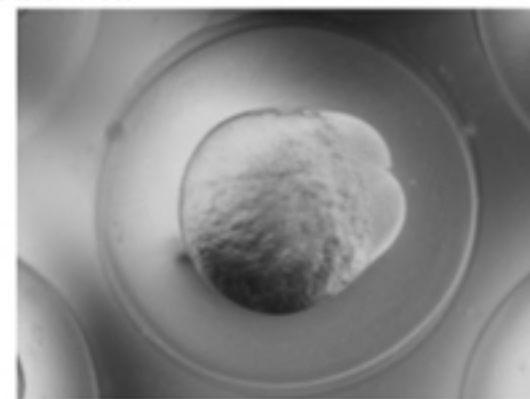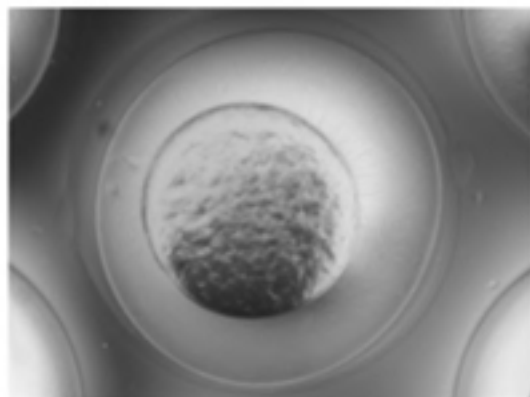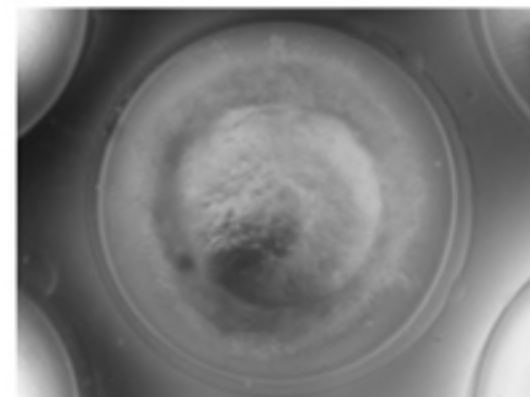**B** Efficiency between automated and manual yolk injections

Fig 2

Empty

Inject

(322, 205)

Two-cell

No-cell

Sick

Fig 3

A

Frequency of the distance between
annotated and computed injection position

B

84 μm

Distance between annotation and prediction (μm)

Fig 4

A Survival and efficiency of CRISPR/Cas9 injections
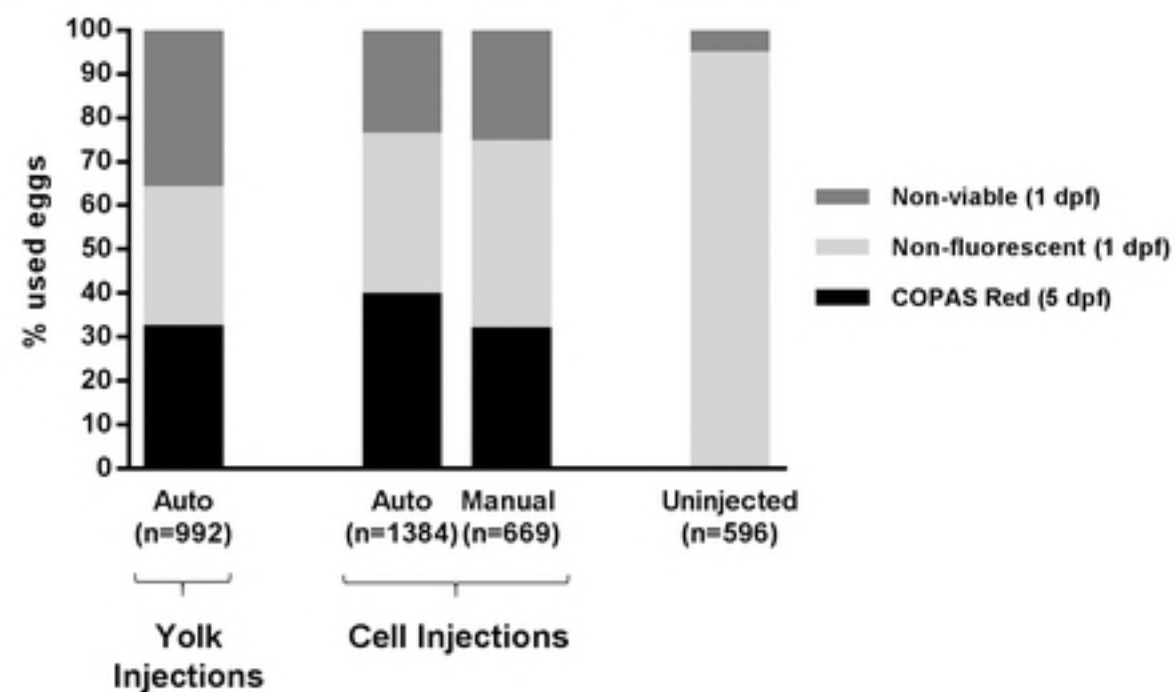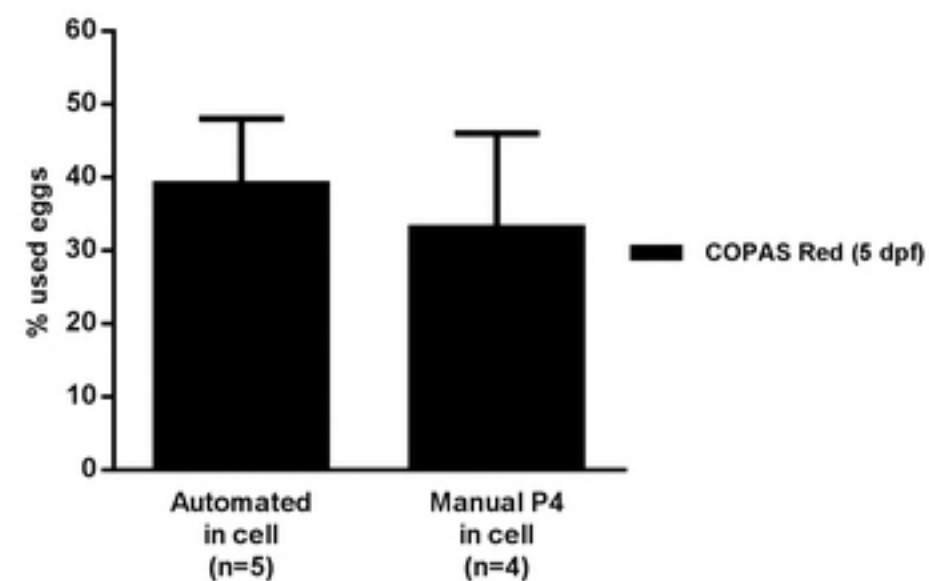
B Efficiency between automated and manual cell injections
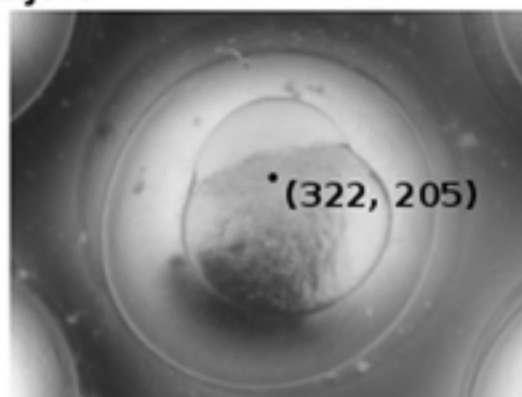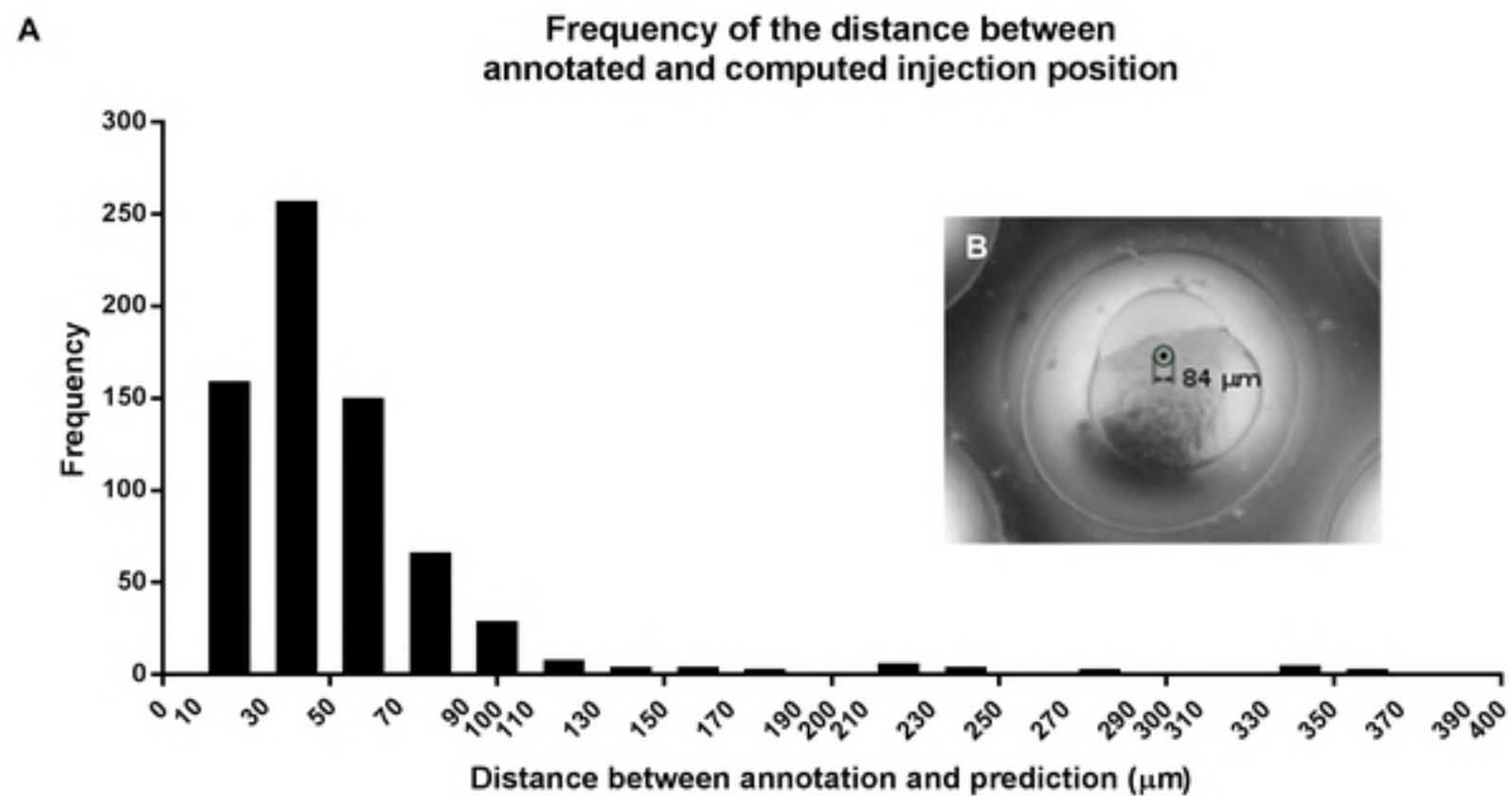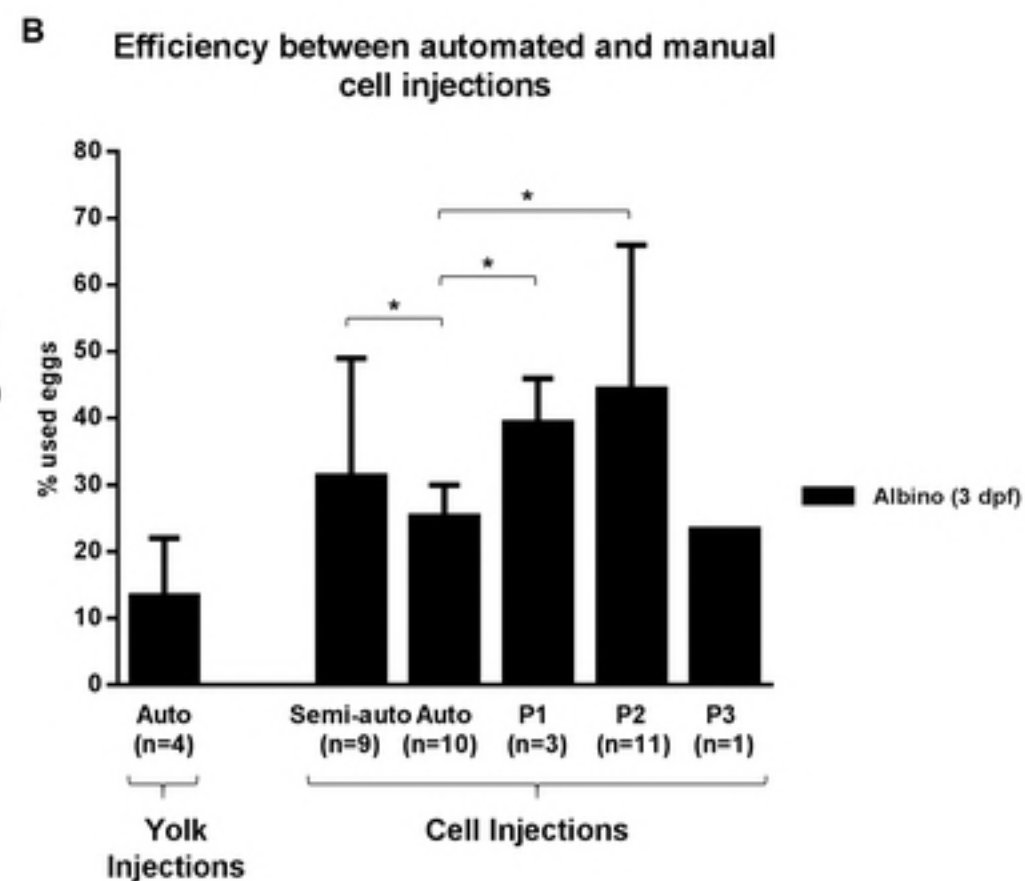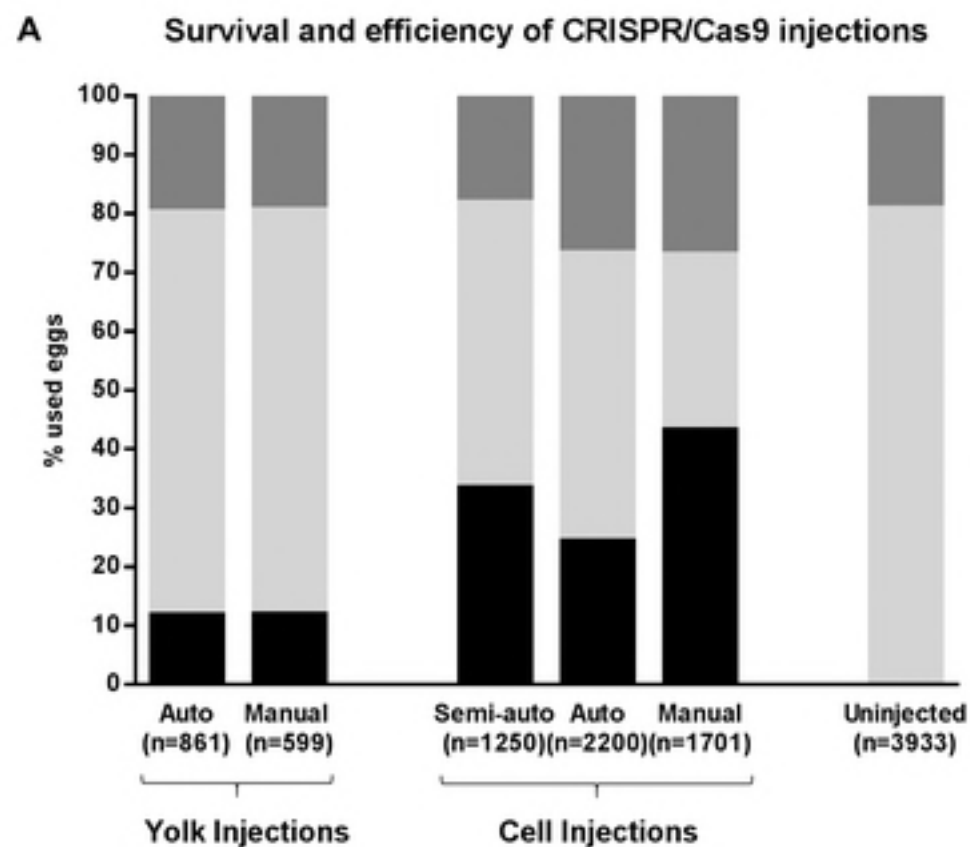
Fig 5

A **Survival and efficiency of DNA injections**

Legend:
- Non-viable (1 dpf)
- Non-fluorescent (1 dpf)
- COPAS Red (5 dpf)

X-axis: Auto (n=992) | Auto (n=1384) | Manual (n=669) | Uninjected (n=596)

Yolk Injections | Cell Injections

B **Efficiency between automated and manual cell injections**

Legend:
- COPAS Red (5 dpf)

X-axis: Automated in cell (n=5) | Manual P4 in cell (n=4)

Fig 6

Microinjection throughput