Maximizing mutagenesis with solubilized CRISPR-Cas9 ribonucleoprotein complexes

Alexa Burger[1,*], Helen Lindsay[1,2,*], Anastasia Felker[1], Christopher Hess[1], Carolin Anders[3], Elena Chiavacci[1], Jonas Zaugg[1], Lukas M. Weber[1,2], Raul Catena[1], Martin Jinek[3], Mark D. Robinson[1,2] and Christian Mosimann[1,‡]

ABSTRACT

CRISPR-Cas9 enables efficient sequence-specific mutagenesis for creating somatic or germline mutants of model organisms. Key constraints *in vivo* remain the expression and delivery of active Cas9-sgRNA ribonucleoprotein complexes (RNPs) with minimal toxicity, variable mutagenesis efficiencies depending on targeting sequence, and high mutation mosaicism. Here, we apply *in vitro* assembled, fluorescent Cas9-sgRNA RNPs in solubilizing salt solution to achieve maximal mutagenesis efficiency in zebrafish embryos. MiSeq-based sequence analysis of targeted loci in individual embryos using CrispRVariants, a customized software tool for mutagenesis quantification and visualization, reveals efficient bi-allelic mutagenesis that reaches saturation at several tested gene loci. Such virtually complete mutagenesis exposes loss-of-function phenotypes for candidate genes in somatic mutant embryos for subsequent generation of stable germline mutants. We further show that targeting of non-coding elements in gene regulatory regions using saturating mutagenesis uncovers functional control elements in transgenic reporters and endogenous genes in injected embryos. Our results establish that optimally solubilized, *in vitro* assembled fluorescent Cas9-sgRNA RNPs provide a reproducible reagent for direct and scalable loss-of-function studies and applications beyond zebrafish experiments that require maximal DNA cutting efficiency *in vivo*.

KEY WORDS: CRISPR-Cas9, Zebrafish, Mutagenesis, Genome editing, RNP, CrispantCal, CrispRVariants

INTRODUCTION

Cas9 nuclease-mediated mutagenesis through non-homologous end joining (NHEJ) repair enables rapid, site-directed mutagenesis of candidate genes in zebrafish for somatic as well as stable germline mutant analysis (Chang et al., 2013; Gagnon et al., 2014; Hwang et al., 2013; Jao et al., 2013; Shah et al., 2015; Varshney et al., 2015). Mutagenesis is routinely performed through microinjection of Cas9-encoding mRNA together with a locus-targeting single-molecule guide RNA (sgRNA). Upon Cas9 translation, folding, and formation of stable Cas9-sgRNA complexes *in vivo*, embryo cells that accumulate sufficient levels of Cas9-sgRNA complex become mutated through imperfect NHEJ at the target locus specified by the sgRNA. This results in a complex genetic mosaic of mutant and wild-type alleles (Jao et al., 2013). Such incomplete mutagenesis is desirable for creating germline mutants, as it warrants embryo survival and yields a spectrum of random alleles to screen for in the next generation (Hruscha et al., 2013; Varshney et al., 2015). More recently, in zebrafish and other models several reports showed increased mutagenesis efficiency upon injection of *in vitro* assembled Cas9-sgRNA ribonucleoprotein complexes (RNPs) that are immediately active upon microinjection and are, variably, more effective (Gagnon et al., 2014; Kotani et al., 2015; Sung et al., 2014).

The somatic mutagenesis efficiency of TALENs and Cas9 allows limited assessment of loss-of-function phenotypes already in the injected F0 generation, potentially providing a promising reverse-genetics tool (Bedell et al., 2012; Dahlem et al., 2012; Jao et al., 2013; Schulte-Merker and Stainier, 2014; Shah et al., 2015). Somatic and tissue-specific mutagenesis has recently been reported for *Ciona* embryos (Stolfi et al., 2014), is possible for assessing tumorigenesis in mice (Platt et al., 2014), and is achievable using mosaic transgene injection in zebrafish (Ablain et al., 2015). Reproducible and significant phenotype penetrance and expressivity in a given cell type or on a whole-embryo scale requires a mutagenesis efficiency close to or reaching saturation, ideally by providing a limited number of alleles. Published efforts using Cas9-mediated mutagenesis in zebrafish have reported a wide range of somatic mosaicism upon injection, with variable numbers of alleles and injection-based mortalities of up to 30% (Auer et al., 2014a; Chang et al., 2013; Gagnon et al., 2014; Hwang et al., 2013; Jao et al., 2013; Moreno-Mateos et al., 2015; Shah et al., 2015). The mutagenesis efficiencies for the different Cas9 applications vary widely in reported studies, with several groups experiencing 50% or more of sgRNAs being ineffective for mutagenesis (Moreno-Mateos et al., 2015; Shah et al., 2015; Sung et al., 2014; Varshney et al., 2015).

Although various online tools are available to assist in design and enable limited target efficiency predictions, the variable mutagenesis efficiency leaves room for optimizing Cas9-mediated mutagenesis in zebrafish and other model organisms. We reasoned that highly pure, pre-assembled Cas9-sgRNA RNPs delivered at optimal conditions into the first cell of the zebrafish embryo would, if well tolerated, mediate saturating mutagenesis within the first few cell divisions. This strategy has the distinct advantage that Cas9-sgRNA RNP assembly is not limited by the amount and rate of Cas9 translation, and pre-loaded sgRNAs are possibly protected from degradation. Here, we describe the *in vitro* assembly of immediately active Cas9-sgRNA RNPs in optimized salt solvent to ensure stability over the course of microinjection. We demonstrate that fluorescently tagged Cas9 protein to monitor RNP delivery is well tolerated by zebrafish embryos with minimal to no injection

[1]Institute of Molecular Life Sciences, University of Zürich, Zürich 8057, Switzerland. [2]SIB Swiss Institute of Bioinformatics, University of Zürich, Zürich 8057, Switzerland. [3]Institute of Biochemistry, University of Zürich, Zürich 8057, Switzerland.
*These authors contributed equally to this work

‡Author for correspondence (christian.mosimann@imls.uzh.ch)

L.M.W., 0000-0002-3282-1730; M.J., 0000-0002-7601-210X; M.D.R., 0000-0002-3048-5518; C.M., 0000-0002-0749-2576

toxicity, while fluorescence provides an instant readout for efficient injections and indicates rapid decay of injected RNPs.

We tested mutagenesis efficiency in individual embryos and find exceedingly high mutagenesis rates that reach 100% for individual target loci, *de facto* generating complete somatic mutants by injection. Despite this exceedingly high efficiency, selective MiSeq analysis of a range of predicted off-targets for highly effective sgRNAs reveal no significant off-target mutagenesis consistent with CRISPR-Cas9 mechanisms, and we observe even a block of mutagenesis when individual polymorphisms are present at a target locus. Although we observe that complete somatic mutagenesis enables loss-of-function readouts in injected embryos, our deep-sequencing results further call for caution in interpreting somatic mutagenesis results due to the wide spectrum of alleles generated with random mutagenesis. Altogether, our approach provides an optimized mutagenesis tool for *in vivo* applications that require maximal mutagenesis efficiency beyond mutagenesis in zebrafish.

RESULTS
Assembly and injection of recombinant, fluorescent Cas9-sgRNA RNPs

We expressed and purified *Streptococcus pyogenes* Cas9 protein fused in-frame with a C-terminal HA epitope tag, nuclear localization signal sequences, and GFP or mCherry for fluorescence detection (Cas9-NLS-GFP and Cas9-NLS-mCherry, respectively) (Fig. 1A, Fig. S1). We further tweaked existing protocols for T7 or SP6 polymerase-driven *in vitro* transcription and purification of sgRNAs (Bassett et al., 2013; Gagnon et al., 2014) to achieve high purity and concentration using standard laboratory protocols for simple and scalable, cloning-free sgRNA production (Fig. S2A). We then combined pure Cas9 protein and sgRNAs and incubated the mix for 5 min at 37°C to reconstitute active RNPs at an injection concentration of at least 800 ng/µl Cas9 (831 ng/µl routinely used in this study). Upon RNP microinjection of a standard volume of 1 pl into zebrafish embryos at the one-cell stage, the fluorescence signal from fluorophore-tagged Cas9 is readily detectable as a homogenous EGFP signal in successfully injected embryos (Fig. 1B,C) and concentrates to nuclei during subsequent cell divisions (Fig. 1D,E, Movie 1). RNP fluorescence in the embryo fades by dilution and possible degradation throughout development and becomes undetectable above background fluorescence before 18 h post-fertilization (hpf).

Consistent with reported salt concentrations for effective Cas9 solubility (Anders et al., 2014), we found that increasing the ionic strength in the reconstitution reaction to at least 300 mM KCl dramatically improved the solubility of the Cas9-sgRNA RNPs and limited aggregation in the injection mix and within the embryo cell (Fig. 1F-I). We found that injections with assembled RNPs at 300 mM KCl were well tolerated by zebrafish embryos: injections by different experimenters of reconstituted Cas9 RNPs into the cell proper of single-cell zebrafish embryos led to no significant lethality or toxicity (*n*=8; Fig. S2B). We optimized the composition of the injection mixture such that only three components and water are required for its preparation: recombinant Cas9 (stored in a protein purification buffer), *in vitro* transcribed sgRNA (dissolved in water), and KCl solution to correct the ionic strength of the mixture to 300 mM KCl to ensure solubility of the Cas9-sgRNA RNPs. To aid in the calculation of the correct amounts needed for complex assembly, we designed CrispantCal, an online tool and complementary smart-phone app to calculate the correct volumes needed for optimal reconstitution (Fig. S3 and Materials and Methods; available at http://lmweber.github.io/CrispantCal/ for desktop and mobile browsers or installable within R; app available through Google Play and iTunes AppStore).

Altogether, fluorescent tracking of the Cas9-sgRNA RNPs provides a simple method to monitor efficient complex solubility and microinjection-based delivery. The easily detectable GFP or mCherry fluorescence further allows convenient sorting of
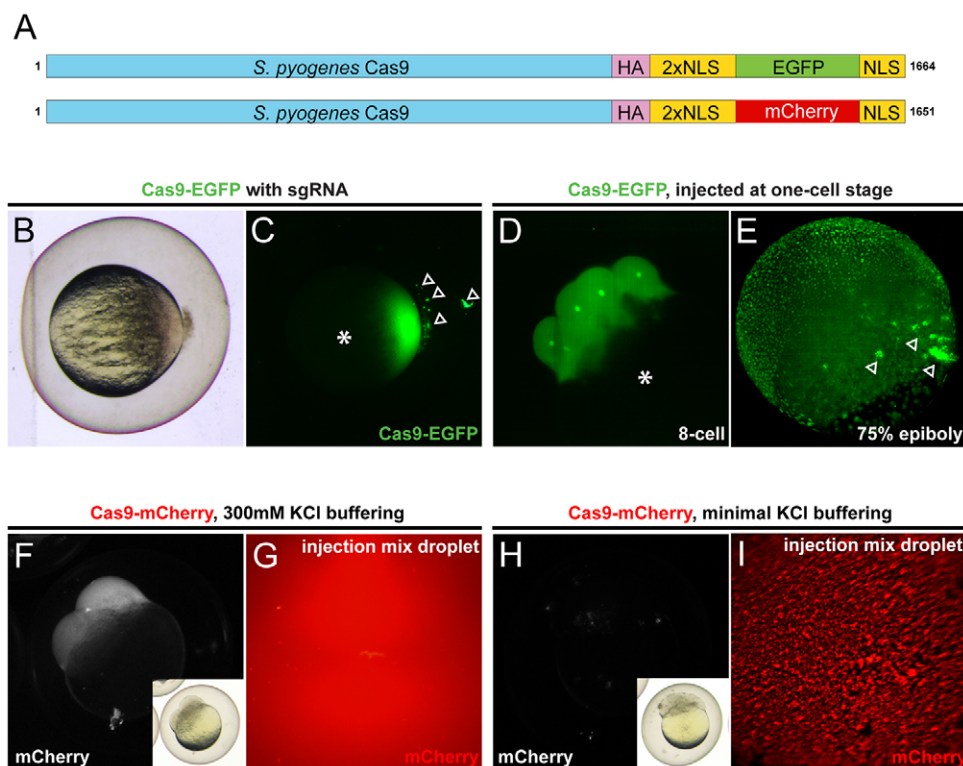

Fig. 1. Microinjection of fluorescent Cas9 fusion proteins into zebrafish embryos. (A) Cas9 fusion proteins used in this study, showing the hemagglutinin tag (HA), two nuclear localization signals (NLS), EGFP or mCherry fluorophores, and a third C-terminal NLS. See Fig. S1 for protein sequences. (B,C) One-cell stage zebrafish embryo following microinjection of reconstituted Cas9-GFP-sgRNA complex. Brightfield (B) and EGFP fluorescence (C) reveal precipitated Cas9 complex outside the embryo (arrowheads, C), and strong EGFP fluorescence in the embryo cell without contribution to the yolk cell (asterisk, C). (D,E) Time-course of injected Cas9-EGFP and nuclear localization. Note the even distribution in the developing embryo without yolk fluorescence (asterisk, D) and minimal precipitate inclusions (arrowheads, E). (F-I) Buffering of the Cas9 protein-sgRNA mix with 300 mM KCl maintains solubility in the injection mix (F,G) and prevents precipitation before injection (H,I). Insets (F,H) show corresponding brightfield images of fluorescence views.

efficiently injected embryos for quality control and experimental reproducibility.

Delivered Cas9-sgRNA complexes efficiently mutate transgene targets

To test the activity of our assembled Cas9-sgRNA complexes, we first targeted a single locus in the genome to assess our RNP efficiency in injected embryos (CRISPR-mediated mutants, or crispants). We targeted a Tol2-based single-copy transgene that drives EGFP expression in all embryo cells under the control of the *ubiquitin* (*ubi*) promoter (*ubi:EGFP*) (Mosimann et al., 2011). We injected RNPs of a previously described, highly efficient sgRNA against *EGFP* (Auer et al., 2014a) into hemizygous *ubi:EGFP* embryos (carrying one transgene copy). In parallel, we co-injected codon-optimized mRNA encoding Cas9-NLS with the same EGFP sgRNA into siblings from the same clutch. In injected *ubi:EGFP* embryos, consistent with previous reports (Auer et al., 2014a), Cas9 mRNA with EGFP sgRNA caused efficient mosaic loss of EGFP signal (*n*=31; Fig. 2A,B,D). By contrast, we consistently failed to detect significant *ubi:EGFP* fluorescence signal in Cas9-sgRNA-injected embryos (*n*=31; Fig. 2A,C,D),

suggesting complete mutagenesis of the single *EGFP* target in these crispants.

This high efficiency is not restricted to *ubi:EGFP*, as we also observe complete absence of EGFP expression in other injected transgenic lines, including *myl7:EGFP* (Huang et al., 2003) (*n*=43; Fig. 2E-J) and *wt1b:EGFP* (Perner et al., 2007) (*n*=5; Fig. 2D,K-M) using our protocol. In addition, we did not detect EGFP protein in RNP-injected embryos by standard western blot analysis (Fig. 2N). Complete mutagenesis depends on direct delivery of Cas9-sgRNA into the embryo cell, as injection into yolk led to EGFP mosaicism akin to Cas9 mRNA injections (*n*=5; Fig. S4). These results reveal that injections of *in vitro* assembled, optimally salt-solubilized Cas9-sgRNA RNPs into the cytoplasm of the initial embryo cell can result in consistent and complete loss of reporter signal from targeted *EGFP* transgenes, suggesting saturating somatic mutagenesis.

Crispants replicate loss-of-function phenotypes

To assess the efficiency of our solubilized RNP delivery at native, bi-allelic genomic targets, we next targeted recessive genes that, upon mutation, cause developmental phenotypes. A gold standard for targeted mutagenesis in zebrafish is *golden* (*gol*; also known as
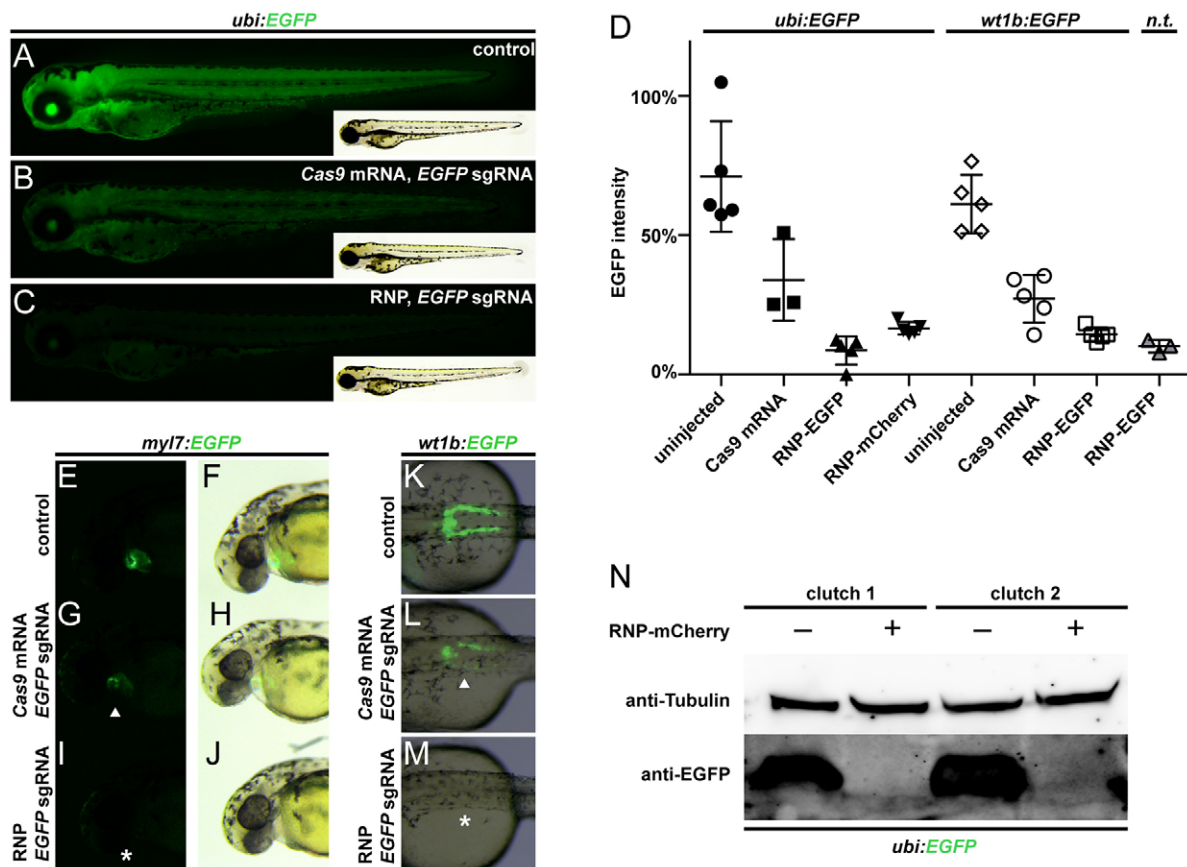

Fig. 2. Efficient mutagenesis of EGFP reporter transgenes. (A-C) Mutagenesis of the *EGFP* open reading frame in transgenic embryos that harbor a single copy of the ubiquitously expressed *ubi:EGFP* transgene using Cas9 and EGFP sgRNA. (A) 5 dpf uninjected heterozygous *ubi:EGFP* control sibling and (B) embryo following injection of Cas9 mRNA with EGFP sgRNA leads to mosaic *EGFP* expression, as compared with (C) complete loss of EGFP signal upon Cas9 protein (Cas9p)-sgRNA RNP injection. Insets are brightfield images of fluorescence views. (D) Relative fluorescence quantification of *ubi:EGFP* and *wt1b:EGFP* mutagenesis versus RNP-injected non-transgenic (*n.t.*) controls. Error bars represent mean with s.e.m. (E-J) *EGFP* mutagenesis in heterozygous *myl7:EGFP* transgenic reporters that express EGFP specifically in heart muscle cells. Left column depicts EGFP fluorescence, right column shows the brightfield view. (E,F) Control sibling, (G,H) Cas9 mRNA and EGFP sgRNA injection causing mosaic EGFP expression (arrowhead, G), and (I,J) reconstituted Cas9p-sgRNA complex injection results in complete loss of EGFP signal (asterisk, I). (K-M) *EGFP* mutagenesis in kidney expressing *wt1b:EGFP* transgenic reporters. Brightfield overlay with EGFP fluorescence. (K) Control sibling, (L) mosaic EGFP expression following Cas9 mRNA and EGFP sgRNA injection (arrowhead), and (M) loss of EGFP following injection of reconstituted Cas9p-sgRNA complex (asterisk). (N) Western blot of EGFP and Tubulin loading control from *ubi:EGFP* transgenic embryos (two examples are shown) injected with EGFP sgRNA-containing Cas9-mCherry RNPs and control embryos.

DEVELOPMENT

slc24a5), which encodes a non-essential ion exchanger involved in skin pigmentation (Lamason et al., 2005). This provides a simple mutagenesis readout, as *gol* mutant zebrafish display markedly lighter melanocyte pigmentation that is readily detectable within 48 hpf (Dahlem et al., 2012; Doyon et al., 2008; Jao et al., 2013). Targeting *gol* with Cas9 mRNA reproduced previously published mosaic phenotypes for TALENs (Dahlem et al., 2012) and Cas9 (Jao et al., 2013) mutagenesis (*n*=57; Fig. 3A,B), albeit with high variability and mortality (34.8% of total clutches), in line with recent reports on delivery of high doses of active Cas9 mRNA and sgRNA (Shah et al., 2015). When we targeted *gol* with reconstituted Cas9-sgRNA complexes, all successfully injected embryos (as judged by Cas9 fluorescence) showed pigment phenotypes without other notable morphological defects (*n*=285; Fig. 3C). As anticipated from random mosaic mutagenesis, these *gol* crispants showed a range of the expected phenotype: we observed complete *gol* phenotype expressivity in nearly 80% of injected embryos, with the remaining 20% showing different degrees of phenotype mosaicism (Fig. 3A-C,P).

Contrary to morpholinos that maintain gene knockdown only for a few days post-injection (Bill et al., 2009), crispants carry mutations in the targeted locus and maintain mutant phenotypes indefinitely. Since *gol* is a non-essential gene, we grew up *gol* crispants of fully penetrant phenotype to adulthood, throughout which the animals preserved the typical pigment phenotype of *gol* mutants (Lamason et al., 2005) (Fig. 3D,E). Upon incrossing of adult *gol* crispants (*n*=6), their F1 offspring showed complete penetrance of the recessive *gol* phenotype in all independently obtained clutches (Fig. 3F,G), revealing that the Cas9-sgRNA-targeted *gol* loci were mutated in the entire germline of all tested crispants.

The seemingly efficient mutagenesis with optimally solubilized RNPs also allows assessment of loss-of-function phenotypes of essential genes. *tbx16* is defective in the *spadetail* (*spt*) mutant, with homozygous *spt* embryos featuring a broadening of the posterior tip of the tail from failed differentiation of multilineage mesoderm progenitor cells (Kimmel et al., 1989). Successfully injected crispants recapitulated the recessive *spt* phenotype, albeit with incomplete penetrance and variable expressivity (Fig. 3H,I,P). We also targeted the lateral plate mesoderm-expressed transcription factor genes *tbx5a* and *hand2*, which when mutated display compound heart and pectoral fin defects (Garrity et al., 2002; Yelon et al., 2000). Analogous to the *tbx5a* mutant *heartstrings* (*hst*) and morpholinos (Ahn et al., 2002; Chiavacci et al., 2012; Garrity et al., 2002), RNP-mediated targeting of *tbx5a* caused a spectrum of phenotype expressivity (*n*=232), including bilateral loss of pectoral fins and elongated heart tubes in efficiently injected embryos (Fig. 3J-L,P). The phenotype penetrance for *hand2* targeting (*n*=308) was less complete but nonetheless replicated the phenotypes of the known *hand2* alleles with variable expressivity (Fig. 3P) (Yelon et al., 2000). We confirmed that targeting of *tbx5a* and *hand2* is responsible for the observed phenotypes in trans-heterozygous mutant F1 embryos derived from F0 incrosses, and homozygous F2 mutants for selected alleles (Fig. 3M-O). These *hand2* and *tbx5a* alleles are, to our knowledge, the first newly derived alleles reported for these key transcription factors.

Altogether, our observations demonstrate that injection of reconstituted Cas9-sgRNA RNPs into one-cell stage zebrafish embryos reproduces loss-of-function phenotypes of targeted genes, albeit with variable phenotype penetrance and expressivity. These results extend previous reports (Gagnon et al., 2014; Jao et al., 2013; Shah et al., 2015) for possible loss-of-function phenotype

assessment of candidate genes using Cas9-sRNA injections and provide a possible framework for phenotype assessment of candidate genes using minimal background toxicity. Using the same reagents, injection of submaximal doses or release of Cas9-sgRNA complexes into the yolk triggers incomplete mutagenesis suitable for germline mutant generation (Fig. 3M-O, Fig. S4).

Cas9 RNPs injected at the one-cell stage can mutate all alleles in zebrafish embryos

Unlike highly efficient morpholinos, phenotype penetrance and expressivity in all injected embryos is variable and incomplete for several essential genes (Fig. 3P). We next sought to quantify and characterize the mutagenesis efficiency in individual crispants to assess whether incomplete phenotype penetrance and expressivity are a function of incomplete mutagenesis or of other factors. Reported deep-sequencing analyses of mutagenesis efficiencies have been performed on pooled embryos with a range of different methods that allow only limited cross-comparison between experiments (Gagnon et al., 2014; Shah et al., 2015). We therefore sought to perform deep-sequencing based on Illumina MiSeq to determine the mutagenesis efficiency in individual embryos.

We devised a scalable analysis pipeline in which we selected individual crispants and PCR-amplified ~350-500 bp genomic regions centered on the individual sgRNA recognition sites to subsequently perform both MiSeq-based deep-sequencing and limited Sanger sequencing of the PCR products. For Sanger sequencing, we developed a column-free workflow to rapidly isolate sequencing-grade DNA from single clones that included subcloning the PCR fragment and performing a colony PCR assay (see Materials and Methods for details and extended protocols).

To establish standardized analysis and interpretation of the mutagenesis spectrum, we devised CrispRVariants, a flexible and scalable R-based software package. CrispRVariants is reproducible, scalable to large data sets, and transparent and flexible about which reads are included in efficiency calculations (Lindsay et al., 2015 preprint). CrispRVariants counts every variant allele and localizes variants with respect to the cut site, enabling simple comparison of the full mutation spectrum between guides and exclusion of pre-existing genomic variants, e.g. in a non-homozygous experimental population. As graphical output, the software provides standardized summaries of individual crispants or any other input sequences for phenotype versus mutagenesis, including quality assessment plus automated illustration of resulting mutant variants and mutagenesis quantification using panel plots to illustrate allele sequences per embryo (Fig. 4A).

Based on our deep-sequencing data for 48 loci (25 on-targets, one control, and a total of 22 predicted off-targets for six on-targets) in up to six individual embryos each (see Materials and Methods for details), our RNP-based mutagenesis approach results in an exceedingly high on-target efficiency upon optimized RNP injection, with selected crispants featuring complete mutagenesis at individual loci (Fig. 4A,B, Table S1, and individual panel plots of MiSeq-analyzed targets in Fig. S5A-O). We found an average mutagenesis rate of 91.26% (median 94.06%, average read count above 32,000 per locus) for the analyzed targets. Consistent with previous reports of Cas9 use in zebrafish (Gagnon et al., 2014; Hruscha et al., 2013; Hwang et al., 2013; Jao et al., 2013), the majority of induced (and PCR-recovered) alleles consist of small deletions and insertions (indels), with fewer larger indels (Fig. 4A, Fig. S5A-O). Several analyzed loci, including *gol*, *atg7*, *camk2g1* and *xirp1*, feature at least one predominant recurring mutation in
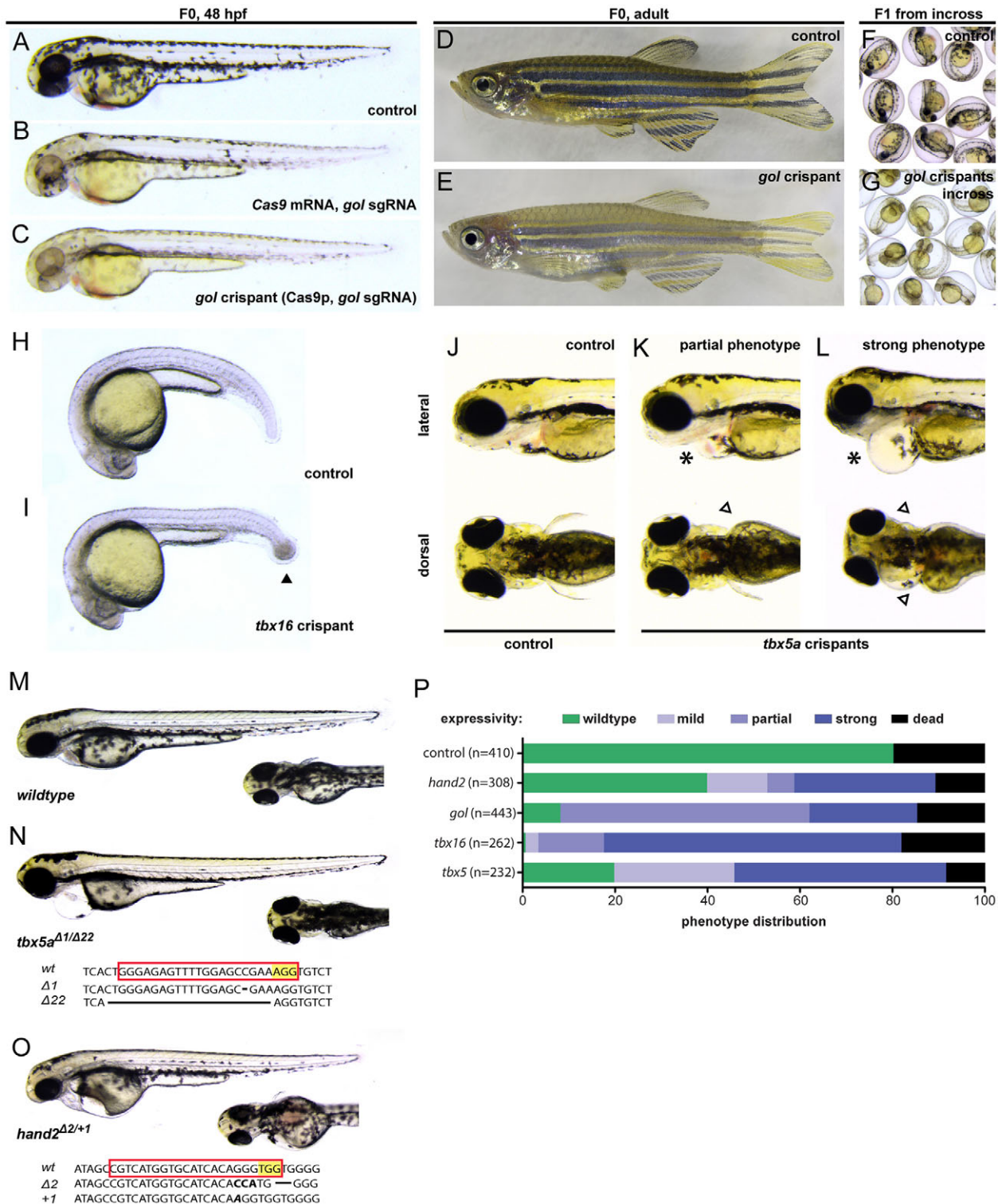
Fig. 3. Recapitulation of mutant and morpholino phenotypes in crispants. (A-G) Mutagenesis of *golden* (*gol*). (A) wild-type control. (B,C) Targeting with Cas9 mRNA causes mosaic pigment cell phenotypes (B), while solubilized RNP injections can result in zebrafish embryos devoid of wild-type melanocytes (C) that completely mimic the *gol* mutant phenotype. (D-G) Whereas control siblings develop characteristic pigment patterns as adults (D), fully penetrant *gol* crispants maintain their mutant phenotype to adulthood (E) and completely transmit mutant *gol* alleles to the F1 generation when incrossed with other *gol* crispants (F,G). Adult zebrafish are shown as composite images. F1 phenotypes were assessed from three crispant incrosses resulting in normally sized (75-100) to big (250-300) clutches. (H,I) Crispants for *tbx16* can form the characteristic *spadetail* (*spt*) phenotype of *tbx16* mutants (arrowhead in I, see also Fig. S5K,L). (J-L) Crispants for *tbx5a* recapitulate the pectoral fin and heart phenotypes of *tbx5a* mutants and morphants, showing a range of phenotype expressivity including cardiac edema (asterisk, K) and unilateral pectoral fin defects (arrowhead, K), or complete heart tube stretching (asterisk, L) and pectoral fin loss (arrowheads, L). (M-O) Phenotype/genotype correlation of *tbx5a* and *hand2* F2 embryos featuring two different alleles. Red box, target sequence; PAM is in yellow. (P) Phenotype penetrance and expressivity for tested crispants. All embryos summarized in the graph were quality controlled for injection by fluorescence microscopy for the tagged RNP directly after injection; non-fluorescent embryos were discarded; dead embryos were counted at 1 dpf and included unfertilized embryos.
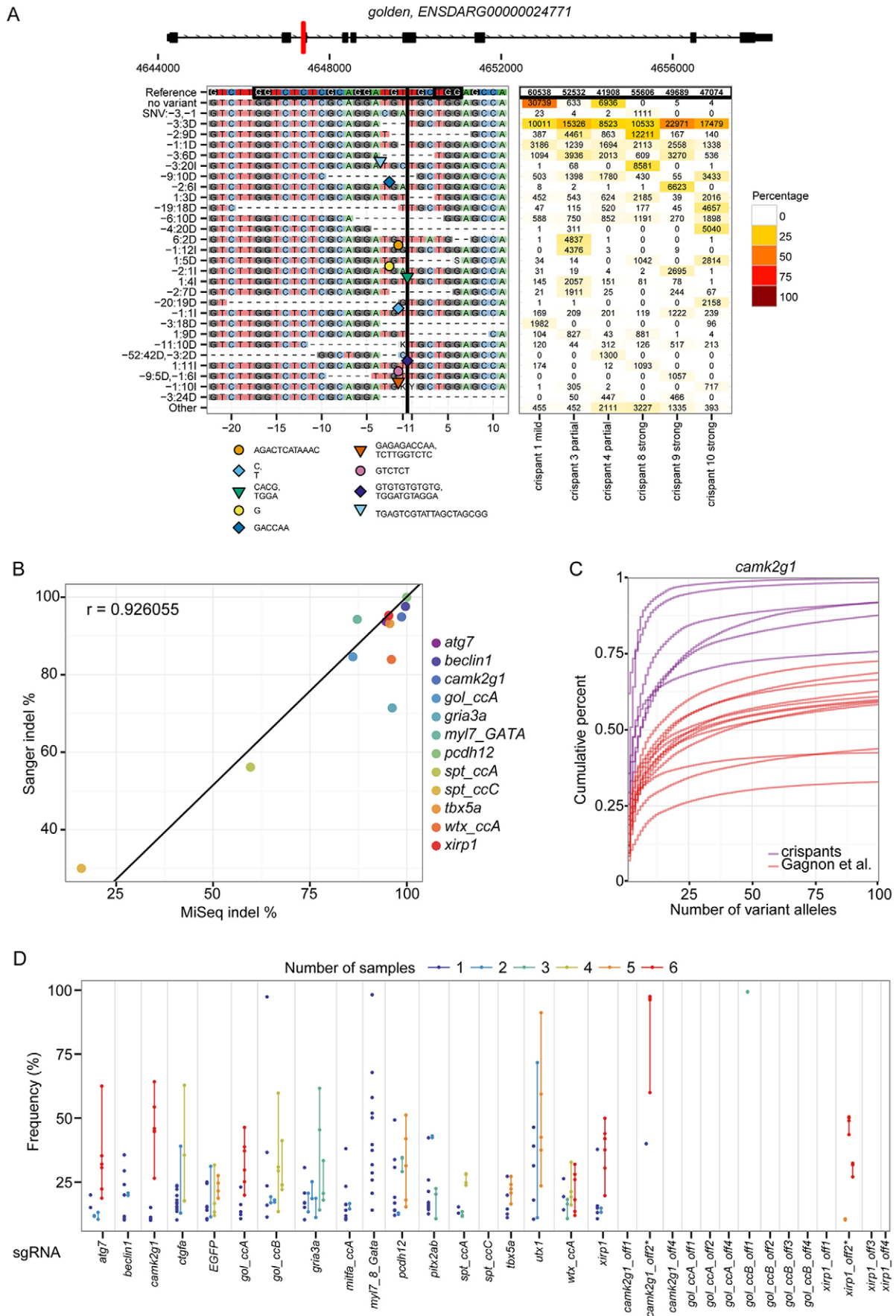
Fig. 4. See next page for legend.

Fig. 4. Mutation spectrum in individual F0 embryos as analyzed by CrispRVariants. (A) Panel plot output of mutagenesis in *gol* crispants as created by CrispRVariants analysis of MiSeq data from individual embryos. The gene schematic at the top shows the location of the sgRNA in red with respect to all overlapping Ensembl transcripts on the reference strand. The left panel shows the pairwise alignment of each variant to the reference genome, oriented so the PAM is 3′ of the alignment. The 20 bp sgRNA and 3 bp PAM sequences are indicated by boxes on the reference sequence, and the cut site is indicated by a vertical line. Deletions are indicated by '-' and insertions by symbols above the corresponding alignments, with the inserted sequence(s) shown underneath the plot. Variants are grouped based on the position of insertions and deletions. The consensus sequence is shown for all sequences sharing a variant, using IUPAC ambiguity codes for sequence differences. The variant locations are numbered by their most upstream base with respect to the cut site, i.e. a 6 bp deletion from 7 to 2 bases upstream of the cut site would be labeled '−7:6D'. The right panel shows the frequency of the variants in each embryo. Colors correspond to variant frequencies. The header shows the total number of MiSeq-derived sequences in each embryo. (B) Mutagenesis reaches saturation (100%) for several targets in individual embryos. Plot depicts the average mutagenesis efficiency (frequency of variants containing indels) in individual embryos for loci analyzed with both Miseq and Sanger sequencing of PCR clones (see Materials and Methods for details). The overall mutagenesis efficiency determined by both methods correlates with a Pearson product-moment correlation coefficient of r=0.926. (C) Optimized RNP mutagenesis occurs early during development. Cumulative proportion of all detected indel variants in single embryos accounted for by the top 100 indel variant alleles for *camk2g1*, ordered from most to least frequent, for crispants generated in this study and RNP-mediated mutagenesis in Gagnon et al. (2014). Each line shows the results for a single embryo (*n*=6 embryos; *n*=12 embryos for Gagnon et al.). Theoretically, the proportion of a variant allele reflects the cell stage at which the variant occurs. A *y*-intercept of 0.5 would represent a variant that occurred at the single-cell stage; in practice, two separate mutation events could generate the same allele or the PCR amplification might not preserve the allele proportions. Higher *y*-intercept indicates earlier mutation occurrence; the maximum *y*-value is the mutation efficiency when chimeric read alignments are excluded. See Fig. S6 for additional examples. (D) The frequency and recurrence of mutant alleles. Each point shows the frequency of a mutant allele among the variant reads for a single sample. Points are colored by the number of samples in which the allele occurred. Repeated observations of the same allele are connected by lines, revealing an abundance of high-frequency reoccurring alleles for most targeted regions. For example, one mutant allele recurred in all six samples for *atg7* and accounted for 20-70% of the total mutant sequences. Six samples were sequenced for each guide except *beclin1*, *spt_ccA* (four embryos) and *spt_ccC* (two embryos). Alleles of at least 1% frequency among the sequenced reads for a guide and of at least 10% frequency among the variant alleles are shown. Off-targets marked by asterisks feature sequencing errors or genetic variation (see text for details). See also panel plots in Fig. S5 for individual sequence details.

independently targeted embryos (Fig. 4A,D, Fig. S5A,C,O), suggesting locus-dependent preferential NHEJ repair following Cas9-mediated double-strand break induction. In contrast to completely mutant crispants, we do recover unmutated alleles in individual crispants that show incomplete phenotype penetrance and variable expressivity for the expected phenotype (Fig. 3P), in particular for *tbx16* (*spt*) (Fig. 4B,D, Table S1, Fig. S5K,L). Of note, CrispRVariants analysis of the *EGFP* open reading frame targeted with RNPs and Cas9 mRNA in *ubi:EGFP* transgenes (Fig. 2) also revealed complete mutagenesis for the sequenced RNP-injected embryos (Fig. S5W).

The average allele variant number per targeted locus ranged from 3.67 to 16.83 different alleles per embryo (Table S1). Assuming that mutagenesis of a given locus continues to generate independent variants as long as the sgRNA template can guide Cas9 to the locus and the locus can be cleaved by Cas9, these results suggest that mutagenesis saturation is reached at different time points across individual loci. Every target we tested for this study harbored an

abundance of alleles, including *gria3a*, which was previously reported as a difficult target for Cas9 mRNA and Cas9 protein mutagenesis (Gagnon et al., 2014) (Fig. 4B-D, Fig. S5F,T, Table S1, Fig. S6A,B). Consistent with immediate activity of optimally reconstituted RNPs, we already detected mutant alleles in individual crispants analyzed at germ ring and 32-cell stage (*gol* and *pcdh12*, respectively), as compared with Cas9 mRNA injections, which showed few if any retrievable mutations in *gol* by the 32-cell stage (Fig. S7).

MiSeq or similar exhaustive deep-sequencing analysis of individual targets is impractical for mutagenesis assessment of single targets during routine experiments. We therefore also performed analysis of the same crispants using Sanger sequencing of a limited number of individual clones from PCR products, as routinely reported to assess mutagenesis efficiency (Auer et al., 2014a; Jao et al., 2013; Stolfi et al., 2014; Varshney et al., 2015). For analyzed targets with good sequence coverage (*n*=12 or more), our analysis revealed that limited Sanger sequencing data strongly correlate with our MiSeq data: for all analyzed loci, Sanger sequencing of subcloned PCR fragments (1) reliably established a mutagenesis efficiency estimate (Fig. 4B) and (2) recovered the most frequent alleles retrieved by deep-sequencing (Fig. S5A). Alleles frequently recovered by MiSeq or Sanger sequencing of F0 embryos targeted for *gol*, *tbx16*, *tbx5*, *hand2* and *xirp1* also transmit through the germline (Fig. 3N,O, Fig. S8), in line with previous reports on a larger cohort of target loci mutated with Cas9 mRNA (Varshney et al., 2015).

The random nature of mutagenesis resulting from efficient NHEJ can result in in-frame lesions (indels with base pair multiples of three) that are predicted to maintain open reading frame integrity. We frequently recovered such in-frame variants in independently targeted genes (Fig. 4A, Fig. S5). This observation reveals that randomly generated in-frame alleles, which potentially maintain open reading frame integrity (yet nonetheless might impact amino acid residues important for protein function), are a major uncontrolled variable in the use of crispants induced by any method to directly assess loss-of-function phenotypes, both on a whole-embryo (Shah et al., 2015; this study) and on a tissue-specific (Ablain et al., 2015; Stolfi et al., 2014) scale.

High sequence fidelity of Cas9-mediated mutagenesis in zebrafish

We also analyzed the possibility of off-target effects. Previous accounts reported remarkably low off-target mutagenesis for analyzed loci (Shah et al., 2015; Varshney et al., 2015). We investigated 22 loci for off-targets [based on CasOT score (Xiao et al., 2014) and proximity to genes], of which we excluded one off-target due to a strain-specific deletion (*camk2g1_off2*) and one due to a possible sequencing error from a homopolymer run (*xirp1_off2*) (Lindsay et al., 2015 preprint). We found no single predicted off-target locus to be mutant above the threshold of the MiSeq error rate (Table S1). Although our analysis is limited to the predicted off-targets that we probed, we also routinely found cases where even single SNPs in target sequences completely abolished the mutagenesis efficiency: our *pitx2ab* target sequence features a SNP adjacent to the Cas9 cut site, resulting in no cutting (Fig. S5J), while the *tbx16* sgRNA *ccC* harbors two SNPs in the WIK strain that completely resisted mutagenesis (Fig. S9). These observations, together with previous reports, differ from findings in other model systems in which higher degrees of sequence divergence are tolerated by Cas9 RNP complexes. We interpret these data as a strict sequence dependence of Cas9-mediated mutagenesis using the

DEVELOPMENT

2031

current methods applied in zebrafish, possibly owing to the limited activity window of injected Cas9 mRNA or RNP complexes during the rapid embryonic development.

Functional assessment of gene regulatory elements

Establishing the developmental contribution of non-coding functional elements in the genome, such as transcription factor binding sites and enhancers, remains experimentally challenging. The flexibility of sgRNA design and efficient mutagenesis using our RNP approach prompted us to assess the feasibility of mutating genomic sequences outside of open reading frames to assess their developmental contribution by direct injection.

The minimal regulatory element of the zebrafish *myl7* gene (formerly *cmlc2*) harbors several predicted transcription factor binding sites that have been implicated in driving *myl7* expression based on transgenic reporter experiments (Huang et al., 2003) (Fig. 5A). We individually targeted the putative GATA factor binding site (−139 to −131 bp upstream from the transcription start) and the MZF motif [−96 to −89 bp (Huang et al., 2003)] of *myl7* with dedicated sgRNAs in embryos carrying a single copy of a transgenic *myl7:EGFP* reporter in addition to the endogenous *myl7* loci (Fig. 5B-D, Fig. S10A,B). At 36 hpf, *myl7:EGFP* reporter fluorescence was severely diminished or abolished in crispants with a targeted GATA site (Fig. 5E,F), while MZF site crispants had strongly decreased reporter activity (Fig. 5G,H). Additionally, embryos with impaired reporter expression developed mild cardiac edema and slowed heart beat (Fig. 5E,G), reminiscent of the reported *myl7* morpholino and mutant phenotype (Huang et al., 2003; Stainier et al., 1996). Targeting the intermittent genomic region between the GATA and the MZF site did not interfere with *myl7:GFP* reporter expression despite efficient mutagenesis (Fig. 5I,J, Fig. S10C), ruling out a non-specific effect of binding site mutagenesis on minimal promoter elements or the *EGFP* transcriptional unit.

Since the targeting sgRNAs recognize both the *myl7:EGFP* reporter insertion and the two native *myl7* loci, we assessed endogenous *myl7* expression by mRNA *in situ* hybridization at 36 hpf (Fig. 5K-N): GATA factor site-targeted crispants invariantly showed mosaic or completely abolished *myl7* expression (77% and 23%, respectively, *n*=98; Fig. 5L,M), while MZF site-targeted animals showed a marked decrease of endogenous *myl7* expression (89%, *n*=19; Fig. 5N). Germline transmission of *myl7:EGFP* reporter transgenes with the mutant GATA binding site confirmed the crispant findings (Fig. 5P-T). These data highlight the potential of crispants to uncover native and reporter-based non-coding regulatory sequences in the zebrafish genome.

DISCUSSION

Genome editing using zinc fingers, TALENs, and now CRISPR-Cas9 has significantly facilitated genome engineering in model organisms. Nonetheless, assessment of loss-of-function phenotypes beyond the contested temporary morpholino-mediated knockdown (Kok et al., 2015) requires the generation of stable mutant strains, which puts a heavy burden on the capacity and costs of animal facilities. Several recent reports suggested the feasibility of direct phenotype readouts from somatic whole-embryo or tissue-specific mutagenesis (Ablain et al., 2015; Bedell et al., 2012; Dahlem et al., 2012; Jao et al., 2013; Platt et al., 2014; Schulte-Merker and Stainier, 2014; Shah et al., 2015; Stolfi et al., 2014). Such F0-based phenotype assessment depends on high mutagenesis efficiency, ideally complete saturation with limited loss-of-function allele mosaicism. Here, we extend and refine previous reports of *in vivo*

mutagenesis and direct phenotype readout upon injection of locus-specific endonucleases in zebrafish by injecting highly pure, pre-assembled, and optimally solubilized Cas9-sgRNA RNPs. Using detailed protocols and dedicated software tools to streamline injection mixes (CrispantCal) and, in particular, data analysis (CrispRVariants) (Lindsay et al., 2015 preprint), our approach provides exceedingly high mutagenesis rates that reach saturation in individual embryos for particular targets. To our knowledge, this is the first report of saturating mutagenesis of individual candidate loci using the CRISPR-Cas9 system in a model organism.

Published reports of CRISPR-Cas9 deployment in zebrafish vary widely in terms of the reported concentrations for sgRNA and Cas9 mRNA or protein, mutagenesis efficiency, and the assessment of resulting mutant alleles (Auer et al., 2014a,b; Chang et al., 2013; Gagnon et al., 2014; Hruscha et al., 2013; Hwang et al., 2013; Irion et al., 2014; Jao et al., 2013; Kimura et al., 2014; Moreno-Mateos et al., 2015; Sung et al., 2014). In our hands, some of this stems from experimental variability in the delicate preparation and handling of *in vitro* transcribed RNA components and in the mechanical process of individual microinjections. Our results highlight that validated working stocks of recombinant Cas9 protein provide a more consistent reagent for CRISPR-Cas9 mutagenesis than long, *in vitro* transcribed and capped Cas9 mRNA. Freshly reconstituted and buffered Cas9-sgRNA RNPs can be appropriately titrated for traditional germline mutagenesis or somatic mutagenesis in crispants. Cas9 fusions with GFP or mCherry facilitate immediate quality control of the injection to minimize experimenter-influenced variability. Of note, appropriate adjustment of ionic strength with 300 mM KCl to solubilize and stabilize the reconstituted injection mixes markedly improves the mutagenesis efficiency compared with previous Cas9 protein-based approaches (Chang et al., 2013; Gagnon et al., 2014) (Fig. 4, Figs S5 and S6). Such optimized efficiency is likely to be of benefit to applications that depend on saturating DNA cutting, such as homologous recombination or insertion of short exogenous DNA sequences. Our solubilized RNPs are also likely to be directly transferable to other model organisms beyond zebrafish that allow injection-based RNP delivery.

Although we cannot exclude the possibility, several observations suggest that off-target mutagenesis of our solubilized Cas9 RNPs in injected zebrafish embryos is minimal: first, SNPs at a sgRNA-targeted locus inhibit Cas9 function (Fig. S9); second, germline-transmitted genomes from highly mutagenized crispants maintain highly specific phenotypes (Fig. 3F,G,M-O, Fig. 5P-S, Fig. S8); and third, although limited to CasOT-predicted key off-targets, we cannot detect any significant off-target mutagenesis, consistent with previous reports and our re-analysis of their data (Gagnon et al., 2014; Lindsay et al., 2015 preprint; Shah et al., 2015). Besides the native sequence fidelity of the CRISPR-Cas9 system, our observations of injected fluorescent Cas9 RNPs suggest a relatively short activity window of only a few hours post-injection (Fig. 4C, Fig. S6B-D). The high numbers of individually injected embryos combined with independent experiments using at least two different sgRNAs per targeted gene would possibly further mitigate off-target effects. Our work also underlines once more that the high fidelity required for complementary sgRNA sequences warrants careful assessment for polymorphisms in the target region; use of sequence-characterized zebrafish strains such as NHGRI-1 (LaFave et al., 2014) will greatly facilitate extended crispant experiments.

Genes or regulatory sequences with promising crispant phenotypes can further be targeted using subsaturating complex concentrations or yolk injections to generate stable mutant alleles
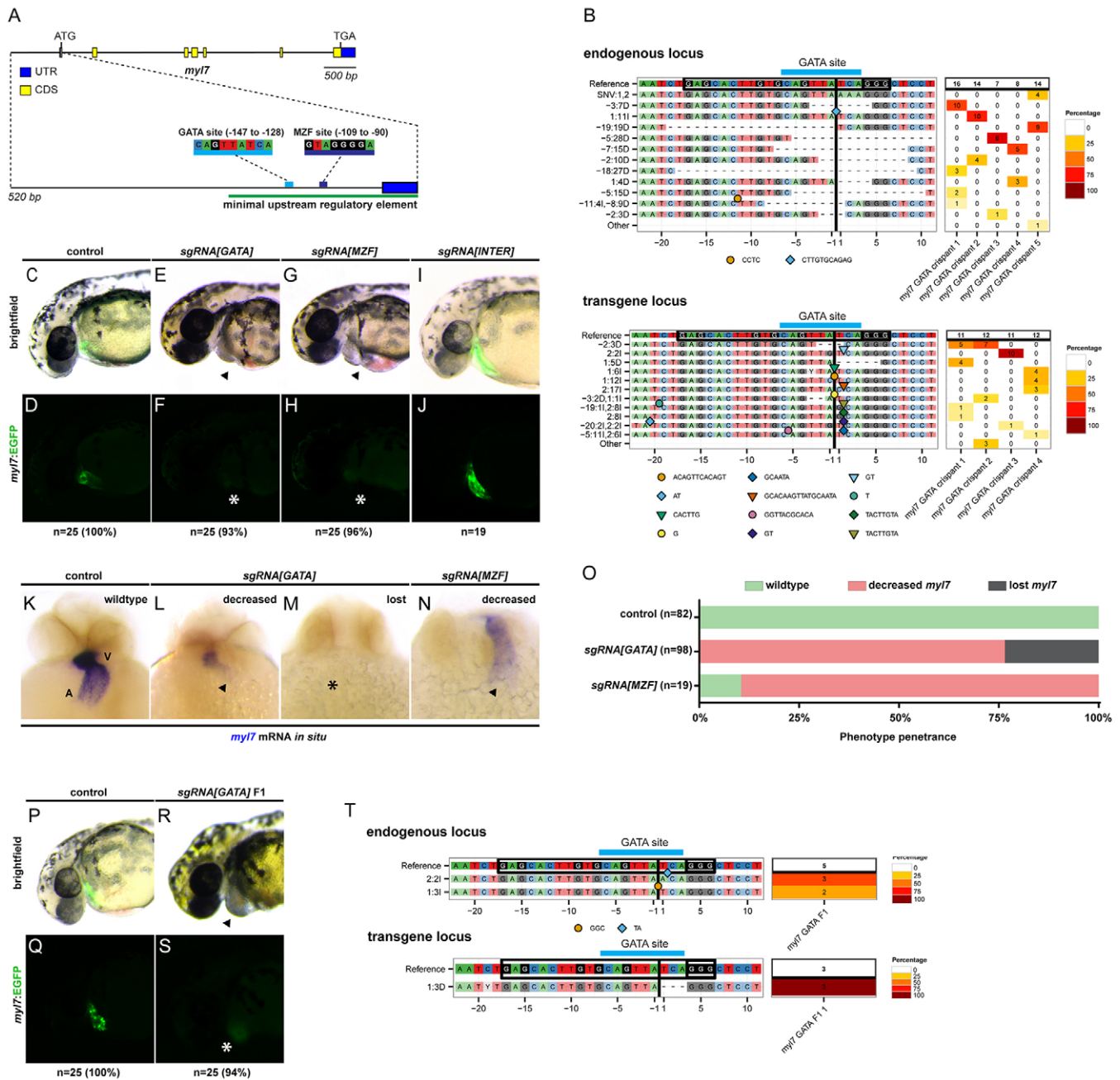
Fig. 5. Crispant analysis of transcription factor binding sites upstream of *myl7*. (A) The zebrafish *myl7* locus encoding cardiac myosin light chain 7, with an enlarged depiction of the upstream regulatory region and putative GATA and MZF transcription factor binding sites. (B) Panel plot depicting Cas9-mediated mutagenesis of the GATA site region in individual representative crispants for the endogenous *myl7* locus and the single-copy *myl7:EGFP* transgene. (C-J) Anterior lateral views at 36 hpf of heterozygous transgenics for *myl7:EGFP* targeted with sgRNAs recognizing the GATA or MZF binding site or the interspersed region between the two sites (INTER) in *myl7* with solubilized RNPs. (C,D) Control sibling with EGFP expression in the heart tube. (E,F) Reduction or loss of EGFP upon targeting of the GATA site with *sgRNA[GATA]*, (G,H) the MZF site using *sgRNA[MZF]* or (I,J) the interspersed sequence with *sgRNA[INTER]*. Arrowheads indicate cardiac edema (E,G); asterisks indicate impaired EGFP signal (F,H). (K-N) mRNA *in situ* hybridization (ISH) for endogenous *myl7* expression at 36 hpf in examples of control siblings (K), GATA site crispants (L,M) and MZF site crispants (N). (K) A, atrium; V, ventricle. Arrowheads indicate decreased *myl7* expression (L,N); asterisk indicates complete loss of ISH signal (M). (O) Phenotype penetrance and expressivity of *myl7* ISH signal in transcription factor site crispants. (P-S) Germline transmission of GATA site mutations. (P,Q) Control *myl7:EGFP* sibling. (R,S) GATA site germline mutant from crispant incross, featuring cardiac edema (arrowhead, R) and impaired EGFP expression (asterisk, S). (T) Allele sequences of endogenous and hemizygous transgene locus in an F1 embryo, featuring lesions in the predicted GATA site.

that can be phenotypically assessed in the F1 using crispant incrosses (Varshney et al., 2015). Despite the high mutagenesis efficiency, a major caveat for phenotype analysis using F0 crispants as phenotype readout remains the occurrence of unpredictable mosaic allele combinations. Previous studies hypothesized a high

somatic allele count (thousands) following Cas9-mediated mutagenesis in somatic tissue (Jao et al., 2013). We find significantly fewer alleles in crispants for selective genes than anticipated (Fig. 4D, Table S1, Figs S5-S7, S10), suggesting that, in the optimal case, mutagenesis by our solubilized RNPs is saturating

after few initial cell divisions. We do consistently recover in-frame indels, which potentially create hypomorphic or even functionally unaffected alleles (Gagnon et al., 2014) (Fig. 4A, Fig. S5). Appropriate design of the sgRNA targeting region in a given sequence might mitigate this issue. Nonetheless, the overall robustness of the DNA triplet code, translation initiation and splicing warrant close examination of every targeted locus. Our observations emphasize the importance of designing sgRNAs against gene regions that encode conserved domains or functional entities in the final protein product or RNA, and to assess crispant phenotypes with at least two distinct sgRNAs. This approach is potentially augmented by using design tools to pick highly efficient sgRNAs (Moreno-Mateos et al., 2015). For example, our mutagenesis of *gol* predominantly results in a 3 bp deletion that invariantly causes complete loss of pigmentation in crispants and mutant F1 zebrafish (Fig. 4A). This 3 bp deletion allele removes Val120 at the edge of the third predicted transmembrane helix of SLC24A5, suggesting functional impairment of protein structure as a result (Lamason et al., 2005). By contrast, our mutagenesis of *tbx16* caused in-frame alleles that have no phenotypic consequence in crispants and the F1 (Fig. S5K,L, Fig. S9). Mutagenesis of essential genes can potentially trigger strong selection against cells with detrimental alleles and the enrichment for cells with mutant alleles that retain function, such as possible downstream start codons when targeting the initial ATG or cryptic splicing upon targeting an exon-intron boundary. sgRNA prediction algorithms that assist in targeting functional features of the resulting proteins in the genomic coding sequence would potentially improve mutant phenotype penetrance and expressivity in the injected F0. Further, positive selection for deletion alleles that maintain protein function represents a potent proxy to uncover functional domains in protein-coding genes using germline mutants. Nonetheless, our observations, despite resulting from exceedingly efficient mutagenesis, call for careful interpretation of somatic mutagenesis analysis in F0 animals performed by somatic or inducible Cas9 expression in zebrafish and other *in vivo* models (Ablain et al., 2015; Stolfi et al., 2014; Yin et al., 2015). Conditional alleles made with floxed exons that will allow Cre-mediated tissue-specific inactivation are now in reach for the zebrafish field and will help in addressing this caveat.

The targeting of non-coding regulatory elements potentially requires less stringent sgRNA design, yet remains challenging owing to the relatively small size of potentially functional chromatin regions, such as transcription factor binding sites. Recent studies have successfully employed systematic enhancer targeting in cell culture systems (Canver et al., 2015; Korkmaz et al., 2016). Our efficient RNP-mediated mutagenesis approach outlined here now paves the way to extend systematic enhancer analysis to *in vivo* model systems. Although labor- and animal number-intensive, the generation of germline mutant allelic series of deletions/insertions spanning a particular transcription factor binding site or whole enhancer will allow for precise assessment of functional non-coding elements. Overall, our results suggest that crispant-based systematic functional assessment of non-coding genome elements is highly efficient in zebrafish (Fig. 5) and possibly represents a high-throughput platform to reveal the developmental contribution of regulatory elements.

MATERIALS AND METHODS
Zebrafish husbandry and experimentation
Zebrafish (*Danio rerio*) were maintained, collected and staged as described (Kimmel et al., 1995). Embryos were raised at 28.5°C if not stated otherwise.

Injections into phenotypically wild-type embryos were performed using WIK, Tü, and mixed WIK/Tü strains, with prior sequence verification of the target locus. Transgenic lines used in this study are ubiquitous for EGFP: *ubi:Switch* [*Tg(−3.5ubb:LOXP-EGFP-LOXP-mCherry)*, cz1701Tg] (Mosimann et al., 2011); reporter for *myl7:EGFP* [*Tg(−3.5ubb:Cre-ERT2, myl7:EGFP)*, cz1702Tg] (Mosimann et al., 2011); and *Tg(wt1b:EGFP)* (Perner et al., 2007). Detection of endogenous *myl7* expression by mRNA *in situ* hybridization was performed as described (Thisse and Thisse, 2008); all embryos were processed in parallel with identical staining times. Brightfield, *in situ* hybridization and basic fluorescence imaging were performed using a Leica M205FA with a DFC450 C camera; selective plane illumination microscopy (SPIM) was performed using a Zeiss Z.1. Images were processed using Leica LAS, ImageJ (NIH), Photoshop CS6 (Adobe) and PaintShop Pro 7 (Corel) software; whole adult captures are composites stitched using Photoshop CS6.

Cas9 protein production and storage
The Cas9-NLS-GFP and Cas9-NLS-mCherry proteins are composed of the polypeptide sequence of *Streptococcus pyogenes* Cas9 fused in-frame with a C-terminal HA epitope tag, a bipartite nuclear localization signal (NLS) sequence, a fluorescent protein (GFP or mCherry) polypeptide and an additional monopartite NLS at the very C-terminus (Fig. S1). The DNA sequence encoding the Cas9-NLS-GFP polypeptide was PCR amplified from plasmid pMJ920 (Jinek et al., 2013) (Addgene) using the following primers (5′-3′): forward, TACTTCCAATCCAATGCCACCATGGACAA-GAAGTACAGCATCGG; reverse, TTATCCACTTCCAATGTTATTACT-CAACTTTTCGTTTTTTCTTAGGTGACCCCTTGTACAGCTCGTCCA-TGCCG. The PCR product was inserted into expression plasmid *2C-T* (gift from S. Gradia, UC Berkeley Macro Lab; information available from Addgene) using ligation-independent cloning (LIC). The resulting expression plasmid (pMJ922, available from Addgene) produces Cas9-NLS-GFP as a fusion with an N-terminal hexahistidine-maltose binding protein (His$_6$-MBP) affinity tag that is removed by cleavage with Tobacco etch virus (TEC) protease during purification. The protein was expressed in *E. coli* Rosetta 2 cells as described (Anders et al., 2014). Cells were resuspended in 20 mM Tris, 250 mM NaCl, 5 mM imidazole pH 8.0 and lysed using a pressure homogenizer (Avestin). Cell lysate was clarified by centrifugation at 40,000 *g* for 45 min and applied to a 10 ml HIS-Select Ni column (Sigma-Aldrich). The column was washed extensively with 20 mM Tris, 250 mM NaCl, 10 mM imidazole pH 8.0 and eluted with 20 mM Tris, 250 mM NaCl, 250 mM imidazole pH 8.0. Eluted protein was dialyzed against 20 mM HEPES, 100 mM KCl, 10% glycerol, 1 mM dithiothreitol (DTT), 1 mM EDTA pH 7.5 overnight at 4°C in the presence of TEV protease to remove the His$_6$-MBP affinity tag. Cleaved protein was bound to a HiTrap SP FF cation exchange column (GE Healthcare) and eluted with a linear gradient of 0.1-1.0 M KCl. In a final polishing step, CAS-NLS-GFP was purified on a Superdex 200 16/600 size exclusion column (GE Healthcare), eluting in 20 mM HEPES, 100 mM KCl pH 7.5. The protein was concentrated to 15 mg/ml using a 50,000 MWCO centrifugal concentrator (Amicon) and 50 µl aliquots were flash-frozen in liquid nitrogen and stored at −80°C. We further made one-time use stocks of 2-3 µl in PCR tubes also stored at −80°C.

sgRNA production
Oligos were obtained from Life Technologies as standard primers except where otherwise noted. sgRNA templates were generated as described (Bassett et al., 2013; Gagnon et al., 2014; Hwang et al., 2013), either using cloning into *pDR274* (Hwang et al., 2013) followed by PCR amplification of the sgRNA template with (5′-3′) forward primer GCACCGCTAGCT-AATACG and reverse primer AAAAGCACCGACTCGGTGC, or oligo-based (Bassett et al., 2013; Gagnon et al., 2014) using the sgRNA forward primer for T7 templates GAAATTAATACGACTCACTATA-N20-GTTT-TAGAGCTAGAAATAGC or SP6 GAAATATTTAGGTGACACTATA-N20-GTTTTAGAGCTAGAAATAGC (with N20 indicating the target site) and the invariant reverse primer AAAAGCACCGACTCGGTGCCACTT-TTTCAAGTTGATAACGGACTAGCCTTATTTTAACTTGCTATTTCT-AGCTCTAAAAC (PAGE-purified, Life Technologies). Our sgRNA nomenclature uses the abbreviation *cc* (*crispr cutter*) followed by an

indexing letter (i.e. *ccA*, *ccB*, etc.) to distinguish sgRNAs targeting the same gene. Primer extension was performed using Phusion polymerase (NEB) followed by QIAquick purification (Qiagen) with elution in DEPC-treated water. *In vitro* transcription of sgRNAs based on the templates above was performed using the MAXIscript T7 or SP6 Kit (Ambion) with the reaction run at 37°C overnight, followed by ammonium acetate precipitation as per the manufacturer's protocol and as described previously (Bassett et al., 2013; Gagnon et al., 2014). We found that adding NTPs at 100 mM instead of the 10 mM recommended in the manufacturer's protocol greatly increases sgRNA yield. Precipitated sgRNA pellets were visualized with GlycoBlue (Life Technologies). Before use, all sgRNAs were quality controlled on denaturing 2.5% MOPS gels. Oligos used in this study are listed in Table S2.

Complex assembly, salt stabilization and microinjection

Injection mixes contained, as standard, 831 ng/μl Cas9-EGFP or Cas9-mCherry, with purified sgRNA added. We recommend 800-900 ng/μl Cas9 per injection mix as a starting amount. 900 ng Cas9-EGFP (191.2 kDa) corresponds to $4.7×10^{-9}$ mol active sgRNA-Cas9 complexes, which were formed by mixing sgRNA and Cas9 protein buffered with KCl (2 M stock added for final concentration of 300 mM) and incubation for 5 min at 37°C. Injection mixes were then used directly without further storage.

To facilitate the setup of injection mixes, we developed the web-based and smart-phone tool CrispantCal to calculate injection mix volumes corresponding to an optimal ratio of gRNA to Cas9 protein molecules. The CrispantCal software allows the user to enter molecular properties and concentrations of gRNA and Cas9 protein samples, volume of Cas9 solution, and desired total volume. Optionally, desired final concentration of Cas9 can be specified instead of total volume. Volumes for an optimal injection mix ratio are then calculated and displayed. Additional volume of KCl diluent required for optimal reaction efficiency can also be calculated. This tool calculates volumes corresponding to a perfect 1:1 mix ratio of gRNA to Cas9 molecules in an injection. The 'KCl diluent' option calculates the additional volume of KCl diluent required to increase the concentration in the injection mix to a desired final value; recommended is 300 mM KCl. Of note, CrispantCal also provides guidelines for generating RNP injection mixes with two sgRNAs to generate mutants with larger targeted deletions. See the main text for a discussion on this approach.

The web-based CrispantCal tool was developed using the Shiny web application framework (RStudio, http://shiny.rstudio.com/) for the statistical programming language R. The tool is accessible online together with further details on usage and calculations at http://lmweber.github.io/CrispantCal/antCal/, or within R using the commands *install.packages ("shiny"); shiny::runGitHub("lmweber/CrispantCal")*. The smart-phone app for Android and iOS platforms was written in Java and Objective-C, respectively. After user input of the concentration of the corresponding Cas9 protein version, the KCl concentration in the Cas9 protein stock, and the sgRNA stock concentration, the tool calculates the amounts of water, protein stock and sgRNA, which are displayed as injection mix. The code for these applications can be found at https://bitbucket.org/raulcatena/crispantcal and https://bitbucket.org/raulcatena/crispantcal-android, and the compiled applications are freely available from Google Play Store (Android) and iTunes AppStore (iOS).

Microinjection

Microinjections were performed using MPPI-3 pressure injector units (ASI) with needles pulled from filamented capillaries (WPI) on a Sutter Instrument P-97 and guided with Narishige M-152 micro-manipulators. Injection droplets were calibrated to ∼100-125 μm at the start of injection, resulting in 0.5-1.5 nl injection mix delivered into the embryo cell, unless noted otherwise.

Fluorescence quantification

The quantification of GFP signal reduction upon sgRNA targeting was performed on images taken with identical settings on a Leica M205FA with a DFC450 C camera by RGB analysis (additive red-green-blue color space)

gated around the embryo outline in ImageJ 1.46r software. Intensity values for each individual gate were statistically analyzed in GraphPad Prism 5.

Immunoblotting

mCherry-tagged RNPs were injected into one-cell stage *ubi:EGFP* embryos and injections were quality controlled for mCherry fluorescence. At 48 hpf, protein was isolated from 30 uninjected control embryos and 30 RNP-injected embryos and processed as SDS samples on MINI-Protean TGX gels (4-20%; BioRad) with running buffer [25 mM Tris, 192 mM glycine, 0.01% (w/v) SDS; if necessary, adjust the pH to 8.3]. Blot transfer was performed using Trans-Blot Turbo nitrocellulose membrane (BioRad), followed by blocking (5% milk) and washing in TBST (150 mM NaCl, 50 mM Tris pH 7.5, 0.1% Tween 20). To probe the blot, anti-GFP (#11814460001, Roche; 1:1000) and mouse anti-Tubulin (sc-32293, Santa Cruz; 1:2000) antibodies were used. HRP-conjugated goat anti-mouse (115-035-003, Jackson Laboratories; 1:5000) secondary antibody was used. The blot was developed using Western Bright ECL solution (Advansta) and detected on an ImageQuant LAS 4000 imager (GE Healthcare Life Sciences).

Molecular analysis

To isolate genomic DNA from crispants or F1 mutants, single embryos of appropriate stages were incubated in 50 μl alkaline lysis buffer (25 mM NaOH, 0.2 mM disodium EDTA, pH 12.0) at 95°C for 30 min (Mosimann et al., 2013). The samples were then quenched on ice and neutralized with 5 μl of 1 M Tris-HCl pH 8.0). Debris was removed by centrifugation at 5000 *g* for 5 min and transferring the supernatant to fresh tubes. The supernatant was kept at 4°C for short-term or at −20°C for long-term storage.

sgRNA target sites were amplified with flanking primers designed to amplify 350-500 bp of the individual genes (see Table S2 for sequences) using GoTaq G2 Green Master Mix (Promega). Reactions were performed in a total volume of 25 μl according to the manufacturer's instructions, using 1 μl template DNA. Annealing temperature and elongation time were adjusted to individual primers and product length according to manufacturer's instructions. Primers used for allele analysis are listed in Table S2.

PCR products were purified with the QIAquick Gel Extraction Kit (Qiagen) and subcloned using the pGEM T-Easy system (Promega). Successful ligation was confirmed with the ready-to-use X-gal solution system (ThermoScientific) with readout as white colonies. Colony PCRs were performed using T7 and SP6 primers using GoTaq G2 Green reaction mix. We refined three protocols for colony PCR and subsequent purification for sequencing.

For procedure 1, white clones were suspended in 20 μl LB medium. Successful clones were confirmed in a 10 μl GoTaq G2 Green reaction using 0.5 μl colony suspension as template. Successful clones were expanded overnight in 2 ml LB medium (Amp 1:1000) and plasmids were purified with the QIAprep Spin MiniPrep Kit (Qiagen). The target region was sequenced with the T7 primer.

For procedure 2, white clones were suspended in 20 μl LB medium. A GoTaq G2 Green PCR reaction was set up with 1 μl suspended colony as template in a total volume of 40 μl. The PCR products were analyzed by agarose gel electrophoresis. Bands of correct size were excised and the DNA-containing gel pieces frozen at −20°C for at least 1 h. Subsequently, the gel pieces were thawed, resulting in the release of the DNA from the gel through its porous matrix. The samples were spun down and the maximum volume of the DNA-containing flow-through was subject to sequencing with the T7 primer.

For procedure 3, which we now perform as standard, a 5 μl GoTaq G2 Green reaction mix was set up as a master mix. White clones were picked and briefly dipped into the 5 μl reaction mix. The PCR reaction was performed according to the manufacturer's instructions. Final PCR products were treated with 10 U exonuclease I (Exo I, ThermoScientific) and 1 U rAPid alkaline phosphatase (Roche Diagnostics) at 37°C for 20 min. The enzymes were then heat inactivated at 95°C for 20 min and the samples sequenced with the T7 primer. This procedure allows simple scaling-up.

Deep-sequencing and computational sequence analysis

sgRNA target sites were amplified as described above and quality controlled on a 2% agarose gel. PCR fragments were then processed by NXT-Dx (Ghent, Belgium) for multiplexed Illumina MiSeq PE250 amplicon sequencing.

MiSeq reads were aligned with bwa mem 0.7.12-r1039 (Li, 2013 preprint) to the zebrafish genome version danRer7 (Zv9). Reads were separated by amplicon sequence after mapping by matching the mapped endpoints of each pair with the genomic locations of the amplicons. Mapped read endpoints were required to be within 5 bases of the expected locations. Sanger sequences were extracted using sangerseqR (Hill et al., 2014) and CrispRVariants (Lindsay et al., 2015 preprint) (see below), with base recalling to resolve ambiguous bases, as we found this method to reduce allele count in informal benchmarking of F1 samples (containing only two alleles). Subsequently, the Sanger data were processed similarly to the MiSeq data, except for the amplicon separation step.

We developed an R software package named CrispRVariants to perform the variant counting and visualization (Lindsay et al., 2015 preprint). Variants were counted within the region from 5 bp upstream of the sgRNA to 5 bp downstream of the protospacer adjacent motif (PAM). As single nucleotide variants (SNVs) near the PAM can prevent cutting, or can result from repairing a cut, sequences without either an insertion or a deletion were separated into those containing a SNV and those matching the reference. SNVs were identified in the region 8 bases upstream to 6 bases downstream of the cut site. For calculating mutation efficiency, we first identified SNVs with a frequency of at least 20% in at least one sample. Less frequent SNVs were considered non-variant sequences, i.e. only sequences with insertions or deletions were counted as variants. By inspecting the CrispRVariants allele summary plots, we further identified and removed three pre-existing insertion variants located away from the cut sites from the efficiency calculations: (1) *gol_ccB_off0 1: 6:1I*, (2) *hand2_ccB_off0 1: -11:16I* and (3) *hand2_ccA,hand2_ccB: 26:9I* (between the two sgRNA locations). When counting the number of (indel) variant alleles, a 1% frequency cutoff was used for the MiSeq data to avoid inflating the allele counts by including rare sequencing errors. No cutoff was used for the Sanger data. Variant locations with respect to Zv9 annotated genes were determined using the R Bioconductor package VariantAnnotation (Obenchain et al., 2014). When variant locations differed between transcripts of a gene, the location most likely to affect the protein sequence was used, with splice sites considered more consequential than exonic sites.

Comparison with the sequencing data from Gagnon et al. (2014) was performed with the original data kindly provided by the authors. All MiSeq data have been deposited at ArrayExpress with accession E-MTAB-4143. CrispRVariants is accessible and explained in detail in Lindsay et al. (2015 preprint).

Author contributions
C.M. and M.J. conceived the project; A.B., A.F., C.H., E.C. and J.Z. performed zebrafish experiments and data analysis; H.L. and M.D.R. performed data analysis and CrispRVariants coding; C.A. performed protein work; L.M.W. coded the CrispantCal web app; R.C. coded the CrispantCal smart-phone apps; A.B., H.L., M.J., M.D.R. and C.M. prepared and edited the manuscript.

Data availability
MiSeq data have been deposited at ArrayExpress with accession E-MTAB-4143.

Supplementary information
Supplementary information available online at
http://dev.biologists.org/lookup/suppl/doi:10.1242/dev.134809/-/DC1

References

Ablain, J., Durand, E. M., Yang, S., Zhou, Y. and Zon, L. I. (2015). A CRISPR/Cas9 vector system for tissue-specific gene disruption in zebrafish. *Dev. Cell* **32**, 756-764.

Ahn, D.-G., Kourakis, M. J., Rohde, L. A., Silver, L. M. and Ho, R. K. (2002). T-box gene tbx5 is essential for formation of the pectoral limb bud. *Nature* **417**, 754-758.

Anders, C., Niewoehner, O., Duerst, A. and Jinek, M. (2014). Structural basis of PAM-dependent target DNA recognition by the Cas9 endonuclease. *Nature* **513**, 569-573.

Auer, T. O., Duroure, K., De Cian, A., Concordet, J.-P. and Del Bene, F. (2014a). Highly efficient CRISPR/Cas9-mediated knock-in in zebrafish by homology-independent DNA repair. *Genome Res.* **24**, 142-153.

Auer, T. O., Duroure, K., Concordet, J.-P. and Del Bene, F. (2014b). CRISPR/Cas9-mediated conversion of eGFP- into Gal4-transgenic lines in zebrafish. *Nat. Protoc.* **9**, 2823-2840.

Bassett, A. R., Tibbit, C., Ponting, C. P. and Liu, J.-L. (2013). Highly efficient targeted mutagenesis of Drosophila with the CRISPR/Cas9 system. *Cell Rep.* **4**, 220-228.

Bedell, V. M., Wang, Y., Campbell, J. M., Poshusta, T. L., Starker, C. G., Krug, R. G., II, Tan, W., Penheiter, S. G., Ma, A. C., Leung, A. Y. H. et al. (2012). In vivo genome editing using a high-efficiency TALEN system. *Nature* **491**, 114-118.

Bill, B. R., Petzold, A. M., Clark, K. J., Schimmenti, L. A. and Ekker, S. C. (2009). A primer for morpholino use in zebrafish. *Zebrafish* **6**, 69-77.

Canver, M. C., Smith, E. C., Sher, F., Pinello, L., Sanjana, N. E., Shalem, O., Chen, D. D., Schupp, P. G., Vinjamur, D. S., Garcia, S. P. et al. (2015). BCL11A enhancer dissection by Cas9-mediated in situ saturating mutagenesis. *Nature* **527**, 192-197.

Chang, N., Sun, C., Gao, L., Zhu, D., Xu, X., Zhu, X., Xiong, J.-W. and Xi, J. J. (2013). Genome editing with RNA-guided Cas9 nuclease in zebrafish embryos. *Cell Res.* **23**, 465-472.

Chiavacci, E., Dolfi, L., Verduci, L., Meghini, F., Gestri, G., Evangelista, A. M. M., Wilson, S. W., Cremisi, F. and Pitto, L. (2012). MicroRNA 218 mediates the effects of Tbx5a over-expression on zebrafish heart development. *PLoS ONE* **7**, e50536.

Dahlem, T. J., Hoshijima, K., Jurynec, M. J., Gunther, D., Starker, C. G., Locke, A. S., Weis, A. M., Voytas, D. F. and Grunwald, D. J. (2012). Simple methods for generating and detecting locus-specific mutations induced with TALENs in the zebrafish genome. *PLoS Genet.* **8**, e1002861.

Doyon, Y., McCammon, J. M., Miller, J. C., Faraji, F., Ngo, C., Katibah, G. E., Amora, R., Hocking, T. D., Zhang, L., Rebar, E. J. et al. (2008). Heritable targeted gene disruption in zebrafish using designed zinc-finger nucleases. *Nat. Biotechnol.* **26**, 702-708.

Gagnon, J. A., Valen, E., Thyme, S. B., Huang, P., Ahkmetova, L., Pauli, A., Montague, T. G., Zimmerman, S., Richter, C. and Schier, A. F. (2014). Efficient mutagenesis by Cas9 protein-mediated oligonucleotide insertion and large-scale assessment of single-guide RNAs. *PLoS ONE* **9**, e98186.

Garrity, D. M., Childs, S. and Fishman, M. C. (2002). The heartstrings mutation in zebrafish causes heart/fin Tbx5 deficiency syndrome. *Development* **129**, 4635-4645.

Hill, J. T., Demarest, B. L., Bisgrove, B. W., Su, Y.-C., Smith, M. and Yost, H. J. (2014). Poly peak parser: method and software for identification of unknown indels using sanger sequencing of polymerase chain reaction products. *Dev. Dyn.* **243**, 1632-1636.

Hruscha, A., Krawitz, P., Rechenberg, A., Heinrich, V., Hecht, J., Haass, C. and Schmid, B. (2013). Efficient CRISPR/Cas9 genome editing with low off-target effects in zebrafish. *Development* **140**, 4982-4987.

Huang, C.-J., Tu, C.-T., Hsiao, C.-D., Hsieh, F.-J. and Tsai, H.-J. (2003). Germ-line transmission of a myocardium-specific GFP transgene reveals critical regulatory elements in the cardiac myosin light chain 2 promoter of zebrafish. *Dev. Dyn.* **228**, 30-40.

Hwang, W. Y., Fu, Y., Reyon, D., Maeder, M. L., Tsai, S. Q., Sander, J. D., Peterson, R. T., Yeh, J.-R. J. and Joung, J. K. (2013). Efficient genome editing in zebrafish using a CRISPR-Cas system. *Nat. Biotechnol.* **31**, 227-229.

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

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

DEVELOPMENT

Jinek, M., East, A., Cheng, A., Lin, S., Ma, E. and Doudna, J. (2013). RNA-programmed genome editing in human cells. *eLife* **2**, e00471.

Kimmel, C. B., Kane, D. A., Walker, C., Warga, R. M. and Rothman, M. B. (1989). A mutation that changes cell movement and cell fate in the zebrafish embryo. *Nature* **337**, 358-362.

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

Kimura, Y., Hisano, Y., Kawahara, A. and Higashijima, S.-i. (2014). Efficient generation of knock-in transgenic zebrafish carrying reporter/driver genes by CRISPR/Cas9-mediated genome engineering. *Sci. Rep.* **4**, 6545.

Kok, F. O., Shin, M., Ni, C.-W., Gupta, A., Grosse, A. S., van Impel, A., Kirchmaier, B. C., Peterson-Maduro, J., Kourkoulis, G., Male, I. et al. (2015). Reverse genetic screening reveals poor correlation between morpholino-induced and mutant phenotypes in zebrafish. *Dev. Cell* **32**, 97-108.

Korkmaz, G., Lopes, R., Ugalde, A. P., Nevedomskaya, E., Han, R., Myacheva, K., Zwart, W., Elkon, R. and Agami, R. (2016). Functional genetic screens for enhancer elements in the human genome using CRISPR-Cas9. *Nat. Biotechnol.* **34**, 192-198.

Kotani, H., Taimatsu, K., Ohga, R., Ota, S. and Kawahara, A. (2015). Efficient multiple genome modifications induced by the crRNAs, tracrRNA and Cas9 protein complex in zebrafish. *PLoS ONE* **10**, e0128319.

LaFave, M. C., Varshney, G. K., Vemulapalli, M., Mullikin, J. C. and Burgess, S. M. (2014). A defined zebrafish line for high-throughput genetics and genomics: NHGRI-1. *Genetics* **198**, 167-170.

Lamason, R. L., Mohideen, M.-A. P. K., Mest, J. R., Wong, A. C., Norton, H. L., Aros, M. C., Jurynec, M. J., Mao, X., Humphreville, V. R., Humbert, J. E. et al. (2005). SLC24A5, a putative cation exchanger, affects pigmentation in zebrafish and humans. *Science* **310**, 1782-1786.

Li, H. (2013). Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXiv* doi:1303.3997.

Lindsay, H., Burger, A., Felker, A., Hess, C., Zaugg, J., Chiavacci, E., Anders, C., Jinek, M., Mosimann, C. and Robinson, M. D. (2015). CrispRVariants: precisely charting the mutation spectrum in genome engineering experiments. *bioRxiv* doi:10.1101/034140.

Moreno-Mateos, M. A., Vejnar, C. E., Beaudoin, J.-D., Fernandez, J. P., Mis, E. K., Khokha, M. K. and Giraldez, A. J. (2015). CRISPRscan: designing highly efficient sgRNAs for CRISPR-Cas9 targeting in vivo. *Nat. Methods* **12**, 982-988.

Mosimann, C., Kaufman, C. K., Li, P., Pugach, E. K., Tamplin, O. J. and Zon, L. I. (2011). Ubiquitous transgene expression and Cre-based recombination driven by the ubiquitin promoter in zebrafish. *Development* **138**, 169-177.

Mosimann, C., Puller, A.-C., Lawson, K. L., Tschopp, P., Amsterdam, A. and Zon, L. I. (2013). Site-directed zebrafish transgenesis into single landing sites with the phiC31 integrase system. *Dev. Dyn.* **242**, 949-963.

Obenchain, V., Lawrence, M., Carey, V., Gogarten, S., Shannon, P. and Morgan, M. (2014). VariantAnnotation: a Bioconductor package for exploration and annotation of genetic variants. *Bioinformatics* **30**, 2076-2078.

Perner, B., Englert, C. and Bollig, F. (2007). The Wilms tumor genes wt1a and wt1b control different steps during formation of the zebrafish pronephros. *Dev. Biol.* **309**, 87-96.

Platt, R. J., Chen, S., Zhou, Y., Yim, M. J., Swiech, L., Kempton, H. R., Dahlman, J. E., Parnas, O., Eisenhaure, T. M., Jovanovic, M. et al. (2014). CRISPR-Cas9 knockin mice for genome editing and cancer modeling. *Cell* **159**, 440-455.

Schulte-Merker, S. and Stainier, D. Y. R. (2014). Out with the old, in with the new: reassessing morpholino knockdowns in light of genome editing technology. *Development* **141**, 3103-3104.

Shah, A. N., Davey, C. F., Whitebirch, A. C., Miller, A. C. and Moens, C. B. (2015). Rapid reverse genetic screening using CRISPR in zebrafish. *Nat. Methods* **12**, 535-540.

Stainier, D. Y., Fouquet, B., Chen, J. N., Warren, K. S., Weinstein, B. M., Meiler, S. E., Mohideen, M. A., Neuhauss, S. C., Solnica-Krezel, L., Schier, A. F. et al. (1996). Mutations affecting the formation and function of the cardiovascular system in the zebrafish embryo. *Development* **123**, 285-292.

Stolfi, A., Gandhi, S., Salek, F. and Christiaen, L. (2014). Tissue-specific genome editing in Ciona embryos by CRISPR/Cas9. *Development* **141**, 4115-4120.

Sung, Y. H., Kim, J. M., Kim, H.-T., Lee, J., Jeon, J., Jin, Y., Choi, J.-H., Ban, Y. H., Ha, S.-J., Kim, C.-H. et al. (2014). Highly efficient gene knockout in mice and zebrafish with RNA-guided endonucleases. *Genome Res.* **24**, 125-131.

Thisse, C. and Thisse, B. (2008). High-resolution in situ hybridization to whole-mount zebrafish embryos. *Nat. Protoc.* **3**, 59-69.

Varshney, G. K., Pei, W., LaFave, M. C., Idol, J., Xu, L., Gallardo, V., Carrington, B., Bishop, K., Jones, M., Li, M. et al. (2015). High-throughput gene targeting and phenotyping in zebrafish using CRISPR/Cas9. *Genome Res.* **25**, 1030-1042.

Xiao, A., Cheng, Z., Kong, L., Zhu, Z., Lin, S., Gao, G. and Zhang, B. (2014). CasOT: a genome-wide Cas9/gRNA off-target searching tool. *Bioinformatics* **30**, 1180-1182.

Yelon, D., Ticho, B., Halpern, M. E., Ruvinsky, I., Ho, R. K., Silver, L. M. and Stainier, D. Y. (2000). The bHLH transcription factor hand2 plays parallel roles in zebrafish heart and pectoral fin development. *Development* **127**, 2573-2582.

Yin, L., Maddison, L. A., Li, M., Kara, N., LaFave, M. C., Varshney, G. K., Burgess, S. M., Patton, J. G. and Chen, W. (2015). Multiplex conditional mutagenesis using transgenic expression of Cas9 and sgRNAs. *Genetics* **200**, 431-441.

DEVELOPMENT