

Development 138, 4555-4564 (2011) doi:10.1242/dev.066779
 © 2011. Published by The Company of Biologists Ltd

Evaluation and application of modularly assembled zinc-finger nucleases in zebrafish

Cong Zhu^{1,*}, Tom Smith^{1,*}, Joseph McNulty¹, Amy L. Rayla¹, Abirami Lakshmanan¹, Arndt F. Siekmann^{1,†}, Matthew Buffardi¹, Xiangdong Meng^{1,‡}, Jimann Shin², Arun Padmanabhan³, Daniel Cifuentes⁴, Antonio J. Giraldez⁴, A. Thomas Look², Jonathan A. Epstein³, Nathan D. Lawson^{1,§} and Scot A. Wolfe^{1,5,§}

SUMMARY

Zinc-finger nucleases (ZFNs) allow targeted gene inactivation in a wide range of model organisms. However, construction of target-specific ZFNs is technically challenging. Here, we evaluate a straightforward modular assembly-based approach for ZFN construction and gene inactivation in zebrafish. From an archive of 27 different zinc-finger modules, we assembled more than 70 different zinc-finger cassettes and evaluated their specificity using a bacterial one-hybrid assay. In parallel, we constructed ZFNs from these cassettes and tested their ability to induce lesions in zebrafish embryos. We found that the majority of zinc-finger proteins assembled from these modules have favorable specificities and nearly one-third of modular ZFNs generated lesions at their targets in the zebrafish genome. To facilitate the application of ZFNs within the zebrafish community we constructed a public database of sites in the zebrafish genome that can be targeted using this archive. Importantly, we generated new germline mutations in eight different genes, confirming that this is a viable platform for heritable gene inactivation in vertebrates. Characterization of one of these mutants, *gata2a*, revealed an unexpected role for this transcription factor in vascular development. This work provides a resource to allow targeted germline gene inactivation in zebrafish and highlights the benefit of a definitive reverse genetic strategy to reveal gene function.

KEY WORDS: *gata2*, Vascular development, Zebrafish, Zinc-finger nuclease

INTRODUCTION

The attributes of the zebrafish have established it as a powerful model for the study of vertebrate development. The accessibility of its externally fertilized embryos allows numerous manipulations, and the transparency of the zebrafish embryo permits direct serial visualization of biological processes in vivo (Beis and Stainier, 2006). These characteristics have enabled forward genetic screens for mutants affecting a wide range of developmental processes (Patton and Zon, 2001). However, the size of the zebrafish genome, the relatively long generation time, and the expense of maintaining large populations preclude forward genetic screening to saturation. Consequently, mutants exist for only a small fraction of zebrafish genes. Although modified antisense oligonucleotides can achieve targeted gene knockdown during embryogenesis (Nasevicius and Ekker, 2000), this technology is problematic with regard to persistence, specificity and penetrance (Robu et al., 2007; Eisen

and Smith, 2008). Thus, definitive reverse genetic strategies are required, especially for the study of later developmental processes, analysis of adult physiology, and for the establishment of disease models.

Zinc-finger nucleases (ZFNs) are chimeric fusions between a zinc-finger protein (ZFP) and the nuclease domain of FokI (Urnov et al., 2010). ZFNs have been employed to achieve heritable targeted gene disruption in the genomes of numerous plant and animal species, including zebrafish (Bibikova et al., 2002; Beumer et al., 2008; Doyon et al., 2008; Meng et al., 2008; Foley et al., 2009; Geurts et al., 2009; Cui et al., 2010; Mashimo et al., 2010; Meyer et al., 2010; Takasu et al., 2010). Gene disruption is achieved through imprecise repair of a ZFN-induced double-strand break within the coding sequence of a target gene (Urnov et al., 2010). ZFNs have also been utilized to achieve tailor-made genomic alterations in animal genomes through stimulation of homology-directed repair from an exogenous donor DNA (Beumer et al., 2008; Cui et al., 2010; Meyer et al., 2010). Although off-target lesions in the genome can result from ZFN treatment, their frequency is significantly lower than at the target site and their location can be predicted based on ZFN specificity (Meng et al., 2008; Perez et al., 2008; Gupta et al., 2011). Together, these characteristics make ZFNs an ideal tool to facilitate reverse genetic studies in zebrafish and other organisms.

Despite their utility, the widespread implementation of ZFNs is hindered by the difficulty in creating highly specific ZFPs. Two general approaches have been employed for this purpose: selection-based methods, which identify ZFPs with a desired specificity from randomized libraries, and assembly-based methods, which utilize archives of predefined zinc fingers to construct ZFPs with a desired specificity. Each approach is hampered by limitations. Selection-based approaches (Greisman and Pabo, 1997; Isalan et al., 1998;

¹Program in Gene Function and Expression, University of Massachusetts Medical School, Worcester, MA 01605, USA. ²Department of Pediatric Oncology, Dana-Farber Cancer Institute, Harvard Medical School, Boston, MA 02115, USA.

³Department of Cell and Developmental Biology, Perelman School of Medicine at the University of Pennsylvania, Philadelphia, PA 19104-6058, USA. ⁴Department of Genetics, Yale University School of Medicine, New Haven, CT 06510, USA.

⁵Department of Biochemistry and Molecular Pharmacology, University of Massachusetts Medical School, Worcester, MA 01605, USA.

*These authors contributed equally to this work

[†]Present address: Max-Planck Institute for Molecular Biomedicine, Laboratory for Cardiovascular Patterning, Roentgenstr. 20, 48149 Muenster, Germany

[‡]Present address: Bio-Rad Laboratories, Gene Expression Division, 2000 Alfred Noble Drive, Hercules, CA 94547, USA

[§]Authors for correspondence (nathan.lawson@umassmed.edu; scot.wolfe@umassmed.edu)

Maeder et al., 2008; Meng et al., 2008), although effective in generating specific ZFPs, can be time-consuming and technically challenging (Kim et al., 2010). Assembly-based approaches are more straightforward (Carroll et al., 2006; Kim et al., 2009) but are dependent on the quality of the available zinc-finger archive and the difficulty in predicting context-dependent effects between neighboring zinc fingers (Desjarlais and Berg, 1993; Greisman and Pabo, 1997; Wolfe et al., 1999). Although assembly-based, or modular, ZFNs have suffered from relatively low success rates (Ramirez et al., 2008), continued characterization of available archives and investigation of context dependence (Carroll et al., 2006; Kim et al., 2009; Sander et al., 2009), along with the generation of new archives with expanded specificity, will make this a more viable approach.

In this study we evaluated an assembly-based approach for creating ZFNs for targeted gene disruption in the zebrafish genome. We constructed an archive of ZFP modules recognizing 27 different triplet sequences and characterized the DNA-binding specificity for three-finger ZFP cassettes assembled from this archive. In parallel, we tested ZFNs constructed from these ZFPs for their ability to induce both somatic and germline lesions in zebrafish embryos. To facilitate the application of this technology in the zebrafish community, we developed a publicly accessible database that assists users in modular ZFN construction using our archive. Finally, we characterized the phenotypes in zebrafish embryos bearing a ZFN-induced deletion in the *gata2a* gene, revealing its role in vascular morphogenesis.

MATERIALS AND METHODS

Fish lines

Zebrafish were handled according to established protocols (Westerfield, 1993) and in accordance with Institutional Animal Care and Use Committee (IACUC) guidelines of participating institutions (University of Massachusetts Medical School, University of Pennsylvania, Dana Farber Cancer Institute and Yale University). The *Tg(kdrl:egfp)^{aj116}* line has been described elsewhere (Choi et al., 2007).

Construction of a modular zinc-finger archive

An archive of zinc-finger modules originating from multiple sources (see Table S1 in the supplementary material and Results) was constructed by subcloning ZFPs into either a pBluescript or pCS2 vector using standard methods. Clones were sequence verified and arrayed into a 96-well plate to generate an address for each module. Plasmids from the archive are available through addgene.org.

Modular assembly of ZFNs

To assemble three-finger ZFPs we first separately amplified modules for each of the three finger positions from plasmid templates by PCR. Distinct primer pairs were used to amplify modules for each position (fingers 1, 2 and 3) (see Table S2 in the supplementary material). PCR was performed using Advantage HF2 polymerase (Clontech) and 20 ng plasmid template. Finger 1 modules were amplified as follows: 94°C for 2 minutes, followed by 20–30 cycles of 94°C for 30 seconds, 60°C for 30 seconds and 68°C for 20 seconds, with a final extension of 68°C for 5 minutes. For finger 2 or 3 modules: 94°C for 2 minutes, then 20–30 cycles of 94°C for 30 seconds, 57°C for 30 seconds and 68°C for 20 seconds, with a final extension of 68°C for 5 minutes. All initial PCR products were gel purified. Three-finger cassettes were constructed by overlapping PCR using 30 ng of each finger module in a 50 µl reaction with Advantage HF2 polymerase. The first five cycles were performed without primers: 94°C for 2 minutes, followed by five cycles of 94°C for 30 seconds, 55°C for 30 seconds and 68°C for 1 minute. We added 2 µl of 10 µM F1 forward and F3 reverse primer and performed the following steps: 24 cycles of 94°C for 15 seconds, 60°C for 30 seconds, 68°C for 1 minute, with a final extension of 68°C for 5 minutes. Final PCR products were gel purified, digested with *Acc65I* and *BamHI* and cloned into

the pCS2 nuclease backbone in frame with the EL/KK heterodimeric variants of FokI (Miller et al., 2007), such that the 5' ZFP was fused to the EL nuclease and the 3' ZFP was fused to the KK nuclease. In several cases, DD/RR nuclease variants (Miller et al., 2007; Szczepek et al., 2007) were used. All constructs were sequence verified.

ZFN injections into zebrafish embryos

Synthesis and injection of ZFN mRNAs, dose optimization (see Table S3 in the supplementary material) and establishment and identification of germline lesions were performed as previously described (Meng et al., 2008; Gupta et al., 2011).

Bacterial one-hybrid binding site selections

Sequence-verified ZFP cassettes were cloned into the 1352-UV2 expression vector using *Acc65I* and *BamHI*. Bacterial one-hybrid (BIH) binding site selections were typically carried out at 5 mM 3-amino-1,2,4-triazole (3-AT), 10 µM IPTG and in the absence of uracil as previously described (Noyes et al., 2008; Gupta et al., 2011). For some ZFPs, lower stringency was required to achieve sufficient enrichment of recognition sequences above background.

Illumina library preparation of BIH-selected binding sites

Preparation of selected binding sites for deep sequencing was performed as described (Gupta et al., 2011), except that amplicons were digested with either *EcoRI* or *NotI* prior to ligation of bar-coded adapters for the identification of sequences associated with each ZFP (see Table S2 in the supplementary material).

Evaluation of the DNA-binding specificity of each ZFP and individual zinc-finger modules

Recognition motifs within the population of unique 28 bp sequences recovered from each binding site selection were identified using MEME (Bailey and Elkan, 1994). Position weight matrices (PWMs) were generated using the Log-Odds method (Hertz and Stormo, 1999) from the aligned sequences within the most statistically significant motif, weighting each sequence based on the number of counts within the Illumina dataset using the formula:

$$S = \sum_{i=1}^9 \log_2 \frac{f_{b,i}}{p_{b,i}}, \quad (1)$$

where S is the PWM score over the 9 bp target site s , b, i is the identity of base b at position i , $f_{b,i}$ is the normalized frequency of occurrence of base b at position i , and $p_{b,i}$ is the probability of observing the same base at the same position in a background model, which is assumed to be an equal distribution. Each ZFP was evaluated by calculating the score for its target site from its derived PWM. Each individual finger module was evaluated by calculating the score for its corresponding 3 bp recognition element (assuming canonical recognition) from this PWM. When data were available for multiple identical modules at identical positions, the scores for these modules were averaged for an overall assessment of the quality of the finger module. Modules with (average) scores that were two standard deviations below the average score of all modules at that position (i.e. finger 1, finger 2 or finger 3) were considered 'questionable' and flagged.

Analysis of somatic lesion frequency

Somatic lesion frequency was determined by Illumina sequencing as previously described (run #1) (Meng et al., 2008; Gupta et al., 2011). A second round of sequencing was performed for a subset of ZFNs from an independent set of injections (run #2). To avoid cross-contamination with samples from the first analysis, a unique pair of PCR primers was designed for each target site that would not amplify the PCR products from the first trial. These primers encoded the Illumina P1 and P2 sequences within each primer pair (see Table S2 in the supplementary material). In this case, PCR products for each ZFN target site were gel purified and amplified using Illumina genomic DNA primers (1.1 and 2.1). The PCR products for each target were then pooled at equal molar ratio, sequenced at 7 pM on a HiSeq 2000 (Illumina), and analyzed in a manner identical to run #1.

Calculation of the occupancy probability for a ZFN pair at its target site

The probability of a ZFP occupying its 9 bp target site was calculated using statistical thermodynamic free energy derived from a modified PWM score based on its B1H-determined DNA-binding specificity:

$$\Delta G = -RT \cdot \left(\sum_{i=1}^9 \ln \frac{f_{b,i}}{p_{b,i}} \right), \quad (2)$$

where $f_{b,i}$ and $p_{b,i}$ are defined as in (1), $R=0.001987$ kcal/(mol.K) and $T=301.5K$ (28.5°C).

All 9mers within the zebrafish genome (Zv8) were extracted considering both strands then the free energy of binding each 9mer by each ZFP was calculated using the scoring function shown in formula (2). These energies were then used to calculate the probability (P) for on-target binding by the ZFP within the zebrafish genome using:

$$P = \frac{e^{-\frac{\Delta G_s}{RT}}}{\sum_{i=1}^l n_i e^{-\frac{\Delta G_{s,i}}{RT}}}, \quad (3)$$

where ΔG_s is the free energy of binding to the 9mer matching the desired ZFP target site, $\Delta G_{s,i}$ is the free energy of binding of the i th 9mer, n_i is the occurrence of the i th 9mer in the genome, and l is total number of 9mers. A threshold for $\Delta G_{s,i}$ was set such that $\Delta G_{s,i} - \Delta G_s \leq 4.0$ to simulate the non-specific binding affinity of each ZFP. The occupancy probability of each ZFN pair was calculated as the product of the occupancy probabilities of the two ZFPs (3' ZFP and 5' ZFP).

Sequence and statistical analyses

All sequence analyses, PWM score calculations, ZFP occupancy calculations and lesion frequency calculations were performed using customized Perl scripts. All statistical analyses, plots and DNA-binding logos were generated using R (<http://www.r-project.org>).

Analysis of *gata2a*^{um27} mutant embryos

For phenotypic analysis, embryos were obtained from individual incrosses of heterozygous carriers. To analyze vascular development, *gata2a*^{um27} was crossed into the *Tg(kdrl:egfp)^{jal116}* background. Vascular morphology and circulatory function were observed by confocal microscopy, confocal microangiography and video microscopy as described elsewhere (Covassin et al., 2006); Quantum Dots were from Invitrogen. Whole-mount in situ hybridization was performed as previously described using antisense riboprobes against *kdrl*, *hey2*, *efnb2a*, *vegfaa*, *flt4*, *tall1* and *gata1* (Lawson et al., 2001; Hart et al., 2007). Following phenotypic analysis or in situ hybridization, DNA was isolated from selected individual embryos as described elsewhere (Roman et al., 2002). The presence of the 10 bp *um27* deletion was determined by PCR amplification (see Table S2 in the supplementary material for primer sequences), followed by gel electrophoresis or analysis on a QiaXL system (Qiagen). A rescue construct was made by amplifying the *gata2a* coding sequence by PCR using primers containing the Gateway attB1 and attB2 sites (see Table S2 in the supplementary material), followed by BP cloning into pDONR221 (Invitrogen). The resulting plasmid, pME-*gata2a*, was used in an LR reaction with pCSDest (Villefranc et al., 2007) to generate pCS*gata2a*, which was linearized with *NotI* and used as a template to synthesize mRNA. Embryos derived from crosses between *gata2a*^{um27} heterozygous carriers were injected at the 1-cell stage with 200 pg *gata2a* mRNA, or a comparable amount of *mcherry* mRNA, and were scored for the presence or absence of trunk circulation at 48 to 55 hours post-fertilization (hpf). Following phenotypic scoring, embryos were genotyped as above.

RESULTS

Overall rationale and experimental approach

To assess ZFNs constructed via modular assembly, we developed the following approach (see Fig. S1 in the supplementary material). We generated a collection of plasmids encoding zinc-finger modules recognizing 27 different triplet sequences. In parallel, we constructed

a database of sites within zebrafish protein-coding genes (Zv7) that could be targeted by ZFNs assembled from this archive. From this database, we chose target genes and generated the corresponding ZFP cassettes. We characterized the specificity of these ZFP cassettes and determined their function in ZFNs by assessing their ability to induce somatic and germline lesions in vivo. Finally, we created zebrafish bearing a truncation allele in *gata2a* and characterized the phenotypes associated with this mutation.

A zinc-finger archive for modular assembly

We compiled an archive of zinc fingers recognizing 27 different triplet sequences in the three finger positions of the Zif268 backbone that can be rapidly assembled by PCR into three-finger ZFPs (see Table S1 in the supplementary material). This archive comprises modules from previously defined finger archives (Segal et al., 1999; Liu et al., 2002; Carroll et al., 2006) or that have been designed based on previously described recognition principles (Isalan et al., 1998; Segal et al., 1999; Wolfe et al., 1999; Dreier et al., 2001; Dreier et al., 2005). We also generated 14 new modules by bacterial one-hybrid (B1H) selection (Meng et al., 2008). In some cases, distinct recognition helices were utilized at different finger positions for a common triplet for the B1H-generated fingers. In general, the archive was focused around GNN recognition elements (see Table S1 in the supplementary material) because of their reliable functional properties (Ramirez et al., 2008). Nine HNG modules with a preference for either G or T at the neighboring 3' position due to an RSD motif at positions -1, 1 and 2 of the recognition helix were included as they should synergistically recognize a composite sequence (NNGGNN) with GNN modules. Finally, AGA and TGT modules were included to expand the set of triplets that could be specified.

Target gene selection

We identified zebrafish genes that could be targeted using ZFNs constructed from the modular archive. In this analysis, each ZFN monomer consists of a three-finger cassette recognizing a 9 bp sequence. Therefore, we searched all coding exons (including the 10 bp flanking each exon) for potential ZFN sites that constitute adjacent, appropriately oriented 9 bp recognition sites compatible with our archive and separated by a 5 or 6 bp gap. Nearly 75% of all annotated zebrafish protein coding genes contained accessible ZFN sites. To generally assess the archive, targets sites were chosen in more than thirty genes (see Table S4 in the supplementary material) by four different laboratories, in which the corresponding ZFNs were constructed. The target genes are involved in multiple biological processes and are located randomly on 19 different chromosomes throughout the zebrafish genome. We restricted target sites to those that should yield a loss-of-function allele (i.e. those in the 5' half of the coding sequence or in a known functional domain) in the event of a ZFN-induced lesion. We gave preference to targets in which a restriction enzyme site overlapped the sequence gap between the ZFP binding sites to allow straightforward lesion identification in founder fish.

Binding specificity of modularly assembled ZFPs

To assess the recognition properties of three-finger ZFPs, we determined their DNA-binding specificities by interrogating each ZFP against a 28 bp randomized library using the B1H system (Noyes et al., 2008). The majority of selections (56 out of 76; see Table S5 in the supplementary material) were successful (i.e. yielded a significant increase in colonies over background) under standard stringency (5 mM or 10 mM 3-AT). Selections were successful for

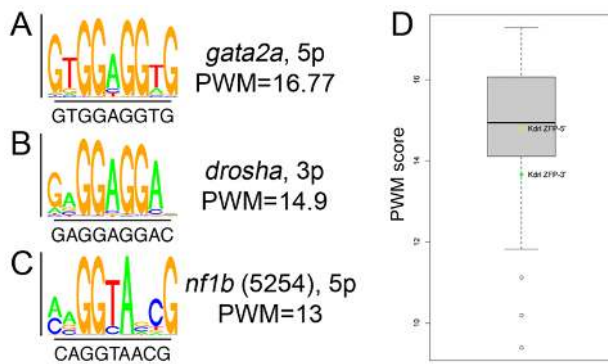


Fig. 1. DNA-binding specificity of modular zinc-finger proteins (ZFPs). (A-C) Sequence logos and position weight matrix (PWM) scores determined for 5' (5p) or 3' (3p) target sites for zebrafish (A) *gata2a*, (B) *drosha* and (C) *nf1b*. (D) Box plot depicting meta-analysis of PWM scores for each target site across all ZFPs for which bacterial one-hybrid (B1H) selections were successful. PWM scores for the previously described *kdrl* ZFPs are shown for reference. Whiskers indicate the largest (smallest) datum still within 1.5 interquartile range (IQR) of the upper (lower) quartile, where outliers are indicated as open circles.

19 of the remaining 20 ZFPs at lower stringency (≤ 2.5 mM 3-AT). For each ZFP a recognition motif (sequence logo) (Schneider and Stephens, 1990) and position weight matrix (PWM) were generated based on binding sites recovered following selection. In general, determined recognition motifs resembled the desired target motif, although there was variability in the overall quality of the PWM score of the target site across all ZFPs (see Fig. S2 in the supplementary material). For example, the 5p ZFP targeting *gata2a* properly specified all 9 bp and displayed among the highest PWM scores, suggesting excellent specificity for its target (Fig. 1A). Most ZFPs exhibited more modest specificity for their target, similar to the *drosha* (*rnasen* – Zebrafish Information Network) 3p ZFP (Fig. 1B), whereas several displayed weaker preference, such as the *nf1b* 5p (Fig. 1C). Overall, the distribution of PWM scores for the modular ZFP cassettes was similar to the engineered ZFPs successfully used to target the zebrafish *kdrl* locus (Fig. 1D) (Meng et al., 2008; Gupta et al., 2011), indicating that the ZFPs constructed using this archive generally display good recognition properties.

We further analyzed the B1H binding site selection data to assess the quality of individual modules, allowing us to identify those with poor or context-dependent specificity. Overall, most modules displayed a strong preference for their recognition triplet based on individually derived PWM scores (Fig. 2A), although modules at fingers 1 and 3 display lower median scores than those at finger 2.

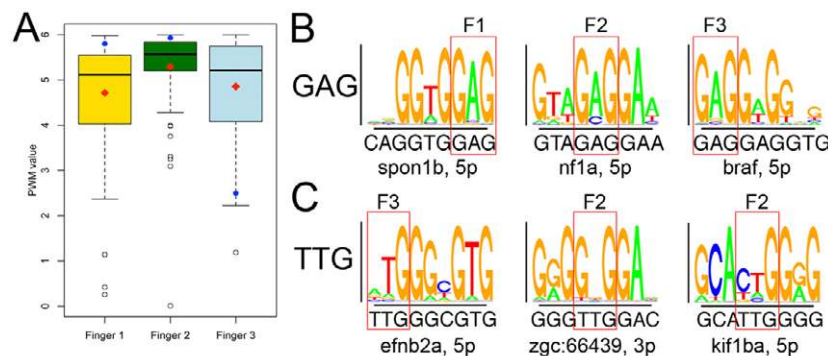


Fig. 2. Assessing the quality and behavior of individual ZFP modules. (A) Box plot depicting meta-analysis of PWM values for individual ZFP modules in each of the three positions in the Zif268 backbone. 5p (red diamonds) and 3p (blue circles) scores for fingers in the *kdrl* ZFPs are shown for reference. Whiskers indicate the largest (smallest) datum still within 1.5 interquartile range (IQR) of the upper (lower) quartile, where outliers are indicated as open circles. (B) Sequence logos for ZFP monomers recognizing GAG in each finger position. (C) Sequence logos for ZFPs containing TTG modules at the indicated finger position.

This is likely to be due to fraying effects at the edges of the protein-DNA complex that reduce the specificity of the determinants at these positions (Choo, 1998). Comparison of PWM values across all individual modules revealed that those targeting purine-rich triplets were among the most robust, regardless of their finger position. For example, the GAG module consistently displayed excellent specificity in multiple finger positions in most ZFPs (Fig. 2B), as did modules targeting GAT, GGG and GGT (see Tables S5 and S6 in the supplementary material). We also identified four modules within the archive that displayed poor specificity at a specific position or in a context-dependent manner. For example, the TTG module displays context-dependent alterations in specificity. We originally utilized the TTG module at finger 3 in ZFPs targeting the zebrafish *kdrl* locus, where it displays moderate preference (NtG) for its triplet (Meng et al., 2008; Gupta et al., 2011), similar to the *efnb2a* 5p ZFP in this study (Fig. 2C). However, when it is placed at the finger 2 position adjacent to a C-terminal module bearing an RSD motif at positions -1, 1 and 2, the specificity shifts to a strong preference for a GNG binding site [Fig. 2C, *zgc:66439* (*clec14a* – Zebrafish Information Network), 3p ZFP]. This dramatic alteration in specificity is not observed when the neighboring C-terminal module lacks an RSD motif (Fig. 2C, *kif1b*, 5p ZFP). Overall, the majority of analyzed modules have favorable recognition properties in multiple contexts, although several might specify incorrect triplets in particular contexts.

In vivo activity of modular ZFNs

We evaluated the ability of ZFNs constructed from our modular archive to induce lesions at a target site within the zebrafish genome. Twenty-nine pairs of ZFNs targeting 28 genes (see Table S7 in the supplementary material) were constructed from 76 ZFPs described above (see Table S5 in the supplementary material) by fusing them to engineered heterodimeric FokI nuclease domains (Miller et al., 2007; Szczepek et al., 2007). We assessed lesion frequency by deep sequencing the target region in normal embryos at 24 hpf following injection of mRNAs encoding ZFNs at an optimal dose (see Table S3 in the supplementary material). Eight of the 29 ZFNs displayed lesions frequencies of 1% or more in at least one set of injections (see Table S7 in the supplementary material). The previously characterized *kdrl* ZFNs were injected as a positive control and generated an in vivo lesion frequency of ~7%, similar to our previous observations using Illumina sequencing (Gupta et al., 2011).

Germline transmission of ZFN-induced mutant alleles

Deep sequencing is likely to underestimate the frequency of ZFN-induced lesions because of short read lengths, and assessment of somatic lesions in embryos may not reflect the frequency of

Table 1. Somatic lesion frequency and germline transmission rate of ZFN-induced mutagenic alleles

ZFN ID	Gene	Somatic lesion frequency (%)	Frequency of founders (%)
ZFNv1_68779	<i>chd5</i>	<0.5	0
ZFNv1_53751	<i>kif1b</i>	<0.5	0
ZFNv1_27773	<i>braf</i>	>1	0
ZFNv1_5254	<i>nf1b</i>	ND	1.3
ZFNv1_5253	<i>nf1b</i>	~1	3
ZFNv1_87654	<i>gata2a</i>	~1	4
ZFNv1_19302	<i>nf1a</i>	<0.5	5.5
ZFNv1_94061	<i>ago2</i>	<0.5	6.5
ZFNv1_83611	<i>cxcr4a</i>	~1	10
ZFNv1_37185	<i>numb-like</i>	~1	13
ZFNv1_53765	<i>kif1b</i>	1-2	18
ZFNv1_25601	<i>gata3</i>	>1	25

ND, not determined.

germline transmission. Therefore, we determined the ability of 12 ZFNs targeting 11 genes to generate founder fish bearing germline lesions at the desired target site (Table 1). For most ZFNs that induced appreciable somatic lesion frequencies, we identified founder fish that transmitted mutant alleles through their germline and did so at a higher frequency than the somatic lesion rate (Table 1). Overall, there was a moderate correlation between the somatic lesion frequency and the founder rate ($R^2=0.71$). Notably, for two different ZFNs [*ago2* (*efj2c2* – Zebrafish Information Network) and *nf1a*], we were able to obtain founders despite lesion rates below 1% in somatic cells. Conversely, embryos injected with ZFNs targeting *braf* failed to yield founders even with a somatic lesion frequency greater than 1% (Table 1, see Table S7 in the supplementary material).

Correlations between ZFP specificity and ZFN activity

We next investigated potential correlations between the recognition properties of the ZFPs and the activity of the ZFNs in vivo. This analysis could provide a basis for estimating the likelihood of success of ZFNs generated from this archive. Surprisingly, we did not observe a significant correlation between DNA-binding specificity of the ZFPs and the activity of the respective ZFNs in vivo (Fig. 3A). Specificity alone may not be predictive of activity because genomic sequence is not random and thus activity might depend on the number of favorable binding sites for each ZFP within the genome. Therefore, we estimated the fractional occupancy of each ZFP monomer at its target site based on its DNA-binding specificity and the number of alternative high-affinity binding sites within the zebrafish genome (see Table S7 in the supplementary material). Again, there was no significant difference between the predicted target site occupancy for the active and inactive groups (Fig. 3B).

Although DNA-binding specificity did not appear to definitively predict in vivo ZFN activity, we did observe a strong correlation ($P=0.001$) between the number of GNN subsites within the recognition sequences and active ZFNs (Fig. 3C), consistent with previous observations (Ramirez et al., 2008). One additional prominent element of our module archive is the inclusion of NNG fingers. There are, on average, more NNG fingers in active ZFNs than inactive ZFNs (Fig. 3D), but the difference is not significant ($P=0.36$). Combined, a significant difference ($P=0.019$) exists between the average total number of GNN and NNG fingers in the

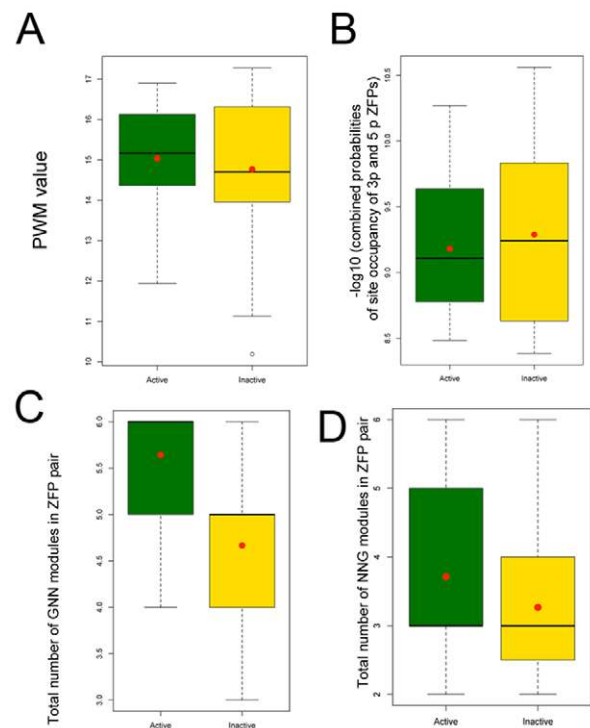


Fig. 3. Correlating ZFP characteristics with in vivo zinc-finger nuclease (ZFN) activity. Box plots depicting (A) PWM score, (B) estimated target site occupancy, (C) number of GNN modules in a ZFN pair and (D) number of NNG modules in a ZFN pair in both active and inactive ZFNs, where red dots indicate the mean of each population. Active ZFNs are those in which somatic lesion frequencies were greater than 0.5% or that successfully generated founders for mutant alleles. Whiskers indicate the largest (smallest) datum still within 1.5 interquartile range (IQR) of the upper (lower) quartile, where outliers are indicated as open circles.

active versus inactive groups of ZFNs, but the majority of the predictive value of this combination is derived from the GNN fingers.

A zebrafish database for modular ZFNs

To facilitate ZFN construction by the zebrafish community, we designed a web-accessible database (pgfe.umassmed.edu/ZFNv1/) that can be used in a standard workflow (see Fig. S3 in the supplementary material), allowing researchers to easily construct ZFNs, generate founder zebrafish bearing targeted lesions, and identify associated phenotypes within 7-12 months. A user can identify ZFN target sites in a gene of interest by entering or uploading an ENSEMBL gene number or RefSeq ID. The resulting output provides a count of ZFN target sites, a gene description and abbreviation, and a link to the locus on the UCSC genome browser (Fig. 4A). The user can view sites within one particular gene or all targets using a sortable list (Fig. 4B) that aids identification of optimal sites based on position (exon rank), presence of a restriction enzyme site in the spacer sequence, or a revised efficacy score. The efficacy score has a maximum of 12 points, where one point is assigned for each guanine at the edge of each recognition triplet (i.e. GNN and NNG fingers), as active ZFNs are enriched for these contacts (see above). The database also notes single-nucleotide polymorphisms in the target site in the reference Tuebingen (Tu) and AB wild-type zebrafish strains that have been

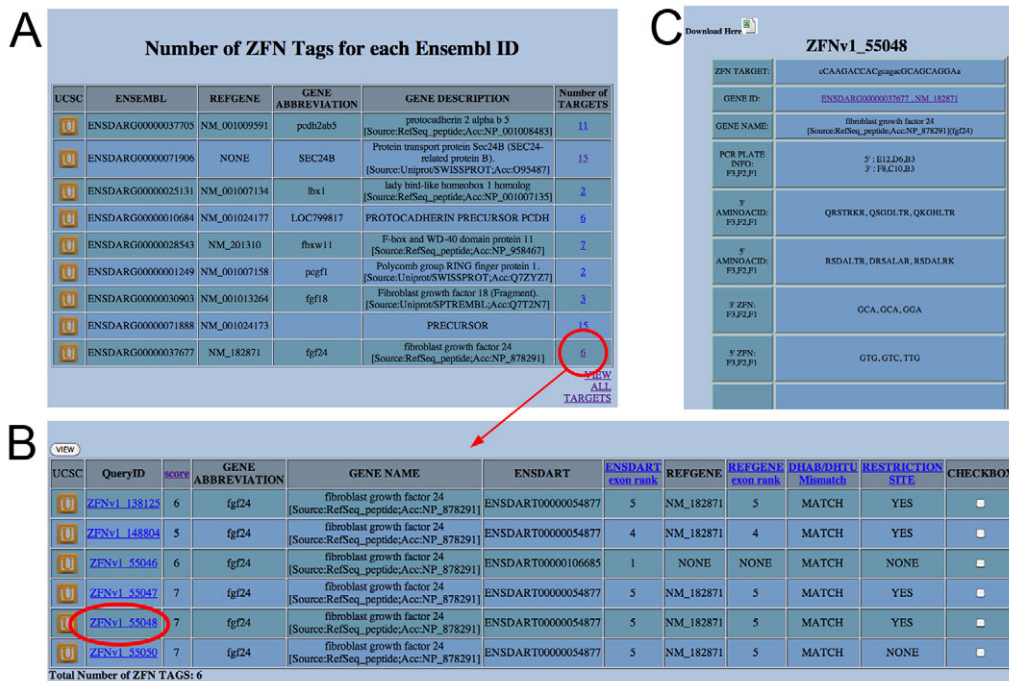


Fig. 4. Screenshots from a modular ZFP/ZFN database.

(A) Example output from a search for ZFN target sites. Circle indicates the link to the page shown in B. (B) List of targets in the *fgf24* gene with associated information. Circled number indicates the link to individual target site and ZFN information shown in C. (C) Example output page with individual target site information.

sequenced by the Sanger Center (D. Stemple, personal communication), although we recommend sequencing the target site in wild-type strains other than the hybrid Tu/AB (SAT) reference line.

Once a target site is chosen, the user can click on the ZFN entry (QueryID; see Fig. 4B) for details about each construct (Fig. 4C), such as the target site sequence, the amino acid sequence of the recognition helices for the assembled ZFNs, and clone IDs for each plasmid in the archive (available at addgene.org) to facilitate overlapping PCR of the appropriate modules (see Table S1 in the supplementary material). Alternatively, the DNA sequence of the ZFP cassette is included for direct synthesis. In either case, the inclusion of the amino acid sequence for each entry (Fig. 4C) aids in sequence validation of ZFN-containing plasmids. ZFNs containing modules with poor or context-dependent specificity (see above) are flagged with an asterisk. For lesion detection, we also include restriction sites, when present, within the spacer region separating the ZFP binding sites, along with sequences of flanking primers for PCR amplification. The primer and restriction enzyme information also facilitates easy genotypic analysis of founders and mutant embryos. Data from these pages can be exported into Excel (Microsoft) for easy reference. A query page (<http://pgfe.umassmed.edu/ZFPmodularsearch.html>) is available to identify ZFN sites in sequences not contained within our database.

A novel role for *gata2a* in vascular development

To demonstrate the application of our archive to interrogate gene function and to provide an example of the workflow shown in Fig. S3 in the supplementary material, we investigated defects associated with a ZFN-induced mutation in the *gata2a* gene. *Gata2* is a zinc-finger transcription factor that is essential for definitive hematopoiesis and maintenance of stem cell progenitors in both embryos and adults (Tsai et al., 1994; Rodrigues et al., 2005). *Gata2* has also been implicated in angiogenesis and is associated with early onset coronary artery disease in humans (Connelly et al., 2006; Mammoto et al., 2009), but its role in vascular development

is unknown. As described above, we constructed ZFNs targeting the fourth exon of *gata2a*. The target sequence is upstream of two zinc fingers required for function of the mammalian *Gata2* ortholog (Minegishi et al., 2003) and, therefore, truncation in this region would be expected to generate a null allele (Fig. 5A). Following injection of ZFNs targeting this site, we identified a founder fish bearing a 10 bp deletion allele, referred to as *gata2a^{um27}*, which causes a frameshift to a premature stop codon (Fig. 5A,B).

To identify phenotypes caused by *gata2a* deficiency we observed embryos derived from an incross of *gata2a^{um27}* heterozygous carriers derived from the original founder. At 24 and 48 hpf, all embryos appeared morphologically normal, including normal development of neural tube, notochord and somites (see Fig. S4 in the supplementary material; data not shown). However, ~25% of embryos failed to display trunk blood vessel circulation by 48 hpf (see Movie 1 in the supplementary material, Table 2) whereas the remaining siblings were normal (see Movie 2 in the supplementary material). Genotypic analysis demonstrated that embryos with defective trunk circulation were homozygous for the *um27* deletion, whereas normal embryos were heterozygous or homozygous for the wild-type allele (Fig. 5C, Table 2). Importantly, injection of mRNA encoding wild-type *Gata2a* could rescue trunk circulation in *gata2a^{um27}* mutant embryos (16 out of 26 mutant embryos in two separate experiments) (see Movie 3 in the supplementary material, Fig. 5D,E), whereas injection with *mcherry* mRNA did not (0 out of 13 mutant embryos in two separate experiments) (see Movie 4 in the supplementary material, Fig. 5D,E). Closer inspection of *gata2a^{um27}* mutant embryos revealed the occurrence of pulsating blood cells trapped in the trunk blood vessels and abnormal circulatory connections, or shunts, between the dorsal aorta and posterior cardinal vein (see Movies 1 and 5 in the supplementary material). Despite these defects, mutant embryos displayed a beating heart, the presence of blood cells and circulation through cranial blood vessels (see Movie 6 in the supplementary material), similar to sibling embryos

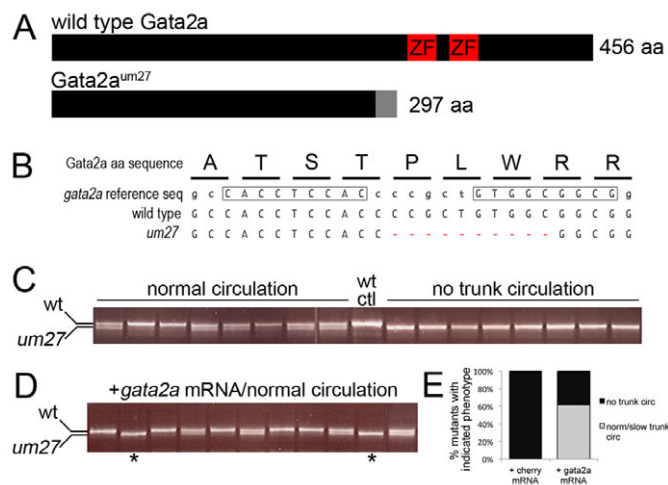


Fig. 5. The ZFN-induced *um27* lesion is a truncation allele of *gata2a*. (A) Location of frameshift caused by the *um27* deletion in the zebrafish *gata2a* gene. ZF, zinc-finger domain. (B) *gata2a* coding and amino acid sequence in the region of the *um27* deletion. ZFN recognition sequences are boxed. (C) Genotype of embryos derived from an incross of *gata2a^{um27}* heterozygous carriers. (D) Genotype of embryos with normal circulation derived from a *gata2a^{um27}* incross injected with 200 pg *gata2a* mRNA. Asterisks denote mutant embryos with normal trunk circulation. (E) Percentage of *gata2a^{um27}* mutant embryos displaying the indicated circulation phenotypes.

with trunk circulation (see Movie 7 in the supplementary material). Together, these observations demonstrate that *gata2a* deficiency leads to a specific defect in circulatory function in embryonic zebrafish.

To further characterize the phenotype caused by loss of *gata2a*, we performed confocal microangiography on wild-type and *gata2a^{um27}* mutant siblings bearing the *Tg(kdrl:egfp)^{la116}* transgene at 50 hpf. Consistent with normal head circulation in *gata2a^{um27}* mutant embryos, we did not note any overt defects in vascular morphology of the cranial vessels when compared with wild-type siblings (Fig. 6A,B). We observed normal formation of arteries and veins within the head, which are fully perfused following angiography (Fig. 6A,B). By contrast, the lateral dorsal aortae in *gata2a^{um27}* mutant embryos are more dilated than in wild-type siblings, although these vessels are perfused, suggesting that lumenization is not affected (Fig. 6C,D). Within the trunk, we

observed specific defects in vascular morphology. Whereas intersegmental vessels appeared largely normal in *gata2a^{um27}* mutant embryos, the dorsal aorta was discontinuous (Fig. 6E). Accordingly, *gata2a^{um27}* mutant trunk vessels were poorly perfused following angiography when compared with their wild-type siblings (Fig. 6E,F). These analyses suggest a specific defect in the morphogenesis of the dorsal aorta in *gata2a^{um27}* mutant embryos, which would be the likely cause of the arteriovenous shunts noted above.

We have previously observed aorta morphogenesis defects and arteriovenous shunts in zebrafish embryos deficient for Notch or Vegf signaling (Lawson et al., 2001; Covassin et al., 2006). In these cases, these defects were associated with loss of arterial endothelial cell identity. However, *gata2a^{um27}* mutant embryos did not display loss of artery markers, such as *efnb2a*, and exhibited normal levels of *vegfaa* expression (data not shown). Similarly, the *hey2* gene, which is also required for proper dorsal aorta morphogenesis (Zhong et al., 2000), was expressed at normal levels in *gata2a^{um27}* mutants (Fig. 7A,B). We also did not observe any significant changes in the expression of blood markers such as *gata1* and *tall* (Fig. 7C,D; data not shown). However, *gata2a^{um27}* mutant embryos displayed a subtle, but consistent, downregulation of *kdrl*, the functional ortholog of *Vegfr2*, when compared with wild-type siblings (Fig. 7E,F). Taken together, our results demonstrate that *gata2a* function is essential for the proper morphogenesis of the dorsal aorta, but may be dispensable for arterial endothelial differentiation.

DISCUSSION

With the increasing characterization of the genomes of model and non-model organisms, there is a broad need for reverse genetic approaches to assess gene function in multiple species. ZFNs allow the direct introduction of targeted germline lesions in vivo, without the need for species-matched embryonic stem cell lines, thus providing a general means to determine gene function. However, the primary limitation hindering the widespread employment of this technology has been the lack of an affordable, simple method for constructing ZFNs that are active in vivo. Modularly assembled ZFNs provide a rapid and effective method for introducing targeted lesions into the genomes of animals. The specificity analysis of our assembled ZFPs indicates that the majority have similar recognition potential to the selected ZFPs incorporated into our *kdrl* ZFNs (Meng et al., 2008; Gupta et al., 2011). Consequently, a sizable fraction (29%) of ZFNs constructed from these modules displayed significant activity ($\geq 1\%$) for the generation of somatic lesions,

Table 2. Linkage of *gata2a^{um27}* to circulatory defects in mutant embryos

Parents	Circulation phenotype	No. embryos	Proportion	<i>gata2a^{um27}</i> genotype		
				+/+	+/-	-/-
A2×A5	Normal	116	0.77	1	10	0
	Head but no trunk	34	0.23	0	0	11
	Normal	59	0.81	3	7	2
	Head but no trunk	14	0.19	1	0	11
A3×A6	Normal	42	0.79	4	7	1
	Head but no trunk	11	0.21	0	0	12
	Normal	67	0.79	3	8	0
	Head but no trunk	18	0.21	0	0	11
Total	Normal	284	0.79	11	32	3
	Head but no trunk	77	0.21	1	0	45

Results are shown for two separate clutches (different days) from each set of parents.

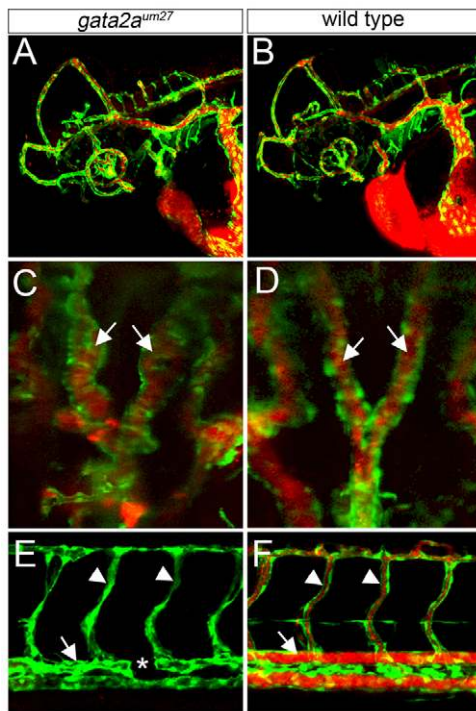


Fig. 6. *gata2a^{um27}* mutant embryos display defects in aorta morphogenesis. Confocal microangiography using Quantum Dots in *Tg(kdrl:egfp)^{la116}* transgenic zebrafish embryos. Endothelial cells are green and vessel perfusion is red. (A,B) Cranial blood vessels in *gata2a^{um27}* mutant (A) and wild-type (B) embryos. Lateral views, anterior to the left, dorsal is up. (C,D) Lateral dorsal aortae (arrows) in *gata2a^{um27}* mutant (C) and wild-type (D) embryos. Dorsal views, anterior is up. (E,F) Trunk blood vessels in *gata2a^{um27}* mutant (E) and wild-type (F) embryos. Lateral views, anterior to the left, dorsal is up. Segmental vessels are indicated by arrowheads and the dorsal aorta by an arrow; the asterisk indicates a region of aorta that failed to form.

establishing the utility of this archive for targeted genome manipulation in complex genomes. This success rate compares favorably with that of other single-finger modular assembly archives for constructing ZFNs (Ramirez et al., 2008; Kim et al., 2009; Kim et al., 2011). Moreover, characterization of DNA-binding specificity provides additional information about the performance of individual modules within this archive and will facilitate its subsequent improvement.

Surprisingly, we did not find a strong correlation between ZFP specificity and in vivo ZFN activity. Previous studies have shown that improvements in DNA-binding specificity can lead to improved ZFN activity (Cornu et al., 2008) and precision (Gupta et al., 2011). It is likely that other factors, such as ZFP affinity, also significantly influence ZFN activity. Whereas the recognition motifs generated using the B1H system provide an estimate of ZFP specificity, our analysis utilized varying stringencies to achieve enrichment of binding sites, possibly bypassing variations that could reflect differences in affinity. That said, seven of eight ZFNs that contained at least one ZFP that displayed low activity in the B1H system (i.e. binding site selections required a stringency below 5 mM 3-AT) displayed low somatic lesion rates (<0.5%). This correlation is consistent with studies utilizing the bacterial two-hybrid system, in which highly active ZFPs often perform well in ZFNs (Ramirez et al., 2008; Lam et al., 2011). Thus, a more

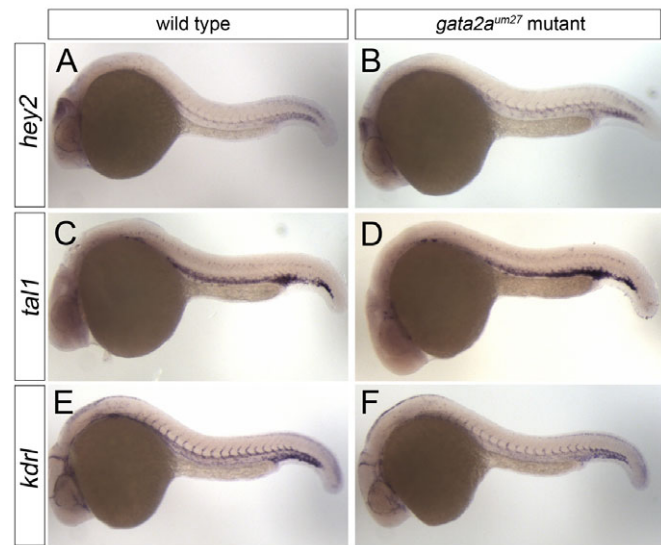


Fig. 7. *gata2a^{um27}* mutant embryos show reduced *kdrl* expression but normal artery and blood differentiation. (A-F) Whole-mount in situ hybridization at 24 hpf using antisense riboprobes against *hey2* (A,B), *tal1* (C,D) and *kdrl* (E,F) in genotypically wild-type zebrafish embryos (A,C,E) and those homozygous for the *um27* deletion (B,D,F). Lateral views, anterior to the left, dorsal is up.

refined analysis of ZFP affinity might provide additional insights into the behavior of our archive. Additionally, ZFN activity could be affected by properties of the endogenous target sequence (e.g. chromatin architecture or DNA methylation status) that could interfere with nuclease recognition or function.

We did observe significantly higher rates of ZFN activity as a function of the number of GNN fingers, consistent with previous observations (Ramirez et al., 2008), and a slight increase in the average number of NNG fingers in the active versus inactive ZFN population. In both sets of modules, arginine-guanine interactions, which can contribute dramatically to binding affinity and specificity (Elrod-Erickson and Pabo, 1999), appear to be critical linchpins for the formation of active ZFN complexes. Consequently, this information is incorporated into a simple scoring function for evaluating our ZFNs that notes the number of these contacts. Consistent with the importance of arginine-guanine interactions, their preservation is a defining feature of active off-target sites for *kdrl* ZFNs in zebrafish (Gupta et al., 2011). Clearly, the presence of a large number of arginine-guanine interactions is not required for the construction of active ZFNs (Hockemeyer et al., 2009), but these particular examples employed larger numbers of fingers, which might compensate for the absence of these favorable interactions.

A limitation of the modular assembly approach highlighted by our analysis is position or context-dependent specificity of individual modules. In other studies, this problem has been partially mitigated by the creation of extensive archives of selected two-finger modules that allow the assembly of multiple ZFPs against a single target site. For example, Sangamo BioSciences employs a proprietary archive of two-finger modules with origins in context-dependent selection for ZFN assembly (Isalan and Choo, 2001). Recently, the Zinc Finger Consortium reported the selection and construction of an archive of two-finger modules, in which three-finger proteins are created from the assembly of overlapping

two-finger modules (CoDA) (Sander et al., 2011). By reducing context-dependent effects both methods are anticipated to improve the success rate for the assembly of functional ZFPs over those assembled from single-finger archives. For the Zinc Finger Consortium-based CoDA ZFPs this has been directly demonstrated on a large scale, in which ~50% of ZFNs were active in vivo (Sander et al., 2011), although only ZFPs that performed well in a bacterial two-hybrid assay were used for in vivo testing in this case. Taking this prefiltering into account, the in vivo efficacy of ZFNs from our archive compares favorably to the CoDA approach. Furthermore, our archive is capable of targeting genomic sequences that are not accessible using CoDA because of differences in the recognition sequences of modules between these archives. Finally, the detailed characterization of ZFPs constructed from our archive allowed us to identify context-dependent or otherwise poorly performing modules, facilitating future improvements to our collection. Taken together, we believe that our archive and the associated computational resources will provide a viable and accessible platform for the generation of gene knockouts in zebrafish, as well as in other organisms.

A number of studies have now described ZFNs that are capable of producing somatic lesions in zebrafish embryos. However, only a handful of new mutants have been generated and there is a paucity of data concerning germline transmission of ZFN-induced alleles. To date, the modular archive described here has been used to generate eight new zebrafish mutants. Significantly, several of these new lines have been used to reveal novel aspects of microRNA biogenesis and vascular development (Siekman et al., 2009; Cifuentes et al., 2010), whereas others, such as *nfla* and *nflb*, will serve as important disease models (Padmanabhan et al., 2009; Lee et al., 2010). Equally importantly, our archive has been used to independently generate these mutant lines in four separate laboratories, suggesting that this approach can be easily implemented within the research community at large to create zebrafish lines bearing mutations in genes of interest.

Among the new mutants that we have generated is a truncation in the *gata2a* gene. We observed that *gata2a*-deficient embryos displayed defects in the formation of the dorsal aorta, yet the remainder of the vascular system is remarkably unaffected. This phenotype is reminiscent of embryos lacking Vegf or Notch signaling components and is often attributed to the loss of properly specified artery and vein endothelial identity (Lawson et al., 2001; Covassin et al., 2006). However, we did not observe obvious loss of artery marker gene expression. These results are somewhat surprising and suggest that vascular morphogenesis and endothelial differentiation might be uncoupled at some point downstream of Vegf. Alternatively, the phenotype of *gata2a^{um27}* mutant embryos might reflect a defect in endothelial mechanosensation that affects morphogenesis independently of differentiation. Indeed, we do not observe any major changes in vascular morphology until the onset of circulation, although *kdrl* expression is slightly downregulated at 24 hpf, prior to blood flow. In addition, GATA2 has recently been implicated in the induction of *VEGFR2* in human endothelial cells in response to increased extracellular matrix stiffening (Mammoto et al., 2009), suggesting a role for this transcription factor in mediating endothelial mechanosensation. In any case, our observations reveal an unexpected and previously undescribed role for *gata2a* in artery morphogenesis. Importantly, the *gata2a^{um27}* mutant generated using modular ZFNs now provides an important tool to further characterize the observed vascular phenotype and to identify relevant target genes that are required for vascular morphogenesis or function.

Acknowledgements

We thank Craig Ceol, Stefania Nicoli and Fatma Kok for critical reading of the manuscript, John Polli for excellent fish care, and Derek Stemple for providing the Tu and AB genome sequences.

Funding

This work was funded by grants from the NIH National Heart, Lung, and Blood Institute to S.A.W. and N.D.L. (R01 HL093766) and to N.D.L. (R01 HL079266) and from the Department of Defense to A.T.L. and J.A.E. (DOD, NF050175), and from the National Institute of General Medical Sciences to A.J.G. (R01 GM081602). Deposited in PMC for release after 12 months.

Competing interests statement

The authors declare no competing financial interests.

Supplementary material

Supplementary material for this article is available at <http://dev.biologists.org/lookup/suppl/doi:10.1242/dev.066779/-DC1>

References

- Bailey, T. L. and Elkan, C. (1994). Fitting a mixture model by expectation maximization to discover motifs in biopolymers. *Proc. Int. Conf. Intell. Syst. Mol. Biol.* **2**, 28-36.
- Beis, D. and Stainier, D. Y. (2006). In vivo cell biology: following the zebrafish trend. *Trends Cell Biol.* **16**, 105-112.
- Beumer, K. J., Trautman, J. K., Bozas, A., Liu, J. L., Rutter, J., Gall, J. G. and Carroll, D. (2008). Efficient gene targeting in *Drosophila* by direct embryo injection with zinc-finger nucleases. *Proc. Natl. Acad. Sci. USA* **105**, 19821-19826.
- Bibikova, M., Golic, M., Golic, K. G. and Carroll, D. (2002). Targeted chromosomal cleavage and mutagenesis in *Drosophila* using zinc-finger nucleases. *Genetics* **161**, 1169-1175.
- Carroll, D., Morton, J. J., Beumer, K. J. and Segal, D. J. (2006). Design, construction and in vitro testing of zinc finger nucleases. *Nat. Protoc.* **1**, 1329-1341.
- Choi, J., Dong, L., Ahn, J., Dao, D., Hammerschmidt, M. and Chen, J. N. (2007). FoxH1 negatively modulates flk1 gene expression and vascular formation in zebrafish. *Dev. Biol.* **304**, 735-744.
- Choo, Y. (1998). End effects in DNA recognition by zinc finger arrays. *Nucleic Acids Res.* **26**, 554-557.
- Cifuentes, D., Xue, H., Taylor, D. W., Patnode, H., Mishima, Y., Cheloufi, S., Ma, E., Mane, S., Hannon, G. J., Lawson, N. D. et al. (2010). A novel miRNA processing pathway independent of Dicer requires Argonaute2 catalytic activity. *Science* **328**, 1694-1698.
- Connelly, J. J., Wang, T., Cox, J. E., Haynes, C., Wang, L., Shah, S. H., Crosslin, D. R., Hale, A. B., Nelson, S., Crossman, D. C. et al. (2006). GATA2 is associated with familial early-onset coronary artery disease. *PLoS Genet.* **2**, e139.
- Cornu, T. I., Thibodeau-Beganny, S., Guhl, E., Alwin, S., Eichinger, M., Jung, J. K. and Cathomen, T. (2008). DNA-binding specificity is a major determinant of the activity and toxicity of zinc-finger nucleases. *Mol. Ther.* **16**, 352-358.
- Covassin, L. D., Villefranc, J. A., Kacergis, M. C., Weinstein, B. M. and Lawson, N. D. (2006). Distinct genetic interactions between multiple Vegf receptors are required for development of different blood vessel types in zebrafish. *Proc. Natl. Acad. Sci. USA* **103**, 6554-6559.
- Cui, X., Ji, D., Fisher, D. A., Wu, Y., Briner, D. M. and Weinstein, E. J. (2010). Targeted integration in rat and mouse embryos with zinc-finger nucleases. *Nat. Biotechnol.* **29**, 64-67.
- Desjarlais, J. R. and Berg, J. M. (1993). Use of a zinc-finger consensus sequence framework and specificity rules to design specific DNA binding proteins. *Proc. Natl. Acad. Sci. USA* **90**, 2256-2260.
- 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.
- Dreier, B., Beerli, R. R., Segal, D. J., Flippin, J. D. and Barbas, C. F., 3rd (2001). Development of zinc finger domains for recognition of the 5'-ANN-3' family of DNA sequences and their use in the construction of artificial transcription factor. *J. Biol. Chem.* **276**, 29466-29478.
- Dreier, B., Fuller, R. P., Segal, D. J., Lund, C. V., Blancfort, P., Huber, A., Koksche, B. and Barbas, C. F., 3rd (2005). Development of zinc finger domains for recognition of the 5'-CNN-3' family DNA sequences and their use in the construction of artificial transcription factors. *J. Biol. Chem.* **280**, 35588-35597.
- Eisen, J. S. and Smith, J. C. (2008). Controlling morpholino experiments: don't stop making antisense. *Development* **135**, 1735-1743.
- Elrod-Erickson, M. and Pabo, C. O. (1999). Binding studies with mutants of Zif268. Contribution of individual side chains to binding affinity and specificity in the Zif268 zinc finger-DNA complex. *J. Biol. Chem.* **274**, 19281-19285.

- Foley, J. E., Yeh, J. R., Maeder, M. L., Reyon, D., Sander, J. D., Peterson, R. T. and Joung, J. K. (2009). Rapid mutation of endogenous zebrafish genes using zinc finger nucleases made by Oligomerized Pool ENgineering (OPEN). *PLoS ONE* **4**, e4348.
- Geurts, A. M., Cost, G. J., Freyvert, Y., Zeitler, B., Miller, J. C., Choi, V. M., Jenkins, S. S., Wood, A., Cui, X., Meng, X. et al. (2009). Knockout rats via embryo microinjection of zinc-finger nucleases. *Science* **325**, 433.
- Greisman, H. A. and Pabo, C. O. (1997). A general strategy for selecting high-affinity zinc finger proteins for diverse DNA target sites. *Science* **275**, 657-661.
- Gupta, A., Meng, X., Zhu, L. J., Lawson, N. D. and Wolfe, S. A. (2011). Zinc finger protein-dependent and -independent contributions to the in vivo off-target activity of zinc finger nucleases. *Nucleic Acids Res.* **39**, 381-392.
- Hart, D. O., Raha, T., Lawson, N. D. and Green, M. R. (2007). Initiation of zebrafish haematopoiesis by the TATA-box-binding protein-related factor Trf3'. *Nature* **450**, 1082-1085.
- Hertz, G. Z. and Stormo, G. D. (1999). Identifying DNA and protein patterns with statistically significant alignments of multiple sequences. *Bioinformatics* **15**, 563-577.
- Hockemeyer, D., Soldner, F., Beard, C., Gao, Q., Mitalipova, M., DeKaveler, R. C., Katibah, G. E., Amora, R., Boydston, E. A., Zeitler, B. et al. (2009). Efficient targeting of expressed and silent genes in human ESCs and iPSCs using zinc-finger nucleases. *Nat. Biotechnol.* **27**, 851-857.
- Isalan, M. and Choo, Y. (2001). Rapid, high-throughput engineering of sequence-specific zinc finger DNA-binding proteins. *Methods Enzymol.* **340**, 593-609.
- Isalan, M., Klug, A. and Choo, Y. (1998). Comprehensive DNA recognition through concerted interactions from adjacent zinc fingers. *Biochemistry* **37**, 12026-12033.
- Kim, H. J., Lee, H. J., Kim, H., Cho, S. W. and Kim, J. S. (2009). Targeted genome editing in human cells with zinc finger nucleases constructed via modular assembly. *Genome Res.* **19**, 1279-1288.
- Kim, S., Kim, E. J. and Kim, J. S. (2010). Construction of combinatorial libraries that encode zinc finger-based transcription factors. *Methods Mol. Biol.* **649**, 133-147.
- Kim, S., Lee, M. J., Kim, H., Kang, M. and Kim, J.-S. (2011). Preassembled zinc-finger arrays for rapid construction of ZFNs. *Nat. Methods* **8**, 7-7.
- Lam, K. N., van Bakel, H., Cote, A. G., van der Ven, A. and Hughes, T. R. (2011). Sequence specificity is obtained from the majority of modular C2H2 zinc-finger arrays. *Nucleic Acids Res.* **39**, 4680-4690.
- Lawson, N. D., Scheer, N., Pham, V. N., Kim, C. H., Chitnis, A. B., Campos-Ortega, J. A. and Weinstein, B. M. (2001). Notch signaling is required for arterial-venous differentiation during embryonic vascular development. *Development* **128**, 3675-3683.
- Lee, J. S., Padmanabhan, A., Shin, J., Zhu, S., Guo, F., Kanki, J. P., Epstein, J. A. and Look, A. T. (2010). Oligodendrocyte progenitor cell numbers and migration are regulated by the zebrafish orthologs of the NF1 tumor suppressor gene. *Hum. Mol. Genet.* **19**, 4643-4653.
- Liu, Q., Xia, Z., Zhong, X. and Case, C. C. (2002). Validated zinc finger protein designs for all 16 GNN DNA triplet targets. *J. Biol. Chem.* **277**, 3850-3856.
- Maeder, M. L., Thibodeau-Beganny, S., Osiaik, A., Wright, D. A., Anthony, R. M., Eichinger, M., Jiang, T., Foley, J. E., Winfrey, R. J., Townsend, J. A. et al. (2008). Rapid 'open-source' engineering of customized zinc-finger nucleases for highly efficient gene modification. *Mol. Cell* **31**, 294-301.
- Mammoto, A., Connor, K. M., Mammoto, T., Yung, C. W., Huh, D., Aderman, C. M., Mostoslavsky, G., Smith, L. E. and Ingber, D. E. (2009). A mechanosensitive transcriptional mechanism that controls angiogenesis. *Nature* **457**, 1103-1108.
- Mashimo, T., Takizawa, A., Voigt, B., Yoshimi, K., Hiai, H., Kuramoto, T. and Serikawa, T. (2010). Generation of knockout rats with X-linked severe combined immunodeficiency (X-SCID) using zinc-finger nucleases. *PLoS ONE* **5**, e8870.
- Meng, X., Noyes, M. B., Zhu, L. J., Lawson, N. D. and Wolfe, S. A. (2008). Targeted gene inactivation in zebrafish using engineered zinc-finger nucleases. *Nat. Biotechnol.* **26**, 695-701.
- Meyer, M., de Angelis, M. H., Wurst, W. and Kuhn, R. (2010). Gene targeting by homologous recombination in mouse zygotes mediated by zinc-finger nucleases. *Proc. Natl. Acad. Sci. USA* **107**, 15022-15026.
- Miller, J. C., Holmes, M. C., Wang, J., Guschin, D. Y., Lee, Y. L., Rupniewski, I., Beausejour, C. M., Waite, A. J., Wang, N. S., Kim, K. A. et al. (2007). An improved zinc-finger nuclease architecture for highly specific genome editing. *Nat. Biotechnol.* **25**, 778-785.
- Minegishi, N., Suzuki, N., Yokomizo, T., Pan, X., Fujimoto, T., Takahashi, S., Hara, T., Miyajima, A., Nishikawa, S. and Yamamoto, M. (2003). Expression and domain-specific function of GATA-2 during differentiation of the hematopoietic precursor cells in midgestation mouse embryos. *Blood* **102**, 896-905.
- Nasevicius, A. and Ekker, S. C. (2000). Effective targeted gene 'knockdown' in zebrafish. *Nat. Genet.* **26**, 216-220.
- Noyes, M. B., Meng, X., Wakabayashi, A., Sinha, S., Brodsky, M. H. and Wolfe, S. A. (2008). A systematic characterization of factors that regulate *Drosophila* segmentation via a bacterial one-hybrid system. *Nucleic Acids Res.* **36**, 2547-2560.
- Padmanabhan, A., Lee, J. S., Ismat, F. A., Lu, M. M., Lawson, N. D., Kanki, J. P., Look, A. T. and Epstein, J. A. (2009). Cardiac and vascular functions of the zebrafish orthologues of the type I neurofibromatosis gene NF1. *Proc. Natl. Acad. Sci. USA* **106**, 22305-22310.
- Patton, E. E. and Zon, L. I. (2001). The art and design of genetic screens: zebrafish. *Nat. Rev. Genet.* **2**, 956-966.
- Perez, E. E., Wang, J., Miller, J. C., Jouvenot, Y., Kim, K. A., Liu, O., Wang, N., Lee, G., Bartsevich, V. V., Lee, Y. L. et al. (2008). Establishment of HIV-1 resistance in CD4+ T cells by genome editing using zinc-finger nucleases. *Nat. Biotechnol.* **26**, 808-816.
- Ramirez, C. L., Foley, J. E., Wright, D. A., Muller-Lerch, F., Rahman, S. H., Cornu, T. I., Winfrey, R. J., Sander, J. D., Fu, F., Townsend, J. A. et al. (2008). Unexpected failure rates for modular assembly of engineered zinc fingers. *Nat. Methods* **5**, 374-375.
- Robu, M. E., Larson, J. D., Nasevicius, A., Beiraghi, S., Brenner, C., Farber, S. A. and Ekker, S. C. (2007). p53 activation by knockdown technologies. *PLoS Genet.* **3**, e78.
- Rodrigues, N. P., Janzen, V., Forkert, R., Dombkowski, D. M., Boyd, A. S., Orkin, S. H., Enver, T., Vyas, P. and Scadden, D. T. (2005). Haploinsufficiency of GATA-2 perturbs adult hematopoietic stem-cell homeostasis. *Blood* **106**, 477-484.
- Roman, B. L., Pham, V. N., Lawson, N. D., Kulik, M., Childs, S., Lekven, A. C., Garrity, D. M., Moon, R. T., Fishman, M. C., Lechleider, R. J. et al. (2002). Disruption of *acvr1l* increases endothelial cell number in zebrafish cranial vessels. *Development* **129**, 3009-3019.
- Sander, J. D., Zaback, P., Joung, J. K., Voytas, D. F. and Dobbs, D. (2009). An affinity-based scoring scheme for predicting DNA-binding activities of modularly assembled zinc-finger proteins. *Nucleic Acids Res.* **37**, 506-515.
- Sander, J. D., Dahlborg, E. J., Goodwin, M. J., Cade, L., Zhang, F., Cifuentes, D., Curtin, S. J., Blackburn, J. S., Thibodeau-Beganny, S., Qi, Y. et al. (2011). Selection-free zinc-finger-nuclease engineering by context-dependent assembly (CoDA). *Nat. Methods* **8**, 67-69.
- Schneider, T. D. and Stephens, R. M. (1990). Sequence logos: a new way to display consensus sequences. *Nucleic Acids Res.* **18**, 6097-6100.
- Segal, D. J., Dreier, B., Beerli, R. R. and Barbas, C. F., 3rd (1999). Toward controlling gene expression at will: selection and design of zinc finger domains recognizing each of the 5'-GNN-3' DNA target sequences. *Proc. Natl. Acad. Sci. USA* **96**, 2758-2763.
- Siekman, A. F., Standley, C., Fogarty, K. E., Wolfe, S. A. and Lawson, N. D. (2009). Chemokine signaling guides regional patterning of the first embryonic artery. *Genes Dev.* **23**, 2272-2277.
- Szczepek, M., Brondani, V., Buchel, J., Serrano, L., Segal, D. J. and Cathomen, T. (2007). Structure-based redesign of the dimerization interface reduces the toxicity of zinc-finger nucleases. *Nat. Biotechnol.* **25**, 786-793.
- Takasu, Y., Kobayashi, I., Beumer, K., Uchino, K., Sezutsu, H., Sajwan, S., Carroll, D., Tamura, T. and Zurovec, M. (2010). Targeted mutagenesis in the silkworm *Bombyx mori* using zinc finger nuclease mRNA injection. *Insect Biochem. Mol. Biol.* **40**, 759-765.
- Tsai, F. Y., Keller, G., Kuo, F. C., Weiss, M., Chen, J., Rosenblatt, M., Alt, F. W. and Orkin, S. H. (1994). An early haematopoietic defect in mice lacking the transcription factor GATA-2. *Nature* **371**, 221-226.
- Urnov, F. D., Rebar, E. J., Holmes, M. C., Zhang, H. S. and Gregory, P. D. (2010). Genome editing with engineered zinc finger nucleases. *Nat. Rev. Genet.* **11**, 636-646.
- Villefranc, J. A., Amigo, J. and Lawson, N. D. (2007). Gateway compatible vectors for analysis of gene function in the zebrafish. *Dev. Dyn.* **236**, 3077-3087.
- Westerfield, M. (1993). *The Zebrafish Book*. Eugene, OR: University of Oregon Press.
- Wolfe, S. A., Greisman, H. A., Ramm, E. I. and Pabo, C. O. (1999). Analysis of zinc fingers optimized via phage display: evaluating the utility of a recognition code. *J. Mol. Biol.* **285**, 1917-1934.
- Zhong, T. P., Rosenberg, M., Mohideen, M. A., Weinstein, B. and Fishman, M. C. (2000). gridlock, an HLH gene required for assembly of the aorta in zebrafish. *Science* **287**, 1820-1824.