Automated high-throughput mapping of promoter-enhancer interactions in zebrafish embryos

Jochen Gehrig[1,2,5], Markus Reischl[3,5], Éva Kalmár[1,5], Marco Ferg[1], Yavor Hadzhiev[1,2], Andreas Zaucker[1,2], Chengyi Song[1,4], Simone Schindler[1], Urban Liebel[1] & Ferenc Müller[1,2]

Zebrafish embryos offer a unique combination of high-throughput capabilities and the complexity of the vertebrate animal for a variety of phenotypic screening applications. However, there is a need for automation of imaging technologies to exploit the potential of the transparent embryo. Here we report a high-throughput pipeline for registering domain-specific reporter expression in zebrafish embryos with the aim of mapping the interactions between *cis*-regulatory modules and core promoters. Automated microscopy coupled with custom-built embryo detection and segmentation software allowed the spatial registration of reporter activity for 202 enhancer-promoter combinations, based on images of thousands of embryos. The diversity of promoter-enhancer interaction specificities underscores the importance of the core promoter sequence in *cis*-regulatory interactions and provides a promoter resource for transgenic reporter studies. The technology described here is also suitable for the spatial analysis of fluorescence readouts in genetic, pharmaceutical or toxicological screens.

High-throughput sequencing of the human and model vertebrate genomes has generated a tremendous amount of sequence information that provides the raw material for the holistic analysis of the molecular organization of organisms. However, sequence data can only be used to decipher genomes if functional assays are designed and implemented with matching high-throughput capabilities. In particular, the development and exploitation of imaging technologies for high-content analysis of gene expression and function in the context of the organism is a primary objective[1]. The zebrafish is an ideal vertebrate system for genome-scale functional assays as well as for drug and toxicology screens owing to its high-throughput capabilities (reviewed in refs. 2–4). Analysis of reporter gene expression in transient and stable transgenic zebrafish embryos provides the opportunity for efficient and rapid detection of tissue-specific gene activity[5–7] and accelerates functional testing of bioinformatically predicted *cis*-regulatory modules (CRMs) of vertebrate developmental genes[8]. However, these studies have remained limited in capacity because of the lack of automated tools for the spatial detection of reporter gene expression. Although advances have been reported in detecting tissue-specific reporter expression[9–11], the ability to automate and simultaneously register reporter gene activity in multiple tissues or domains in tens of thousands of fish embryos would greatly enhance the potential of zebrafish for functional analysis of CRMs.

CRMs are found in scattered locations, often at a large distance from genes[11–13] and often contain disease-causing mutations[12,13]. Correct identification of the cognate targets of CRMs requires an understanding of the mechanisms specifically linking CRMs and their target gene promoters. The recent discovery of an unexpected diversity of core promoter features[14] and of transcription factors that can bind to core promoters suggests a specific regulatory role for CRM interactions (reviewed in refs. 15,16). However, insight into the specificity of enhancer-promoter interactions is lacking in the context of the vertebrate embryo. For this purpose, we developed a systems biology approach to document reporter expression in transient transgenic zebrafish embryos.

Here we demonstrate a pipeline based on automated image acquisition of thousands of arrayed live transgenic zebrafish embryos and the warping of experimental embryos onto a two-dimensional reference shape. The warped overlay images allowed the registration of mosaic and continuous fluorescence signals in transient and stable transgenics at different developmental stages. Key embryo domains relevant for high-throughput expression screening were scored and expression in these domains was quantitatively processed. The analysis of over 200 reporter constructs revealed a diversity of enhancer-promoter interactions and provides a core promoter resource for transgenic applications such as enhancer trapping or functional analysis of bioinformatically predicted CRMs.

RESULTS

Embryo transgenesis and imaging

To address the interaction specificity between enhancers and a diverse set of core promoters, we designed a pipeline for a reporter

[1]Institute of Toxicology and Genetics, Forschungszentrum Karlsruhe, Eggenstein-Leopoldshafen, Germany. [2]Department of Medical and Molecular Genetics, School of Clinical and Experimental Medicine, College of Medical and Dental Sciences, University of Birmingham, Birmingham, UK. [3]Institute of Applied Computer Science, Forschungszentrum Karlsruhe, Eggenstein Leopoldshafen, Germany. [4]Present address: College of Animal Science and Technology, Yangzhou University, Jiangsu, China. [5]These authors contributed equally to this work. Correspondence should be addressed to F.M. (f.mueller@bham.ac.uk) or U.L. (urban.liebel@kit.edu).

RECEIVED 8 JULY; ACCEPTED 22 SEPTEMBER; PUBLISHED ONLINE 8 NOVEMBER 2009; DOI:10.1038/NMETH.1396

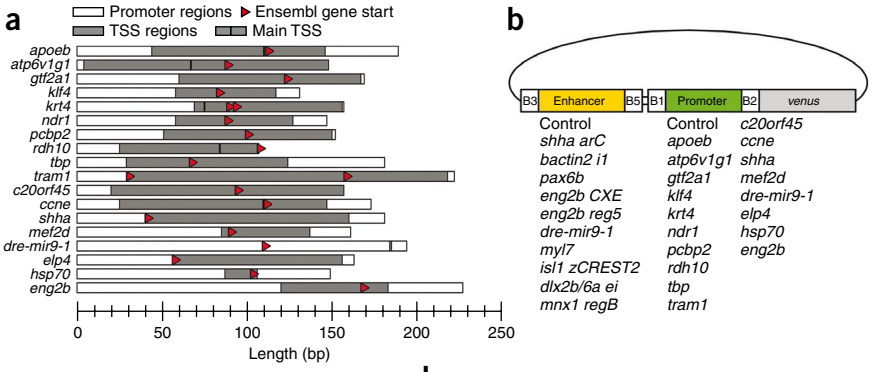

Figure 1 | A pipeline for automated spatial registration of tissue-specific reporter gene activity in zebrafish embryos. (**a**) Core promoters used in this study are represented as horizontal bars indicating length. Gray areas represent transcription start site (TSS) regions covered by 5′ ends of expressed sequence tags or GenBank mRNAs that map in the promoter region[27]. (**b**) Schematic of the reporter constructs with a list of 19 core promoter fragments and 11 enhancers that we used with a *venus* reporter in Multisite Gateway expression constructs[28]. B1, B2, B3 and B5 indicate recombination sites. (**c**) We sorted dechorionated and CFP-positive embryos into agarose coated plates (left) at prim-20 stage (up to 92 per plate) and oriented them for lateral view by a 0.5 mm hole holding the yolk ball (right). (**d**) For each embryo, we acquired brightfield images (left) as well as cyan (middle) and yellow (right) fluorescence channel views. Scale bars, 100 µm.

expression screen of microinjected zebrafish embryos (**Fig. 1**). We chose core promoters to represent a diverse set of gene expression profiles in the embryo; we defined the exact locations of the core promoters on the basis of transcription start site distribution (**Fig. 1a**, Online Methods and **Supplementary Table 1**). To test the capacity of core promoters to interact with enhancers, we linked each of 10 CRMs (enhancers), representing various tissue-specific activities present in the embryo, (**Supplementary Table 2**) plus a negative control, to each core promoter and a *venus* reporter (**Fig. 1b** and **Supplementary Table 3**). We injected zygote stage embryos with the enhancer-promoter constructs and with *in vitro* synthesized *cfp* mRNA as an injection control. We manually selected embryos for CFP activity at prim-6 stage. We imaged laterally positioned prim-20 or long-pec stage embryos in 4 *z*-dimension slices, from which an extended focus image was generated (**Fig. 1c**,**d**). The expression of Venus was mosaic and varied among individuals in an injection batch, as expected for plasmid DNA directly injected into embryos[5,6,8]. Thus, we imaged a total of 17,793 embryos representing 202 enhancer-promoter constructs in a total of 213,516 images.

Registration and projection tool

To identify the domain specificity of mosaic Venus activity in embryos at high throughput, we developed a two-dimensional embryo registration and projection tool (Online Methods and **Supplementary Fig. 1**). We used the cyan channel images of individual embryos to detect the embryo outline (**Supplementary Note**) after which the embryos were oriented automatically by the software. We used a set of anchor points to determine the outline and several key morphological features such as the yolk cell, cerebellum and notochord (**Supplementary Fig. 1**). Landmarking generated a two-dimensional reference embryo shape by averaging 12,582 embryo images (**Supplementary Fig. 2a**). We projected all CFP and YFP images that matched our quality control of experimental embryos (**Supplementary Note**) onto the reference embryo shape by warping (**Supplementary Table 4**). This step resulted in a single overlay image for each construct (**Fig. 2**). Individual embryos injected, for example, with an *isl1 zCREST2* enhancer and an *eng2b* promoter construct resulted in a small number of Venus-expressing cells with limited information about overall specificity of CRM activity (**Fig. 2d**). However, the maximum projection images generated from all experimental embryos injected with this combination revealed a comprehensive pattern of neural-specific Venus activity (**Fig. 2f**), confirming the published function of the *isl1 zCREST2* enhancer[17]. Projection carried out on the CFP images, in which the reporter was expressed throughout the embryo, demonstrated the quality of the overlay (**Fig. 2a–c**).

To test whether the embryo registration protocol can also be applied to other developmental stages with different overall

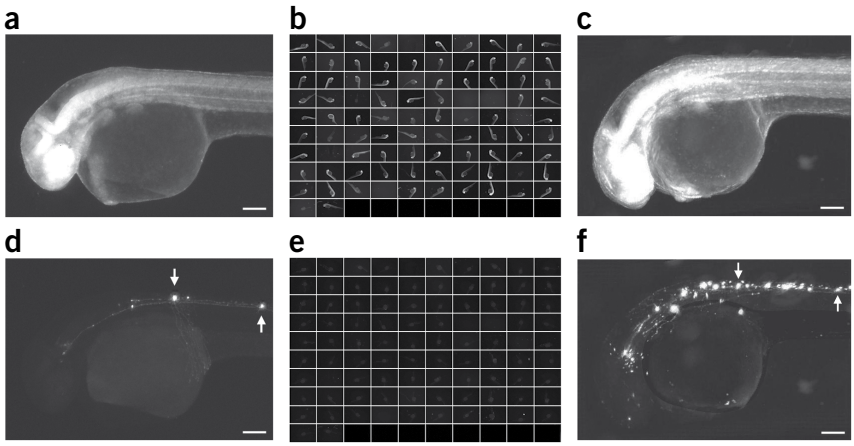

Figure 2 | Two-dimensional embryo image warping tool to detect mosaic reporter expression patterns in groups of embryos. (**a**–**f**) Cyan (**a**–**c**) and yellow (**d**–**f**) fluorescence filter views of the same embryos injected with a Venus reporter construct containing the *isl1 zCREST2* enhancer linked to an *eng2b* minimal promoter and with *cfp* mRNA. Shown are images of a single transient transgenic embryo (**a**,**d**); thumbnail images of multiple embryos (**b**,**e**); and maximum projections of the embryos in **b** and **e** warped onto the reference embryo shape (**c**,**f**). Arrows in **d** and **f** indicate single neurons expressing Venus. Scale bars, 100 µm.

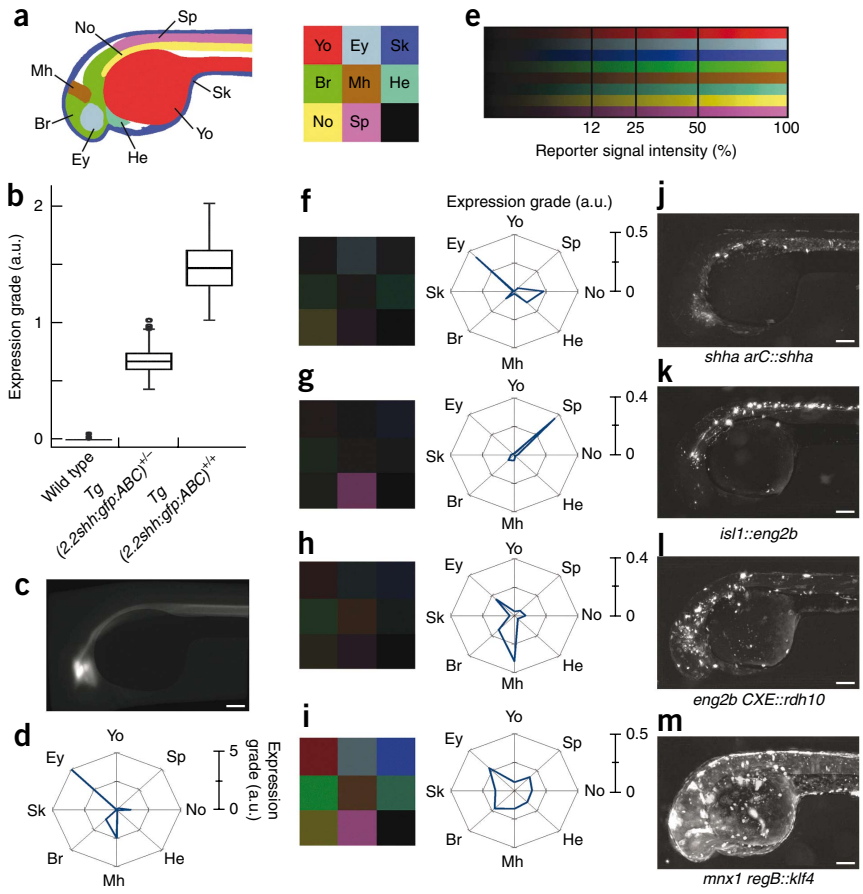

Figure 3 | Registration of domain-specific expression in transgenic embryos. (**a**) Superimposed arbitrary domain definitions from 26 manually segmented embryos warped onto the reference embryo shape (left) and color code to indicate domains (right). Yo, yolk ball and extension domain; Ey, eye domain; Sk, embryonic skin domain; Br, brain domain; Mh, midbrain-hindbrain boundary domain; He, heart domain; Sc, spinal cord domain; and No, notochord domain. (**b**) Box plot indicating distributions of measured GFP signal intensities of nontransgenic wild type ($n = 76$), *Tg(2.2shh:gfp:ABC)*$^{+/-}$ ($n = 80$) and *Tg(2.2shh:gfp:ABC)*$^{+/+}$ ($n = 40$) embryos. For definition of expression grade see equations 5–9 in the **Supplementary Note**. (**c**) Mean projection generated from embryos of the *Tg(2.2shh:gfp:ABC)* stable transgenic line. (**d**) Domain-specific distribution of GFP signal intensities of the transgenic line normalized to the domain size demonstrated by a radar plot of segmentation domains. Expression grade values are plotted along the axes for each segmentation domain and connected with a blue line. (**e**) Color scale illustrating color values and corresponding reporter signal intensities. (**f–i**) Distribution of Venus signal intensities normalized to domain size, indicated by color codes (left) and radar plots (right) for *shha arC::shha* (**f**), *isl1::eng2b* (**g**), *eng2b CXE::rdh10* (**h**) and *mnx1 regB::klf4* (**i**). (**j–m**) Maximum projections of distinct expression profiles generated by transient expression of the indicated constructs. Scale bars, 100 µm.

morphology, we used the embryo detection and segmentation tools on long-pec stage embryos (**Supplementary Fig. 2d–f** and **Supplementary Fig. 3m–o**).

Domain specificity of reporter expression

To achieve rapid spatial registration of reporter gene activity for hundreds of DNA constructs, we warped embryo images onto the two-dimensional reference shape, which we arbitrarily pre-segmented into domains, each of them representing several overlapping embryo tissues (**Supplementary Note** and **Fig. 3a**). We verified the quality of the automated domain assignment by comparing it to manually segmented embryo images (**Supplementary Fig. 2g–n** and **Supplementary Table 5**). We then measured pixel intensity in individual extended-focus embryo images. To test whether the pixel intensity count provides reliable quantification of fluorescence signal measurement, we used the image analysis tool to measure GFP reporter activity in hemizygous and homozygous embryos of a *Tg(2.2shh:gfp:ABC)* stable transgenic line[18]. The median pixel intensity count was proportional to the allelic copy number of the GFP transgene and indicated on average double fluorescence intensity in homozygotes as compared to hemizygotes (**Fig. 3b**). We displayed the reporter activity in batches of *Tg(2.2shh:gfp:ABC)* stable transgenic zebrafish embryos by mean projection overlay (**Fig. 3c**). To obtain a tissue-specific readout of reporter activity, we normalized the sum of Venus pixel intensity to the respective domain size and to the embryo numbers and presented the data in a radar plot. We observed reporter activity in the notochord, retina and brain compartments (**Fig. 3d**). Mosaic expression in transient transgenic embryos injected with the same construct as that used to generate the *Tg(2.2shh:gfp:ABC)* stable transgenic zebrafish line resulted in overlay patterns comparable to those in the stable line (**Supplementary Fig. 3g–l**).

We analyzed mosaic expression in microinjected embryos by using radar plots and a color intensity code (**Fig. 3a**,**e–i**). The domain-specific fluorescent signal quantification carried out on individual extended focus images indicated reporter activity in the same domains as seen in the maximum overlay projections in the same set of microinjected embryos (**Fig. 3f–m**). The spatial analysis of Venus expression resulted in reproducible patterns of domain specificity in independent repeats of the microinjection experiments with the *shha arC::krt4* construct (**Supplementary Table 6** and **Supplementary Fig. 4**). Overall, the maximum projections provided cellular resolution suitable for mosaic reporter activity, and mean projections allowed averaging of signals for demonstrating transgene activity in the stable transgenic line (**Supplementary Fig. 3**).

To test whether the warping and segmentation technology can be extended to registering other developmental stages, we imaged and processed long-pec stage *Tg(2.2shh:gfp:ABC)* embryos. This resulted in a comparable mean projection overlay and similar accuracy in domain registration to that of the same transgenic line at prim-20 stage (**Supplementary Fig. 3m–o** and **Supplementary Table 5**).

Promoter-enhancer interaction mapping

We injected 202 combinations of core promoters and enhancers into embryos and analyzed them for domain-specific reporter activity. The results are listed in **Supplementary Table 4**, and we

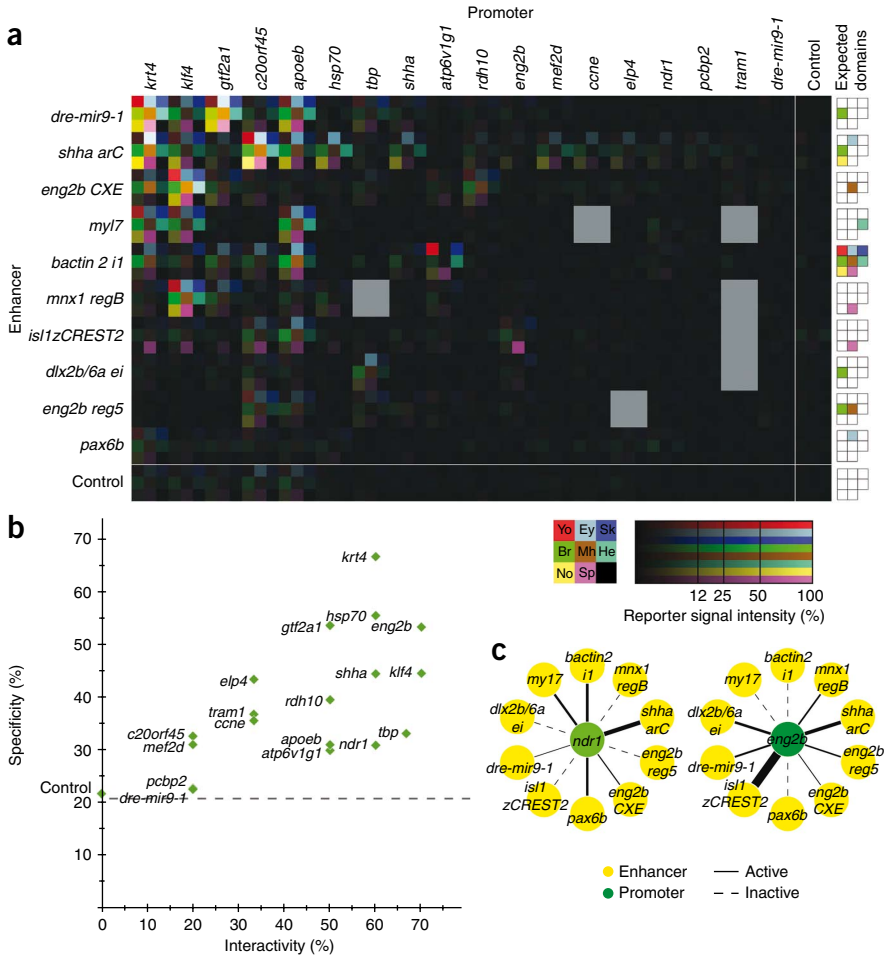

Figure 4 | Activities of enhancer-promoter combinations. (**a**) Interaction matrix of all 202 reporter constructs reflecting a total of 12,582 transient transgenic embryos. We arranged color intensity codes into rows (enhancers) and columns (promoters) and sorted them in descending order according to the average expression value of the enhancers or promoters. The white lines demarcate intensity values of negative control constructs. Gray boxes indicate enhancer-promoter combinations not assayed. The expected domain specificity of the enhancers is schematized on the right. (**b**) Plot of domain specificity (mean percentage) of reporter signal in the expected domains in conjunction with the *shha ar-C, eng2b CXE* and *isl1 zCREST2* enhancers versus interactivity (percentage of enhancer combinations for which fluorescence exceeds by at least twofold the maximum signal intensity of the corresponding control experiments) of core promoters and enhancers. Expected domains are indicated in **a**. Dashed line indicates the expected proportion of fluorescence signal by random mosaic expression. (**c**) Interactions of the *ndr1* and *eng2b* promoters with enhancers. Thickness of black lines indicates relative strength of interaction proportional to the total sum of brightness of pixels. 'Active' indicates that signal intensity exceeds two times the maximum of the relevant controls.

created an overview color-coded interaction matrix (**Fig. 4a**), in which we ranked constructs by their activity. Several core promoters had basal activity (for example, *c20orf45* and *apoeb*), whereas most promoters were inactive with a negative control enhancer. However, all core promoters without apparent basal activity (for example, *eng2b* and *ccne*) were activated by at least one enhancer, indicating that these core promoters were functional, with the ability to respond to CRM input. All tested enhancers activated at least one core promoter above the activity observed with the negative control enhancer fragment (**Fig. 4a**) confirming enhancer functionality. Upregulation of the core promoters by enhancers varied greatly, suggesting sequence-specific interactions.

Enhancer-core promoter combinations resulted in reporter activity in the expected domain with varying efficiency. To aid comparison of the observed activity with the expected activity of enhancers, we created a schematic diagram (based on published transgenic enhancer assays; **Supplementary Table 2**). For instance, *shha ar-C* was highly active with most promoters tested as expected in the notochord, eye and brain domains. We also clearly observed tissue specificity for other enhancers (for example, *isl1 zCREST2* and *eng2b CXE*), whereas the brain enhancer[12] *dre-mir9-1*, showed no apparent domain restriction in this assay. Moreover, enhancers did not necessarily interact with their cognate promoters more efficiently than with heterologous promoters. For example, the *shha ar-C* enhancer was more active with the *krt4* promoter than with the *shha* promoter, even though it mostly retained tissue specificity (**Fig. 4a**).

The differential efficiency of promoters in interacting with enhancers in the expected domains is an important consideration for promoter design in enhancer trap and transgenic assays of CRMs. To examine the reliability of a core promoter in responding to an enhancer in the expected domains, we tested all 19 promoters with 3 highly specific enhancers and ranked the results by the mean percentage of the activity in the expected enhancer-specific expression domains (**Fig. 4b**). The plot diagram also demonstrates, for each promoter, its capacity to interact with enhancers in general (interactivity) as demonstrated by the percentage of enhancer combinations in which the given core promoter showed activity above the basal level. The interaction matrix and interaction plot identified sets of promoters (for example, *krt4*, *hsp70* and *eng2b*) with the majority of Venus expression in the expected domains and broad capability of interactions with the majority of enhancers.

A notable observation was the apparent differential interaction specificity of individual enhancer-promoter combinations. For example, the *ndr1* and *eng2b* core promoters interacted with most enhancers differently (**Fig. 4c**). Thus, the interactivity as well as the domain or tissue specificity of enhancer-promoter interactions was highly dependent on the core promoters.

Detection of expression phenotypes

Beside the analysis of *cis*-regulatory function, our image acquisition and processing system may also serve as a tool for other screening applications based on reporter gene expression.

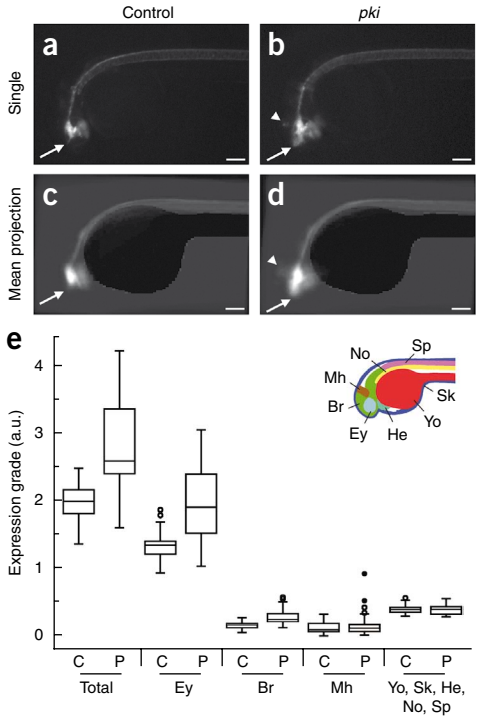

Figure 5 | Detection of domain specific changes of reporter gene expression in a stable transgenic line. (**a**,**b**) GFP expression in single heterozygous embryos of the stable transgenic line *Tg(2.2shh:gfp:ABC)* injected with control mRNA (**a**) or *pki* mRNA (**b**). (**c**,**d**) Mean projections of YFP images of control mRNA–injected embryos ($n = 27$; **c**) or *pki* mRNA–injected embryos ($n = 32$; **d**). Arrows point at the zona limitans intrathalamica with expansion of reporter gene expression in *pki* mRNA–injected embryos. Arrowheads indicate the additional reporter gene expression in the dorsal midbrain and hindbrain in *pki* mRNA–injected embryos. Scale bars, 100 µm. (**e**) Box plots illustrating the domain-specific changes of reporter gene expression in *pki* mRNA injected embryos. The distributions of signal intensities of total embryos and of selected affected domains are shown. Abbreviations of domains are as described in **Figure 3a**. C, control mRNA–injected ($n = 27$). P, *pki* mRNA–injected ($n = 32$).

To test whether the image analysis system is sufficiently sensitive to distinguish expression phenotypes that may emerge in screening situations, we overexpressed a dominant negative protein kinase A (PKI) that mimics ectopic Shh signaling, resulting in the expansion of *shh* expression in the dorsal brain[19,20]. Accordingly, we detected the expansion of GFP into the dorsal brain in the transgenic line *Tg(2.2shh:gfp:ABC)* injected with *pki* mRNA. The embryo registration assay applied on the *pki* injected embryos detected and quantified tissue-specific differences in the GFP signal (**Fig. 5** and **Supplementary Table 7**). This result indicates that minor changes in reporter activity affecting a small proportion of tissues without obvious change in the gross embryo morphology can be automatically detected and demonstrates the utility of the screening system in gene expression phenotype analysis.

DISCUSSION

In our system, the simplification of image analysis by two-dimensional warping allows short computing time, keeps hardware requirements low and enables the rapid analysis of up to tens of thousands of zebrafish embryos. The customized self-learning algorithm provides embryo feature parameters that are not determined by fixed thresholds but that are extracted individually out of the context of each image or an assembly of images of a training dataset. Thus, domain definitions within the reference embryo shape can be user defined and flexible. At the same time, the domains are of sufficient detail to detect change in tissue-specificity of well-defined organs and tissues (such as the retina, notochord or spinal cord).

The shape and size of both embryos and domains varied greatly among individuals owing to the natural shape variation and irregularity of developmental stages used. This emphasizes the importance of generating embryos at a well-defined developmental stage to maintain a high standard in domain registration at high throughput. However, the validation experiment demonstrated that the efficiency of the domain registration can be very high (up to 98% efficiency in the yolk domain).

The overlay technique overcomes the effect of reporter expression mosaicism in an analogous fashion to previously used manual expression plotting protocols[6,8], but with the added benefit of automation. This facilitated the analysis of fluorescence signals in transient transgenic embryos with comparable results to stable transgenics. However, direct comparison of stable and transient transgenics needs to be treated with caution because of the specific limitations of both systems. For example, stable transgenics vary in pattern and intensity among lines due to position effects (for example, ref. 21), whereas transient expression is prone to loss of some regulatory control of transgene expression that may require the chromosomal context. Keeping the above limitations in mind, transient transgenics provide a rapid first-level screening assay particularly suited for high throughput.

Beside the automated imaging pipeline, we generated a core promoter resource. Our assay, which brought enhancers in close proximity to core promoters outside of the chromosomal context demonstrated that core promoter identity influenced its ability to interact with CRMs. Thus the ranking of the interaction specificity of 19 different core promoters will aid in choosing promoters for CRM function and enhancer trapping assays, as well as in conventional transgenic reporter experiments to generate cell type–specific reporter transgenic lines[22–25]. Notably, strong promoters were efficiently activated by enhancers, but they also generated more ectopic expression (for example, *klf4*). In contrast, some weaker promoters (such as *hsp70* and *eng2b*) were also likely to interact with a variety of enhancers and in a domain-specific manner and thus may be useful in enhancer studies. We recommend using several core promoters with a given enhancer to detect potential promoter bias of enhancer specificity. We found no significant association between interactivity and core promoter sequence properties (data not shown), and additional work will be required to extend the promoter assay to a larger, more diverse set of core promoters to elucidate sequence determinants of specificity and interactivity.

Suggested applications for the automated imaging system include high-throughput functional analysis of new putative CRMs predicted by bioinformatics[8] and detection of gene expression phenotypes caused by toxic and pharmacological compounds. As demonstrated in our model experiment through the

modulation of the Hedgehog signaling pathway, genetic screens[26] could use our method for detecting gene expression phenotypes in the living zebrafish embryo.

The current technology allows image acquisition for up to 2,000 embryos within 4 h by a single microscope system. This speed also allows the image processing to follow within a similar time frame. The screening system may be scaled up to 10,000 embryos per day without substantial technological change (for example, by increasing the rate of embryo production and extending the period of imaging); however, for this level of throughput, the technology would greatly benefit from novel automation tools for the initial steps of embryo plating and orienting. Thus, domain registration in two-dimensional embryos may ultimately support the application of zebrafish in large-scale biotechnology projects.

METHODS

Methods and any associated references are available in the online version of the paper at http://www.nature.com/naturemethods/.

Note: Supplementary information is available on the Nature Methods website.

ACKNOWLEDGMENTS

We thank N. Borel for fish care, K. Straatman, N. Groebner and P. Kobor for technical support, T. Becker (Brain and Mind Research Institute, University of Sydney) for the *dre-mir9-1* enhancer and R. Sanges and E. Stupka for unpublished promoter analysis. This work was supported by FP6 projects EUTRACC and TRANSCODE, by the European Commission and the Deutsche Forschungsgemeinschaft to F.M. and programme grant by the Helmholtz Association of German Research Centres (HGF) to U.L. We thank the technical support team of Olympus Germany and Olympus UK, U. Strähle for general support and N. Foulkes and A. Cullinane for advice and critical reading of the manuscript.

AUTHOR CONTRIBUTIONS

E.K. created the constructs, E.K., J.G. led the screening and imaging of embryos, M.R. designed and created automation software and automated image processing, F.M. and U.L. supervised the screen, M.F., Y.H., A.Z., C.S. and S.S. participated in the embryo screen. J.G., M.R., E.K. and F.M. analyzed data, U.L. designed and constructed the screening platform, F.M., E.K., J.G., M.R. and U.L. designed the study and J.G., M.R., E.K. and F.M. wrote the manuscript.

Published online at http://www.nature.com/naturemethods/.
Reprints and permissions information is available online at http://npg.nature.com/reprintsandpermissions/.

1. Pepperkok, R. & Ellenberg, J. High-throughput fluorescence microscopy for systems biology. *Nat. Rev. Mol. Cell Biol.* **7**, 690–696 (2006).
2. Zon, L.I. & Peterson, R.T. *In vivo* drug discovery in the zebrafish. *Nat. Rev. Drug Discov.* **4**, 35–44 (2005).
3. Lieschke, G.J. & Currie, P.D. Animal models of human disease: zebrafish swim into view. *Nat. Rev. Genet.* **8**, 353–367 (2007).
4. Yang, L. *et al.* Zebrafish embryos as models for embryotoxic and teratological effects of chemicals. *Reprod. Toxicol.* **28**, 245–253 (2009).
5. Westerfield, M., Wegner, J., Jegalian, B.G., DeRobertis, E.M. & Puschel, A.W. Specific activation of mammalian Hox promoters in mosaic transgenic zebrafish. *Genes Dev.* **6**, 591–598 (1992).
6. Müller, F. *et al.* Intronic enhancers control expression of zebrafish sonic hedgehog in floor plate and notochord. *Development* **126**, 2103–2116 (1999).
7. Barton, L.M. *et al.* Regulation of the stem cell leukemia (SCL) gene: a tale of two fishes. *Proc. Natl. Acad. Sci. USA* **98**, 6747–6752 (2001).
8. Woolfe, A. *et al.* Highly conserved non-coding sequences are associated with vertebrate development. *PLoS Biol.* **3**, e7 (2005).
9. Tran, T.C. *et al.* Automated, quantitative screening assay for antiangiogenic compounds using transgenic zebrafish. *Cancer Res.* **67**, 11386–11392 (2007).
10. Liu, T. *et al.* Computerized image analysis for quantitative neuronal phenotyping in zebrafish. *J. Neurosci. Methods* **153**, 190–202 (2006).
11. Vogt, A. *et al.* Automated image-based phenotypic analysis in zebrafish embryos. *Dev. Dyn.* **238**, 656–663 (2009).
12. Kleinjan, D.A., Seawright, A., Childs, A.J. & van Heyningen, V. Conserved elements in Pax6 intron 7 involved in (auto)regulation and alternative transcription. *Dev. Biol.* **265**, 462–477 (2004).
13. Lettice, L.A. *et al.* A long-range Shh enhancer regulates expression in the developing limb and fin and is associated with preaxial polydactyly. *Hum. Mol. Genet.* **12**, 1725–1735 (2003).
14. Carninci, P. *et al.* Genome-wide analysis of mammalian promoter architecture and evolution. *Nat. Genet.* **38**, 626–635 (2006).
15. Sandelin, A. *et al.* Mammalian RNA polymerase II core promoters: insights from genome-wide studies. *Nat. Rev. Genet.* **8**, 424–436 (2007).
16. Juven-Gershon, T., Hsu, J.Y., Theisen, J.W. & Kadonaga, J.T. The RNA polymerase II core promoter—the gateway to transcription. *Curr. Opin. Cell Biol.* **20**, 253–259 (2008).
17. Uemura, O. *et al.* Comparative functional genomics revealed conservation and diversification of three enhancers of the isl1 gene for motor and sensory neuron-specific expression. *Dev. Biol.* **278**, 587–606 (2005).
18. Shkumatava, A., Fischer, S., Müller, F., Strahle, U. & Neumann, C.J. Sonic hedgehog, secreted by amacrine cells, acts as a short-range signal to direct differentiation and lamination in the zebrafish retina. *Development* **131**, 3849–3858 (2004).
19. Strahle, U., Fischer, N. & Blader, P. Expression and regulation of a netrin homologue in the zebrafish embryo. *Mech. Dev.* **62**, 147–160 (1997).
20. Müller, F. *et al.* Direct action of the nodal-related signal cyclops in induction of sonic hedgehog in the ventral midline of the CNS. *Development* **127**, 3889–3897 (2000).
21. Hadzhiev, Y. *et al.* Hedgehog signaling patterns the outgrowth of unpaired skeletal appendages in zebrafish. *BMC Dev. Biol.* **7**, 75 (2007).
22. Ellingsen, S. *et al.* Large-scale enhancer detection in the zebrafish genome. *Development* **132**, 3799–3811 (2005).
23. Parinov, S., Kondrichin, I., Korzh, V. & Emelyanov, A. Tol2 transposon-mediated enhancer trap to identify developmentally regulated zebrafish genes in vivo. *Dev. Dyn.* **231**, 449–459 (2004).
24. Scott, E.K. *et al.* Targeting neural circuitry in zebrafish using GAL4 enhancer trapping. *Nat. Methods* **4**, 323–326 (2007).
25. Asakawa, K. *et al.* Genetic dissection of neural circuits by Tol2 transposon-mediated Gal4 gene and enhancer trapping in zebrafish. *Proc. Natl. Acad. Sci. USA* **105**, 1255–1260 (2008).
26. Jin, S.W. *et al.* A transgene-assisted genetic screen identifies essential regulators of vascular development in vertebrate embryos. *Dev. Biol.* **307**, 29–42 (2007).
27. Kuhn, R.M. *et al.* The UCSC genome browser database: update 2007. *Nucleic Acids Res.* **35**, D668–D673 (2007).
28. Roure, A. *et al.* A multicassette Gateway vector set for high throughput and comparative analyses in ciona and vertebrate embryos. *PLoS One* **2**, e916 (2007).

ONLINE METHODS

Core promoters and enhancers. We chose the pool of core promoters from genes that are expressed in the early embryo[29,30], and we isolated and functionally verified these promoters by reporter assays[30] (J.G. *et al.*, unpublished data; **Supplementary Table 1**). We applied the following criteria when choosing the promoters: (i) at least 1 verified transcriptional start site (TSS) was annotated to the genome sequence (zv7), (ii) promoters represented a diversity of gene classes such as spatially restricted (for example, *apoeb*) and ubiquitous (for example, *tbp*). As negative control, we used the *pbFOG* promoter of *Ciona intestinalis*[28], which is transcriptionally silent in *Danio rerio* (data not shown). We predicted the location of the main TSS region of genes by using the DBTSS database[31] and the ensembl genome browser (**Supplementary Table 1**). Additionally, we confirmed the promoter regions by analyzing the distribution of 5′ ends of expressed sequence tags in the University of California Santa Cruz (UCSC) genome browser[27] in regions flanking the defined TSS region (**Fig. 1a** and **Supplementary Table 1**). We designed PCR primers to flank a region ranging between 107–227 base pairs incorporating the predicted main TSS region in cases where a dominant TSS peak was evident (**Fig. 1a** and **Supplementary Table 1**). For most promoters, however, no TSS peak was evident; in these cases we designed the primers to include the highest possible number of potential TSSs within size constraints (**Supplementary Table 3**).

We selected enhancers with previously described spatial patterns of activity matching or overlapping with the corresponding endogenous gene (**Supplementary Table 2**): *shha ar-C*[6], β-actin intron 1 (ref. 32), *pax6b eye*[8], *eng2b CXE*[33], *eng2b reg5* (this report), *dre-mir9-1* (ref. 34), *myl7* (ref. 35), *islet1 zCREST2* (ref. 17), *dlx2b/dlx6a ei*[36] and *mnx1 regB*[37]. We used *VC_909*, a nonconserved, noncoding segment without enhancer activity from *Takifugu rubripes* as a negative control[38]. We verified the activity of the PCR-isolated enhancers by reporter injections in fish embryos in conjunction with a test core promoter (data not shown). The insertion of gateway vector–derived DNA sequences between regulatory sequences and the *venus* reporter gene did not affect enhancer-promoter interaction[28].

Isolation of zebrafish core promoters and enhancers. We PCR-amplified enhancer and core promoter fragments using Triple Master polymerase enzyme mix (Eppendorf) from zebrafish genomic DNA isolated with DNeasy Tissue kit (Qiagen) (primer sequences are listed in **Supplementary Table 3**). We agarose gel–purified the PCR fragments by SV Gel and PCR Clean-Up system (Promega).

Generation of multisite Gateway entry clones and expression vectors. To generate entry clones (ECs), we carried out PB recombination reactions between the PCR products and pDONR 221 donor vectors as described[28]. For the generation of pSP72-B3-enhancer-B5::B1:core promoter:B2-Venus expression vectors (ExV), we carried out LR recombination reactions between promoter and enhancer containing ECs and the destination vector as described[28]. Gateway expression vectors with enhancer-promoter combinations are referred to as '*enhancer::core promoter*'. After transformation of the BP and LR reaction solutions, we subjected bacterial colonies to colony-PCR performed with EC or ExV specific primers (**Supplementary Table 3**). We isolated plasmids containing the correct fragment size (QiaPrep Spin Miniprep kit; Qiagen) and verified them by sequencing (ECs) or by restriction digestion (ExVs).

Fish keeping. We maintained zebrafish (*Danio rerio*, AB) stocks at 28 °C, according to ref. 39. Zebrafish were kept and used according to Home Office (UK) license 30/4141. We determined the developmental stages of the embryos by morphological features, as described previously[40]. We collected eggs from pairwise and batch crossings within a short time window of 5–15 min after fertilization.

Microinjection and embryo handling. We diluted plasmids for injection solutions with nuclease-free water to a final concentration of 5 ng μl^{-1} and 0.1% phenol red and 15 ng μl^{-1} *cfp* mRNA produced from the *pCS2+::CFP* construct by *in vitro* transcription (Message Machine; Ambion). We microinjected the solutions manually through the chorion into the cytoplasm of one cell-stage zygotes 15–25 min after fertilization. We raised the embryos at 30 °C in fish water containing 0.003% phenylthiourea (PTU). We dechorionated the embryos at prim-5 stage with 10 mg ml^{-1} Pronase (Sigma) and discarded CFP fluorescence–negative embryos from further analysis. We anaesthetized microinjected embryos using 0.03% tricaine, plated them in 96-well agarose embedded plates and oriented them laterally by inserting the yolk ball of embryos into a concentric depression (500 μm diameter) in the agarose (**Fig. 1c**).

HTS microscopy, imaging, data acquisition and data storage and processing. We imaged the 96-well plates on a Scan^R high content screening microscope[41] (Olympus Biosystems) with a SWAP plate gripper (Hamilton) and a ×2.5 objective (Plan-Apo) and an Olympus Biosystems DB-1 (1,300 × 1,024 pixels) camera in bright field and with CFP, YFP filter cubes. We fixed image integration times (180 ms for CFP and 1,000 ms for YFP). The light source was an ultra stable MT-20 (Olympus Biosystems) with a 150 W xenon lamp. We detected the focal plane of the embryo by an object detection autofocus algorithm. We acquired images of each embryo in four *z*-dimension slices (55 μm). We carried out data management, thumbnail gallery generation and data compression via an assembly of LabView software modules (National Instruments).

Embryo detection and segmentation, domain specificity analysis. For detailed description of the algorithms applied, see the **Supplementary Note**. In brief, we used a combination of bright field and CFP-images to register domains within a simplified two dimensional embryo outline. We adjusted embryos by rotation and reflection to a standard orientation and excluded erroneous images (for example, malformed embryos). We combined four *z*-dimension slices to obtain maximum sharpness. We introduced a mathematical model to describe the ventral and the dorsal body curvature. We derived landmarks (anchor points) to determine characteristic points in each embryo, which we used in subsequent warping of individual images. We generated landmarks for (i) the outline of the embryo, (ii) the outline of the yolk cell which defines gross embryo morphology, (iii) key morphology points including a horizontal set of landmarks allowing detection of the body axis (notochord) and a set of radial points (for example, cerebellum base) for detecting circumferential variation

in morphology during development. We trained a learning classifier on the landmarks of 51 well-registered images and excluded erroneous registrations in the whole dataset. We averaged domain activities for all acquired images of the same construct (background yolk fluorescence and saturated areas were deduced). To visualize the outcome of a single construct we warped all images onto a reference embryo and projected images of a given experiment onto each other. Based on key embryonic tissues we manually segmented a small set of embryos (**Supplementary Note**) to generate characteristic domains within the embryo. The two-dimensional nature of segmentation domains mean that fluorescence signal from transversely underlying and overlying tissues are also counted together with signal emanating from the embryonic structure which gives name to the domain. We then transferred the segmentation domains onto the reference shape by warping. Thus we used the segmented reference shape for further analysis of fluorescence distribution of experimental embryos. The image processing algorithms are not restricted to images recorded by ScanR but can handle images taken on a simple stereo fluorescence microscope (data not shown).

Color plot and interaction map. We assigned a color to each domain and the reporter activity within the domain was shown by color intensity. We assigned the maximum Venus intensity value of the given domain within all experiments carried out as 100%. The color gradient was nonlinear, obtained by the root of the average domain-specific signal, and resulted in black for no activity. We assigned enhancers and promoters with an expressivity value by averaging all eight domain activities for all constructs and subtracting the activity of the control experiments with the same enhancer or promoter, respectively.

Specificity and interactivity. We determined for each promoter the percentage of signal within the expected domains for the *shha arC*, *eng2b CXE* and the *isl1 zCREST2* enhancer. We took the mean of these percentages as the specificity value reflecting the proportion of signal matching expected enhancer activity. Also we calculated the theoretical mean proportion of random signal within expected domains. Interactivity describes the percentage of experiments with a given promoter where fluorescence exceeded the maximum of the negative promoter- and the negative enhancer control by the factor 2. The specificity and interactivity values were plotted on an *x*-*y* plot.

Maximum and mean projections. To visualize the tissue distribution of reporter activity by a given construct, we projected all warped images (size, 549 × 359 pixels) of this construct to one 549 × 359 pixel image. We derived the pixel values of the projection image from all pixels located at the same position in the warped images. We carried out projections either by taking the maximum value or taking the mean value of each pixel (**Supplementary Fig. 3**).

Quantitation of fluorescence signal in a *shh:gfp* stable transgenic line. We crossed individual males and females of the stable transgenic line *Tg(2.2shh:gfp:ABC)*[18] with wild type adults and determined the parental genotype by the mendelian frequency of transgenic embryos among the offspring. We obtained heterozygous and homozygous embryos from a single identified homozygous female by two independent crosses with either a wild type or a homozygous transgenic male, respectively. We microinjected the obtained embryos with *cfp* mRNA, imaged them and processed the images as described in the **Supplementary Note**.

Generation of transgene expression phenotype by mRNA injection. We used the pCSdnReg plasmid to synthesize *pki* mRNA[20]. We used the tol2 transposase gene[23] with no biological effect on development as control mRNA. We injected both mRNAs at 10 ng μl^{-1} into heterozygous zygote stage embryos of the stable transgenic line *Tg(2.2shh:gfp:ABC)*[18] in conjunction with *cfp* mRNA (15 ng μl^{-1}). We carried out analysis of fluorescent reporter signal at prim-20 stage as described above and in **Supplementary Note**.

Hardware and software used. We processed image data using Matlab 5.3 (Mathworks Inc.). We made all algorithms custom without use of external toolboxes. We stored the whole dataset (1.29 terabyte) on a Thecus 5200 device and image processing took 40 h using seven standard PC computers. We derived all visualizations directly from the data using Matlab scripts. Algorithms and visualization routines are available as **Supplementary Software.** Help with installation is available upon request from the authors and see http://hwiki.fzk.de/wiki/index.php/Zebrafish-screen.

29. O'Boyle, S., Bree, R.T., McLoughlin, S., Grealy, M. & Byrnes, L. Identification of zygotic genes expressed at the midblastula transition in zebrafish. *Biochem. Biophys. Res. Commun.* **358**, 462–468 (2007).
30. Ferg, M. *et al.* The TATA-binding protein regulates maternal mRNA degradation and differential zygotic transcription in zebrafish. *EMBO J.* **26**, 3945–3956 (2007).
31. Wakaguri, H., Yamashita, R., Suzuki, Y., Sugano, S. & Nakai, K. DBTSS: database of transcription start sites, progress report 2008. *Nucleic Acids Res.* **36**, D97–D101 (2008).
32. Müller, F. *et al.* Activator effect of coinjected enhancers on the muscle-specific expression of promoters in zebrafish embryos. *Mol. Reprod. Dev.* **47**, 404–412 (1997).
33. Song, D.L., Chalepakis, G., Gruss, P. & Joyner, A.L. Two Pax-binding sites are required for early embryonic brain expression of an Engrailed-2 transgene. *Development* **122**, 627–635 (1996).
34. Kikuta, H. *et al.* Genomic regulatory blocks encompass multiple neighboring genes and maintain conserved synteny in vertebrates. *Genome Res.* **17**, 545–555 (2007).
35. Huang, C.J., Tu, C.T., Hsiao, C.D., Hsieh, F.J. & Tsai, H.J. 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 (2003).
36. Zerucha, T. *et al.* A highly conserved enhancer in the Dlx5/Dlx6 intergenic region is the site of cross-regulatory interactions between Dlx genes in the embryonic forebrain. *J. Neurosci.* **20**, 709–721 (2000).
37. Nakano, T., Windrem, M., Zappavigna, V. & Goldman, S.A. Identification of a conserved 125 base-pair Hb9 enhancer that specifies gene expression to spinal motor neurons. *Dev. Biol.* **283**, 474–485 (2005).
38. Sanges, R. *et al.* Shuffling of cis-regulatory elements is a pervasive feature of the vertebrate lineage. *Genome Biol.* **7**, R56 (2006).
39. Westerfield, M. *The Zebrafish Book*. (University of Oregon Press, Eugene, Oregon, USA, 1995).
40. Kimmel, C.B., Ballard, W.W., Kimmel, S.R., Ullmann, B. & Schilling, T.F. Stages of embryonic development of the zebrafish. *Dev. Dyn.* **203**, 253–310 (1995).
41. Liebel, U. *et al.* A microscope-based screening platform for large-scale functional protein analysis in intact cells. *FEBS Lett.* **554**, 394–398 (2003).

doi:10.1038/nmeth.1396