Published in final edited form as:
*Genomics*. 2010 June ; 95(6): 363–369. doi:10.1016/j.ygeno.2010.02.011.

# Efficient discovery of *ASCL1* regulatory sequences through transgene pooling

**David M. McGaughey**[1] and **Andrew S. McCallion**[1,2]

David M. McGaughey: davidm@jhmi.edu; Andrew S. McCallion: andy@jhmi.edu

[1] McKusick - Nathans Institute of Genetic Medicine, Johns Hopkins University School of Medicine, 733 N. Broadway, BRB Suite 449, Baltimore, MD 21205, USA

[2] Department of Molecular and Comparative Pathobiology, Johns Hopkins University School of Medicine, Baltimore, MD 21205, USA

## Abstract

Zebrafish transgenesis is a powerful and increasingly common strategy to assay vertebrate transcriptional regulatory control. Several challenges remain, however, to the broader application of this technique; they include increasing the rate with which transgenes can be analyzed and maximizing the informational value of the data generated. Presently, many rely on the injection of individual constructs and the analysis of resulting reporter expression in mosaic G0 embryos. Here, we contrast these approaches, examining whether injecting pooled transgene constructs can increase the efficiency with which regulatory sequences can be assayed, restricting analysis to the offspring of germ line transmitting transgenic zebrafish in an effort to reduce potential subjectivity. We selected a 64 kb interval encompassing the human *ASCL1* locus as our model interval and report the analysis of 9 highly conserved putative enhancers therein. We identified 32 transgene-positive zebrafish, transmitting one or more independent constructs displaying *ASCL1*-like regulatory control. Through examination of embryos harboring one or more transgenes, we demonstrate that five of the nine sequences account for the observed control and describe their likely roles in *ASCL1* regulation. These data demonstrate the utility of this approach and its potential for further adaptation and higher throughput application.

## Introduction

Cis-regulation of transcription by noncoding DNA sequence plays crucial roles in development [1–5], homeostasis [6,7], inter-species variation [8–12], and disease risk [13–19]. In recent years regulatory sequences such as enhancers have garnered much research interest and commentary [20,21] and the repertoire of published enhancers has been expanded by an increasing number of mid and large-scale transgenic analyses performed *in vivo* [1,22–26]. These studies have recently been complemented by efforts to integrate sequence conservation and expression data with computational motif identification and also by analyses that have implemented emerging technologies like chromatin immunoprecipitation (ChIP) based assays

Correspondence to: Andrew S. McCallion, andy@jhmi.edu.

**Authors' contributions**

ASM conceived this study. DMM performed amplification and cloning of sequences, injections of zebrafish. DMM and ASM performed analysis and imaging of zebrafish embryos. DMM and ASM wrote the manuscript.

[27,28]. In combination these new technologies have shown significant promise in predicting tissue-dependent enhancer function on a genome-wide scale. Importantly, efforts to identify or validate predictions of regulatory sequences are in large part dependent on transgenic strategies applied in multiple vertebrate organisms and have been significantly facilitated in their application by recent improvements in technology and scale [29,30].

Mouse has for some time been considered the gold standard for functional analyses, and as such has been favored by many labs for transgenic studies of putative regulatory sequences. However, the efficacy of mouse transgenesis in high throughput applications is blunted by cost and time constraints that cause many studies to be restricted to transgenic analyses performed in G0 embryos at a single developmental time point. Perhaps for this reason, among others, transgenesis in non-mammalian vertebrates such as zebrafish has become an increasingly popular and powerful tool in these types of studies. These organisms provide significant cost benefits and facilitate analyses by live imaging at multiple time points during development due to their external fertilization and transparent embryos. As with mice, transgenic studies in zebrafish frequently rely on analyses performed in G0 embryos, which in the case of zebrafish can be highly mosaic. While this is a rapid and powerful approach, the mosaic nature of transgene expression makes it difficult to thoroughly characterize the regulatory control of a particular sequence. Interpretation of these mosaic expression patterns relies upon the documentation and integration of overlapping data from significant numbers of independent G0 embryos for any single construct (Table S1). This yields a composite imputation of expression that is inherently incomplete and makes scaling up to greater numbers of elements all the more challenging [31].

By contrast, stable transgene transmission through the germ allows a complete view of the tissue and temporal specific expression pattern directed by each regulatory sequence. Its application in large-scale studies has however been limited, likely due to the added time required to raise and screen offspring from identified transgene “founders” and the inherent increased cost and space. Taken in combination these issues compromise the rate at which one may comprehensively assay sequences on the increasing scale required by contemporary genomic analyses.

We wanted to assess whether a collection of putative regulatory sequences could be reliably assayed in a single experiment, in contrast to standard methods that introduce only one transgene per injection. In an effort to address these issues we set out to develop an efficient strategy that focuses on analysis post germ line transmission and pools constructs for injection. We have focused our efforts on the human *ASCL1* gene, encoding the Achaete-schute homolog 1. ASCL1 is a member of the basic-helix-loop-helix (bHLH) family of transcription factors that is required for development of many neural precursors, including components of noradrenergic, serotonergic, sympathetic, parasympathetic, and enteric neuronal populations [32–37]. Mutations in *ASCL1* have been associated with neuroendocrine tumors, Central Congenital Hypoventilation Syndrome (CCHS), and Parkinson’s disease [38–40]. This locus provides an ideal model for this effort for a number of reasons. One enhancer has already been identified at the mouse *Ascl1* locus, and an interval encompassing the entire *Ascl1* mouse locus was also shown to recapitulate much of the endogenous expression [41,42]. These previous studies define an interval in which we search for *ASCL1* enhancers. Additionally, the relatively small number of highly conserved sequences flanking the *ASCL1*, its tightly controlled expression during early development and the well-documented expression of the *ASCL1* orthologs in zebrafish and mice make it a good test case for this novel strategy [32,35].

We report the application of a novel transgenic pooling strategy in the analysis of the human *ASCL1* locus. We demonstrate that this method allows for rapid validation of *ASCL1* enhancers in stable transgenic zebrafish lines. The resulting transgenic composition of identified zebrafish

is readily established by a PCR-based assay, simplifying the necessary analyses and interpretation. We identify five enhancers directing expression that overlaps *ASCL1* and incompletely overlap one another, postulating that they may act cooperatively to yield the spectrum of regulatory control displayed by the endogenous *ASCL1*. In light of these data we conclude that this method can be used to efficiently analyze the regulatory potential of numerous sequences in the offspring of germ-line transmitting zebrafish and eliminates many issues related to mosaic analyses. We, however, observe several complicating factors in these analyses and propose several additional modifications that would facilitate scaling to systematically address larger sets of sequences.

## Results

### Development of a pooling transgenesis strategy

Zebrafish transgenesis is an established and powerful strategy to analyze transcriptional regulatory control but most common implementations share several limitations and bottlenecks. We, and others, most frequently inject a single amplicon into 50–200 embryos, creating mosaic transgenics. Currently, studies use either the transgene expression profile solely in the mosaic embryos or raises select transgene-positive embryos to sexual maturity for more comprehensive reporter analysis. Although mosaic embryos can be rapidly processed, their analysis is dependent on the determination of composite signal across many embryos, leaving the interpretation of their output somewhat subjective and incomplete. When analysis of germ-line transmitted offspring is required, embryos injected with individual constructs are raised discretely from other constructs, which, in large numbers, can represent a strain on zebrafish system capacity. To increase the efficiency with which potential regulatory noncoding sequences can be evaluated we set out to determine the efficacy of assaying pools of cloned sequences, injecting multiple constructs simultaneously into zebrafish embryos.

In this pilot pooling study, illustrated in Figure 1a, we used as our test case the human *ASCL1* locus. We selected and pooled together ten amplicons, nine selected sequences proximal to the *ASCL1* locus and a positive control sequence (zebrafish *phox2b* −11.2; [26]) previously demonstrated to direct robust expression discretely in the ventral anterior spinal cord by 48 hours post fertilization (hpf). The nine test sequences comprised the most highly conserved noncoding sequences within a 64 kb interval encompassing *ASCL1*, scored by 28-species MultiZ alignment with PhastCons [43]. Sequences ranged in size from 2.3 kilobases (kb) (*ASCL1*+54.4; the names are the sequence's distance in kb from the transcriptional start site of *ASCL1*) to 0.3 kb (*ASCL1*−1.4; Figure 1b and Supplemental Table 2)

Previous analysis at the mouse *Ascl1* locus established several transgenic LacZ reporter mouse lines containing up to 36 kb encompassing this gene [41]. The largest transgene, J1A, directed near complete *Ascl1*-like expression; the J1A interval is aligned to the human *ASCL1* locus (Figure 1b) using the UCSC BLAT tool [44]. Additionally, smaller portions of the J1A transgene were subsequently assessed, demonstrating that a 1.2 kb fragment (Transgene 14) also directed tissue specific expression [42]. We generated an amplicon encompassing sequence orthologous to Transgene 14 (*ASCL1*−6.1) among our nine selected amplicons. All nine sequences were subcloned into the pGWcfosGFP reporter construct [45].

We injected this pooled group of DNA into greater than 1000 zebrafish embryos, screening all viable injected embryos at 24, 48, and 72 hpf. During screening, we selected ~250 embryos displaying any mosaic GFP reporter signal to raise to sexual maturity. These zebrafish were then out-crossed to AB stocks and their offspring were screened for tissue specific expression of the GFP reporter. The patterns of reporter expression displayed by the offspring from all identified transgene-positive founder zebrafish were documented (Figure 2 and Supplemental Figure 1) and compared to endogenous *ascl1* expression.

### Expression pattern of zebrafish *ascl1a* and *asclb*

Zebrafish possess two orthologs of *ASCL1*, termed *ascl1a* and *ascl1b*. The expression patterns of *ascl1a* and *ascl1b* have already been extensively characterized [32]. *ascl1a* is expressed prominently in the telencephalon, diencephalon, midbrain, and hindbrain at 24 and 48 hpf, accompanied by weaker expression in the spinal cord. Additional expression can be found in the epiphysis, retina, and sympathetic chain. The expression pattern of *ascl1b* is similar, but expression in the spinal cord is more prominent than for *ascl1a*. To aid our analyses we conducted *in-situ* hybridization of *ascl1a* (because of its higher sequence identity to human *ASCL1*) on 24 and 48 hpf zebrafish embryos (Figure 3).

### Pooling transgenesis identifies five putative enhancers directing *ASCL1* consistent central neuronal expression patterns

Of the ~250 fish we identified in our G0 screen, we out-crossed 82 of the surviving mature male G0 zebrafish with AB females; their offspring were analyzed at 24, 48, and 72 hpf for tissue specific expression of GFP. 36/82 (44%) of established crosses resulted in the identification of embryos that displayed tissue specific expression. To determine which of the putative *ASCL1* enhancers could contribute to reporter expression in each transgenic line, DNA was extracted from GFP-positive G1 embryos and amplified independently using primers specific to each amplicon. These genotyping results, along with the tissues with GFP expression, are reported in Table 1. Seven of the nine human *ASCL1* amplicons (−7.3, −6.1, −1.4, +3.3, +7.0, +13.2, +54.4, Table 1) and the positive control sequence (*phox2b*−11.2) were present in the identified G1 founders. Four of the 36 identified founders contained only the zebrafish *phox2b* −11.2 sequence and displayed reporter expression as previously reported [26]; thus 32 fish remained to be analyzed for *ASCL1*-consistent enhancer activity. All seven *ASCL1* amplicons were identified in transgene-positive G1 embryos displaying reporter expression in *ascl1a/b* appropriate tissues, including the telencephalon, diencephalon, midbrain, and hindbrain. The majority of transgene transmitting embryos displayed reporter expression in hindbrain neuronal populations (27/32). Similarly the midbrain was marked in 26/32 and the spinal cord in 17/32, consistent with ascl1a/b and ascl1b respectively and with mammalian *ASCL1* orthologs [34]. *ASCL1*+3.3 was also identified in offspring from two "founder" G0 zebrafish displaying expression in the pronephric duct (Supplemental Table 2 and Supplemental Figure 1). Although not a domain of endogenous Ascl1 expression, this may reflect a dual role for this element in the regulation of neighboring genes in addition to potentially regulating *ascl1*. Indeed, the *PAH* gene encoding phenyalanine hydroxylase, which lies 40 kb upstream of the human *ASCL1* is expressed in the pronephric duct of embryonic zebrafish [46]. Alternatively this observation may simply reflect the analysis of a sequence out of its genomic context, consistent with observations in similar transgenic enhancer studies [23,26,47–49].

Tol2 transposon mediated transgenesis yields single and multiple independent integrations within the genome of injected embryos with varying frequency [29]. G1 embryos thus harbor one or more transgenes that may contribute to the observed reporter expression and may segregate independently in subsequent generations. Of the 36 identified G0 "founders", 16 generated embryos harboring only one transgene, 12 generated embryos harboring two transgenes, and offspring from the remaining 8 "founders" had three or more transgenes each (Supplemental Table 2). Because of the large number of transgenic zebrafish created we are able to illuminate the regulatory potential of five of the nine regulatory amplicons. The proportion of embryos expressing GFP and their corresponding spatial reporter patterns for each construct (independently or in combination with others) are reported in Table 1 (*ASCL1* −7.3, −6.1, −1.4, +3.3, +13.2; Figure 3). Images of the 28 transgenic embryos not displayed in the main text are provided in Supplemental Figure 1.

Two of the nine *ASCL1*-specific amplicons were not detected among transgene positive G1 embryos. This may reflect their inability or failure to direct tissue specific expression at the times evaluated and thus were not among those raised or that they do contain enhancer activity and that an insufficient number of founders were screened. To eliminate the second possibility, the two amplicons −0.4, and +19.9 were each re-injected into >200 embryos and screened independently for mosaic transgene expression at 24, 48, and 72 hpf. No tissue-specific expression was detected, suggesting that these regions do not function as enhancers in the assay. We also note that the +7.0 and +54.4 amplicons were not present alone in any GFP positive embryos; to confirm whether or not these amplicons could direct tissue specific expression these two were injected independently into >200 embryos. None of the embryos assayed between 24 and 72 hpf displayed tissue specific GFP expression, suggesting that these two do not contribute to ascl1 transcription as enhancers during early development.

Of the remaining five enhancers (*ASCL1* −7.3, −6.1, −1.4, +3.3, and +13.2) all direct *ASCL1* consistent expression patterns in embryonic zebrafish (Table 1). Representative images for five of the enhancers (−7.3, −6.1, −1.4, +3.3, and +13.2) are displayed in Figure 3. One founder was identified for the −7.3 amplicon, which has expression in the diencephalon, midbrain, and hindbrain (Table 1, Supplemental Figure 1).

Four independent founders were identified that integrated the element −6.1. In total, the four sequences directed expression across the telencephalon, diencephalon, midbrain, hindbrain, and spinal cord. However, while only one of the four founders displayed expression in the telencephalon and diencephalon, two or more showed expression in the midbrain, hindbrain, and spinal cord, indicating that this sequence has stronger specificity for the later tissues. −6.1 was designed to encompass the human ortholog of mouse *Ascl1* Tg14 [42]. This transgene directed expression restricted to the diencephalon, midbrain, and spinal cord. Variants of Tg14 that mutated a putative repressor E-box (CAGGTG) directed expression in a less restricted manner throughout the entire CNS [42]. Our −6.1 amplicon similarly contains this E-box repressor and is also tightly restricted to *ASCL1* tissues (midbrain, hindbrain, spinal cord).

Two founders were identified containing only the *ASCL1*−1.4 sequence. These founders showed expression in the diencephalon, midbrain, hindbrain, and spinal cord with differing frequencies. When analyzing all founders containing −1.4, most directed expression in the midbrain, hindbrain, and spinal cord (Figure 2). However, only 1/6 showed diencephalon-specific expression, suggesting this expression domain may not reflect endogenous control by this enhancer. By contrast founders carrying only the +3.3 element show a strong tendency to direct expression in the midbrain, hindbrain, and spinal cord, with 4/4 founders showing expression in these tissues. Only one founder showed expression in the diencephalon.

Finally, the element +13.2 has a single unique founder that directs expression in the midbrain, hindbrain, and spinal cord. It also has several founders that contain only +13.2 and *phox2b* −11.2, the latter is expressed generally in the ventral anterior spinal cord (Supplemental Table 2). This allows for better confidence in seeing whether +13.2 can direct expression in other tissues. The majority of embryos with multiple insertions that also contain +13.2 (6/9) direct expression in the telencephalon, suggesting that this sequence also contains enhancers who are active in this tissue as well as confirm expression in the midbrain (9/9), hindbrain (7/9), and spinal cord (5/9).

In summary, all CNS sites of expression discovered in our 32 transgene founders overlap *ASCL1* expression, suggesting they represent true endogenous *ASCL1* enhancers.

### *In-situ* hybridization reveals temporal specific expression patterns

*ASCL1* expression is temporally dynamic and tightly spatially controlled within the developing nervous system. We have previously seen how enhancers at a single locus can display overlapping spatial control but discrete temporal control; thus to better determine whether these enhancers also display temporal-specific expression in developing zebrafish we completed GFP *in-situ* hybridization on four lines (−6.1, −1.4, +3.3, +13.2, Figure 5). *In* situ hybridization of zebrafish *ascl1a*, tyrosine hydroxylase (*th*), and dopamine beta hydroxylase (*dbh*), were also performed to provide landmarks that overlap some sites of *ASCL1* expression and would aid in annotating the sites GFP expression (Supplemental Figure 2). *th* is expressed at both 24 hpf and 48 hpf in the diencephalon and locus coeruleus. *dbh* is expressed at 24 hpf and 48 hpf in the locus coeruleus.

We found that while −6.1 remains inactive at 24 hpf (data not shown), it directs expression specific to the midbrain and sympathetic chain at 48 hpf (Figure 3a). Similarly, +3.3 also shows no GFP expression at 24 hpf (data not shown) but directs expression across multiple *ASCL1*-specific tissues at 48 hpf, including the diencephalon and hindbrain (Figure 3c). In contrast, the elements −1.4 and +13.3 direct very strong and broad CNS expression at 24 hpf becoming more restricted at 48 hpf (Figure 3b, d). Amplicon −1.4 directs expression in a large number of *ASCL1*-specific tissues, including very strong expression in the telencephalon, diencephalon, midbrain, hindbrain, and spinal cord (Figure 3b). However, at 48 hpf, the expression is tightly restricted to the telencephalon and locus coeruleus. The element +13.3 also shows very strong expression at 24 hpf in many tissues including the telencephalon, diencephalon, midbrain, and hindbrain. At 48 hpf the expression tapers significantly to only the epiphysis and hindbrain (Figure 3d). Collectively these data suggest that the identified enhancers comprise *ASCL1* regulatory elements with incompletely overlapping temporal and spatial control perhaps reflecting differing requirements during development.

## Discussion

Transgenesis in vertebrate organisms provides a robust system in which to evaluate putative noncoding cis-regulatory sequences. However, the pace at which these animals can be created and analyzed is readily outpaced by the *in silico* prediction of potential regulatory elements. We set out to test a new approach that we hoped would allow more rapid and comprehensive analysis of a single locus or many loci. We report a method for pooling constructs for injection and analysis in germline transmitted zebrafish embryos. Having previously demonstrated that mammalian regulatory sequences can be reliably assayed in zebrafish [1,23,25], we chose to apply this method on the human *ASCL1* locus.

We screened 82 potential transgenic founder zebrafish, identifying 32 that transmitted *ASCL1*-derived constructs and reporting GFP in their central nervous system in a manner consistent with the endogenous *ascl1a/b* orthologs. These enhancer sequences, in sum, recapitulate almost the entire *ASCL1* endogenous expression pattern in the telencephalon, diencephalon, midbrain, hindbrain, sympathetic chain, and spinal cord. Control of expression in the midbrain and hindbrain predominated, suggesting that development of these neuronal populations may require especially precise control of *ASCL1* expression consistent with the requirement for *ascl1* in these regions, particularly in noradrenergic neurons. We also found reporter expression in non-*ASCL1* specific tissues, including the heart, and the pronephric duct in a small number (<5/36) of transgenic lines (Supplemental Table 2). Some of these expression domains however may reflect additional roles in the regulation of neighboring genes like *PAH*, which lies 40 kb upstream of *ASCL1* is expressed in the pronephric duct of embryonic zebrafish or position effect of insertion [46].

In undertaking this study we tried to improve both the yield and rate of transgenic analysis in zebrafish. We compare the approaches and the time required at each step in Supplemental Table 1. The established mosaic strategies offer a rapid initial screen of regulatory control, available within days. The resulting data, however, is not comprehensive and requires the analysis of many representative embryos, thus significantly increasing documentation time in assembling a composite description of regulatory control (Supplemental Table 1). For a more comprehensive description of developmental regulatory control one must consider germline transmission-based analyses. By contrast, although not as immediate as the data generated in mosaic analyses of individual constructs (Supplemental Table 1), our approach does have several advantages, including: increased scaling potential, the ease of raising complex pools of embryos in common tanks for subsequent identification, and the non-mosaic nature of the resulting analyses. The improvement in efficiencies of scale are clear when one considers larger data sets; one may inject many more constructs and, somewhat like a mutagenesis screen, identify the molecular origins of only those displaying the phenotype (expression pattern) of interest. Using a PCR-based approach for the identification of discrete lines means that the injected embryos do not need to be raised in separate tanks, significantly reducing the numbers of tanks utilized for this purpose. Furthermore, moving away from analyses in mosaic embryos to analyze larger numbers of constructs in germ line transmitted embryos in our hands has proved to be less prone to subjective determination of the anatomical location of signal as compared to analysis in G0 mosaic embryos and also significantly reduces time spent screening for and analyzing reporter expression (Supplemental Table 1). Real comparisons between these approaches are difficult; one is trading near immediate data for a pipeline that produces more comprehensive data on a larger scale.

However, this study has also revealed an additional layer of complexity; significant numbers of zebrafish harbor two or more transgenes, occasionally complicating analysis. We believe that the robust nature of this study largely overcomes this issue though the generation of large numbers of transgene-positive founder zebrafish. We do, however, feel that in the light of the comparison of these strategies further improvement is readily attainable. One might imagine the following: First, the injection of single transgenes into 150–200 zebrafish embryos along with co-population of embryos from five or more other transgene injections for raising in single 5 liter tank would eliminate the issue of zebrafish containing multiple transgenes and reduce pressure on facility occupancy. Second, the co-population of tanks combined with the straightforward PCR/sequence-based identification of transgenes should facilitate efficient screens of selected loci to identify specific regulatory phenotypes of interest. Based on our estimates of a study of 50 constructs, the time taken in such an approach is the same as that for a similar sized pooling effort; it retains the analytical advantages of germline transmission and yet is complimented by the reduced complexity of single injection-based traditional approaches (Supplemental Table 1). One may then in theory simply screen through offspring for expression patterns of interest and sequence the contributing enhancers post-hoc. We believe that such an approach may prove to be particularly useful in the validation of large numbers of sequences identified through ChIP-Seq assays or the identification of biologically relevant regulatory sequences within intervals implicated in disease through human association studies. These and other advances will continue to expand the platform on which functional analyses of genomic datasets are predicated.

## Methods

### Selection and amplification of human noncoding sequence

The sequences studied were in the regions corresponding to chr12:101,869,373–101,932,015 in the human March 2006 (hg18) build. Using standard PCR conditions, the nine most highly constrained sequences as defined by PhastCons (Supplementary Table S1) were amplified off

of human genomic DNA and separately subcloned into the pT2GWcfosEGFP, a Tol2-based transgenic reporter construct [23, 26, 45]. We, and others, have previously shown this to be a reliable screen for enhancer activity [23, 45, 49].

### Fish Care

All zebrafish were raised, bred, and staged according to standard protocols at 28° C [50,51].

### Embryo injections and analysis

Putative regulatory elements subcloned into the pT2GWcfosEGFP reporter construct were injected into wild-type G0 AB zebrafish embryos [23,45]. Reporter expression directed by each construct was then evaluated in >1000 live G0 mosaic embryos at 24, 48, and 72 hpf. Approximately 250 fish were raised to sexual maturity. 65 males were outcrossed to AB females and the offspring were screened at 24, 48, and 72 hpf for reporter expression. Analysis of embryos was conducted using a Carl Zeiss Lumar V12 Stereo microscope with AxioVision version 4.6 software. Images captured were further cropped and levels adjusted in Adobe Photoshop CS4.

**In-situ Hybridization—**For the *ascl1a, th, and dbh* ISH, embryos were collected from matings of AB zebrafish at 24 and 48 hpf and fixed for ISH using standard protocols. For the GFP ISH, GFP positive embryos were collected from the matings of G0 males and AB females. The *ascl1a* riboprobe was generated by topo cloning sequence amplified with the forward primer ACGACTTGGTTGTTCATGC and the reverse pririmer GGATCCATTAACCCTCACTAAAGGGAATTGACTGCAACACGTAAAGC off zebrafish genomic into the vector pCRII-TOPO. The plasmids used to create the *th* and *dbh* riboprobes were created were acquired from, respectively, Zygogen and Steve Wilson.

### Genotyping

DNA was extracted from G1 embryos using standard protocols. PCRs were done off the genomic DNA using a forward primer designed off the pT2GWcfosEGFP backbone (CAATCCTGCAGTGCTGAAAA) while the 10 reverse primers are designed off the 10 unique sequences being analyzed (Supplemental Table 3). The primers were tested by their ability to amplify off a mixture of zebrafish genomic DNA and the appropriate transgene vectors diluted down to a concentration appropriate for a single insertion.

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

## Acknowledgments

We thank Samantha Maragh for *in-situ hybridization* imaging assistance, Takeshi Matsui for assistance with the outcrossings, and the Johns Hopkins Zebrafish Core ZIRC for fish care. This study was supported by funds from the NIH (NIGMS: R01GM071648 and NINDS: R01NS062972) to A.S.M. D.M.M. was also supported by NIH pre-doctoral training grant 5T32GM07814.

## References

1. Antonellis, A., et al. Identification of neural crest and glial enhancers at the mouse Sox10 locus through transgenesis in zebrafish; PLoS Genetics. 2008. p. e1000174%U http://www.ncbi.nlm.nih.gov/pubmed/18773071
2. Chandler, KJ.; Chandler, RL.; Mortlock, DP. Identification of an ancient Bmp4 mesoderm enhancer located 46 kb from the promoter; Developmental Biology. 2009. p. 590-602.%U http://www.ncbi.nlm.nih.gov/pubmed/19159624

3. Loots, GG. Genomic identification of regulatory elements by evolutionary sequence comparison and functional analysis; Advances in Genetics. 2008. p. 269-293.%U http://www.ncbi.nlm.nih.gov/pubmed/18282510
4. Prabhakar, S., et al. Accelerated evolution of conserved noncoding sequences in humans; Science (New York, NY). 2006. p. 786%U http://www.ncbi.nlm.nih.gov/pubmed/17082449
5. Prabhakar, S., et al. Human-specific gain of function in a developmental enhancer; Science (New York, NY). 2008. p. 1346-1350.%U http://www.ncbi.nlm.nih.gov/pubmed/18772437
6. Jantzen HM, et al. Cooperativity of glucocorticoid response elements located far upstream of the tyrosine aminotransferase gene. Cell 1987;49(1):29–38. [PubMed: 2881624]
7. Xiao H, Lis JT. Germline transformation used to define key features of heat-shock response elements. Science 1988;239(4844):1139–42. [PubMed: 3125608]
8. Cretekos, CJ., et al. Regulatory divergence modifies limb length between mammals; Genes & Development. 2008. p. 141-151.%U http://www.ncbi.nlm.nih.gov/pubmed/18198333
9. Davidson, EH.; Erwin, DH. Gene regulatory networks and the evolution of animal body plans; Science (New York, NY). 2006. p. 796-800.%U http://www.ncbi.nlm.nih.gov/pubmed/16469913
10. He, J.; Deem, MW. Hierarchical evolution of animal body plans. Developmental Biology. 2009. %U http://www.ncbi.nlm.nih.gov/pubmed/19799894
11. Levine, M.; Tjian, R. Transcription regulation and animal diversity; Nature. 2003. p. 147-151.%U http://www.ncbi.nlm.nih.gov/pubmed/12853946
12. Stathopoulos, A.; Levine, M. Genomic regulatory networks and animal development; Developmental Cell. 2005. p. 449-462.%U http://www.ncbi.nlm.nih.gov/pubmed/16198288
13. Emison, ES., et al. A common sex-dependent mutation in a RET enhancer underlies Hirschsprung disease risk; Nature. 2005. p. 857-863.%U http://www.ncbi.nlm.nih.gov/pubmed/15829955
14. Haigh, CL.; Wright, JA.; Brown, DR. Regulation of prion protein expression by noncoding regions of the Prnp gene; Journal of Molecular Biology. 2007. p. 915-927.%U http://www.ncbi.nlm.nih.gov/pubmed/17376480
15. Kleinjan, DA.; van Heyningen, V. Long-range control of gene expression: emerging mechanisms and disruption in disease; American Journal of Human Genetics. 2005. p. 8-32.%U http://www.ncbi.nlm.nih.gov/pubmed/15549674
16. Landry, J-R., et al. Expression of the leukemia oncogene Lmo2 is controlled by an array of tissue-specific elements dispersed over 100 kb and bound by Tal1/Lmo2, Ets, and Gata factors; Blood. 2009. p. 5783-5792.%U http://www.ncbi.nlm.nih.gov/pubmed/19171877
17. Lettice, LA., et al. A long-range Shh enhancer regulates expression in the developing limb and fin and is associated with preaxial polydactyly; Human Molecular Genetics. 2003. p. 1725-1735.%U http://www.ncbi.nlm.nih.gov/pubmed/12837695
18. Lodder, EM., et al. Implication of long-distance regulation of the HOXA cluster in a patient with postaxial polydactyly; Chromosome Research: An International Journal on the Molecular, Supramolecular and Evolutionary Aspects of Chromosome Biology. 2009. p. 737-744.%U http://www.ncbi.nlm.nih.gov/pubmed/19672683
19. Verlaan, DJ., et al. Allele-specific chromatin remodeling in the ZPBP2/GSDMB/ORMDL3 locus associated with the risk of asthma and autoimmune disease; American Journal of Human Genetics. 2009. p. 377-393.%U http://www.ncbi.nlm.nih.gov/pubmed/19732864
20. Visel, A.; Rubin, EM.; Pennacchio, LA. Genomic views of distant-acting enhancers; Nature. 2009. p. 199-205.%U http://dx.doi.org/10.1038/nature08451
21. Elgar G, Vavouri T. Tuning in to the signals: noncoding sequence conservation in vertebrate genomes. Trends Genet 2008;24(7):344–52. [PubMed: 18514361]
22. Allende, ML., et al. Cracking the genome's second code: enhancer detection by combined phylogenetic footprinting and transgenic fish and frog embryos; Methods (San Diego, Calif). 2006. p. 212-219.%U http://www.ncbi.nlm.nih.gov/pubmed/16806968
23. Fisher S, et al. Conservation of RET regulatory function from human to zebrafish without sequence similarity. Science 2006;312(5771):276–9. [PubMed: 16556802]
24. Li, Q., et al. A systematic approach to identify functional motifs within vertebrate developmental enhancers. Developmental Biology. 2009. %U http://www.ncbi.nlm.nih.gov/pubmed/19850031

25. McGaughey, D., et al. Asymmetrical distribution of non-conserved regulatory sequences at PHOX2B is reflected at the ENCODE loci and illuminates a possible genome-wide trend; BMC Genomics. 2009. p. 8%U http://www.biomedcentral.com/1471-2164/10/8
26. McGaughey, DM., et al. Metrics of sequence constraint overlook regulatory sequences in an exhaustive analysis at phox2b; Genome Research. 2008. p. 252-260.%U http://genome.cshlp.org/content/18/2/252.abstract
27. Yaragatti M, Basilico C, Dailey L. Identification of active transcriptional regulatory modules by the functional assay of DNA from nucleosome-free regions. Genome Res 2008;18(6):930–8. [PubMed: 18441229]
28. Visel A, et al. ChIP-seq accurately predicts tissue-specific activity of enhancers. Nature 2009;457 (7231):854–8. [PubMed: 19212405]
29. Kawakami K. Tol2: a versatile gene transfer vector in vertebrates. Genome Biol 2007;8(Suppl 1):S7. [PubMed: 18047699]
30. Korzh V. Transposons as tools for enhancer trap screens in vertebrates. Genome Biol 2007;8(Suppl 1):S8. [PubMed: 18047700]
31. Kawakami K, Shima A, Kawakami N. Identification of a functional transposase of the Tol2 element, an Ac-like element from the Japanese medaka fish, and its transposition in the zebrafish germ lineage. Proc Natl Acad Sci U S A 2000;97(21):11403–8. [PubMed: 11027340]
32. Allende, ML.; Weinberg, ES. The Expression Pattern of Two Zebrafish achaete-scute Homolog (ash) Genes Is Altered in the Embryonic Brain of the cyclops Mutant; Developmental Biology. 1994. p. 509-530.%U http://www.sciencedirect.com/science/article/B6WDG-45P0WXV-1F/2/7d5c10e394e2ba0a28199a8839c8ab3d
33. Blaugrund, E., et al. Distinct subpopulations of enteric neuronal progenitors defined by time of development, sympathoadrenal lineage markers and Mash-1-dependence; Development (Cambridge, England). 1996. p. 309-320.%U http://www.ncbi.nlm.nih.gov/pubmed/8565843
34. Guillemot, F., et al. Mammalian achaete-scute homolog 1 is required for the early development of olfactory and autonomic neurons; Cell. 1993. p. 463-476.%U http://www.sciencedirect.com/science/article/B6WSN-4C6BN4C-2C/2/adc805d65ed05a9b25f048199e44f096
35. Lo, LC., et al. Mammalian achaete-scute homolog 1 is transiently expressed by spatially restricted subsets of early neuroepithelial and neural crest cells; Genes & Development. 1991. p. 1524-1537. %U http://genesdev.cshlp.org/content/5/9/1524.abstract
36. Meredith, A.; Johnson, JE. Negative autoregulation of Mash1 expression in CNS development; Developmental Biology. 2000. p. 336-346.%U http://www.ncbi.nlm.nih.gov/pubmed/10837123
37. Pattyn, A., et al. Ascl1/Mash1 is required for the development of central serotonergic neurons; Nature Neuroscience. 2004. p. 589-595.%U http://www.ncbi.nlm.nih.gov/pubmed/15133515
38. Shida T, et al. Aberrant expression of human achaete-scute homologue gene 1 in the gastrointestinal neuroendocrine carcinomas. Clin Cancer Res 2005;11(2 Pt 1):450–8. [PubMed: 15701827]
39. de Pontual L, et al. Noradrenergic neuronal development is impaired by mutation of the proneural HASH-1 gene in congenital central hypoventilation syndrome (Ondine's curse). Hum Mol Genet 2003;12(23):3173–80. [PubMed: 14532329]
40. Ide M, et al. Genetic association analyses of PHOX2B and ASCL1 in neuropsychiatric disorders: evidence for association of ASCL1 with Parkinson's disease. Hum Genet 2005;117(6):520–7. [PubMed: 16021468]
41. Verma-Kurvari, S., et al. Lineage-Specific Regulation of the Neural Differentiation GeneMASH1; Developmental Biology. 1996. p. 605-617.%U http://www.sciencedirect.com/science/article/B6WDG-45MGVMT-R/2/5dfb5b143c77906de6004d01b4c5cb39
42. Verma-Kurvari, S., et al. Multiple elements regulate Mash1 expression in the developing CNS; Developmental Biology. 1998. p. 106-116.%U http://www.ncbi.nlm.nih.gov/pubmed/9578622
43. Siepel A, et al. Evolutionarily conserved elements in vertebrate, insect, worm, and yeast genomes. Genome Res 2005;15(8):1034–50. [PubMed: 16024819]

44. Kent WJ. BLAT--the BLAST-like alignment tool. Genome Res 2002;12(4):656–64. [PubMed: 11932250]
45. Fisher S, et al. Evaluating the biological relevance of putative enhancers using Tol2 transposon-mediated transgenesis in zebrafish. Nat Protoc 2006;1(3):1297–305. [PubMed: 17406414]
46. Thisse B, Pflumio S, Fürthauer M, Loppin B, Heyer V, Degrave A, Woehl R, Lux A, Steffan T, Charbonnier XQ, Thisse C. Expression of the zebrafish genome during embryogenesis. ZFIN Direct Data Submission. 2001
47. Nobrega, MA.; Pennacchio, LA. Comparative genomic analysis as a tool for biological discovery; The Journal of Physiology. 2004. p. 31-39.%U http://www.ncbi.nlm.nih.gov/pubmed/14678488
48. Pennacchio LA, et al. Predicting tissue-specific enhancers in the human genome. Genome Res 2007;17(2):201–11. [PubMed: 17210927]
49. Woolfe A, et al. Highly conserved non-coding sequences are associated with vertebrate development. PLoS Biol 2005;3(1):e7. [PubMed: 15630479]
50. Kimmel CB, et al. Stages of embryonic development of the zebrafish. Dev Dyn 1995;203(3):253–310. [PubMed: 8589427]
51. Westerfield, M. The Zebrafish Book: a Guide for the Laboratory Use of Zebrafish (Danio rerio). 4. Eugene, Oregeon: University of Oregon Press; 2000.

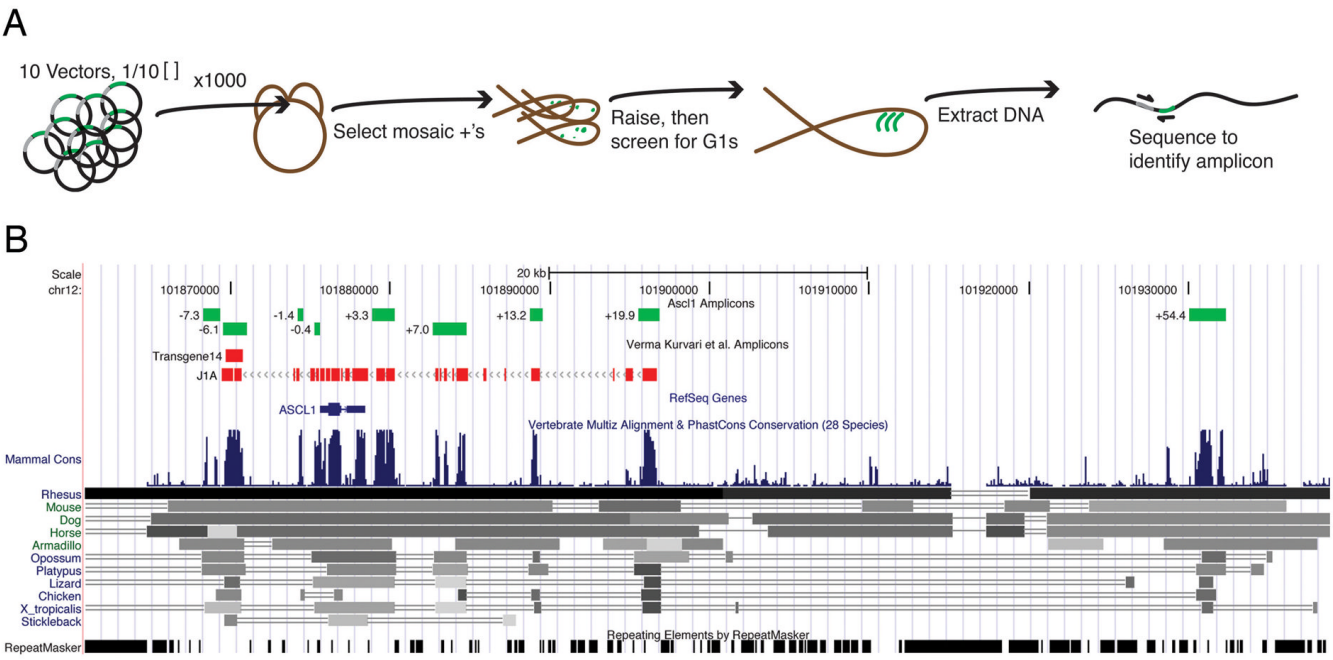


**Figure 1. Schematic of pooling strategy and selected amplicons of *ASCL1* locus**

A, Overview of pooling strategy. B, The UCSC Genome Browser (genome.ucsc.edu) custom track (hg18) of the nine selected highly conserved amplicons are shown in green while two of the Verma-Kurvari et al. [42] orthologues are displayed in red overlayed onto the *ASCL1* locus.

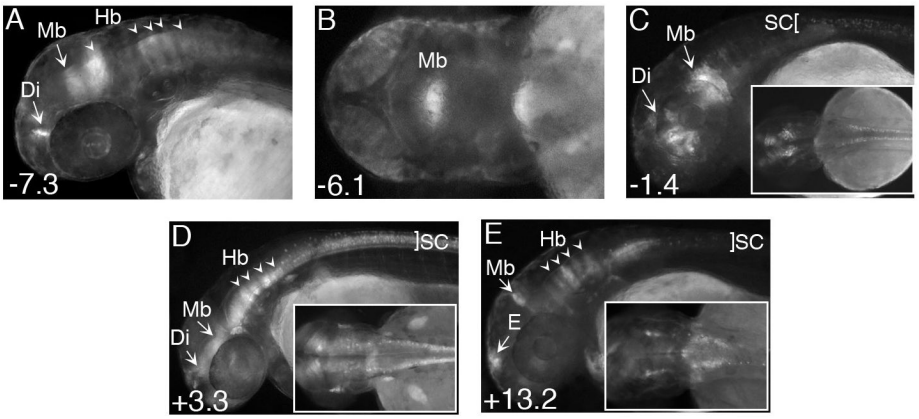


**Figure 2. Five *ASCL1* amplicons direct tissue specific expression in embryonic zebrafish**

Lateral and dorsal (inset) fluorescent images of GFP expression in 48 hpf zebrafish. All images oriented with the anterior to the left and posterior to the right. A, The −7.3 amplicon in zebrafish line 46 directs expression in the diencephalon, midbrain, and hindbrain. B, The −6.1 amplicon in zebrafish line 50 directs expression in the midbrain. C, The −1.4 amplicon in zebrafish line 39 directs expression in the diencephalon, midbrain, and spinal cord. D, The +3.3 amplicon in zebrafish line 10 directs expression in the diencephalon, midbrain, hindbrain, and spinal cord. E, The +13.2 amplicon in zebrafish line 19 directs expression in the epiphysis, midbrain, hindbrain, and spinal cord. Te, telencephalon; E, epiphysis; Di, diencephalon; Hb, hindbrain; arrowheads, rhombomeres of hindbrain; SC, spinal cord; [ marking SC expression region.

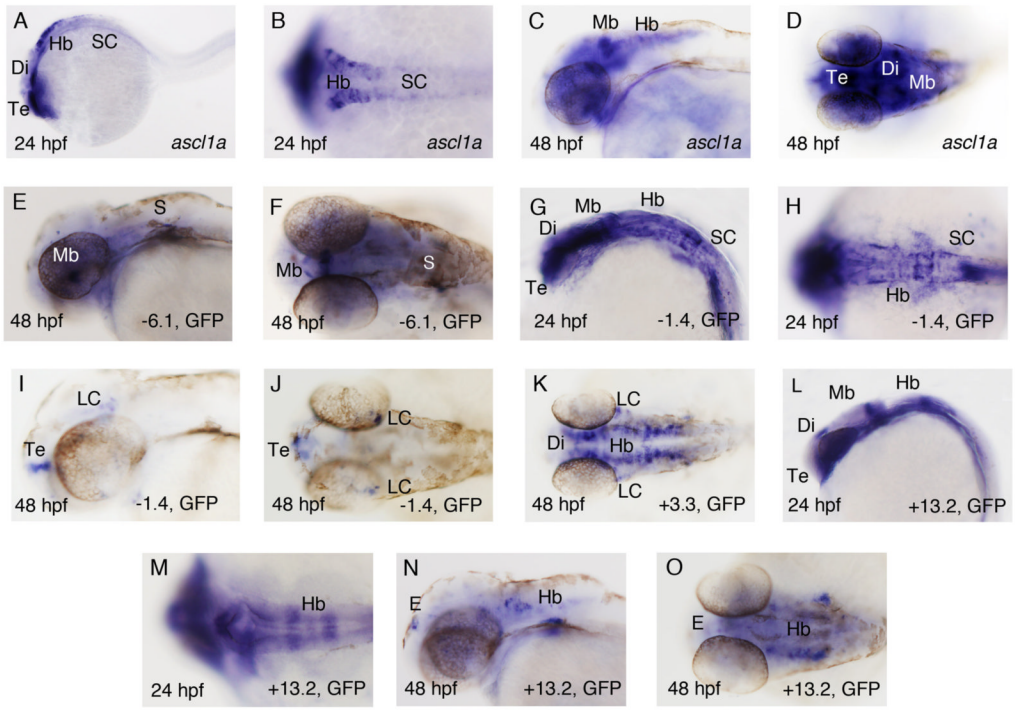


**Figure 3. *in situ* hybridization images of endogenous *ascl1a* and GFP expression**

Lateral and dorsal (inset) bright field images of *ascl1a* and GFP expression in 24 hpf and 48 hpf zebrafish. All images oriented with the anterior to the left and posterior to the right. A–D, 24 and 48 hpf lateral and dorsal images of endogenous *ascl1a* expression. E and F, 48 hpf images of GFP expression in −6.1 containing zebrafish line 50. G–J, 24 and 48 hpf images of GFP expression in −1.4 containing zebrafish line 39. K 48 hpf image of GFP expression in +3.3 containing zebrafish line 10. L–O, 24 and 48 hpf images of GFP expression in +13.2 containing zebrafish line 19. Te, telencephalon; Di, diencephalon; E, epiphysis; Mb, midbrain; Hb, hindbrain; S, sympathetic ganglia; SC, spinal cord

**Table 1**

Overview of expression patterns for identified founders

The ratio of founders containing a single amplicon displaying GFP expression in specific tissues over the total number of founders containing a single amplicon. In parentheses are the numbers of founders containing multiple amplicons (which also include the given amplicon) showing GFP expression specific tissues over the total number of founders containing multiple amplicons.

| Amplicon | Telencephalon | Diencephalon | Midbrain | Hindbrain | Spinal Cord |
|---|---|---|---|---|---|
| −7.3 | 0/1 (3/8) | 1/1 (2/8) | 1/1 (7/8) | 1/1 (7/8) | 0/1 (6/9) |
| −6.1 | 1/4 (4/7) | 1/4 (2/7) | 3/4 (6/7) | 2/4 (5/7) | 2/4 (3/7) |
| −1.4 | 0/2 (3/6) | 1/2 (1/6) | 1/2 (4/6) | 1/2 (5/6) | 2/2 (5/6) |
| −0.4 | | | | | |
| +3.3 | 0/4 (4/8) | 1/4 (3/8) | 4/4 (5/8) | 4/4 (8/8) | 4/4 (7/8) |
| +7.0 | (4/4) | (1/4) | (4/4) | (3/4) | (4/4) |
| +13.2 | 0/1 (6/9) | 0/1 (2/9) | 1/1 (9/9) | 1/1 (7/9) | 1/1 (5/9) |
| +19.9 | | | | | |
| +54.4 | (1/1) | (0/1) | (1/1) | (1/1) | (1/1) |
| −11.2 | 2/4 (6/9) | 0/4 (3/9) | 1/4 (6/9) | 4/4 (9/9) | 2/4 (6/9) |
| Sum* | 12/32 [38%] | 10/32 [31%] | 26/32 [81%] | 27/32 [84%] | 17/32 [53%] |

* includes all founders except those carrying only the −11.2 amplicon (fish 26, 56, 62, 79)