Genomics 95 (2010) 363–369

Contents lists available at ScienceDirect

Genomics

journal homepage: www.elsevier.com/locate/ygeno

Efficient discovery of *ASCL1* regulatory sequences through transgene pooling

David M. McGaughey [a], Andrew S. McCallion [a,b,*]

[a] McKusick - Nathans Institute of Genetic Medicine, Johns Hopkins University School of Medicine, 733 N. Broadway, BRB Suite 449, Baltimore, MD 21205, USA
[b] Department of Molecular and Comparative Pathobiology, Johns Hopkins University School of Medicine, Baltimore, MD 21205, USA

ARTICLE INFO

Article history:
Received 8 December 2009
Accepted 25 February 2010
Available online 4 March 2010

Keywords:
Regulation
Noncoding
Cis-regulatory
ASCL1
Development
Transgenesis
Zebrafish
Neural
Enhancers
Germ-line

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 [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-line allows a complete view of the tissue and temporal specific expression pattern directed by each regulatory sequence. Its application in large-scale studies however, has 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

* Corresponding author. McKusick - Nathans Institute of Genetic Medicine, Johns Hopkins University School of Medicine, 733 N. Broadway, BRB Suite 449, Baltimore, MD 21205, USA.
E-mail addresses: davidm@jhmi.edu (D.M. McGaughey), andy@jhmi.edu (A.S. McCallion).

0888-7543/$ – see front matter © 2010 Elsevier Inc. All rights reserved.
doi:10.1016/j.ygeno.2010.02.011

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 however, 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 raise selected 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 Fig. 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; Fig. 1B and Supplemental Table 2).

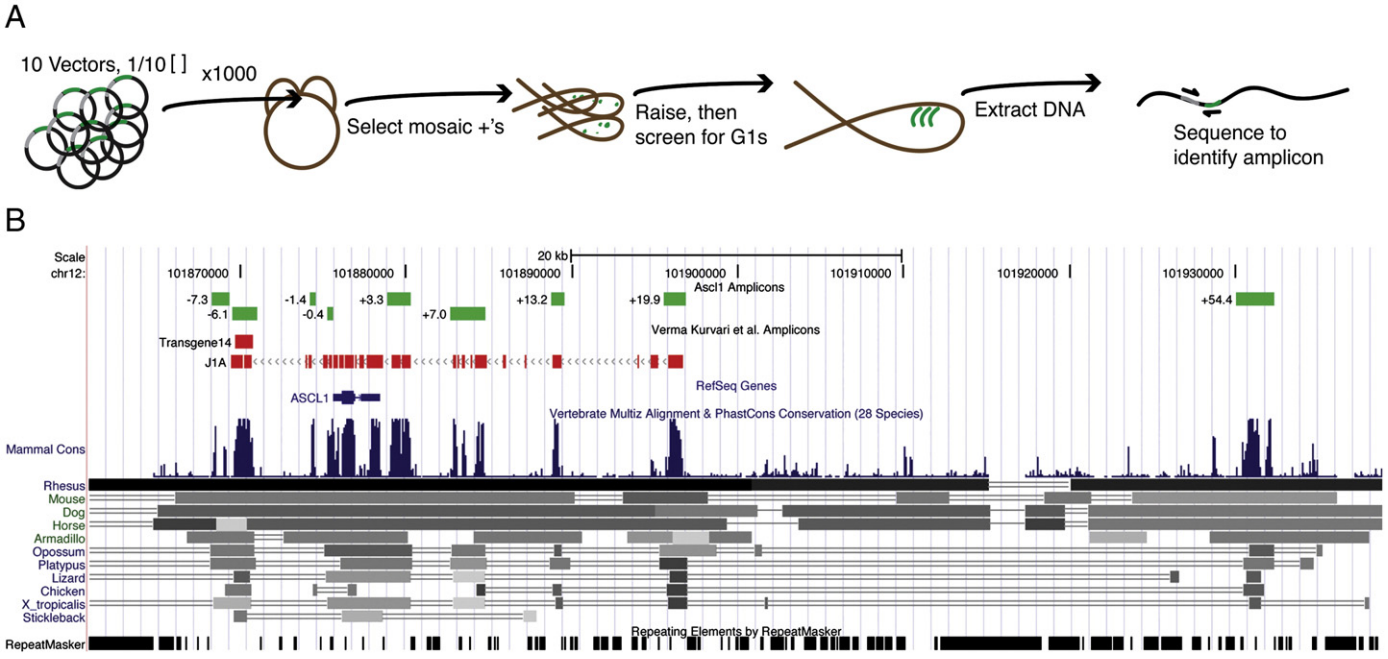

Fig. 1. Schematic of pooling strategy and selected amplicons of *ASCL1* locus. A, Overview of pooling strategy. B, The UCSC Genome Browser (genome.ucsc.edu) custom track (hg18) of the nine selected highly conserved amplicons are shown in green while two of the Verma-Kurvari et al. [42] orthologues are displayed in red overlayed onto the *ASCL1* locus.