Developmental Biology 350 (2011) 239–254

Contents lists available at ScienceDirect

Developmental Biology

journal homepage: www.elsevier.com/developmentalbiology

Review

When needles look like hay: How to find tissue-specific enhancers in model organism genomes

Maximilian Haeussler *, Jean-Stéphane Joly

U1126 MSNC INRA Group, UPR3294 NED, Institut Fessard, CNRS, 91 198 Gif-sur-Yvette, France

ARTICLE INFO

Article history:
Received for publication 14 April 2010
Revised 11 November 2010
Accepted 22 November 2010
Available online 3 December 2010

Keywords:
Cis-regulatory element
Fish
Zebrafish
Medaka
Transgenesis
Cis-regulation
Transcriptional regulation
Non-coding elements
Genome analysis
Enhancers

ABSTRACT

A major prerequisite for the investigation of tissue-specific processes is the identification of *cis*-regulatory elements. No generally applicable technique is available to distinguish them from any other type of genomic non-coding sequence. Therefore, researchers often have to identify these elements by elaborate *in vivo* screens, testing individual regions until the right one is found.

Here, based on many examples from the literature, we summarize how functional enhancers have been isolated from other elements in the genome and how they have been characterized in transgenic animals. Covering computational and experimental studies, we provide an overview of the global properties of *cis*-regulatory elements, like their specific interactions with promoters and target gene distances. We describe conserved non-coding elements (CNEs) and their internal structure, nucleotide composition, binding site clustering and overlap, with a special focus on developmental enhancers. Conflicting data and unresolved questions on the nature of these elements are highlighted. Our comprehensive overview of the experimental shortcuts that have been found in the different model organism communities and the new field of high-throughput assays should help during the preparation phase of a screen for enhancers. The review is accompanied by a list of general guidelines for such a project.

Introduction

Activating tissue-specific *cis*-regulatory elements – called "enhancers" (Banerji et al., 1981) – trigger gene expression in a given cell type, at the right developmental time and in the necessary quantity. They are tools of fundamental importance in diverse domains of biology. Cloned upstream of a fluorescent reporter gene, for example, they allow sorting of dissociated cells and tracking cell fate during embryogenesis with laser-scanning microscopes. They permit the analysis of essential genes by limiting the effect of functional assays to targeted cell populations: ectopic or over-expression of genes, knock-down with RNA interference or dominant-negative proteins or activation of Cre/Lox constructs can be performed in a tissue-specific manner. Finally, sequences of *cis*-regulatory elements can give clues about the trans-activating factor, helping to identify tissue-specific selector genes (Hobert, 2008).

In a more general sense, *cis*-regulatory elements also represent one big gap in our understanding of genomes, especially the huge non-coding parts: How much of the DNA is "junk" and what functions does the rest fulfill? What types of different functions are there? Which regions are implicated in human diseases (Kleinjan and Coutinho, 2009)? Although there are various types of *cis*-regulatory elements – reviewed by e.g. Arnosti (2003) and Maston et al. (2006) – robust assays are only available for enhancers. This is why the large-scale screens focus almost exclusively on these. Systemic tests have been conducted on one single locus at a time (Ishihara et al., 2008; Uchikawa et al., 2003) or on regions sampled from the whole genome (Woolfe et al., 2005). At the time of writing, the biggest project has screened around 1300 elements in thousands of mouse embryos (Pennacchio et al., 2006).

However, in many cases, researchers are interested in an element with a specific expression pattern . Given the size of the genome, relatively few regions have been tested already. Researchers are therefore often obliged to dissect the *cis*-regulatory landscape of a gene themselves and have to select a strategy on how to proceed. In the following, we provide guidelines for an enhancer screen targeting a single locus in a standard model organism. We summarize how various experimental improvements can be integrated in order to simplify the *in vivo* testing with transgenic model organisms. We describe some algorithms that predict the expression pattern from DNA sequences and point out their limits in the context of an enhancer screen. Finally, we highlight several topics deserving further investigation and comment on the importance of systematic *cis*-regulatory data collection.

* Corresponding author. Michael Smith Building, Faculty of Life Sciences, University of Manchester, Manchester M15 5RP, UK.
E-mail address: maximilianh@gmail.com (M. Haeussler).

0012-1606/$ – see front matter © 2010 Elsevier Inc. All rights reserved.
doi:10.1016/j.ydbio.2010.11.026

Main types of *cis*-regulatory elements and experimental testing

The high price of in vivo testing

To validate active individual regulatory elements, small DNA fragments (up to around 10 kbp) are cloned into plasmids one by one and tested for their activity with a reporter gene. As a result, elements active in tissues with available cell cultures are the ones best described in the literature. *In vivo*, however, there exists no experimental technique to screen large nucleotide sequences efficiently for their *cis*-regulatory potential at kilo base pair resolution. Complete testing of all randomly sheared fragments within a genomic locus is only feasible in simple model organisms such as ascidians or sea urchins (Keys et al., 2005; Cameron et al., 2004).

Nevertheless, protocols for other animals have been streamlined during the last years: observation of F0 embryos in mice is often sufficient (Loots, 2008). In zebrafish and nematodes, cloning can be avoided altogether by injecting PCR products (Woolfe et al., 2005; Hobert, 2002), although with an increase in mosaicism. In zebrafish, the number of assays can be reduced by testing genomic DNA from *Takifugu rubripes*, which is four times more compact while assumed to harbor similar regulatory elements (Barton et al., 2001). These experiments are still expensive in vertebrates, ranging between several hundred dollars per tested element in flies and fish to several thousand in mice (Table 1). For well-conserved elements, a time- and money-saving strategy might generate transgenic mice only with elements that have been tested first in fish. In many cases, conserved sequences yielded comparable expression patterns in both animals (Aparicio et al., 1995; Navratilova et al., 2009; Suster et al., 2009; Kimura-Yoshida et al., 2004).

If a given region does not fit into a plasmid, larger vectors, notably cosmids and BACs (Long and Miano, 2007), are more difficult to handle but can contain up to 40 kbp and 300 kbp of genomic data. Thanks to optimized protocols and better selectable markers, they can now be efficiently modified within 1 week (Sharan et al., 2009; Tursun et al., 2009; Smith, 2008; Venken et al., 2009; Ejsmont et al., 2009). Modifications of BACs (and more expensive knock-out mice) are the only way to remove elements from their context and find out if they are really necessary for a given expression pattern. Protocols and reagents are freely available from the National Cancer Institute at Frederick (NCICRF). Instead of screening individual DNA fragments to find the *cis*-regulatory element of interest, a BAC-clone with the gene replaced by a fluorescent protein coding sequence should often be sufficient to mark a cell type for subsequent analyses (Bouchard et al., 2005). Mouse lines for 800 BACs with an inserted GFP can be ordered through the GENSAT consortium (Geschwind, 2004). If large vectors are not an option, then individual regions in a locus have to be selected for testing.

Enhancer–promoter interactions

Where are enhancer elements found around a gene? In genes, a largely described location is in the introns, mainly in the first one. A handful of tissue-specific enhancers have also been described in coding regions. Several were described in 5′ untranslated regions (e.g. in the first exons of Pax6 (Zheng et al., 2001), IGF-1 (McLellan et al., 2006) and TH (Arányi et al., 2005)). Some have been recently discovered in translated coding exons (Hoxa2 (Tümpel et al., 2008; Lampe et al., 2008), Adamts5 (Barthel and Liu, 2008)). In addition, genome analyses found widespread non-coding selective pressure on coding regions (Woltering and Duboule, 2009; Chen and Blanchette, 2007; Kural et al., 2009) and exonic remnants after genome duplications and duplicated genes are known where all but one exon have disappeared (Dong et al., 2009b). This suggests that there might be more enhancers in transcribed and translated regions than is currently acknowledged but most are still expected to reside within the flanking non-coding regions around a gene or in introns within it.

The closest functional sequence here, directly upstream at around 50–100 bp, is the core or basal promoter (Juven-Gershon and Kadonaga, 2010). It used to be and still is often considered an essential but non-specific element of regulation, merely guiding the polymerase (Smale, 2001; Frith et al., 2008). Such a flexible structure with less sequence constraints might explain why the most conserved elements are located further upstream (Blanchette et al., 2006).Core promoters seem to be interchangeable between genes, as various studies in vertebrates have found a similar ratio of active enhancers although they used different core promoters (see Table 1).

But with more experimental data, the difference between the core promoter, the general “gateway to transcription” (Juven-Gershon et al., 2008), and tissue-specific elements has become less clear (summarized by e.g. Smale, 2001; Ohler and Wassarman, 2010). When assaying *cis*-regulatory sequences in invertebrates, not every enhancer could activate any promoter: for *Drosophila*, enhancers of gsb, gsbn, ant, bx require a certain type of promoter (DPE- or TATA-containing) (Li and Noll, 1994; Ohtsuki et al., 1998; Butler and Kadonaga, 2001). A mutation of the *yellow* or *oaf* promoters can change the interactions with enhancers (Lee and Wu, 2006; Merli et al., 1996). In *C. elegans*, a neural motif is not active when combined with non-neural promoters (Wenick and Hobert, 2004). Mammalian genome analyses found in roughly one third of the cases a relationship between the direct upstream sequence and the cell type where a gene is expressed (Smith et al., 2007; Roider et al., 2009; Vandenbon and Nakai, 2010). In cell cultures, the expression response to p53 depends on the type of basal promoter (Morachis et al., 2010) and specific transcription factors like E2F bind to a large proportion of all core promoter regions (Xu et al., 2007). In an extreme case, a tissue-specific element in sea urchin showed two different expression patterns, depending on the basal promoter it was combined with (Kobayashi et al., 2007).

The dependence on the basal promoter can lead to problems in medium-scale enhancer screens that test elements genome-wide, from various loci. In such experiments, non-coding fragments are combined with one standardized promoter, typically pHsp or pBeta-globin. In *Drosophila*, the possible incompatibilities motivated the development of an artificial *Super Core Promoter*, a mix of several different sequences with the goal of increased enhancer compatibility and high expression levels (Juven-Gershon et al., 2006). For some mammalian cell lines, optimized sequences have been synthesized that perform better than the CMV minimal promoter (Schlabach et al., 2010). In zebrafish, (Gehrig et al., 2009) analyzed almost all

Table 1

Organism	Delivery	Time from experiment to observation	Price transgenesis, academic rate	Source
D. melanogaster	injection	1–2 weeks	$250	the best Gene.com
C. elegans	injection	1–2 days	$150–$250	C. elegans Core, NTHU, Taiwan
C. intestinalis	electroporation	1 day	No core	
Zebrafish	injection	1–2 days	$450	Amagen Platform, CNRS, France
Chicken	electroporation (not all cells)	1 day	No core	
Mouse	injection	7–13 days	$2200	Ohio State Univ., Mouse Core