PostDoc Journal
Vol. 1, No. 10, October 2013

Journal of Postdoctoral Research
www.postdocjournal.com

Identifying developmental enhancers: changing perspectives with advancing techniques

Patricia Gongal
Institut de Biologie de l'Ecole Normale Supérieure
46 rue d'Ulm, 75005, Paris, France
Current address: The English Edition, Aberystwyth, UK (www.theenglishedition.com)
Email: pgongal@theenglishedition.com

Abstract
Establishing precise, dynamic patterns of gene expression, critical during embryonic development, is under the control of non-coding regulatory DNA sequences. Identifying these regulatory elements, including enhancers, for individual genes of interest remains difficult. Recent genomic studies based on new techniques have provided major insights into the properties and functions of enhancers, highlighting new approaches for their identification. These studies have yielded a deeper understanding of transcription factor binding, chromatin structure, and chromatin dynamics in developmental contexts, facilitating the search for enhancers that regulate individual genes in the embryo. After identifying putative enhancer sequences, however, confirming their target genes in vivo is critical, and can be done by combining multiple experimental approaches. This review examines how recent work has clarified the most effective approaches for identifying the enhancers that help drive development, how to match them to target genes, and confirm their role in the embryo.

Introduction

Precise spatial and temporal regulation of gene expression is critical for the development and function of all organisms. Inappropriate, absent, or mistimed expression can have numerous harmful developmental effects: abnormal distribution of cell types, abnormal organ sizes, abnormal tissue organization, or cell death. The ability of noncoding DNA to regulate transcription was first identified over thirty years ago, in experiments transfecting viral sequences into cells (Banerji et al., 1981; Moreau et al., 1981). Since that time, the properties and functions of non-coding regulatory elements have been intensely investigated. The best-studied type of regulatory element, the enhancer, plays a critical role in development and disease (Borok et al., 2010; Lee and Young, 2013; Sakabe et al., 2012; Visel et al., 2009a).

Over the last decade, our understanding of enhancer function during development has increased enormously. There have been a series of major advances on how correct gene expression in space and time is achieved and on the underlying mechanics. Transcription factors, non-coding DNA, and non-coding RNAs together drive precise and robust transcriptional regulation: activating, repressing, and fine-tuning tissue-specific expression. Many of the molecular aspects of these processes have been detailed, including sequence-specific transcription factor binding and cooperation (Spitz and Furlong, 2012), marking of chromatin by specific histone modifications (Bonn et al., 2012; Rada-Iglesias et al., 2012), recruitment of chromatin remodelling or co-activator complexes (Ansari and Morse, 2013; Sudarsanam and Winston, 2000), looping of DNA to bring the promoter near an enhancer (Sanyal et al., 2012), and transcription and role of non-coding RNAs (eRNAs or lncRNAs; Faust et al., 2012; Li et al., 2013), together culminating in the activation of RNA polymerase or recruitment of RNA polymerase elongation factors (Fuda et al., 2009). This increased understanding of gene regulation has been made possible by impressive technical advances. Modern genomic and bioinformatic techniques, including chromatin immunoprecipitation combined with microarray analysis (Zeitlinger et al., 2007) or high-throughput sequencing (Visel et al., 2009b), and chromatin conformation capture (and all its high-throughput variants; Andrey et al., 2013; Noordermeer et al., 2011), have revealed a great deal about the nature of gene regulation.

Despite these advances, for "small-scale" biologists working in model systems who are interested in the regulation of individual genes, it remains difficult to identify tissue-specific regulatory sequences and to verify an endogenous function. Multiple approaches have been developed to identify enhancers based on our understanding of their features and functions, rooted in the vast datasets now available. These have included identification of conserved non-coding sequences, finding clusters of transcription factor binding sites or events, and locating features of chromatin structure that are characteristic of enhancers. Many putative regulatory elements have been identified in this way, and many have been shown to be capable of driving activity in transgenic assays. However, few studies have gone beyond an initial validation, to demonstrate an endogenous role for putative elements. Demonstrating a functional role for identified sequences can be difficult, yet is critical for identifying the elements that are truly biologically important. This review will examine how recent research is providing new information that will optimize the success of enhancer identification, and outline the current challenges in the field.

Methods to identify enhancers

Conservation of non-coding sequences

Based on the idea that enhancer elements are conserved over evolution (Boffelli et al., 2003; Pennacchio et al., 2006), to search for a particular gene's enhancers, many laboratories have taken the straightforward approach of examining regional sequence conservation. Many studies have begun with sequence alignments, followed by manual selection of highly conserved regions around a gene of interest, cloning into a reporter-containing vector, and testing by transgenesis. Although it is intuitive that sequence conservation should be a ubiquitous feature of enhancers (particularly when the genes in question are highly conserved), recent work suggests that this is not necessarily the case.

Studies of transcription factor binding across species have shown that the majority of transcription factor-DNA binding events are species-specific, and only rarely do binding events on putative regulatory elements align between phylogenies (Schmidt et al., 2010). Further, sequence conservation of enhancers seems to strongly differ between tissues. For example, there is very little conservation of cardiac-specific enhancers, but higher rates for those specific for the nervous system (Blow et al., 2010; Pennacchio et al., 2006; Visel et al., 2008). That enhancers are functionally conserved in the absence of sequence conservation highlights the limits of our current understanding of how enhancers actually activate spatially- and temporally-specific transcription. A focus on conservation will therefore miss enhancers with conserved function but not sequence.

Further, researchers studying conserved sequence blocks tend to focus on the region surrounding the coding sequence. However, enhancers can act over a large range of distances, so this approach necessarily excludes distant elements that may be nevertheless biologically crucial. One of the first long-distance enhancers identified was that which regulates the expression of *Sonic hedgehog*, at a distance of 1MB (Lettice et al., 2003). At even greater distances, several studies have identified regulatory interactions in *trans*, between genes and noncoding elements located on different chromosomes (Lomvardas et al., 2006; Ronshaugen and Levine, 2004; Spilianakis et al., 2005). These studies reflect that important, very long distance interactions occur, but whether these represent a significant proportion of all regulatory elements remains debated.

Recent studies have given some clues to the prevalence of long-distance regulatory events. Although it is common in large-scale enhancer identification studies to assign putative regulatory elements to their nearest ORF, recent studies suggest this approach may be misguided, and that long-distance interactions are the rule rather than the exception. Assuming that physical interactions between regulatory elements and transcription start sites reflect endogenous regulatory activity, a large-scale analysis of these interactions (Sanyal et al., 2012) has shown that distal elements interact with the nearest transcription start site

rather infrequently. Further, less than half of distal enhancers interact with the nearest expressed open reading frame at all. Intriguingly, the same study detected a bias towards transcription start sites interacting with regions 120 kilobases upstream. The significance of this particular zone remains to be determined. Overall, it is common that long distances separate enhancers from their target genes, spanning other genes and functional DNA elements. Focusing on a gene's immediate region in an enhancer search may therefore exclude key sequences.

When searching for conserved non-coding sequences, it is nevertheless necessary to focus on a reasonably-sized subset of the genome. Different methods have suggested different sizes for a window around a gene of interest that is most likely to contain the enhancers. Hwang and colleagues, in their study of physical interactions involving promoters, histone modifications, and protein-binding data, argue for an effective radius around a gene of 1MB (Hwang et al., 2013). In contrast, bioinformatic methods based on sequence duplication and paralogy suggest that half of all regulatory elements fall within 250kb of genes (Vavouri et al., 2006). In either case, these distances are very large, making a comprehensive search and screen for an individual gene's regulatory elements difficult.

Together, the individually-identified enhancers acting at long distances, statistical approaches, and physical interaction studies suggest that location of a gene is not likely to be very informative for the identification of regulatory elements, and the efficiency of targeting a gene's immediate vicinity in a search for enhancer elements is doubtful. Therefore, when working on a single gene, particularly in cell types such as the heart, relying on sequence conservation in the immediate region for enhancer identification may not be the most effective approach.

In silico clustering of transcription factor binding sites

Because groups of transcription factors regulate genes, often in modules, one hypothesis has been that regions with clusters of transcription factor binding sites likely represent regulatory elements. Many in silico detection methods have been developed based on this idea, and have identified clusters of putative binding sites for selected transcription factors, which may represent functional regulatory sequences (reviewed in Hardison and Taylor, 2012; Van Loo and Marynen, 2009).

Methods of detecting transcription factor binding site clustering are based on position weight matrices of selected transcription factors. These thus depend on the reliability of databases of transcription factor binding, which tend to rely on binding data from in vitro (ex. gel shift assays, protein microarrays) rather than in vivo (ex. chromatin immunoprecipitation, ChIP) experiments. However, the frequencies of in vivo binding are much lower than in vitro predictions (Hardison and Taylor, 2012), possibly due to the degeneracy, flexibility, and cooperative nature of some binding events. This disconnect between in vitro and in vivo binding predictions/data may limit the accuracy of enhancer predictions by in silico clustering detection methods, and may explain why they vary widely in sensitivity and false positive rates (Hardison and Taylor, 2012; Van Loo and Marynen, 2009).

Papers that report binding site clustering usually include at least one method of in vivo validation, that is, the method has successfully predicted a genomic region with enhancer activity. However, despite this success, these methods have not been widely adopted by the developmental biology community. This may be partly due to a communication issue: bioinformatic methods are not often published in accessible, user-friendly formats which developmental biologists can easily understand and adopt for their system or gene of interest. Increased collaboration between these disciplines could have a major impact on success in identifying developmental regulatory elements.

In vivo transcription factor binding events

That enhancers are bound by specific transcription factors also underpins identification methods based not on predicted binding sites, but in vivo binding data. The first major study using this

approach in the embryo performed ChIP-chip at multiple time points for five different transcription factors required for mesoderm development in Drosophila (Zinzen et al., 2009). Putative regulatory modules were defined based on this comprehensive body of binding data, and a subset were tested in transgenic assays. Essentially all tested elements (97%) were sufficient to drive transcription in transgenic assays, suggesting that this method of identifying bona fide enhancers is highly effective.

Binding information from just a single transcription factor is informative for enhancer prediction, although subsequent studies in different systems have shown variable success rates. For example, Jin and colleagues report a >70% prediction success rate with genome-wide Tinman binding data in the Drosophila mesoderm (Jin et al., 2013), while Corbo and colleagues report around a success rate of about 50% based on CRX (cone-rod homeobox) in mouse photoreceptors (Corbo et al., 2010). Despite this variability, these success rates are still substantially greater than those seen with in silico methods.

Because of its high success rate, transcription factor occupancy is likely to be the first-choice method to identify regulatory elements across the genome in the future. Nevertheless, once regulatory regions are identified, matching them with their target genes is still a major challenge. These studies identify putative regulatory elements by the thousands, and as discussed above, assigning them a gene based on proximity is problematic.

To find an individual gene's regulatory elements based on this approach, ChIP-seq or ChIP-chip could be performed for known regulators of the gene, but this will only be successful if those regulators are direct. Another limitation is that effective antibodies for at least one (direct) regulator are required. While overexpressing tagged proteins followed by ChIP using antibodies for the tag (most commonly HA) can circumvent this problem, in embryos, this requires the additional process of establishing a stable transgenic line with consistent and homogenous expression, and runs the risk of increasing the false-positive rate of identified binding sites.

An additional issue relevant to small-scale biologists is that it is unclear whether whole-embryo ChIP can detect transcription factor binding to enhancers only active in a very small population of cells. Although predicted enhancers in one cell type are, in some cases, informative for others (Wang et al., 2012), this may reflect transcription factor binding to sites regardless of activity (Li et al., 2008). For some genes, identifying a tissue-specific enhancer may require transcription factor binding profiles from embryonic cells in which the enhancer is active, a formidable technical challenge for genes expressed in a small number of cells.

Chromatin structure

ChIP has been used not only to understand the genome-wide binding profiles of specific transcription factors, but also the behaviour of RNA polymerase and chromatin structural proteins, including specific histone variants and histones bearing post-translational modifications. The majority of studies have been completed using cell lines, where distinct chromatin signatures correlating with functional or active enhancers have been identified. (Heintzman et al., 2007, 2009; Rada-Iglesias et al., 2011, 2012)

Overall, enhancers correlate with a number of distinct histone variants, namely H3.3 and H2A.Z (Goldberg et al., 2010; Jin and Felsenfeld, 2007). In addition, the co-enrichment of the post-translational modifications H3K4me1 and H3K27ac strongly correlate with active enhancers (reviewed in Calo and Wysocka, 2013), although this is not the case for all cell types (Wang et al., 2008). Essentially all chromatin marks described on enhancers except H3K4me1 vary according to whether the enhancer is active (reviewed in Calo and Wysocka, 2013). This makes it more difficult to predict enhancers for a particular gene: in order to use deposited ChIP-seq data to find a particular gene's enhancer, selecting regions showing the correct histone marks requires knowledge of whether the enhancer is active in the cell line or system used. Expression of the target gene

is not necessarily reflective of a tissue-specific enhancer's activity: even if a gene is expressed in a given cell line, genes typically have multiple enhancers, and therefore the one of interest may not be active.

Whether the chromatin dynamics of enhancers in cell cultures reflect mechanisms in developing embryos was an open question until very recently. There has been an increasing number of high-throughput chromatin structure studies in model organisms during development (for example, Bonn et al., 2012), despite the inherent technical challenges. These challenges are multi-fold. First, a sufficient amount of starting material is more difficult to obtain in the form of embryos than cultured cells. Particularly for vertebrates, obtaining a large pool of embryos synchronized at a particular stage is difficult. Examining the temporal dynamics of chromatin structure as enhancers are turned on and off requires even more starting material, as chromatin samples from multiple developmental stages are necessary. Further, embryos represent mixed cell populations with highly dynamic chromatin. To understand the differences in enhancer behaviour between cell lineages, separating cell types is necessary, but difficult. Several laboratories have developed ways to overcome these challenges, either using cell systems that closely mimic the progression of a particular cell lineage (Wamstad et al., 2012), or in whole embryos with innovative cell-sorting techniques (Bonn et al., 2012; Deal and Henikoff, 2011).

It is a major challenge to separate distinct cell types from embryos in sufficient quantities for genomic studies. Two solutions to this issue have been developed: batch-isolation of tissue-specific chromatin (BiTS; Bonn et al., 2012), and isolation of nuclei tagged in specific cell types (INTACT; Deal and Henikoff, 2011). Both techniques are based on the principle of in vivo labelling of nuclei in specific tissues, and then sorting them into pure populations. While the INTACT procedure labels nuclei with biotin, which can then be pulled down with affinity purification, BiTS uses fluorescent labelling and FACS sorting. Both of these techniques applied to embryos have much promise for understanding the differences between activated and non-activated enhancers in different cell types.

Using BiTS to sort mesodermal and non-mesodermal lineages, and performing ChIP-seq on the two populations of nuclei, Bonn and colleagues have shown that many (but not all) properties of enhancer chromatin discovered in cultured cells are applicable to the embryo (Bonn et al., 2012). In both cell cultures and Drosophila embryos, H3K4me1 marks most enhancers, regardless of whether they are activated (it is present on 77% of mesodermal enhancers). This mark, however, is not exclusive to enhancers, as it is also present on promoters. Other modifications correlate with enhancer activity. Notably, in contrast to ES cells in a differentiation model of neural crest (Rada-Iglesias et al., 2012), H3K27me3 is correlated with inactive, not poised, enhancers in the Drosophila mesoderm. Other studies also support a diversity in enhancer marks in different systems: in differentiating cardiomyocytes, there is little genomic overlap between the distribution of H3K27me3 and enhancers defined by H3K4me1 or H3K27ac (Wamstad et al., 2012), suggesting H3K27me3 does not correlate with enhancers (nor their functional state) in this cell type. Clarifying these cell type- or species-specific differences will be essential for it to be possible to identify a developmental enhancer of a gene of interest from chromatin features.

Nevertheless, these studies are a strong beginning to understand the properties of enhancers, and thus improve our ability to identify them. In the Drosophila mesoderm, H3K27ac and H3K79me3 are temporally correlated with enhancer activation and show changes in distribution (rather than simply appearing or disappearing) as the activity of an enhancer changes. The binding of RNA Pol II is also highly correlated with the timing of enhancer activation. Pol II binding not only predicts enhancer activity, but it is strongly correlated with specific transcription factor binding, and has been suggested to be the most specific mark to precisely localize a regulatory element (Bonn et al., 2012). For individual laboratories to be able to more easily utilize these vast and extremely

useful datasets for the identification of elements of interest, the data should be integrated with existing genome browsers.

Combined approaches

These approaches need not be considered mutually exclusively, and the combination of transcription factor motifs, transcription factor binding data, and chromatin signatures may be very effective for enhancer prediction. For example, combining histone modification and multiple transcription factor occupancy has been used to predict novel enhancers in mouse ES cells, although these were not tested in vivo (Chen et al., 2012). Further, Rada-Iglesias and colleagues have demonstrated that comparing the positions of transcription factor binding motifs with chromatin signatures is a productive approach for enhancer prediction in the neural crest (Rada-Iglesias et al., 2012). Lastly, Wilczynski and collegues have reported higher success rates for enhancer predictions when incorporating data on transcription factor binding and chromatin states (Wilczynski et al., 2012). As information on the properties and functions of enhancers becomes more detailed, these integrative approaches will help narrow candidate regulatory regions, making it easier to match enhancers with their target genes.

Validating the activity of predicted enhancers

Transgenesis

A key step in enhancer identification is confirming that a putative element plays an in vivo role. The most common method of validation is transgenic analysis: fusing a putative element to a reporter protein (usually GFP) to confirm that the sequence can drive activity. This technique clearly demonstrates whether an element is sufficient to drive expression in a particular pattern. However, a negative result should not necessarily be interpreted as the sequence being unimportant. Other explanations could be that the sequence acts in conjunction with other sequences, or cannot function with the standard promoter sequence included in the expression construct.

Conversely, a positive result in a transgenic assay does not necessarily mean that the sequence plays an endogenous role in the regulation of a gene of interest. It may regulate another gene (or genes) with an overlapping expression pattern, or drive expression that it simply does not in vivo. Recent work by White and colleagues (White et al., 2013) highlights that sequences unbound by transcription factors in vivo nevertheless often have the ability to strongly drive transcription. Spurious results may be particularly likely in the case of auto-regulatory elements. These elements are expected to include a series of binding sites for their target gene, sequences which are likely to be able to drive tissue-specific transcription in transgenic assays (ie. in the tissues where the gene is expressed), regardless of whether they play an endogenous role. Particular caution is therefore warranted when interpreting positive results for putative autoregulatory elements in transgenic assays.

Evidence for enhancer-gene matches: shared regulators, physical interactions, and effect of enhancer knockout on endogenous expression

After demonstrating that a sequence can drive tissue-specific expression that correlates with that of a gene of interest, confirming that the enhancer actually regulates that gene is not straightforward. First, there must be an overlap between a gene's regulators and an enhancer's direct regulators. This requires confirming that putative transcription factor binding sites on the enhancer are genuine (if identified in silico, for example), and that loss of these sites and the relevant transcription factor affect enhancer function. Those transcription factors should affect the putative target gene with similar spatial and temporal dynamics.

Additional strong evidence of a enhancer-gene match is that they physically interact (tested by chromatin conformation capture or fluorescent in situ hybridization). Evidence is growing that physical interactions are a key feature of enhancer function (Montavon and Duboule, 2012), and a physical interaction, particularly if this is dependent on specific transcription factors (Palstra, 2012), along with shared regulators and

a correlating expression pattern constitutes solid evidence than an enhancer regulates a particular gene.

Further, mutations or deletion of a regulatory sequence should have an impact on endogenous gene expression. Enhancer knockout is perhaps the most convincing evidence that a non-coding DNA sequence actually regulates a gene. However, as this is technically difficult and/or expensive in vertebrate systems, enhancer knockouts have been only rarely reported. In addition, an enhancer knockout experiment assumes that losing a single enhancer will have significant and detectable impact on expression. After an enhancer is mutated, loss of endogenous gene expression is evidence that it is a bona fide enhancer, but unchanged or only subtly changed gene expression can be more difficult to interpret. It is now becoming apparent, however, that it is common for genes to have multiple enhancers that overlap in space and time (Perry et al., 2011). Knockout of one of a group of semi-redundant or redundant enhancers may thus cause very subtle or no effects on gene expression under standard laboratory conditions. A clear example of this is the pair of shadow enhancers described for the Drosophila gene snail. Perry and colleagues (Perry et al., 2011) introduced a BAC transgene containing the snail locus in a snail mutant background, permitting its detailed functional study. Mutation of a single enhancer has essentially no effect on snail expression under optimal laboratory conditions. Under stress conditions, however, both enhancers are necessary for invariant, reliable, and precise gene expression. Unidentified redundant elements may therefore pose a major problem for validation of enhancer-gene matches.

Future studies may require large-scale, comprehensive approaches to overcome the challenges of enhancer validation, for example, combining techniques such as chromatin conformation, chromatin structure, specific transcription factor binding, and mutagenesis. Together, these methods can not only identify enhancers, but bring remarkable new insights into gene regulation. For example, for the Hoxd cluster in mice, the combination of chromatin conformation studies and enhancer knockouts, complemented by existing chromatin structure data, has demonstrated that semi-redundant enhancers and dynamic chromatin conformation underpins digit patterning and development (Montavon et al., 2011).

Conclusion

With technical developments, our understanding of how enhancers function has greatly advanced. It is now much more clear, on a genomic scale, how specific transcription factors bind DNA, how chromatin structure is linked to gene expression, and how chromatin conformation and looping are related to transcriptional activation. These advances are clarifying the most effective ways to identify bona fide regulatory elements and match them to their target genes, to ultimately understand how the embryo establishes precise, robust, spatial and temporal patterns of gene expression. Advances in the field of developmental gene regulation will significantly increase our understanding of not only the normal development of cell types, tissues, and organs, but how these processes can go wrong. As many mutations associated with non-coding regions are linked to birth defects and disease (Kleinjan and van Heyningen, 2005), this work will have significant implications for multiple aspects of human health.

Acknowledgements

The author's postdoctoral research was supported by the Pierre-Gilles de Gennes Foundation, the Fondation de Recherche Medicale (FRM), and the Natural Sciences and Engineering Research Council of Canada (NSERC).

References

1. Andrey, G., Montavon, T., Mascrez, B., Gonzalez, F., Noordermeer, D., Leleu, M., Trono, D., Spitz, F., and Duboule, D. (2013). A Switch Between Topological Domains Underlies HoxD Genes Collinearity in Mouse Limbs. Science 340.

2. Ansari, S.A., and Morse, R.H. (2013). Mechanisms of Mediator complex action in transcriptional activation. Cell. Mol. Life Sci.

3. Banerji, J., Rusconi, S., and Schaffner, W. (1981).

Expression of a β-globin gene is enhanced by remote SV40 DNA sequences. Cell 27, 299–308.

4. Blow, M.J., McCulley, D.J., Li, Z., Zhang, T., Akiyama, J.A., Holt, A., Plajzer-Frick, I., Shoukry, M., Wright, C., Chen, F., et al. (2010). ChIP-Seq identification of weakly conserved heart enhancers. Nat. Genet. 42, 806–810.

5. Boffelli, D., McAuliffe, J., Ovcharenko, D., Lewis, K.D., Ovcharenko, I., Pachter, L., and Rubin, E.M. (2003). Phylogenetic Shadowing of Primate Sequences to Find Functional Regions of the Human Genome. Science 299, 1391–1394.

6. Bonn, S., Zinzen, R.P., Girardot, C., Gustafson, E.H., Perez-Gonzalez, A., Delhomme, N., Ghavi-Helm, Y., Wilczyński, B., Riddell, A., and Furlong, E.E.M. (2012). Tissue-specific analysis of chromatin state identifies temporal signatures of enhancer activity during embryonic development. Nat. Genet. 44, 148–156.

7. Borok, M.J., Tran, D.A., Ho, M.C.W., and Drewell, R.A. (2010). Dissecting the regulatory switches of development: lessons from enhancer evolution in Drosophila. Development 137, 5–13.

8. Calo, E., and Wysocka, J. (2013). Modification of Enhancer Chromatin: What, How, and Why? Mol. Cell 49, 825–837.

9. Chen, C., Morris, Q., and Mitchell, J.A. (2012). Enhancer identification in mouse embryonic stem cells using integrative modeling of chromatin and genomic features. Bmc Genomics 13, 152.

10. Corbo, J.C., Lawrence, K.A., Karlstetter, M., Myers, C.A., Abdelaziz, M., Dirkes, W., Weigelt, K., Seifert, M., Benes, V., Fritsche, L.G., et al. (2010). CRX ChIP-seq reveals the cis-regulatory architecture of mouse photoreceptors. Genome Res. 20, 1512–1525.

11. Deal, R.B., and Henikoff, S. (2011). The INTACT method for cell type-specific gene expression and chromatin profiling in Arabidopsis thaliana. Nat. Protoc. 6, 56–68.

12. Faust, T., Frankel, A., and D'Orso, I. (2012). Transcription control by long non-coding RNAs. Transcription 3, 78–86.

13. Fuda, N.J., Ardehali, M.B., and Lis, J.T. (2009). Defining mechanisms that regulate RNA polymerase II transcription in vivo. Nature 461, 186–192.

14. Goldberg, A.D., Banaszynski, L.A., Noh, K.-M., Lewis, P.W., Elsaesser, S.J., Stadler, S., Dewell, S., Law, M., Guo, X., Li, X., et al. (2010). Distinct factors control histone variant H3.3 localization at specific genomic regions. Cell 140, 678–691.

15. Hardison, R.C., and Taylor, J. (2012). Genomic approaches towards finding cis-regulatory modules in animals. Nat. Rev. Genet. 13, 469–483.

16. Heintzman, N.D., Stuart, R.K., Hon, G., Fu, Y., Ching, C.W., Hawkins, R.D., Barrera, L.O., Calcar, S.V., Qu, C., Ching, K.A., et al. (2007). Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. Nat. Genet. 39, 311–318.

17. Heintzman, N.D., Hon, G.C., Hawkins, R.D., Kheradpour, P., Stark, A., Harp, L.F., Ye, Z., Lee, L.K., Stuart, R.K., Ching, C.W., et al. (2009). Histone modifications at human enhancers reflect global cell-type-specific gene expression. Nature 459, 108–112.

18. Hwang, Y.-C., Zheng, Q., Gregory, B.D., and Wang, L.-S. (2013). High-throughput identification of long-range regulatory elements and their target promoters in the human genome. Nucleic Acids Res. 41, 4835–4846.

19. Jin, C., and Felsenfeld, G. (2007). Nucleosome stability mediated by histone variants H3.3 and H2A.Z. Genes Dev. 21, 1519–1529.

20. Jin, H., Stojnic, R., Adryan, B., Ozdemir, A., Stathopoulos, A., and Frasch, M. (2013). Genome-Wide Screens for In Vivo Tinman Binding Sites Identify Cardiac Enhancers with Diverse Functional Architectures. Plos Genet 9, e1003195.

21. Kleinjan, D.A., and van Heyningen, V. (2005). Long-Range Control of Gene Expression: Emerging Mechanisms and Disruption in Disease. Am. J. Hum. Genet. 76, 8–32.

22. Lee, T.I., and Young, R.A. (2013). Transcriptional Regulation and Its Misregulation in Disease. Cell 152, 1237–1251.

23. Lettice, L.A., Heaney, S.J.H., Purdie, L.A., Li, L., Beer, P. de, Oostra, B.A., Goode, D., Elgar, G., Hill, R.E., and Graaff, E. de (2003). 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.

24. Li, W., Notani, D., Ma, Q., Tanasa, B., Nunez, E., Chen, A.Y., Merkurjev, D., Zhang, J., Ohgi, K., Song, X., et al. (2013). Functional roles of enhancer RNAs for oestrogen-dependent transcriptional activation. Nature 498, 516–520.

25. Li, X., MacArthur, S., Bourgon, R., Nix, D., Pollard, D.A., Iyer, V.N., Hechmer, A., Simirenko, L., Stapleton, M., Hendriks, C.L.L., et al. (2008). Transcription Factors Bind Thousands of Active and Inactive Regions in the Drosophila Blastoderm. Plos Biol 6, e27.

26. Lomvardas, S., Barnea, G., Pisapia, D.J., Mendelsohn, M., Kirkland, J., and Axel, R. (2006). Interchromosomal Interactions and Olfactory Receptor Choice. Cell 126, 403–413.

27. Van Loo, P., and Marynen, P. (2009). Computational methods for the detection of cis-regulatory modules. Brief. Bioinform. 10, 509–524.

28. Montavon, T., and Duboule, D. (2012). Landscapes and archipelagos: spatial organization of gene regulation in vertebrates. Trends Cell Biol. 22, 347–354.

29. Montavon, T., Soshnikova, N., Mascrez, B., Joye, E., Thevenet, L., Splinter, E., de Laat, W., Spitz, F., and Duboule, D. (2011). A Regulatory Archipelago Controls Hox Genes Transcription in Digits. Cell 147, 1132–1145.

30. Moreau, P., Hen, R., Wasylyk, B., Everett, R., Gaub, M.P., and Chambon, P. (1981). The SV40 72 base repair repeat has a striking effect on gene expression both in SV40 and other chimeric recombinants. Nucleic Acids Res. 9, 6047–6068.

31. Noordermeer, D., Leleu, M., Splinter, E., Rougemont, J., Laat, W.D., and Duboule, D. (2011). The Dynamic Architecture of Hox Gene Clusters. Science 334, 222–225.

32. Palstra, R.-J. (2012). Transcription factor binding at enhancers: shaping a genomic regulatory landscape in flux. Front. Epigenomics Epigenetics 3, 195.

33. Pennacchio, L.A., Ahituv, N., Moses, A.M., Prabhakar, S., Nobrega, M.A., Shoukry, M., Minovitsky, S., Dubchak, I., Holt, A., Lewis, K.D., et al. (2006). In vivo enhancer analysis of human conserved non-coding sequences. Nature 444, 499–502.

34. Perry, M.W., Boettiger, A.N., and Levine, M. (2011). Multiple enhancers ensure precision of gap gene-expression patterns in the Drosophila embryo. Proc. Natl. Acad. Sci. 108, 13570–13575.

35. Rada-Iglesias, A., Bajpai, R., Swigut, T., Brugmann, S.A., Flynn, R.A., and Wysocka, J. (2011). A unique chromatin signature uncovers early developmental enhancers in humans. Nature 470, 279–283.

36. Rada-Iglesias, A., Bajpai, R., Prescott, S., Brugmann, S.A., Swigut, T., and Wysocka, J. (2012). Epigenomic Annotation of Enhancers Predicts Transcriptional Regulators of Human Neural Crest. Cell Stem Cell 11, 633–648.

37. Ronshaugen, M., and Levine, M. (2004). Visualization of trans-Homolog Enhancer-Promoter Interactions at the Abd-B Hox Locus in the Drosophila Embryo. Dev. Cell 7, 925–932.

38. Sakabe, N.J., Savic, D., and Nobrega, M.A. (2012). Transcriptional enhancers in development and disease. Genome Biol. 13, 238.

39. Sanyal, A., Lajoie, B.R., Jain, G., and Dekker, J. (2012). The long-range interaction landscape of gene promoters. Nature 489, 109–113.

40. Schmidt, D., Wilson, M.D., Ballester, B., Schwalie, P.C., Brown, G.D., Marshall, A., Kutter, C., Watt, S., Martinez-Jimenez, C.P., Mackay, S., et al. (2010). Five-vertebrate ChIP-seq reveals the evolutionary dynamics of transcription factor binding. Science 328, 1036–1040.

41. Spilianakis, C.G., Lalioti, M.D., Town, T., Lee, G.R., and Flavell, R.A. (2005). Interchromosomal associations between alternatively expressed loci. Nature 435, 637–645.

42. Spitz, F., and Furlong, E.E.M. (2012). Transcription factors: from enhancer binding to developmental control. Nat. Rev. Genet. 13, 613–626.

43. Sudarsanam, P., and Winston, F. (2000). The Swi/Snf family: nucleosome-remodeling complexes and transcriptional control. Trends Genet. 16, 345–351.

44. Vavouri, T., McEwen, G.K., Woolfe, A., Gilks, W.R., and Elgar, G. (2006). Defining a genomic

radius for long-range enhancer action: duplicated conserved non-coding elements hold the key. Trends Genet. 22, 5–10.

45. Visel, A., Prabhakar, S., Akiyama, J.A., Shoukry, M., Lewis, K.D., Holt, A., Plajzer-Frick, I., Afzal, V., Rubin, E.M., and Pennacchio, L.A. (2008). Ultraconservation identifies a small subset of extremely constrained developmental enhancers. Nat. Genet. 40, 158–160.

46. Visel, A., Rubin, E.M., and Pennacchio, L.A. (2009a). Genomic views of distant-acting enhancers. Nature 461, 199–205.

47. Visel, A., Blow, M.J., Li, Z., Zhang, T., Akiyama, J.A., Holt, A., Plajzer-Frick, I., Shoukry, M., Wright, C., Chen, F., et al. (2009b). ChIP-seq accurately predicts tissue-specific activity of enhancers. Nature 457, 854–858.

48. Wamstad, J.A., Alexander, J.M., Truty, R.M., Shrikumar, A., Li, F., Eilertson, K.E., Ding, H., Wylie, J.N., Pico, A.R., Capra, J.A., et al. (2012). Dynamic and Coordinated Epigenetic Regulation of Developmental Transitions in the Cardiac Lineage. Cell 151, 206–220.

49. Wang, D., Rendon, A., Ouwehand, W., and Wernisch, L. (2012). Transcription factor co-localization patterns affect human cell type-specific gene expression. Bmc Genomics 13, 263.

50. Wang, Z., Zang, C., Rosenfeld, J.A., Schones, D.E., Barski, A., Cuddapah, S., Cui, K., Roh, T.-Y., Peng, W., Zhang, M.Q., et al. (2008). Combinatorial patterns of histone acetylations and methylations in the human genome. Nat. Genet. 40, 897–903.

51. White, M.A., Myers, C.A., Corbo, J.C., and Cohen, B.A. (2013). Massively parallel in vivo enhancer assay reveals that highly local features determine the cis-regulatory function of ChIP-seq peaks. Proc. Natl. Acad. Sci. 110, 11952–11957.

52. Wilczynski, B., Liu, Y.-H., Yeo, Z.X., and Furlong, E.E.M. (2012). Predicting Spatial and Temporal Gene Expression Using an Integrative Model of Transcription Factor Occupancy and Chromatin State. Plos Comput Biol 8, e1002798.

53. Zeitlinger, J., Zinzen, R.P., Stark, A., Kellis, M., Zhang, H., Young, R.A., and Levine, M. (2007). Whole-genome ChIP–chip analysis of Dorsal, Twist, and Snail suggests integration of diverse patterning processes in the Drosophila embryo. Genes Dev. 21, 385–390.

54. Zinzen, R.P., Girardot, C., Gagneur, J., Braun, M., and Furlong, E.E.M. (2009). Combinatorial binding predicts spatio-temporal cis-regulatory activity. Nature 462, 65–70.