

Minireview

Divergence in *cis*-regulatory networks: taking the 'species' out of cross-species analysis

Robert P Zinzen and Eileen EM Furlong

Address: European Molecular Biology Laboratory, D-69117 Heidelberg, Germany.

Correspondence: Eileen EM Furlong. Email: furlong@embl.de

Published: 4 November 2008

Genome Biology 2008, **9**:240 (doi:10.1186/gb-2008-9-11-240)

The electronic version of this article is the complete one and can be found online at <http://genomebiology.com/2008/9/11/240>

© 2008 BioMed Central Ltd

Abstract

Many essential transcription factors have conserved roles in regulating biological programs, yet their genomic occupancy can diverge significantly. A new study demonstrates that such variations are primarily due to *cis*-regulatory sequences, rather than differences between the regulators or nuclear environments.

Genetic studies in a range of organisms reveal that essential transcription factors (TFs) tend not only to be conserved in sequence but also in function. For example, the NKx2.5 TFs are essential for heart development in species as diverse as mice [1], zebrafish [2], *Xenopus* [3], humans [4] and *Drosophila* [5]. At a structural level, the DNA-binding domains of many orthologous TFs are highly similar over large evolutionary distances, allowing them to bind to identical DNA motifs. In fact, cross-species experiments demonstrate that orthologous TFs can regulate the same target genes and even rescue some mutant phenotypes [6,7]. It is thus reasonable to assume that conserved TFs, which lead to the development and maintenance of orthologous tissues [8], regulate conserved sets of downstream target genes as part of conserved gene-regulatory networks.

It therefore came as a surprise when recent studies on DNA binding of the TFs Zeste among *Drosophila* species [9] and Ste12 and Tec1 across yeast species [10] indicated that individual binding events turn over rapidly during evolution. A similar discovery has been made for liver-specific TFs among vertebrates [11]. Mouse and human hepatocytes have a similar complement of gene expression [11] and are defined by a set of highly conserved TFs [8], yet the underlying *cis*-regulatory network appears to have diverged extensively. Odom *et al.* [11] showed that relatively few TF-binding events - perhaps even a small minority in some cases - are conserved between the two species. Their results

indicate that the target genes of hepatocyte TFs differ significantly from mouse to human, and even when orthologous genes are targeted by the same TF, the exact pattern of binding events at the conserved DNA motifs is different. These results, together with those from *Drosophila* and yeast, argue that binding events are subject to less selective pressure than previously anticipated, which has important implications for the degree of divergence in *cis*-regulatory networks.

Eliminating experimental variables when assaying cross-species TF binding

Despite the high conservation of the TFs assayed in the studies mentioned above, it is conceivable that the differences in binding signatures between species were due to differential interaction with cofactors (owing to differences in protein-protein interactions or cofactor availability), other species-specific nuclear conditions, or simply because of experimental variables. Alternatively, the genomic sequences themselves might be different enough to trigger species-specific TF-binding signatures. A new study by Wilson *et al.* [12] addresses precisely this question by using a mouse model for human trisomy 21. This partially mosaic 'Tc1' mouse line carries most of human chromosome 21 in addition to the entire murine chromosome complement [13]. Assaying TF binding to both the mouse and human chromosomes in the same cells eliminates many technical variables,

as well as variables pertaining to interspecies differences in nuclear environment. Importantly, all assayed TFs are derived from the mouse genome, as none of them, nor any known cofactors or other hepatocyte-specific factors, are encoded on human chromosome 21 [12]. The authors were therefore able to ask: 'Does a human chromosome in the murine nuclear context exhibit human-like, mouse-like, or a mixture of TF binding signatures?' In other words, does the human genetic material direct where TFs bind, or do mouse TFs bind elsewhere - maybe even to sites orthologous to the cognate mouse chromosome sites?

The authors focus on the binding events exhibited by three hepatocyte-specific TFs (HNF1a, HNF4a, and HNF6) across the orthologous regions of human chromosome 21 (WT-HsChr21) in human liver tissue, human chromosome 21 in mice (Tc1-HsChr21) and mouse chromosome 16 (Tc1-MmChr16) [12]. Only about a third to a half of identified bound regions are shared among all three chromosomes, confirming the stark differences in TF-binding events between mouse and human observed previously [11]. Importantly, the vast majority of the remaining peaks on human chromosome 21 are not found on the mouse chromosome, but rather recapitulate peaks found on chromatin isolated from human liver tissue [12]. The fact that mouse TFs, in the mouse nuclear environment, still recapitulate human-like binding signatures on a human-derived chromosome strongly indicate that it is the human chromosomal sequence that is primarily responsible for the placement of transcription factors (*cis*-directed), rather than changes in the regulators or the regulative environment (*trans*-directed). It is interesting to note that a small number of peaks (5 out of 173 non-shared peaks) appear to be *trans*-directed (Tc1-HsChr21 peaks align with Tc1-MmChr16 peaks), and may warrant further investigation in their own right.

Cis-regulation of RNA polymerase loading and transcription

Having established that the TFs are placed on the DNA in a species-specific sequence-dependent manner, the authors examined an event downstream of TF recruitment - the placement of the basal transcriptional machinery. They did this by chromatin immunoprecipitation followed by microarray analysis (ChIP-chip) against the trimethylated state of lysine 4 on histone H3 (H3K4me3) [14]. Whereas the majority of the H3K4me3 peaks detected can be identified in equivalent positions on human chromosome 21 and the corresponding mouse regions, some of these methylation marks appear species-specific, as indicated previously [15].

In Tc1 mice, the authors report 78 alignable H3K4me3 marks, of which about two-thirds (53) are shared between mouse and human. Of the remaining 25 peaks, 18 Tc1-HsChr21 peaks were also found on the WT-HsChr21 (*cis*-directed, mostly not at transcriptional start sites (TSSs)),

indicating that the human chromosomal sequence plays a significant (albeit not necessarily direct) role in the placement of at least some epigenetic marks [12]. Curiously, the remaining seven H3K4me3 marks appear *trans*-directed (also found on Tc1-MmChr16, mostly at TSSs) and may represent cases where human chromosomal regions are recognized and treated by the mouse nuclear environment in a mouse-specific manner. Finally, the authors find that the transcriptional profile of human chromosome 21 genes in Tc1 mice resembles their transcription in the native human environment, rather than the transcriptional profile of their murine orthologs [12].

Insights into cis-regulatory evolution

Studies of *cis*-evolution have largely focused on individual enhancers or *cis*-regulatory modules (CRMs) [16-19]; however, more recent studies venture to identify *cis*-regulatory differences on a global scale [10,11,20]. The use of the trans-chromosomal Tc1 mice [12] to address species-specific differences in transcriptional regulation is certainly elegant, and one wonders if, in principle, a similar system might be extendable to other chromosomes, transcription factors, tissues, developmental contexts and species.

The study by Wilson *et al.* [12] provides strong evidence that it is the genomic sequence, rather than differences in nuclear environment, which is primarily responsible for the differences in mouse versus human TF occupancy. This underlines the importance of measuring TF binding directly rather than inferring occupancy through sequence and phylogenomic analysis. The ability of murine hepatocyte TFs to 'read' the transcriptional program of a human chromosome, even when placed in the nuclear environment of the mouse, a species separated from humans by approximately 75-100 million years, adds to the growing evidence that *cis*-regulatory changes are a major (if not the) driving force of evolutionary change [21].

As with all interspecies comparisons, the conclusions that can be drawn from these studies are largely dependent on reliable alignment of the genomes and the faithful mapping of orthologous regions [22]. For example, misalignment of ChIP peaks will skew data, as orthologous peaks could easily be misannotated as *trans*-, rather than *cis*-directed. The task of sequence alignment is relatively tractable when performing interspecies comparisons of coding regions, but the challenge is exponentially more difficult when comparing noncoding regions. Even with largely syntenic chromosomes (such as mouse chromosome 16 *versus* human chromosome 21), defining orthologous peaks is very difficult. Choosing the proper species for cross-species analyses is extremely important and depends on the precise question being asked (for example, [17]): whereas comparisons over large evolutionary distances might yield insights into gross changes in gene regulatory networks [10,12], comparisons over smaller

distances might be more fruitful when dissecting differences in the underlying *cis*-regulatory networks [9,16].

One important remaining question from the hepatocyte studies [11,12] concerns the functional activity of species-specific TF binding. Although the authors show by Solexa sequencing that most of the species-unique H3K4me3 marks are associated with transcription, a precise analysis of the overlap of TF-bound regions with regions of active transcription (deduced from either H3K4me3 marks or expression profiling) was not presented. Do the genomic regions bound in both human and mouse correspond to regulatory regions in the vicinity of active transcription (that is, in close proximity to shared H3K4me3 peaks), whereas uniquely bound regions do not? In other words, do conserved binding events represent the functional sites? If this is the case, it suggests that once 'functional' *cis*-binding events are distilled from non-functional ones, there may be significant conservation in *cis*-regulatory networks. Alternatively, although the general properties of gene regulatory networks are conserved, the underlying *cis*-regulatory networks may have undergone significant divergence. No doubt future *cis*-evolutionary studies, both at individual loci and genome-wide, will begin to unravel this question and provide exciting insights into the general principles underlying the changes in *cis*-regulatory networks during speciation.

References

- Lyons I, Parsons LM, Hartley L, Li R, Andrews JE, Robb L, Harvey RP: **Myogenic and morphogenetic defects in the heart tubes of murine embryos lacking the homeo box gene Nkx2-5.** *Genes Dev* 1995, **9**:1654-1666.
- Chen JN, Fishman MC: **Zebrafish tinman homolog demarcates the heart field and initiates myocardial differentiation.** *Development* 1996, **122**:3809-3816.
- Fu Y, Yan W, Mohun TJ, Evans SM: **Vertebrate tinman homologues XNkx2-3 and XNkx2-5 are required for heart formation in a functionally redundant manner.** *Development* 1998, **125**:4439-4449.
- Schott JJ, Benson DW, Basson CT, Pease W, Silberbach GM, Moak JP, Maron BJ, Seidman CE, Seidman JG: **Congenital heart disease caused by mutations in the transcription factor NKX2-5.** *Science* 1998, **281**:108-111.
- Azpiazu N, Frasch M: **tinman and bagpipe: two homeo box genes that determine cell fates in the dorsal mesoderm of Drosophila.** *Genes Dev* 1993, **7**:1325-1340.
- Haun C, Alexander J, Stainier DY, Okkema PG: **Rescue of Caenorhabditis elegans pharyngeal development by a vertebrate heart specification gene.** *Proc Natl Acad Sci USA* 1998, **95**:5072-5075.
- Zaffran S, Reim I, Qian L, Lo PC, Bodmer R, Frasch M: **Cardioblast-intrinsic Tinman activity controls proper diversification and differentiation of myocardial cells in Drosophila.** *Development* 2006, **133**:4073-4083.
- Zaret KS: **Regulatory phases of early liver development: paradigms of organogenesis.** *Nat Rev Genet* 2002, **3**:499-512.
- Moses AM, Pollard DA, Nix DA, Iyer VN, Li XY, Biggin MD, Eisen MB: **Large-scale turnover of functional transcription factor binding sites in Drosophila.** *PLoS Comput Biol* 2006, **2**:e130.
- Borneman AR, Gianoulis TA, Zhang ZD, Yu H, Rozowsky J, Seringhaus MR, Wang LY, Gerstein M, Snyder M: **Divergence of transcription factor binding sites across related yeast species.** *Science* 2007, **317**:815-819.
- Odom DT, Dowell RD, Jacobsen ES, Gordon W, Danford TW, MacIsaac KD, Rolfe PA, Conboy CM, Gifford DK, Fraenkel E: **Tissue-specific transcriptional regulation has diverged significantly between human and mouse.** *Nat Genet* 2007, **39**:730-732.
- Wilson MD, Barbosa-Morais NL, Schmidt D, Conboy CM, Vanes L, Tybulewicz VL, Fisher EM, Tavare S, Odom DT: **Species-specific transcription in mice carrying human chromosome 21.** *Science* 2008, **322**:434-438.
- O'Doherty A, Ruf S, Mulligan C, Hildreth V, Errington ML, Cooke S, Sesay A, Modino S, Vanes L, Hernandez D, Linehan JM, Sharpe PT, Brandner S, Bliss TV, Henderson DJ, Nizetic D, Tybulewicz VL, Fisher EM: **An aneuploid mouse strain carrying human chromosome 21 with Down syndrome phenotypes.** *Science* 2005, **309**:2033-2037.
- Guenther MG, Levine SS, Boyer LA, Jaenisch R, Young RA: **A chromatin landmark and transcription initiation at most promoters in human cells.** *Cell* 2007, **130**:77-88.
- Bernstein BE, Kamal M, Lindblad-Toh K, Bekiranov S, Bailey DK, Huebert DJ, McMahon S, Karlsson EK, Kulbokas EJ 3rd, Gingeras TR, Schreiber SL, Lander ES: **Genomic maps and comparative analysis of histone modifications in human and mouse.** *Cell* 2005, **120**:169-181.
- Gompel N, Prud'homme B, Wittkopp PJ, Kassner VA, Carroll SB: **Chance caught on the wing: cis-regulatory evolution and the origin of pigment patterns in Drosophila.** *Nature* 2005, **433**:481-487.
- Hare EE, Peterson BK, Iyer VN, Meier R, Eisen MB: **Sepsid even-skipped enhancers are functionally conserved in Drosophila despite lack of sequence conservation.** *PLoS Genet* 2008, **4**:e1000106.
- Ludwig MZ, Palsson A, Alekseeva E, Bergman CM, Nathan J, Kreitman M: **Functional evolution of a cis-regulatory module.** *PLoS Biol* 2005, **3**:e93.
- Zinzen RP, Cande J, Ronshaugen M, Papatsenko D, Levine M: **Evolution of the ventral midline in insect embryos.** *Dev Cell* 2006, **11**:895-902.
- Tirosh I, Weinberger A, Bezalel D, Kaganovich M, Barkai N: **On the relation between promoter divergence and gene expression evolution.** *Mol Syst Biol* 2008, **4**:159.
- Carroll SB: **Evo-devo and an expanding evolutionary synthesis: a genetic theory of morphological evolution.** *Cell* 2008, **134**:25-36.
- Tirosh I, Bilu Y, Barkai N: **Comparative biology: beyond sequence analysis.** *Curr Opin Biotechnol* 2007, **18**:371-377.