Molecular Systems Biology (2006) doi:10.1038/msb4100066

www.molecularsystemsbiology.com
Article number: 2006.0021

molecular systems biology

NEWS & VIEWS

Stability and evolution of transcriptional regulatory networks

Kenneth S Zaret

Cell and Developmental Biology Program, Fox Chase Cancer Center, Philadelphia, PA, USA

Molecular Systems Biology 2 May 2006; doi:10.1038/msb4100066

How do transcriptional regulatory networks elicit stable gene expression patterns? How do such networks evolve? These topics have kept molecular biologists occupied for years. The recent elucidation of regulatory networks that control endoderm development, including 'genomic location analysis' of transcription factors in endoderm-derived liver cells, provides more comprehensive views than in the past. But how much more have we learned?

Genome location analysis, or ChIP-on-chip, first involves crosslinking chromatin in native cells, breaking it into small fragments (e.g., 0.5–1 kb), immunoprecipitating the fragments that contain a specific transcription factor antigen (ChIP), and performing ligation-mediated PCR to amplify the ChIP fragments. The amplified DNA pool is then labeled and hybridized to a microarray (chip) containing single-stranded DNA probes for the genomic regions of interest; in the present case, regions spanning the promoters of most genes. The resulting information, when compared to the appropriate controls, yields semi-quantitative information about how frequently in the cell population a transcription factor is bound to each promoter queried on the chip.

In an article currently published in *Molecular Systems Biology*, Odom *et al* (2006) performed genomic location analysis on six endodermal and liver transcriptional activators (FOXA2, HNF1α, HNF4α, HNF6, CREB1 and USF1) at 18 000 promoters in isolated human liver cells. They analyzed 10 kb spanning each promoter at 250 bp resolution, allowing them to determine whether the bound factors were clustered at specific promoters in the genome. Many more promoters were bound by multiple regulators than predicted by random assortment, confirming the importance of combinatorial control at the genomic level (also see Rada-Iglesias *et al*, 2005). Combinatorial control has been known from individual gene promoter ('one-off') studies for many years (Yamamoto, 1985), reinforcing the idea that if factors work in various combinations, far fewer factors are required to specify distinct gene expression patterns for different cell types than if each expression pattern in a cell type is specified by a distinct factor.

Odom *et al* (2006) also found a strong positive correlation between the number of regulators bound to a promoter and the extent of gene activity. This observation could only be derived with confidence from the statistical power of genomic studies. It extends an earlier prediction that activators (i.e. intrinsic activation) would figure more prominently than repressors (i.e., derepression) at promoters for which high levels of gene expression are necessary (Savageau, 1977). In evolutionary terms, where high gene activity is selected positively, it could be simpler for additional activator binding sites to arise at a promoter than for the mutational changes required to enhance the magnitude of activation by pre-existing binding factors. For promoters that require extremes of activity (e.g., tissue specificity), a high degree of functional cooperativity among binding factors may be necessary.

Perhaps the most interesting findings of Odom *et al* (2006) were the cross- and auto-regulatory relationships among the transcription factor genes themselves. Notably, five of the six factors tested bound to their own promoters, suggesting that they autoregulate their own expression. By contrast, earlier studies from the Young laboratory found that fewer than 10% of transcription factors in yeast bind their own promoters (Harbison *et al*, 2004), but re-analysis of the data by the present authors (Odom *et al*, 2006) indicated that the master regulators of yeast cell processes more frequently bound their own promoters than other factors; that is, were apparent autoregulators. Of the apparent autoregulators studied by Odom *et al* (2006), FOXA2 and HNF6 first function in embryonic endodermal progenitors to the liver, and HNF4α and HNF1α first function in newly specified hepatic cells; thus, all four are initiators of the regulatory networks for endoderm or liver gene expression, apart from being involved in maintenance (Zaret, 2002). The fifth self-promoter binding factor is CREB, which responds to elevated cyclic AMP during hormonal stimulation, while the factor that did not bind its own promoter is USF1, which is ubiquitously expressed. The authors cite theoretical and experimental studies showing that autoregulation is crucial for providing stability to gene expression patterns. While previous one-off and genomic studies had shown autoregulation for some of the factors studied here, Odom *et al* (2006) provide a more comparative view from a single experimental platform.

Importantly, findings from genome location analysis need to be confirmed by conditional expression and/or genetic studies. A recent study of the glucocorticoid receptor found that over 75% of the genes to which the receptor bound did not exhibit altered gene expression in the presence of glucocorticoids (Phuc Le *et al*, 2005). Furthermore, previous genome location studies of Odom *et al* (2004) demonstrated that HNF4α bound to 12% of the 13 000 queried promoters in liver chromatin, which is vastly greater than the maximum of 2.5% bound that they observed for other transcription factors. The apparent excess of target sites for certain factors that are identified by genomic location analysis may reflect technical

issues, such as nonspecific activities of antibodies used for ChIP. Studies of HNF4α-null cells may be necessary to resolve the issue.

Two other possibilities can explain the large number of genomic binding sites for certain transcription factors. First, the factors may have unanticipated functions. Second, binding to a subset of genomic sites may be functionally neutral. Taking the view that present-day organisms represent works in progress, rather than terminal evolutionary states, excess factor binding sites may be tolerated in the genome, as well as provide the opportunity for selective advantage under unusual conditions that may later become fixed for a population.

Another perspective on transcriptional regulatory networks is gained by asking which networks are the most conserved. Davidson and Erwin (2006) recently noted that FOXA transcription factors similarly autoregulate their genes in animals that have been separated by over half a billion years of evolution. More significantly, such autoregulation is linked to a detailed cross-regulatory network with five transcription factors that control endoderm development and are similarly conserved. Davidson and Erwin (2006) termed such a highly conserved network a 'kernel' and suggest that disruption of any kernel gene would be disastrous for tissue (e.g., endoderm) function. They further defined various regulatory network 'plug-ins' that link to the kernel network and specify the differences in endoderm development among metazoan organisms.

In summary, detailed studies of regulatory networks within individual tissues of single organisms and cross-species analyses are beginning to provide a comprehensive view of the wiring that underlies gene expression patterns, and how the wiring evolved. We clearly have a lot to learn, but the day may not be far off when such principles may be used to manipulate and design genetic networks and, hence, tissues for biomedical needs and interests.

References

Davidson EH, Erwin DH (2006) Gene regulatory networks and the evolution of animal body plans. *Science* **311:** 796–800

Harbison CT, Gordon DB, Lee TI, Rinaldi NJ, Macisaac KD, Danford TW, Hannett NM, Tagne JB, Reynolds DB, Yoo J, Jennings EG, Zeitlinger J, Pokholok DK, Kellis M, Rolfe PA, Takusagawa KT, Lander ES, Gifford DK, Fraenkel E, Young RA (2004) Transcriptional regulatory code of a eukaryotic genome. *Nature* **431:** 99–104

Odom DT, Dowell RD, Jacobsen ES, Nekludova L, Rolfe PA, Danford TW, Gifford DK, Fraenkel E, Bell GI, Young RA (2006) Core transcriptional regulatory circuitry in human hepatocytes. *Mol Syst Biol* (doi:10.1038/msb4100059)

Odom DT, Zizlsperger N, Gordon DB, Bell GW, Rinaldi NJ, Murray HL, Volkert TL, Schreiber J, Rolfe PA, Gifford DK, Fraenkel E, Bell GI, Young RA (2004) Control of pancreas and liver gene expression by HNF transcription factors. *Science* **303:** 1378–1381

Phuc Le P, Friedman JR, Schug J, Brestelli JE, Parker JB, Bochkis IM, Kaestner KH (2005) Glucocorticoid receptor-dependent gene regulatory networks. *PLoS Genet* **1:** e16

Rada-Iglesias A, Wallerman O, Koch C, Ameur A, Enroth S, Clelland G, Wester K, Wilcox S, Dovey OM, Ellis PD, Wraight VL, James K, Andrews R, Langford C, Dhami P, Carter N, Vetrie D, Ponten F, Komorowski J, Dunham I, Wadelius C (2005) Binding sites for metabolic disease related transcription factors inferred at base pair resolution by chromatin immunoprecipitation and genomic microarrays. *Hum Mol Genet* **14:** 3435–3447

Savageau MA (1977) Design of molecular control mechanisms and the demand for gene expression. *Proc Natl Acad Sci USA* **74:** 5647–5651

Yamamoto KR (1985) Steroid receptor regulated transcription of specific genes and gene networks. *Ann Rev Genet* **19:** 209–252

Zaret KS (2002) Regulatory phases of early liver development: paradigms of organogenesis. *Nat Rev Genet* **3:** 499–512