Cell

Leading Edge
Perspective

Making Sense of Transcription Networks

Trevor R. Sorrells[1] and Alexander D. Johnson[1,*]
[1]Department of Biochemistry & Biophysics and Department of Microbiology & Immunology, Tetrad Graduate Program, University of California, San Francisco, San Francisco, CA 94158, USA
*Correspondence: ajohnson@cgl.ucsf.edu
http://dx.doi.org/10.1016/j.cell.2015.04.014

When transcription regulatory networks are compared among distantly related eukaryotes, a number of striking similarities are observed: a larger-than-expected number of genes, extensive overlapping connections, and an apparently high degree of functional redundancy. It is often assumed that the complexity of these networks represents optimized solutions, precisely sculpted by natural selection; their common features are often asserted to be adaptive. Here, we discuss support for an alternative hypothesis: the common structural features of transcription networks arise from evolutionary trajectories of "least resistance"—that is, the relative ease with which certain types of network structures are formed during their evolution.

Introduction

The complexity of cells continues to fascinate scientists. Two broad views are often advanced to account for such complexity. In one, it is assumed that any complexity must necessarily benefit the cell. Some cell and molecular biologists go even further and discuss how a particular mechanism was "designed" by evolution to be perfectly matched to its task. As with a machine, it is assumed that every molecular nut and bolt must have a purpose. Because this view seems intuitive and relatively simple (after all, examples abound of animals, plants, and microbes adapted to their environments), it is often invoked to explain any aspect of cell and molecular biology. A different view, the one we elaborate here, is embodied in Dobzhansky's famous line, now a cliché, "nothing in biology makes sense except in the light of evolution." According to this view, any rationalization of a modern cellular mechanism depends critically on understanding its evolutionary history. We argue that this emphasis on evolutionary history is especially appropriate for analyzing transcription circuits and for rationalizing their structures. This view has explanatory power in that it can readily account for some of the more bewildering and counterintuitive features of modern transcription circuits; it also gives us insight into the best ways to describe and study such circuits.

In this Perspective, we first review common features of transcription network structures—observed across diverse species—and argue that these similarities cannot be the result of descent from a single ancestral circuit possessing these characteristics. Next, we consider key biochemical and biophysical properties of transcription regulators and *cis*-regulatory sequences that make certain evolutionary pathways much more probable than others, in part because they circumvent fitness barriers. Finally, we argue that many aspects of transcription circuits, particularly those that seem overly complex and counterintuitive, can be understood as relatively crude products of high-probability evolutionary trajectories rather than as highly optimized, specific solutions.

The arguments discussed in this perspective rely heavily on prior ideas advanced by evolutionary biologists, particularly those ideas concerning the role of non-adaptive mutations in generating complexity (Covello and Gray, 1993; Doolittle, 2013; Force et al., 1999; Gray et al., 2010; Lukeš et al., 2011; Lynch, 2007a, 2007b, 2014; Stoltzfus, 1999; Zuckerkandl, 1997). Although sometimes dismissed as unimportant (or uninteresting), non-adaptive mutations can have a profound role in generating evolutionary novelty. Of particular importance is the idea, sometimes called "constructive neutral evolution," that changes that arise neutrally can open up new evolutionary pathways; in some cases, changes that arose non-adaptively can become essential for function if they are incorporated into subsequent layers of evolutionary change. Through this sequence of events, molecular and organismal complexity can be increased through non-adaptive mutations. As we discuss, the biochemical and biophysical properties of transcription network components support the idea that their evolutionary trajectories—which depend on mutation, selection, and genetic drift—lead to specific types of structures. Because their components are highly conserved across eukaryotes, we argue that it is inevitable that networks across a wide variety of species tend to converge on similar structures. We propose that these common structures are not likely to represent optimized solutions but are, in a sense, "default" evolutionary products.

Depictions of Transcription Networks

For the most part, genome-wide studies of transcriptional network structures have been largely descriptive, often culminating in large "hairball" diagrams such as those depicted in Figure 1. Their complexity has made it difficult to formulate simple conclusions regarding the logic or outputs of these networks, particularly since quantitative parameters and dynamic measurements are typically lacking.

Although there are many components of gene expression networks, we will focus here on only two key elements, transcription regulators and *cis*-regulatory sequences. We define transcription regulators as sequence-specific DNA-binding proteins that control the transcription of specific genes by binding to *cis*-regulatory sequences, short (typically 6–15 nucleotides) DNA

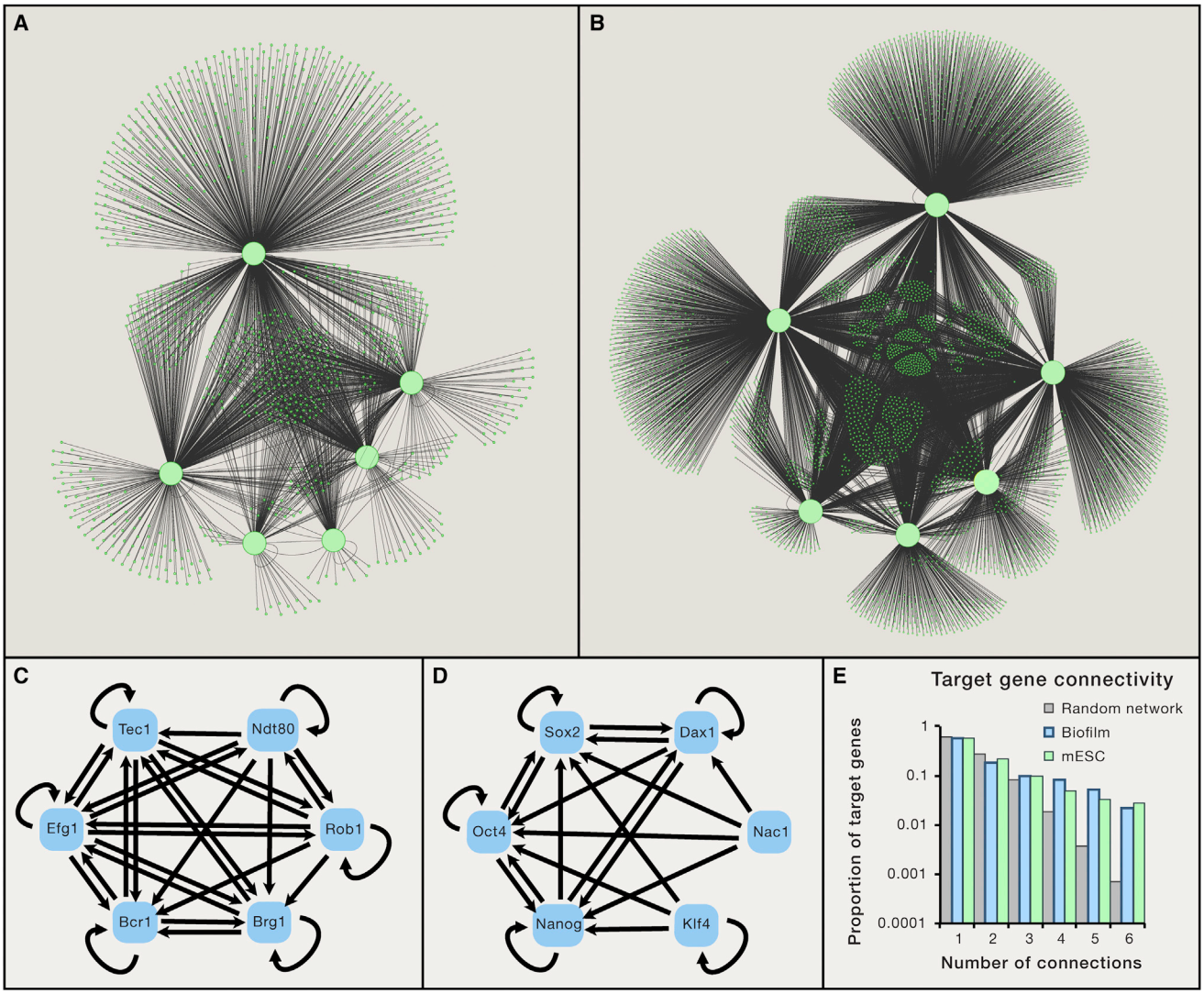

Figure 1. Typical Depictions of Transcription Regulatory Networks

(A and B) (A) The *C. albicans* biofilm network (Nobile et al., 2012) and (B) the *M. musculus* embryonic stem cell network (Kim et al., 2008) are depicted as graphs where balls represent genes and lines represent the binding of transcription regulators to intergenic regions. Master transcription regulators (defined in the text) are shown as large balls, and "target genes" are shown as small balls. For the stem cell network, only the six most heavily connected transcription regulators are shown.

(C and D) Close-up of the core of each network, showing only the binding connections between the master transcription regulators. Directionality of the connection is indicated by arrows. Note that the arrows refer only to binding connections and do not imply that the connection activates the recipient gene. (C) *C. albicans* biofilm, (D) mouse stem cell networks.

(E) The degree of connectivity for nodes in the two networks. The two biological networks show a larger proportion of nodes with high connectivity than would be found in a random network (Lee et al., 2002).

sequences. It is the distribution of these *cis*-regulatory sequences across the genome that largely specifies the time, place, and rate of each gene's transcription; this information is "read" by transcription regulators, whose binding to DNA specifies, often through a complex series of downstream steps, the rate of transcription of the gene. Although in many eukaryotic species, *cis*-regulatory sequences are typically located within several thousand nucleotide pairs of the genes they control, in plants and animals, they can be spread out over hundreds of thousands of nucleotide pairs. Nearly all eukaryotic genes are directly controlled by more than one transcription regulator, and most genes respond to dozens of regulators, specified by the identity and arrangement of their *cis*-regulatory sequences. We also know, from decades of "promoter bashing" experiments, that *cis*-regulatory sequences can be moved from one gene to another (and from one species to another) and still retain much of their specificity to direct transcription. Finally, transcription regulators typically bind cooperatively to DNA, a fundamental property that, as we shall discuss, has important implications for network evolution.

Many additional proteins besides transcription regulators are needed to transcribe a gene (for example, RNA polymerase