

NEWS AND VIEWS

Modeling gene expression control using Ohm's Law

Harmen J Bussemaker

Department of Biological Sciences and Center for Computational Biology and Bioinformatics, Columbia University, New York, NY, USA

Molecular Systems Biology 18 April 2006; doi:10.1038/msb4100055

The binding of transcription factors (TFs) to specific sites in the genome is a crucial step in the molecular process controlling gene expression. The *in vitro* sequence specificity of these regulatory proteins can generally be well represented by consensus DNA motifs or slightly more sophisticated sequence profiles called position-specific scoring matrices. These are widely used to scan genome sequences in order to find novel transcriptional target genes. Unfortunately, usually only a small fraction of the 'hits' thus obtained are functional *in vivo*, where local chromatin structure and TF-TF interactions come into play. Taking into account the context provided by the surrounding noncoding DNA is therefore essential. In a recent study currently published in *Molecular Systems Biology*, Nguyen and D'haeseleer (2006) present a promising strategy for determining which context features are most important for a given TF binding motif. Their approach belongs to a growing class of methods that fit simple mathematical models of transcription regulation to DNA microarray data to map gene regulation networks.

Many of the molecular players that govern gene expression are known, but our knowledge about their interactions with the DNA and with each other is very incomplete. Information about the gene regulatory network is only implicitly represented in the large volume of functional genomics data now available to us. The strengths of the 'arrows' between TFs and their target genes and the condition-specific activities of the regulatory 'nodes' need to be inferred by computational means. A detailed mathematical model that accurately describes the molecular computations performed by the cell would greatly deepen our understanding of cellular physiology, and provide a framework for analyzing regulatory pathways or predicting the effects of genetic variation between individuals.

While the activity of a TF is often represented by its mRNA expression level (Segal *et al.*, 2003), regulatory control is more often than not exerted at the level of subcellular localization or covalent modification of the protein, or the presence/absence of ligands. These variables really define the regulatory state of the cell, but they are much harder to measure experimentally than mRNA expression levels and therefore usually remain 'hidden'. Nguyen and D'haeseleer use multivariate linear modeling to computationally infer the hidden post-translational activity of each TF from the mRNA expression levels of its target genes, ignoring the mRNA expression level of the TF itself. This model-based approach was previously introduced (Bussemaker *et al.*, 2001) as an alternative to clustering-based analysis of microarray data (Eisen *et al.*, 1998; Beer and Tavazoie, 2004), and has been extended to include TF deletion data (Wang *et al.*, 2002),

position-specific scoring matrices (Conlon *et al.*, 2003; Foat *et al.*, 2005), and TF-TF interactions (Das *et al.*, 2004). Since each individual microarray experiment is analyzed by itself, TF activities can be inferred in a condition-specific manner.

The ability to infer condition-specific TF activities makes it possible to estimate the regulatory coupling strength between a TF and a putative target gene, by comparing the mRNA expression profile of the gene with the inferred TF activity profile across a large number of microarray experiments. This approach has previously been used (Liao *et al.*, 2003; Gao *et al.*, 2004) to refine the gene regulatory network structure derived from genome-wide TF occupancy data (Harbison *et al.*, 2004). Nguyen and D'haeseleer derive their initial guess of the network connectivity from matches to TF binding motifs in noncoding sequence, and subsequently use a modified version of the method of Liao *et al.* (2003) to self-consistently infer a matrix of inferred activities of every TF in every condition and a matrix of regulatory coupling strengths between every TF and every gene. Their approach provides an alternative to the use of evolutionary conservation to distinguish functional DNA motifs from nonfunctional ones (Kellis *et al.*, 2003). While this is already interesting *per se*, the unique insight of the authors is that the inferred regulatory couplings can in turn be analyzed to determine which aspects of the promoter context cause the same motif to be functional in one gene and nonfunctional in another. They use this approach to gain insight into the role of promoter geometry and the interplay between two elusive motifs called PAC and rRPE.

An appealing analogy exists between the linear model for transcription regulation used by Nguyen and D'haeseleer and the well-known linear equation called Ohm's Law, $I=GV$, which states that the electrical current (I) through a resistor is proportional to the voltage (V) across it. In the cell, TF activities play the role of the voltage and transcription rates that of the current, while the regulatory coupling between a TF and a target gene corresponds to the conductivity (G) of the resistor (see Figure 1). Changes in the mRNA expression level of all genes (often called the 'transcriptome') are interpreted as a response to changes in the regulatory activity of all TFs (which we might call the 'transfactome'), and this relationship is modeled by a linear equation one might refer to as 'Ohm's Law'. Nguyen and D'haeseleer show that Ohm's Law allows them to predict condition-specific expression levels that were held out from the data set used to fit their model parameters more accurately than the method of Beer and Tavazoie (2004).

Electrical engineers will be surprised to learn that, in biology, the observed conductivity of a resistor strongly

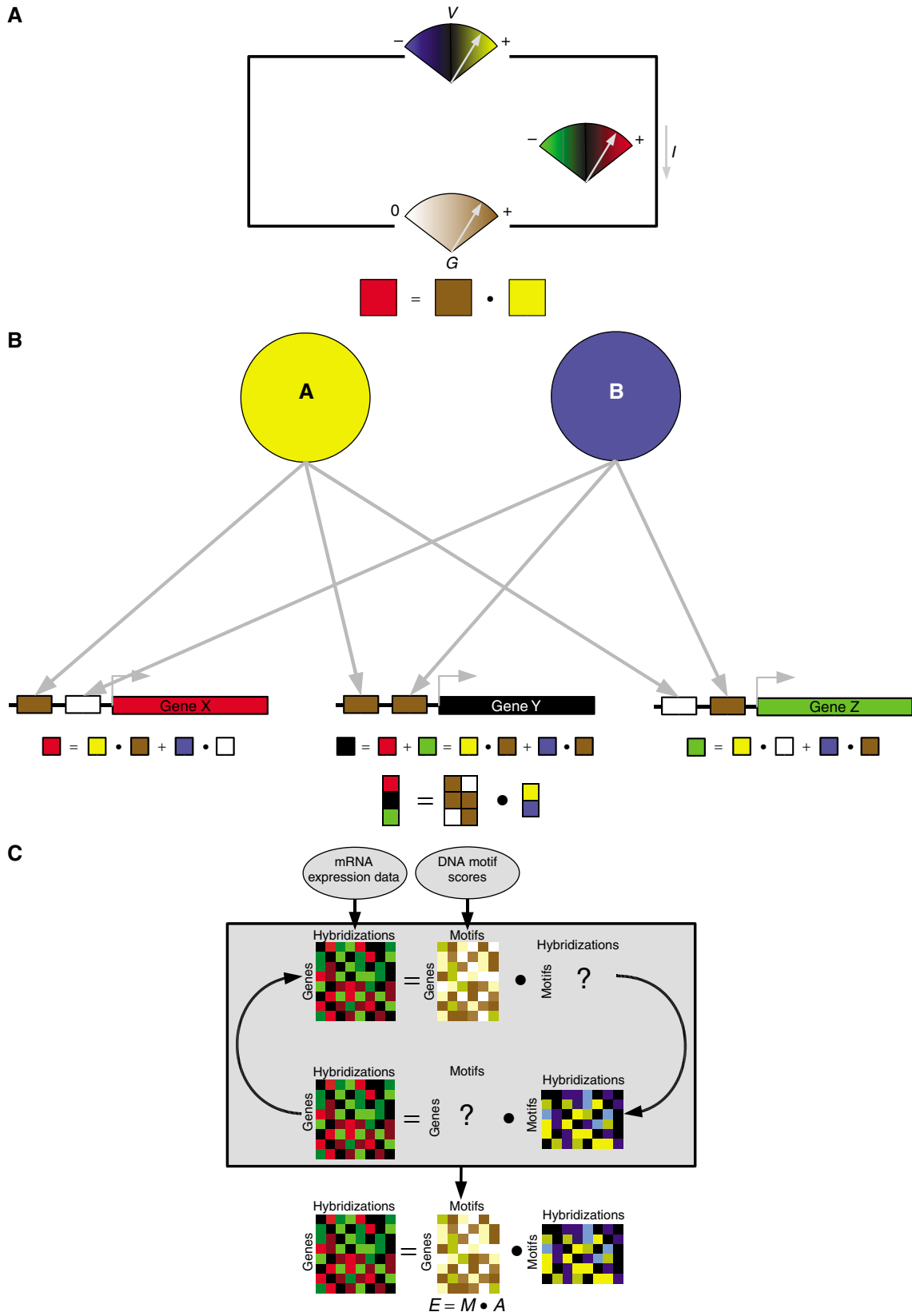


Figure 1 Illustrating the analogy between Ohm's Law and 'Omes Law'. The work of Nguyen and D'haeseleer extends a class of linear models for gene expression regulation that has a very direct and useful analogy to basic electricity theory. **(A)** In an electrical circuit, Ohm's Law, $I = GV$, describes the linear relation that exists between the current (I) through a resistor and the voltage (V) that drives it, the constant of proportionality being the conductivity (G) of the resistor. We use a color-coding scheme where a scale from white ($G=0$) to dark brown ($G>0$) represents conductivity, green ($I<0$) to red ($I>0$) via black ($I=0$) represents current, and blue ($V<0$) to yellow ($V>0$) via black ($V=0$) represents voltage. **(B)** In a gene regulatory network, changes in 'hidden' post-translational TF activity play the role of the voltage, while the resulting changes in mRNA expression level play that of the current. For any given TF, the regulatory strength of DNA binding sites in the upstream region, or 'conductivity', varies greatly between genes. The change in mRNA expression for a given gene is a weighted combination of the changes in activity of the TFs that bind to its upstream region. In the example shown, gene X is only controlled by factor A, while gene Y is controlled by both factor A and factor B. Therefore, while gene X is upregulated (red) in response to the increase in the activity of factor A (yellow), the decrease in the activity of factor B (blue) causes the net change in expression of gene Y to be zero (black). The many-to-many relationship between TF activities and mRNA expression levels can be summarized in the form of a linear matrix equation ('Omes Law'). **(C)** Schematic depiction of the iterative procedure used by Nguyen and D'haeseleer to simultaneously infer a matrix of condition-specific TF activity changes (blue/yellow) and a matrix of gene-specific motif strengths (white/brown), which together optimally explain the observed mRNA expression changes (green/red).

depends on where it gets inserted into the electronic circuit. With the work of Nguyen and D'haeseleer, we now have a computational strategy to systematically analyze how genomic context influences the *in vivo* responsiveness of TF binding sites.

References

- Beer MA, Tavazoie S (2004) Predicting gene expression from sequence. *Cell* **117**: 185–198
- Bussemaker HJ, Li H, Siggia ED (2001) Regulatory element detection using correlation with expression. *Nat Genet* **27**: 167–171
- Conlon EM, Liu XS, Lieb JD, Liu JS (2003) Integrating regulatory motif discovery and genome-wide expression analysis. *Proc Natl Acad Sci USA* **100**: 3339–3344
- Das D, Banerjee N, Zhang MQ (2004) Interacting models of cooperative gene regulation. *Proc Natl Acad Sci USA* **101**: 16234–16239
- Eisen MB, Spellman PT, Brown PO, Botstein D (1998) Cluster analysis and display of genome-wide expression patterns. *Proc Natl Acad Sci USA* **95**: 14863–14868
- Foat BC, Houshmandi SS, Olivas WM, Bussemaker HJ (2005) Profiling condition-specific, genome-wide regulation of mRNA stability in yeast. *Proc Natl Acad Sci USA* **102**: 17675–17680
- Gao F, Foat BC, Bussemaker HJ (2004) Defining transcriptional networks through integrative modeling of mRNA expression and transcription factor binding data. *BMC Bioinform* **5**: 31
- 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
- Kellis M, Patterson N, Endrizzi M, Birren B, Lander ES (2003) Sequencing and comparison of yeast species to identify genes and regulatory elements. *Nature* **423**: 241–254
- Liao JC, Boscolo R, Yang YL, Tran LM, Sabatti C, Roychowdhury VP (2003) Network component analysis: reconstruction of regulatory signals in biological systems. *Proc Natl Acad Sci USA* **100**: 15522–15527
- Nguyen DH, D'haeseleer P (2006) Deciphering principles of transcription regulation in eukaryotic genomes. *Mol Syst Biol* doi:10.1038/msb4100054
- Segal E, Shapira M, Regev A, Pe'er D, Botstein D, Koller D, Friedman N (2003) Module networks: identifying regulatory modules and their condition-specific regulators from gene expression data. *Nat Genet* **34**: 166–176
- Wang W, Cherry JM, Botstein D, Li H (2002) A systematic approach to reconstructing transcription networks in *Saccharomyces cerevisiae*. *Proc Natl Acad Sci USA* **99**: 16893–16898