Using Network Dynamics Toward a Top-Down Approach to Gene Expression Analysis

Sandy Shaw
Prime Genomics
San Francisco, CA, USA
Email: sshaw@primegenomics.com

Many studies involving gene expression data choose a "bottom-up" approach, looking to isolate effects from a single gene or group of genes for their impact on the phenotype or pathology of the organism in question. At the same time, the starting point for medical examination of complex organisms, such as the human body, often begins with a "top-down" approach, checking homeostatic equilibrium by using indicators such as body temperature or pulse to probe for abnormalities within the underlying system. In this paper, we use recent results in complex network theory to show evidence of a parameter that can be used for such a top-down approach to genomic analysis as well. We apply these methods to the well-studied Golub leukemia dataset and a more recent dataset involving resistance to HIV-1 infection to indicate that there is linkage between this parameter and phenotype.

1. INTRODUCTION

Many studies involving gene expression data choose a "bottom-up" approach to variations in clinical phenotypes, looking to isolate causality to a single gene or group of genes. At the same time, everyday clinical medicine begins with a top-down approach, checking systemic homeostatic equilibrium using simple measurements, such as body temperature or pulse, to probe for variations within the underlying phenotype.

We use recent results in complex network theory and current studies of the role of noise in cellular processes[1-3] to show evidence of a parameter that can be used for such a top-down systemic approach in genomics. This parameter, which we call the characteristic noise parameter or CNP, quantifies the scaling of an informative type of genome-wide biochemical noise (the analogous "temperature"). We evaluated CNP from microarray data samples in the well-studied Golub leukemia dataset and samples from a more recent dataset involving resistance to HIV-1 infection. Both tests indicated linkage between CNP and phenotype. The HIV study pointed to the utility of this approach for identification of HIV-1 resistance and its possible causes.

2. METHODS

• We examined two competing processes within the dynamics of the genome. One of synchronization between gene network "hubs" and one of desynchronization caused by noise from individual genes or smaller gene networks[1-3,5].

• The scaling of correlated gene cluster sizes in S. cerevisiae microarray data (Spellman 1998) produced evidence of a "homeostatic" dynamic equilibrium between these two tendencies. This equilibrium manifests as a type of global biochemical noise throughout the genome.

• We then applied similar methods to compare scaling of datasets against phenotype for subjects with two types of leukemia (AML and ALL) in Golub (1999) and subjects identified as being resistant or non-resistant to HIV-1 infection in McLaren (2009). The role of noise has been strongly implicated in HIV-1 infection[4].

2.1. The Noise Parameter (CNP)

The equation describing the scaling of this form of global noise involves three premises:

1. Synchronization is a type of percolation process in the genome.

2. Desynchronization is a type of "anti-percolation" process[3].

3. The topology of gene networks is scale-free[5,6].

$$N_{av} = (M/\log(M))^{S} \quad (1)$$

Eq. (1) represents a stochastic equilibrium for conditions 1-3. N_{av} and M are dependent on a correlation threshold C. The scaling parameter S is a constant. In the case of gene expression networks:

• N_{av} is the average number of links to a group of G gene expression values from all other groups of size G. A link is defined as a correlation value between it and

another group of size G above a given correlation threshold C (e.g. > C = .95).

• M is the largest number of links found for any group. S we call the characteristic noise parameter (CNP).

3. RESULTS

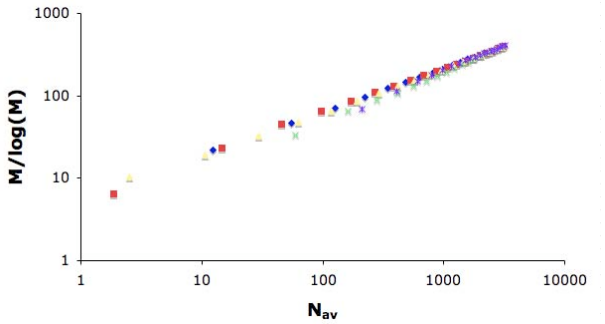

Fig. 1. This log-log graph represents the power law scaling (linearity) predicted in Eq. (1). The data used was temporal microarray data from yeast. Each line (color) represents a different time slice of the data, varied by starting time and size G. To generate the lines 19 different values of C ranging from .95 to .05 in .05 increments were used.

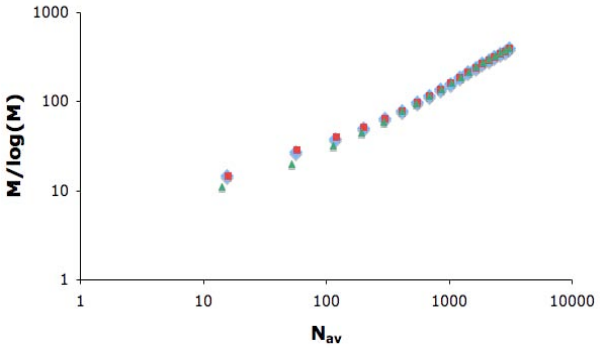

Fig. 2. For this figure we used an ensemble of 6-gene groups (G=6) and static gene expression values. The blue and the red markings represent versions of the same ensemble where the gene ordering has been randomized. The green marking represents a Monte Carlo model of the same data. This graph gives evidence that Eq. (1) represents an equilibrium for a type of global noise in the (yeast) genome.

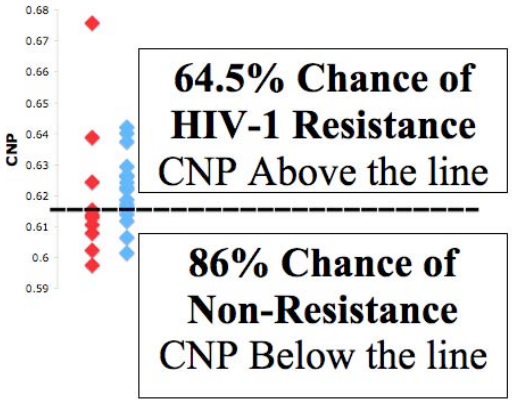

Fig. 3. The blue markings represent CNP values derived from microarray data sampled from 17 subjects identified as HIV-1 resistant. The red markings represent CNP values derived from 10 subjects who were identified as non-resistant. Note that the resistance group is shifted to higher CNP values than the non-resistant group with a clear delineation. The percentages on the graph denoting confidence of HIV-1 resistance versus CNP value are normalized for the difference in sample size.

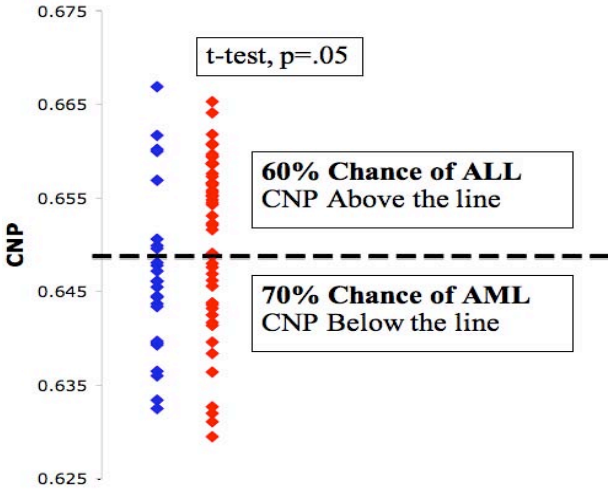

Fig. 4. The red markings represent the CNP values derived from microarray data obtained from 45 subjects identified with Acute Lymphoblastic Leukemia (ALL). The blue markings represent the CNP values from 25 subjects identified with Acute Myeloid Leukemia (AML). The two groupings are shifted relative to each other (ttest, p ~.05) and there is delineation. The percentages on the graph are normalized for sample size.

4. CONCLUSIONS

There is evidence that CNP is a single parameter measurement giving characteristic information about the genome and its relation to phenotype. CNP could be of particular importance in diseases that have been linked to genetic noise, such as AIDS.

References

1. Raj A, van Oudenaarden A. Nature, nurture, or chance: stochastic gene expression and its consequences. *Cell* 2008; **135**: 216.
2. Wang X, Guan S, Lai YC, Li B, Lai CH. Desynchronization and on-off intermittency in complex networks. *Europhysics Letters* 2009; **88**: 28001+**.**
3. Luque B, Miramontes O, Lacasa L. Number Theoretic Example of Scale-Free Topology Inducing Self-Organized Criticality. *Physical Review Letters* 2008; **101**: 158702**.**
4. Singh A, Razooky B, Cox CD, Simpson ML, Weinberger LS. Transcriptional Bursting from the HIV-1 Promoter Is A Significant Source of Stochastic Noise in HIV-1 Gene Expression. *Biophysical Journal* 2010; **98**: L32**.**
5. Albert R. Scale-free networks in cell biology. *J Cell Sci* 2005; **118**: 4847**.**
6. Shaw S. Evidence of Scale-Free Topology and Dynamics in Gene Regulatory Networks. *12th International Conference on Intelligent and Adaptive Systems and Software Engineering* 2003; 4847**.**