

A small regulatory core that discriminates the gene expression profiles of cancer points to a 'Medusa' structure of the genomic network

Over the past decades, researchers seeking to understand molecular mechanisms underlying various diseases, notably cancer, have taken advantage of DNA microarrays to interrogate tissue specimen of patients for the expression status of thousands of genes at once. Jointly, such gene expression status of each gene in the genome, measured as the level of their transcripts, constitutes the gene expression profile. Since each of the tens of thousands of genes can be switched on or off, a gene expression profile contains complex information, akin to a huge bar code with tens of thousands of digits for every sample. While microarray data were initially used by gene hunters to identify novel genes, such as those which are only active in samples of particular cancer tissues, researchers have later learned to employ sophisticated computational tools to classify these bar codes into subgroups and to find subgroup-specific signatures. In cancer research such statistical analysis of gene expression patterns can serve to identify new cancer subtypes and help classify patients more accurately. However, there is only so much that such brute force computational pattern recognition can offer. Biologists also would like to understand: Where the particular gene expression pattern comes from? How does the cell know how to 'write' the long bar code, defining the expression level of gene after gene, across tens of thousands of genes, in such a reliable manner to encode cell types and cancer cell phenotypes? Although not often asked by computational biologists using statistical analysis to extract information, this is a central and basic biological question.

A team of researchers at Harvard Medical School's Children's Hospital led by Sui Huang (who is now at the University of Calgary) have analyzed gene expression profiles with precisely this question in mind. In the work published on page 628 of this issue of *Experimental Biology and Medicine*, Dr Huang and his students, Guo, Feng and Trivedi, offer a first step towards understanding the source of the stable pattern of gene expression profiles by testing whether gene expression profiles are indeed established by a gene regulatory network that has the structure of a 'medusa,' with a command and control 'head' and an enslaved periphery, as proposed by theoreticians.

'We tend to take gene expression profiles for granted – much like the forensic examiner looks at finger prints without ever asking how they are produced in development,' Dr Huang says. The expression of a gene is regulated by particular types of proteins, the transcription factors (TFs), of which there are 2000 or so in the human genome.

Thus, obviously, the entire gene expression profile, the tens of thousands digit bar code, is determined by the collective activity of these TFs. Since they also control the expression of each other, this subset of TF genes forms a 'core network' of mutual regulation. In addition, they must also control the 'non-TF' work horse proteins of the cell, such as cytoskeletal proteins or metabolic enzymes which are also regulated by TFs (as are all genes) but do not regulate the expression of other genes. In this elementary picture, the pattern of the gene expression bar code would be determined essentially by the core network which represents the medusa head and controls the peripheral, regulated but not-regulating genes, the medusa arms (tentacles).

If the entire gene expression profile, the bar code that characterizes the phenotype of cell types, is controlled by the core of a just few thousands genes rather than the entire genome of tens of thousands of genes, then as Dr Huang explains 'this would have practical consequences beyond theoretical biology, for it would facilitate gene expression pattern based disease characterisation and diagnosis by allowing efficient computation focused on the regulatory core.'

Huang's team has now used a set of gene expression profiles of lung cancer tissues from a group of patients to show that the expression patterns are consistent with a medusa network. They found using various statistical tests that instead of the entire set of almost 10,000 genes available on the DNA microarrays less than a thousand TF genes were sufficient to classify the patient lung cancer samples according to the diagnosed cancer types. The subset of a few hundred TFs performed as well or better than the set of nearly 10,000 genes that represent much of the genome. The effect persisted after correction for gene number and expression levels. Conversely, metabolic genes that would correspond to the subordinate arms of the medusa, and hence should have minimal influence on the gene expression profile, performed most poorly in the same comparison. Interestingly, Huang and his group also found that microRNAs, a class of regulatory transcripts that do not encode for proteins but contain nucleotide sequence complementary to protein coding transcripts that allow them to specifically target the latter and prevent translation into proteins, were even more powerful than TFs. Since microRNAs are part of the regulatory core, this was not entirely surprising. But why did they perform so much better than TFs? As Huang explains, continuing to draw the analogy of

molecular regulation to computing the gene expression patterns, microRNAs act as 'canalizing Boolean functions' of networks – an old idea first proposed by Stuart Kauffman in the 1970s. MicroRNAs, through their action at the post-transcriptional level, override the input of other regulators. Their occurrence suppresses what in network theory is called chaotic dynamics and allows the network to produce multiple stable gene expression patterns capable of self-organization, so-called attractor states. The coexistence of many such stable attractor states is the very basis of multicellularity and of cancer, since these cell phenotypes possess distinct stable gene expression profiles. Perhaps it is

no coincidence that microRNAs appeared in evolution when gene regulatory networks became highly complex and just before the emergence of multicellular organisms.

Dr Steven Goodman, Editor-in-Chief of *Experimental Biology and Medicine*, said 'The work by Huang and colleagues supporting the organizing role of the Medusa network suggests that gene expression profile interpretation can be performed in the context of our increased understanding of the relationship between the expression profile and the underlying network.'

DOI: 10.1258/ebm.2011.011f05