



Boyd, Scott D.

[28]

## Targets of the p53-related transcription factors

Scott D. Boyd<sup>1,2</sup>, Elsa R. Flores<sup>3</sup> & Tyler Jacks<sup>1,2</sup><sup>1</sup>Howard Hughes Medical Institute and Department of Biology, MIT Center for Cancer Research, Cambridge, Massachusetts, USA<sup>2</sup>Harvard Medical School, Cambridge, Massachusetts, USA<sup>3</sup>Department of Biology, MIT Center for Cancer Research, Cambridge, Massachusetts, USA

The p53 tumor suppressor protein is the most intensively studied member of a family of three related transcription factors in mammals. The protein and its homologues, p63 and p73, share a common domain structure, with particularly high sequence homology in the DNA-binding domain. Most of the DNA-contacting residues in p53 are conserved in p63 and p73, and all three proteins have been shown to bind to similar DNA sequences. However, the phenotypes of mutant mice carrying inactivated alleles of p53, p63 or p73 indicate that these genes carry out highly distinct functions: p63- and p73-null mice exhibit pleiotropic defects in embryonic development, whereas p53 deficiency results in tumorigenesis in early adult life. The basis for this specificity of action is unclear. We are conducting a series of experiments to determine the spectrum of genes regulated by each of the p53 family members. We have generated cell lines carrying inducible constructs for expression of each of these transcription factors, and we are assaying them with oligonucleotide chip arrays. In addition, we are measuring changes in gene expression in cells derived from mutant mice. Better understanding of the transcriptional targets of the p53-related genes may indicate which downstream genes are specific to the tumor-suppressing activity of p53 and provide insights into the structural basis for target gene specificity.

Brenner, Charles

[29]

## Genomic approaches to elucidation of the Fhit pathway in worms and yeast: Rosetta Stone and synthetic lethals

Charles Brenner

Kimmel Cancer Center, Thomas Jefferson University, Philadelphia, Pennsylvania 19107, USA

The Fhit protein is lost early in the development of many human cancers. Loss of Fhit in the mouse predisposes the organism to an array of tumors, including stomach and sebaceous carcinomas that resemble human Muir-Torre syndrome. Re-expression of Fhit in cancer cells unable to synthesize the protein induces apoptosis by an unknown mechanism. We have taken two orthogonal genomic approaches to identify additional proteins in the Fhit pathway. First, by cloning Fhit homologues from a wide variety of genomes, we discovered that Fhit is made as a natural fusion or Rosetta Stone protein in invertebrates. We will show phylogenetic profiles, expression analysis, and the crystal structure of the 200-kD NitFhit tetramer from worms to demonstrate that the fusion partners Nit and Fhit have a conserved interaction. Second, we have developed a new method by which an entire genome can be interrogated for synthetic genetic interactions. Synthetic lethals are a valuable means of finding interacting proteins because, beyond identifying the proteins in a pathway, the product of a gene that is synthetically lethal with loss of Fhit is arguably the best drug target for specific elimination of cancer cells unable to synthesize Fhit. We will demonstrate how we use array hybridization to conduct a genome-wide synthetic screen in yeast and give an example of a conserved kinase pathway that is synthetically lethal with mutations in a yeast histidine triad protein.

Butte, Atul J.

[30]

## Analyzing functional genomic differences yields oncogenes and chromosomal breakpoints in ALL and AML

Atul J. Butte, Ben Reis, Alvin Kho, Yao Sun &amp; Isaac S. Kohane

Informatics Program, Children's Hospital, Boston, Massachusetts 02115, USA

The promise of functional genomics is to find the genes responsible for phenotypes by carefully analyzing patterns of expression. Calculation of the fold-differences in gene expression between two diseases has been used to find those genes whose expression successfully differentiates the diseases. However, generating a list of independent genes with differential expression in two diseases does not necessarily help determine how regulatory differences contribute to disease etiology. Genes with the highest differences in expression levels may represent dissimilarities at levels far downstream, compared with subtle changes or no changes at all in expression levels in the central controlling genes. Focusing on differences in models of genetic regulatory networks in acute lymphocytic leukemia and acute myelogenous leukemia, rather than on differences in expression levels, yields a short list of well-known oncogenes, genes known to be involved in leukemogenesis, genes near known chromosomal breakpoints associated with leukemia and genes with unknown but possibly related functions.

Cano, Pedro

[31]

## Information theory in the analysis of gene expression microarray data

Pedro Cano

University of Texas M. D. Anderson Cancer Center, Houston, Texas 77030, USA

The use of propositional logic in the analysis of gene expression microarray data enables the interrelations of gene expression variables to be used in answering biological or clinical questions. The main objection to this approach has been that the range of values of expression for a given gene is artificially reduced to the extremes of "high" and "low" expression, losing important analytical detail. Propositional calculus demands such bipolarity; nevertheless, propositional functions are best evaluated by means of  $2 \times 2$  contingency tables, and it is precisely in this evaluation that the analytical detail can be recovered by applying the principles of information theory. I introduce a method for evaluating  $2 \times 2$  contingency tables, not counting the number of cases, as is regularly done, but measuring and adding the amount of information for the cases. In this way each gene expression datum has two components: (1) a bipolar value ("high" or "low" expression) and (2) the amount of information (the decrease in uncertainty provided by the message, that is, the change in entropy). The first component is used to build propositional functions, and the second is used to evaluate the predictive power of these functions. The measurement of this second component is based on the principle that the amount of information for a variable with a particular value is inversely related to the probability of obtaining that value. Application of this method to empirical data reveals relations that had been lost in the process of bipolar data conversion.