SMP: RNA

Project: Systematic Analysis of Functional Networks and Key Mediators

Holger Sültmann - German Cancer Research Center (DKFZ), Heidelberg - h.sueltmann@dkfz.de

Introduction

Global approaches such as gene expression profiling deliver molecules that represent potential targets for therapeutic intervention. It is generally accepted that complex interdependencies between deregulated genes and proteins cause the respective disease phenotype. Therefore, the functional validation of candidate genes in appropriate *in vitro* cellular systems is an imperative step towards the identification of potential drug targets. Since biological processes occur in networks of interactions between genes, proteins and organelles, the identification of interaction partners is a crucial step towards fully understanding these processes.

In this project, we generate whole transcriptome gene expression profiles from cells after perturbation of certain genes, which were found to be relevant to human diseases. In this way, we broaden the "gene expression space" of disease-relevant genes by identifying their downstream effects. Modeling of the correlations between the expression of genes in these networks will lead to the identification of central "nodes", i. e. novel key elements in the biological processes relevant to the disease. This will result in improved understanding of the biological processes associated with the disease and may provide novel candidate targets for future therapies.

Results/Project Status

We analyze the effects of gene induction and suppression by whole genome cDNA microarray expression profiling of transfected cells. The selection of gene sets and the cell lines to be used for this study are based on the results obtained from previous (NGFN-1) as well as current gene expression profiling experiments (SMP-RNA) and functional analyses (SMP-Cell).

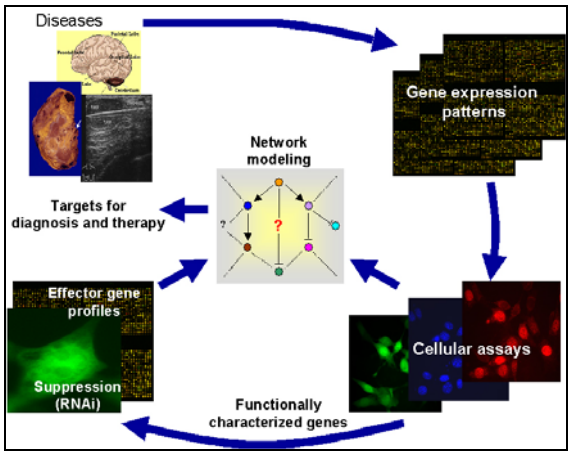

Fig 1: *Creating gene networks by functional characterization and gene expression analysis. Expression profiling of clinically relevant entities and disease stages is performed in collaboration with clinical partners from KGs. Prioritized candidate genes are systematically investigated with regard to their functions (e. g. proliferation, apoptosis, chemoresistance, calcium-signalling) in the SMP-Cell. Expression profiling of transfected cells yields signatures which are associated with the prioritized genes. Construction of gene expression networks using modeling algorithms provides clues to the coregulation of genes in regulatory networks.*

Genes are silenced by transfection of cells with siRNA, and the degree of silencing is measured by quantitative RT-PCR. RNA is isolated from the cells, and labelled cDNA is used for hybridization on whole genome cDNA microarrays. Comparison with control transfections yields genes which are associated with the gene expression status of the silenced gene. Complementary to gene silencing, we are using overexpression of the corresponding proteins after transfection with an ORF and measure the respective gene activity. The results are used for the modeling of gene expression networks (Figure 1).

Strategy

In the first phase of the project, we have established and optimized methodological parameters, such as:

- Determination of appropriate cell line(s) with respect to the biological process that is investigated
- Determination of appropriate cell line(s) with respect to transfection and gene silencing efficiency
- Time course experiments to determine the optimal time point for sampling with respect to protein turnover
- Concentration variation experiments to determine the optimal concentration of siRNA for transfection
- Analysis of control systems (e.g. p53) in pilot studies
- Establishment of a universal experimental design scheme w. r. t. transfections and microarray experiments.

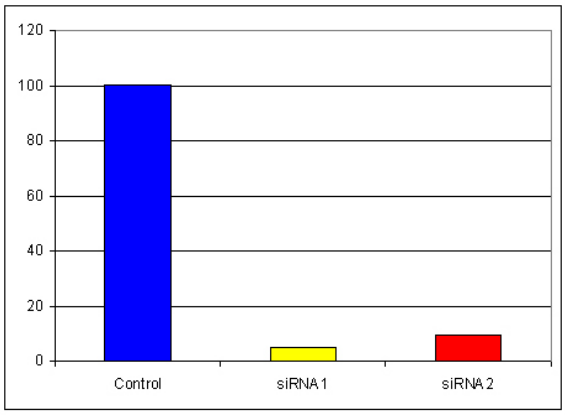

Fig. 2: *Silencing efficiency of p53 in HeLa cells. Two different siRNAs were used to silence p53. A control siRNA was used in parallel. p53 gene expression was then measured using quantitative RT-PCR. Residual p53 expression is less than 5-10%, depending on the transfected siRNA.*

After successfully establishing these parameters (Figure 2), in the second phase of the project we are focussing on selected biological processes with relevance in diseases, particularly in cancer. The basis for these experiments is provided by our own whole genome microarray studies in breast, kidney, and brain cancer as well as data from the literature. Prioritized genes will be knocked down by specific siRNAs, and the global gene expression pattern of the perturbed cells will be measured by microarray analysis.

Modeling of gene expression networks

The primary goal of the project is the identification of disease-relevant genes and the determination of their interactions. In this respect, the focus of the project is on the development and implementation of novel methods for the reconstruction of biological networks. To this end, we are developing novel methods for network construction from microarray data. Here, the major challenge is to distinguish between direct interactions of two genes and those that are mediated by third players. To account for the inherent experimental noise level, we address this problem with a probabilistic method for the detection of gene regulation. A Bayesian statistics is used to infer the probability of the presence of a direct interaction between each pair of genes, and to control the false discovery rate. Subsequently, a pruning algorithm decides whether a detour regulatory pathway is more likely than the direct interaction. This leads to a significant increase of the positive predictive value for proposed edges (which is, from an experimentalist's point of view, the most important quantity for validation purposes). A first attempt to reconstruct the edges purely from microarray data is shown in Figure 3. Further – and more sophisticated – procedures for network modeling and data visualization are being developed.

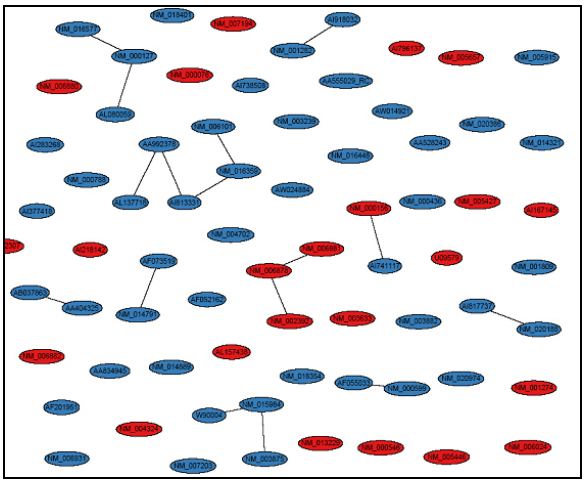

Fig. 3: *Modeling of a gene interaction graph from microarray data. Red and blue ellipses (nodes) indicate up- and downregulated genes, respectively. Lines denote edges between the nodes, i.e. partial correlation of genes on the transcriptional level.*

Outlook

Comparison of the data obtained from the transfectants combined with mathematical modeling will enable the establishment of previously unknown interdependencies between genes and allow for a functional annotation of genes based on their presence in larger networks. Genes that are deregulated in the respective cellular systems will immediately be funneled into a further round of transfection and gene expression profiling. With the iteration of this experimental process, our knowlegde on the functional and transcriptional networks underlying cancer and other diseases will rapidly increase. Finally, the knowledge about synergism and antagonism of the genes will become evident, and this will provide criteria for the selection of novel and more promising targets for future drug development.