Reconstruction of transformation events in lung cancer on gene expression level

Ruzanov P, Varhol R, Zuyderduyn S, Jones S

British Columbia Cancer Agency
Vancouver, British Columbia, Canada

Genome Sciences Centre

www.bcgsc.ca
info@bcgsc.ca

Introduction

This poster presents the results of our work on reconstruction of transformation events in lung cancer. The ability to circumscribe the list of the genes, characteristic for different stages of tumorogenesis, would allow to alleviate the process of extracting 'mission critical' events, because it narrows the list of possible oncogenic triggers. We used the data for the SAGE (Serial Analysis of Gene Expression) libraries, generated from samples, obtained from patients with lung cancer (6 libraries, including one generated from invasive carcinoma sample) and normal lung (3 libraries). Using the phenetic tree building approach we built a diagram, visualizing the relationship between different lung SAGE libraries. As far as we had the information about the biological nature of the samples, we could say that the resulting tree was biologically meaningful and might be used for further analysis.
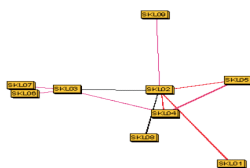
Phenetic Tree Building Technique

A

XY	SKL01	SKL02	SKL03	SKL04	SKL05	SKL06	SKL07	SKL08	SKL09
SKL01	X	313	574	374	584	796	776	456	637
SKL02		X	324	174	216	387	392	227	344
SKL03			X	363	492	111	124	401	479
SKL04				X	321	444	482	216	503
SKL05					X	665	722	396	619
SKL06						X	32	510	645
SKL07							X	539	602
SKL08								X	512
SKL09									X

B

(A) distance matrix, generated using Chi-square test of SAGE data. Values in the table show the number of differently expressed genes for each possible pair of SAGE libraries. These numbers were used as a 'distance' between libraries when the phenetic tree was built. (B) Diagram, produced with the aid of a small Java application, developed by authors. The application takes the values, representing the distance between libraries from distance matrix, and then tries to relax the diagram, minimizing the tension between all the nodes. Rectangles represent the SAGE libraries, used in analysis. Libraries SKL03, SKL06 and SKL07 are prepared from normal lung, while the rest of the libraries - from malignant tissue. Sample, used for preparation of SKL05 library, was characterized as invasive form of lung carcinoma. The resulting diagram was used for extracting the chain of transformation events with the consequent analysis of genes, which expression affects different stages of tumorogenesis in lung.


Here, the screenshots of tree-building application are shown. Application allows to load the data from a text file, and than – visualize the relationships between analyzed data sets.

Phenetic trees allow to build hierarchical diagrams, which allow to see the relationships between different data sets, in our case – SAGE libraries. Diagrams were produced with the aid of a small Java application, which allows interactively adjust the structure of the tree in order to achieve the optimal organization. The application takes the values representing the distance between libraries from distance matrix (shown above), and then tries to relax the diagram, minimizing the tension between all the nodes.

## Deriving Waves of Transformation Events
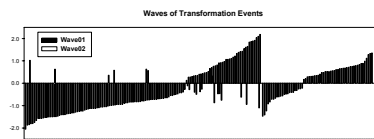
Next, we identified the genes, significantly changed during cell transformation. Apparently, such genes are characteristic for malignant cell transformation and represent the potential drug targets and transformation markers. As far as in the case of phenetic trees the evolutionary events might be reconstructed, in the case of expression data it is possible to extract the alterations in expression profile, which drive the cellular processes such as malignant transformation.

We chose to extract the lists of genes with expression level significantly changed during two consequent stages of transformation: a) transformation from normal tissue to carcinoma and b) transformation of carcinoma into invasive form. Totally 194 genes with UniGene id were extracted, 132 for normal-carcinoma transition and 82 (of which 20 genes from the first list) for transformation of carcinoma into invasive phenotype

Annotation of genes using GO terms

For the characterization of the derived genes we used Gene Ontology (GO) terms. Gene Ontology Consortium is developing a dynamic controlled vocabulary that can be used for gene annotation.GO vocabulary has hierarchical structure, allowing to annotate genes using a range of detail level from a general (or high-level) to very descriptive low-level, highly specific definition. As the GO terms are a set of nodes, organized in a rooted acyclic graph, it could be treated as a tree, where the parent nodes provide more general definitions than those of its children.
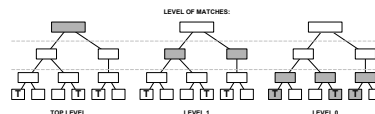
Above, shown the scheme of counting the matches between genes on different levels, using GO terms. 'T' stands for terminal GO entries (without any descendants)

Below, sorting of the derived genes. (Top panel) Sorting genes by the GO terms, using *molecular function* and (Bottom Panel) *biological process* GOsubsets. Numbers of genes, having at least one match at certain level of similarity are shown along the axis. Wave01 and Wave02 represent two stages of malignant transformation: from normal to carcinoma (non-invasive) and from carcinoma to invasive-stage carcinoma.
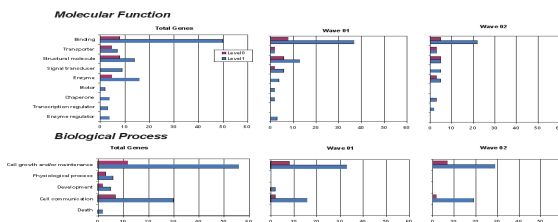
Acknowledgements and References

Planet, P.J., DeSalle, R., Siddal, M., Bael, T., Sarkar, I.N. and Stanley, S.E. (2001) Systematic Analysis of DNA Microarray Data: Ordering and Interpreting Patterns of Gene Expression. *Genome Research*, **11**, 1149-1155.
John Avise, C. (1994) *Methods of phenetic tree building*. Chapman & Hall, New York London.