Network: Systems Biology of Embryonal Tumors – Neuroblastoma as Model

Project: Bioinformatics/Data Management

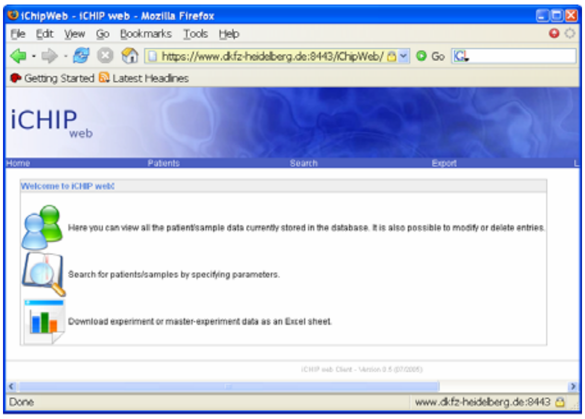
Roland Eils - German Cancer Research Center (DKFZ), Heidelberg - r.eils@dkfz.de

Introduction

Novel molecular data like those from gene expression profiling by DNA microarrays, or from protein profiling by mass spectrometry, are high-dimensional and require special methods both with regard to storage and to biostatistical analysis. Our database application for the maintenance of microarray data, iCHIP (http://www.dkfz.de/ibios/ngfnServices), is installed at several locations within the clinical networks. In addition, it has also been established as a local decentralized version. The conceptual design of iCHIP is in accordance with the MIAME- and MAGE-standards of the international MGED consortium. Originally iCHIP was developed to function as a gene expression database suitable for data generated from DNA array screening, using both oligonucleotides and cDNAs. iCHIP has been flexibly and modularly constructed, making iCHIP very much an analysis toolbox.

With regard to analysis, we have developed workflows to deal with the high dimensionality and low sample numbers. Such workflows, based on nested cross-validation, have been implemented in an Bioconductor package (1) based on the statistical environment R (www.r-project.org). A typical application is classification analysis, where patterns are determined that allow for accurate prediction of a given class distinction. That is, given a set of training data of known classes, algorithms like support vector machines are trained to predict these classes in new data sets. The methods may be combined with feature selection procedures like recursive feature elimination (2) to obtain comprehensive patterns with high discriminative power.

Typical questions to address in the context of neuroblastoma would be, e.g., which genes change expression in samples with high TrkA expression compared to those with high TrkB expression, or which genes are associated with good or bad prognosis, or which genes are associated with established markers like N-Myc amplification, advanced age or deletion of 1p.


iChipWeb - iCHIP web - Mozilla Firefox
File Edit View Go Bookmarks Tools Help
Go
Getting Started
Latest Headlines
iCHIP web
Home
Patients
Search
Export
Done

https://www.dkfz-heidelberg.de:8443/iChipWeb/

Welcome to iCHIP web!

Here you can view all the patient/sample data currently stored in the database. It is also possible to modify or delete entries.

Search for patients/samples by specifying parameters.

Download experiment or master-experiment data as an Excel sheet.

iCHIP web Client - Version 0.5 (07/2005)

www.dkfz-heidelberg.de:8443

Fig 1: *Screenshot of iCHIP database web frontend.*

Results/Project Status

Extension of iCHIP database

The iCHIP database was originially designed to store data from gene expression profiling studies based on cDNA microarray or oligonucleotide array technology. We have extended the database to be able to deal with tissue microarray, single nucleotide polymorphism and proteome profiling data based on SELDI technology. The database is already partially filled with such data, including the most relevant associated clinical data of those patients whose tumor specimens have been investigated. Access restrictions and security guidelines prevent unauthorized access to the data, which are stored in anonymized form. A web-based frontend to the database exists to facilitate interaction with the database (Fig. 1). Export to flat files and reports are available.

Analysis of gene expression data

We have applied our workflows to gene expression data from neuroblastoma to find differntially expressed genes associated with established clinical markers (3) and to find patterns that are associated with prognosis (Fig.2 and ref. 4). Prediction of early relapse was possible with an accuracy of 85% when using support vector machines as predictors of early relapse (4). Since others have also presented classification analyses in neuroblastoma (5,6), we have started to analyse the commonalities in the different studies. One approach was to construct a custom neuroblastoma array that integrates data from cDNA arrays, oligonucleotide arrays, SAGE, literature search, frequently affected genomic regions and genes important for analysis of signal transduction networks. This custom array has been tested on a retrospective series of tumors (manuscript in preparation) and will be validated by prospective analysis on a large cohort of patients during the next years.

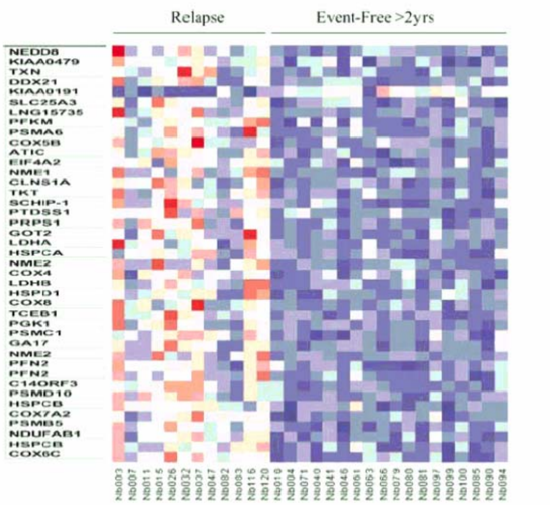


Fig 2: *Gene expression pattern from (4) which discriminates between early relapse (< 2 years) and event-free survival for langer than 2 years. Gene expression is coded on a color scale, red: high, blue: low expression. Genes shown were obtained by Prediction Analysis of Microarrays (PAM), the top 39 genes are shown. For details, see (4).*

Analysis of proteome profiles from serum

Within the work package "Proteomics and Methylation", serum protein profiles have been obtained by surface-enhanced laser desorption/ionization mass spectrometry. We have developed a preprocessing workflow ("TBI workflow") to perform baseline correction, spectrum normalization, denoising, peak detection and inter-spectrum peak alignment (Fig. 3). This workflow is implemented in R/Bioconductor and

GEFÖRDERT VOM
Bundesministerium für Bildung und Forschung

will be provided as a Bioconductor package. Resulting peak lists from spectra are used in a biomarker discovery analysis involving support vector machines coupled to recursive feature elimination (2). Preliminary results indicate feasability of the method while accuracies may be still improved. In collaboration with the work package "Proteomics and Methylation", we are checking which surfaces and preceding chromatographic steps are necessary for optimal results in serum protein profiling.

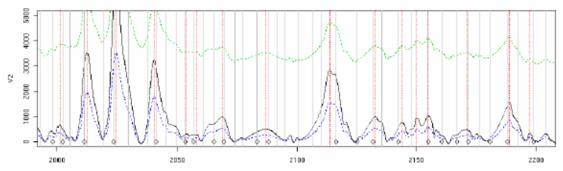


Fig 3: *Surface-enhanced mass spectra from serum proteins after processing by TBI workflow. M/z ratios are on the x-axis, intensities on the y axis. Red dotted lines denote centers of detected peaks, corresponding grey lines the region used for alignment of different spectra.*

Outlook

We will extend the work on the iCHIP database to include more experimental data, better annotation of novel data types (SNP, proteomics data, tissue micorarray data), extended frontends to the database, export of MAGE-ML data, and connection to the integrative bioinformatics system BioRS (BioMax, Munich). With respect to analysis, we will continue our biostatistical analyses on new data arising in other work packages. We will use methods of meta analysis and cross-platform analysis to get a consistent view on all publicly available data together with those of the project. This will lead to an improved understanding of neuroblastoma progression, spontaneous regression, and resistance to therapy. Proteomics data will help to contrast already existing gene expression profiles and get an idea on changes at the protein level. The prognostic study using our custom neuroblastoma-specific microarray will be a first step to introduce this as a novel diagnostic means which may help to improve prognosis for patients, especially of intermediate risk groups which are currently suspected to be under- or overtreated due to lack of reliable prognostic information.

Lit.: ***1.*** *Ruschhaupt M et al. A Compendium to ensure computational reproducibility in high-dimensional classification tasks. Stat Appl Genet Mol Biol. 2004; 3(1):37.* ***2****. Guyon I et al. Gene selection for cancer classification using support vector machines. Machine Learning. 2002; 46:389-422.* ***3.*** *Schulte JH et al. Microarray analysis reveals differential gene expression patterns and regulation of single target genes contributing to the opposing phenotype of TrkA- and TrkB-expressing neuroblastomas. Oncogene. 2005 Jan 6; 24(1):165-177.* ***4.*** *Schramm A et al. Prediction of clinical outcome and biological characterization of neuroblastoma by expression profiling. Oncogene. 2005 Aug 15; [Epub ahead of print].* ***5.*** *Wei JS et al. Prediction of clinical outcome using gene expression profiling and artificial neural networks for patients with neuroblastoma. Cancer Res. 2004 Oct 1;64(19):6883-6891.* ***6.*** *Ohira M et al. Expression profiling using a tumor-specific cDNA microarray predicts the prognosis of intermediate risk neuroblastomas. Cancer Cell. 2005 Apr;7(4):337-350.*

GEFÖRDERT VOM
Bundesministerium für Bildung und Forschung