Gene expression based prognostic classification of solid tumors

Synopsis of PhD thesis

Zsófia Sztupinszki, M.D.

Semmelweis University
Doctoral School of Pathological Sciences

Supervisor: Balázs Győrffy, M.D., Ph.D, D.Sc.

Official reviewers: Rónai Zsolt, Ph.D.
            Tamás Korcsmáros, Ph.D.

Head of the Final Examination Committee:
            Attila Zalatnai, M.D., Ph.D.

Members of the Final Examination Committee:
            Zoltán Szeltner, Ph.D.
            Attila Ambrus, Ph.D.

Budapest
2017

1. Introduction

In my theses I investigate the classification of solid tumors, breast and colorectal cancer, and the identification of patients with high risk or poor prognosis. By predicting the patient's response to a given therapy, the efficacy, cost-effectiveness of the therapy can be maximized and drug loads and side effects can be reduced. Today we does not consider breast and colon tumors homogeneous disease, the different subtypes have different biological and clinical features. Therefore, their identification and characterization allows a deeper understanding of tumorigenesis, the identification of new drug targets and the selection of better therapies. Predictive and prognostic biomarkers in the context of current therapeutic options are important elements of rational therapeutic planning and their widespread use is expected to be have large effect on treatment protocols.

In my dissertation I present the result of four experiments. The topics are the gene expression based classification of colorectal cancer, new prognostic classification of breast cancer, prediction lymph node status of breast cancer patients and evaluating reproducibility of siRNA based gene silencing.

2. Objectives

During my PhD work my primarily interest was identification and validation of biomarkers in high risk subgroups of breast and colon cancer.

My objectives are the following:

1. Comparing classification of colon cancer and identifying the most prognostic one.
2. Identifying cell lines corresponding to the molecular subtypes of colon cancers.
3. New gene expression based classification of breast cancer and comparing its results with previously published prognostic tests.
4. Predicting lymph node involvement in breast cancer based on the primary tumor's gene expression.
5. Comparing reproducibility of combined siRNA and micorarray experiments.

3. Methods

3.1. Colon cancer – identifying poor prognosis patients

In this study, my goal was to evaluate the published molecular subtypes, prognostic transcriptomic signatures of colorectal cancer using the large set of independent patients.

3.1.1. Database construction and preprocessing steps

To create my own database, I searched the public GEO database to identify datasets with gene expression data measured on Affymetrix arrays in colorectal cancer samples. As one of the aims of this study was to reproduce the classifiers, I also their followed the detailed descriptions of the original publications in the preprocessing steps.

3.1.2. Identification of previously published classifiers

In order to identify previously published molecular classifiers of colorectal cancer in the scientific literature I searched the PubMed

(http://www.pubmed.com) database according to PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines.

3.1.3. Comparison of classifiers

In case of the classifiers the hazard ratio between the survival of the worst and the best prognostic groups were compared using Cox proportional hazards regression, the p-values were calculated with logrank test, and the results were plotted using Kaplan-Meier method. Multivariate analyses were carried out using the following prognostic variables: MSI-status, sex, gene expression of MKI67 and CDX2. The results of the classification methods were compared with Cramer-V.

3.1.4. Most central genes

To compare the significance of the utilized genes, for each gene, I computed a score. This score was computed as: "gene score" = [number of classifiers containing the gene]* $\sum$[gene proportion in each classifier].

3.1.5. Preclinical models - cell lines

One of my goals was to assign cell lines to the closest molecular subtypes. Gene expression signatures of colon cancer cell lines were screened in GEO and Array Express and samples with available raw gene chip data using Affymetrix HGU 133 plus 2.0 microarrays were collected.

3.2. Breast cancer – identifying poor prognosis patients

The aim of this study was to develop a new prognostic test, which is able to identity breast cancer patients with bad prognosis.

3.2.1. Independent validation – sample collection

For the independent validation of the model 325 fresh frozen validation samples were collected of early stage breast cancers at the

Departments of Gynecology and Obstetrics at the University Hospitals in Frankfurt and Hamburg, Germany. The gene expression profiling was done with Affymetrix Human Genome U133A microarrays. The raw .CEL files and clinical data have been deposited in the GEO database under the accession numbers GSE4611 (Frankfurt dataset) and GSE46184 (Hamburg dataset).

3.2.2. Database construction and preprocessing steps

The construction of the database was performed similar to the steps described previously. The raw microarrays of the breast cancer samples with sufficient clinical data were normalized with MAS5 algorithm. For predictor building only probe sets that were measured by both HG-U133A and HG-U133 Plus 2.0 arrays (n = 22,277) were used. For genes targeted by multiple probe sets only the JetSet best probe sets was used. The final number of probe sets/genes included in the training database for each case was n = 9,886.

3.2.3. Classification of the patients

In this study the classification consists of two main steps. During the molecular classification in the group of similar patients to the examined sample the prognostic power of all of the genes are calculated. The best genes are selected and the sample is classified according to the average expression of these best genes. In case of the clinical classification the prognosis of the training set is compared to all the remaining patients of the database. The final, consensus classification was based on the combination of the molecular and the clinical classification.

3.2.4. Comparison of the dynamic predictor to previously published classifiers

The new Dynamic Predictor was compared to three previously published multigene test: the 97-gene Genomic Grade Index (GGI), the 70-

gene Mammaprint, and the 21-gene Oncotype DX. These were determined either with the genefu R package or based on former publication of our group. The relapse-free survival of the prognostic groups were compared in all of the patients and in clinical subgroups.

3.3. Breast cancer – prediction of lymph node status

Axillary lymph node involvement is one of the most important prognostic factors in breast cancer. In case of the lymph node negative cases the biopsy of the sentinel lymph nodes does not have a therapeutic effect. In this experiment my aim was to predict the lymph node status based on the primary tumor's gene expression signature.

3.3.1. Sample collection for independent validation

As a part of an international collaboration formalin fixed paraffin embedded sample of breast cancer patients operated between 2004 and 2010, and corresponding clinical data were collected from the biobank of the Atossa company (Seattle, USA).

3.3.2. Database construction and preprocessing steps

The database was established similar to the previously described way. I screened the GEO for experiments using HG-U133 Plus 2.0 or HG-U133A microarrays in breast cancer, where clinical information about lymph node involvement is available. In this study I excluded all of the patients who received primary systemic (neoadjuvant) therapy, because this treatment may result in downstaging of the axillary region.

The data was fRMA normalized with an alternative annotation. For the alternative annotation I used probe sets which bind to the 5' end of the transcript, as in the FFPE validation samples high degree of RNA degradation is expected. Following additional filtering steps the best probes were selected

with the JetSet method. The further analysis was carried out using 9462 genes.

3.3.3. Classification of the samples

The patients were enrolled into three groups: ER-negative, ER-positive / MKI67-positive, and ER-positive / MKI67-negative cohort. In the interest of proper teaching and validation of the model one half of the patients were randomly assigned to the training set, while the others were assigned to the internal validation set. The differently expressed genes between the lymph node negative and positive groups were identified using RankProducts algorithm. These genes were used in the boosted random forest model separately for each of the cohort with the caret R package.

3.4. Comparison of reproducibility of biomarkers

Validation of transcription effects of biomarkers are important experiments of oncology research The effectiveness of gene silencing with small interfering RNA (siRNA) may differ between cell line and therefore, I aimed to evaluate the silencing efficacy in experiments where gene silencing with siRNA and gene expression microarrays were combined.

In the first step I used the GEOquery R package to identify studies where silencing with siRNA was carried out in cancer cell lines and the gene expression were measured using Affymetrix HG-U133A or HG-U133Plus 2.0 microarrays. Further manual curation of the hits was done. The raw data was normalized with MAS5 algorithm and the best probe sets were selected with the JetSet method. The efficacy of the silencing was evaluated with t-test and fold-change, where fold-change= (expression of target gene [silenced])/ (expression of target gene [control])

4. Results

4.1. Colon cancer – identifying poor prognosis patients

4.1.1. Database

The final database consists of gene expression and clinical data of 2166 patient from 12 datasets. Relapse-free survival data were available for 1405 cases. The vast majority of the patients, 74%, are from stage 2 or 3 disease. This is really important, because prognostic tests have the highest clinical significance in this group of patients.

4.1.2. Classification algorithms

After screening 282 papers I could reproduce 22 classifiers. Reproducibility was mostly hampered by incomplete documentation and lack of availability of training sets and clinical data.

4.1.3. Comparison of classifiers

For the stage 2 and 3 patients Yuen et al.'s classification – using only gene expression of three genes – had the highest prognostic value (HR=2.9). The second best was Marisa's classifier (HR=2.60), and the third most efficient method was the Chang95 (HR=2.35). It is important to point out that the results of commercially available tests showed low level of association, for example the Cramer-index of association between the Oncotype DX and the ColoGuideEx was only 0.03.

4.1.4. Comparison of used genes

The 22 compared classifiers utilized altogether 2001 genes. Only five genes (REG4, ASCL2, VAV3, C10orf99 and CYPB1) were included in at least six classifications. According to the gene-score CTGF, GADD45B, FAP genes are the most important ones.

4.1.5. Preclinical models

From the CCLE (Cancer Cell Line Encyclopedia), Cancer Cell Line Project, GSE8332, and GSE32474 datasets 151 gene arrays from 61 unique cell lines were collected. For each of the cell lines the closest molecular subtypes were determined. Compared to the classification of primary tumors it is evident that some of the classifiers are able to assign subtypes to the cell lines, whereas other are not.

4.2. Breast cancer– identifying poor prognosis patients

4.2.1. Sample collection for independent validation

In case of the independent validation cohort of 325 patients, the average follow-up time was 66 months. 81% of these patients were ER-positive, 39% had lymph node metastasis.

4.2.2. Database

In the GEO database I could identify 3534 samples in 22 datasets with raw gene expression measurement and relapse-free survival time.

4.2.3. Comparison with other tests

In comparison with the three multigene tests, comparing the best and worst prognostic groups the new Dynamic Predictor performed the best in the group of all patients (HR=3.68), and only 40% of the patients were classified into the poor prognostic group. Both in the cohort of ER-positive, HER2-negative untreated cases and patients with history of adjuvant therapy the Dynamic Classifier proved to be the most efficient one (HR=4.61 and HR=4.51). In the ER- and HER2-negative treated group the old prognostic tests were not able to distinguish bad and good prognostic groups, whereas the Dynamic Classifier worked well also in this cohort (HR=3.0). Considering the 5-year relapse-free survival status as an endpoint the 70-gene

test was the most sensitive one, although its specificity was really low. The Dynamic Classifier had the highest specificity and the positive predictive value, while its sensitivity and the negative predictive value remains acceptable.

4.2.4. Independent validation

In case all of cases of the 325 independent samples the Dynamic Classifier was the best (HR=3.02). In the ER-positive, lymph node negative cohort only the 21-gene test (HR=3.0) and the Dynamic Prediction (HR=2.21) were able to classify the patients, but only the result of the Dynamic Classifier was significant.

4.3. Breast cancer – predicting lymph node status

4.3.1. Database

In the GEO database from 16 datasets data for 2341 patients were collected. The 21% of the patients were lymph node positive and 16% were ER-negative.

4.3.2. Predicting lymph node status

The performance of the prediction model was evaluated in of the internal validation set and the cohort of the 100 patients. In the internal validation group both in the ER-negative (NPV=0.85) and in the ER-positive / MKI67-positive patients (NPV=0.78) negative predictive value were high, while the accuracy was above 75%. In the case of the formalin fixed paraffin embedded independent validation samples the negative predictive values were 0.92 and 1.0, that is, patients who were predicted not to have lymph node involvement most likely had no lymph node metastasis.

4.4. Comparison of reproducibility of biomarkers

In the GEO database for 8 cell lines silencing studies for 15 genes were identified. A total of 441 microarrays (289 siRNA-treated, 152 control samples) were further analyzed. After MAS5 normalization the expression of the target gene was compared between the siRNA treated and the control samples. The silencing was not effective in three cases: HeLa cell line CTNBB1 gene, MCF7 line CTNBB1 gene, IMR32 line CHAF1A gene. This result also demonstrates that, besides q-PCR, Western blot validation, it is also important to check the expression of the target gene on the gene chips.

5. Conclusions

1. Based on my research regarding colorectal cancer by analyzing microarray and clinical data of 2,166 patients, the highest efficacy to predict progression free survival in stage II-III patients was achieved by the 3-gene Yuen (HR = 2.9) classifier. The second best is Marisa's algorithm (HR=2.60), which is also the best one when investigating all patients regardless of stage (HR=3.20). It is an important observation that even in case of classifiers with good performance overlap between the patients of the same (poor or good) prognostic groups were small.

2. For each of the subtypes the best preclinical models were determined based on the analysis of 151 microarrays from 61 unique colon cancer cell lines. Based on my results it is clear that some of the classifiers are not able to assign molecular subtypes to the cell lines.

3. For breast cancer patients – based on the data of 3524 patients – I developed a new Dynamic Predictor, which utilizes both the prognosis of patients with similar gene expression profiles to the investigated sample and the gene expression signature of the sample. Its performance was compared

to three previously published multigene prognostic tests (Oncotype DX, Mammaprint, Genomic Grade Index). Our prognostic discrimination was the highest for all cases (Dynamic Predictor: HR=3.2, p=7.0*10^{-54}, Oncotype DX HR=1,4, p=4,3*10^{-39}, Mammaprint: HR=3,4, p=1,5*10^{-15}, Genomic Grade Index HR=2,2, p=2,2*10^{-38}). In the clinical subtypes the Dynamic Classifier also outperformed the others. None of the previous test were able to classify ER-, HER2-negative, pretreated patients, while the performance of the Dynamic Classifier was HR=3,9, p=4,8*10^{-4}. The model was also validated in 325 independent cases, and outperformed the three multigene classifiers.

4. Prediction of lymph node metastases based on the primary tumor's gene expression was developed using data from 2341 patients. The model uses boosted random forest classification, and was validated in an independent cohort of 100 patients. In the internal validation set of ER-negative patients the predictor achieved good accuracy (ACC): 85% and negative predictive value (NPV): 88%. For the ER-positive / MKI67-positive group accuracy was 90% and NPV was 77%. In case of the independent validation set the prediction model performed also well: for the ER-negative cohort: ACC=0.73, NPV=0.92, and for the ER-positive and MKI67-positive group: ACC=0.86, NPV=1.0.

5. On the basis of evaluation of siRNA gene silencing and microarray experiments, after comparing the gene chip-based microarray profiles before and after silencing, it can be concluded that microarray based assays are reliable methods for examining the effect of gene silencing.

6. Bibliography of the candidate's publications

6.1. Publications related to the thesis:

Sztupinszki, Z., B. Gyorffy. (2016) Colon cancer subtypes: concordance, effect on survival and selection of the most representative preclinical models. Sci Rep, 6: 37169.   IF=5,228

Munkacsy, G.*, **Z. Sztupinszki***, P. Herman, B. Ban, Z. Penzvalto, N. SzarvasB. Gyorffy. (2016) Validation of RNAi Silencing Efficiency Using Gene Array Data shows 18.5% Failure Rate across 429 Independent Experiments. Mol Ther Nucleic Acids, 5: e366   IF=5,048

Gyorffy, B., T. Karn, **Z. Sztupinszki**, B. Weltz, V. MullerL. Pusztai. (2015) Dynamic classification using case-specific training cohorts outperforms static gene expression signatures in breast cancer. Int J Cancer, 136: 2091-8. IF=5,53

6.2. Publications not related to the thesis

Beres, N.J., Z. Kiss, **Z. Sztupinszki**, G. Lendvai, A. Arato, E. Sziksz, A. Vannay, A.J. Szabo, K.E. Muller, A. Cseh, K. BorosG. Veres. (2016) Altered mucosal expression of microRNAs in pediatric patients with inflammatory bowel disease. Dig Liver Dis.   IF=2,719

Lee, S., K. Vargova, I. Hizoh, Z. Horvath, P. Gulacsi-Bardos, **Z. Sztupinszki**, A. Apro, A. Kovacs, I. Preda, E. Toth-ZsambokiR.G. Kiss. (2014) High on clopidogrel treatment platelet reactivity is frequent in acute and rare in

elective stenting and can be functionally overcome by switch of therapy. Thromb Res, 133: 257-64.   IF=2,32

Szasz, A.M., B. Acs, E. Agoston, **Z. Sztupinszki**, A.M. Tokes, L. Szittya, B. Szekely, M. Szendroi, Q. Li, L. Harsanyi, J. Timar, Z. Szallasi, C. Swanton, B. GyorffyJ. Kulka. (2013) [Simplified, low-cost gene expression profiling for the prediction of outcome in breast cancer based on routine histologic specimens]. Orv Hetil, 154: 627-32. IF= 0,291

Szasz, A.M., Q. Li, A.C. Eklund, **Z. Sztupinszki**, A. Rowan, A.M. Tokes, B. Szekely, A. Kiss, M. Szendroi, B. Gyorffy, Z. Szallasi, C. SwantonJ. Kulka. (2013) The CIN4 chromosomal instability qPCR classifier defines tumor aneuploidy and stratifies outcome in grade 2 breast cancer. PLoS One, 8: e56707. IF=3,057

Penzvalto, Z., B. Tegze, A.M. Szasz, **Z. Sztupinszki**, I. Liko, A. Szendroi, R. SchaferB. Gyorffy. (2013) Identifying resistance mechanisms against five tyrosine kinase inhibitors targeting the ERBB/RAS pathway in 45 cancer cell lines. PLoS One, 8: e59503. IF=3,057

Mihaly, Z., **Z. Sztupinszki**, P. SurowiakB. Gyorffy. (2012) A comprehensive overview of targeted therapy in metastatic renal cell carcinoma. Curr Cancer Drug Targets, 12: 857-72. IF=3,707