Predictive transcriptomic analysis identified specific deregulated pathways before and after renal transplantation

Maglietta R[1], Zaza G[2], Grandaliano G[2], Stifanelli P[1], Schena FP[2], Ancona N[1]

Motivation

Delayed graft function (DGF) represents the most common complication occuring in the immediate post-transplantation period. It is most commonly defined as the need for dialysis within the first week after transplantation. The molecular mechanisms underlying this event are still poorly defined. Thus, we are currently missing specific markers to predict this event. In this contest, genome-wide expression analysis may represent a valuable approach in the search for disease-specific molecular markers. However, the standard associative statistic method currently used to perform data analysis, with a single gene approach, is generally unable to analyze the multi-genetic influence on a specific phenotype. Therefore, to identify specific biological fingerprints associated to renal transplantation and able to identify DGF patients versus those with early graft function (EGF), a new analytical method, that evaluates microarray data at the level of functional gene sets or pathways has been developed.

Results

For our study, we analyzed the transcriptomic profile (expression level of approximatly 20,000 gene probe sets by HG U133A, Affymetrix platform) of pheripheral blood mononuclear cells isolated from 14 chronic kidney disease patients before (Time 0) and 24 hours after renal transplantation (Time 24). For the statistical analysis, we estimated the generalization error evaluated through the Leave-K-Out Cross Validation procedure of Regularized Least Square (RLS) classifiers and the statistical significance of the prediction error by the p-value computed using a nonparametric permutation test. In this analysis, in which we used 19 samples to train the RLS classifier, we obtained a prediction error rate of 7.2% with a p-value=0.0008 demostrating a clear difference between the transcriptome before and after transplantation. Additionally, using the t-test, we found that 2720 gene probe sets clearly descriminated the two conditions (p-value<0.05, fdr<37%). Subsequently, at both Time 0 and 24, we divided the patients in two groups: DGF group (n=7) and EGF group (n=7). Analyzing the difference beween the whole genomic profile of DGF and EGF at time 0, we found an error rate of 32% with a p-value=0.31 and the t-test statistics identified 1666

[1] Istituto di Studi sui Sistemi Intelligenti per l'Automazione, CNR, Via Amendola 122/D-I, 70126, Bari. [2] Unità di Nefrologia, Dialisi e Trapianto-Dipartimento dell'Emergenza e Trapianto, Università di Bari, Piazza Giulio Cesare 11, 70124, Bari

differentially expressed probes (p-value<0.05, fdr<52%). Analogue results were found in the analysis performed at Time 24. In fact, the error rate was 38% with p-value=0.30 and the t-test statistics showed 1036 differentially expressed gene probe sets (p-value<0.05, fdr<81%). The estimated error rate although not statistically significant shows weak differences between the phenotypes analyzed which the classical single gene analysis is unable to highlight. These results showed the inability of the single gene method to discriminate at a genomic level the two study groups and underlined the necessity to utilize different statistical approachs to address this point. Therefore, we decided to follow a new strategy in which sets of genes coding particular pathways or cellular functions are jointly analyzed. In particular, we adopted a Random Set method to analyze the difference between the global levels of expression of 825 biological processes (listed in the c5 collection of Molecular Signature Database, http://www.broadinstitute.org/gsea/msigdb/index.jsp). Using this method, we found 83 statistically significant biological processes discriminating all the 14 patients at Time 0 versus Time 24, 92 discriminating DGF versus EGF group at Time 0 and 45 at Time 24, respectively. Interestingly, at Time 0, NLS_BEARING_SUBSTRATE_IMPORT_INTO _NUCLEUS (including 37 gene probe sets) was the most statistically significant biological process discriminating DGF versus EGF (p-value=0.003 and fdr=33%). Moreover, the same pathway was found correlated to the phenotype in the analysis of the dataset at Time 24 (p-value= 0.003, fdr=56%). In conclusion, this new analytic approach, avoiding the biases of the traditional associative statistic metodology, was able to identify important biological mechanisms potentially involved in the pathogenesis of DGF and may represent a valuable tool for biological and medical transcriptomic research strategies.

Contact e-mail
ancona@ba.issia.cnr.it

Supplementary information

Funding
This work was supported by grants from Regione Puglia, Progetto Strategico PS_012 and Progetto Reti di Laboratori Pubblici di Ricerca BISIMANE.