

Novel Global Network Scores to Analyze Kinase Inhibitor Profiles

Jacques Colinge Uwe Rix Giulio Superti-Furga

CeMM – Center for Molecular Medicine of the Austrian Academy of Sciences,
1090 Vienna, Austria, {jcolinge,urix,gsuperti-furga}@cemm.oeaw.ac.at

Abstract We propose new developments of global network scoring methods previously introduced to estimate the importance of genes in diseases. We apply these methods to drug proteomics profiles, which consist of drug targets, to determine the part of the human interactome perturbed by a drug. Drug and disease network scores can be combined to obtain novel side-effect and pathway association prediction strategies. We illustrate our methods comparing four kinase inhibitor profiles (dasatinib, bosutinib, imatinib, bafetinib) ranging from very promiscuous to highly specific. We predict and identify the cause of plausible side-effects of bosutinib.

Keywords proteomics, drugs, systems biology, CML, kinase inhibitors

1 Introduction

To understand the mechanisms of action of drugs and to be able to predict *in silico* their potential side-effects is of prime importance in drug discovery and molecular medicine. In addition to the obvious benefit of developing better and safer drugs, emergence of resistance to treatment, e.g. in cancer, further calls for detailed modeling of drug actions to adapt clinical practice to patient specificities.

Proteomics approaches combining drug affinity purification techniques (AP) and tandem mass spectrometry (MS/MS) make it possible to measure drug targets in cell lines or patient material in an unbiased manner [1-3] as opposed to classical binding assays. Such analyses usually reveal larger than expected target spectra, thus making obsolete the naïve view of a single main target. Such complex target spectra can be used to understand the interaction between drugs and patients at a molecular level. In order to try identifying mechanisms of action and side-effects, we adopt a systems biology view, hence considering drug target spectra in the context of the whole human interactome or pathways [4].

The analysis of genetic data, e.g. WGA screens, often requires the identification of relevant gene(s) from a rather large portion of the genome, i.e. from a long list of non relevant genes. To link the candidate genes with previously known disease-associated genes through protein-protein interaction (PPI) data has been shown to be promising [5]. Nonetheless, in such networks, certain proteins are highly connected (hubs) and it is unclear whether their many interacting partners retain any specificity for the disease, see Figure 1. Therefore, global scores, taking

the whole interactome topology into account and limiting the importance of hubs have been proposed [6].

Recently, Berger et al. [7] used a related method to associate drugs with a heart pathology (EQTS) that is often observed as side-effect. They postulated that if a drug targets disease causing genes, then side-effects similar to the disease phenotype can be expected. They first assigned relevance (global) scores to genes from known genes causing EQTS. Then, only considering drug target genes, they took the maximum disease relevance score as a measure of side-effect risk.

Although Berger et al. obtained satisfying results testing many drugs with target spectra taken from public databases, certain kinase inhibitors can have a large number of targets and to only consider a single one is risky, especially if the targets synergize with each other. It is also important to mention that target spectra from public sources are often quite incomplete and thus small. Moreover, we want to exploit information regarding drug-target affinities, which can be estimated from the proteomics data or available from other sources.

Therefore, we introduce a novel disease/drug association score combining the influence of a drug and a disease over the human interactome. Our score statistical significance is explicitly determined. We show the improvement brought by the new scores and use them to predict dasatinib, bosutinib, imatinib, and bafetinib relative efficacy in chronic myeloid leukemia (CML) treatment and new areas of application. We finally adapt the our score to search KEGG [8] and identify likely side-effects of Bosutinib.

2 Materials and Methods

2.1 Generation of Experimental Data

Drug profiles were obtained from K562 cells. Compounds have been modified to be coupled to NHS-activate Sepharose beads and drug affinity purification has been performed following a protocol described in previous publications [2, 9, 10]. Pulldown samples were submitted to mass spectrometry (MS) analysis and bioinformatics protein identification. Elimination of non-specific binders has been achieved by subtracting pulldown data of unrelated compounds (kanamycin, daunorubicin, ciprofloxacin, amphotericin B, paroxetine) and considering kinases only (except NQO2 for imatinib [1]). There were no direct measurement of drug affinities but we have shown that proteomics data, e.g. protein sequence coverage, can provide a good estimate of IC_{50} values [10], which we use then as a proxy for binding strengths.

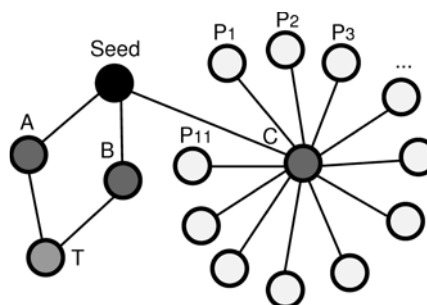


Figure 1. Seed node influencing its neighbors. Second interactor T is still specific, though P_1, \dots, P_{11} have lost specificity because of C high degree of connectivity.

2.2 Network Scores

Scores associating a handful of seed nodes with the rest of the network through random walks have been described by other authors already [6, 7, 11]. We use the method of Köhler, et al.: Given a human interactome represented as a graph $G = (E, V)$ and a set S of k seed (disease causing) genes $S = \{v_{s_1}, \dots, v_{s_k}\} \subset V$, we define initial node probabilities x^0 , with $x_i^0 = 1/k$ if $i \in \{s_1, \dots, s_k\}$, $x_i^0 = 0$ otherwise. We then define a random walk (Markov chain) with restart $x^{n+1} = (1-r)Px^n + rx^0$, with P the row-normalized adjacency matrix of G and $r \in [0,1]$. The relevance of genes in V , with respect to S , is defined by the steady-state probability vector x^∞ . This algorithm is similar to ‘‘PageRank with Priors’’ [11]. Köhler et al. have shown good performance in finding disease associated genes (AUROC > 90%, one disease gene removed from S). We do not repeat these validations here. Chen et al. have shown that $r=0.3$ is a good (robust) choice.

Given a drug target profile S' , we can apply a similar iteration assigning initial probabilities x'^0 proportional to protein sequence coverage. The latter is an approximation of the protein abundance [12] in the pulldown sample and it is correlated to the real affinity with the compound [10].

Probabilities in x^∞ and x'^∞ are defined over the whole set V and we can naturally multiply these vectors to measure how a drug impacts the human interactome under the influence of a disease. A natural global score for disease/drug interaction would thus be the inner product $I = \langle x^\infty, x'^\infty \rangle$, which sums all the node scores. To avoid introducing excessive noise in the global score, we first determine significant sub-graphs spanned by x^∞ and x'^∞ : by means of bootstrap simulations we estimate critical node probabilities at the α level of significance ($\alpha=5\%$ typically). We thus obtain sub-graphs $G_S = (E_S, V_S)$ and $G_{S'} = (E_{S'}, V_{S'})$ and we set to 0 the probabilities of nodes not in the significant sub-graphs:

$$\text{Disease}_{\text{local},i} = \begin{cases} x_i^\infty, & i \in V_S \\ 0, & i \in V - V_S \end{cases} \quad \text{and} \quad \text{Drug}_{\text{local},i} = \begin{cases} x'_i{}^\infty, & i \in V_{S'} \\ 0, & i \in V - V_{S'} \end{cases}. \quad (1)$$

The score is finally given by the inner product $I_{\text{local}} = \langle \text{Disease}_{\text{local}}, \text{Drug}_{\text{local}} \rangle$. We also consider an extremely localized score I_{max} defined as the maximum of component wise product of $\text{Disease}_{\text{local}}$ and $\text{Drug}_{\text{local}}$. It informs on the strongest protein/drug pair. We name the score of Berger et al. I_{single} .

P-values for global scores are obtained by a bootstrap on the drug network. Random scores fit a Gamma distribution accurately, which is hence used to compute P-values.

The human interactome G was built integrating the IntAct, HPRD, MINT, and BioGRID databases [13-16], complemented with Bcr-Abl interactions measured by our laboratory [17]. All the protein accession codes were mapped to UniProtKB/SwissProt and we obtained 11,406 nodes and 71,387 edges. We extracted the maximum connex component to define G (11,303 nodes, 71,328 edges).

All the computations were implemented in R with some data preparation in Perl.

3 Results

3.1 Score Performance

Köhler et al. introduced a list of 110 manually curated disease-gene associations [6]. We used this list to generate disease networks. As CML is not present in this list, we complemented it with a CML network generated by BCR-ABL, STAT5, and Gab2 [18-20]. We also generated drug networks for the four kinase inhibitors on the basis of our proteomics data and scored disease associations. Table 1 reports 5% significant diseases by giving their rank as obtained with I_{local} . The first 4 non cancer diseases are together with most cancers in a strong cluster if one computes a disease distance similarity matrix with the same global scores. It is thus not a surprise they are not well separated from cancers. The next 3 listed are their immediate neighbors in the next cluster. Only the last 3 are really non related.

Table 1. Ranks (Ilocal) of the diseases selected at a 5% false positive rate. For CML, we indicate in brackets the rank obtained adding Lyn to the disease causing genes. Grayed cells indicate current clinical trials (ClinicalTrials.org). (*) Dasatinib CML rank with LYN is close to one as the scores of first and second positions are almost identical.

	Imatinib	Bafetinib	Dasatinib	Bosutinib
Cancers				
Chronic myeloid leukemia (Lyn)	1 (3)	1 (1)	2 (2*)	5 (1)
Lung cancer	3	3	3	2
Hepatoacelluar carcinoma	2	4	4	3
Breast cancer familial	5	6	6	4
Juvenile myelomonocytic leukemia	12	5	7	7
Pancreatic carcinoma	6	11		6
Glioma of brain, familial	8			10
Bladder Cancer	15	8	11	15
Pheochromocytoma		14	12	9
Esophageal carcinoma	9			
Medulloblastoma		13		13
Others				
Noonan Syndrome	7	2	5	1
Alzheimer Disease	14	9		
Elliptocytosis		7		14
Spinocerebellar Ataxia			10	11
Cutis laxa	11	10	9	
Stickler syndrome	4			
Ehlers Danlos syndrome	10			
Charcot Marie Tooth Disease	13	12	8	6
Night-blindness, congenital stationary			1	
Atypical mycobacteriosis, familial				12

We found that the local score I_{local} gave better results than I . We have also observed the score I_{single} of Berger et al. did perform worse. It is worth noting they used mean first passage times as relevance score and we used steady-state probabilities.

I_{single} score performed better always giving CML with rank 1 but it was more distracted by the closest 5 non cancer diseases and reported much weaker cancer associations. Its good performance on CML is due to the fact that one of the causing

genes (BCR-ABL) is a strong target of all 4 compounds. If strong targets are not used as seeds, such as in the discovery of side-effects, it is likely not to happen and we conclude that I_{single} is a worse choice. Scores I and I_{max} were worse in ranking CML first and other cancers at the top.

3.2 Comparing Four Kinase Inhibitors

The four compounds we consider in this study are very different. Imatinib was designed to be extremely specific and, accordingly, it only targets a few proteins. Dasatinib and bosutinib have been designed as dual SRC/BCR-ABL inhibitors. However, subsequent analysis, by us and others, has shown that they target a wide range of tyrosine, receptor tyrosine, and serine/threonine kinases in CML cells at relevant drug concentrations. Bafetinib, on the other hand, is based on the imatinib structural scaffold and achieves higher target selectivity than Dasatinib and Bosutinib, while retaining the ability to potently inhibit BCR-ABL and certain SRC family kinases, such as LYN that participates in maintaining the disease in certain blast phase imatinib resistant patients [21].

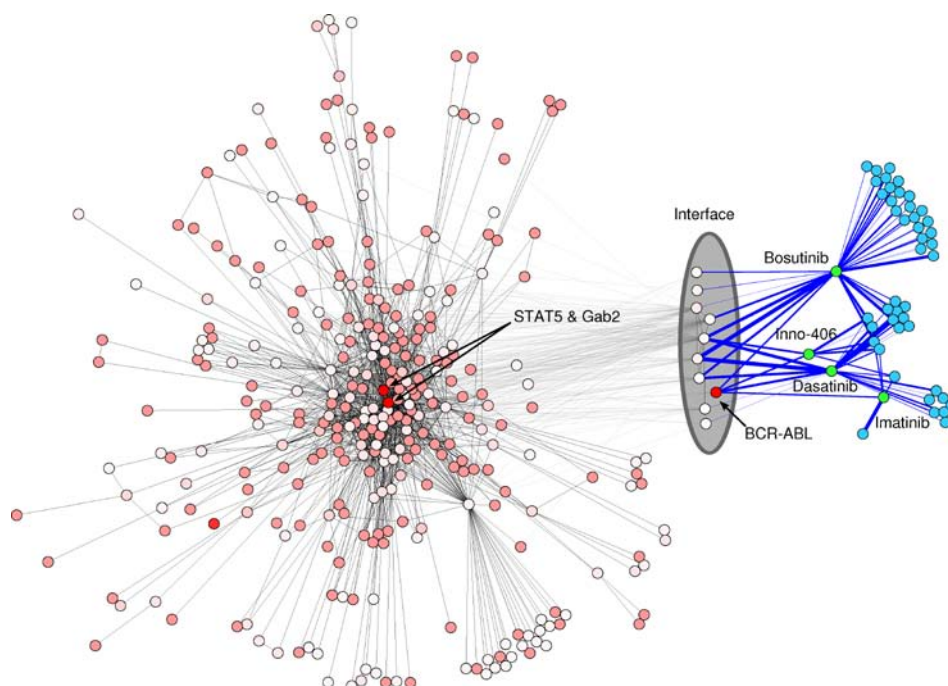


Figure 2. CML network with the kinase inhibitors targets. Several targets are part of CML network but only BCR-ABL is strongly relevant (color reflects relevance, red highly relevant, white at the lowest limit of significant relevance). Non CML relevant targets are in blue. To include Lyn as one disease causing gene turns it into a strongly relevant node in the “interface” grey ellipse and remodel the CML network (1/3 larger and more connected).

To be able to generate a global approximation of the disease influence over the whole human interactome, gives us the opportunity to compare these 4 compounds in the context of the disease. We have generated comparative target profiles of all 4 compounds with estimated affinity strengths and color-coded disease relevance (Figure 2). We see that several targets are entry points into the CML network, but of these only BCR-ABL is sufficiently relevant for CML.

We can apply our score I_{local} to compare treatment efficacies. To normalize I_{local} for the variable number of targets we assume that compounds are available a sufficient concentrations in patient cells to bind with their interactors only based on their affinity (very likely to be true given measured maximum blood concentrations). Therefore, we multiply I_{local} of each drug by the sum of the target sequence coverages and we obtain that drugs are ranked as

$$\text{dasatinib } (I_{local}=0.11) > \text{bafetinib } (0.09) \approx \text{bosutinib } (0.07) > \text{imatinib } (0.05)$$

which matches clinical observations and further indicate that our score captures a reasonable part of the disease-drug association.

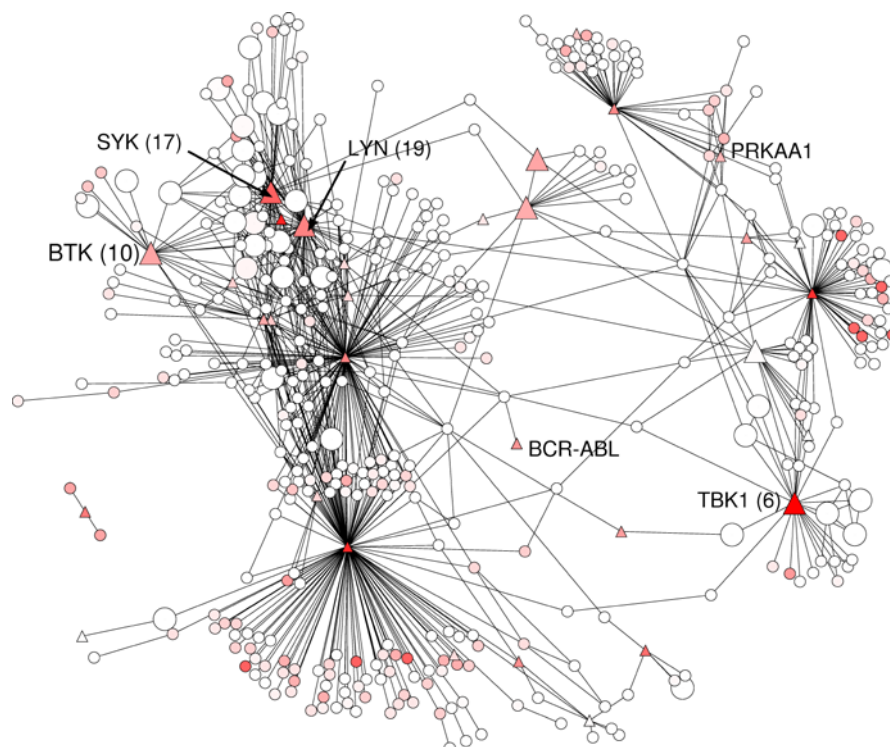


Figure 3. The Bosutinib network with targets depicted as triangles and drug strength of influence color coded (red=strong). Proteins located in the selected immune system pathways are represented with larger nodes. Main immune system related targets are named with – in brackets – the number of interactions with other immune system proteins found in the bosutinib network. BCR-ABL is isolated in this context. PRKAA1 is involved in the endocrine system interaction.

3.3 Bosutinib and Focused KEGG Searches

Dasatinib has been shown to impact the immune system [22, 23]. Bosutinib that is even more promiscuous and used at higher maximum blood concentrations (500nM vs 100nM) is likely to have side-effects as well, in particular on the immune system. We also indicated in a previous work that it might interact with the endocrine system additionally [10].

Köhler et al. list was not appropriate to identify such side-effects. We rather use KEGG as it contains many signaling and metabolic pathways. To directly use the target list for classical pathway enrichment analysis in KEGG is too local and neglects extended influence of the drug. To use the human interactome to collect all bosutinib targets direct interactors yields an enormous list of 1306 proteins, whose relations with the drug are unweighted, and 69 pathways at the 5% significance level. This is too unspecific.

We adapted I_{local} score by summing the drug network steady-state probabilities over the proteins included in a given pathway. We naturally find cancer pathways (CML, ErbB signaling, Pathways in cancer) and several cell growth and proliferation pathways. We selected all the immune system pathways included in a 5% significant list of hits and (B cell receptor signaling, Fc epsilon RI signaling, Fc gamma R-mediated phagocytosis, Chemokine signaling, Toll-like receptor signaling, Natural killer cell mediated cytotoxicity, and RIG-I-like receptor signaling) and we annotated the bosutinib network to localize their interaction. See Figure 3. We did the same with the two endocrine system pathways identified: GnRH and Insulin signaling.

We note the strong interaction of bosutinib with immune system actors and notably LYN and BTK that play an important role in dasatinib immune system side-effects [22, 23]. We also note important roles of TBK1 [24] and SYK [25, 26] in immunodeficiency. Potential interaction with the endocrine system is more localized and mediated by PRKAA1, which regulates fatty acids and cholesterol synthesis and is considered as a stress-sensing metabolic regulator.

4 Discussion

We have introduced novel drug/disease association methods by extending the use of global network scores to drug profiles integrating information about protein drug affinities. Such affinities were obtained from semi-quantitative proteomics data but the method presented could also take more precise measurements, e.g. affinity constants or IC_{50} 's. Since we define drug and disease global score as probabilities, the two can be naturally combined to measure coincidence of disease and drug influence over an interactome. Disease and drug network sizes are defined by means of statistical significance. We have shown that the new disease-drug score performs competitively.

Comparison of 4 kinase inhibitors used in CML treatment in the context of the CML network show that broad spectrum compounds are not likely to be advantageous as BCR-ABL is their only highly relevant target, which is confirmed by existing unpublished data of our institute. Additional targets such as LYN can be useful in certain resistant cases but, in general, they are more likely to cause side-effects or open opportunities for additional indications. In Table 1, we note that

several clinical trials are running for imatinib, dasatinib, and bosutinib in cancer types highly ranked by our score. In particular, Table 1 suggests testing the 4 compounds against hepatocellular carcinoma and bafetinib against lung and breast cancers.

Finally, adapted drug global scores allowed us to search for side-effects against KEGG pathways and we could find strong indication that bosutinib has an effect on the immune system as dasatinib does.

To conclude, we believe that the global scores pioneered by Köhler et al. and Berger et al., are useful to understand interactions with diseases and potential side-effect. Our work extended their ideas and proposed new tools for understanding drug mechanisms of action at a system level.

Acknowledgements

We thank Oliver Hantschel for helping us defining a minimal list of CML causing genes. This work was partially supported by GEN-AU grants (APP and BIN) from the Austrian Ministry of Sciences.

References

- [1] Rix, U., et al. Chemical proteomic profiles of the BCR-ABL inhibitors imatinib, nilotinib, and dasatinib reveal novel kinase and nonkinase targets. *Blood*, 2007, 110(12): 4055-4063.
- [2] Fernbach, N.V., et al. Acid elution and one-dimensional shotgun analysis on an Orbitrap mass spectrometer: an application to drug affinity chromatography. *J Proteome Res*, 2009, 8(10): 4753-4765.
- [3] Huang, S.M., et al. Tankyrase inhibition stabilizes axin and antagonizes Wnt signalling. *Nature*, 2009, 461(7264): 614-620.
- [4] Berger, S.I., Iyengar, R. Network analyses in systems pharmacology. *Bioinformatics*, 2009, 25(19): 2466-2472.
- [5] Lage, K., et al. A human phenome-interactome network of protein complexes implicated in genetic disorders. *Nat Biotechnol*, 2007, 25(3): 309-316.
- [6] Kohler, S., et al. Walking the interactome for prioritization of candidate disease genes. *Am J Hum Genet*, 2008, 82(4): 949-958.
- [7] Berger, S.I., et al. Systems pharmacology of arrhythmias. *Sci Signal*, 3(118): ra30.
- [8] Kanehisa, M., et al. KEGG for linking genomes to life and the environment. *Nucleic Acids Res*, 2008, 36(Database issue): D480-484.
- [9] Rix, U., et al. A comprehensive target selectivity survey of the BCR-ABL kinase inhibitor INNO-406 by kinase profiling and chemical proteomics in chronic myeloid leukemia cells. *Leukemia*, 2009.
- [10] Rix, U., et al. Global target profile of the kinase inhibitor bosutinib in primary chronic myeloid leukemia cells. *Leukemia*, 2009, 23(3): 477-485.
- [11] Chen, J., et al. Improved human disease candidate gene prioritization using mouse phenotype. *BMC Bioinformatics*, 2007, 8: 392.
- [12] Ishihama, Y., et al. Exponentially modified protein abundance index (emPAI) for estimation of absolute protein amount in proteomics by the number of sequenced peptides per protein. *Mol Cell Proteomics*, 2005, 4(9): 1265-1272.
- [13] Kerrien, S., et al. IntAct--open source resource for molecular interaction data. *Nucleic Acids Res*, 2007, 35(Database issue): D561-565.
- [14] Breitkreutz, B.J., et al. The BioGRID Interaction Database: 2008 update. *Nucleic Acids Res*, 2008, 36(Database issue): D637-640.

- [15] Mishra, G.R., et al. Human protein reference database--2006 update. *Nucleic Acids Res*, 2006, 34(Database issue): D411-414.
- [16] Cesareni, G., et al. Searching the MINT database for protein interaction information. *Curr Protoc Bioinformatics*, 2008, Chapter 8: Unit 8 5.
- [17] Brehme, M., et al. Charting the molecular network of the drug target Bcr-Abl. *Proc Natl Acad Sci U S A*, 2009, 106(18): 7414-7419.
- [18] Sawyers, C.L. Chronic myeloid leukemia. *N Engl J Med*, 1999, 340(17): 1330-1340.
- [19] Shuai, K., et al. Constitutive activation of STAT5 by the BCR-ABL oncogene in chronic myelogenous leukemia. *Oncogene*, 1996, 13(2): 247-254.
- [20] Sattler, M., et al. Critical role for Gab2 in transformation by BCR/ABL. *Cancer Cell*, 2002, 1(5): 479-492.
- [21] Donato, N.J., et al. BCR-ABL independence and LYN kinase overexpression in chronic myelogenous leukemia cells selected for resistance to STI571. *Blood*, 2003, 101(2): 690-698.
- [22] Sillaber, C., et al. Immunosuppression and atypical infections in CML patients treated with dasatinib at 140 mg daily. *Eur J Clin Invest*, 2009, 39(12): 1098-1109.
- [23] Hantschel, O., et al. The Btk tyrosine kinase is a major target of the Bcr-Abl inhibitor dasatinib. *Proc Natl Acad Sci U S A*, 2007, 104(33): 13283-13288.
- [24] Ishii, K.J., et al. TANK-binding kinase-1 delineates innate and adaptive immune responses to DNA vaccines. *Nature*, 2008, 451(7179): 725-729.
- [25] Hussain, S.F., et al. A novel small molecule inhibitor of signal transducers and activators of transcription 3 reverses immune tolerance in malignant glioma patients. *Cancer Res*, 2007, 67(20): 9630-9636.
- [26] Deuse, T., et al. Novel immunosuppression: R348, a JAK3- and Syk-inhibitor attenuates acute cardiac allograft rejection. *Transplantation*, 2008, 85(6): 885-892.