Using probabilistic models of signaling pathways to predict *in vivo* drug activity.

Patricia Sebastián-León[1] and Joaquín Dopazo[1,2,3,*]

1 Department of Computational Genomics, Centro de Investigación Príncipe Felipe (CIPF), Valencia, 46012, Spain.
2 CIBER de Enfermedades Raras (CIBERER), Valencia, 46012, Spain,
3 Functional Genomics Node (INB) at CIPF, Valencia, 46012, Spain;
* To whom correspondence should be addressed. Tel: +34 963289680; Email: jdopazo@cipf.es

Abstract

Signaling pathways constitute a valuable source of information that allows interpreting the way in which the cell respond to external stimulus and the aspects of the cell functionality affected by these. Here we explore the effect of drugs in cell signaling and the feasibility of using signaling to predict drug effect. A simple probabilistic model of 23 rat KEGG signaling pathways is used to compare the impact of drugs *in vitro* and *in vivo*. Our results document that almost all the pathways (20 out of the 23 pathways modeled) were affected in one or more stimulus-response circuits in the same way both in cell lines and in the *in vivo* experiment. This effect was observed for half of the drugs tried. Therefore models of cell signaling can be used as predictor of *in vivo* activity from *in vitro* activity in a reasonable number of cases. The advantage of using such models is that they permit an unprecedented insight into the mechanisms of drug effect and also understanding the differences between the *in vitro* and the *in vivo* systems.

Introduction

Signaling pathways represent the way in which the combined effect of gene activity elicits cell-level responses by activating/deactivating specific functionalities in response to particular stimulus through a chain of intermediate molecules. Drugs can either act as external stimulus or directly interfere with the genes of the pathway, causing changes in the "normal" responses. Such changes can be used to understand the biological consequences of the effect of drug in the cell, as well as to give clues on the drug mechanism of action. Despite a different behavior of signaling pathways is expected when cell lines are compared to organs or tissues, some affected signaling mechanisms could be common to certain drugs and could be used to predict *in vivo* activity. We have used the KEGG (Kanehisa, et al., 2012) repository, which contain detailed information about pathways, to obtain the templates for the derivation of the probabilistic models.

Methods

If the individual probabilities of protein presence/absence of all the proteins in the pathway are known, a simple probabilistic model of the pathway can be used to calculate the probabilities for signal transmission from any receptor protein to any final effector protein (taking into account all the intermediate activator and/or repressor proteins in between). Here, we take gene expression values as proxies of gene activity

and, consequently, presence/absence of the corresponding protein (Efroni, et al., 2007). We have used more than 10,000 Affymetrix microarrays downloaded from the GEO database (Barrett, et al., 2013) to derive the empirical distributions of presence/absence for each probe, that are further used to calculate the probability of presence/absence for the to the genes involved in the studied pathways (Efroni, et al., 2007; Sebastian-Leon, et al., 2013). Nodes have been treated in different ways depending on whether they were composed by alternative proteins (redundancy: only one of them keeps the node working) or complexes (all proteins are indispensable to keep the node working). This simplification has proven to be useful in practical terms (Sales, et al., 2012). Therefore, given the measurements of gene expressions in a particular experiment, the reference distributions can be sued to estimate the probabilities of presence/absence of each protein (and each node) of the pathway.

Once such probabilities have been estimated, the probability of signal transmission along a stimulus-response circuit can easily be inferred from the probabilities of activation of all the connecting nodes that constitute the circuit (providing that inhibitor nodes allow signal transmission when they are deactivated). The circuits are defined by the 23 KEGG pathways of rat used here (see Table 1). Therefore, the stimulus-response circuits of any of the pathways can easily be modeled by means of a simple product of probabilities (using the principle of inclusion/exclusion when bi- or multi-furcating stretches are present) (Sebastian-Leon, et al., 2013). This provides a straightforward approach to estimate the probability of signal transmission from gene expression values. However, such probabilities of signal transmission when out of context are not informative. What is interesting is the comparison of such probabilities in two different conditions (typically cases versus controls). We apply a Wilcoxon test (Wilcoxon, 1945) that allows detecting which stimulus-response circuits significantly change their probabilities of signal transmission between the compared conditions.

Here, we compare the changes induced by a collection of 132 drugs from the TGP dataset from the Japanese Toxicogenomics Project (Uehara, et al., 2010) in the different circuits of different pathways both, *in vitro* and *in vivo*.

The models of the pathways have recently been published (Sebastian-Leon, et al., 2013) and are available at: http://pathiways.babelomics.org/

Results

For each drug, we carried out all the comparisons between the doses tried *in vitro* and *in vivo*, independently. For any of these comparisons, we studied which circuits in which pathways displayed a significant change in the activity induced by the drug, as well as the type of change experimented (activation or inhibition). Table 1 shows the pathways in which the drugs caused the same type of alterations in one or several stimulus-response circuits. A total of 931 different circuits from all the pathways were affected by one or more drugs. Cell lines are more affected by drugs than the corresponding *in vivo* counterparts (by more than a 25% in average). However, only 207 stimulus-response circuits, corresponding to almost all the pathways (20 out of a total of 23

modeled) represented in Table 1 display coincident patterns of activation in response to several of the drugs tried. Almost half of the drugs tried (58 out of 132) caused an identical effect both *in vitro* and *in vivo* in at least one circuit of at least one pathway.

KEGG ID	Name	Drugs
rno03320	PPAR SIGNALING PATHWAY	bendazac, benzbromarone, benziodarone, clofibrate, fenofibrate, furosemide, gemfibrozil, simvastatin, sulfasalazine, WY-14643
rno04115	p53 SIGNALING PATHWAY	colchicine, disopyramide, ethionine, moxisylyte, nitrosodiethylamine, propylthiouracil, puromycin_aminonucleoside, quinidine
rno04060	CYTOKINE-CYTOKINE RECEPTOR INTERACTION	diazepam
rno04210	APOPTOSIS	hydroxyzine, nitrofurantoin
rno04340	HEDGEHOG SIGNALING PATHWAY	
rno04514	CELL ADHESION MOLECULES	caffeine, aproxen, nitrofurazone, tacrine, colchicine, gentamicin
rno04612	ANTIGEN PROCESING AND PRESENTATION	flutamide, puromycin_aminonucleoside
rno04662	B CELL RECEPTOR SIGNALING PATHWAY	nimesulide, nitrofurazone, chloramphenicol, colchicine, mexiletine, gentamicin, hydroxyzine, sulpiride
rno04916	MELANOGENESIS	Doxorubicin, isoniazid
rno04012	ERBB SIGNALING PATHWAY	hydroxyzine, nitrofurantoin, colchicine, ethionine, colchicine, caffeine
rno04310	WNT SIGNALING PATHWAY	Caffeine, ibuprofen
rno04370	VEGF SIGNALING PATHWAY	Acetamidofluorene, cyclophosphamide, danazol, diazepam, ethambutol, ethinylestradiol, ibuprofen, cyclosporine_A, diazepam, ajmaline, ethinylestradiol, ethambutol, nitrofurantoin, nitrofurantoin
rno04530	TIGHT JUNCTION	caffeine, cisplatin, naproxen, sulindac, ethionine, gentamicin, monocrotaline, puromycin_aminonucleoside
rno04630	JAK-STAT SIGNALING PATHWAY	diclofenac, disopyramide, furosemide, ibuprofen, sulindac
rno04664	Fc EPSILON RI SIGNALING PATHWAY	colchicine, ethionine, gentamicin, penicillamine, valproic_acid
rno04920	ADIPOCYTOKINE SIGNALING PATHWAY	diclofenac, naphthyl_isothiocyanate, naproxen, colchicine,
rno04020	CALCIUM SIGNALING PATHWAY	ethionine, hydroxyzine, caffeine
rno04330	NOTCH SIGNALING PATHWAY	Methimazole, naproxen
rno04512	ECM-RECEPTOR INTERACTION	nifedipine
rno04540	GAP JUNCTION	carbon_tetrachloride
rno04660	T CELL RECEPTOR SIGNALING PATHWAY	colchicine, ethionine, gentamicin, penicillamine, valproic_acid, caffeine, disopyramide, naproxen, sulindac, naphthyl_isothiocyanate, hydroxyzine, coumarin
rno04912	GnRH SIGNALING PATHWAY	disopyramide, naproxen, iproniazid

Figure 1 shows in detail the activity of several drugs in the PPAR signaling pathway. A total of twelve drugs significantly trigger the activation of the lipid metabolism and the adipocyte differentiation both *in vitro* and *in vivo*. This functional activation is attained through the activation three main stimulus-response circuits (in red in the figure). The detail provided by the model allows understanding the ways through the drugs are acting in the cell, as well as detecting other valuable collateral drug effects, as side effects, drug resistances, etc., providing these have a significant impact in any of the modeled pathways.

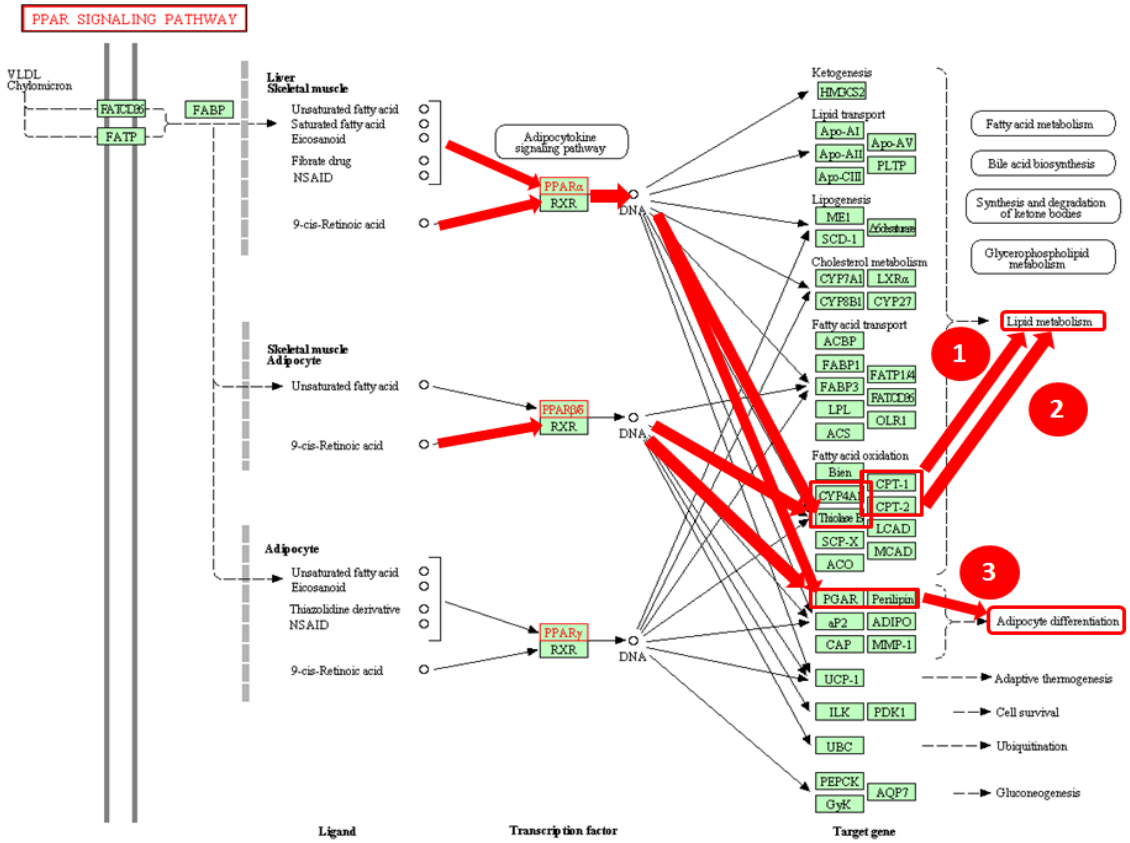

Figure 1. Rat pathway PPAR (rno03320) with red arrows indicating activation of signaling circuits by different drugs. 1) Circuit activated by: benzbromarone, clofibrate, fenofibrate, naproxen, WY-14643, omeprazole; 2) circuit activated by: benziodarone, sulfasalazine; 3) circuit activated by: bendazac, benzbromarone, fenofibrate, gemfibrozil, simvastatin, WY-14643. The effect of the drugs is an activation of the lipid metabolism and the adipocyte differentiation.

Discussion

Cell lines have extensively been used for initial *in vitro* testing of drugs. However, its validity as models of *in vivo* systems is questionable. Recent studies demonstrate that the global pattern of gene expression of cell lines is completely different to any other cell type, either healthy or diseased (Lukk, et al., 2010). However, this quantitative observation does not provide any information about the extent at which cell lines still retain similar functionalities of the cell type from which they have been derived from. Here we have used a simple probabilistic model that transforms gene expression levels into probabilities of signal transmission across signaling pathways, from receptor nodes, which receive the stimulus, to the effector nodes that trigger the corresponding response. In this way, gene expression data, of often difficult interpretation, are transformed into meaningful functional information regarding changes in the different pathway responses triggered by particular stimulus.

Our observations document a different behavior of cell lines with respect to their *in vivo* counterparts. However, such differences are not as radical as the behaviors described for the global gene expression (Lukk, et al., 2010) and only affects to about a 25% of the

signaling circuits in the average. This indicates that, despite the disparity in global gene expression, the global behaviors are, probably, not so dissimilar.

Using pathways to assess drug responses have a number of limitations. Firstly, there are drugs (half of the drugs tested here) that will not affect to the set signaling pathways modeled and therefore their effects will remain undetectable. In other cases, extensive responses, mainly observed *in vitro*, mask the induction or repression of common circuits that might be useful to predict drug activity.

Despite the described limitations, our results suggest that the use of models of pathways can offer an interesting alternative to other "black box" methods for drug activity prediction. More detailed modeling of cell activity, including metabolic pathways and other aspects such as regulation, protein interaction, etc., will probably increase the predictive accuracy offering, at the same time, valuable information on the drug action mechanisms.

Funding

This work is supported by grants BIO2011-27069 from the Spanish Ministry of Economy and Competitiveness (MINECO), PROMETEO/2010/001 from the Conselleria de Educacio of the Valencia Community. We also thank the support of the National Institute of Bioinformatics (www.inab.org), the CIBER de Enfermedades Raras (CIBERER), both initiatives of the ISCIII, MINECO and the Bull Chair in Computational Genomics (http://bioinfo.cipf.es/chair_compgenom).

References

Barrett, T., *et al.* (2013) NCBI GEO: archive for functional genomics data sets--update, *Nucleic Acids Res*, **41**, D991-995.
Efroni, S., Schaefer, C.F. and Buetow, K.H. (2007) Identification of key processes underlying cancer phenotypes using biologic pathway analysis, *PLoS ONE*, **2**, e425.
Kanehisa, M., *et al.* (2012) KEGG for integration and interpretation of large-scale molecular data sets, *Nucleic Acids Res*, **40**, D109-114.
Lukk, M., *et al.* (2010) A global map of human gene expression, *Nat Biotechnol*, **28**, 322-324.
Sales, G., *et al.* (2012) graphite - a Bioconductor package to convert pathway topology to gene network, *BMC Bioinformatics*, **13**, 20.
Sebastian-Leon, P., *et al.* (2013) Inferring the functional effect of gene expression changes in signaling pathways, *Nucleic Acids Res*, **In press**.
Uehara, T., *et al.* (2010) The Japanese toxicogenomics project: application of toxicogenomics, *Mol Nutr Food Res*, **54**, 218-227.
Wilcoxon, F. (1945) Individual comparisons by ranking methods, *Biometrics Bulletin*, **1**, 80-83.