

Actionable pathways: interactive discovery of therapeutic targets using signaling pathway models

Francisco Salavert^{1,2}, Marta R. Hidago¹, Alicia Amadoz¹, Cankut Çubuk¹, Ignacio Medina³, Daniel Crespo¹, Jose Carbonell-Caballero¹ and Joaquín Dopazo^{1,2,4,*}

¹Computational Genomics Department, Centro de Investigación Príncipe Felipe (CIPF), Valencia, 46012, Spain, ²Bioinformatics in Rare Diseases (BiER), Centro de Investigación Biomédica en Red de Enfermedades Raras (CIBERER), Valencia, 46012, Spain, ³HPC Service, University Information Services, University of Cambridge, Cambridge, CB3 0RB, UK and ⁴Functional Genomics Node, (INB, PRB2, ISCIII) at CIPF, Valencia 46012, Spain

Received February 8, 2016; Revised April 13, 2016; Accepted April 22, 2016

ABSTRACT

The discovery of actionable targets is crucial for targeted therapies and is also a constituent part of the drug discovery process. The success of an intervention over a target depends critically on its contribution, within the complex network of gene interactions, to the cellular processes responsible for disease progression or therapeutic response. Here we present PathAct, a web server that predicts the effect that interventions over genes (inhibitions or activations that simulate knock-outs, drug treatments or over-expressions) can have over signal transmission within signaling pathways and, ultimately, over the cell functionalities triggered by them. PathAct implements an advanced graphical interface that provides a unique interactive working environment in which the suitability of potentially actionable genes, that could eventually become drug targets for personalized or individualized therapies, can be easily tested. The PathAct tool can be found at: <http://pathact.babelomics.org>.

INTRODUCTION

Diagnostic strategies are rapidly changing in cancer and other diseases because of the availability of increasingly affordable genomic analysis (1). Therapies that specifically target genetic alterations are probing to be safer and more effective than traditional chemotherapies when used in the adequate patient population (2). Actionable targets with therapeutic potential are discovered through empirical associations between the abundance of specific proteins (or transcripts) or the presence of specific mutations and clinical outcomes, and can be assessed by massive sequencing (3). However, single-gene biomarkers have a limited predictive power and frequently only partially account for the

fundamental cellular processes responsible for tumorigenesis or therapeutic response (4). Since signaling pathways play a crucial role in these processes, the analysis of its activity should provide better resolution in the development of biomarkers linked to cellular function. We recently proposed the use of mechanism-based biomarkers derived from models of cell signaling activity that account for disease mechanisms or for drug mechanisms of action (5,6). Such models are based on the analysis of the collective contribution of genes to the final signal transmission across signaling pathways. The individual contributions are deduced from gene expression values (6) but gene mutations can also be easily integrated in the model (7). Actually, different types of mechanism-based biomarkers have recently proven to be superior to conventional biomarkers in predicting complex clinical parameters such as bad prognostic (4) or drug sensitivity (8). Thus, the use of models of signaling networks constitutes a promising strategy for the prediction of disease outcomes or responses to therapeutic interventions.

Here we present PathAct, a web server that assess how interventions over genes (knockouts—KOs, over-expressions or drug treatments) can affect to signaling pathways and, ultimately, to the cell functionalities triggered by them. PathAct implements improved robust models of signaling pathways, taken from KEGG (9) and based on our previous work (5,6), within an advanced graphical interface that provide a unique interactive working environment in which potentially actionable genes, that could eventually become drug targets, can be easily assayed alone or in combinations.

PROGRAM STRUCTURE

PathAct uses the measurements of gene expression in a given condition (diseased tissue biopsy, cell line, etc.) as reference and calculates the signaling activity of all the signaling circuits represented in the pathways. Then, within the interactive workspace, users can easily make interventions on this reference condition. Interventions consist on gene deactivations (simulating KOs) or gene activations (simulating

*To whom correspondence should be addressed. Tel: +34 96 328 96 80; Fax: +34 96 328 97 01; Email: jdopazo@cipf.es

over-expressions). The consequences of these interventions are calculated and the resulting condition is compared to the reference condition, highlighting the differences.

By default, PathAct is used in anonymous mode, which implies that after the session is finished, the results obtained are lost. Alternatively, users can choose sign up for an account. The functionalities are exactly the same but the all the results are stored in the server and available for future sessions.

Data upload

PathAct uses normalized gene expression values that can be uploaded through the *Jobs* panel (Figure 1A). The input file format is a simple text file with two columns: standard gene identifier (Ensemble gene, gene name, Entrez ID and the most common microarray probe identifiers) and normalized expression value. If probe expressions are provided, the corresponding gene expression values are calculated as the mean value of all the probes mapping in each gene. If more columns with additional samples are provided, the user has to choose the sample to analyze within the input file window. The data file can easily be exported from the typical excel file in which results of microarray or RNA-seq experiments are provided to the users.

Since only one sample at a time is analyzed, normalization, that typically requires of several samples, must be carried out previously by external programs such as the Babylomics (10) or others.

Gene interventions

Once the file corresponding to the reference condition is uploaded, PathAct calculates the activity status of all the signaling circuits in which the 58 KEGG pathways analyzed can be decomposed. Here, we focus on effector proteins, at the end of the pathways, which are the ultimate responsible for the cellular response to stimulus by triggering specific cell functionalities. Therefore, a signaling circuit is defined here as the sequence of proteins that connect an effector protein back to all the possible receptor proteins from which the signal transmission is initiated upon stimulus reception (5,6). PathAct uses a simplified version of the probabilistic models previously reported (5,6) in which the normalized gene expression values are directly taken as proxies of gene activity (instead of using empirical distributions as in the original formulation of the method), which expands the use of the method to other technologies beyond Affymetrix microarrays, such as RNA-seq. Then, the signal is propagated from the receptor nodes to the final effector nodes in a similar way that in the probabilistic model, as a product of probabilities of gene activity, considering that both activations and inhibitions can coexist within circuits. The circuits, colored according to their activity status, are represented within the pathways in the graphical working environment (*Pathway viewer* panel, Figure 1B). The *Pathway list* panel (Figure 1C) contains the list of pathways. Clicking on them will bring the corresponding pathway to the *Pathway viewer*.

The intervention on a particular gene can be directly carried out within the working environment by setting the cursor over the node targeted, which immediately prompts the

normalized value of gene expression (see Figure 1D, *Activity value window*). Such value constitutes its contribution to the signal transmission that could be very relevant if it is a bottleneck for the signal or irrelevant if, for example, lies in one branch of a redundant bifurcation. In addition, the gene can be present in more than one pathway which means that it can be contributing to several signaling circuits simultaneously. The value prompted can be modified to simulate interventions. The selected genes appear in the *Add genes* panel (Figure 1E), where the value of the intervention can be defined. Thus, a KO can be simulated by setting a gene contribution to 0 (or to a low value). Conversely, an over-expression of an inactive gene can be simulated by setting its contribution value to 1 (or to a high value). Additionally, a gene name can be typed in the *Add genes* search box (Figure 1F). Once all the desired interventions have been made, the button *Update* (Figure 1G) recalculates the predicted signaling status of the resulting simulated condition. Then, the simulated condition is compared to the reference condition and circuits are colored according to the activity status changes (red indicates a significant increase in the signaling status and blue a significant decrease). The significance of the change, in absence of a conventional testing scenario, is given by a user-defined threshold, which is set to 2 by default. The *Pathway list* panel summarizes result of the comparison by highlighting pathways in which one or more signaling circuits have significantly changed. The *Circuit List* panel (Figure 1H) displays the significant circuits found within the pathway selected. Clicking on the circuits will interactively highlight them in the *Pathway viewer* panel.

Simulating drug effects

Also the effect of drugs with known targets (as described in DrugBank) over the different signaling pathways can be studied. The *Add drugs* search box (Figure 1I) allows selecting drugs for simulating the effect on the uploaded system. The genes targeted by the selected drugs appear in the *Genes affected by drugs* panel (Figure 1J). Additionally, any time a gene is selected the *Select gene related drug* list panel (Figure 1K) displays all the drugs known to target such gene. When a drug is selected its known targets are displayed in the *Genes affected by drugs* panel (Figure 1J). If the *Update* button (Figure 1G) is pressed then the effect of the drug over the corresponding gene targets in the *Genes affected by drugs* panel is simulated. Contrarily to the case of gene KOs or over-expressions, in the case of drug action, the predicted gene activity is obtained by multiplying the *drug action weight* by the original normalized target gene expression value. Thus a weight of 0 will set the gene activity to 0 but a coefficient of 0.1 will turn the original gene activity to its 10% (not to a gene activity value of 0.1). Since gene activity values must range between 0 and 1, activity values over 1 after the application of an agonist weight are trimmed to 1. The intensity of the effect produced by the drugs can be modulated in the *Configure Drug Action Weights* window invoked using the *Settings* option of the main menu.

Since many genes participate in more than one pathway and many drugs affect to more than one gene but, simultaneously, signaling circuits are wired with a substantial level

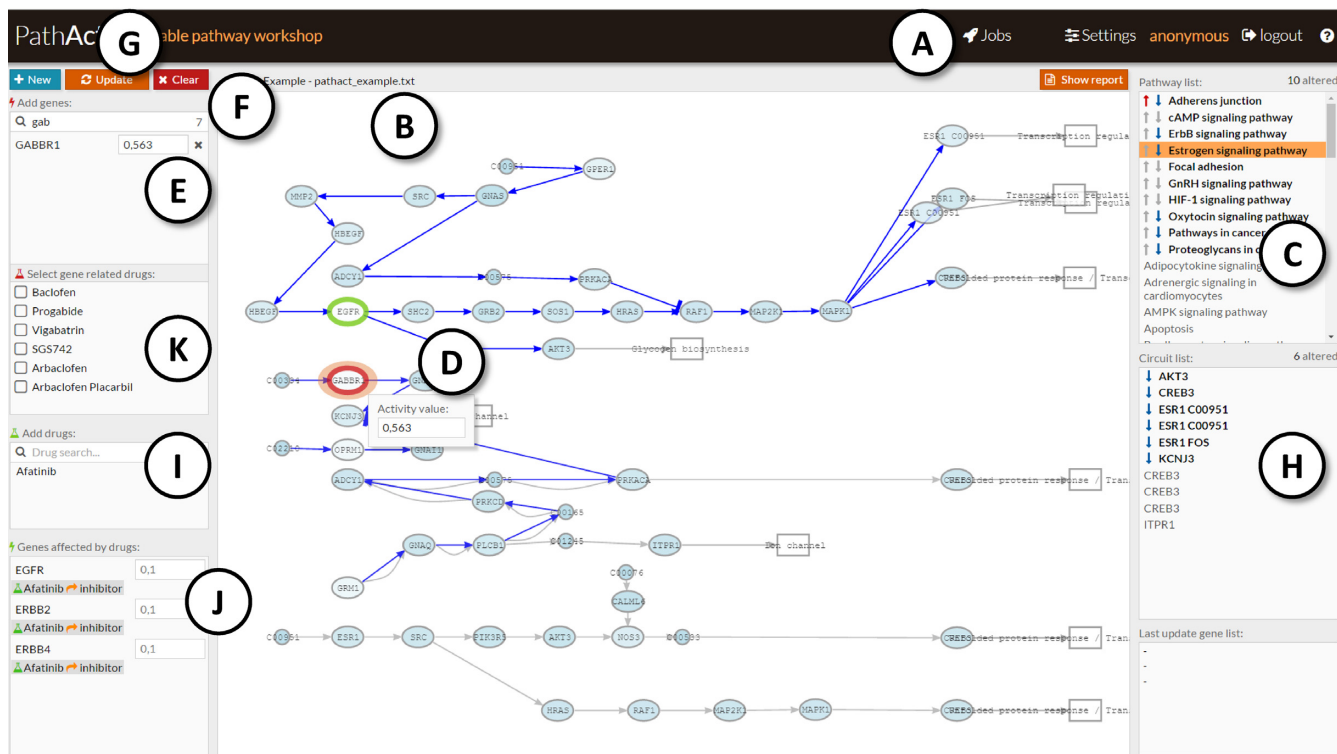


Figure 1. Snapshot of the PathAct working environment. (A) *Jobs* panel, to upload data; (B) *Pathway viewer* panel, where pathways are displayed; (C) *Pathway list* panel, where the pathway to be displayed is selected; (D) *Activity value* window, that reports the gene activity value; (E) *Add gene* panel, lists the gene selected for the intervention; (F) *Add gene* box, where genes can be typed; (G) *Update* button, simulates the intervention; (H) *Circuit List* panel, contains the circuits in the selected pathway; (I) *Add drugs* panel displays drugs typed in the *Add drugs* box; (J) the *Genes affected by drugs* panel contain the genes targeted by the selected drugs; (K) *Select gene related drug list* panel contains drugs known to target genes selected in the *Add genes* list.

of redundancy, the predicted results of drug effects are often unexpected. This fact highlights the importance of comprehensive holistic simulation approaches like the one presented here.

Obviously, unknown off-target effects cannot be predicted. However, they could eventually be inferred by comparing the predicted condition to the real condition measured upon the application of the drug.

An example of the simulation of the activity of a drug, Sorafenib, is illustrated in Figure 2. Sorafenib is known to inhibit tumor growth by a dual mechanism that involves either a direct effect on the tumor by inhibiting the proliferation in several signaling pathways and/or on an indirect effect, preventing tumor angiogenesis by means of the inhibition of VEGF and other signaling pathways (11,12) (see also FDA prescription information: http://www.accessdata.fda.gov/drugsatfda_docs/label/2013/021923s0161bl.pdf). We use the data in the program example, taken from the TCGA (ID: TCGA.BH.A1FM.11B.23R.A13Q.07), which corresponds to a woman with breast cancer, aged 44, that died 1388 days after diagnostic. Once the data are ready, we choose Sorafenib in the *Add drugs* search box (Figure 1I), which includes nine genes affected: *BRAF*, *RAF1*, *FLT3*, *FLT4*, *PDGFRB*, *KIT*, *FGFR1*, *RET* and *FLT1*. VEGF signaling pathway, one the canonical angiogenesis pathways, results with most of their circuits down-activated, with Angiogenesis or Cell adhesion functions directly inhibited (Figure 2A). Moreover, the inhibition of *AKT3*, which is an inhibitor

of *CASP9* and *BAD*, produces an activation of Apoptosis, thus completing the antitumoral effect of the drug. In addition to VEGF signaling pathway other pathways like HIF-1 signaling pathway (by inhibiting *FLT1* circuit; see Figure 2B) is affected and Angiogenesis is again inhibited. In the Focal adhesion pathway the circuit that triggers Apoptosis undergoes a net activation (Figure 2C). Sorafenib also binds to different MAPK proteins, limiting the tumor growth and inducing apoptosis. Some MAPK are found in different pathways, like RAP-1 or Sphingolipid signaling pathways, and their inhibition provokes inhibition of cell cycle function (see more detail in the worked example 4 in the PathAct documentation). Thus, the prediction of the drug effect fits remarkably well with the description of the effect of the drug in patients and xenografts (11,12).

Combinations of drugs and gene KOs or over-expressions can also be simulated. Therefore, when the *Update* button is pressed, the corresponding interventions on genes in *Add genes* panels and KOs over the genes in the *Genes affected by drugs* panel are simulated.

Technical details

PathAct client has been implemented in JavaScript using the HTML5 and SVG standards, which provide a rich and user-friendly interface built with Polymer web components (<https://www.polymer-project.org/1.0/>) and using CellMaps (<http://cellmaps.babelomics.org>) functionalities. The back-

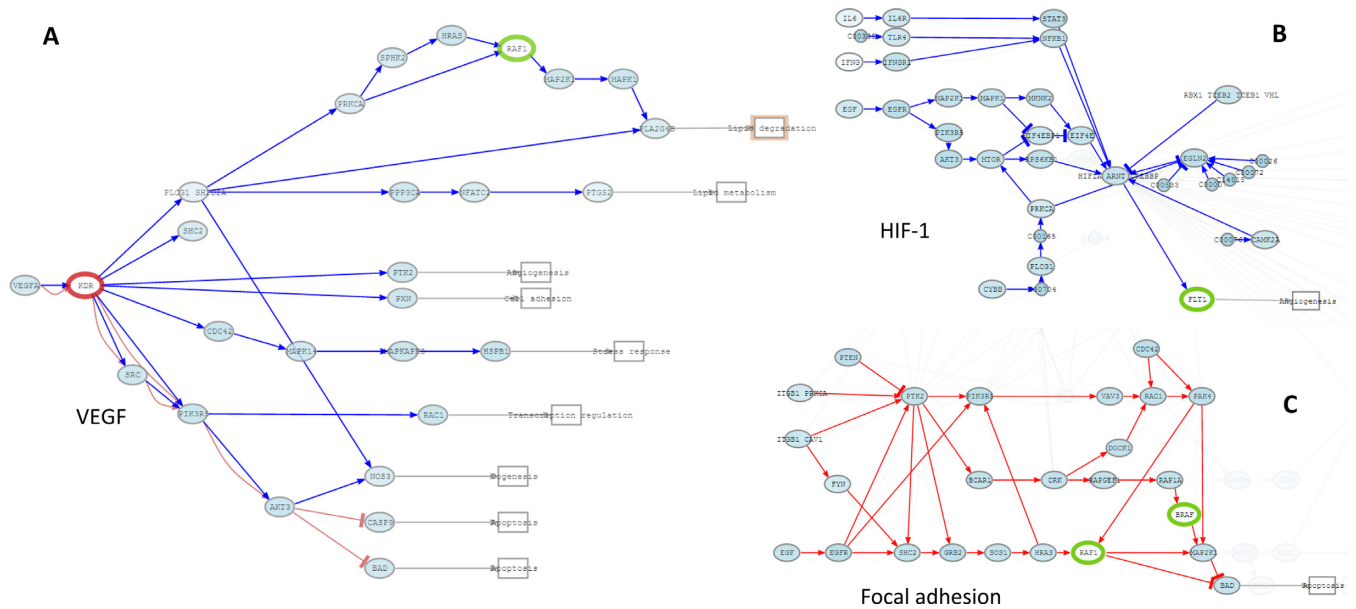


Figure 2. Some pathways affected by Sorafenib. (A) VEGF signaling pathway with Angiogenesis inhibited and Apoptosis activated. (B) *FLT1* circuit of the HIF-1 signaling pathway with Angiogenesis inhibited. (C) *BAD* circuit of the Focal adhesion pathway with Apoptosis activated.

end, which makes the calculations of pathway activities, is written in R. Drugs that target the selected genes are taken from DrugBank (13) through the CellBase (14) webservices. Pathways are taken from KEGG (9).

DISCUSSION

PathAct has been designed to provide a comprehensive summary of the whole (and often unexpected) effect that interventions on one or several genes have over all the signaling circuits in all the pathways. Moreover, since signals received by receptor proteins trigger functions mediated by the effector proteins at the end of the signaling circuits, both the direct and the long-distance functional consequences of interventions over the genes of the circuit can be straightforwardly revealed within this actionable pathway scenario. PathAct actively contributes to save an enormous amount of time and resources in trial-and-error experiments by allowing highly focused testing of hypothesis of intervention over a reduced number of signaling circuits. In this way this tool will decisively contribute to the acceleration of the discovery of new drug targets and can enormously facilitate application of targeted therapies in personalized treatments. To our knowledge, there are no other tools available that offer a similar functionality.

FUNDING

Spanish Ministry of Economy and Competitiveness [BIO2014-57291-R]; ‘Plataforma de Recursos Biomoleculares y Bioinformáticos’ PT 13/0001/0030 from the ISCIII; co-funded with European Regional Development Funds (ERDF); Generalitat Valenciana (GVA-FEDER) [PROMETEOII/2014/025]; Fundació la Marató TV3 [20133134]; EU FP7-People ITN Marie Curie Project [ref 316861]; EU H2020-INFRADEV-1-2015-1 ELIXIR-EXCELERATE [ref 676559]. Funding for open access

charge: Spanish Ministry of Economy and Competitiveness [BIO2014-57291-R];

Conflict of interest statement. None declared.

REFERENCES

1. Jones,S., Anagnostou,V., Lytle,K., Parpart-Li,S., Nesselbush,M., Riley,D.R., Shukla,M., Chesnick,B., Kadan,M., Papp,E. *et al.* (2015) Personalized genomic analyses for cancer mutation discovery and interpretation. *Sci. Transl. Med.*, **7**, 283ra253.
2. Stegmeier,F., Warmuth,M., Sellers,W.R. and Dorsch,M. (2010) Targeted cancer therapies in the twenty-first century: lessons from imatinib. *Clin. Pharmacol. Ther.*, **87**, 543–552.
3. Wagle,N., Berger,M.F., Davis,M.J., Blumenstiel,B., Defelice,M., Pochanard,P., Ducar,M., Van Hummelen,P., Macconail,L.E., Hahn,W.C. *et al.* (2012) High-throughput detection of actionable genomic alterations in clinical tumor samples by targeted, massively parallel sequencing. *Cancer Discov.*, **2**, 82–93.
4. Fey,D., Halasz,M., Dreidax,D., Kennedy,S.P., Hastings,J.F., Rauch,N., Munoz,A.G., Pilkington,R., Fischer,M., Westermann,F. *et al.* (2015) Signaling pathway models as biomarkers: patient-specific simulations of JNK activity predict the survival of neuroblastoma patients. *Sci. Signal.*, **8**, ra130
5. Sebastian-Leon,P., Carbonell,J., Salavert,F., Sanchez,R., Medina,I. and Dopazo,J. (2013) Inferring the functional effect of gene expression changes in signaling pathways. *Nucleic Acids Res.*, **41**, W213–W217.
6. Sebastian-Leon,P., Vidal,E., Minguez,P., Conesa,A., Tarazona,S., Amadoz,A., Armero,C., Salavert,F., Vidal-Puig,A., Montaner,D. *et al.* (2014) Understanding disease mechanisms with models of signaling pathway activities. *BMC Syst. Biol.*, **8**, 121.
7. Hernansaiz-Ballesteros,R.D., Salavert,F., Sebastian-Leon,P., Aleman,A., Medina,I. and Dopazo,J. (2015) Assessing the impact of mutations found in next generation sequencing data over human signaling pathways. *Nucleic Acids Res.*, **43**, W270–W275.
8. Amadoz,A., Sebastian-Leon,P., Vidal,E., Salavert,F. and Dopazo,J. (2015) Using activation status of signaling pathways as mechanism-based biomarkers to predict drug sensitivity. *Sci. Rep.*, **5**, 18494.
9. Kanehisa,M., Goto,S., Sato,Y., Kawashima,M., Furumichi,M. and Tanabe,M. (2014) Data, information, knowledge and principle: back to metabolism in KEGG. *Nucleic Acids Res.*, **42**, D199–D205.

10. Alonso,R., Salavert,F., Garcia-Garcia,F., Carbonell-Caballero,J., Bleda,M., Garcia-Alonso,L., Sanchis-Juan,A., Perez-Gil,D., Marin-Garcia,P., Sanchez,R. *et al.* (2015) Babelomics 5.0: functional interpretation for new generations of genomic data. *Nucleic Acids Res.*, **43**, W117–W121.
11. Adnane,L., Trail,P.A., Taylor,I. and Wilhelm,S.M. (2006) Sorafenib (BAY 43-9006, Nexavar®), a dual-action inhibitor that targets RAF/MEK/ERK pathway in tumor cells and tyrosine kinases VEGFR/PDGFR in tumor vasculature. *Methods Enzymol.*, **407**, 597–612.
12. Wilhelm,S.M., Carter,C., Tang,L., Wilkie,D., McNabola,A., Rong,H., Chen,C., Zhang,X., Vincent,P. and McHugh,M. (2004) BAY 43-9006 exhibits broad spectrum oral antitumor activity and targets the RAF/MEK/ERK pathway and receptor tyrosine kinases involved in tumor progression and angiogenesis. *Cancer Res.*, **64**, 7099–7109.
13. Wishart,D.S., Knox,C., Guo,A.C., Shrivastava,S., Hassanali,M., Stothard,P., Chang,Z. and Woolsey,J. (2006) DrugBank: a comprehensive resource for in silico drug discovery and exploration. *Nucleic Acids Res.*, **34**, D668–D672.
14. Bleda,M., Tarraga,J., de Maria,A., Salavert,F., Garcia-Alonso,L., Celma,M., Martin,A., Dopazo,J. and Medina,I. (2012) CellBase, a comprehensive collection of RESTful web services for retrieving relevant biological information from heterogeneous sources. *Nucleic Acids Res.*, **40**, W609–W614.