Universitat Politècnica de València
Escola Tècnica Superior d'Enginyeria Agronòmica i del Medi Natural

Pathway Activity Analysis (PAA) as a new class of mechanistic biomarker to predict drug responses in drug repositioning for cancer patients

Biotechnology Bachelor's Thesis
(Trabajo Fin de Grado en Biotecnología)
Academic year 2016 - 2017

STUDENT: Anna Bailach Adsuara
TUTOR: Prof. Lynne Yenush
EXTERNAL TUTOR: Dr. Joaquín Dopazo

València, June 2017

Abstract

Abstract – *In recent years, progress in new technologies has resulted in the capacity to generate massive amounts of data, this is known as the "Omics Age". The challenge now is data integration and analysis. Thus, Systems Biology emerges as a solution; where, previously, genetic studies estimated the impact of a single gene, now all gene data can be integrated. This allows for more precise conclusions since diseases and drug responses are caused by different combinations of genetic perturbations. Furthermore, it allows for simulations that would otherwise be prohibitively costly in terms of time and resources.*

In this context, here is presented a new method for the integration of available data for each element of a signalling pathway in the end result of said pathway, the phenotype. This system acts as a mechanistic biomarker, since the difference in activation level present in a pathway, when comparing samples, serves to expose more information about the mechanisms which act in a different manner. A much more informative method than descriptive biomarkers. Additionally, this method allows simulations. When inputting information about a drug's effects, the activity level of the pathway can be modified and an estimation of the desirability of the effects can be made.

Cancer patients frequently respond in an undesirable manner to therapy, a great problem in oncology that is thought to be due to a lack of predictive biomarkers. The activity of pathways in cancerous cells can be used as mechanistic biomarkers. This project intends to exploit this new tool to reposition drugs for cancer patients.

Key Words – *Computational Drug Repositioning, Mechanistic Biomarker, Pathway Activity, Cancer.*

Resumen

Resumen – *En los últimos años, los avances en nuevas tecnologías han permitido generar enormes cantidades de datos, la conocida "Era de las Ómicas". El reto ahora es la integración de datos y su análisis. Así, la Biología de Sistemas emerge como una solución. Dónde los estudios genéticos una vez estimaban el impacto de un solo gen, ahora todos los datos disponibles para todos los genes se pueden integrar. Esto permite llegar a conclusiones más precisas, puesto que las enfermedades y las respuestas a fármacos están causadas por distintas combinaciones de perturbaciones genéticas. Incluso mejor, permite hacer simulaciones que de cualquier otro modo serían increíblemente costosas en términos de tiempo y recursos.*

En este contexto, se presenta aquí un nuevo método para integrar los datos disponibles para cada elemento de un camino de señalización en la actividad final resultante de dicho camino, el fenotipo. Este sistema sirve como un biomarcador mecanístico, puesto que el diferente nivel de activación que presente un camino, al comparar muestras, sirve para indicar mucha más información sobre los mecanismos que están funcionando de forma distinta. Un método mucho más informativo que los biomarcadores descriptivos. Además, el método permite realizar simulaciones. Al introducir información sobre los efectos de un fármaco, se puede modificar el nivel de actividad del camino y estimar si sus efectos son deseados.

Los pacientes con cáncer a menudo no responden deseablemente a una terapia, un gran problema en la oncología que se piensa es debido a la falta de biomarcadores predictivos. La actividad de los caminos de señalización en células cancerígenas puede utilizarse como biomarcador mecanístico. Este proyecto pretende emplear esta nueva herramienta para el reposicionamiento de fármacos en pacientes con cáncer.

Palabras Clave – *Reposicionamiento de fármacos computacional, Biomarcador mecanístico, Actividad de Pathway, Cancer.*

Resum

Resum – *En els últims anys els avanços en les noves tecnologies han permés generar enormes quantitats de dades, la coneguda "Era de les Òmiques". El repte ara és la integració de dades i el seu análisis. Així, la Biología de Sistemes surt com una sol·lució. On els estudis genètics estimaven l'impacte de un sol gen, ara totes les dades disponibles per a tots els genes es poden integrar. Açò permet arribar a conclusions més precisses, doncs tant les enfermetats com les respostes a fàrmcs están causades per distintes combinacions de perturbacions genètiques. El que és millor, permet realitzar simulacions que de qualsevol altra manera serien increïblement costoses en temps i recursos.*

En aquest contexte, es presenta ací un nou mètod per a integrar les dades disponibles per a cada element d'un camí de senyalització en l'activitat final resultant d'aquest camí. Aquest sistema s'empra com a biomarcador mecanístic perque el diferent grau d'activació que presente un camí de senyalización representa información sobre els mecanismes que están funcionant de forma diferent. Un mètode molt més informatiu que els biomarcadors descriptius convencionals. A més, el mètode permet realitzar simulacions. A l'introduir dades sobre fàrmacs el nivell d'activitat del camí es modifica i es pot estimar si els efectes del fàrmac son desitjats.

Els pacients amb càncer, solen no respondre de la manera desitjada a teràpies, en gran mesura per la falta de biomarcadors predictius. L'activitat dels camins de senyalización en cèlules cancerígenes es pot emprar com a biomarcador mecanístic. Aquest projecte pretén emprar aquest nou mètode per al reposicionament de fàrmacs en pacients amb càncer.

Paraules Clau – *Reposicionamient de fàmacs computacional, Biomarcador mecanístic, Activitat de Pathway, Càncer.*

Acknowledgements

I would like to thank the Computational Genomics laboratory at the Príncipe Felipe Research Center. Specially to Dr. Joaquín Dopazo for letting me work in such a new field and to Dr. Jose Carbonell, for supervising my work.

I would specially like to thank Lynne Yenush and Ismael Rodrigo, for being there when I needed academic counsel and for supporting me in my career decisions, and to Javier Forment for introducing me to Bioinformatics.

General Index

Table Index

Figure Index

Abbreviations

BRCA	Breast Cancer BRCA
CCAA	Canonical Circuit Activity Analysis
CNA	Cancer Network Activity
ComBat	Combined Association Test for Genes
CSV	Comma-Separated Values
DIRPP	Drug Intervention Response Prediction with Paradigm
DMC	DrugMap Central
GO	Gene Ontology
KEGG	Kyoto Encyclopedia of Genes and Genomes
KIRC	Kidney Renal Clear Cell Carcinoma
KO	Knock Out
MoA	Mode of Action
NGS	Next Generation Sequencing
PAA	Pathway Activity Analysis
PheWas	Phenome-Wide association
TCGA	The Cancer Genome Atlas

1. Introduction

1.1. Systems Biology

Since genome sequencing became possible after Frederick Sanger developed the first sequencing method, many advancements in the field have been made, and many sequences from several genomes have been obtained, thanks to the advancements in various omics disciplines. From that moment until now, Bioinformatics has been essential to deal with the increasing amount of information available by developing algorithms to assess the relationships among large data sets in order to understand complex biological processes. The applications of this field go from sequence assembly and evolutionary biology to simulations on cell biology, also known as Systems Biology.

A living cell can be viewed as a dynamic system in which a large number of different substances react continuously with one another. In order to understand the behaviour of a dynamic system with numerous interacting parts, it is usually insufficient to study the behaviour of each part in isolation. Instead, the behaviour must be analysed as a whole (Kitano, 2001). Driven by advances in biology, engineering, and informatics, Systems Biology emerges as the engineering field that studies biological systems.

1.1.1. Computational Drug Repositioning

Traditional drug discovery has focused either on phenotypic effects or on target-based activities of specific molecules. But simply knowing a single target and its interactions with a drug is not sufficient to predict the clinical success of the drug. Instead, it is the effect produced by the interaction of a drug with its target within the context of a complex biological system that has been corrupted by pathophysiological mechanisms what best predicts the success (Waldman & Terzic, 2013). Systems Biology processes different types of data, such as genes, their mutations, their level of expression or any other desired variable and thus, allows for enhanced comprehension of biological problems. A comprehensive, systems-level approach poses an optimized strategy for drug discovery. It is important to take a systems-level approach, in the context of highly validated predictive models, to maximize the discovery of effective therapeutics (Waldman & Terzic, 2013). But not only is this integration essential to success in drug discovery, but it is also critical to repurpose established agents for new therapeutic indications (Waldman & Terzic, 2013).

Developing a new drug from original idea to launch of a finished product is a complex process which can take 12–15 years and cost an excess of $1 billion (Hughes, Rees, Kalindjian, & Philpott, 2011). *De novo* drug discovery has been experiencing rising costs (Booth & Zemmel, 2004) while the number of new drugs approved has plateaued (Li et al., 2016). Drug repositioning concerns the detection of new clinical indications for drugs already in the market, which decreases the length of the process tenfold (Ashburn & Thor, 2004). With the accumulation of large volumes of omics data, bioinformatics plays an important role in the discovery of new indications for drugs, since it allows for new repositioning strategies and approaches (Li et al., 2016). Given the large number of

druggable protein targets and existing drugs, it is infeasible to set up assays to test every interaction in the laboratory (Li & Jones, 2012).

1.1.2. Computational Drug Repositioning Approaches

The first studies on drug repositioning focused on exploring common characteristics among drugs. One of the first aspects considered was chemical structure. It is assumed that chemical structures often affect proteins in similar ways, therefore similar drugs in their structure share a similar mode of action (MoA). But the rapid advancements in genomics have led to new sources of information to be considered: genetic data, transcriptomic data, and phenotype data. Genetic mutations can serve as biomarkers for drug responses and transcriptional data allows comparison of gene expression profiles to make drug-disease pairs that can be used for drug repositioning purposes (Li et al., 2016). On the other hand, Phenome-Wide association studies (PheWas) detect associations between genetic markers and human diseases. Because the phenome reflects information on clinical side effects (Rastegar-Mojarad, Ye, Kolesar, Hebbring, & Lin, 2015), based on the assumption that similar side-effects may share similar therapeutic properties, drugs can be repositioned using this approach.

Computational drug repositioning studies more and more aim to integrate information from all the above-mentioned categories, because strategies based on a sole category are incapable of capturing associations not manifested at the assessed level (Gottlieb, Stein, Ruppin, & Sharan, 2014). For instance, the lack of resolution of structural data for targets makes drug-target predictions by chemical structure not very accurate (Li et al., 2016). To develop drug repositioning models, different computational approaches can be used. The most well-known being text mining, machine learning, and network analysis.

Semantic Inference/Text Mining
The available information from literature and databases for any disease, drug or protein is huge. Biological ontology makes it possible for the comparison and analysis of biological information from different sources. Text-mining approaches automatically retrieve relevant information on the disease, proteins and cell processes. From the retrieved knowledge, new knowledge can be inferred (Tari & Patel, 2014). By finding relevant knowledge through text mining approaches it is possible to detect novel indications for existing drugs (Li et al., 2016). For instance, DrugMap Central (DMC), an online tool, enables the users to integrate, query, visualize, interrogate, and download multi-level data of known drugs or compounds quickly for drug repositioning studies all within one system (Fu et al., 2013).

Machine learning
Machine learning is the subfield of computer science that confers computers the ability to learn without being explicitly programmed for a task, so that when presented with new data the model is updated. For instance, the DDR unified computational framework constructs drug similarity and disease similarity matrixes based on genome, phenome and chemical structure, weighing information sources based on their contributions to the prediction, to

finally solve the drug–disease network analysis as an optimization problem (Zhang, Wang, & Hu, 2014).

Network analysis

Molecular interactions in the biological systems can be modelled, such as protein interactions, drug-target interactions or signalling networks. These networks have proven very useful in the identification of therapeutic targets or characteristics of drug targets, thus providing new opportunities for drug discovery or repositioning (Li et al., 2016). For example, the Cancer Network Activity (CNA) developed by Serra-Musach *et al.* (Serra-Musach et al., 2016) uses the human interactome network with gene expression measurements from cancer cell lines whose sensitivity to cancer drugs is determined. Then, it assigns a score to each cell line and these scores are evaluated for their correlations with types of drugs and therapies (Serra-Musach et al., 2016). The main limitation of this approach is that it is based on assays performed with drugs given to cancer cell lines, it does not use data from cancer patients. Even if the simulation is performed on the human interactome, it is not as realistic a model as it could, especially for cancer, because of the huge inter-patient and intra-tumour heterogeneity that is characteristic of the disease. The Same problem is found in the Drug Intervention Response Prediction with Paradigm (DIRPP) developed by Brubaker *et al.* which assesses the response of cell lines to a drug intervention from molecular data to predict drug-resistant cancer cell lines and pathway mechanisms of resistance (Brubaker et al., 2014).

The problem with the aforementioned approaches is that experimental variation from the different sources of data, for instance across batches for gene expression experiments, causes 'noise'. In addition to this, those strategies that assess several variables become increasingly difficult to employ as the number of variables increases, leading to enormous matrixes and complex algorithms to deal with them. Thus, the network-based analysis seems to be the best approach.

1.2. Mechanistic Biomarkers

A biomarker is a characteristic that can be objectively measured as an indicator of normal or pathologic biological processes, or as an indicator of response to therapy. The concept of an actionable biomarker is based on the expectation that results of biomarker testing can be used to guide clinical management of disease (Robinson, Lindstrom, Cheung, & Sokolove, 2013).

Some biomarkers are products of the disease itself or of disease-induced damage. These descriptive biomarkers reflect the state of a disease but are not directly involved in disease pathogenesis. Since such biomarkers are by-products rather than players in the disease, they are less powerful in obtaining reliable pharmacodynamic, diagnostic or prognostic information (Robinson & Mao, 2016).

On the other hand, there are biomarkers which are rooted in the biologic mechanisms of disease, the mechanistic biomarkers. These biomarkers have the greatest potential for guiding clinical decision making and are superior to descriptive biomarkers for several reasons (Robinson, Lindstrom, Cheung, & Sokolove, 2013):

1. Since mechanistic biomarkers are involved in the pathogenesis of a disease, they are more likely to be specific to that disease and thus proves more powerful in diagnosing the disease.
2. Similarly, can be used to classify the subtype of the disease.
3. Because it is rooted in the mechanisms, it is also a better pharmacodynamic biomarker, better informing of the efficacy of a treatment, rather than simply improving the symptoms.
4. Similarly, in pharmacodynamic studies, provides information on the mechanism that is working or failing in a therapy, allowing for a more rational design of a therapy.
5. Ideal for personalized medicine. Personalized medicine requires predictive data about the disease and its sensitivity to treatment.

Mechanistic biomarkers can take the form of several different molecules or cell types, which may all contribute to the pathogenesis of a disease: genes, cytokines, antibodies, immune cells or even cell signalling. From these, the most promising biomarker seems to be cell signalling. To study a disease first one needs to consider its phenotype is multigenic. Single gene biomarkers frequently lack any mechanistic link to the fundamental processes responsible for disease progression or therapeutic response. Such processes are better understood as pathological alterations in the normal operation of cells caused by different combinations of gene perturbations (mutations or gene expression changes) (Hidalgo et al., 2017).

1.3. Cell Signalling Pathways as Mechanistic Biomarkers in Cancer

A major challenge in anticancer drug development is the inability to identify cancers that are most likely to respond to a treatment, which exposes patients to the risks of ineffective treatments (Fang, Mehran, Heymach, & Swisher, 2015).

Signalling pathways control cell fate decisions that ultimately determine the behavior of cells. These cascades trigger particular effects after the transmission of a signal through all the elements that conform the cascade, which are all functional proteins. Therefore, pathological alterations are the result of different combinations of gene perturbations and the consequent alteration of the normal communication among proteins in the signalling cascades. Particularly in cancer, the alteration of signalling pathways plays a key role in its origin and progression. This was proven in the model developed by Fey *et al.* (Fey et al., 2015) in which they modelled the JNK pathway in neuroblastoma cells revealing its central role in the disease. In addition to this, they successfully employed it as a biomarker to stratify neuroblastoma patients across different individual molecular backgrounds and

showed that the activity of the pathway better correlated with patients' mortality than the activity of the conventional biomarker gene *MICN* (Fey et al., 2015).

Signalling pathways are functional modules in the cell which provide a more informative insight of cellular function than conventional methods because their activity reflects how the information has been modified by all its components. It is the final level of activity of a pathway what will command a cell to carry out a certain function. Therefore, the dynamics of pathway activity may contain relevant information on prognosis different from that contained in the static nature of other types of biomarkers, as shown above.

One of the great advantages of working from a network approach using signalling pathways is that the activity of each element in a network can be modified depending on the variables considered (such as mutations or level of expression of a gene), but the essential information comes from the output element, overcoming the problem of the computational strategies that lead to complex algorithms as the number of variables increases. These variables do not need to be weighed, since the final activity of the pathway is a result of the active proteins in the pathway. Thus, Pathway Activity Analysis (PAA) emerges as an alternative way of defining a new class of mechanistic biomarkers, whose activity is related to the molecular mechanisms that account for disease progression or drug response (Hidalgo et al., 2017).

1.3.1. Canonical Circuit Activity Analysis (CCAA)

The Canonical Circuit Activity Analysis (CCAA) strategy developed by Hidalgo *et al.* proved to be a very powerful computational network-based approach to find mechanistic biomarkers based on PAA. In the paper, it demonstrated a high diagnostic value and related to disease outcomes in an extensive analysis involving 5640 patients from 12 different cancer types. One of the most important findings was that the expression profiles from different cancers resulted in the activation of the same final functions in the cell, which also further supported the notion that signalling pathways determine the cell fate of cancer cells, validating them as useful mechanistic biomarkers (Hidalgo et al., 2017). In this case, the proof of concept was done in the context of cancer, but it could be used to model and study any other disease. In fact, two user-friendly websites are available for free for this purpose.

The first one is the web tool hiPathia, for the interpretation of the consequences of the combined changes of gene expression levels and/or genomic mutations in the context of signalling pathways. It transforms uninformative gene expression and/or genomic variation data into signalling circuit activities, which carry information on the different cell functionalities triggered by them. Such signalling activities not only account for the underlying molecular mechanisms of diseases or the mode of action of drugs but they can also be used as mechanistic features for the prediction of complex phenotypes (http://hipathia.babelomics.org).

The second one is PathAct (http://pathact.babelomics.org), which enables the study of the consequences that Knockouts (KOs) or over-expressions of genes can have over signalling

pathways. PathAct implements robust models of signalling pathways within an advanced graphical interface that provide a unique interactive working environment in which actionable genes, that could become potential drug targets, can be easily assayed alone or in combinations. Also, the effect of drugs with known targets over the different signalling pathways can be studied. Since signals trigger functions across the pathways, the direct and long-distance functional consequences of interventions over genes can be straightforwardly revealed through this actionable pathway scenario (Salavert et al., 2016).

When compared to other methods, CCAA performed better. It is highly specific, meaning it has a very low false positive rate, and has a high sensitivity, high true positive rate. In addition, this method handles loops in the signal transmission, which better represents the functioning of signalling pathways in the cell. But the most important aspect of the method is that the association of the activity of a circuit to mortality of a patient was higher than the individual association of any of the genes that formed the circuit (Hidalgo et al., 2017).

1.3.2. Experimental Design

The aim of this work is to employ the CCAA using PAA as a mechanistic biomarker for drug repositioning for cancer patients. In order to do so, drug responses are modelled at a functional level using the same data which validated the method in the original paper (Hidalgo et al., 2017) from Kidney Renal Clear Cell Carcinoma (KIRC) and Breast Cancer BRCA patients (BRCA). The idea is to see how cellular functions in cancer cells can be reversed to a normal profile upon applying a drug.

The program requires the functional annotation of the final proteins of each pathway. The involvement of several cell functionalities in cancer pathogenesis was validated in the study (Hidalgo et al., 2017). The key words given to the program should therefore reflect relevant functions in the context of cancer. Annotation of drugs is also required because there is no data on drug-target activity shift as a result of administering the drug, so all drugs need to be classified either as having an activating or inhibitory effect or removed from the assay. A matrix is created containing the information on the drug's real name, code, effect, targets and a translated target ID for the program to identify which element to modify.

The program is expanded to estimate how signalling pathways are affected by a drug and uses functional annotations to transform values of applicable signalling pathways to values for each annotated function. Statistical analysis is then used to assess the difference in function value for normal and tumour data, and the effect drugs have on said function value. The program has can be used in two possible settings: either as a personalized tool or for massive screening of drugs for repositioning.

2. Objectives

The aim of the project is to develop a program that employs the CCAA method using PAA as mechanistic biomarkers for drug repositioning purposes. To achieve this, several goals need to be accomplished first.

1. Annotate relevant functions in cancer to improve the program's predictions.
2. Improve the algorithm that will transform the values obtained from signalling pathways into values of functions relevant in cancer.
3. Categorize drugs into activating and inhibiting depending on their characteristics and targets to provide the program with information.
4. Program the adequate formatting of drug data.
5. Program the algorithm that modifies the patients' data upon administration of a drug.
6. Program the statistical analysis to have a complete tool usable in two settings, either for a personalized approach and the other one at cohort level.
7. Validate the method as a tool for drug repositioning purposes: find potential drugs for KIRC and BRCA cancers.

3. Materials and Methods

3.1. Pathway Activity Analysis

Pathway Activity Analysis (PAA) refers to the analysis of the collective contribution of genes to the final signal transmission across signalling pathways. Individual contributions are deduced from gene expression values, mutations or any other variable which might contribute to a change on its activity. The method is written in the open source R programming language (www.R-project.org) and available in the GitHub repository for download (www.github.com/babelomics/hipathia).

3.1.1. Modelling Framework

The method requires a description of the relationship among the proteins that form a pathway. The Kyoto Encyclopedia of Genes and Genomes (KEGG) is a database resource for understanding high-level functions and utilities of the biological system, such as the cell, the organism and the ecosystem, from molecular-level information, specially large-scale molecular datasets generated by high-throughput experimental technologies (Kanehisa, Sato, Kawashima, Furumichi, & Tanabe, 2016). From KEGG, sixty pathways related to signalling are selected to model the framework.

The library igraph is freely available in a GitHub repository, and is the basis on which the original program is developed (https://github.com/igraph/). It is a R package that provides routines for simple graphs and network analysis. It can handle large graphs very well and provides functions for generating random and regular graphs, graph visualization, centrality methods and much more. With this package, KEGG networks are built in R.

3.1.2. Signal Propagation

In order to transmit a signal along a pathway a protein needs to be present and functional, and to be activated by another protein. The activity of each protein is quantified as a normalized value between 0 and 1. These values are inferred from proteomic data, phosphoproteomic data, transcriptomic data, or any other data that can be interpreted as level of activity of a protein. In this case the amount of mRNA corresponding to a protein is used as presence of a protein. However, the intensity of the signal transmitted does not correspond to the level of activity of the protein itself, but is modulated by the level of the signal that arrives to it. Ultimately the level of the signal transmitted across the pathway corresponds to the final value transmitted by the last protein of the pathway, called the effector protein. These are the proteins that ultimately trigger the cell functions activated by the pathway (Hidalgo et al., 2017).

In a pathway, an incoming signal initiates the signalling cascade and the signal is transmitted along the pathway until it reaches the final effector proteins, triggering the cell to perform a function. However, many different incoming signals may reach the same effector. Those sub-pathways that transmit an input signal from a unique receptor to a

unique effector node are called circuits. The effector circuits then are the combination of all those circuits that trigger the same output (activating the same effector), which will perform a function in a cell. Therefore, the triggered function is the combination of the resulting signals from all effectors that trigger said function (See Figure 1).

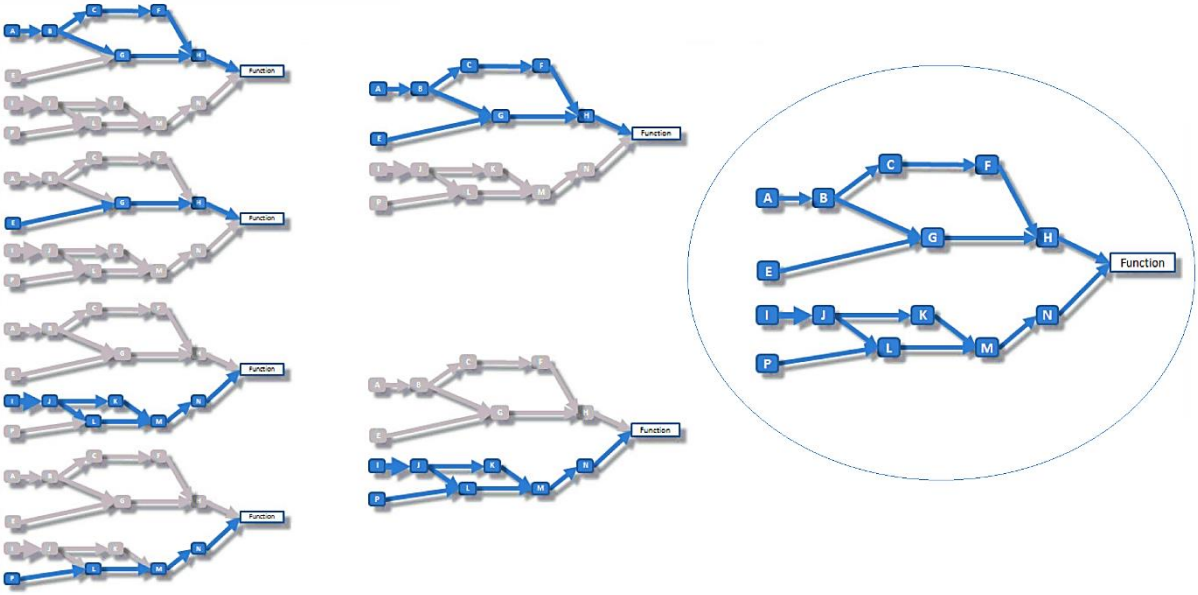

FIGURE 1. Relationship among circuits, effector circuits and functions (Hidalgo et al., 2017). Left: signalling circuits, the sub-pathways that transmit signals from a unique receptor to a unique effector node. Centre: effector circuits that represent the combined activity of all the signals that converge into a unique effector node. Right: functional activity that represents the combined effect of the signal received by all the effectors that trigger a particular cell function.

To compute the signal arriving to the effector proteins, the first step is to calculate the node activity of each node in the network. The node composition of the network is provided by KEGG. A node can be plain (with one or more proteins) or complex (as in a protein complex). For plain nodes, the value is obtained as the percentile 90 of the values of the proteins contained in it. For complex nodes, the limiting compound of the complex is taken as the node activity value. Then, the circuit is initialized with a standard input signal of one, and the propagation of the signal across the pathway is calculated with the following iterative algorithm (Hidalgo et al., 2017):

$$S_n = V_n \cdot \left[1 - \prod_{S_a \in A}(1 - S_a)\right] \cdot \prod_{S_i \in I}(1 - S_i) \qquad (1)$$

Where:

Sn is the signal intensity for the current node ***n*** and ***V*** its normalized value.
A is the total number of activation signals (***Sa***) arriving to the current node.
I is the total number of inhibitory signals (***Si***) arriving to the node.

On Figure 2 there is a schematic representation on how the node values are modified depending on the signals received and the final signal it transmits to the next node. To handle with the loops that are commonly found in signalling pathways, whenever a signal arrives to a node, its value is updated only if the difference between the new value and the

previous value is greater than a threshold (Hidalgo et al., 2017). The overall process is presented in Figure 3.

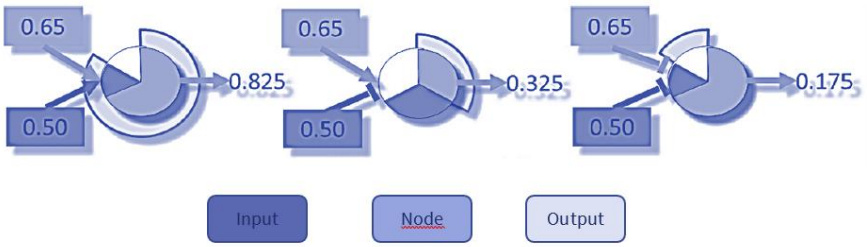

FIGURE 2. Representation of the modification of nodes during propagation of the signal (Hidalgo et al., 2017). The three possibilities are represented. Left: combined activity of two activators. Centre: combined activity of an activator and an inhibitor. Right: combined activity of two inhibitors.

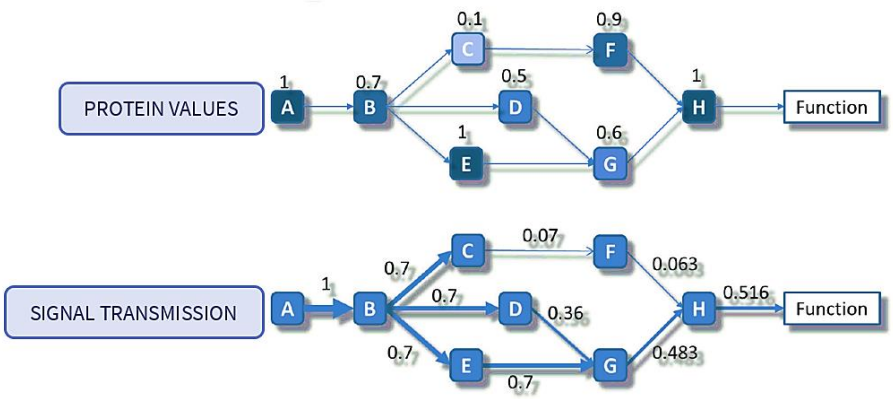

FIGURE 3. Representation of signal propagation (Hidalgo et al., 2017). On top, the normalized values of gene expression are assigned to the corresponding nodes in the circuits. At the bottom, the signal is propagated and the node values are modified.

Many protein effectors can contribute to a cellular function. Therefore, an algorithm is required to estimate the overall contribution each effector and signalling pathways has on said function. The original method used an algorithm similar to the one used to estimate the contributions of all signals on a node. However, it seemed to saturate the activities of functions too much. An improved algorithm needs to be developed to solve this (Annex 3, Script 1).

3.2. Data Sources and Processing

The program requires a file with the final effector proteins from the pathways and their functions in order to perform the functional analysis. Since the context is cancer, the functional annotation will be based on those pathways found relevant for the disease, based on the previous work by Hidalgo *et al.* and KEGG's human pathways on cancer. For instance, cell proliferation is a cancer hallmark and therefore pathways such as cell cycle progression need to be annotated. These annotations are done manually based on bibliographic research and information from KEGG (Kanehisa, Sato, Kawashima, Furumichi,

& Tanabe, 2016) and UniProt, a high-quality and freely accessible resource for protein sequence and functional information (The UniProt Consortium, 2017). The 17 annotated pathways are listed in Table 1 and the complete functional annotation is found on Annex 1, where the pathways are organized in separate tables and each table contains information on the subpathway or condition and their final protein effectors, an explanation on their functions and the Key Words that represent those functions.

TABLE 1. Annotated pathways relevant in cancer.

KEGG Identifier	KEGG Pathways
hsa04010	MAPK Signalling Pathway
has04115	p53 Signalling Pathway
hsa04310	Wnt Signalling Pathway
hsa04350	TGF-β Signalling Pathway
hsa04510	Focal Adhesion Pathway
hsa04520	Adherence Junction Pathway
hsa04530	Tight Junction Pathway
hsa04150	mTOR Signalling Pathway
hsa04152	AMPK Signalling Pathway
hsa04151	PI3K-AKT Signalling Pathway
hsa03320	PPAR Signalling Pathway
hsa04370	VEGF Signalling Pathway
hsa04630	Jak-STAT Signalling Pathway
hsa04024	cAMP Signalling Pathway
hsa04340	Hedgehog Signalling Pathway
hsa04110	Cell Cycle
hsa04210	Apoptosis

The most useful resources for computational methods for drug studies are datasets of known interactions (Li & Jones, 2012). In this case, DrugBank (Wishart, 2006) was used both as the source of drugs and for their annotation, along with bibliographic information. Many drugs have known effects and interactions, but often the mechanism behind them is not clear. Also, some of their effects are not descriptive of the impact they would have in a signalling context, meaning it is not clear whether they would contribute to an activating or inhibitory effect on a node in the signalling network. For these reasons, many terms need to be manually curated. In the Annex 2, the terms used by DrugBank are explained and classified into “Activating” or “Inhibiting” and some particular cases are removed. The compiled data from DrugBank (Unpublished data from study, Untitled work) is processed with the Script 2 on Annex 3 to remove drugs which lack information on their effect and to format them into a manageable data frame. This data frame is then further expanded with the annotated activating and inhibiting actions each drug has on their targets.

To perform the analysis, expression values from KIRC and BRCA cancers are employed. The Cancer Genome Atlas (TCGA) is a project to catalogue genetic data on cancer, using high-throughput genome analysis techniques, and has a free repository to download data (https://tcga-data.nci.nih.gov/tcga/). The expression data from KIRC and BRCA cancers contained RNA-seq samples sequenced from tumour biopsy and samples from normal adjacent tissue. KIRC data has a total of 526 primary tumours and 72 of normal adjacent

tissue, while BRCA has 1057 primary tumours and 113 normal adjacent tissue. RNA-seq consists on Next-Generation Sequencing (NGS) of the transcripts found in a sample, so that expression values can be obtained by quantifying these transcripts. One of the advantages of having the patients' expression profile in RNA-seq is that the batch effect caused by the different origin of each data can be solved applying the Combined Association Test for Genes (ComBat) method (Johnson, Li, & Rabinovic, 2007).

To model the impact a drug has on the altered signalling pathways from tumours, the expression value of the drug's target is multiplied by 0.001 for an inhibitory drug or modified to 0.99 in the case of an activating drug (See Annex 3, Script 3). With these parameters, the potential consequences of a highly effective drug can be simulated, so that conclusions can be drawn on the impact it could have on the altered functions of a tumour after processing of the signal.

3.3. Combined code to run the program

All steps need to be brought together in a coherent and structured way, and precise instructions are given to the program to perform the desired analysis (Annex 3, Script 4).

The program requires as input: the experimental data and design (expression values and conditions), a data frame with the drug information (generated running Annex 3, Script 2), the annotation file for functions and a file containing the desired drugs to test. The output is a data frame containing the function values for each analyzed individual/sample.

With the experimental data, the program creates a data frame containing all samples/individuals and their expression values for each gene. Then this data frame is extended by adding the modified Tumour data for each drug present in the given file, as described in the previous section (Annex 3, Scripts 3 and 4). The example on Table 2 represents a data matrix generated from expression values for genes A – E for two individuals and each condition in an assay with two drugs.

TABLE 2. Example for an Expression Data Frame. Columns are conditions and individuals and rows are genes.

	Normal	**Normal**	**Tumour**	**Tumour**	**Drug 1**	**Drug 1**	**Drug 2**	**Drug 2**
	1	2	1	2	1	2	1	2
Gene A								
Gene B								
Gene C								
Gene D								
Gene E								

Applying the algorithm described in Equation 1, a data frame is generated like in Table 3. In the example, the original expression data contributes to the Signalling Pathways A – C.

TABLE 3. Example for Signalling Pathways Data Frame. Columns are conditions and individuals and rows are signalling pathways.

	Normal	Normal	Tumour	Tumour	Drug 1	Drug 1	Drug 2	Drug 2
	1	2	1	2	1	2	1	2
Pathway A								
Pathway B								
Pathway C								

Finally, with the annotation file and the pathway values are used to estimate the values for each function performed by these pathways (See Annex 3, Scripts 1 and 4). In the example on Table 4, the pathways have contributed to Functions A and B.

TABLE 4. Example for a Functions Data Frame. Columns are conditions and individuals and rows are signalling pathways.

	Normal	Normal	Tumour	Tumour	Drug 1	Drug 1	Drug 2	Drug 2
	1	2	1	2	1	2	1	2
Function A								
Function B								

3.4. Statistics

To estimate the effect each drug has on the tumour profile, a Mann-Whitney-Wilcoxon Test is performed. It is a non-parametric test that can be used to determine whether two dependent samples were selected from populations having the same distribution. This test is usually used when a normal distribution cannot be assumed, it would be the equivalent of a t-Test in a normal distribution. The null hypothesis for this test is that the populations are identical. With a cut-off value of 0.05, when comparing two distributions, if the p-value is lower than 0.05, then the distributions are significantly different (Kruschke, 2011).

With this test, Normal and Tumour distributions can be compared and those functions which are significantly different in a Tumour condition are then compared to the Drug distributions. R has a function called *wilcox.test* is implemented in the code to perform the analysis (Annex 3, Script 5).

3.5. Plots

To see the resulting distributions the plotly library for R is used (https://plot.ly). Plotly provides online graphing, analytics, and statistics tools for individuals and collaboration, as well as scientific graphing libraries for Python, R, MATLAB, Perl, Julia, Arduino, and REST. The free version allows for customizing graphs working in R but also online.

4. Results and Discussion

4.1. Functional Annotation

One of the purposes of this project was to analyse the effect on cell functions and not on the signaling pathways. Therefore, a file containing the functions performed by all the effector proteins of all pathways is required. Since the disease being studied is cancer, a manual annotation of the effectors of relevant pathways for the disease was required to have functions that are descriptive or relevant to the disease, based on literature. The written annotations can be found on Annex 1 and the final annotation file was named *annot_manual_hsa.annot* (Annex 3, Script 4) and saved into the annotations folder of the original method. The format of the annotation file is a tab-separated file with the first column containing the gene names in upper case and the second column containing the key words chosen to describe their functions. This file needs to be provided for the functional analysis as seen on Annex 3, Script 4.

4.2. Function Analysis Algorithm

In the functional analysis, when using the original algorithm, it seemed to saturate the activities of functions. The new algorithm sets the level of activity of a function as the averaged values of all pathways that contribute to said function (Annex 3, Script 1). This approach seems to solve the problem because lower values now decrease the overall final activity instead of adding to it. However, this is a compromise because in a cell not all pathways are equally important in the triggering of a function, some might be more relevant than others.

A way to try to model this, that has not been assessed here, would be to set the value of the most active pathway as the value for the function. The logic behind this is that it can be assumed that a pathway that leads to a higher function value will have either highly activated effectors or more effectors contributing to the function than the other pathways, so that the pathway could be considered as the most important for the function.

4.3. Drug Compilation and Categorization

For the drug analysis, a repository of drugs with the information is formatted to contain the information the program needs. The original Drug List, was a *DrugList.RData* file of 23.9Mb containing 8054 drug candidates downloaded from DrugBank (Unpublished data from study, Untitled Work). In this list, many drugs lacked targets or pharmacological effect. Many others didn't, but the label provided by DrugBank for their pharmacological action was not informative of their effects on targets and for some others it was unknown. For this reason, Drugs were annotated (Annex 2) based on the descriptions from DrugBank (Wishart, 2006) and other literature and classified as activators on inhibitors, or were removed.

The resulting 4316 drug candidates were organized into a data frame with their name, target and effect (Annex 3, Script 2), providing the program with quite an extensive repository.

4.4. Programming Drug Modifications on Experimental Data

The developed program uses the initial experimental design and expression values and expands them with the new conditions modelled. These conditions are the drugs to be assayed.

The function *drug.exp* is designed to modify expression values taking as input the tumours' expression values and a drug (Annex 3, Script 3). Then, it looks for the drug targets in the compiled drug data matrix and for its effects on those targets. The expression values for those targets in the Tumour condition field of the expression values matrix are modified to 0.99 if the drug is an activator or multiplied by 0.001 if it is an inhibitor. The function returns the modified data for the drug.

The function *drug.des* is designed to modify the experimental design taking as input the tumour conditions from the experimental design and a drug (Annex 3, Script3). It appends the drug code to the samples code and sets it as a new condition.

To run the program, first the experimental design and expression values are loaded together with the drug data and the file containing the drugs to be assessed. The file named *drugs_to_test.txt* contains the name of one drug per line. Since the genes are written with a unique code, the named targets in the drug matrix need to be changed to the same code, so that the program can recognize them. All those targets that are not present in the experimental design are removed and then each drug is given a unique code. For each drug to be assayed, the original tumour matrix and design are given to the *drug.exp* and *drug.des* functions, respectively, and the returned matrixes containing the modified information are appended to the original matrixes (Annex 3, Script 4).

The process is exemplified in Figure 4, which depicts how the *DrugA* activates its target *Gene1*, whose code is *G1*, and therefore modifies the expression value for the tumour condition to 0.99, to finally add it to the original data frame, where it goes by its code name *D1*. In the experimental design data frame, the drug will appear as the new *D1_Tumour_1*, describing it has modified the values for Tumour_1 as the new D1 condition.

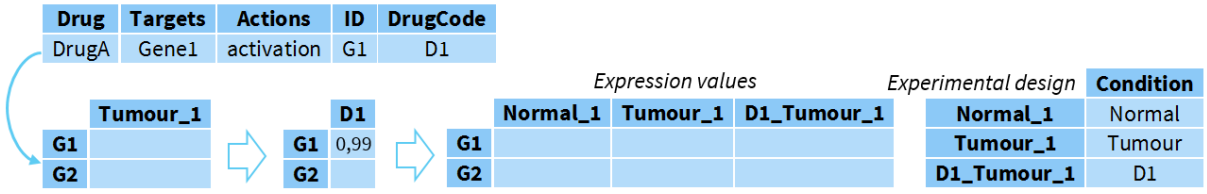

FIGURE 4. Result of applying an activating drug.

4.5. Statistical Analysis and Program Output

After expanding and modifying the experimental data with the drugs to be assayed, the signaling pathways values are calculated and then transformed into cellular function values (Annex 3, Script 4). For analysing the results, the *wilcox.function* is developed to apply the wilcox test to the conditions assayed and the *drug.type.per.function* is developed to classify the drugs into four descriptive categories: Optimum, Underdose, Overdose, Undesired (Annex3, Script 5).

In the first step, the Normal and Tumour conditions are compared and a small matrix called *NT* is created, which contains the p-values resulting from the wilcox test per function, and a column describing its significance. When the p-value is smaller than the cut-off value of 0.05, then the distributions are considered significantly different and *Y* will appear in the significance column, meaning *Yes*. Otherwise, it will display *N* for *No* (See Table 5).

TABLE 5. Result of a wilcox test for the Tumour – Normal comparison of three functions and significance.

	pvalueNT	SignificanceNT
Function A	0.002	Y
Function B	0.3	N
Function C	0.001	Y

Then, the rest of Conditions (drugs) are analyzed. The same process done for the Normal and Tumour conditions is repeated for both the Tumour and Drug comparison and the Normal and Drug comparison. For each drug, a *Resume* matrix is then created containing the p-values and significance for all functions and comparisons (See Table 6).

TABLE 6. Result of a wilcox test for all comparisons on three functions and significance.

	Pvalue NT	Significance NT	Pvalue DT	Significance DT	Pvalue DN	Significance DN
FunctionA	0.002	Y	0.001	Y	0.08	N
FunctionB	0.3	N	0.4	N	0.07	N
FunctionC	0.001	Y	0.003	Y	0.01	Y

From this table the significantly altered functions can be detected. All those functions which are significantly different between Normal and Tumour, but also between Tumour and Drug, are considered altered by the Drug. A *Significant* table is then written containing only those functions. Using the example on Table 6, the *Significant* table would contain Function A and Function C. With this information now the effect of the drug can be classified using the *drug.type.per.function*, which uses the level of significance and the means for all distributions. The results are then written in the *DrugType* table:

- Optimum: the distributions for Drug and Normal conditions are not significantly different. This means the drug has reversed the Tumour condition into a Normal

condition. For instance, from Table 6 it can be deduced the drug has an Optimum effect on Function A.

- Underdose: the distributions for Drug and Normal conditions are significantly different but the Drug distribution has shifted from Tumour towards Normal. The impact of the Drug is positive, however it is not powerful enough. It could be argued an optimum effect could be attained by changing the dosage.

- Overdose: the distributions for Drug and Normal conditions are significantly different but the Drug distribution has shifted from the Tumour towards Normal and beyond. The impact of the Drug is positive, however it is too powerful. It could be argued an optimum effect could be attained by changing the dosage.

- Undesired: the distributions for Drug and Normal conditions are significantly different but the Drug distribution has shifted from Normal towards Tumour and beyond. This means the drug has caused a function to be more tumorous than it is., therefore the drug should be discarded as a repurposing candidate.

The descriptive tables for a drug are written in a new directory created for each drug in a Comma-Separated Values (CSV) format (readable by Excel). Finally, the *All_Drugs_Resume* table is written, resuming the information for all functions and drugs, also in a CSV format.

To see the Normal and Tumour distributions, plotly is used to plot a boxplot for each function using the Script 5 on Annex 3. Also, to further visualize the effect of a Drug, its functions can be plotted and, since plotly allows for modification of the transparency of each box, overlaps and shifts in the distribution can be observed for all functions at once.

4.6. Validating the Method

This method has been developed to be used in the context of cancer. To validate it, expression values from KIRC and BRCA cancer patients were used. The idea was to perform an assay with most of the drugs contained in the repository created. However, for such an assay a powerful server is required and due to some constrains at the laboratory this was not possible.

Instead, the drug Goserelin was applied to KIRC's data. Goserelin is a synthetic hormone that stops the production of the hormone testosterone in men, which may stimulate the growth of cancer cells. In women, Goserelin decreases the production of the hormone estradiol (which may stimulate the growth of cancer cells) to levels similar to a postmenopausal state. When the medication is stopped, hormone levels return to normal (Wishart, 2006). This drug is therefore used to treat prostate cancer in men and breast cancer in women. Using this drug on BRCA data should provide a good validation of the method, and if used in KIRC with good results, it would also become good candidate for repurposing.

However, when the method was used with BRCA data, many functions showed too much variation. In Figure 5, it can be seen how the box plots for some functions are too wide, specially in the Tumour condition (in blue) where some values have a range of approximately 0.2. There is too much variability.

The reason for this might be these data needed to be classified first into three subtypes of BRCA . It has been seen that there are three main types of BRCA Cancers (Tonin et al., 1998), which therefore have differential expression profiles. When assayed together, they contribute to the high variability observed. Having so much variability when trying to model the effect of Goserelin, limited the fiability of the results. Also, barely two functions were outputted in the *Resume* file.

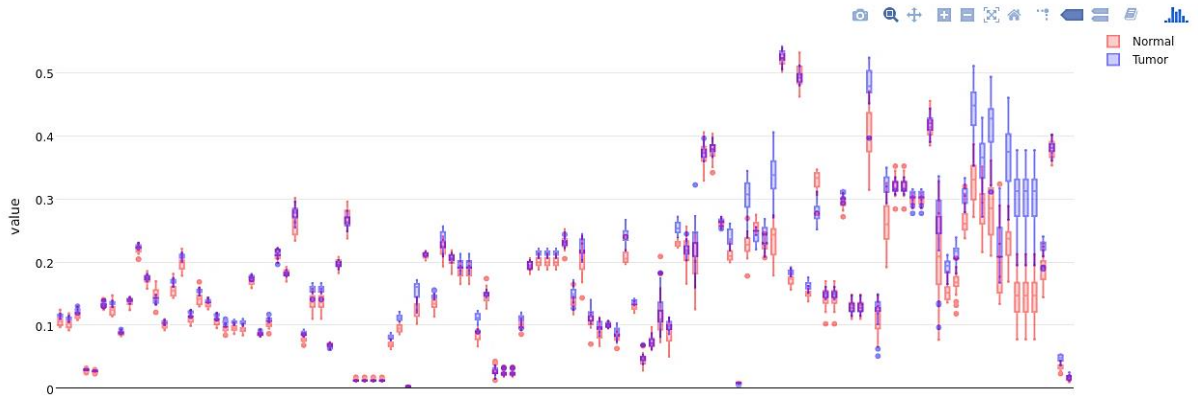

FIGURE 5. BRCA Boxplot for Normal and Tumour conditions. Red: Normal. Blue: Tumor.

However, when the method was used with KIRC data, the picture changed. The variability in these data is not so dramatic. The method worked beautifully. The new algorithm to estimate the values of the functions seemed to have solved the saturation of the signal, which can also be seen in the Figure 5 of BRCA. Each function has a different value of activity, and they do not fall too high on the 0 – 1 scale used to evaluate them (Figure 6. Also, drug modifications on the original data were correctly introduced, proving the Drug Repository is well designed. Together with the functional annotation, the modelling of the actions of Goserelin seemed to make sense. The *DrugType* resulting table can be seen on Table 7.

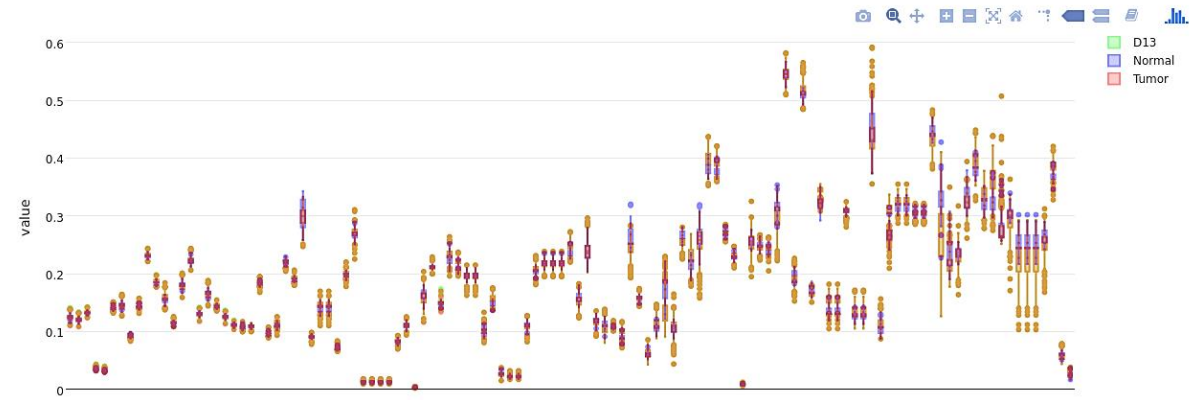

FIGURE 6. KIRC and Goserelin Boxplot for Normal and Tumour conditions. Blue: Normal. Red: Tumour. Green: Goserelin (D13).

TABLE 7. Goserelin classification of drug effect in KIRC.

	pvalue_DN	SignificantDN	Classification
Phagocytosis	0,98	N	Optimum
Cancer Growth Repressor	2,75E-22	Y	Undesired
Transcription	4,14E-22	Y	Undesired
Antiapoptosis	1,17E+02	Y	Underdose
Cell Survival	3,78E-05	Y	Underdose
Cell Cycle Progression	6,76E-13	Y	Underdose
Cancer Drug Resistance	0,84	N	Optimum
Apoptotic Stress Response	0,88	N	Optimum
Inflammatory Response	8,16E-06	Y	Underdose
Cancer Transcriptional Misregulation	1,64E-07	Y	Underdose
Cell Polarity	5,00E-01	Y	Underdose
Cell Growth	5,60E+04	Y	Underdose
Stem Cell Self Renewal	5,60E+04	Y	Underdose
Stress Response	6,55E-13	Y	Underdose
Quiescense	5,96E-15	Y	Underdose
Cell Migration	3,08E-09	Y	Underdose
Antiinflammatory	1,11E+04	Y	Underdose
Metabolism	1,70E-07	Y	Underdose
Cytoskeletal Organization	1,64E-04	Y	Underdose
Cancer Survival	3,53E+02	Y	Underdose
Proliferation Arrest	9,08E-06	Y	Underdose
Chloride Ion Secretion	4,10E-05	Y	Underdose
Lipid Biosynthesis	4,97E-06	Y	Underdose
DNA Biosynthesis	1,63E-12	Y	Underdose
Inhibition of mTOR pathway	9,82E-12	Y	Underdose
Growth inhibition	9,82E-12	Y	Underdose
Cancer Invasion Inhibition	6,98E-12	Y	Underdose
Cell Differentiation	4,08E-09	Y	Underdose
Proliferation	4,08E-09	Y	Underdose
Lipid Metabolism	2,72E+04	Y	Underdose
Adipocyte Differentiation	1,70E+02	Y	Underdose
Glucose Metabolism	1,81E-05	Y	Underdose
Ubiquitination	6,17E-09	Y	Underdose

One of the great things about plotly is that it is an interactive plot that allows for zooming in and out of the plot to focus on the desired region and pops the name of the cellular function when placing the mouse on top of a box. When looking at the plotted data, the cellular functions that had shifted from the tumour are consistent with the description performed by the method in Table 7 (See Figure 7). Although the Cancer Growth Repressor and Transcription Functions are described as Undesired in Table 7, most of the functions

named in Figure 7 do correspond to the original effect the drug has on its prescribed cancers. It can be observed how the Goserelin data has shifted towards the Normal data for the named functions. One of the limitations of the method is that when a cellular function requires the presence of many proteins, the presence of only one of them is enough to trigger the function. This is the case of Transcription, for example, because the presence of one activating factor that works in combination with others is enough to consider the function is taking place. This limitation might explain the classification of Transcription as undesired.

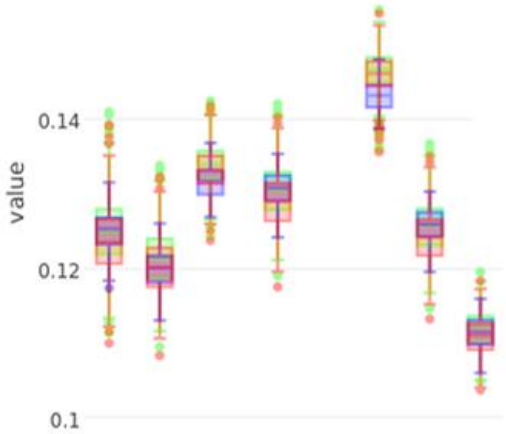

FIGURE 7. Relevant altered functions by Goserelin. Pink: Tumour condition. Blue: Normal condition. Green: Goserelin. The functions from left to right: Phagocytosis, Platelet Activation, Immune Response, Cancer Drug Resistance, Proliferation Arrest, Apoptotic Stress Response, Cancer Proliferation.

Even if these results seem promising and the developed method works as designed, the method needs to be further validated. To do so, the complete assay should be performed with the data from BRCA cancers and including more drugs.

This method could also be more fine-tuned by trying a different algorithm for the calculation of the cellular functions. A comparison between the current algorithm and one that would set the maximum value among pathways as the value for the function should be performed. One notable drawback in terms of time and effort is that the functional annotation is manual, although terms from Gene Ontology (GO) could also be used. The drug repository developed here is quite big, but any user could also expand it.

Further validation would also come from uploading the developed methodology into a GitHub repository, freely available for the public, to be employed in a different context, in a different disease.

Nevertheless, the potential applications of this method are many. Since it is a method that is rooted in the mechanisms of the biological processes happening in cells, it is a more reliable method. Its applications include: modelling drug repositioning for any disease, it allows for the possibility of personalized medicine drug repositioning and evaluation of a therapy, and drug synergies could also be found. Drug repositioning has been proven to be most applicable to cancer and personalized medicine with great success, but current drug

combination strategies rely on clinical and empirical evidence solely, so there is a high demand for computational methods like the one here developed (Li et al., 2016).

The advantages of this method in personalized medicine are many, including it can be easily used by any user, who has minimum knowledge of bioinformatics, to assay many drugs in a home computer. To perform drug repositioning analysis of many drugs on many patients, like it was the original purpose for validating this method, a server is required. Usually such an analysis is performed at centres where such servers are available, so this does not pose a problem.

This model here developed allows for integration of many different types of data (gene expression, proteomics data, mutations, etc) and employs signaling pathways as mechanistic biomarkers to perform any analysis desired (for any disease, big scale or personalized medicine). When compared to the previous computational models for drug repositioning, this is computationally the best one, because each variable does not need to be scored and analyzed, all the information is integrated in the cellular functional outcome, simulating the dynamics of a real cell. It is a reliable and versatile method.

5. Conclusions

One of the big problems faced nowadays is the enormous time and costs for the development of new drugs. Here is presented a new method for the integration of data from any *Omic* into activation values for signalling pathway. It is a method based on mechanistic biomarkers, it is the signalling pathway the one responsible for an effect in the cell, the phenotype. Therefore, it is a more informative method with higher predictive value, but also computationally more effective. All the objectives regarding the development of the method have been met successfully. The method, together with a repository of drug annotations containing more than 4000 drugs, could be freely available online. But before that, further validation of the method is required.

The method is presented in the context of cancer. All the effector proteins in pathways relevant in cancer are here annotated, with relevant functions for the disease. Cancer patients frequently respond poorly for therapy and are in more need of drugs that are taking too long to be developed. Instead, repositioning methods can offer them with drugs in one or two years from their discovery.

Not only is this method reliable, it is also versatile in terms of its applications. It can be used to study any disease, for drug repositioning approaches or to look for synergies. And it can be implemented at any scale: both for big studies and for personalized medicine.

6. Bibliography

Ashburn, T. T., & Thor, K. B. (2004). Drug repositioning: identifying and developing new uses for existing drugs. *Nature Reviews Drug Discovery*, *3*(8), 673–683. https://doi.org/10.1038/nrd1468

Booth, B., & Zemmel, R. (2004). Opinion/Outlook: Prospects for productivity. *Nature Reviews Drug Discovery*, *3*(5), 451–456. https://doi.org/10.1038/nrd1384

Brubaker, D., Difeo, A., Chen, Y., Pearl, T., Zhai, K., Bebek, G., … Barnholtz-Sloan, J. (2014). Drug Intervention Response Predictions with PARADIGM (DIRPP) identifies drug resistant cancer cell lines and pathway mechanisms of resistance. *Pacific Symposium on Biocomputing. Pacific Symposium on Biocomputing*, 125–135.

Fang, B., Mehran, R. J., Heymach, J. V., & Swisher, S. G. (2015). Predictive biomarkers in precision medicine and drug development against lung cancer. *Chinese Journal of Cancer*, *34*(3). https://doi.org/10.1186/s40880-015-0028-4

Fey, D., Halasz, M., Dreidax, D., Kennedy, S. P., Hastings, J. F., Rauch, N., … Croucher, D. R. (2015). Signaling pathway models as biomarkers: Patient-specific simulations of JNK activity predict the survival of neuroblastoma patients. *Science Signaling*, *8*(408), ra130-ra130. https://doi.org/10.1126/scisignal.aab0990

Fu, C., Jin, G., Gao, J., Zhu, R., Ballesteros-villagrana, E., & Wong, S. T. C. (2013). DrugMap Central: an on-line query and visualization tool to facilitate drug repositioning studies. *Bioinformatics*, *29*(14), 1834–1836. https://doi.org/10.1093/bioinformatics/btt279

Gottlieb, A., Stein, G. Y., Ruppin, E., & Sharan, R. (2014). PREDICT: a method for inferring novel drug indications with application to personalized medicine. *Molecular Systems Biology*, *7*(1), 496–496. https://doi.org/10.1038/msb.2011.26

Hidalgo, M. R., Cubuk, C., Amadoz, A., Salavert, F., Carbonell-Caballero, J., & Dopazo, J. (2017). High throughput estimation of functional cell activities reveals disease mechanisms and predicts relevant clinical outcomes. *Oncotarget*. https://doi.org/10.18632/oncotarget.14107

hiPathia. (n.d.). R. Retrieved from www.github.com/babelomics/hipathia

Hughes, J., Rees, S., Kalindjian, S., & Philpott, K. (2011). Principles of early drug discovery: Principles of early drug discovery. *British Journal of Pharmacology*, *162*(6), 1239–1249. https://doi.org/10.1111/j.1476-5381.2010.01127.x

igraph. (n.d.). Retrieved from https://github.com/igraph/

Johnson, W. E., Li, C., & Rabinovic, A. (2007). Adjusting batch effects in microarray expression data using empirical Bayes methods. *Biostatistics*, *8*(1), 118–127. https://doi.org/10.1093/biostatistics/kxj037

Kanehisa, M., Sato, Y., Kawashima, M., Furumichi, M., & Tanabe, M. (2016). KEGG as a reference resource for gene and protein annotation. *Nucleic Acids Research*, *44*(D1), D457–D462. https://doi.org/10.1093/nar/gkv1070

Kitano, H. (Ed.). (2001). *Foundations of systems biology*. Cambridge, Mass: MIT Press.

Kruschke, J. K. (2011). *Doing bayesian data analysis: a tutorial with R and BUGS*. Burlington, MA: Academic Press.

Li, J., Zheng, S., Chen, B., Butte, A. J., Swamidass, S. J., & Lu, Z. (2016). A survey of current trends in computational drug repositioning. *Briefings in Bioinformatics*, *17*(1), 2–12. https://doi.org/10.1093/bib/bbv020

Li, Y. Y., & Jones, S. J. (2012). Drug repositioning for personalized medicine. *Genome Medicine*, *4*(3), 27. https://doi.org/10.1186/gm326

PathAct. (n.d.). Retrieved from http://pathact.babelomics.org

*R Core Team (2013). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org/*Rastegar-Mojarad, M.,

Ye, Z., Kolesar, J. M., Hebbring, S. J., & Lin, S. M. (2015). Opportunities for drug repositioning from phenome-wide association studies. *Nature Biotechnology*, *33*(4), 342–345. https://doi.org/10.1038/nbt.3183

Robinson, W. H., Lindstrom, T. M., Cheung, R. K., & Sokolove, J. (2013). Mechanistic biomarkers for clinical decision making in rheumatic diseases. *Nature Reviews Rheumatology*, *9*(5), 267–276. https://doi.org/10.1038/nrrheum.2013.14

Robinson, W. H., & Mao, R. (2016). Biomarkers to guide clinical therapeutics in rheumatology?: *Current Opinion in Rheumatology*, *28*(2), 168–175. https://doi.org/10.1097/BOR.0000000000000250

Salavert, F., Hidago, M. R., Amadoz, A., ?ubuk, C., Medina, I., Crespo, D., … Dopazo, J. (2016). Actionable pathways: interactive discovery of therapeutic targets using signaling pathway models. *Nucleic Acids Research*, *44*(W1), W212–W216. https://doi.org/10.1093/nar/gkw369

Serra-Musach, J., Mateo, F., Capdevila-Busquets, E., de Garibay, G. R., Zhang, X., Guha, R., … Pujana, M. ?ngel. (2016). Cancer network activity associated with therapeutic response and synergism. *Genome Medicine*, *8*(1). https://doi.org/10.1186/s13073-016-0340-x

Tari, L. B., & Patel, J. H. (2014). Systematic Drug Repurposing Through Text Mining. In V. D. Kumar & H. J. Tipney (Eds.), *Biomedical Literature Mining* (Vol. 1159, pp. 253–267). New York, NY: Springer New York. https://doi.org/10.1007/978-1-4939-0709-0_14

Tonin, P. N., Mes-Masson, A.-M., Futreal, P. A., Morgan, K., Mahon, M., Foulkes, W. D., … Narod, S. A. (1998). Founder BRCA1 and BRCA2 Mutations in French Canadian Breast and Ovarian Cancer Families. *The American Journal of Human Genetics*, *63*(5), 1341–1351. https://doi.org/10.1086/302099

Waldman, S. A., & Terzic, A. (2013). Systems-Based Discovery Advances Drug Development. *Clinical Pharmacology & Therapeutics*, *93*(4), 285–287. https://doi.org/10.1038/clpt.2013.21

Wishart, D. S. (2006). DrugBank: a comprehensive resource for in silico drug discovery and exploration. *Nucleic Acids Research*, *34*(90001), D668–D672. https://doi.org/10.1093/nar/gkj067

Zhang, P., Wang, F., & Hu, J. (2014). Towards drug repositioning: a unified computational framework for integrating multiple aspects of drug similarity and disease similarity. *AMIA ... Annual Symposium Proceedings. AMIA Symposium*, *2014*, 1258–1267.

Pathway Activity Analysis (PAA) as a new class of mechanistic biomarker to predict drug responses in drug repositioning for cancer patients

---

ANNEX 1. Signalling Pathways

Functional Annotation

Index

Signalling Pathways Functional Annotation

TABLE 1. Functional annotation of the final effectors of the MAPK Signalling Pathway.

Condition	Protein	Function	Key Word
JNK and p38	NFKB1	NF-kB acts through the transcription of anti-apoptotic proteins, leading to increased proliferation of cells and tumour growth (Escárcega, Fuentes-Alexandro, García-Carrasco, Gatica, & Zamora, 2007).	Antiapoptosis
		Misregulated transcription in cancer (Escárcega et al., 2007).	Cancer Transcriptional Misregulation
		Regulates osteoclast formation, function, and survival and is essential for osteoclast precursors to differentiate into TRAP+ osteoclasts (Soysa & Alles, 2009)	Differentiation
		Plays a central role in inflammation through its ability to induce transcription of proinflammatory genes (Tak & Firestein, 2001).	Inflammatory Response
Classical and JNK	MAPT	Tubulin assembly and microtubule stabilization in the nervous system (Cleveland, Hwo, & Kirschner, 1977).	Microtubule Stabilization
	STMN1	Microtubule destabilization and transition from microtubule growth to shortening (Cassimeris, 2002).	Microtubule Destabilization
		Plays an inhibitory role in classically activated macrophages and its down-regulation is required for the phenotypic changes and activation of macrophages (Xu & Harrison, 2015)	Immunity Regulation
		Merkel cell polyomavirus (MCPyV) drives Merkel Cell Cancer (MCC), and its highly metastatic nature is due to the increased expression and microtubule destabilization of stathmin (Whitehouse & Macdonald, 2015).	Viral Carcinogenesis
	PLA2G4A	Selectively hydrolyzes arachidonyl phospholipids in the sn-2 position releasing arachidonic acid. Together with its lysophospholipid activity, it is implicated in the initiation of the inflammatory response (The UniProt Consortium, 2017).	Inflammatory Response

(Continued)

Condition	Protein	Function	Key Word
Classical and JNK	PLA2G4A	Requirement for eicosanoid synthesis and subsequent platelet activation (Kirkby et al., 2015)	Platelet Activation
	ATF4	In concert with DDIT3/CHOP, activates the transcription of TRIB3 and promotes ER stress-induced neuronal apoptosis by regulating the transcriptional induction of BBC3/PUMA (The UniProt Consortium, 2017).	Apoptotic Stress Response
		Role in multidrug resistance through glutathione-dependent redox system (Igarashi et al., 2007)	Cancer Drug Resistance
		Biomarker for Esophageal Squamous Cell Carcinoma (ESCC) prognosis, its dysregulation correlates with cell invasion and metastasis (Zhu et al., 2014)	Cancer Invasion
		Overexpressed in solid tumours, its inhibition reduced proliferation (Ye et al., 2010)	Proliferation
	FOS	Osteoblast (Grigoriadis et al., 1994) and adipocyte differentiation (Luther et al., 2014)	Differentiation
		Increased expression in response to growth factors, leading to proliferation via the Activating Protein-1 (AP1) complex (Angel & Karin, 1991).	Proliferation
	MYC	Dependent and independently of TP53, through death receptor pathways at multiple junctions and amplifies apoptotic signalling at the mitochondria (Hoffman & Liebermann, 2008).	Apoptosis
		Overexpression correlates with multiple myeloma (A. G. Szabo et al., 2016).	Cancer Transcriptional Misregulation
		Induces positive cell cycle regulators required for initiating replication, binds replication origins, antagonizes cell cycle inhibitors p21 and p27 (Bretones, Delgado, & León, 2015).	Cell Cycle Progression

(Continued)

Condition	Protein	Function	Key Word
Classical and JNK	MYC	Activates the transcription of growth-related genes (Dang, 1999).	Cell Growth
		Differentiation of epidermal stem cells (Gandarillas & Watt, 1997). It also controls the balance between Hematopoietic Stem Cell (HSC) renewal and differentiation (Wilson, 2004).	Differentiation
		HSCs are activated to self-renew and to differentiate at the interface between the niche and non-niche microenvironments (Murphy, Wilson, & Trumpp, 2005).	Stem Cell Self Renewal
Classical activates JNK and p38 inhibit	NFATC1	Participates in the cardiovascular system development (Horsley & Pavlath, 2002) and osteoclast formation (Teitelbaum, 2007).	Differentiation
		Transcription of cytokine genes and other genes involved in the immune response (Rao, Luo, & Hogan, 1997).	Immune Response
	NFATC3	Regulation of gene expression in T cells and thymocytes, specially cytokine IL-2 (The UniProt Consortium, 2017).	Immune Response
JNK and p38	JUN	DNAzymes targeting c-Jun act as inhibitors of angiogenesis (Folkman, 2004). Activated c-Jun is predominantly expressed at the invasive front in breast cancer and is associated with proliferation and angiogenesis (Vleugel, Greijer, Bos, Wall, & Diest, 2006).	Angiogenesis
		Targets the tumour suppresor TP53, which has an increased expression during cell division and in response to growth factors (Shaulian & Karin, 2001).	Proliferation
		G1 progression through repression of tumour suppressor genes and induction of CCND1 transcription (Shaulian & Karin, 2001).	Cell Cycle Progression
		Expression is altered early during lung and liver carcinogenesis (E. Szabo, Riffe, Steinberg, Birrer, & Linnoila, 1996).	Cancer Transcriptional Misregulation
		NFAT/Fos/Jun is a critical osteoclastogenic complex, and deletion of any of the three arrests osteoclast formation (Teitelbaum, 2007).	Differentiation

(Continued)

Condition	Protein	Function	Key Word
JNK and p38	JUND	Protects cells from TP53 dependent apoptosis (Ameyar, Wisniewska, & Weitzman, 2003).	Antiapoptosis
		Negative regulator of cell growth by maintaining the cells in a quiescent state ("The mammalian Jun proteins," 2001).	Quiescence
	ATF2	In response to stress, ATF-2, a member of the ATF/cAMP response element-binding protein family, is phosphorylated by p38/Jun NH2-terminal protein kinase and activates the transcription of apoptosis-related genes (Makino, Sano, Shinagawa, Millar, & Ishii, 2006)	Apoptotic Stress Response
		The phosphorylated form (mediated by ATM) plays a role in the DNA damage response and is involved in the ionizing radiation (IR)-induced S phase checkpoint control and in the recruitment of the MRN complex into the IR-induced foci (IRIF) (The UniProt Consortium, 2017).	DNA Damage Response
		HBZ activates transcription of ATF2 pro-survival genes (Mesri, Feitelson, & Munger, 2014).	Viral Carcinogenesis
	ELK1	One of the mechanisms by which BRCA1a/1b proteins function as growth/tumour suppressors is through inhibition of the expression of Elk-1 target genes like FOS (Chai et al., 2001).	Growth Repressor
		Transcription factor that binds to purine-rich DNA sequences. Forms a ternary complex with SRF and the ETS and SRF motifs of the serum response element (SRE) on the promoter region of immediate early genes such as FOS and IER2 (The UniProt Consortium, 2017).	Transcription
	TTP53	(See TP53 Signalling Pathway)	Apoptosis
			Cell Cycle Arrest
	ELK4	Forms a ternary complex with the serum response factor (SRF). Interaction with SIRT7 leads to recruitment its stabilization at promoters, followed by deacetylation of histone H3 at Lys-18 and subsequent transcription repression (The UniProt Consortium, 2017).	Transcription Regulation

(Continued)

Condition	Protein	Function	Key Word
JNK and p38	DDIT3	Inducible inhibitor of adipocytic differentiation in response to metabolic stress (Batchvarova, Wang, & Ron, 1995).	Differentiation Inhibition
	DDIT3	Intrinsic pathway. Response to toxic and metabolic insults that perturb function of the endoplasmic reticulum (ER stress) (Zinszner et al., 1998).	Apoptotic Stress Response
		Nuclear DDIT3 causes cell cycle arrest at the G1/S (Jauhiainen et al., 2012).	Cell Cycle Arrest
		Cytoplasmic DDIT3 inhibits cell migration (Jauhiainen et al., 2012).	Cell Migration
		Enhances differentiation in erythroid cells (Cui, Coutts, Stahl, & Sytkowski, 2000).	Erythropoiesis
		Induction of CASP4/CASP11 which induces the activation of CASP1 and both of these increase the activation of pro-IL1B to mature IL1B (The UniProt Consortium, 2017).	Immune Response
	MAX	Transcription regulator. Forms a sequence-specific DNA-binding protein complex with MYC or MAD which recognizes the core sequence 5-CAC[GA]TG-3. The MYC-MAX complex is a transcriptional activator, whereas the MAD:MAX complex is a repressor (The UniProt Consortium, 2017).	Transcription Regulation
		(Kanehisa, Sato, Kawashima, Furumichi, & Tanabe, 2016)	Differentiation Proliferation
	MEF2C	Heart development, skeletal muscle differentiation, dendrite morphogenesis, control of vascular integrity, T-cell development, neuronal differentiation and survival (Potthoff & Olson, 2007).	Differentiation
		Required for B-cell survival and proliferation in response to BCR stimulation (Wilker et al., 2008).	Immune Response
	HSPB1	Inhibits apoptotic and necrotic pathways under stress (Takayama, Reed, & Homma, 2003).	Antiapoptotic Stress Response

(Continued)

Condition	Protein	Function	Key Word
JNK and p38	ATF4	(See Classical Route)	Apoptotic Stress Response
			Cancer Drug Resistance
			Cancer Invasion
	ATF4	(See Classical Route)	Proliferation
	NLK	Kinase that regulates a number of transcription factors (The UniProt Consortium, 2017).	Transcription Regulation
MAPK7 pathway	NR4A1	Via migration to the mitochondrial outer membrane, converts anti-apoptotic BCL2 into a pro-apoptotic protein (Pawlak, Strzadala, & Kalas, 2015).	Apoptosis

TABLE 2. Functional annotation of the final effectors of the TP53 Signalling Pathway.

Condition	Protein	Function	Key Word
Stress Signalling	CCND1	Forms a complex with and functions as regulatory subunit of CDK4 and CDK6, whose activity is required for cell cycle G1/S transition (Hydbring, Malumbres, & Sicinski, 2016).	Cell Cycle Progression
	CCNE1	Essential master regulator of the G1/S transition which also cooperates with Cdc6 to allow prereplication complexes to form during the G0/S transition (Hwang & Clurman, 2005).	Cell Cycle Progression
	CCNB1	G2/M transition (Hydbring et al., 2016).	Cell Cycle Progression
	CASP3	Activated by the extrinsic and intrinsic pathways, cleaves and activates caspases 6, 7 and 9, and it is processed itself is by caspases 8, 9 and 10 (Salvesen, 2002).	Apoptosis

(Continued)

Condition	Protein	Function	Key Word
Stress Signalling	IGF1	Regulates BCL2 family proteins, inhibitors of caspases and Signalling of death-inducing receptors inhibiting apoptosis in many cell types and in the presence of different apoptogenic stimuli (Kooijman, 2006).	Antiapoptosis
		Involved both in prenatal and postnatal development, enhances proliferation (Kemp, 2009).	Development
		Important growth hormone, endocrine when secreted by the liver or paracrine in cartilagenous cells important in protein anabolism (Laron, 2001).	Cell Growth
		Local repair mechanisms: promotion of cell recruitment to the injured muscle and the subsequent resolution of the inflammatory response (Mourkioti & Rosenthal, 2005).	Inflammatory Response
		Enhances proliferation and survival of mesenchymal stem cells before differentiation to neural progenitor-like cells (Huat et al., 2014).	Proliferation
	SERPINB5	The tumour suppressor activity of SERPINB5 may depend in large part on its ability to inhibit angiogenesis (Zhang, Volpert, Shi, & Bouck, 2000).	Antiangiogenesis
		Blocks the growth, invasion, and metastatic properties of mammary tumours (Streuli, 2002).	Cancer Invasion Inhibition
	SESN1	Antioxidant defense in response to oxidative stress (The UniProt Consortium, 2017).	Antioxidant
	IGFBP3	Induces apoptosis and mediates the effects of TGFb1 on programmed cell death through TP53 and IGF mechanisms (Rajah, 1997).	Apoptosis
		Inhibits proliferation of neural progenitor cells (Kalluri & Dempsey, 2011).	Antiproliferation
		Modulates the early stages of keratinocyte differentiation (Edmondson et al., 2005).	Differentiation

(Continued)

Condition	Protein	Function	Key Word
Stress Signalling	IGFBP3	Prolongs IGF1 and IGF2 half-life in circulation and regulates the available amount for interaction with their receptors (Cerri, Gonzales, Ballard, & Cohen, 1999).	Growth Inhibition
	STEAP3	Enhances susceptibility to apoptosis cooperating with Nix (Passer et al., 2003).	Apoptosis
		Augments MYT1 activity, a negative regulator of G2/M transition (Passer et al., 2003).	Cell Cycle Inhibition
		Transferrin uptake in erythroid cells (Sendamarai, Ohgami, Fleming, & Lawrence, 2008).	Ion Transporter
	TP73	In response to DNA damage (Allocati, Di Ilio, & De Laurenzi, 2012).	Apoptosis

TABLE 3. Functional annotation of the final effectors of the Wnt Signalling Pathway.

Condition	Protein	Function	Key Word
Canonical	MYC	(See MAPK Signalling Pathway)	Apoptosis
			Cancer Transcriptional Misregulation
			Cell Cycle Progression
			Cell Growth
			Differentiation
			Stem Cell Renewal
	JUN	(See MAPK Signalling Pathway)	Angiogenesis
			Proliferation
			Cell Cycle Progression
			Cancer Transcriptional Misregulation

(Continued)

Condition	Protein	Function	Key Word
Canonical	JUN	(See MAPK Signalling Pathway)	Differentiation
	FOSL1	High Fra-1 expression is associated with a more malignant cell phenotype (Belguise, Kersual, Galtier, & Chalbos, 2004).	Cancer Invasion Cancer Proliferation
		Negatively regulates LPS-induced responses in macrophages and inhibits fracture-induced ossification through suppression of inflammation-induced chondrogenesis (Morishita et al., 2009).	Antiinflammatory
		Dimerizes with Jun family proteins forming the transcription factor AP-1 (Shaulian & Karin, 2002).	Proliferation
	CCND1	(See TP53 Signalling pathway)	Cell Cycle Progression
	PPARD	Pronounced anti-inflamatory effects (Kilgore & Billin, 2008).	Antiinflammatory
		Redirects fatty acids from adipose tissue to skeletal muscle and increases its oxidative capacity, genetic variations determine change in aerobic physical fitness and insulin resistance (Stefan et al., 2007).	Lipid Metabolism
	MMP7	Breakdown of extracellular matrix in development, reproduction and tissue remodeling. Degrades casein, gelatins of types I, III, IV, and V, and fibronectin. Activates procollagenase. (The UniProt Consortium, 2017).	Extracellular Matrix Degradation
	ROCK2	Acts as a negative regulator of VEGF-induced angiogenic endothelial cell activation (The UniProt Consortium, 2017).	Antiangiogenesis
		Involved in regulation of smooth muscle contraction, actin cytoskeleton organization, focal adhesion formation, neurite retraction and motility via phosphorylation of ADD1, BRCA2, CNN1, EZR, DPYSL2, EP300, MSN, MYL9/MLC2, NPM1, RDX, PPP1R12A and VIM. The UniProt Consortium, 2017).	Cell Motility Regulation Cell Polarity

(Continued)

Condition	Protein	Function	Key Word
Canonical	ROCK2	Plays an important role in the timely initiation of centrosome duplication (The UniProt Consortium, 2017).	Centrosome Amplification
Wnt/Ca2+	NFATC1	(See MAPK Signalling Pathway)	Differentiation
			Immune Response
	CAMK2A	Plasticity at glutamatergic synapses, hippocampal long-term potentiation (LTP) and spatial learning (The UniProt Consortium, 2017).	Learning
	PRKCA	Required for full endothelial cell migration, adhesion to vitronectin (VTN) and VEGFA-dependent regulation of kinase activation and vascular tube formation. Involved in the stabilization of VEGFA mRNA level and mediates VEGFA-induced cell proliferation (The UniProt Consortium, 2017).	Angiogenesis
		Phosphorylates BCL2, required for its antiapoptotic activity (Ruvolo, Deng, Carr, & May, 1998).	Antiapoptosis
		Translocates from focal contacts to lamellipodia and participates in the modulation of desmosomal adhesion. Plays a role in cell motility by phosphorylating CSPG4, which induces association of CSPG4 with extensive lamellipodia at the cell periphery and cell polarity of the cell accompanied by increases in cell motility (The UniProt Consortium, 2017).	Cell Motility
		Calcium-induced platelet aggregation, mediates signals from the CD36/GP4 receptor for granule release, and activates the integrin heterodimer ITGA2B-ITGB3 through the RAP1GAP pathway for adhesion (The UniProt Consortium, 2017).	Platelet Aggregation
		Activates Raf1, Rap1 and Ras in the classical MAPK pathway, which leads to activation of STMN1, cPLA2, ATF4 and FOS. (Via FOS)	Angiogenesis Differentiation

(Continued)

Condition	Protein	Function	Key Word
Wnt/Ca2+	PRKCA	(Via ATF4 and FOS)	Proliferation
		(Via ATF4)	Apoptotic Stress Response
			Cancer Drug Resistance
			Cancer Invasion
		(Via STMN1)	Microtubule Destabilization
	PRKCA	(Via STMN1)	Viral Carcinogenesis
			Immunity Regulation
		(Via cPLA2)	Immune Response
			Phagocytosis
			Platelet Activation
		Promotes cell growth by phosphorylating and activating RAF1 (The UniProt Consortium, 2017).	Cell Growth

TABLE 4. Functional annotation of the final effectors of the TGF-β Signalling Pathway.

Condition	Protein	Function	Key Word
BMP	ID1	Can inhibit the DNA binding and transcriptional activation ability of basic HLH proteins with which it interacts. Regulates: cellular growth, senescence, differentiation, apoptosis, angiogenesis, and neoplastic transformation. Inhibits skeletal muscle and cardiac myocyte differentiation. Leads to osteoblast differentiation, neurogenesis, neurogenesis, ventral mesoderm specification (The UniProt Consortium, 2017).	Transcription Regulation
TGFβ	CDKN2B	Interacts strongly with CDK4 and CDK6 preventing their action, thus causing a cell cycle arrest at the G1 phase (Hydbring et al., 2016).	Cell Cycle Arrest

(Continued)

Condition	Protein	Function	Key Word
TGFβ	ROCK1	Acts as a negative regulator of VEGF-induced angiogenic endothelial cell activation, reducing cell interactions and mediates angiogenic processes (Bryan et al., 2010).	Angiogenesis
		Suppressor of inflammatory cell migration by regulating PTEN phosphorylation and stability (Vemula, Shi, Hanneman, Wei, & Kapur, 2010).	Antiinflammatory
		Rho-kinases are modulators of processes involving cytoskeletal rearrangement: regulation of smooth muscle contraction, cell adhesion and motility via phosphorylation of DAPK3, GFAP, LIMK1, LIMK2, MYL9/MYL2, PFN1 and PPP1R12A, promotes src-dependent blebbing (The UniProt Consortium, 2017).	Cytoskeletal Organization
Activin	SMAD2	(Kanehisa, Sato, Kawashima, Furumichi, & Tanabe, 2016)	Gonadal Growth
	SMAD3		Embryo Differentiation
	SMAD4		Placenta Formation
Nodal	PITX2	Transcriptional regulator involved in basal and hormone-regulated activity of prolactin. Involved in the development of the eye, tooth and abdominal organs. During embryonic development, exerts a role in the expansion of muscle progenitors (The UniProt Consortium, 2017).	Morphogenesis

TABLE 5. Functional annotation of the final effectors of the Focal Adhesion Pathway.

Condition	Protein	Function	Key Word
ECM – Receptor Interaction	ZYX	Adhesion plaque protein. Binds α-actinin and the CRP protein. Important for targeting TES and ENA/VASP family members to focal adhesions and for the formation of actin-rich structures. (The UniProt Consortium, 2017).	Cytoskeleton Organization Scaffold

(Continued)

Condition	Protein	Function	Key Word
ECM – Receptor Interaction	ZYX	Abnormal regulation of the actin cytoskeleton leads to the invasive and metastatic phenotypes of malignant cancer cells (Olson & Sahai, 2009).	Cancer Invasion
		Molecules that link migratory signals to the actin cytoskeleton are upregulated in invasive and metastatic cancer cells leading to formation of invasive protrusions used by tumour cells, such as lamellipodia and invadopodia (Yamaguchi & Condeelis, 2007)	Cancer Invasion
	VASP	Ena/VASP proteins are actin-associated proteins involved in a range of processes dependent on cytoskeleton remodeling and cell polarity such as axon guidance, lamellipodial and filopodial dynamics, platelet activation and cell migration. VASP promotes actin filament elongation. It protects the barbed end of growing actin filaments against capping and increases the rate of actin polymerization in the presence of capping protein. VASP stimulates actin filament elongation by promoting the transfer of profilin-bound actin monomers onto the barbed end of growing actin filaments. Plays a role in actin-based mobility of Listeria monocytogenes in host cells. Regulates actin dynamics in platelets and plays an important role in regulating platelet aggregation (The UniProt Consortium, 2017).	Cytoskeletal Organization Platelet Aggregation
		Abnormal regulation of the actin cytoskeleton leads to the invasive and metastatic phenotypes of malignant cancer cells (Olson & Sahai, 2009). Molecules that link migratory signals to the actin cytoskeleton are upregulated in invasive and metastatic cancer cells leading to formation of invasive protrusions used by tumour cells, such as lamellipodia and invadopodia (Yamaguchi & Condeelis, 2007).	Cancer Invasion

(Continued)

Condition	Protein	Function	Key Word
ECM – Receptor Interaction or Interaction of Cytokines and their Receptors	ACTA1	Involved in cell motility, structure, and integrity. It is found in two main states: G-actin is the globular monomeric form and F-actin forms helical polymers. Both G- and F-actin are intrinsically flexible structures.	Cytoskeleton Organization
ECM – Receptor Interaction or Interaction of Cytokines and their Receptors	ACTA1	Actin polymerization leads to formation of lamellipodia and filopodia (The UniProt Consortium, 2017).	Motility Scaffold
		Abnormal regulation of the actin cytoskeleton leads to invasive and metastatic phenotypes of malignant cancer cells (Olson & Sahai, 2009). Molecules that link migratory signals to the actin cytoskeleton are upregulated in invasive and metastatic cancer cells leading to formation of invasive protrusions used by tumour cells, such as lamellipodia and invadopodia (Yamaguchi & Condeelis, 2007).	Cancer Invasion
	FLNB	Connects cell membrane constituents to the actin cytoskeleton. May promote orthogonal branching of actin filaments and links actin filaments to membrane glycoproteins. Anchors various transmembrane proteins to the actin cytoskeleton. Interaction with FLNA may allow neuroblast migration from the ventricular zone into the cortical plate. Various interactions and localizations of isoforms affect myotube morphology and myogenesis (The UniProt Consortium, 2017).	Cellular Community Morphogenesis Cytoskeleton Organization Motility Scaffold
		Abnormal regulation of the actin cytoskeleton leads to the invasive and metastatic phenotypes of malignant cancer cells (Olson & Sahai, 2009). Molecules that link migratory signals to the actin cytoskeleton are upregulated in invasive and metastatic cancer cells leading to formation of invasive protrusions used by tumour cells, such as lamellipodia and invadopodia (Yamaguchi & Condeelis, 2007).	Cancer Invasion

(Continued)

Condition	Protein	Function	Key Word
ECM – Receptor Interaction or Interaction of Cytokines and their Receptors	FLNB	Tumour-promoting effect by interacting with signalling molecules. At the nucleus, interacts with transcription factors suppressing tumour growth and inhibit metastasis. It correlates patient prognosis, depending on its localization and cancer type (Savoy & Ghosh, 2013).	Tumour Suppressor
	PAK4	Prevents caspase-8 binding to death domain receptors (Gnesutta, Qu, & Minden, 2001).	Antiapoptosis
		Roll in cell-cycle progression by phosphorylating RAN, if silenced induces a blockade at the G2/M transition (Bompard et al., 2010).	Cell Cycle Progression
		Stimulates cell survival by phosphorylating the BCL2 antagonist of cell death BAD (The UniProt Consortium, 2017).	Cell Survival
		Phosphorylates and inactivates the protein phosphatase SSH1 and LIMK1, leading to increased inhibitory phosphorylation of the actin binding/depolymerizing factor cofilin. Decreased cofilin activity may lead to stabilization of actin filaments. Also phosphorylates ARHGEF2 and activates the downstream target RHOA that plays a role in the regulation of assembly of focal adhesions and actin stress fibres (The UniProt Consortium, 2017).	Cytoskeleton Regulation
	CCND1	(See TP53 Signalling Pathway)	Cell Cycle Progression
	BIRC2	Inhibits apoptosis by binding to TRAF1 and TRAF2 and targets caspases for inactivation acting as a E3 ubiquitin-protein ligase (Darding et al., 2011).	Antiapoptosis
		Protects from spontaneous formation of the ripoptosome, a large multi-protein complex that has the capability to kill cancer cells, by ubiquitinating RIPK1 and CASP8 (Tenev et al., 2011).	Cancer Survival

(Continued)

Condition	Protein	Function	Key Word
ECM – Receptor Interaction or Interaction of Cytokines and their Receptors	BIRC2	Acts as an important regulator of innate immune signalling via regulation of Toll-like receptors (TLRs), Nodlike receptors (NLRs) and RIG-I like receptors (RLRs), collectively referred to as pattern recognition receptors (PRRs) (Sharma, Kaufmann, & Biswas, 2017).	Inflammatory Response
	BCL2	This gene encodes an integral outer mitochondrial membrane protein that blocks the apoptotic death of some cells, such as lymphocytes (The UniProt Consortium, 2017).	Antiapoptosis

TABLE 6. Functional annotation of the final effectors of the Adherens Junction Pathway.

Condition	Protein	Function	Key Word
Nectin	PARD3	Adapter protein involved in asymmetrical cell division and cell polarity processes, also has a central role in the formation of epithelial tight junctions (The UniProt Consortium, 2017).	Cell Polarity Scaffold
	WASL	Regulates actin polymerization by stimulating the actin-nucleating activity of the Arp2/3 complex. It is also involved in mitosis and cytokinesis. (The UniProt Consortium, 2017).	Cytoskeleton Organization
	WAS	Interacts with the Arp2/3 complex to induce actin polimerization (The UniProt Consortium, 2017).	Cytoskeleton Organization
		Mediates actin filament reorganization and the formation of actin pedestals upon infection by pathogenic bacteria (The UniProt Consortium, 2017).	Immune Response
	IQGAP1	It interacts with components of the cytoskeleton and adhesion molecules and other molecules to regulate cell morphology and motility (The UniProt Consortium, 2017).	Cytoskeleton Organization
		Contributes to the transformed cancer cell phenotype by regulating signalling pathways involved in cell proliferation and transformation,	Cancer Invasion

(Continued)

Condition	Protein	Function	Key Word
Nectin	IQGAP1	weakening of cell – cell adhesion contacts and stimulation of cell motility and invasion (Johnson, Sharma, & Henderson, 2009).	Cancer Invasion
	BAIAP2	Associated with formation of stress fibres and cytokinesis. Involved in lamellipodia and filopodia formation in motile cells acting synergistically with ENAH. Pays a role in neurite growth (The UniProt Consortium, 2017).	Cytoskeleton Organization
		Plays a role in the reorganization of the actin cytoskeleton in response to bacterial infection. Participates in actin bundling when associated with EPS8, promoting filopodial protrusions (The UniProt Consortium, 2017).	Immune Response
	WASF2	Associated with formation of stress fibres and cytokinesis. Involved in lamellipodia and filopodia formation in motile cells acting synergistically with ENAH. Pays a role in neurite growth (The UniProt Consortium, 2017).	Cytoskeleton Organization
		Plays a role in the reorganization of the actin cytoskeleton in response to bacterial infection. Participates in actin bundling when associated with EPS8, promoting filopodial protrusions (The UniProt Consortium, 2017).	Immune Response
Nectin and Cadherin	ACTA1	(See Focal Adhesion Pathway)	Cytoskeleton Organization Motility Scaffold Cancer Invasion
Cadherin	CTNNB1	Mutations are commonly found in various cancer such as colorectal and ovarian cancer, pilomatrixoma or medulloblastoma (Forbes et al., 2010). These mutations lead to impossible degradation of the protein and its translocation to the nucleus in the absence of external stimulus, where it continuously drives transcription of its target genes (Stamos & Weis, 2013).	Cancer Development

(Continued)

Condition	Protein	Function	Key Word
Cadherin	CTNNB1	It anchors the actin cytoskeleton and may be responsible for transmitting the contact inhibition signal that causes cells to stop dividing once the epithelial sheet is complete (The UniProt Consortium, 2017).	Cell Proliferation Inhibition
		Part of the complex of proteins that constitute adherens junctions (AJs), necessary for the creation and maintenance of epithelial cell layers by regulating cell growth and adhesion between cells	Scaffold
	LEF1	Hair cell differentiation and follicle morphogenesis (The UniProt Consortium, 2017).	Cell Differentiation
	SNAI2	Critical role for SNAI2 in the pathogenesis of mesenchymal tumours and contributes to cancer progression (Pérez-Mancera et al., 2005). Forced expression of SNAI2 in collaboration with Sox9 in breast cancer cells can efficiently induce entrance into the Tumour Initiating Cell (TIC) state (Ye et al., 2015).	Cancer Development
		Involved in epithelial-mesenchymal transitions and has antiapoptotic activity. Transcriptional repressor, represses BRCA2 expression in breast cells and ITGA3 in keratinocytes. Involved in the regulation of ITGB1 and ITGB4 expression and cell adhesion and proliferation in epidermal keratinocytes. Represses E-Cadherin transcription (The UniProt Consortium, 2017).	EMT Transcription Regulation

TABLE 7. Functional annotation of the final effectors of the Tight Junction Pathway.

Condition	Protein	Function	Key Word
CRB3	MPP4	Localized to the outer limiting membrane in the retina, functions in photoreceptor polarity and the organization of specialized intercellular junctions (The UniProt Consortium, 2017).	Polarity Scaffold

(Continued)

Condition	Protein	Function	Key Word
CRB3	CDC42	RAC1 and CDC42 are active in their GTP-bound state. When active, they bind to a variety of effector proteins to regulate epithelial cell polarity. Furthermore, RAC1 is responsible for growth-factor induced formation of membrane ruffles and promotes cell migration and invasion in glioma cells (Ensign et al., 2013).	Cancer Invasion Cell Polarity Proliferation
	RAC1	RAC1 and CDC42 are active in their GTP-bound state. When active, they bind to a variety of effector proteins to regulate epithelial cell polarity. Furthermore, RAC1 is responsible for growth-factor induced formation of membrane ruffles and promotes cell migration and invasion in glioma cells (Ensign et al., 2013).	Cancer Invasion Cell Polarity Proliferation
	DLG1	Multi-domain scaffolding protein, recruits channels, receptors and Signalling molecules to discrete plasma membrane domains in polarized cells (The UniProt Consortium, 2017).	Cell Polarity Scaffold
	SCRIB	Scaffold protein involved in different aspects of cell Cell Polarity processes, regulating epithelial and neuronal morphogenesis, such as the establishment of apico-basal cell polarity (The UniProt Consortium, 2017).	Cell Polarity
Claudin with Claudin	MPP4	(See CRB3 above)	Polarity Scaffold
	CDC42	(See CRB3 above)	Cancer Invasion Cell Polarity Proliferation

(Continued)

Condition	Protein	Function	Key Word
Claudin with Claudin	DLG1	(See CRB3 above)	Cell Polarity Scaffold
	SCRIB	(See CRB3 above)	Cell Polarity
	RAC1	(See CRB3 above)	Cancer Invasion Cell Polarity Proliferation
	DLG2	Regulates surface expression of NMDA receptors, is part of the postsynaptic protein scaffold of excitatory synapses (The UniProt Consortium, 2017).	Cell Polarity Scaffold
	DLG3	Clustering of NMDA receptors at excitatory synapses, regulates surface expression of NMDA receptors, is part of the postsynaptic protein scaffold of excitatory synapses (The UniProt Consortium, 2017).	Cell Polarity Scaffold
	CLDN2	Major integral membrane proteins localized exclusively at tight junctions. Its increased expression leads to less paracellular permeability between cells (The UniProt Consortium, 2017).	Decreased Permeability Scaffold
	WAS	(See Adherens Junction Pathway)	Cytoskeleton Organization Immune Response
	ACTN1	Actin is a ubiquitous globular protein that is one of the most highly-conserved proteins known. It is found in two main states: G-actin is the globular monomeric form, whereas F-actin forms helical polymers. Both G- and F-actin are intrinsically flexible structures (The UniProt Consortium, 2017).	Cytoskeleton Organization

(Continued)

Condition	Protein	Function	Key Word
Claudin with Claudin	Arp2/3	Arp2/3-mediated actin polymerization both at the Golgi apparatus and along tubular membrane and actin branching (The UniProt Consortium, 2017).	Cytoskeleton Organization
	MYL2	Very important for cardiac muscle contraction via tight junction assembly, it increases myosin lever arm stiffness and myosin head diffusion (The UniProt Consortium, 2017).	Scaffold
	VASP	(See Focal Adhesion Pathway)	Cytoskeleton Organization
			Cancer Invasion
			Platelet Aggregation
Occludin with Occludin	RAC1	(See CRB3 above)	Cancer Invasion
			Cell Polarity
			Proliferation
	DLG2	(See Claudins above)	Cell Polarity
			Scaffold
	DLG3		Cell Polarity
			Scaffold
Occludin with Occludin	CLDN2	(See Claudins above)	Decreased Permeability
			Scaffold
	WAS	(See Adherens Junction Pathway)	Cytoskeleton Organization
			Immune Response
	ACTN1	(See Claudins above)	Cytoskeleton Organization

(Continued)

Condition	Protein	Function	Key Word
Occludin with Occludin	Arp2/3	(See Claudins above)	Cytoskeleton Organization
	MYL2	(See Claudins above)	Scaffold
	VASP	(See Claudins above)	Cytoskeleton Organization
			Scaffold
	PCNA	Involved in the control of eukaryotic DNA replication by increasing the polymerases processibility during elongation of the leading strand (The UniProt Consortium, 2017).	Proliferation
	CCND1	(See TP53 Signalling Pathway)	Cell Cycle Progression
	ERBB2	Regulates outgrowth and stabilization of peripheral microtubules. Implicated in transcriptional activation of CDKN1A; the function involves STAT3 and SRC. Involved in the transcription of rRNA genes by RNA Pol I and enhances protein synthesis and cell growth (The UniProt Consortium, 2017).	Growth
			Cytoskeleton Organization
	Runx1	Runx1 is the alpha subunit of CBF, which is involved in the development of normal hematopoiesis (Hart & Foroni, 2002).	Hematopoiesis
JAM/JAM	RAC1	(See CRB3 above)	Cancer Invasion
			Cell Polarity
			Proliferation
	DLG3	(See Claudins above)	Cell Polarity
			Scaffold
	DLG2	(See Claudins above)	Cell Polarity

(Continued)

Condition	Protein	Function	Key Word
JAM/JAM	DLG2	(See Claudins above)	Scaffold
	CLDN2	(See Claudins above)	Decreased Permeability
			Scaffold
	Integrin	Integrins are receptors for collagen and support cell migration through collagen rich extracellular matrix. They mediate dynamic interactions between the extracellular matrix and the actin cytoskeleton during cell motility (Huttenlocher & Horwitz, 2011).	Cell Migration
	MYL2	(See Claudins above)	Decreased Permeability
			Scaffold
Bves/Bves	CLDN2	(See Claudins above)	Decreased Permeability
			Scaffold
MarvelD3/MarvelD3	CD1	Mediate the presentation of primarily lipid and glycolipid antigens of self or microbial origin to T cells (Barral & Brenner, 2007).	Cell Survival
			Immune Response
CFTR	PCNA/CCND1	(See Occludins above)	Proliferation
	ERBB2	(See Occludins above)	Growth
			Cytoskeleton Organization
	Runx1	(See Occludins above)	Hematopoiesis
	PRKCE	Assembly of the Tight Junction, PCK epsilon triggers the anchorage of the actin cytoskeleton to the plasma membrane via moesin (Newton & Messing, 2010).	Scaffold
aPKC/PAR6	MYL2	(See Claudins Above)	Scaffold

(Continued)

Condition	Protein	Function	Key Word
JAM4/JAM4	RAB8A	Regulators of intracellular membrane trafficking, from the formation of transport vesicles to their fusion with membranes (The UniProt Consortium, 2017).	AJ Assembly Vesicular Transport
JAM4/JAM4	RAB13	Regulates transport to the plasma membrane of transmembrane proteins, thereby, it regulates the assembly and the activity of tight junction. Key regulator of intracellular membrane trafficking, from the formation of transport vesicles to their fusion with membranes. Plays also a role in angiogenesis through regulation of endothelial cells chemotaxis (The UniProt Consortium, 2017).	Angiogenesis Tight Junction Assembly Vesicular Transport
Tricellulin/Tricellulin	WAS	(See Adherens Junction Pathway)	Cytoskeleton Organization Immune Response
	ACTN1	(See Claudins above)	Cytoskeleton Organization
	Arp2/3		Cytoskeleton Organization

TABLE 8. Functional annotation of the final effectors of the mTOR Signalling Pathway.

Condition	Protein	Function	Key Word
mTORC1	CLIP1	Promotes microtubule growth and microtubule bundling. Links cytoplasmic vesicles to microtubules and thereby plays an important role in intracellular vesicle trafficking (The UniProt Consortium, 2017).	Microtubule Organization
	GRB10	Key regulator of adiposity, thermogenesis, and energy expenditure (Liu et al., 2014).	Lipolysis

(Continued)

Condition	Protein	Function	Key Word
mTORC1	LPIN1	Catalyzes the penultimate step in triglyceride synthesis. Expression of this gene is required for adipocyte differentiation (The UniProt Consortium, 2017).	Differentiation Lipid Biosynthesis
	ULK1	Involved in autophagy in response to starvation. Acts upstream of phosphatidylinositol 3-kinase PIK3C3 to regulate the formation of autophagophores, the precursors of autophagosomes. Part of regulatory feedback loops in autophagy: downstream effector and a negative regulator of mTORC1 (The UniProt Consortium, 2017).	Autophagy
	EIF4E	component of the eukaryotic translation initiation factor 4F complexids in translation initiation by recruiting ribosomes to the 5'-cap structure (The UniProt Consortium, 2017).	Protein Synthesis
	EIF4B	Required for the binding of mRNA to ribosomes.inds near the 5-terminal cap of mRNA in presence of EIF-4F and ATP. Promotes the ATPase activity and the ATP-dependent RNA unwinding activity of both EIF4-A and EIF4-F (The UniProt Consortium, 2017).	Protein Synthesis
	RPS6KB1	Cytoplasmic ribosomal protein that is a component of the 40S subunit (The UniProt Consortium, 2017).	Protein Synthesis
mTORC2	RHOA	Role in the regulation of assembly of focal adhesions and actin stress fibres (Sit & Manser, 2011).	Cytoskeleton Organization
	PKC	(See Wnt Signalling Pathway)	Angiogenesis Apoptosis Apoptotic Stress Response

(Continued)

Condition	Protein	Function	Key Word
mTORC2	PKC	(See Wnt Signalling Pathway)	Cancer Drug Resistance
			Cancer Invasion
			Cancer Proliferation
			Cytoskeleton Organization
			Cytoskeleton Regulation
			Differentiation
			Immunity Regulation
			Immune Response
			Phagocytosis
			Platelet Activation
			Proliferation
			Viral Carcinogenesis
	SGK1	SGKs are related to Akt (also called PKB), a serine/threonine kinase that plays a crucial role in promoting cell survival. Like Akt, SGKs are activated by the phosphoinositide-3 kinase (PI3K) and translocate to the nucleus upon growth factor stimulationt SGK1, like Akt, promotes cell survival and that it does so in part by phosphorylating and inactivating FKHRL1 (Brunet et al., 2001).	Cell Survival

TABLE 9. Functional annotation of the final effectors of the AMPK Signalling Pathway.

Condition	Protein	Function	Key Word
AMPK	PFKFB1	Catalyses the synthesis and degradation of fructose 2,6-bisphosphate, phosphorylated by AMPK leads to increased glycolysis (The UniProt Consortium, 2017).	Glycolysis
	G6PC	Hydrolyzes glucose-6-phosphate to glucose in the endoplasmic reticulum. Forms with the glucose-6-phosphate transporter complex responsible for glucose production through glycogenolysis and gluconeogenesis. Hence, it is the key enzyme in homeostatic regulation of blood glucose level (The UniProt Consortium, 2017).	Gluconeogenesis
	PCK2	Catalyzes the conversion of oxaloacetate (OAA) to phosphoenolpyruvate (PEP), the rate-limiting step in the metabolic pathway that produces glucose from lactate and other precursors derived from the citric acid cycle (The UniProt Consortium, 2017).	Gluconeogenesis
	PPARGC1A	Transcriptional coactivator, can regulate key mitochondrial genes that contribute to the program of adaptive thermogenesis. Plays an essential role in metabolic reprogramming in response to dietary availability through coordination of the expression of a wide array of genes involved in glucose and fatty acid metabolism (The UniProt Consortium, 2017).	Gluconeogenesis
	CCND1	(See TP53 Signalling Pathway)	Cell Cycle Progression
	CCNA	Essential for the control of the cell cycle at the G1/S (start) and the G2/M (mitosis) transitions (The UniProt Consortium, 2017).	Cell Cycle Progression
	EEF2	Essential factor for protein synthesis, it promotes the GTP-dependent translocation of the nascent protein chain from the A-site to the P-site of the ribosome (The UniProt Consortium, 2017).	Protein Synthesis

(Continued)

Condition	Protein	Function	Key Word
AMPK	SIRT1	Sirt1's deacetylation of Peroxisome proliferator-activated receptor Gamma Coactivator-1α (PGC-1α) has been extensively implicated in metabolic control and mitochondrial biogenesis (Tang, 2016).	Mitochondrial Biogenesis
	SLC2A4	Insulin-regulated facilitative glucose transporter, within minutes of insulin stimulation, the protein moves to the cell surface and begins to transport glucose across the cell membrane (The UniProt Consortium, 2017).	Glucose Transporter
	GYS	Catalyzes the addition of glucose monomers to the growing glycogen molecule through the formation of alpha-1,4-glycoside linkages	Glycogen Synthesis
	LIPE	In adipose tissue and heart, it primarily hydrolyzes stored triglycerides to free fatty acids, while in steroidogenic tissues, it converts cholesteryl esters to free cholesterol for steroid hormone production (The UniProt Consortium, 2017).	Lipolysis
	HMGCR	Catalyzes the conversion of 3-hydroxy-3-methyl-glutaryl-CoA to mevalonic acid, the rate limiting step in cholesterol biosynthesis (The UniProt Consortium, 2017).	Cholesterol Synthesis
	FAS	Catalyzes the conversion of acetyl-CoA and malonyl-CoA to the 16-carbon fatty acid palmitate (The UniProt Consortium, 2017).	Lipid Biosynthesis
		FASN overexpression and hyperactivity is commonly associated with malignant cells (Menendez & Lupu, 2007).	Cancer Transcriptional Misregulation
	ACACA	Catalyzes the rate-limiting reaction in the biogenesis of long-chain fatty acids (The UniProt Consortium, 2017).	Lipid Biosynthesis
	SCD	Enzyme involved in fatty acid biosynthesis, primarily the synthesis of oleic acid (The UniProt Consortium, 2017).	Lipid Biosynthesis

(Continued)

Condition	Protein	Function	Key Word
AMPK	CFTR	Transporter, secretes chloride outside of the cell (The UniProt Consortium, 2017).	Chloride Secretion
	ACACB	Thought to control fatty acid oxidation by means of the ability of malonyl-CoA to inhibit carnitine-palmitoyl-CoA transferase I, the rate-limiting step in fatty acid uptake and oxidation by mitochondria (The UniProt Consortium, 2017).	Fatty Acid Oxidation
	MLYCD	Catalyses the breakdown of malonyl-CoA to acetyl-CoA and carbon dioxide. Malonyl-CoA is an intermediate in fatty acid biosynthesis, also inhibits the transport of fatty acyl CoAs into mitochondria. Consequently, the encoded protein acts to increase the rate of fatty acid oxidation (The UniProt Consortium, 2017).	Fatty Acid Oxidation
	CPT1A	Regulates the beta-oxidation and transport of long-chain fatty acids into mitochondria (The UniProt Consortium, 2017).	Fatty Acid Oxidation Regulation
	RPS6KB1	(See mTOR Signalling Pathway)	Protein Synthesis
	EIF4E	(See mTOR Signalling Pathway)	Protein Synthesis
	PPARG	Nuclear receptor that binds peroxisome proliferators such as fatty acids. Binds to DNA specific PPAR response elements (PPRE) and modulates the transcription of its target genes, such as acyl-CoA oxidase. It therefore controls the peroxisomal beta-oxidation pathway of fatty acids. Key regulator of adipocyte differentiation and glucose homeostasis (The UniProt Consortium, 2017).	Differentiation
	ULK1	(See mTOR Signalling Pathway)	Autophagy

TABLE 10. Functional annotation of the final effectors of the PI3K-AKT Signalling Pathway.

Condition	Protein	Function	Key Word
Hypoxia/AMP/LKB1	EIF4E	(See mTOR Signalling Pathway)	Protein Synthesis
	EIF4B	(See mTOR Signalling Pathway)	Protein Synthesis
	RPS6KB1	(See mTOR Signalling Pathway)	Protein Synthesis
GF	MAPK1	Translocates to the nucleus of the stimulated cells, where it phosphorylates nuclear target, mediating diverse biological functions such as cell growth, adhesion, survival and differentiation. Plays also a role in initiation and regulation of meiosis, mitosis, and post-mitotic functions in differentiated cells (Sato et al., 2011). Endothelial angiogenic response via frequency-sensitive MAPK/ERK pathway activation (Sheikh et al., 2013).	Angiogenesis DNA Repair Proliferation
PIP3	PKC	(See Wnt Signalling Pathway)	Angiogenesis Apoptosis Apoptotic Stress Response Cancer Drug Resistance Cancer Invasion Cancer Proliferation Cytoskeleton Organization Cytoskeleton Regulation Differentiation
PIP3	PKC	(See Wnt Signalling Pathway)	Immunity Regulation Immune Response Phagocytosis

(Continued)

Condition	Protein	Function	Key Word
PIP3	PKC	(See Wnt Signalling Pathway)	Platelet Activation
			Proliferation
			Viral Carcinogenesis
	PKN3	Regulation of cell adhesion, cell cycle progression, actin cytoskeleton assembly, cell migration (The UniProt Consortium, 2017).	Scaffold
		Mediates malignant cell growth contributing to metastatic cell growth, required for invasive prostate cancer (Leenders et al., 2004).	Cancer Invasion
	SGK3	(See mTOR Signalling Pathway)	Cell Survival
	NOS3	Leads to NO production, which mediates VEGF-induced angiogenesis in coronary vessels and promotes blood clotting through the activation of platelets	Angiogenesis
			Platelet Activation
		NO• is a physiological regulator of cell growth, directly involved in the control of cell cycle progression, including the expression and activity of diverse cyclins and cyclin-dependent kinases, their physiological inhibitors, and the master transcriptional regulator retinoblastoma protein (Villalobo, 2006).	Proliferation Arrest
	BRCA1	Genomic stability, tumor suppressor. Plays a role in DNA repair of ds-breaks, and recombination. Regulates centrosomal microtubule nucleation. Required for cell cycle progression G2/M. Required for cell cycle arrest after ionizing irradiation in both the S-phase and the G2 phase of the cell cycle. Involved in transcriptional regulation of P21 in response to DNA damage (The UniProt Consortium, 2017).	DNA Repair

(Continued)

Condition	Protein	Function	Key Word
PIP3	BRCA1	Probable predisposing mutations have been detected in five of eight kindreds presumed to segregate BRCA1 susceptibility alleles (Miki et al., 1994).	Cancer Development
	GYS	(See AMPK Signalling Pathway)	Glycogen Synthesis
	PCK2	(See AMPK Signalling Pathway)	Gluconeogenesis
	G6Pace	(See AMPK Signalling Pathway)	Gluconeogenesis
	MYC	(See MAPK Signalling Pathway)	Cell Cycle Progression Cell Growth Differentiation Stem Cell Self Renewal
	CCND1	(See Wnt Signalling Pathway)	Cell Cycle Progression
	CDKN1B	Regulator of cell cycle progression. Involved in G1 arrest. Potent inhibitor of cyclin E- and CCNA-CDK2 complexes. Forms a complex with cyclin type D-CDK4 complexes and is involved in the assembly, stability, and modulation of CCND1-CDK4 complex activation. Controls the cell cycle progression at G1. The degradation of this protein is required for the cellular transition from quiescence to the proliferative state (The UniProt Consortium, 2017).	Cell Cycle Arrest
	RBL2	Key regulator of entry into cell division. Directly involved in heterochromatin formation by maintaining overall chromatin structure. In particular, that of constitutive heterochromatin by stabilizing histone methylation (The UniProt Consortium, 2017).	Cell Cycle Progression
	FASLG	Induction of apoptosis triggered by binding to FAS (The UniProt Consortium, 2017).	Apoptosis

(Continued)

Condition	Protein	Function	Key Word
PIP3	FASLG	The Fas/FASLG system plays a significant role in tumorigenesis. Research has shown that its impairment in cancer cells may lead to apoptosis resistance and contribute to tumor progression (Villa-Morales & Fernández-Piqueras, 2012).	Cancer Development
		Essential for immune system regulation, including activation-induced cell death (AICD) of T cells and cytotoxic T lymphocyte induced cell death (The UniProt Consortium, 2017).	Immune Response
	BCL2L11	It has been shown to interact with other members of the BCL2 protein family and to act as an apoptotic activator. The expression of this gene can be induced by nerve growth factor (NGF), as well as by the forkhead transcription factor FKHR-L1, which suggests a role of this gene in neuronal and lymphocyte apoptosis. Transgenic studies of the mouse counterpart suggested that this gene functions as an essential initiator of apoptosis in thymocyte-negative selection (The UniProt Consortium, 2017).	Apoptosis
	BCL2L1	Potent inhibitor of cell death. Inhibits activation of caspases. Appears to regulate cell death by blocking the voltage-dependent anion channel (VDAC) by binding to it and preventing the release of the caspase activator, CYC1, from the mitochondrial membrane (The UniProt Consortium, 2017).	Antiapoptosis
	BCL2	(See Focal Adhesion Pathway)	Antiapoptosis

(Continued)

Condition	Protein	Function	Key Word
	CASP9	This protein can undergo autoproteolytic processing and activation by the apoptosome; this step is thought to be one of the earliest in the caspase activation cascade (The UniProt Consortium, 2017).	Apoptosis
	MCL1	Antiapoptotic, regulates apoptosis versus cell survival, and maintenance of viability. Mediates its effects by interacting with other regulators of apoptosis (The UniProt Consortium, 2017).	Antiapoptosis
	MYB	Essential role in hematopoiesis regulation (The UniProt Consortium, 2017).	Hematopoiesis
		Aberrantly expressed or rearranged in leukemias and lymphomas, an oncogene (Ramsay & Gonda, 2008).	Cancer Development
	TP53	Induces growth arrest or apoptosis depending on the physiological circumstances and cell type. Involved in cell cycle regulation as a trans-activator that negatively regulates cell division. Apoptosis induction seems to be mediated either by stimulation of BAX and FAS antigen expression, or by repression of BCL2 expression. With PPIF is involved in activating oxidative stress-induced necrosis. Induces transcription of long intergenic non-coding RNA p21, which participates in TTP53-dependent transcriptional repression leading to apoptosis and seems to have to effect on cell-cycle regulation. Implicated in Notch signalling cross-over. Prevents CDK7 kinase activity when associated to CAK complex in response to DNA damage, thus stopping cell cycle progression (The UniProt Consortium, 2017).	Apoptosis

TABLE 11. Functional annotation of the final effectors of the PPAR Signalling Pathway.

Condition	Protein	Function	Key Word
PPARα/RXR	HMGCS2	Catalyzes the first reaction of ketogenesis, a metabolic pathway that provides lipid-derived energy for various organs during times of carbohydrate deprivation, such as fasting (The UniProt Consortium, 2017).	Ketogenesis
	APOA1	Components of the high density lipoprotein (HDL) and very low density lipoproteins (VLDL) in plasma, promotes cholesterol efflux from tissues to the liver for excretion into plasma (The UniProt Consortium, 2017).	Lipid Transport
	APOA2		Lipid Transport
	APOC3		Lipid Transport
	APOA5		Lipid Transport
	PLTP		Lipid Transport
PPARα/RXR or PPARγ/RXR	ME1	Generates NADPH for fatty acid biosynthesis (The UniProt Consortium, 2017).	Lipid Biosynthesis
	SCD	Stearyl-CoA desaturase that utilizes O(2) and electrons from reduced cytochrome b5 to introduce the first double bond into saturated fatty acyl-CoA substrates (The UniProt Consortium, 2017).	Lipid Biosynthesis
	FADS2	Catalyzes biosynthesis of highly unsaturated fatty acidsfrom precursor essential polyunsaturated fatty acids (The UniProt Consortium, 2017).	Lipid Biosynthesis
	CYP7A1	Catalyzes the first reaction in the cholesterol catabolic pathway in the liver, which converts cholesterol to bile acids. which is the primary mechanism for the removal of cholesterol from the body (The UniProt Consortium, 2017).	Cholesterol Catabolism
	CYP8B1	Catalyzes many reactions involved in drug metabolism and synthesis of cholesterol, steroids and other lipids. Determines the relative amounts of cholic acid and chenodeoxycholic acid, both secreted in the bile and affect the solubility of cholesterol (The UniProt Consortium, 2017).	Cholesterol Metabolism

(Continued)

Condition	Protein	Function	Key Word
PPARα/RXR or PPARγ/RXR	CYP27A1	Catalyzes a rate-limiting step in cholesterol catabolism and bile acid biosynthesis (The UniProt Consortium, 2017).	Cholesterol Catabolism
	NR1H3	The NR1 family members are key regulators of macrophage function, controlling transcriptional programs involved in lipid homeostasis and inflammation. It forms a heterodimer with retinoid X receptor (RXR), and regulates expression of target genes containing retinoid response elements. Plays an important role in the regulation of cholesterol homeostasis (The UniProt Consortium, 2017).	Cholesterol Metabolism Immune Response
PPARα/RXR or PPARγ/RXR or	ACBP	Binds medium- and long-chain acyl-CoA esters with very high affinity and may function as an intracellular carrier of acyl-CoA esters (The UniProt Consortium, 2017).	Fatty Acid Transport
PPARβδ/RXR	FABP1	Plays a role in lipoprotein-mediated cholesterol uptake in hepatocytes. Binds cholesterol and free fatty acids and their coenzyme A derivatives, bilirubin, and some other small molecules in the cytoplasm. May be involved in intracellular lipid transport (The UniProt Consortium, 2017).	Fatty Acid Transport
	FABP1/4	Lipid transport protein in adipocytes. Binds both long chain fatty acids and retinoic acid. Delivers them to their cognate receptors in the nucleus (The UniProt Consortium, 2017).	Fatty Acid Transport
	FABP3	Intracellular transport of long-chain fatty acids and their acyl-CoA esters (The UniProt Consortium, 2017).	Fatty Acid Transport
	CD36	Glycoprotein on platelet surfaces, serves as a receptor for thrombospondin in platelets and various cell lines, binds to collagen, thrombospondin, anionic phospholipids, oxidized LDL and long chain fatty acids and functions in the transport and/or as a regulator of fatty acid transport (The UniProt Consortium, 2017).	Fatty Acid Transport

(Continued)

Condition	Protein	Function	Key Word
PPARα/RXR or PPARγ/RXR or PPARβδ/RXR	LPL	LPL functions as a homodimer, and has the dual functions of triglyceride hydrolase and ligand/bridging factor for receptor-mediated lipoprotein uptake (The UniProt Consortium, 2017).	Fatty Acid Transport
	ACSL1	Activation of long-chain fatty acids for both synthesis of cellular lipids, and degradation via beta-oxidation (The UniProt Consortium, 2017).	Lipid Metabolism
	OLR1	Receptor that mediates the recognition, internalization and degradation of oxidatively modified low density lipoprotein (oxLDL) by vascular endothelial cells. Its association with oxLDL induces the activation of NF-kappa-B through an increased production of intracellular reactive oxygen and a variety of pro-atherogenic cellular responses including a reduction of nitric oxide (NO) release, monocyte adhesion and apoptosis (The UniProt Consortium, 2017).	Apoptosis Fatty Acid Transport
		Receptor for the HSP70 protein involved in antigen cross-presentation to naive T-cells in dendritic cells, thereby participating in cell-mediated antigen cross-presentation. Involved in inflammatory process, by acting as a leukocyte-adhesion molecule at the vascular interface in endotoxin-induced inflammation. Also, acts as a receptor for advanced glycation end (AGE) products, activated platelets, monocytes, apoptotic cells and both Gram-negative and Gram-positive bacteria (The UniProt Consortium, 2017).	Immune Response
	EHHADH	Bifunctional enzyme and is one of the four enzymes of the peroxisomal beta-oxidation pathway (The UniProt Consortium, 2017).	Fatty Acid Oxidation
	ACAA1	Enzyme operative in the beta-oxidation system of the peroxisomes (The UniProt Consortium, 2017).	Fatty Acid Oxidation

(Continued)

Condition	Protein	Function	Key Word
PPARα/RXR or PPARγ/RXR	SCP2	Peroxisome-associated thiolase that is involved in the oxidation of branched chain fatty acids	Fatty Acid Oxidation
or PPARβδ/RXR	ACOX1	First enzyme of the fatty acid beta-oxidation pathway (The UniProt Consortium, 2017).	Fatty Acid Oxidation
	CPT1A	Rate-controlling enzyme of the long-chain fatty acid beta-oxidation pathway in muscle mitochondria, required for the net transport of long-chain fatty acyl-CoAs from the cytoplasm into the mitochondria (The UniProt Consortium, 2017).	Fatty Acid Oxidation
	CPT2	Oxidizes long-chain fatty acids in the mitochondria (The UniProt Consortium, 2017).	Fatty Acid Oxidation
	ACADL	Mitochondrial flavoenzyme involved in fatty acid and branched chain amino-acid metabolism. Catalyzes the initial step of mitochondrial beta-oxidation of straight-chain fatty acid (The UniProt Consortium, 2017).	Fatty Acid Oxidation
	ACADM	Acyl-CoA dehydrogenase specific for acyl chain lengths of 4 to 16 that catalyzes the initial step of fatty acid beta-oxidation (The UniProt Consortium, 2017).	Fatty Acid Oxidation
	ANGPTL4	Hypoxia-induced expression in endothelial cells. May act as a regulator of angiogenesis and modulate tumorigenesis. Can prevent metastasis by inhibiting vascular growth and tumor cell invasion. In response to hypoxia, the unprocessed form of the protein accumulates in the subendothelial extracellular matrix, which limits the formation of actin stress fibers and focal contacts in the adhering endothelial cells and inhibits their adhesion. It also decreases motility of endothelial cells and inhibits the sprouting and tube formation (Galaup et al., 2006).	Antiproliferation Cancer Invasion Inhibition

(Continued)

Condition	Protein	Function	Key Word
PPARα/RXR or PPARγ/RXR	ANGPTL4	Serum hormone that regulates glucose homeostasis, lipid metabolism, and insulin sensitivity (The UniProt Consortium, 2017).	Glucose Metabolism
Or			Lipid Metabolism
PPARβδ/RXR		(Kanehisa, Sato, Kawashima, Furumichi, & Tanabe, 2016)	Adipocyte Differentiation
	FABP4	Lipid transport protein in adipocytes. Delivers long-chain fatty acids and retinoic acid to their cognate receptors in the nucleus (The UniProt Consortium, 2017).	Lipid Transport
		(Kanehisa et al., 2016)	Adipocyte Differentiation
	SORBS1	CBL-associated protein which functions in the signalling and stimulation of insulin. Required for insulin-stimulated glucose transport (The UniProt Consortium, 2017).	Glucose Transport
		(Kanehisa et al., 2016)	Adipocyte Differentiation
	PLIN2	Involved in development and maintenance of adipose tissue (The UniProt Consortium, 2017).	Adipocyte Differentiation
	ADIPOQ	Antagonizes TNF-alpha by negatively regulating its expression in various tissues such as liver and macrophages, and also by counteracting its effects. Inhibits endothelial NF-kappa-B signalling through a cAMP-dependent pathway (The UniProt Consortium, 2017).	Antiinflammatory
		Control of fat metabolism and insulin sensitivity, with direct anti-diabetic, anti-atherogenic and anti-inflammatory activities. In the liver and the skeletal muscle, enhances glucose utilization and fatty-acid combustion (The UniProt Consortium, 2017).	Glucose Metabolism Lipid Metabolism
		(Kanehisa et al., 2016)	Adipocyte Differentiation

(Continued)

Condition	Protein	Function	Key Word
PPARα/RXR or PPARγ/RXR Or PPARβδ/RXR	MMP1	Secreted protease, breaks down the interstitial collagens, including types I, II, and III. Overexpression has a role in initiating tumorigenesis by degrading the stroma, facilitating metastasis (Poola et al., 2005).	Cancer Invasion Cancer Transcriptional Misregulation
		(Kanehisa et al., 2016)	Adipocyte Differentiation
	UCP1	Separates oxidative phosphorylation from ATP synthesis with energy dissipated as heat, also referred to as the mitochondrial proton leak. Is responsible for thermogenic respiration, a specialized capacity of brown adipose tissue and beige fat that participates to non-shivering adaptive thermogenesis to temperature and diet variations and more generally to the regulation of energy balance (The UniProt Consortium, 2017).	Adaptive Thermogenesis
	ILK	Important in the epithelial to mesenchymal transition, and over-expression of this gene is implicated in tumor growth and metastasis (The UniProt Consortium, 2017).	Cancer Invasion
		Mediator of inside-out integrin signalling. Focal adhesion protein part of the complex ILK-PINCH. This complex is considered to be one of the convergence points of integrin- and growth factor-signalling pathway (The UniProt Consortium, 2017).	Cell Survival
	PDK1	Important role in cellular responses to hypoxia and is important for cell proliferation under hypoxia. Protects cells against apoptosis in response to hypoxia and oxidative stress (The UniProt Consortium, 2017).	Antiapoptosis

(Continued)

Condition	Protein	Function	Key Word
	UBC	If polyubiquitin chains are attached in the 6th Lys, involved in DNA repair. Lys-11-linked, involved in endoplasmic reticulum-associated degradation and in cell-cycle regulation. Lys-29-linked is involved in lysosomal degradation. Lys-33-linked is involved in kinase modification. Lys-48-linked is involved in protein degradation via the proteasome. Lys-63-linked is involved in endocytosis, DNA-damage responses as well as in signalling processes leading to activation of the transcription factor NF-kappa-B. Linear polymer chains formed via attachment by the initiator Met lead to cell signalling (The UniProt Consortium, 2017).	Ubiquitination
PPARγ/RXR	PCK2	(See AMPK Pathway)	Gluconeogenesis
	GK	Key enzyme in the regulation of glycerol uptake and metabolism (The UniProt Consortium, 2017).	Glycerol Metabolism
	AQP7	The encoded protein localizes to the plasma membrane and allows movement of water, glycerol and urea across cell membranes. This gene is highly expressed in the adipose tissue where the encoded protein facilitates efflux of glycerol. In the proximal straight tubules of kidney, the encoded protein is localized to the apical membrane and prevents excretion of glycerol into urine (The UniProt Consortium, 2017).	Glycerol Transport Urea Transport Water Transport

TABLE 12. Functional annotation of the final effectors of the VEGF Signalling Pathway.

Condition	Protein	Function	Key Word
VEGFR2	MAPK1	(See PIK3-AKT Signalling Pathway)	Angiogenesis DNA Repair Proliferation *(Continued)*

Condition	Protein	Function	Key Word
VEGFR2	cPLA2	(See MAPK Signalling Pathway)	Immune Response Phagocytosis Platelet Activation
	MT-CO2	PTGS2 is responsible for production of inflammatory prostaglandins. Up-regulation of PTGS2 is also associated with increased cell adhesion, phenotypic changes, resistance to apoptosis and tumour angiogenesis (The UniProt Consortium, 2017).	Antiapoptosis Cancer Angiogenesis Cell Adhesion Inflammatory Response
	PTK2	Essential role in regulating cell migration, adhesion, spreading, reorganization of the actin cytoskeleton, formation and disassembly of focal adhesions and cell protrusions (The UniProt Consortium, 2017).	Cell Migration
	PXN	Cytoskeletal protein involved in actin-membrane attachment at sites of cell adhesion to the extracellular matrix (focal adhesion) (The UniProt Consortium, 2017).	Cell Migration
	HSPB1	(See MAPK Signalling Pathway)	Antiapoptotic Stress Response
	RAC1	(See Tight Junction Pathway)	Cancer Invasion Cell Polarity Proliferation
	NOS3	(See PIK3-AKT Signalling Pathway)	Angiogenesis Platelet Activation Proliferation Arrest
	CASP9	(See PIK3-AKT Signalling Pathway)	Apoptosis
	BAD	Positively regulates cell apoptosis by forming heterodimers with BCL2L1 and BCL2, and reversing their death repressor activity (The UniProt Consortium, 2017).	Apoptosis

TABLE 13. Functional annotation of the final effectors of the Jak-STAT Signalling Pathway.

Condition	Protein	Function	Key Word
STAT/STAT	BCL2	(See Focal Adhesion Pathway)	Antiapoptosis
	MCL1	(See PIK3-AKT Signalling Pathway)	Antiapoptosis
	BCL2L1	(See PIK3-AKT Signalling Pathway)	Antiapoptosis
	PIM1	Contributes to cell proliferation and survival, and thus provides a selective advantage in tumorigenesis (The UniProt Consortium, 2017).	Antiapoptosis
		Overexpressed in hematopoietic malignancies and in prostate cancer (Brasó-Maristany et al., 2016).	Cancer Transcriptional Misregulation
	MYC	(See MAPK Signalling Pathway)	Apoptosis
			Cancer Transcriptional Misregulation
			Cell Cycle Progression
			Cell Growth
			Differentiation
			Stem Cell Self Renewal
	CCND1	(See TP53 Signalling Pathway)	Cell Cycle Progression
	CDKN1A	Regulator of cell cycle progression at G1 (The UniProt Consortium, 2017).	Cell Cycle Arrest
	AOX1	Oxidase with broad substrate specificity, oxidizing aromatic azaheterocyclesas well as aldehydes. roduces hydrogen peroxide and, under certain conditions, can catalyze the formation of superoxide (The UniProt Consortium, 2017).	Lipid Metabolism
	GFAP	One of the major intermediate filament proteins of mature astrocytes. It is used as a marker to distinguish astrocytes from other glial cells during development (The UniProt Consortium, 2017).	Differentiation

TABLE 14. Functional annotation of the final effectors of the cAMP Signalling Pathway.

Condition	Protein	Function	Key Word
cAMP	PLD1	Yields Phosphatidic Acid, precursor for the biosynthesis of many other lipids, influencing the membrane (The UniProt Consortium, 2017).	Cytoskeleton Organization
	PLCE1	Catalyzes the hydrolysis of phosphatidylinositol-4,5-bisphosphate to generate two second messengers: inositol 1,4,5-triphosphate (IP3) and diacylglycerol (DAG), which are involved in calcium handling (The UniProt Consortium, 2017).	Calcium Handling
	PAK1	Phosphorylates BAD and protects cells against apoptosis (The UniProt Consortium, 2017).	Antiapoptosis
		Plays a role in RUFY3-mediated facilitating gastric cancer cells migration and invasion (Wang et al., 2015).	Cancer Invasion
		Involved in the reorganization of the actin cytoskeleton, actin stress fibers and of focal adhesion complexes. Phosphorylates the tubulin chaperone TBCB and thereby plays a role in the regulation of microtubule biogenesis and organization of the tubulin cytoskeleton. Part of a ternary complex that contains PAK1, DVL1 and MUSK that is important for MUSK-dependent regulation of AChR clustering during the formation of the neuromuscular junction (NMJ) (The UniProt Consortium, 2017).	Cytoskeleton Organization
		Plays a role in the regulation of insulin secretion in response to elevated glucose levels (The UniProt Consortium, 2017).	Insulin Secretion
	RHOA	(See mTOR Signalling Pathway)	Cytoskeleton Organization
	AFDN	Belongs to the E-cadherin-catenin system, which plays a role in the organization of homotypic, interneuronal and heterotypic cell-cell adherens junctions (AJs). Nectin- and actin-filament-binding protein that connects nectin to the actin cytoskeleton (The UniProt Consortium, 2017).	Cytoskeleton Organization

(Continued)

Condition	Protein	Function	Key Word
cAMP	BDNF	Binding of this protein to its cognate receptor promotes neuronal survival in the adult brain (The UniProt Consortium, 2017).	Cell Survival
	FOS	(See MAKP Signalling Pathway)	Differentiation
			Proliferation
	JUN	(See MAKP Signalling Pathway)	Angiogenesis
			Proliferation
			Cell Cycle Progression
			Cancer Transcriptional Misregulation
			Differentiation
			Angiogenesis
	GLI1	The encoded transcription factor is activated by the sonic hedgehog signal transduction cascade and regulates stem cell proliferation	Stem Cell Proliferation Regulation
	PTCH1	The encoded protein is the receptor for sonic hedgehog, a secreted molecule implicated in the formation of embryonic structures (The UniProt Consortium, 2017).	Embryonic Structures Formation
		Inactivation of this protein is probably a necessary, if not sufficient step for tumorigenesis. Mutations of this gene have been associated with basal cell nevus syndrome, esophageal squamous cell carcinoma, trichoepitheliomas, transitional cell carcinomas of the bladder (The UniProt Consortium, 2017).	Tumour Suppressor
	HIP1	Membrane-associated protein that functions in clathrin-mediated endocytosis and protein trafficking within the cell. The encoded protein binds to the huntingtin protein in the brain (The UniProt Consortium, 2017).	Endocytosis
			Protein Trafficking

(Continued)

Condition	Protein	Function	Key Word
cAMP	AMH	This complex binds to the anti-Mullerian hormone receptor type 2 and causes the regression of Mullerian ducts in the male embryo that would otherwise differentiate into the uterus and fallopian tubes. This protein also plays a role in Leydig cell differentiation and function and follicular development in adult females (The UniProt Consortium, 2017).	Development
		Able to inhibit the growth of tumours derived from tissues of Muellerian duct origin (The UniProt Consortium, 2017).	Tumour Suppressor
	ACOX1	(See PPAR Signalling Pathway)	Fatty Acid Oxidation
	F2R	Coagulation factor II receptor is a 7-transmembrane receptor involved in the regulation of thrombotic response (The UniProt Consortium, 2017).	Thrombotic Response Regulation
	BAD	(See VEGF Signalling Pathway)	Apoptosis
	LIPE	(See AMPK Signalling Pathway)	Lipolysis
	MYL2	(See tight Junction Pathway)	Scaffold
	TNNI3	Inhibitory subunit of troponin, the thin filament regulatory complex which confers calcium-sensitivity to striated muscle actomyosin ATPase activity, blocking actin-myosin interactions and thereby mediating striated muscle relaxation (The UniProt Consortium, 2017).	Cardiac Myocyte Relaxation
	PLN	Major substrate for the cAMP-dependent protein kinase in cardiac muscle. The encoded protein is an inhibitor of cardiac muscle sarcoplasmic reticulum Ca(2+)-ATPase in the unphosphorylated state, but inhibition is relieved upon phosphorylation of the protein. The subsequent activation of the Ca(2+) pump leads to enhanced muscle relaxation rates, thereby contributing to the inotropic response elicited in heart by beta-agonists. The encoded protein is a key regulator of cardiac diastolic function (The UniProt Consortium, 2017).	Cardiac Diastolic Function

(Continued)

Condition	Protein	Function	Key Word
cAMP	RYR2	Calcium channel that mediates the release of Ca(2+) from the sarcoplasmic reticulum into the cytoplasm and thereby plays a key role in triggering cardiac muscle contraction (The UniProt Consortium, 2017).	Cardiac Myocyte Contraction
	GRIN3A	Glutamate-regulated ion channels, and function in physiological and pathological processes in the central nervous system, such as long term potentiation (Lüscher & Malenka, 2012).	Long Term Potentiation
	CACNG8	Ionotropic glutamate receptor. L-glutamate acts as an excitatory neurotransmitter at many synapses in the central nervous system. Binding of the excitatory neurotransmitter L-glutamate induces a conformation change, leading to the opening of the cation channel, and thereby converts the chemical signal to an electrical impulse (The UniProt Consortium, 2017).	Excitatory Synapses
	CFTR	Involved in the excretion of chloride ions (Kanehisa et al., 2016).	Chloride Excretion
	FXYD1	Ion transporter, uses ATP to contribute to the uptake of 2K+ coupled with the secretion of 3Na+ (Kanehisa et al., 2016).	Ion Transporter
	SLC9A1	Na+/H+ antiporter, is a plasma membrane transporter expressed in the kidney and intestine. Plays a central role in regulating pH homeostasis (The UniProt Consortium, 2017).	Ion Transporter
	ORAI1	Calcium channel, primary way for calcium influx into T-cells. Defects in this gene are a cause of immune dysfunction with T-cell inactivation due to calcium entry defect type (The UniProt Consortium, 2017).	Ion Transporter
	ATP2B1	Catalyses ATP coupled with the transport of calcium out of the cell (The UniProt Consortium, 2017).	Ion Transporter
	CACNA1C	Calcium channel, mediates the influx of calcium ions into the cell upon membrane polarization (Kanehisa et al., 2016).	Ion Transporter

(Continued)

Condition	Protein	Function	Key Word
	HCN2	HCN (hyperpolarization-activated, cyclic nucleotide-gated) channels are members of the cyclic nucleotide-regulated channel family along with cyclic nucleotide-gated (CNG) channels. They are cationic channels that open under hyperpolarization (The UniProt Consortium, 2017).	Ion Transporter
	HCN4	HCN (hyperpolarization-activated, cyclic nucleotide-gated) channels are members of the cyclic nucleotide-regulated channel family along with cyclic nucleotide-gated (CNG) channels. They are cationic channels that open under hyperpolarization (The UniProt Consortium, 2017).	Ion Transporter

TABLE 15. Functional annotation of the final effectors of the Hedgehog Signalling Pathway.

Condition	Protein	Function	Key Word
With Hh	GLI1	(See c-AMP Signalling Pathway)	Stem Cell Proliferation Regulation
	PTCH1	(See c-AMP Signalling Pathway)	Embryonic Structures Formation
			Tumour Suppressor
	HHIP	Interacts with all three HH family members, SHH, IHH and DHH. Modulates hedgehog signalling in several cell types including brain and lung through direct interaction with members of the hedgehog family (The UniProt Consortium, 2017).	Transcription Regulation
	CCND1	(See TP53 Signalling Pathway)	Cell Cycle Progression
	BCL2	Integral outer mitochondrial membrane protein that blocks the apoptotic death of some cells such as lymphocytes (The UniProt Consortium, 2017).	Antiapoptosis

TABLE 16. Functional annotation of the final effectors of the Cell Cycle.

Condition	Protein	Function	Key Word
DNA damage checkpoint	TP53	(See TP53 Signalling Pathway)	Apoptosis Cell Cycle Arrest
Cyclin D and CDK4/6	E2F4 E2F5	Inhibit DNA transcription of S-phase proteins (Dimova & Dyson, 2005).	Cell Cycle Arrest
	E2F2 E2F3 E2F1	DNA transcription of S-phase proteins and Cyclin E (Dimova & Dyson, 2005).	Cell Cycle Progression
CCNA and CDK2	CDC6	DNA synthesis, during S phase levels of CDK increase which phosphorylate Cdc6 and Cdt1, they lose their affinity for the compex and the DNA, so that the helicase stops being inhibited, leading to the double helix to be available for initiation of replication (The UniProt Consortium, 2017).	Cell Cycle Progression DNA Biosynthesis
	ORC	DNA synthesis, during S phase levels of CDK increase which phosphorylate Cdc6 and Cdt1, they lose their affinity for the compex and the DNA, so that the helicase stops being inhibited, leading to the double helix to be available for initiation of replication (The UniProt Consortium, 2017).	Cell Cycle Progression DNA Biosynthesis

TABLE 17. Functional annotation of the final effectors of Apoptosis.

Condition	Protein	Function	Key Word
Extrinsic	GZMB	Disruption of alpha - tubulin leads to disruption of microtubule function, disruption of Mcl-1 leads to BCL2L11-mediated mitochondrial apoptotic events, cleavage of Lamin leads to loss of integrity of the nuclear membrane, cleavage of PARP leads to low synthesis of poly-ADP-ribose, and of ICAD/CAD to DNA fragmentation (The UniProt Consortium, 2017).	Apoptosis

(Continued)

Condition	Protein	Function	Key Word
Extrinsic	GZMB	Granizyme B is secreted by natural killer (NK) cells and cytotoxic T lymphocytes (CTLs) and proteolytically processed to generate the active protease, which induces target cell apoptosis. This protein processes cytokines and degrades extracellular matrix proteins, and these roles are implicated in chronic inflammation and wound healing (The UniProt Consortium, 2017).	Inflammation
	CASP3	Disruption of Lamin leads to loss of integrity of the nuclear membrane, of PARP leads to low synthesis of poly-ADP-ribose (The UniProt Consortium, 2017).	Apoptosis
	JUN	Leads to transcription of pro-apoptotic genes (Fan & Chambers, 2001).	Apoptosis
		(See Wnt Signalling Pathway)	Angiogenesis
			Proliferation
			Cell Cycle Progression
			Cancer Transcriptional Misregulation
			Differentiation
			Cell Polarity
			Cell Migration
	AP1	(See Wnt Signalling Pathway)	Angiogenesis
			Proliferation
			Cell Cycle Progression
			Cancer Transcriptional Misregulation
			Differentiation
			Cell Polarity
			Cell Migration
		Leads to transcription of pro-apoptotic genes (Fan & Chambers, 2001).	Apoptosis

(Continued)

Condition	Protein	Function	Key Word
Extrinsic and Intrinsic	CASP6	Disruption of Actin and Fodrin leads to cell shrinkage and membrane blebbing (The UniProt Consortium, 2017).	Apoptosis
	CASP7	Cleavage of ICAD/CAD leads to DNA fragmentation (The UniProt Consortium, 2017).	Apoptosis
Intrinsic	BAK1	Undergoes conformational change that induces its oligomerization leading to increased permeability of the external mitochondrial membrane, leakage of intermembrane factors such as cytochrome-c (The UniProt Consortium, 2017).	Apoptosis
	BAX	After apoptotic Signalling is inserted in the external membrane of the mitochondria leading to increased permeability of the external mitochondrial membrane, leakage of intermembrane factors such as cytochrome-c (The UniProt Consortium, 2017).	Apoptosis
	BCL2L1	(See PI3K/AKT Signalling Pathway)	Antiapoptosis
	BCL2		
	ENDOG	DNA fragmentation (The UniProt Consortium, 2017).	Apoptosis
	AIFM1		
NFKB1	NFKB1	(See MAPK Signalling Pathway)	Antiapoptosis
			Cancer Transcriptional Misregulation
			Differentiation
			Inflammatory Response
IL-3	BCL2	(See PI3K/AKT Signalling Pathway)	Antiapoptosis

Bibliography

Barral, D. C., & Brenner, M. B. (2007). CD1 antigen presentation: how it works. *Nature Reviews Immunology*, *7*(12), 929–941. https://doi.org/10.1038/nri2191

Bompard, G., Rabeharivelo, G., Frank, M., Cau, J., Delsert, C., & Morin, N. (2010). Subgroup II PAK-mediated phosphorylation regulates Ran activity during mitosis. *The Journal of Cell Biology*, *190*(5), 807–822. https://doi.org/10.1083/jcb.200912056

Brasó-Maristany, F., Filosto, S., Catchpole, S., Marlow, R., Quist, J., Francesch-Domenech, E., … Tutt, A. N. (2016). PIM1 kinase regulates cell death, tumor growth and chemotherapy response in triple-negative breast cancer. *Nature Medicine*, *22*(11), 1303–1313. https://doi.org/10.1038/nm.4198

Brunet, A., Park, J., Tran, H., Hu, L. S., Hemmings, B. A., & Greenberg, M. E. (2001). Protein Kinase SGK Mediates Survival Signals by Phosphorylating the Forkhead Transcription Factor FKHRL1 (FOXO3a). *Molecular and Cellular Biology*, *21*(3), 952–965. https://doi.org/10.1128/MCB.21.3.952-965.2001

Darding, M., Feltham, R., Tenev, T., Bianchi, K., Benetatos, C., Silke, J., & Meier, P. (2011). Molecular determinants of Smac mimetic induced degradation of cIAP1 and cIAP2. *Cell Death and Differentiation*, *18*(8), 1376–1386. https://doi.org/10.1038/cdd.2011.10

Dimova, D. K., & Dyson, N. J. (2005). The E2F transcriptional network: old acquaintances with new faces. *Oncogene*, *24*(17), 2810–2826. https://doi.org/10.1038/sj.onc.1208612

Ensign, F., Patricia, S., Mathews, I., Symons, M., Berens, M., & Tran, N. L. (2013). Implications of Rho GTPase Signalling in Glioma Cell Invasion and Tumor Progression. *Frontiers in Oncology*, *3*. https://doi.org/10.3389/fonc.2013.00241

Fan, M., & Chambers, T. C. (2001). Role of mitogen-activated protein kinases in the response of tumor cells to chemotherapy. *Drug Resistance Updates*, *4*(4), 253–267. https://doi.org/10.1054/drup.2001.0214

Forbes, S. A., Tang, G., Bindal, N., Bamford, S., Dawson, E., Cole, C., … Futreal, P. A. (2010). COSMIC (the Catalogue of Somatic Mutations in Cancer): a resource to investigate acquired mutations in human cancer. *Nucleic Acids Research*, *38*(Database issue), D652–D657. https://doi.org/10.1093/nar/gkp995

Galaup, A., Cazes, A., Le Jan, S., Philippe, J., Connault, E., Le Coz, E., … Germain, S. (2006). Angiopoietin-like 4 prevents metastasis through inhibition of vascular permeability and tumor cell motility and invasiveness. *Proceedings of the National Academy of Sciences of the United States of America*, *103*(49), 18721–18726. https://doi.org/10.1073/pnas.0609025103

Gnesutta, N., Qu, J., & Minden, A. (2001). The Serine/Threonine Kinase PAK4 Prevents Caspase Activation and Protects Cells from Apoptosis. *Journal of Biological Chemistry*, *276*(17), 14414–14419. https://doi.org/10.1074/jbc.M011046200

Goodsell, D. S. (2002). The Molecular Perspective: Tamoxifen and the Estrogen Receptor. *The Oncologist*, *7*(2), 163–164. https://doi.org/10.1634/theoncologist.7-2-163

Hart, S. M., & Foroni, L. (2002). Core binding factor genes and human leukemia. *Haematologica*, *87*(12), 1307–1323. Retrieved from http://www.haematologica.org/content/87/12/1307

Huttenlocher, A., & Horwitz, A. R. (2011). Integrins in Cell Migration. *Cold Spring Harbor Perspectives in Biology*, *3*(9), a005074. https://doi.org/10.1101/cshperspect.a005074

Johnson, M., Sharma, M., & Henderson, B. R. (2009). IQGAP1 regulation and roles in cancer. *Cellular Signalling*, *21*(10), 1471–1478. https://doi.org/10.1016/j.cellsig.2009.02.023

Kanehisa, M., Sato, Y., Kawashima, M., Furumichi, M., & Tanabe, M. (2016). KEGG as a reference resource for gene and protein annotation. *Nucleic Acids Research*, *44*(D1), D457–D462. https://doi.org/10.1093/nar/gkv1070

Koonce, C. J., Walf, A. A., & Frye, C. A. (2009). Trilostane exerts antidepressive effects among wild-type, but not estrogen receptor [beta] knockout mice. *Neuroreport*, *20*(12), 1047–1050.

Leenders, F., Möpert, K., Schmiedeknecht, A., Santel, A., Czauderna, F., Aleku, M., … Klippel, A. (2004). PKN3 is required for malignant prostate cell growth downstream of activated PI 3-kinase. *The EMBO Journal*, *23*(16), 3303–3313. https://doi.org/10.1038/sj.emboj.7600345

Liu, M., Bai, J., He, S., Villarreal, R., Hu, D., Zhang, C., … Liu, F. (2014). Grb10 Promotes Lipolysis and Thermogenesis by Phosphorylation-Dependent Feedback Inhibition of mTORC1. *Cell Metabolism*, *19*(6), 967–980. https://doi.org/10.1016/j.cmet.2014.03.018

Lüscher, C., & Malenka, R. C. (2012). NMDA Receptor-Dependent Long-Term Potentiation and Long-Term Depression (LTP/LTD). *Cold Spring Harbor Perspectives in Biology*, *4*(6). https://doi.org/10.1101/cshperspect.a005710

Menendez, J. A., & Lupu, R. (2007). Fatty acid synthase and the lipogenic phenotype in cancer pathogenesis. *Nature Reviews Cancer*, *7*(10), 763–777. https://doi.org/10.1038/nrc2222

Miki, Y., Swensen, J., Shattuck-Eidens, D., Futreal, P. A., Harshman, K., Tavtigian, S., … Et, A. (1994). A strong candidate for the breast and ovarian cancer susceptibility gene BRCA1. *Science*, *266*(5182), 66–71. https://doi.org/10.1126/science.7545954

Newton, P. M., & Messing, R. O. (2010). The substrates and binding partners of protein kinase Cε. *The Biochemical Journal*, *427*(2), 189–196. https://doi.org/10.1042/BJ20091302

Olson, M. F., & Sahai, E. (2009). The actin cytoskeleton in cancer cell motility. *Clinical & Experimental Metastasis*, *26*(4), 273. https://doi.org/10.1007/s10585-008-9174-2

Poola, I., DeWitty, R. L., Marshalleck, J. J., Bhatnagar, R., Abraham, J., & Leffall, L. D. (2005). Identification of MMP-1 as a putative breast cancer predictive marker by global gene expression analysis. *Nature Medicine*, *11*(5), 481–483. https://doi.org/10.1038/nm1243

Ramsay, R. G., & Gonda, T. J. (2008). MYB function in normal and cancer cells. *Nature Reviews Cancer*, *8*(7), 523–534. https://doi.org/10.1038/nrc2439

Sato, A., Sunayama, J., Matsuda, K., Seino, S., Suzuki, K., Watanabe, E., … Kitanaka, C. (2011). MEK-ERK Signalling Dictates DNA-Repair Gene MGMT Expression and Temozolomide Resistance of Stem-Like Glioblastoma Cells via the MDM2-p53 Axis. *STEM CELLS*, *29*(12), 1942–1951. https://doi.org/10.1002/stem.753

Savoy, R. M., & Ghosh, P. M. (2013). The dual role of filamin A in cancer: can't live with (too much of) it, can't live without it. *Endocrine-Related Cancer*, *20*(6), R341–R356. https://doi.org/10.1530/ERC-13-0364

Sharma, S., Kaufmann, T., & Biswas, S. (2017). Impact of inhibitor of apoptosis proteins on immune modulation and inflammation. *Immunology and Cell Biology*, *95*(3), 236–243. https://doi.org/10.1038/icb.2016.101

Sheikh, A. Q., Taghian, T., Hemingway, B., Cho, H., Kogan, A. B., & Narmoneva, D. A. (2013). Regulation of endothelial MAPK/ERK signalling and capillary morphogenesis by low-amplitude electric field. *Journal of The Royal Society Interface*, *10*(78), 20120548. https://doi.org/10.1098/rsif.2012.0548

Sit, S.-T., & Manser, E. (2011). Rho GTPases and their role in organizing the actin cytoskeleton. *J Cell Sci*, *124*(5), 679–683. https://doi.org/10.1242/jcs.064964

Stamos, J. L., & Weis, W. I. (2013). The β-Catenin Destruction Complex. *Cold Spring Harbor Perspectives in Biology*, *5*(1). https://doi.org/10.1101/cshperspect.a007898

Tang, B. L. (2016). Sirt1 and the Mitochondria. *Moleucles and Cells*, *39*(2), 87–95. https://doi.org/10.14348/molcells.2016.2318

Tenev, T., Bianchi, K., Darding, M., Broemer, M., Langlais, C., Wallberg, F., … Meier, P. (2011). The Ripoptosome, a Signalling Platform that Assembles in Response to Genotoxic Stress and Loss of IAPs. *Molecular Cell*, *43*(3), 432–448. https://doi.org/10.1016/j.molcel.2011.06.006

The UniProt Consortium. (2017). UniProt: the universal protein knowledgebase. *Nucleic Acids Research*, *45*(D1), D158–D169. https://doi.org/10.1093/nar/gkw1099

Villalobo, A. (2006). REVIEW ARTICLE: Nitric oxide and cell proliferation: Nitric oxide and cell proliferation. *FEBS Journal*, *273*(11), 2329–2344. https://doi.org/10.1111/j.1742-4658.2006.05250.x

Villa-Morales, M., & Fernández-Piqueras, J. (2012). Targeting the Fas/FasL signalling pathway in cancer therapy. *Expert Opinion on Therapeutic Targets*, *16*(1), 85–101. https://doi.org/10.1517/14728222.2011.628937

Wang, G., Zhang, Q., Song, Y., Wang, X., Guo, Q., Zhang, J., … Li, F. (2015). PAK1 regulates RUFY3-mediated gastric cancer cell migration and invasion. *Cell Death & Disease*, *6*(3), e1682. https://doi.org/10.1038/cddis.2015.50

Wishart, D. S. (2006). DrugBank: a comprehensive resource for in silico drug discovery and exploration. *Nucleic Acids Research*, *34*(90001), D668–D672. https://doi.org/10.1093/nar/gkj067

Yamaguchi, H., & Condeelis, J. (2007). Regulation of the actin cytoskeleton in cancer cell migration and invasion. *Biochimica et Biophysica Acta*, *1773*(5), 642–652. https://doi.org/10.1016/j.bbamcr.2006.07.001

Ye, X., Tam, W. L., Shibue, T., Kaygusuz, Y., Reinhardt, F., Ng Eaton, E., & Weinberg, R. A. (2015). Distinct EMT programs control normal mammary stem cells and tumour-initiating cells. *Nature*, *525*(7568), 256–260. https://doi.org/10.1038/nature14897

Pathway Activity Analysis (PAA) as a new class of mechanistic biomarker to predict drug responses in drug repositioning for cancer patients

ANNEX 2. Drug Actions Annotation

TABLE 1. Categorization of drugs into activating or inhibiting based on Drug Bank terms and bibliography.

DrugBank Term	Drug-Target Explanation	Action
Acetylation	Acetylsalicilic Acid - TP53 Not known pharmacological action (Wishart, 2006).	Remove
Activator	Triggers the target.	Activation
Adduct	Tigecycline Binds 30S ribosome, inhibiting translation (Wishart, 2006). To remove because it affects bacteria.	Remove
	Vinblastine Binds tubulin, inhibiting mitosis at metaphase (Wishart, 2006).	Inhibition
	Ethionamide Inhibits InhA, the enoyl reductase from Mycobacterium tuberculosis, by forming a covalent adduct with the NAD cofactor (Wishart, 2006). To remove because affects bacteria.	Remove
	Isoniazid Inhibits InhA, the enoyl reductase from Mycobacterium tuberculosis, by forming a covalent adduct with the NAD cofactor (Wishart, 2006). To remove because affects bacteria.	
	Gentamicin Irreversibly binds to specific 30S-subunit proteins and 16S rRNA (Wishart, 2006). To remove because affects bacteria.	
Agonist	Activates a receptor	Activation
Agonistinhibitor	Inhibits a receptor upon binding	Inhibition
Agonistmodulator	Aldesleukin Binds to the IL-2 receptor which leads to heterodimerization of the cytoplasmic domains of the IL-2R beta and gamma(c) chains, activation of the tyrosine kinase Jak3, and phosphorylation of tyrosine residues on the IL-2R beta chain. These events led to the creation of an activated receptor complex, to which various cytoplasmic signalling	Activation

(Continued)

DrugBank Term	Drug-Target Explanation	Action
Agonistmodulator	Molecules are recruited and become substrates for regulatory enzymes (especially tyrosine kinases) that are associated with the receptor. These events stimulate growth and differentiation of T cells (Wishart, 2006).	Activation
	Quinestrol 3-cyclopentyl ether of ethinyl estradiol (the active metabolite). After gastrointestinal absorption, it is stored in adipose tissue where it is slowly released and metabolized principally to the parent compound, ethinyl estradiol. Ethinyl estradiol is a synthetic derivative of the natural estrogen estradiol (Wishart, 2006).	
Agonistpartial agonist	Ergotamine Unknown pharmacological action on Alpha-2B adrenergic receptors (Wishart, 2006).	Remove
	Pseudoephedrine Unknown pharmacological action on Beta-1 adrenergic receptors (Wishart, 2006).	
	Ketamine Unknown pharmacological action on D(2) dopamine receptors (Wishart, 2006).	
	Cariprazine Agonist on Dopamine receptors (Wishart, 2006).	Activation
Allosteric Modulator	Carglumic Acid Synthetic structural analogue of N-acetylglutamate (NAG), an essential allosteric activator of the liver enzyme carbamoyl phosphate synthetase 1 (Wishart, 2006).	Activation
	Galantamine Not known pharmacological action (Wishart, 2006).	Remove
	Vardenafil Unknown pharmacological action (Wishart, 2006).	
	Trilostane In breast cancer, exerts estrogen-like actions through ERβ (Koonce, Walf, & Frye, 2009).	Activation
Allosteric Modulator	Halothane Unknown action (Wishart, 2006).	Remove

(Continued)

DrugBank Term	Drug-Target Explanation	Action
Antagonist	Binds a receptor dampening a biological response	Inhibition
Antagonistagonist	Olanzapine Antagonism at D2 receptors (Wishart, 2006).	Inhibition
	Ergoloid mesylate Unknown pharmacological action at Dopamine receptors (Wishart, 2006). To remove.	Remove
	Tamoxifen Activates and inhibits (Goodsell, 2002). To remove.	
	Risperidone Unknown pharmacological action at lpha-2B adrenergic receptors (Wishart, 2006). To remove.	
	Clomifene Both estrogenic and anti-estrogenic properties (Wishart, 2006). To remove.	
	Ospemifene Ospemifene is a next generation SERM (selective estrogen receptor modulator) that selectively binds to estrogen receptors and either stimulates or blocks estrogen's activity in different tissue types. It has an agonistic effect on the endometrium (Wishart, 2006). To remove.	
Antagonistantibody	Antibody which agonizes its target.	Inhibition
Antagonistbinder	Antagonizes upon binding.	Inhibition
Antagonistblocker	Antagonizes by blocking the target.	Inhibition
Antagonistinhibitor	Antagonizes, inhibits the target.	Inhibition
AntagonistInhibitor	Antagonizes, inhibits the target.	Inhibition
Antagonistinhibitor, competitive	Competes with the target.	Inhibition
Antagonistinhibitory allosteric modulator	Induces a conformational change which diminishes the target's effect.	Inhibition
Antagonistmultitarget	Antagonist with many targets.	Inhibition
Antagonistother/ unknown	Antagonizes.	Inhibition
Antagonistpartial agonist	Mivacurium No pharmacological actions on Muscarinic acetylcholine receptor M2 and M3 (Wishart, 2006). To remove.	Remove

(Continued)

DrugBank Term	Drug-Target Explanation	Action
Antagonistpartial agonist	Tegaserod Agonist of human 5-HT4 receptors (Wishart, 2006).	Activation
	Aripiprazole Antipsychotic activity is likely due to a combination of antagonism at D2 receptors in the mesolimbic pathway and 5HT2A receptors in the frontal cortex. Antagonism at D2 receptors relieves positive symptoms while antagonism at 5HT2A receptors relieves negative symptoms of schizophrenia (Wishart, 2006).	Inhibition
	Penbutolol Blocks the catecholamine activation of β1 adrenergic receptors (Wishart, 2006).	
Antibody	Etanercept Etanercept is a dimeric soluble form of the p75 TNF receptor that can bind to two TNF molecules, thereby effectively removing them from circulation (Wishart, 2006).	Inhibition
	Adalimumab Binds to TNF-alpha and blocks its interaction with the p55 and p75 cell surface TNF receptors (Wishart, 2006).	
	Basiliximab Blocking the interleukin-2 receptor α-chain (Wishart, 2006).	
	Efalizumab Binds to CD11a, a subunit of leukocyte function antigen-1 (LFA-1), expressed on all leukocytes, as a result decreases cell surface expression of CD11a (Wishart, 2006).	
	Natalizumab Binds α4b1 and α4b7 integrins expressed on the all leukocytes except neutrophils, and inhibits theiadhesion of leukocytes to their counter-receptor(s) (Wishart, 2006).	
Antibody	Daclizumab IL-2 receptor antagonist (Wishart, 2006).	Inhibition
	Eculizumab Blocks C5 cleavage (Wishart, 2006).	

(Continued)

DrugBank Term	Drug-Target Explanation	Action
Antibody	Farletuzumab Against the FOLR1 which is overexpressed in ovarian cancer (Jelovac & Armstrong, 2012).	Inhibition
	Tocilizumab Binds soluble and membrane-bound IL-6 receptors, inhibiting IL-6-mediated signalling (Wishart, 2006).	
	Pertuzumab Binds to the HER2 receptor and inhibits the ability of HER2 to interact with other HER family members (Wishart, 2006).	
	Denosumab Prevents RANKL from activating its receptor (Wishart, 2006).	
	Golimumab Inhibits soluble and transmembrane human TNFα (Wishart, 2006).	
	Raxibacumab Inhibits the binding of PA to its cellular receptors (Wishart, 2006).	
	Vedolizumab Inhibits the α4β7 integrin (Wishart, 2006).	
	Nivolumab Binds programmed cell death 1 (PD-1) receptor, blocks interaction with its ligands PD-L1 and PD-L2 (Wishart, 2006).	
	Gemtuzumab ozogamicin Against the CD33 antigen expressed by hematopoietic cells. Binding of the anti-CD33 antibody portion of Mylotarg with the CD33 antigen results in the formation of a complex that is internalized. Upon internalization, the calicheamicin derivative is released inside the lysosomes of the myeloid cell. The released calicheamicin derivative binds to DNA in the minor groove resulting in DNA double strand breaks and cell death (Wishart, 2006). Too broad a function.	Remove
	Trastuzumab Binds to the HER2 cells, leading to their antibody mediated killing (Wishart, 2006). Too broad a function.	

(Continued)

DrugBank Term	Drug-Target Explanation	Action
Antibody	Rituximab Binds to the CD20 antigen, which is predominantly expressed on mature B cells and on >90% of B-cell non-Hodgkin's lympohomas. The antibody leads to selective killing of B-cells (Wishart, 2006). Too broad a function. Ibritumomab Targets CD20 on B-cells, radioactive yttrium to destroy the cell via production of beta particles (Wishart, 2006). Too broad a function. Tositumomab CD20, binding appears to induce apoptosis, complement-dependent cytotoxicity and cell death through ionizing radiation (Wishart, 2006). Too broad a function. Alemtuzumab CD52 on B and T, antibody-dependent lysis of leukemic cells (Wishart, 2006). Too broad a function. Methyl aminolevulinate Topical application, porphyrins accumulate intracellularly in the treated skin lesions, upon light activation in the presence of oxygen, singlet oxygen is formed which causes damage to mitochondria, phototoxicity (Wishart, 2006). Too broad a function. ado-trastuzumab emtansine Binds to the HER2 (or c-erbB2) proto-oncogene, an EGF receptor-like protein found on 20-30% of breast cancer cells, leads to antibody mediated killing of the positive cells (Wishart, 2006). Too broad a function. Obinutuzumab Binds to type II CD20, higher induction of antibody-dependant cytotoxicity and direct cytotoxic effect (Wishart, 2006). Too broad a function.	Remove

(Continued)

DrugBank Term	Drug-Target Explanation	Action
Binder	Antihemophilic Factor ASGR2 Unknown pharmacological action (Wishart, 2006).	Remove
	Antihemophilic Factor VWF - Promotes adhesion and aggregation of platelets at wound sites, thereby inducing platelet plug formation. Antihemophilic Factor binds it (Wishart, 2006).	Activation
	Menotropins Bind the FSH Receptor, which results in ovulation in the absence of sufficient endogenous LH (Wishart, 2006).	
	Interferon gamma-1b Agonist of IFNGR1, leading to a complex of IFNGR1 and IFNGR2 and activation of the pathway (Wishart, 2006).	
	Palifermin Binds the KGF receptor and activates it (Wishart, 2006).	
	Somatropin recombinant Agonist of GH (Wishart, 2006).	
	Interferon alfacon-1 Interferon alfa receptors agonist (Wishart, 2006).	
	Insulin, porcine Activates the insulin receptor (Wishart, 2006).	
	Choriogonadotropin alfa FSHR agonist (Wishart, 2006).	
	Octreotide Octreotide exerts pharmacologic actions similar to the natural hormone, somatostatin (Wishart, 2006).	
	Interferon Alfa-2b, Recombinant Exerts actions like the natural Interferon alfa-2b (Wishart, 2006).	
	Oxytocin Unknown pharmacological action on Oxytocin-neurophysin 1 (Wishart, 2006).	
	Thiamine Binds its transporter SLC19A (Wishart, 2006). To remove.	
	NADH. Remove	*(Continued)*

DrugBank Term	Drug-Target Explanation	Action
Binder	Folic Acid	Activation
	Binds its transporter (Wishart, 2006).	
	Levonorgestrel	
	Synthetic form of the naturally occurring female sex hormone, progesterone (Wishart, 2006).	
	Niacin	
	Precursor of NAD and NADP (Wishart, 2006).	
	Guanabenz	
	α-2 adrenergic agonist (Wishart, 2006).	
	Methoxamine	
	Acting as a pure alpha-1 adrenergic receptor agonist (Wishart, 2006).	
	Potassium Chloride	
	SLC12A1 transporter(Wishart, 2006).	
	Alglucosidase alfa	
	Exogenous source of GAA (Wishart, 2006).	
	Teriparatide	
	Substitute of endogenous PTH (Wishart, 2006).	
	Tesamorelin	
	Stimulates production and release of the endogenous hormone (hGRF) (Wishart, 2006).	
	Cyclosporine	Inhibition
	CAMLG binds to cyclophilin, the complex then inhibits calcineurin (Wishart, 2006).	
	PPIF	Remove
	Unknown pharmacological action (Wishart, 2006). To remove.	
	Chlorpromazine	Inhibition
	Inhibits DRD4 (Wishart, 2006).	
	HTR2A, ORM1, HTR6/7 and HRH4	Remove
	Unknown pharmacological action (Wishart, 2006). To remove.	
Binder	Loxapine	Remove
	Unknown pharmacological action (Wishart, 2006). To remove.	
	Intravenous Immunoglobulin for C3/4A/4B/5	Inhibition
	Inhibits the complement cascade (Wishart, 2006).	

(Continued)

DrugBank Term	Drug-Target Explanation	Action
Binder	Botulinum Toxin Type B Binds to and cleaves the synaptic VAMP, inhibits acetylcholine release at the neuromuscular junction (Wishart, 2006).	Inhibition
	Citalopram Inhibits HRH1 ADRA1A CHRM1 with less affinity than to SLC6A4 (Wishart, 2006)	
	Trimipramine Unknown pharmacological actions (Wishart, 2006). To remove.	Remove
	Metyrosine Inhibits tyrosine hydroxylase (Wishart, 2006).	Inhibition
	Gliclazide ABCC8 blockade of the channels (Wishart, 2006).	
	Halothane Alters the flow of potassium in cells, functions as anesthetic (Wishart, 2006). Too broad.	Remove
	Ciclopirox Acts via chelation of polyvalent metal cations, leading to inactivation of the enzymes which use them (Wishart, 2006).	Inhibition
	Glatiramer Acetate Strong and promiscuous binding to MHC molecules, and consequent competition with various myelin antigens for their presentation to T cells (Wishart, 2006).	
	Canakinumab Neutralizes IL1B (Wishart, 2006).	
	Rilonacept Blocks IL-1 (Wishart, 2006).	
	Cabazitaxel Microtubule inhibitor (Wishart, 2006).	
	Aflibercept Anti-VEGF drug (Wishart, 2006).	
Binder	Denileukin diftitox The diphtheria toxin associated with Ontak then selectively kills the IL-2 bearing cells (Wishart, 2006).	Remove
	Pegademase bovine Unknown pharmacological action (Wishart, 2006).	

(Continued)

DrugBank Term	Drug-Target Explanation	Action
Binder	Muromonab Binds to the T-cell CD3 epsilon chain. Kills CD-3 positive cells by inducing Fc mediated apoptosis, antibody mediated cytotoxicity and complement-dependent cytotoxicity (Wishart, 2006). Too broad an action. L-Phenylalanine Unknown pharmacological action (Wishart, 2006). To remove. L-Tyrosine Unknown pharmacological action (Wishart, 2006). To remove. L-Proline Unknown pharmacological action (Wishart, 2006). To remove. Amphetamine Unknown pharmacological action on SLC6A4 and DR2 (Wishart, 2006). Methysergide Unknown pharmacological action on HTR1B/F/W (Wishart, 2006). Cabergoline Unknown pharmacological action on ADRA1A/1B/1D ADRB1/2 (Wishart, 2006). Atomoxetine Unknown pharmacological action on SLC6A3 Amitriptyline Unknown pharmacological action on HRH4, HTR2C, HTR1D, OPRM1, HTR1B, ADRB1 (Wishart, 2006). To remove. Terfenadine Unknown pharmacological action on CHRM1/2/4/5 (Wishart, 2006). To remove. Norepinephrine Unknown pharmacological action on SLC18A1/2 (Wishart, 2006). To remove. Mirtazapine Unknown pharmacological action on ADRA1A/2C, ADRB1/2/, DRD1/2, SLC6A2/3/4, HTR2B (Wishart, 2006). To remove.	Remove

(Continued)

DrugBank Term	Drug-Target Explanation	Action
Binder	Spironolactone Unknown pharmacological action on SHBG (Wishart, 2006). To remove. Pethidine Unknown pharmacological action on CHRM1 (Wishart, 2006). To remove. Prazosin Unknown pharmacological action on ADRA2A/B (Wishart, 2006). To remove. Imipramine Unknown pharmacological action on DRD1/2 AND HTR6 (Wishart, 2006). To remove. Oxycodone Unknown pharmacological action on ORM1 (Wishart, 2006). To remove. Dextromethorphan Unknown pharmacological action on PGRMC1 (Wishart, 2006). To remove. Nortriptyline Unknown pharmacological action on HTR6, PGRMC1 (Wishart, 2006). To remove. Amoxapine Unknown pharmacological action on SLC6A3, HRH4, GABRA1 Cinnarizine Unknown pharmacological action on DRD1, CHRM1 (Wishart, 2006). To remove. Insulin Unknown pharmacological action on CFT (Wishart, 2006). To remove. Procaine Unknown pharmacological action on KCNMA1 (Wishart, 2006). To remove. Tolazoline Unknown pharmacological action on ADRA2B/C (Wishart, 2006). To remove. Cysteamine Unknown pharmacological action on SST (Wishart, 2006). To remove. Maprotiline Unknown pharmacological action on HTR2A/C and DRD2 (Wishart, 2006). To remove.	Remove

(Continued)

DrugBank Term	Drug-Target Explanation	Action
Binder	Oxymetazoline Unknown pharmacological action on ADRA2B (Wishart, 2006). To remove. Glycopyrrolate Unknown pharmacological action on CHRM2 (Wishart, 2006). To remove. Dopamine Unknown pharmacological action on HTR1A and HTR7 (Wishart, 2006). To remove. Guanfacine Unknown pharmacological action on ADRA2B (Wishart, 2006). To remove. Ketoconazole Unknown pharmacological action on AR (Wishart, 2006). To remove. Thalidomide Unknown pharmacological action on ORM1 (Wishart, 2006). To remove. Memantine Unknown pharmacological action on GRIN1 Ibuprofen Unknown pharmacological action on FABP2 (Wishart, 2006). To remove. Doxepin Unknown pharmacological action on HTR6 HRH4 (Wishart, 2006). To remove. Desipramine Unknown pharmacological action on HTR1A/C DRD2 ADRA2A (Wishart, 2006). To remove. Ketamine Unknown pharmacological action on OPRD1 OPRM1 CHRM1 (Wishart, 2006). To remove. Bepridil Unknown pharmacological action on CALM1 (Wishart, 2006). To remove. Docetaxel Unknown pharmacological action on NR1I2 (Wishart, 2006). To remove. Aluminium Astringent (Wishart, 2006). To remove.	Remove

(Continued)

DrugBank Term	Drug-Target Explanation	Action
Binder	Dehydroepiandrosterone Unknown pharmacological action on ESR1 (Wishart, 2006). To remove. Indirubin-3'-Monoxime Unknown pharmacological action on CDK1 (Wishart, 2006). To remove. Olomoucine Unknown pharmacological action on CDK1 (Wishart, 2006). To remove. SU9516 - Not found Nicotinamide Unknown pharmacological action on PARP1 (Wishart, 2006). To remove. Nesiritide Unknown pharmacological action on NPR1 (Wishart, 2006). To remove. Mianserin Unknown pharmacological action on HRH4/6 HTR2B/1F DRD1/3 SLC6A3 (Wishart, 2006). To remove. Tinzaparin Unknown pharmacological action on CXCL12 (Wishart, 2006). To remove. Hyaluronic acid Unknown pharmacological action on VCAN C1QBP HAPLN1 HAPLN3 HAPLN4 HABP2 LAYN STAB2 TNFAIP6 IMPG2 HABP4 (Wishart, 2006). To remove.	Remove
Binding	Mirtazapine Unknown pharmacological action on HTR7 (Wishart, 2006). To remove. Pethidine Unknown pharmacological action on SLC6A4 (Wishart, 2006). To remove. Dimethyl fumarate Unknown pharmacological action on KEAP1 (Wishart, 2006). To remove. Acetylsalicylic acid Unknown pharmacological action on HSPA5 (Wishart, 2006). To remove.	Remove
Blocker	Inhibits the target's action upon binding.	Inhibition

(Continued)

DrugBank Term	Drug-Target Explanation	Action
Chaperone	Aids in folding, may help stabilize the protein or may tag the protein for degradation. To remove.	Remove
Cleavage	Breakage of peptide bonds.	Inhibition
Cofactor	Necessary component to carry out the effect.	Activation
Desensitize the target	Diminishes the response.	Inhibition
Inducer	Transcriptional activator.	Activation
inhibitor	Decreases the target's activity.	Inhibition
Inhibitor	Decreases the target's activity.	Inhibition
Inhibitor, competitive	Competes with the substrate with the active site of an enzyme.	Inhibition
Inverse agonist	Opposite action of an agonist.	Inhibition
Intercalation	Insertion between planar structures.	Inhibition
Ligand	Binds its target to form an active complex.	Activation
Modulator	Toremifene compete with estrogen for binding sites in the cancer (Wishart, 2006). Rufinamide prolongs the inactive state of voltage gated sodium channels (Wishart, 2006).	Inhibition
	Antihemophilic Factor Unknown pharmacologic action on LRP1 MACFD2 (Wishart, 2006).	Remove
Modulator	Loperamide Unknown pharmacologic action on POMC (Wishart, 2006).	Remove
	Glyburide Unknown pharmacologic action on ABCC8/9 KCNJ11 (Wishart, 2006).	
	Minocycline Unknown pharmacologic action on ILB1 (Wishart, 2006). To remove.	
	Ibuprofen Unknown pharmacologic action on BCL2 THBD (Wishart, 2006). To remove.	
	Carvedilol Unknown pharmacologic action on HIF1A (Wishart, 2006). To remove.	
	Fingolimod Binds with high affinity to sphingosine 1-phosphate receptors 1, 3, 4, and 5. The mechanism in multiple sclerosis may involve reduction of lymphocyte migration into the central nervous system (Wishart, 2006).	

(Continued)

DrugBank Term	Drug-Target Explanation	Action
Multitarget	Drotrecogin alfa	Inhibition
	Inhibits factor Va and VIIIa (Wishart, 2006).	
	Dasatinib	
	Unknown pharmacologic action for SRC, ABLS2, FYN, LCK, ABL1 (Wishart, 2006). To remove.	Inhibition
	Sunitinib	Remove
	Unknown pharmacological action on KDR, FLT3 (Wishart, 2006). To remove	
	Ramelteon	Activation
	Melatonin receptor agonist (Wishart, 2006).	
Negative Modulator	Allosteric negative modulator.	Inhibition
Neutralizer	Inhibits the target's effect.	Inhibition
Other	Lovastatin	Remove
	Unknown pharmacological action on HDAC2 (Wishart, 2006). To remove.	
	Vancomycin	
	Unknown pharmacological action on GFTA (Wishart, 2006). To remove.	
	Simvastatin	Remove
	Unknown pharmacological action on ITGB2 (Wishart, 2006). To remove.	
	Oseltamivir	Activation
	Binds CES1, which converts it to active form (Wishart, 2006).	
	Levonorgestrel	Remove
	The affinity for the ESR1 is very low (Wishart, 2006).	
	Sirolimus binds to FKBP-12 activating its immunosuppressive actions (Wishart, 2006).	Activation
	Riboflavin	Remove
	Binding to ribE on Ecoli. To remove. (Wishart, 2006).	
	Lorazepam	
	Binds the transporter TSPO (Wishart, 2006). To remove.	
	Hydroxocobalamin	
	Unknown pharmacological action on MTRR TCN1 AMN CUBN MMAB (Wishart, 2006). To remove.	
	Temazepam	
	Binds the transporter protein TSPO (Wishart, 2006). To remove	

(Continued)

DrugBank Term	Drug-Target Explanation	Action
Other	Chloramphenicol Unknown pharmacological action on CD55 (Wishart, 2006). To remove.	Remove
	Quinine Unknown pharmacological action on GP9 (Wishart, 2006). To remove.	
	Aminocaproic Acid Unknown pharmacological action on LPA (Wishart, 2006). To remove.	
	Lactulose Binds ebgA , a sugar receptor on EColi (Wishart, 2006). To remove.	
	Diclofenac Unknown pharmacological action on KCNQ2/3 (Wishart, 2006). To remove.	
	Verapamil Unknown pharmacological action on SCN5A (Wishart, 2006). To remove.	
	Sufentanil Unknown pharmacological action on OPRK1 (Wishart, 2006). To remove.	
	Adefovir Dipivoxil Unknown pharmacological action on virus HBV (Wishart, 2006). To remove.	
	Pentamidine Unknown pharmacological action on TRDMT1 (Wishart, 2006). To remove.	
	Etodolac Unknown pharmacological action on RXRA (Wishart, 2006). To remove.	
	Triazolam Binds translocator protein TSPO (Wishart, 2006). To remove.	
	Meclofenamic acid Unknown pharmacological action on KCNQ2/3 (Wishart, 2006). To remove.	
	Zaleplon Binds transporter TSPO (Wishart, 2006). To remove.	
	Ezetimibe Unknown pharmacological action on ANPEP (Wishart, 2006). To remove.	

(Continued)

DrugBank Term	Drug-Target Explanation	Action
Other	Ketoprofen Unknown pharmacological action on CXCR1 (Wishart, 2006). To remove. Felodipine Unknown pharmacological action on CALM1 TNNC2 TNNC1 (Wishart, 2006). To remove. Procainamide Unknown pharmacological action on DNMT1 (Wishart, 2006). To remove. Flucytosine Unknown pharmacological action on DNMT1 (Wishart, 2006). To remove. Diazoxide Unknown pharmacological action on ATP1A1 KCNMA1 (Wishart, 2006). To remove. Carvedilol Unknown pharmacological action on EGFA NPPB GJA1 (Wishart, 2006). To remove. Desipramine Unknown pharmacological action on ADRB1 (Wishart, 2006). To remove. Halothane Unknown pharmacological action on RHO (Wishart, 2006). To remove. Bepridil Unknown pharmacological action on TNNC1 (Wishart, 2006). To remove. Chenodeoxycholic acid Binds NR1H4, suppresses hepatic synthesis of both cholesterol and cholic acid, gradually replacing the latter and its metabolite Chenodeoxycholic acid (Wishart, 2006). To remove.	Remove
Other/unknown	Unknown.	Remove
Partial agonist	Partial activating effect on the target.	Activation
Partial antagonist	Partial inhibiting effect on the target.	Inhibition
Positive allosteric modulator	Binds to a site distinct from that of the orthosteric agonist binding site inducing the activation of the target.	Activation
Positive modulator	Binds to a site distinct from that of the orthosteric agonist binding site inducing the activation of the target.	Activation

(Continued)

DrugBank Term	Drug-Target Explanation	Action
Potentiator	Enhances sensitization.	Activation
Product of	Unknown.	Remove
Stimulator	Excites the functional activity.	Activation
Suppressor	Inhibits the target's effects.	Inhibition
Unknown	Unknown.	Remove

Bibliography

Goodsell, D. S. (2002). The Molecular Perspective: Tamoxifen and the Estrogen Receptor. The Oncologist, 7(2), 163–164. https://doi.org/10.1634/theoncologist.7-2-163

Jelovac, D., & Armstrong, D. K. (2012). Role of farletuzumab in epithelial ovarian carcinoma. Current Pharmaceutical Design, 18(25), 3812–3815.

Koonce, C. J., Walf, A. A., & Frye, C. A. (2009). Trilostane exerts antidepressive effects among wild-type, but not estrogen receptor [beta] knockout mice. Neuroreport, 20(12), 1047–1050.

Wishart, D. S. (2006). DrugBank: a comprehensive resource for in silico drug discovery and exploration. Nucleic Acids Research, 34(90001), D668–D672. https://doi.org/10.1093/nar/gkj067

Pathway Activity Analysis (PAA) as a new class of mechanistic biomarker to predict drug responses in drug repositioning for cancer patients

---

ANNEX 3. R Scripts

Index

Script 1. Functions

```
# FUNCTIONS TO BE SOURCED

# Read experimental data
read.expression.matrix <- function(file){
    data.matrix(read.table(file, header=T, sep="\t", stringsAsFactors=F, row.names=1,
comment.char=""))
}

read.experimental.design <- function(file){
    des <- read.table(file,header=F,stringsAsFactors=F,row.names=1)
    colnames(des)[1] <- c("Condition")
    return(des)
}


# Functional analysis function
prettyfuns.average <- function(results, pathigraphs, dbannot, entrez2hgnc){

    # Calculates a value per function by estimating the average of the values obtained in
those pathways that lead to that function
    annofuns <- do.call("rbind", lapply(pathigraphs,function(pathigraph){
        new.pathigraph <- pathigraph
        new.pathigraph$graph <- induced.subgraph(new.pathigraph$graph,
V(new.pathigraph$graph)$name[!grepl("_func", V(new.pathigraph$graph)$name)])
        funs <- get.pathway.functions(new.pathigraph, dbannot, entrez2hgnc,
use.last.nodes=T)
        paths <- lapply(names(funs), function(path) rep(paste0(pathigraph$path.id, "__",
path), times=length(funs[[path]])))
        df <- data.frame(paths = unlist(paths), funs=unlist(funs), stringsAsFactors=F)
    }))

    annofuns <- annofuns[!is.na(annofuns$funs),]
    fun.names <- unique(annofuns$funs)
    fun.vals <- matrix(0, ncol=ncol(results$all$effector.path.vals), nrow =
length(fun.names), dimnames = list(fun.names, colnames(results$all$effector.path.vals)) )

    # By using the mean value, saturation of the signal is avoided
    for( fun in fun.names){
        print(fun)
        paths <- annofuns$paths[annofuns$funs==fun]
        minimat <- results$all$effector.path.vals[paths,,drop=F]
        fun.vals[fun,] <- apply(minimat, 2, mean)

    }
    print(dim(fun.vals))
    return(fun.vals)
}
```

Script 2. Drugs into Data Frame

```
# Drug Data comes in a complex list that needs to be organized
# Drugs that do not meet the criteria are removed: have targets and actions
# Actions are categorized as “activation” or “inhibition”

drugs.into.data.frame <- function(data){
# Selects drugs for which targets and actions are known
# Returns data.frame

drugs <- names(data)
    d <- data.frame(Drug = NA, Targets = NA, Actions = NA)

    for(i in 1:length(drugs)){
        targets <- names(data[[i]]$targets)
        if(is.character(targets)){
            drug <- drugs[i]
            for(j in 1:length(targets)){
                target_name <- data[[i]]$targets[[j]]$geneName
                drug_action <- data[[i]]$targets[[j]]$actions
                lista <- c(Drug = drug, Targets = target_name, Actions = drug_action)
                d <- rbind(d, lista)
            }
        }
    }


    # Remove NA and empty strings
    d <- d[complete.cases(d),]
    d <- d[!apply(d, 1, function(x) any(x == "")),]


    # See all posible actions drugs have to decide if they activate or not
    Actions <- sort(unique(d$Actions))


    # Manually curated activating and inhibition actions
    activatingActions <- c("acetylation", "activator", "agonist", "agonistimodulator",
                           "agonistpartial agonist", "cofactor", "inducer", "ligand",
                           "partial agonist", "positive allosteric modulator",
                           "positive modulator", "potentiator", "stimulator" )

    inhibitingActions <- c("adduct", "agonistinhibitor", "antagonist",
                           "antagonistantibody", "antagonistbinder", "antagonistblocker",
                           "antagonistinhibitor", "antagonistInhibitor",
                           "antagonistinhibitor, competitive",
                           "antagonistinhibitory allosteric modulator",
"antagonistmultitarget",
                           "antagonistother/unknown", "antibody","blocker", "cleavage",
                           "desensitize the target", "inhibitor", "Inhibitor",
                           "inhibitor, competitive", "inverse agonist", "intercalation",
                           "negative modulator", "neutralizer", "partial antagonist",
"suppressor")


    # Some actions need to be removed
    actionsToDelete <- c("acetylation","other/unknown", "product of", "unknown", "binding",
"chaperone")
    d <- subset(d, !d$Actions %in% actionsToDelete )


    # Replace actions on the data.frame to inhibiting or activating
```

```
activatingRows <- d$Actions %in% activatingActions
d$Actions[activatingRows] <- "activation"

inhibitingRows <- d$Actions %in% inhibitingActions
d$Actions[inhibitingRows] <- "inhibition"


# Special Cases that need to be removed
d <- d[!(d$Actions == "allosteric modulator" & d$Drug == "Vardenafil"),]
d <-  d[!(d$Actions == "allosteric modulator" & d$Drug == "Halothane"),]
d <-  d[!(d$Actions == "antagonistagonist" & d$Drug == "Risperidone"),]
d <-  d[!(d$Actions == "antagonistagonist" & d$Drug == "Tamoxifen"),]
d <-  d[!(d$Actions == "antagonistagonist" & d$Drug == "Clomifene"),]
d <-  d[!(d$Actions == "antagonistagonist" & d$Drug == "Ospemifene"),]
d <-  d[!(d$Actions == "antagonistpartial agonist" & d$Drug == "Mivacurium"),]
d <-  d[!(d$Actions == "agonistpartial agonist" & d$Drug == "Ergotamine"),]
d <-  d[!(d$Actions == "agonistpartial agonist" & d$Drug == "Pseudoephedrine"),]
d <-  d[!(d$Actions == "agonistpartial agonist" & d$Drug == "Ketamine"),]
d <-  d[!(d$Actions == "antibody" & d$Drug == "Trastuzumab"),]
d <-  d[!(d$Actions == "antibody" & d$Drug == "Rituximab"),]
d <-  d[!(d$Actions == "antibody" & d$Drug == "Ibritumomab"),]
d <-  d[!(d$Actions == "antibody" & d$Drug == "Tositumomab"),]
d <-  d[!(d$Actions == "antibody" & d$Drug == "Alemtuzumab"),]
d <-  d[!(d$Actions == "antibody" & d$Drug == "Methyl aminolevulinate"),]
d <-  d[!(d$Actions == "antibody" & d$Drug == "ado-trastuzumab emtansine"),]
d <-  d[!(d$Actions == "antibody" & d$Drug == "Obinutuzumab"),]
d <-  d[!(d$Actions == "adduct" & d$Drug == "Ethionamide"),]
d <-  d[!(d$Actions == "adduct" & d$Drug == "Isoniazid"),]
d <-  d[!(d$Actions == "adduct" & d$Drug == "Gentamicin"),]
d <-  d[!(d$Actions == "adduct" & d$Drug == "Tigecycline "),]
d <-  d[!(d$Targets == "DRD1" & d$Drug == "Aripiprazole"),]
d <-  d[!(d$Targets == "DRD5" & d$Drug == "Aripiprazole"),]
d <-  d[!(d$Targets == "DRD3" & d$Drug == "Aripiprazole"),]
d <-  d[!(d$Targets == "DRD4" & d$Drug == "Aripiprazole"),]
d <-  d[!(d$Targets == "ATP1A1" & d$Drug == "Aripiprazole"),]
d <-  d[!(d$Targets == "ASGR2" & d$Drug == "Antihemophilic Factor"),]
d <-  d[!(d$Actions == "binder" & d$Drug == "Oxytocin"),]
d <-  d[!(d$Drug == "NADH"),]
d <-  d[!(d$Actions == "binder" & d$Drug == "Niacin"),]
d <-  d[!(d$Targets == "PPIF" & d$Drug == "Cyclosporine"),]
d <-  d[!(d$Targets == "HTR2C" & d$Drug == "Chlorpromazine"),]
d <-  d[!(d$Targets == "HTR2A" & d$Drug == "Chlorpromazine"),]
d <-  d[!(d$Targets == "ORM1" & d$Drug == "Chlorpromazine"),]
d <-  d[!(d$Targets == "HTR6" & d$Drug == "Chlorpromazine"),]
d <-  d[!(d$Targets == "HTR7" & d$Drug == "Chlorpromazine"),]
d <-  d[!(d$Targets == "HRH4" & d$Drug == "Chlorpromazine"),]
d <-  d[!(d$Actions == "binder" & d$Drug == "Loxapine"),]
d <-  d[!(d$Actions == "binder" & d$Drug == "Trimipramine"),]
d <-  d[!(d$Actions == "binder" & d$Drug == "Thiamine"),]
d <-  d[!(d$Actions == "binder" & d$Drug == "Folic Acid"),]
d <-  d[!(d$Drug == "Metyrosine"),]
d <-  d[!(d$Actions == "binder" & d$Drug == "Halothane"),]
d <-  d[!(d$Actions == "binder" & d$Drug == "Pegademase bovine"),]
d <-  d[!(d$Drug == "L-Phenylalanine"),]
d <-  d[!(d$Drug == "L-Tyrosine"),]
d <-  d[!(d$Drug == "L-Proline"),]
d <-  d[!(d$Targets == "SLC6A4" & d$Drug == "Amphetamine"),]
d <-  d[!(d$Targets == "DRD2" & d$Drug == "Amphetamine"),]
d <-  d[!(d$Actions == "binder" & d$Drug == "Methysergide"),]
d <-  d[!(d$Actions == "binder" & d$Drug == "Cabergoline"),]
d <-  d[!(d$Actions == "binder" & d$Drug == "Atomoxetine"),]
d <-  d[!(d$Actions == "binder" & d$Drug == "Amitriptyline"),]
d <-  d[!(d$Actions == "binder" & d$Drug == "Terfenadine"),]
d <-  d[!(d$Actions == "binder" & d$Drug == "Norepinephrine"),]
```

```
d <-  d[!(d$Actions == "binder" & d$Drug == "Mirtazapine"),]
d <-  d[!(d$Actions == "binder" & d$Drug == "Spironolactone"),]
d <-  d[!(d$Actions == "binder" & d$Drug == "Pethidine"),]
d <-  d[!(d$Actions == "binder" & d$Drug == "Prazosin"),]
d <-  d[!(d$Actions == "binder" & d$Drug == "Imipramine"),]
d <-  d[!(d$Actions == "binder" & d$Drug == "Oxycodone"),]
d <-  d[!(d$Actions == "binder" & d$Drug == "Dextromethorphan"),]
d <-  d[!(d$Actions == "binder" & d$Drug == "Nortriptyline"),]
d <-  d[!(d$Actions == "binder" & d$Drug == "Amoxapine"),]
d <-  d[!(d$Actions == "binder" & d$Drug == "Cinnarizine"),]
d <-  d[!(d$Actions == "binder" & d$Drug == "Inulin"),]
d <-  d[!(d$Actions == "binder" & d$Drug == "Procaine"),]
d <-  d[!(d$Actions == "binder" & d$Drug == "Tolazoline"),]
d <-  d[!(d$Actions == "binder" & d$Drug == "Cysteamine"),]
d <-  d[!(d$Actions == "binder" & d$Drug == "Maprotiline"),]
d <-  d[!(d$Actions == "binder" & d$Drug == "Oxymetazoline"),]
d <-  d[!(d$Actions == "binder" & d$Drug == "Glycopyrrolate"),]
d <-  d[!(d$Actions == "binder" & d$Drug == "Dopamine"),]
d <-  d[!(d$Actions == "binder" & d$Drug == "Guanfacine"),]
d <-  d[!(d$Actions == "binder" & d$Drug == "Ketoconazole"),]
d <-  d[!(d$Actions == "binder" & d$Drug == "Thalidomide"),]
d <-  d[!(d$Actions == "binder" & d$Drug == "Memantine"),]
d <-  d[!(d$Actions == "binder" & d$Drug == "Ibuprofen"),]
d <-  d[!(d$Actions == "binder" & d$Drug == "Doxepin"),]
d <-  d[!(d$Actions == "binder" & d$Drug == "Desipramine"),]
d <-  d[!(d$Actions == "binder" & d$Drug == "Ketamine"),]
d <-  d[!(d$Actions == "binder" & d$Drug == "Bepridil"),]
d <-  d[!(d$Actions == "binder" & d$Drug == "Docetaxel"),]
d <-  d[!(d$Actions == "binder" & d$Drug == "Aluminium"),]
d <-  d[!(d$Actions == "binder" & d$Drug == "Almitrine"),]
d <-  d[!(d$Actions == "binder" & d$Drug == "Dehydroepiandrosterone"),]
d <-  d[!(d$Actions == "binder" & d$Drug == "Indirubin-3'-Monoxime"),]
d <-  d[!(d$Actions == "binder" & d$Drug == "Olomoucine"),]
d <-  d[!(d$Actions == "binder" & d$Drug == "Nicotinamide"),]
d <-  d[!(d$Actions == "binder" & d$Drug == "Brentuximab vedotin"),]
d <-  d[!(d$Drug == "SU9516"),]
d <-  d[!(d$Actions == "binder" & d$Drug == "Nesiritide"),]
d <-  d[!(d$Actions == "binder" & d$Drug == "Mianserin"),]
d <-  d[!(d$Actions == "binder" & d$Drug == "Tinzaparin"),]
d <-  d[!(d$Actions == "binder" & d$Drug == "Hyaluronic acid"),]
d <-  d[!(d$Actions == "binder" & d$Drug == "Dimethyl fumarate"),]
d <-  d[!(d$Actions == "binder" & d$Drug == "Muromonab"),]
d <-  d[!(d$Actions == "binder" & d$Drug == "Denileukin diftitox"),]
d <-  d[!(d$Actions == "modulator" & d$Drug == "Antihemophilic Factor"),]
d <-  d[!(d$Actions == "modulator" & d$Drug == "Loperamide"),]
d <-  d[!(d$Actions == "modulator" & d$Drug == "Glyburide"),]
d <-  d[!(d$Actions == "modulator" & d$Drug == "Minocycline"),]
d <-  d[!(d$Actions == "modulator" & d$Drug == "Ibuprofen"),]
d <-  d[!(d$Actions == "modulator" & d$Drug == "Carvedilol"),]
d <-  d[!(d$Actions == "modulator" & d$Drug == "Fingolimod"),]
d <-  d[!(d$Actions == "multitarget" & d$Drug == "Dasatinib"),]
d <-  d[!(d$Actions == "multitarget" & d$Drug == "Sunitinib"),]
d <-  d[!(d$Actions == "other" & d$Drug == "Lovastatin"),]
d <-  d[!(d$Actions == "other" & d$Drug == "Vancomycin"),]
d <-  d[!(d$Actions == "other" & d$Drug == "Simvastatin"),]
d <-  d[!(d$Actions == "other" & d$Drug == "Levonorgestrel"),]
d <-  d[!(d$Actions == "other" & d$Drug == "Sirolimus"),]
d <-  d[!(d$Actions == "other" & d$Drug == "Riboflavin"),]
d <-  d[!(d$Actions == "other" & d$Drug == "Lorazepam"),]
d <-  d[!(d$Actions == "other" & d$Drug == "Hydroxocobalamin"),]
d <-  d[!(d$Actions == "other" & d$Drug == "Temazepam"),]
d <-  d[!(d$Actions == "other" & d$Drug == "Chloramphenicol"),]
d <-  d[!(d$Actions == "other" & d$Drug == "Quinine"),]
d <-  d[!(d$Actions == "other" & d$Drug == "Aminocaproic Acid"),]
```

```
    d <-  d[!(d$Actions == "other" & d$Drug == "Lactulose"),]
    d <-  d[!(d$Actions == "other" & d$Drug == "Diclofenac"),]
    d <-  d[!(d$Actions == "other" & d$Drug == "Verapamil"),]
    d <-  d[!(d$Actions == "other" & d$Drug == "Sufentanil"),]
    d <-  d[!(d$Actions == "other" & d$Drug == "Adefovir"),]
    d <-  d[!(d$Actions == "other" & d$Drug == "Adefovir Dipivoxil"),]
    d <-  d[!(d$Actions == "other" & d$Drug == "Pentamidine"),]
    d <-  d[!(d$Actions == "other" & d$Drug == "Etodolac"),]
    d <-  d[!(d$Actions == "other" & d$Drug == "Triazolam"),]
    d <-  d[!(d$Actions == "other" & d$Drug == "Meclofenamic acid"),]
    d <-  d[!(d$Actions == "other" & d$Drug == "Zaleplon"),]
    d <-  d[!(d$Actions == "other" & d$Drug == "Ezetimibe"),]
    d <-  d[!(d$Actions == "other" & d$Drug == "Ketoprofen"),]
    d <-  d[!(d$Actions == "other" & d$Drug == "Felodipine"),]
    d <-  d[!(d$Actions == "other" & d$Drug == "Procainamide"),]
    d <-  d[!(d$Actions == "other" & d$Drug == "Flucytosine"),]
    d <-  d[!(d$Actions == "other" & d$Drug == "Diazoxide"),]
    d <-  d[!(d$Actions == "other" & d$Drug == "Carvedilol"),]
    d <-  d[!(d$Actions == "other" & d$Drug == "Desipramine"),]
    d <-  d[!(d$Actions == "other" & d$Drug == "Halothane"),]
    d <-  d[!(d$Actions == "other" & d$Drug == "Bepridil"),]
    d <-  d[!(d$Actions == "other" & d$Drug == "Chenodeoxycholic acid"),]
    d <-  d[!(d$Actions == "other" & d$Drug == "Iron Dextran"),]


    #  Special Cases that need to be modified
    d <- within(d, Actions[ Drug == "Galantamine" & Actions == "allosteric modulator" ] <-
"inhibition")
    d <- within(d, Actions[ Drug == "Trilostane" & Actions == "allosteric modulator" ] <-
"activation")
    d <- within(d, Actions[ Drug == "Aldesleukin" & Actions == "agonistmodulator" ] <-
"activation")
    d <- within(d, Actions[ Drug == "Carglumic Acid" & Actions == "allosteric modulator" ]
<- "activation")
    d <- within(d, Actions[ Drug == "Quinestrol" & Actions == "agonistmodulator" ] <-
"activation")
    d <- within(d, Actions[ Drug == "Olanzapine" & Actions == "antagonistagonist" ] <-
"inhibition")
    d <- within(d, Actions[ Drug == "Ergoloid mesylate" & Actions == "antagonistagonist" ]
<- "activation")
    d <- within(d, Actions[ Drug == "Tegaserod" & Actions == "antagonistpartial agonist" ]
<- "activation")
    d <- within(d, Actions[ Drug == "Aripiprazole" & Targets == "DRD2" ] <- "inhibition")
    d <- within(d, Actions[ Drug == "Aripiprazole" & Targets == "HTR1A" ] <- "inhibition")
    d <- within(d, Actions[ Drug == "Penbutolol" & Actions == "antagonistpartial agonist" ]
<- "inhibition")
    d <- within(d, Actions[ Drug == "Menotropins"] <- "activation")
    d <- within(d, Actions[ Drug == "Interferon gamma-1b"] <- "activation")
    d <- within(d, Actions[ Drug == "Palifermin"] <- "activation")
    d <- within(d, Actions[ Drug == "Somatropin recombinant"] <- "activation")
    d <- within(d, Actions[ Drug == "Interferon alfacon-1"] <- "activation")
    d <- within(d, Actions[ Drug == "Insulin, porcine"] <- "activation")
    d <- within(d, Actions[ Drug == "Choriogonadotropin alfa"] <- "activation")
    d <- within(d, Actions[ Drug == "Octreotide"] <- "activation")
    d <- within(d, Actions[ Drug == "Interferon Alfa-2b, Recombinant"] <- "activation")
    d <- within(d, Actions[ Drug == "Thiamine" & Actions == "binder"] <- "activation")
    d <- within(d, Actions[ Drug == "Antihemophilic Factor" & Targets == "VWF"] <-
"activation")
    d <- within(d, Actions[ Drug == "Folic Acid" & Actions == "binder"] <- "activation")
    d <- within(d, Actions[ Drug == "Levonorgestrel" & Actions == "binder"] <-
"activation")
    d <- within(d, Actions[ Drug == "Guanabenz" & Actions == "binder"] <- "activation")
    d <- within(d, Actions[ Drug == "Methoxamine" & Actions == "binder"] <- "activation")
    d <- within(d, Actions[ Drug == "Potassium Chloride"] <- "activation")
```

```
    d <- within(d, Actions[ Drug == "Alglucosidase alfa"] <- "activation")
    d <- within(d, Actions[ Drug == "Teriparatide"] <- "activation")
    d <- within(d, Actions[ Drug == "Tesamorelin"] <- "activation")
    d <- within(d, Actions[ Drug == "Cyclosporine" & Actions == "binder"] <- "inhibition")
    d <- within(d, Actions[ Drug == "Chlorpromazine" & Actions == "binder"] <-
"inhibition")
    d <- within(d, Actions[ Drug == "Intravenous Immunoglobulin" & Actions == "binder"] <-
"inhibition")
    d <- within(d, Actions[ Drug == "Botulinum Toxin Type B" ] <- "inhibition")
    d <- within(d, Actions[ Drug == "Citalopram"] <- "inhibition")
    d <- within(d, Actions[ Drug == "Gliclazide"] <- "inhibition")
    d <- within(d, Actions[ Drug == "Ciclopirox"] <- "inhibition")
    d <- within(d, Actions[ Drug == "Glatiramer Acetate"] <- "inhibition")
    d <- within(d, Actions[ Drug == "Canakinumab"] <- "inhibition")
    d <- within(d, Actions[ Drug == "Rilonacept"] <- "inhibition")
    d <- within(d, Actions[ Drug == "Cabazitaxel"] <- "inhibition")
    d <- within(d, Actions[ Drug == "Aflibercept"] <- "inhibition")
    d <- within(d, Actions[ Drug == "Toremifene"] <- "inhibition")
    d <- within(d, Actions[ Drug == "Rufinamide" & Actions == "modulator"] <- "inhibition")
    d <- within(d, Actions[ Drug == "Drotrecogin alfa"] <- "inhibition")
    d <- within(d, Actions[ Drug == "Ramelteon"] <- "activation")
    d <- within(d, Actions[ Drug == "Oseltamivir" & Actions == "other"] <- "activation")
    d <- within(d, Actions[ Drug == "Sirolimus" & Actions == "other"] <- "activation")
    }
```

Script 3. Functions for the Drug Analysis

```
# MODIFY EXP
drug.exp <- function(drug, expD = expT, effect = drugEffect){

    # Drug targets in xref
    affected_genes <- effect[which(effect$Drug == drug),]$ID

    # Drug effect on targets
    effect_on_genes <- effect[which(effect$Drug == drug),]$Actions

    # Copies the original exp for tumors and adds the drug name to the column name
    colnames(expD) <- paste(unique(drugEffect$DrugCode[drugEffect$Drug == drug]),
colnames(expD), sep = "_")


    # Calculate new values for the affected genes
    iteration <- 1
    for(gen in affected_genes){

        action <- effect_on_genes[iteration]

        if(action == "inhibition"){expD[rownames(expD) %in% gen, ] <-
expD[rownames(expD) %in% gen, ]*0.001}
        else if(action == "activation"){expD[rownames(expD) %in% gen, ] <- 0.99}

        iteration <- iteration +1
    }

    # Returns a modified exp after applying a drug
    return(expD)
}


# MODIFY DES
drug.des <- function(drug, desD = desT, effect = drugEffect){

    d <- unique(effect$DrugCode[effect$Drug == drug])
    # Condition is drug name
    desD$Condition <- d

    # rows maintain name and add drug as a prefix
    rownames(desD) <- paste(unique(effect$DrugCode[effect$Drug == drug]), rownames(desD),
sep = "_")

    # Returns a modified des after applying a drug
    return(desD)
```

Script 4. Code to Run the Program

```
# ORIGINAL CODE NEEDS TO BE SOURCED (GitHub HiPathia repository)
# igraph version 0.7.0.
library("igraph")
source("prettyways.R")
source("stats.R")
source("utils.R")
source("functions.r")

# Source developed code
source("drugs.into.data.frame.R")
source("functions.drug.analysis.R")
source("drug.statistics.R")
source("other.functions.R")

# OBTAIN DATA
# Load graphs and xref
load("files/meta_graph_info_hsa.RData")
for(pw in names(metaginfo$pathigraphs)){
    metaginfo$pathigraphs[[pw]]$graph
   for(eff in names(metaginfo$pathigraphs[[pw]]$effector.subgraphs)){
       metaginfo$pathigraphs[[pw]]$effector.subgraphs[[eff]]
    }
}

load("files/xref/hsa/xref.rdata")

# Load experimental data
des <- read.experimental.design("KIRC_des.txt")
exp <-
read.expression.matrix("exp_KIRC__hiseq_data_combat_WO_batch_corrected_by_gcc_and_plate.txt
")

# Load drug data
load("drugList.RData")
drugEffect <- drugs.into.data.frame(drugList)


# PREPARE EXPERIMENTAL DATA
exp <- translate.matrix(exp,xref)
exp <- normalize.data(exp, by.quantiles=F, by.gene=F, percentil=F)
exp <- add.missing.genes(exp, genes=metaginfo$all.genes)


# PREPARE DRUG DATA
# Eliminate targets that are not in the experimental data
toRemove <- setdiff(unique(ID), rownames(exp))
drugEffect <- drugEffect[which( !drugEffect$ID %in% toRemove), ]

# Add column with targets translated to xref
drugEffect$ID <- unname(translate.ids(drugEffect$Targets, xref)$translation)

# Set unique code for each drug
drugEffect <- transform(drugEffect, DrugCode = match(Drug, unique(Drug)))
drugEffect$DrugCode <- sprintf('D%i', drugEffect$DrugCode)


# EXTEND EXPERIMENTAL DATA WITH DRUG MODIFICATIONS TO PATIENTS
# File with the drugs to be tested
dat <- readLines("drugs_to_test.txt")
```

```r
# Copy the original data for the tumor conditions to modify
expT <- exp[,which(des$Condition == "Tumor")[1]:nrow(des)]
desT <- des[des$Condition %in% c("Tumor"),1, drop = FALSE]

# Create a new modified matrix per drug and add it to the original data
for(d in dat){
    # For each drug given creates a new matrix modifying the original tumor data
    new_expD <- drug.exp(drug = d)

    # Append the new matrix to the full matrix
    exp <- cbind2(exp, new_expD)

    # Same for des
    new_desD <- drug.des(drug = d)
    des <- rbind2(des, new_desD)
}


# PATHWAY ANALYSIS
pathway_data <- prettyways(exp, metaginfo$pathigraphs, verbose=T)


# FUNCTIONAL ANALYSIS
# Load the manual annotations file
entrez2hgnc <-
read.table("files/annotations/hsa/entrez_hgnc_hsa.annot",header=F,sep="\t",stringsAsFactors
=F)
annot <- load.annot.file("files/annotations/hsa/annot_manual_hsa.annot")
annot <- unique(annot)

# Pathways into Functions
functions <- prettyfuns.average(pathway_data, metaginfo$pathigraphs, annot, entrez2hgnc)
```

Script 5. Statistics and Output

```
# wilcox function
wilcox.function <- function(A, B, c){
    column <- data.frame(ncol = NA)
    colnames(column) <- c
    for(i in 1:nrow(A)){
        column <- rbind2(column, wilcox.test(A[i,], B[i,])$p.value)}
    column <- na.omit(column)
    rownames(column) <- rownames(A)
    return(column)
}


# Classification function
drug.type.per.function <- function(mD, mN, mT, df){
    # Per function it estimates the type of drug it is
    for(i in 1:nrow(df)){
        if(df$SignificantDN[i] == "N"){df$Classification[i] <- c("Optimum")}
        if(df$SignificantDN[i] == "Y"){
            if(mN[i,] < mT[i,]){
                if(mD[i,] < mN[i,]){df$Classification[i] <-c("Overdose")}
                else if(mN[i,] < mD[i,] & mD[i,] < mT[i,]){df$Classification[i] <-
c("Underdose")}
                else if(mT[i,] < mD[i,]){df$Classification[i] <-c("Undesired")}
            }
            if(mN[i,] > mT[i,]){
                if(mD[i,] > mN[i,]){df$Classification[i] <-c("Overdose")}
                else if(mN[i,] > mD[i,] & mD[i,] > mT[i,]){df$Classification[i] <-
c("Underdose")}
                else if(mT[i,] > mD[i,]){df$Classification[i] <-c("Undesired")}
            }
        }
    }

    return(df)
}


# TUMOR & NORMAL ANALYSIS
Tumor <- functions[,which(des$Condition == "Tumor")]
Normal <- functions[,which(des$Condition == "Normal")]

# Functions significantly different are set as Y
NT <- wilcox.function(Normal, Tumor, c = "pvalue_NT")
NT$SignificantNT[NT$pvalue <= 0.05]<- c("Y")
NT$SignificantNT[NT$pvalue > 0.05] <- c("N")


# DRUGS ANALYSIS
Conditions <- unique(des$Condition)
Conditions <- setdiff(Conditions, c("Tumor", "Normal"))

# All_Drugs_resume will oontain pvalues for all drugs compared
All_Drugs_Resume <- NT

for(Condition in Conditions){

    dir.create(Condition)

    D <- functions[, which(des$Condition == Condition)]
```

```
    # 1. COMPARISONS
    # Wilcoxt test Drug vs Tumor
    DT <- wilcox.function(D, Tumor, c = "pvalue_DT")
    DT$SignificantDT[DT$pvalue <= 0.05]<- c("Y")
    DT$SignificantDT[DT$pvalue > 0.05] <- c("N")

    # Wilcox test Drug vs Normal
    DN <- wilcox.function(D, Normal, c = "pvalue_DN")
    DN$SignificantDN[DN$pvalue <= 0.05]<- c("Y")
    DN$SignificantDN[DN$pvalue > 0.05] <- c("N")

    # Combine the information in aResume file of the drug for the folder
    # Combine the Resume with the global Resume for all Drugs
    Resume <- cbind2(NT, DT)
    Resume <- cbind2(Resume, DN)
    All_Drugs_Resume <- cbind2(All_Drugs_Resume,DT)
    All_Drugs_Resume <- cbind2(All_Drugs_Resume, DN)

    # Significant functions are those different between NT and DT
    Significant <- Resumen[Resumen$pvalue_NT <= 0.05,]
    Significant <- Significant[Significant$pvalue_DT <= 0.05,]

    # 2. HOW IS THE DIFFERENCE WITH THE NORMAL?
    # Get the average per function for those tumoral functions
    # Original data reduced to siginificant for TN and DT comparisons
    Tum <- Tumor[which(NT[,1] <= 0.05 & DT[,1] <= 0.05),]
    Nor <- Normal[which(NT[,1] <= 0.05 & DT[,1] <= 0.05),]
    D <- D[which(NT[,1] <= 0.05 & DT[,1] <= 0.05),]

    # The means for N and T data and D per function:
    meansN <- data.frame(unname(rowMeans(Nor, na.rm = FALSE, dims = 1)))
    rownames(meansN) <- rownames(Nor)
    colnames(meansN) <- c("meansN")

    meansT <- data.frame(unname(rowMeans(Tum, na.rm = FALSE, dims = 1)))
    rownames(meansN) <- rownames(Tum)
    colnames(meansN) <- c("meansT")

    meansD <- data.frame(unname(rowMeans(D, na.rm = FALSE, dims = 1)))
    rownames(meansD) <- rownames(D)
    colnames(meansD) <- Condition

    DrugTypeData <- DN[which(NT[,1] <= 0.05 & DT[,1] <= 0.05),]
    DrugType <- drug.type.per.function(mD = meansD, mN = meansN, mT=meansT, df =
DrugTypeData)

    # 3. WRITE THE OUTPUT
    write.csv(Resume, 'Resume.csv', row.names=T)
    write.csv(Significant, 'Significant.csv', row.names=T)
    write.csv(DrugType, 'DrugType.csv', row.names=T)

    setwd('Prettyways')

    }


write.csv(Resume, 'All_Drugs_Resume.csv', row.names=T)


# Get Plot for Normal and Tumor
m <- cbind(Normal,Tumor)
m <- t(m)
m <- data.frame(m)
m$Condition <- des$Condition
```

```
#### PLOTS

# Load Libraries
library(reshape)
library("plotly")

m <- melt(m, id.var = "Condition" )
X <- list(title = "")
Plot<- m, x = ~variable, y = ~value, color = ~Condition, colors = c("red", "blue"), type =
"box", opacity = 0.4) %>% layout(xaxis = ax)
```