Precision medicine in the era of big data and artificial intelligence

Joaquín Dopazo

Clinical Bioinformatics research Area,
Fundación Progreso y Salud,

Computational Systems Biomedicine Group,
Institute of Biomedicine of Sevilla

Functional Genomics Node, (INB-ELIXIR-es),

Bioinformatics in Rare Diseases (BiER-CIBERER),

Sevilla, Spain.

http://www.clinbioinfosspa.es
https://www.ibis-sevilla.es/
http://www. babelomics.org
@xdopazo

IV Congreso Nacional de Jóvenes Investigadores en Biomedicina, Virtual, 4 Noviembre, 2020

Setting the problem in context:

The transition to precision medicine

Degree of personalization

Intuitive Based on trial and error	Identification of probabilistic patterns	Decisions and actions based on knowledge
Intuitive Medicine	**Empirical Medicine**	**Systems Medicine**

Today

Tomorrow

Molecular biomarkers

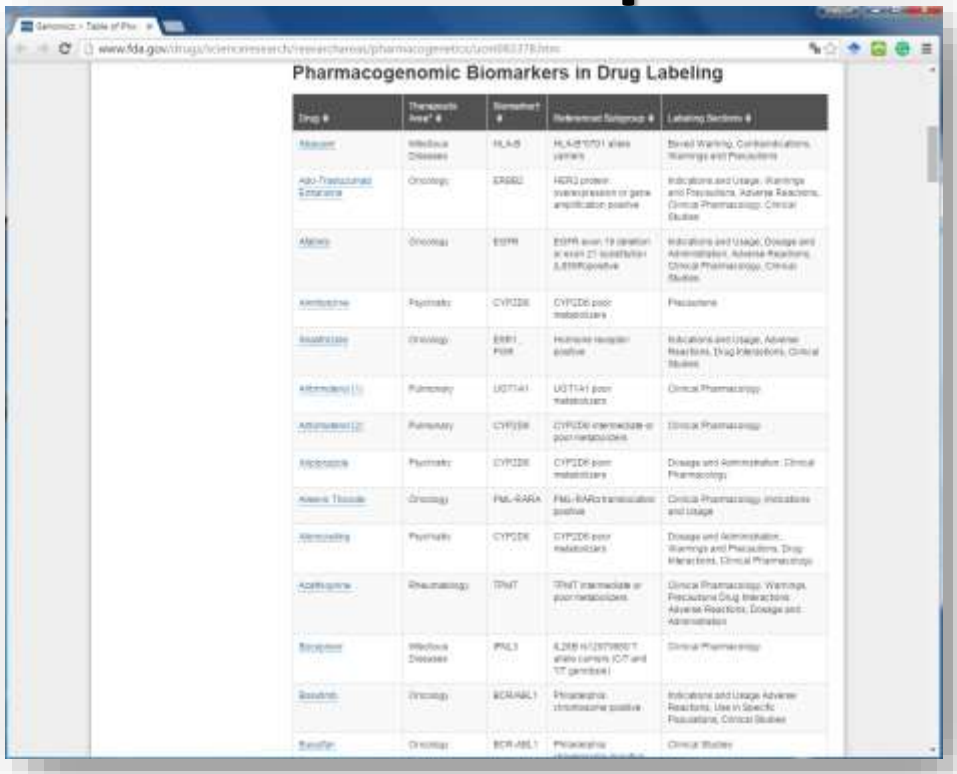
Genomic biomarkers

Mechanistic biomarkers

Precision medicine is based on a better knowledge of phenotype-genotype relationships. This ultimately involves the knowledge of **disease** and **drug action mechanisms**

Requires of a better way of defining diseases by introducing **genomic** technologies in the **diagnostic** procedures and **treatment decisions**

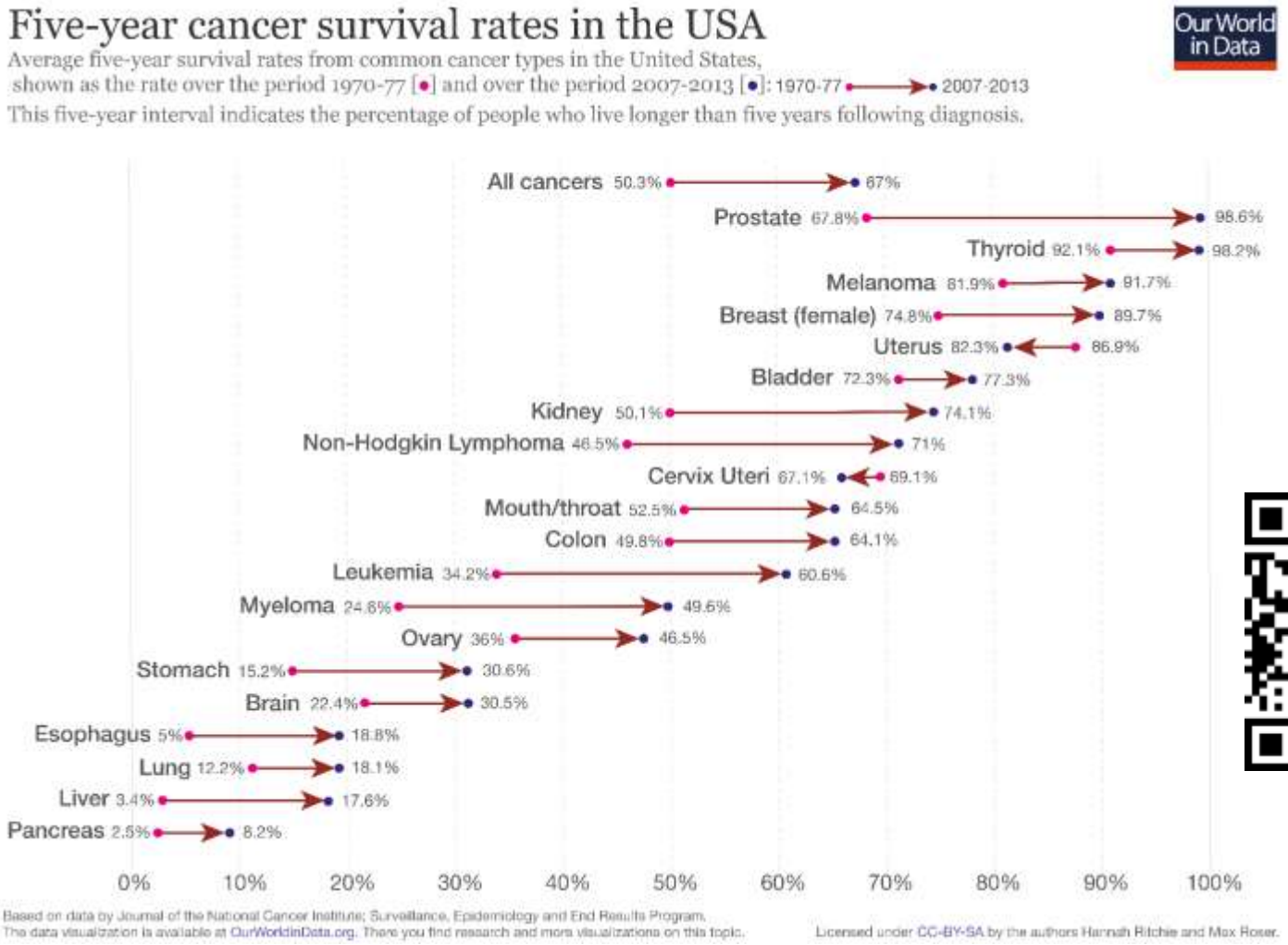
Single-gene biomarkers are the result of probabilistic associations and have a clear clinical impact



http://www.fda.gov/drugs/scienceresearch/researchareas/pharmacogenetics/ucm083378.htm

Most “personalized” therapies are based on this type of biomarkers

Changes in five-year survival rate by using single-gene biomarkers

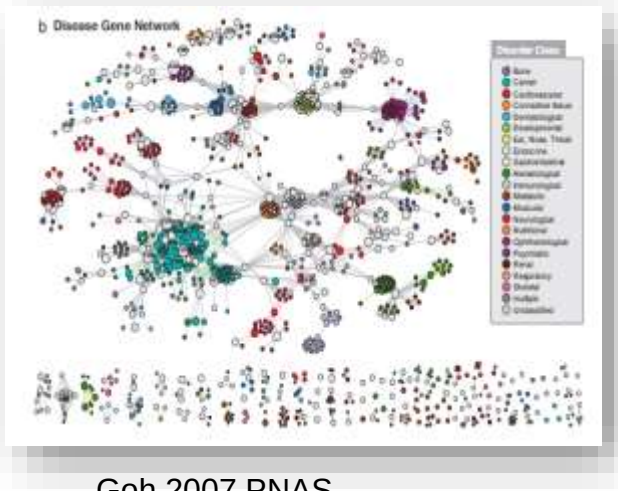


https://ourworldindata.org/cancer-death-rates-are-falling-five-year-survival-rates-are-rising

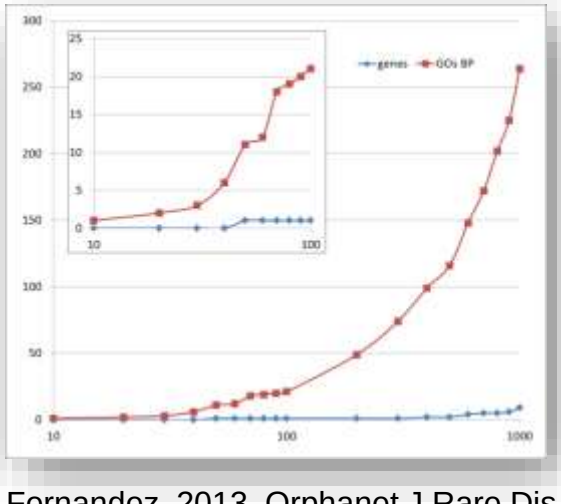
Despite most biomarkers used are single gene variants, most human genetic diseases (and almost all traits) have a modular nature

- Conventional single-gene biomarkers have a demonstrated clinical utility. However, their **success is purely probabilistic**, often **modest** and frequently **lack any mechanistic anchoring** to the fundamental cellular processes responsible for the disease or therapeutic response.
- **Modular nature of genetic diseases: Causative genes** for the same or phenotypically similar diseases may generally reside in the same **biological module**, either a **protein complex** (Lage et al, 2007), a **sub-network** of protein interactions (Lim et al, 2006) , or a **pathway** (Wood et al, 2007)

Disease genes are close in the interactome



Goh 2007 PNAS



Fernandez, 2013, Orphanet J Rare Dis.

Same disease in different populations is caused by different genes affecting the same functions

There are exceptions: MammaPrint, an example of successful breast cancer decision support test based on a multigenic biomarker

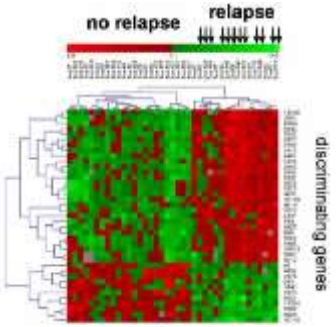
1

Finding genes

Gene expression profiling predicts clinical outcome of breast cancer

Laura J. van 't Veer*†, Hongyue Dai†‡, Marc J. van de Vijver*†, Yudong D. He‡, Augustinus A. M. Hart*, Mao Mao‡, Hans L. Peterse*, Karin van der Kooy*, Matthew J. Marton‡, Anke T. Witteveen*, George J. Schreiber‡, Ron M. Kerkhoven*, Chris Roberts‡, Peter S. Linsley‡, René Bernards* & Stephen H. Friend‡

* Divisions of Diagnostic Oncology, Radiotherapy and Molecular Carcinogenesis and Center for Biomedical Genetics, The Netherlands Cancer Institute, 121 Plesmanlaan, 1066 CX Amsterdam, The Netherlands
‡ Rosetta Inpharmatics, 12040 115th Avenue NE, Kirkland, Washington 98034, USA
† These authors contributed equally to this work



2

Assessing functions

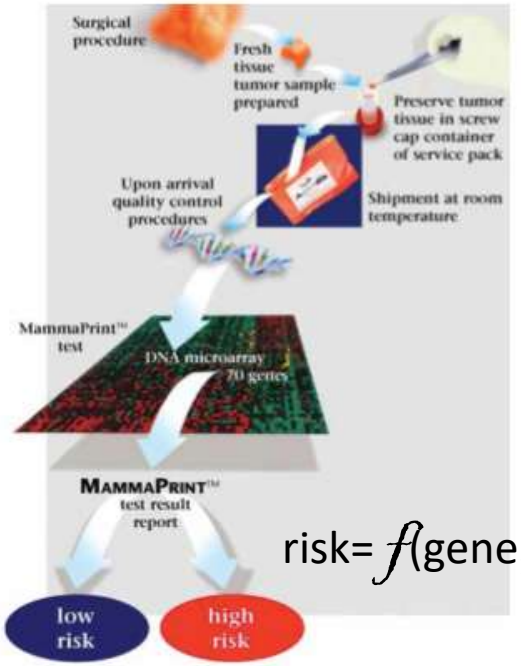
By historic reasons genes were first selected and their functionalities were assessed afterwards.

INSIGHT REVIEW

Enabling personalized cancer medicine through analysis of gene-expression patterns

Laura J. van 't Veer[1,2,3] & René Bernards[1,3,4,5]

Therapies for patients with cancer have changed gradually over the past decade, moving away from the administration of broadly acting cytotoxic drugs towards the use of more-specific therapies that are targeted to each tumour. To facilitate this shift, tests need to be developed to identify those individuals who require therapy and those who are most likely to benefit from certain therapies. In particular, tests that predict the clinical outcome for patients on the basis of the genes expressed by their tumours are likely to increasingly affect patient management, heralding a new era of personalized medicine.



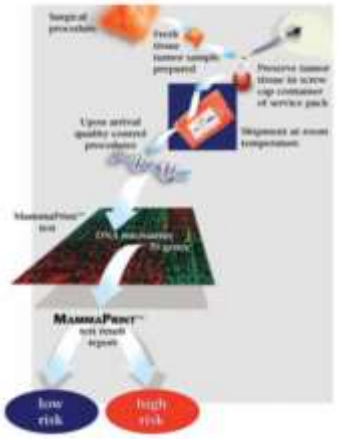
Biological Function	MammaPrint Gene Count
Metabolism	7
Cell cycle and DNA replication	12
Extracellular matrix adhesion and remodeling	5
Growth, proliferation, transformation and cell death	17
General signal transduction and intracellular transport	8
Growth factor	7
Motility or actin filament related	5
Intracellular hydrolase	1
Immune response	1
Neuropeptide	1
Predicted transmembrane protein with unknown function	2
Predicted transcriptional control or DNA binding proteins	5
Unknown function	6
Total Gene Count	70

The strength of this approach is that it is unbiased: there are no assumptions about which genes are likely to be involved in the process of interest. For example, in a **data-driven** study of the prognosis of patients with breast cancer, **little was known about the function** of 15 of the 70 genes that were found to constitute a prognostic gene-expression signature[4]. A drawback of this approach is that the **outcome relies solely on the quality of the data** (and the samples).
By contrast, using the **knowledge-driven** approach, genes that are thought to be relevant to a particular cancer trait are selected on the basis of the scientific literature.

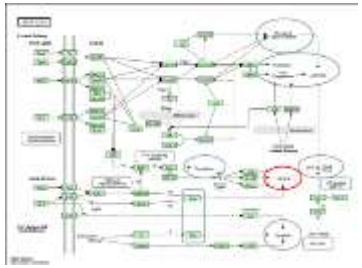
Risk is calculated as a function of the 70 gene expression levels

$$risk = f(gene_1, gene_2, \ldots gene_{70})$$

Change in the paradigm

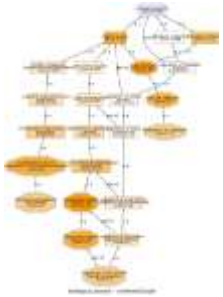


MammaPrint and other multigenic biomarkers: bottom up, from genes to functions that define one (or several) biological modules.



Models of cell functionality: top-down mechanism-based biomarkers, from biological modules to genes

Two problems: defining functional modules and modeling their behavior



Definition

Behavior

Gene ontology: descriptive; unstructured functional labels

Enrichment methods. GO, etc. (simple statistical tests). No information on how components relate among them

Interactome: relationships among components but unknown function

Connectivity models. Protein-protein, protein-DNA and protein-small molecule interactions (tests on network properties). No information the functional roles of the components

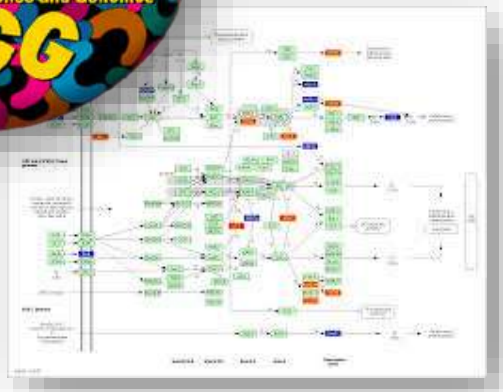
Pathways: relationships among components and their functional roles

Mathematical models. Kinetic models including stoichiometry, balancing reactions, etc.

Computational models. Models of signalling pathways, metabolic pathways, regulatory pathways, etc. (executable models)

Defining the module: Pathways: maps of cell activity (in sickness and in health)

www.genome.jp/kegg



disease-maps.org/

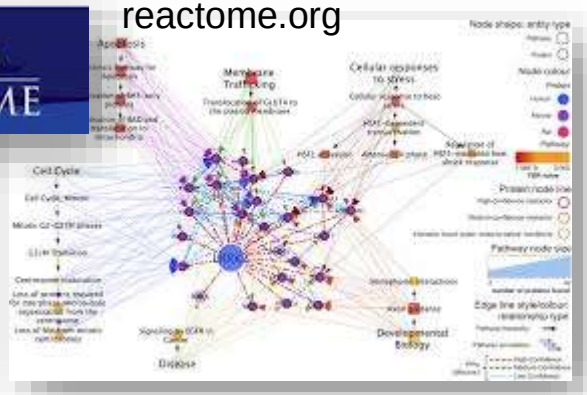
Oncogénesis

Alzheimer

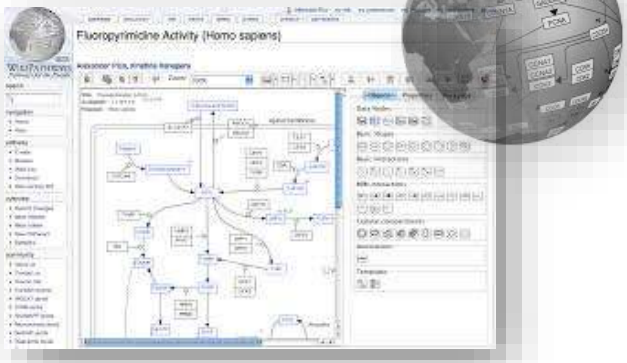
More disease maps…

Parkinson

reactome.org



www.wikipathways.org/

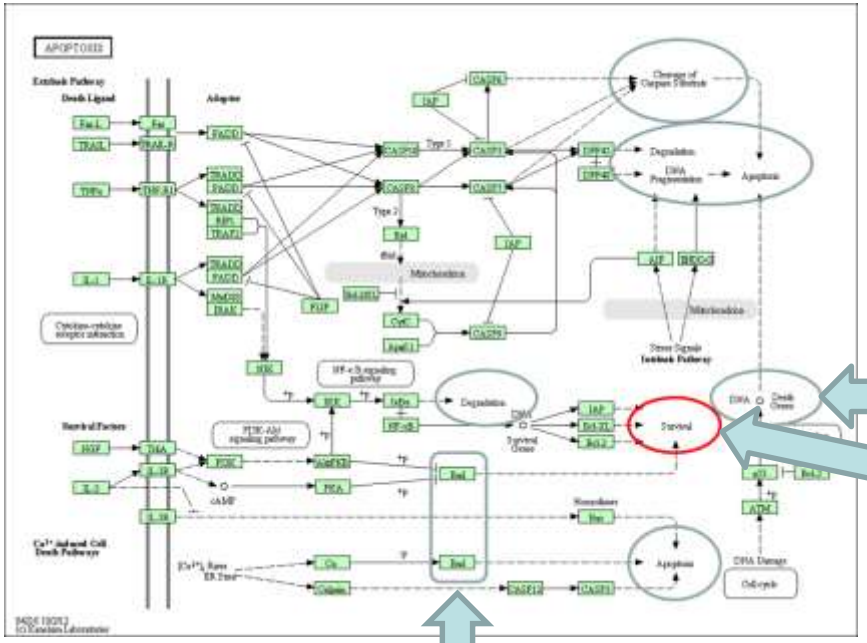


Defining pathway activity

We first need a map: pathways are defined in different repositories (**KEGG**, Reactome, Biocarta, disease maps, etc.)

What pathway level makes a real biological meaning?

Gene ⇨ sub-pathway ⇨ pathway



Enrichment methods (pathway-level): Different and often opposite cell behaviors are triggered by the same **pathway**.
E.g.: death and survival

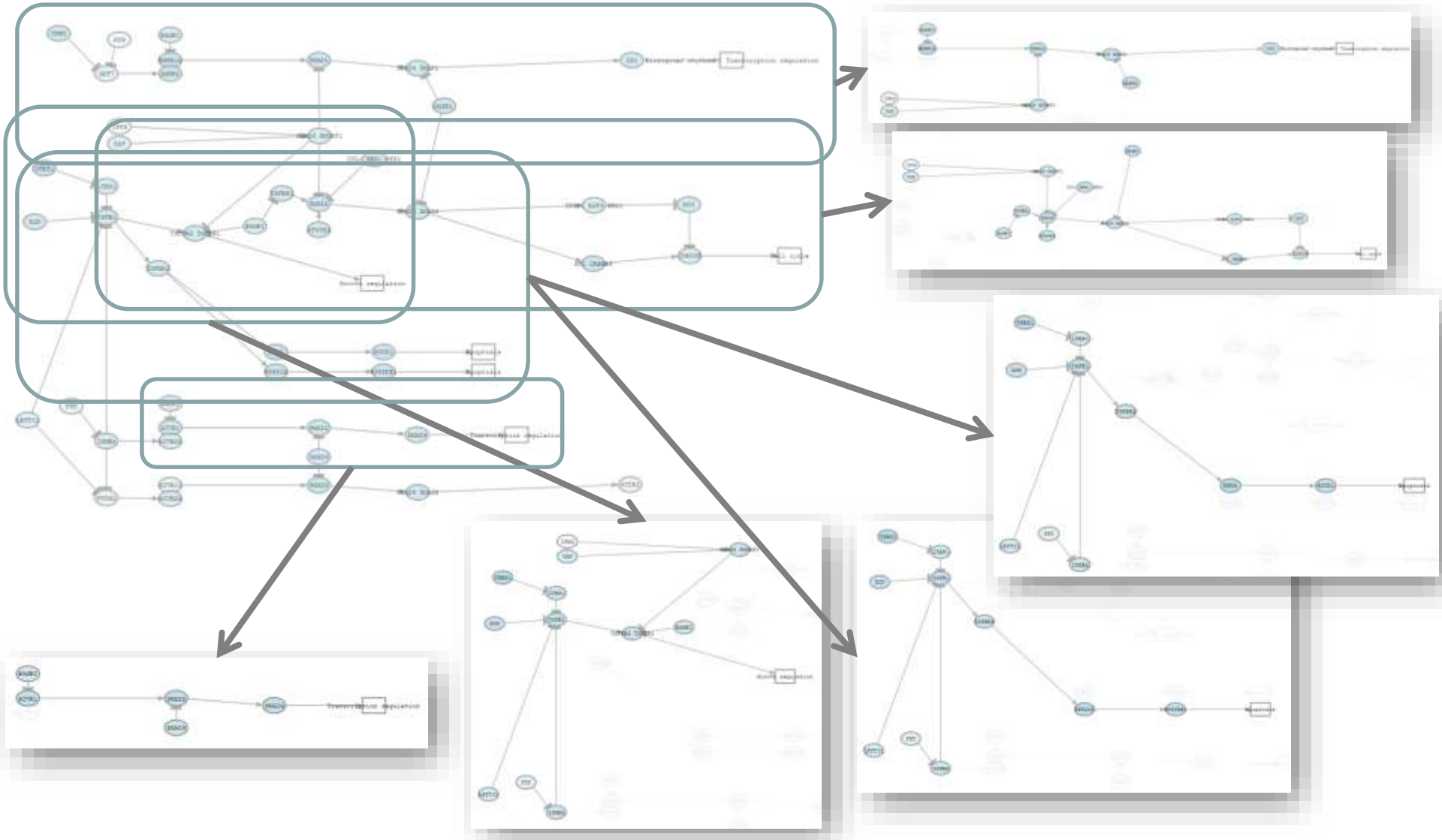
Death

Survival

Gene level: The same gene can trigger different (and often opposite) responses, depending on the stimulus

Sub-pathway (elementary circuit) connects stimulus to response

Decomposition of a pathway into their elementary circuits



How realistic are models of pathway activity?

RESEARCH ARTICLE

CANCER

Signaling pathway models as biomarkers: Patient-specific simulations of JNK activity predict the survival of neuroblastoma patients

Dirk Fey,[1] Melinda Halasz,[1] Daniel Dreidax,[2] Sean P. Kennedy,[1] Jordan F. Hastings,[3] Nora Rauch,[1] Amaya Garcia Munoz,[1] Ruth Pilkington,[1] Matthias Fischer,[4,5,6] Frank Westermann,[2] Walter Kolch,[1,7,8] Boris N. Kholodenko,[1,7,8]* David R. Croucher[1,3,9]*

Signaling pathways control cell fate decisions that ultimately determine the behavior of cancer cells. Therefore, the dynamics of pathway activity may contain prognostically relevant information different from that contained in the static nature of other types of biomarkers. To investigate this hypothesis, we characterized the network that regulated stress signaling by the c-Jun N-terminal kinase (JNK) pathway in neuroblastoma cells. We generated an experimentally calibrated and validated computational model of this network and used the model to extract prognostic information from neuroblastoma patient–specific simulations of JNK activation. Switch-like JNK activation mediates cell death by apoptosis. An inability to initiate switch-like JNK activation in the simulations was significantly associated with poor overall survival for patients with neuroblastoma with or without *MYCN* amplification, indicating that patient-specific simulations of JNK activation could stratify patients. Furthermore, our analysis demonstrated that extracting information about a signaling pathway to develop a prognostically useful model requires understanding of not only components and disease-associated changes in the abundance or activity of the components but also how those changes affect pathway dynamics.

Beyond static biomarkers—The activity of signalling networks as an alternate biomarker?

Fey et al., Sci. Signal. 8, ra130 (2015).

Inability of JNK activation (that mediates apoptosis) is associated to bad prognostic, irrespective of *MYCN* amplification status

Problem: ODE can efficiently solve only small systems

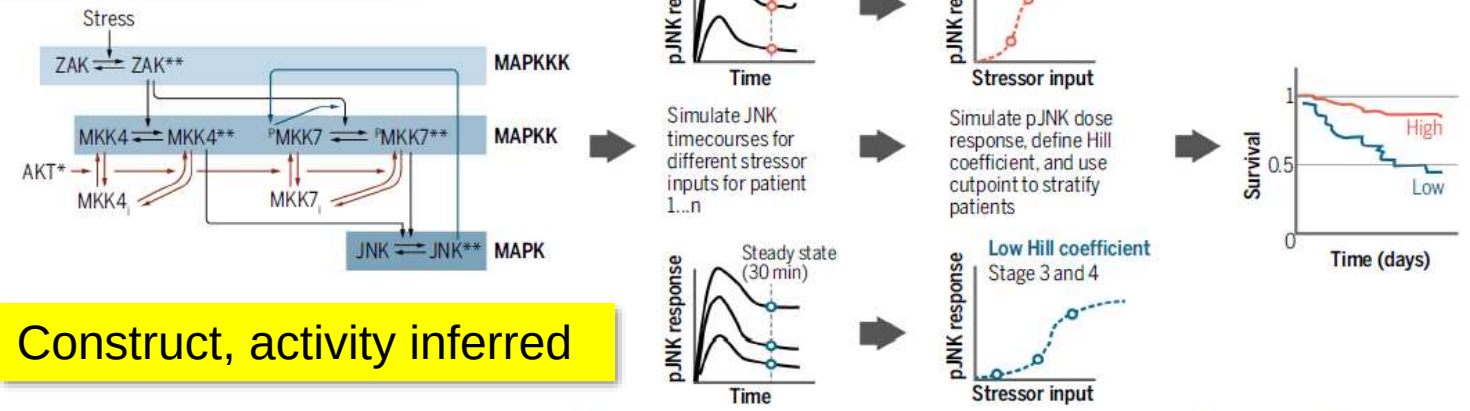
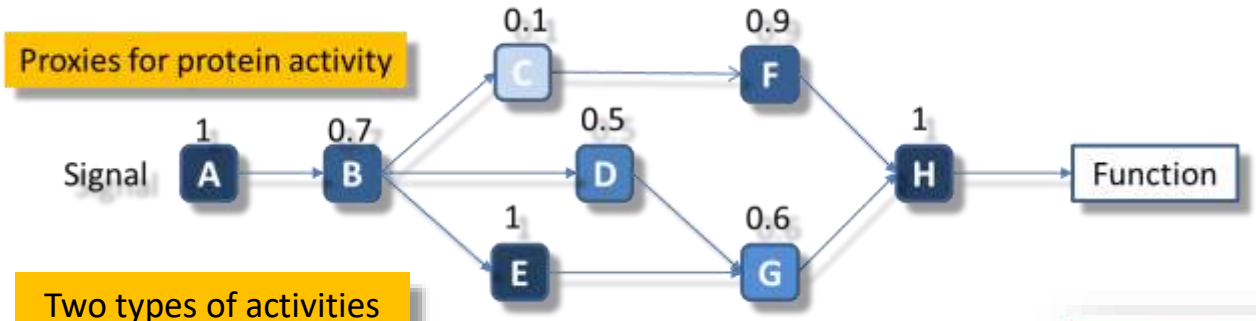
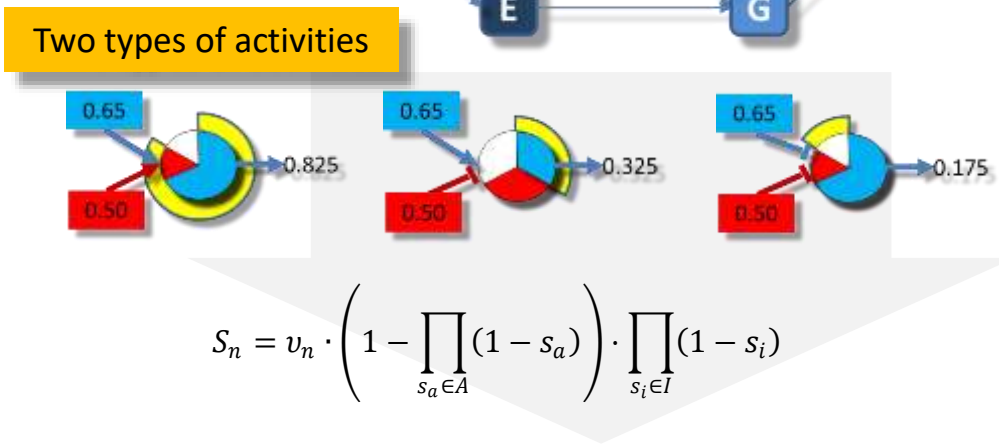


Fig. 1. Using network descriptors of signaling pathway activation potential to predict patient response. After construction of a computational model based on the validated network topology and that reproduces the signaling pathway dynamics, the model can be used to identify network descriptors, such as the Hill coefficient, that are calculated from the dynamic simulation of the activation of a signaling pathway. These in silico biomarkers cannot be directly measured.

Modeling signaling pathways with signal propagation models



From individual gene expression profiles…




www.impactjournals.com/oncotarget/ Oncotarget, Advance Publications 2016

High throughput estimation of functional cell activities reveals disease mechanisms and predicts relevant clinical outcomes

Marta R. Hidalgo[1], Cankut Cubuk[1], Alicia Amadoz[1,2], Francisco Salavert[1,3], José Carbonell-Caballero[1], Joaquin Dopazo[1,2,3]

[1]Computational Genomics Department, Centro de Investigación Príncipe Felipe (CIPF), Valencia, 46012, Spain

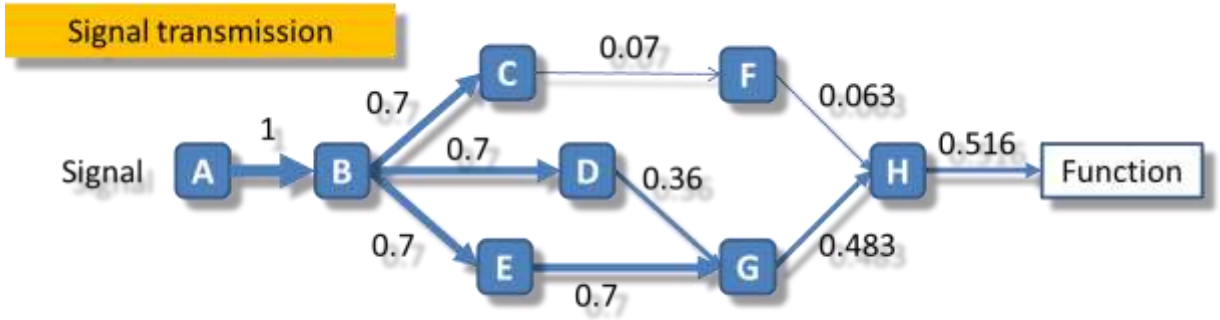
[2]Functional Genomics Node (INB-ELIXIR-es), Valencia, 46012, Spain

[3]Bioinformatics in Rare Diseases (BIER), Centro de Investigación Biomédica en Red de Enfermedades Raras (CIBERER), Valencia, 46012, Spain

Correspondence to: Joaquín Dopazo, **email:** jdopazo@cipf.es

Keywords: signaling pathway, disease mechanism, prognostic, survival, biomarker

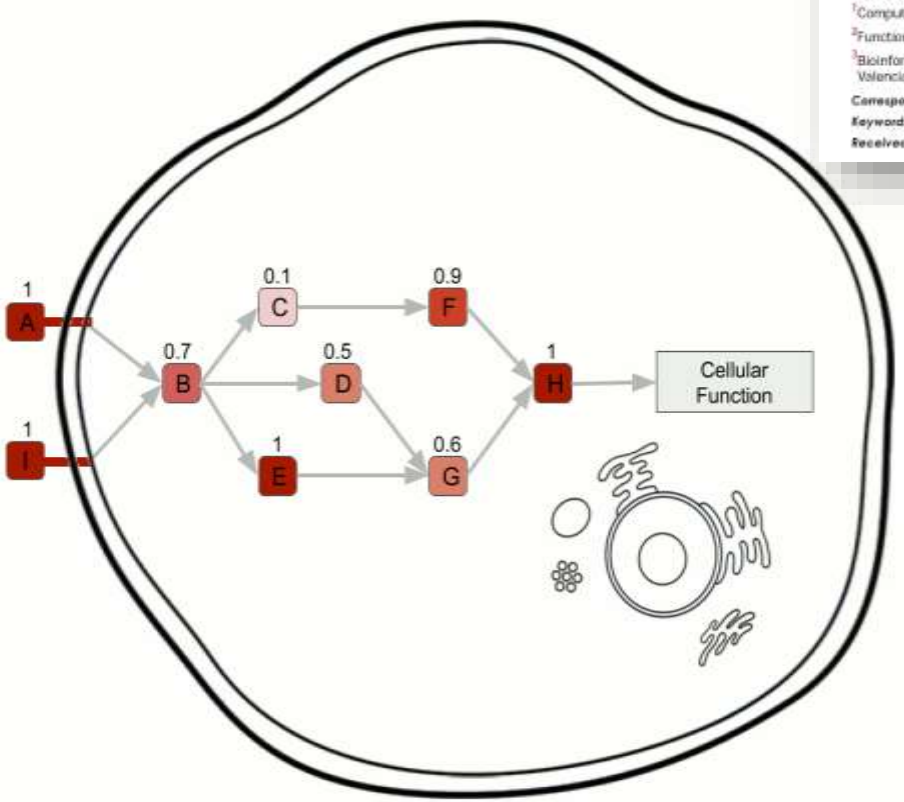
Received: September 01, 2016 **Accepted:** November 21, 2016 **Published:**



…to profiles of circuit activity (and functional activity)

Are scalable

Signal propagation models are mechanistic



www.impactjournals.com/oncotarget/ Oncotarget, Advance Publications 2016

High throughput estimation of functional cell activities reveals disease mechanisms and predicts relevant clinical outcomes

Marta R. Hidalgo[1], Cankut Cubuk[1], Alicia Amadoz[1,2], Francisco Salavert[1,3], José Carbonell-Caballero[1], Joaquin Dopazo[1,2,3]

[1]Computational Genomics Department, Centro de Investigación Príncipe Felipe (CIPF), Valencia, 46012, Spain

[2]Functional Genomics Node (INB-ELIXIR-es), Valencia, 46012, Spain

[3]Bioinformatics in Rare Diseases (BiER), Centro de Investigación Biomédica en Red de Enfermedades Raras (CIBERER), Valencia, 46012, Spain

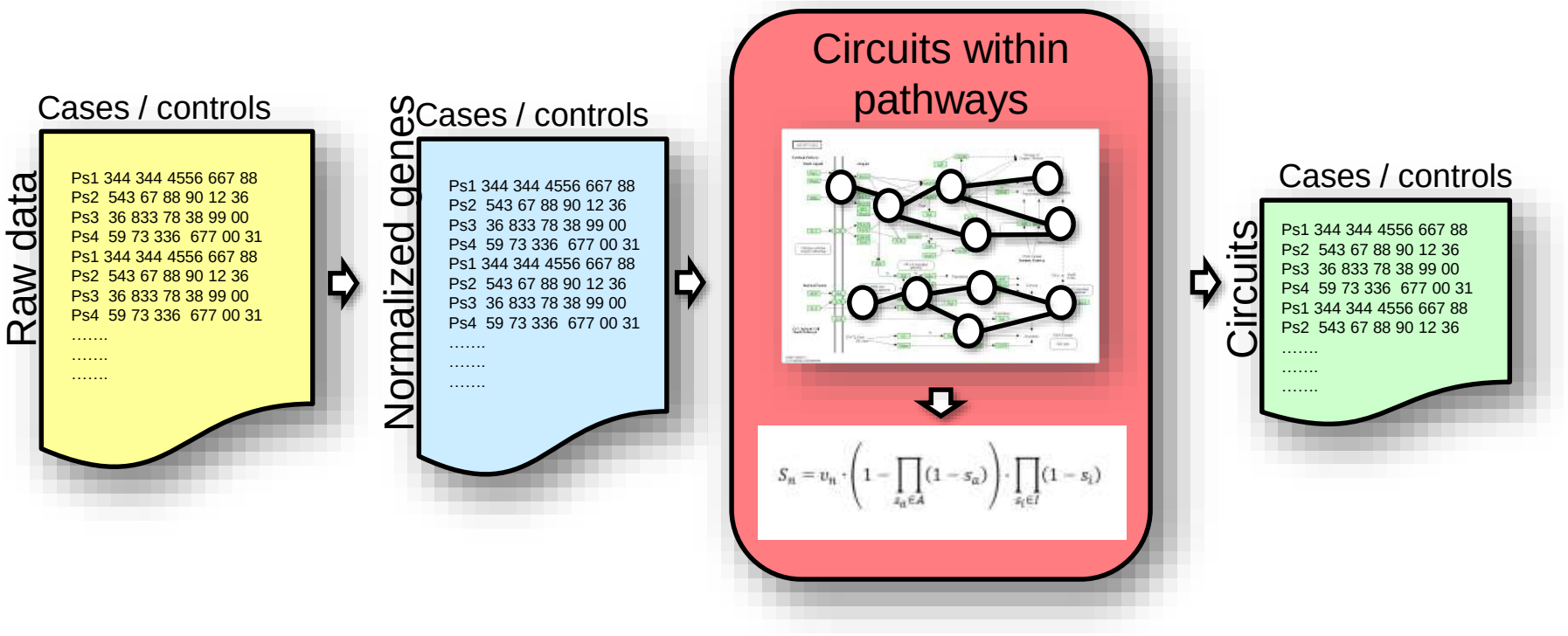
Correspondence to: Joaquín Dopazo, **email:** jdopazo@cipf.es

Keywords: signaling pathway, disease mechanism, prognostic, survival, biomarker

Received: September 01, 2016 **Accepted:** November 21, 2016 **Published:**

The transduction of the signal has a **causal** consequence over cell decisions. The models quantitatively predict cell activities

Gene expression data are transformed into signal activity intensities

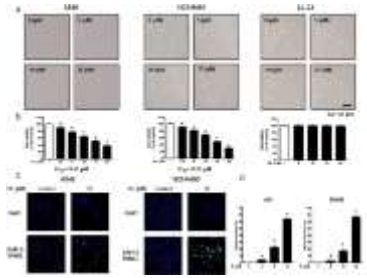


A simple transformation of raw data (normalization) and an algorithm for signal propagation results in accurate estimations of circuit activities.

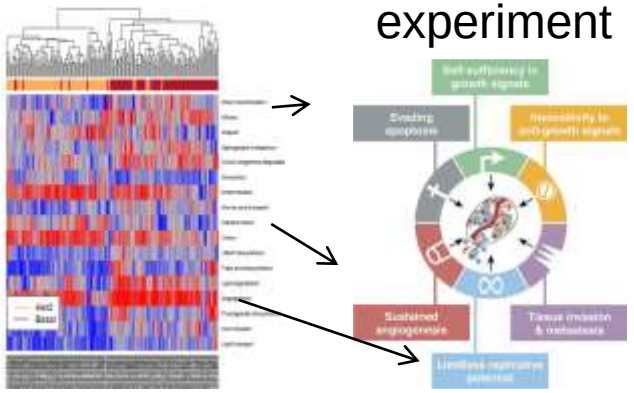
¡ The same concept that MammaPrint, risk= $f(\text{gene}_1, \text{gene}_2, \ldots \text{gene}_{70})$ but based on biological knowledge, is used here to estimate cell functional activity !

Models of signaling activity provide high-throughput estimations of intensity activation of cell functions from gene expression measurements

Some (not all) conventional cell function can be studied, <u>one at a time,</u> in individual assays

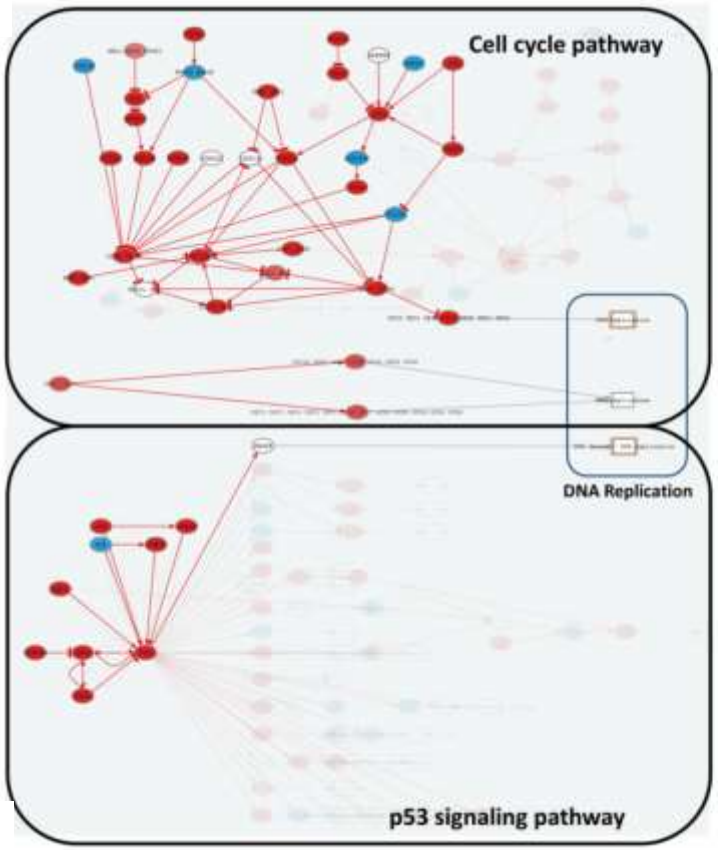


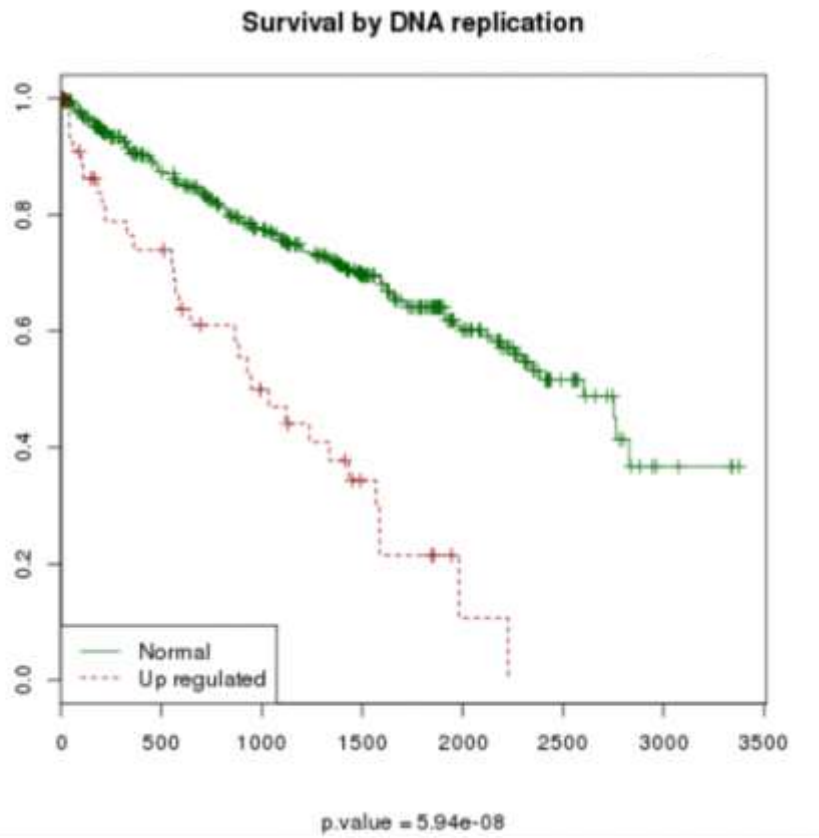
With Signaling models, the intensity at which the <u>whole repertoire of cell functions</u> is triggered can be measured in only <u>one individual</u> experiment



Hypothesis: the intensity at which functions are triggered by the signaling system of the cell is more related to phenotypes than the intensity of gene expression

Signaling activity trigger cell functions directly related to cancer progression





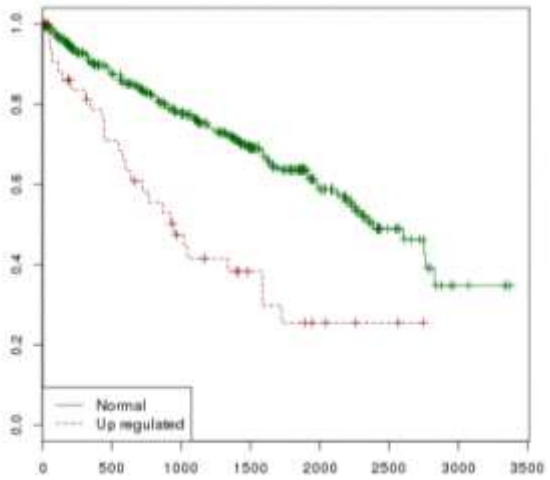
DNA replication= f(gene$_1$, gene$_2$, … gene$_n$)

Hidalgo et al., 2017 Oncotarget

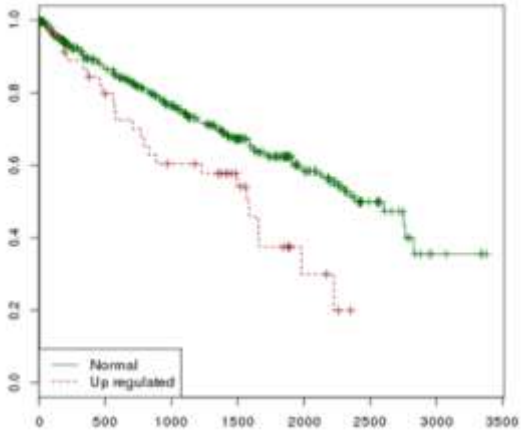
DNA replication function is a **construct**: the activity is **inferred** not measured

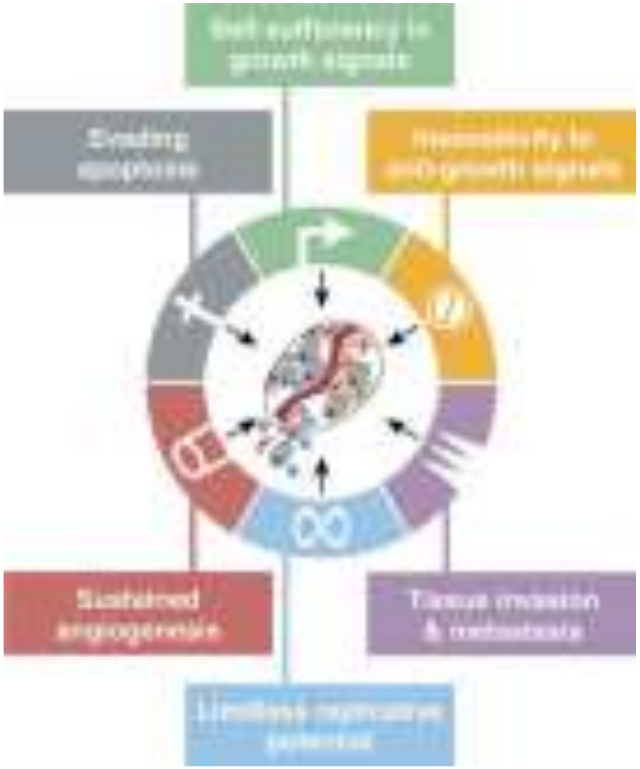
Actually, signal activity triggers all the cancer hallmarks

Negative regulation of release of cytochrome c from mitochondria (inhibition of apoptosis)



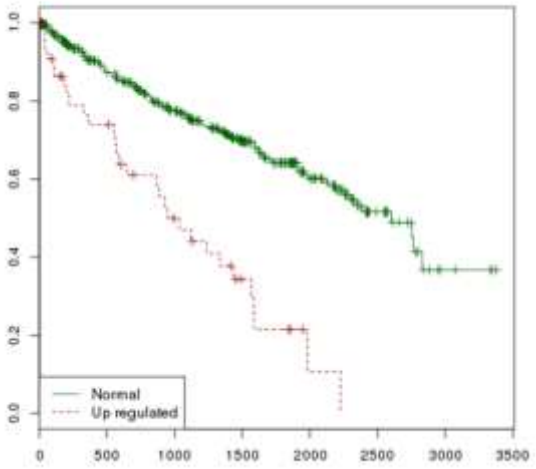
positive regulation of angiogenesis



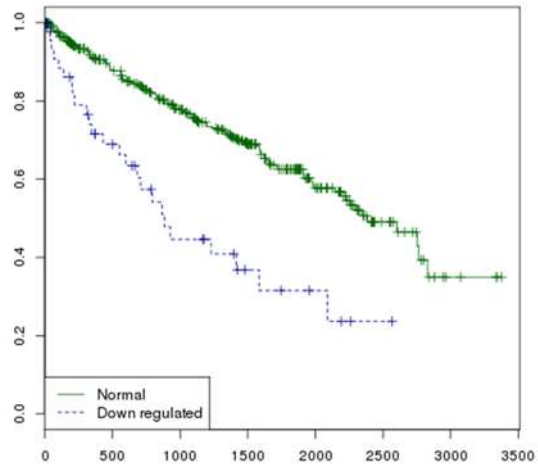


Hanahan, Weinberg, 2011
Hallmarks of cancer: the next generation. Cell 144, 646

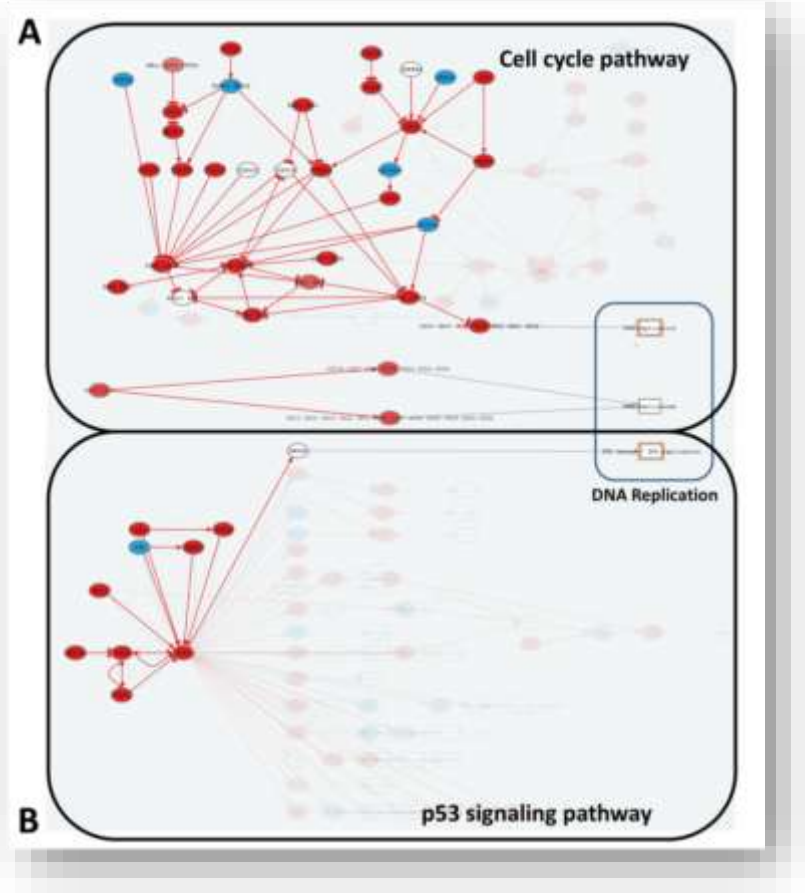
DNA replication

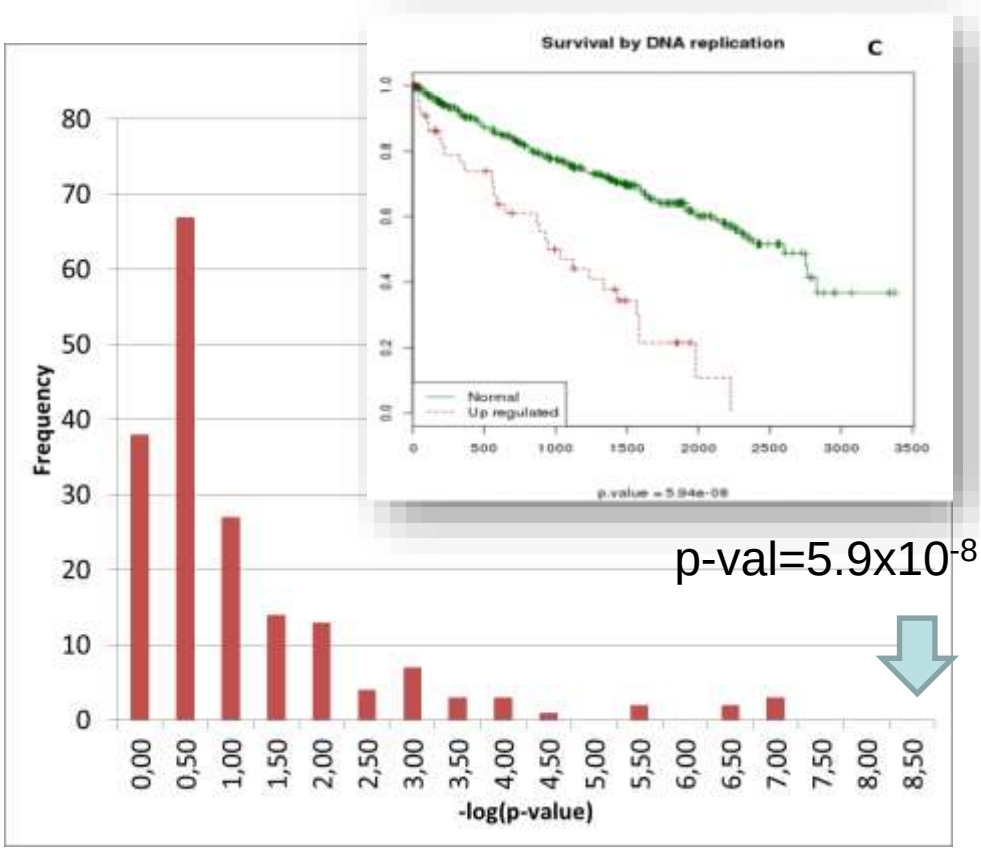


Cell adhesion



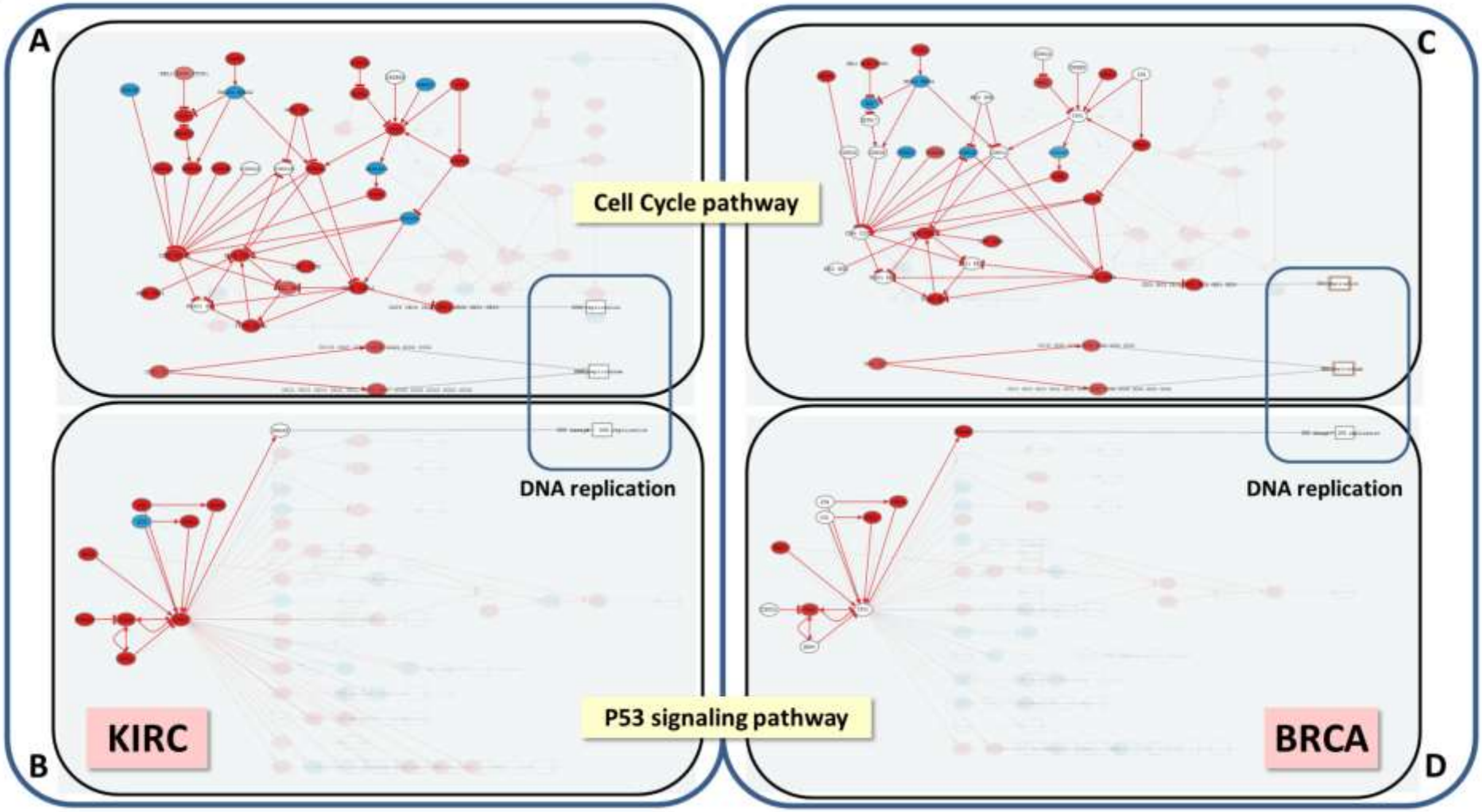
The inferred function activity (mechanistic biomarker) is more correlated to survival than the activity of any gene (conventional biomarker) in the circuit





The system is more than the sum of its parts!

Different cancer use different gene expression programs to activate the same functions



Mechanistic models and causality

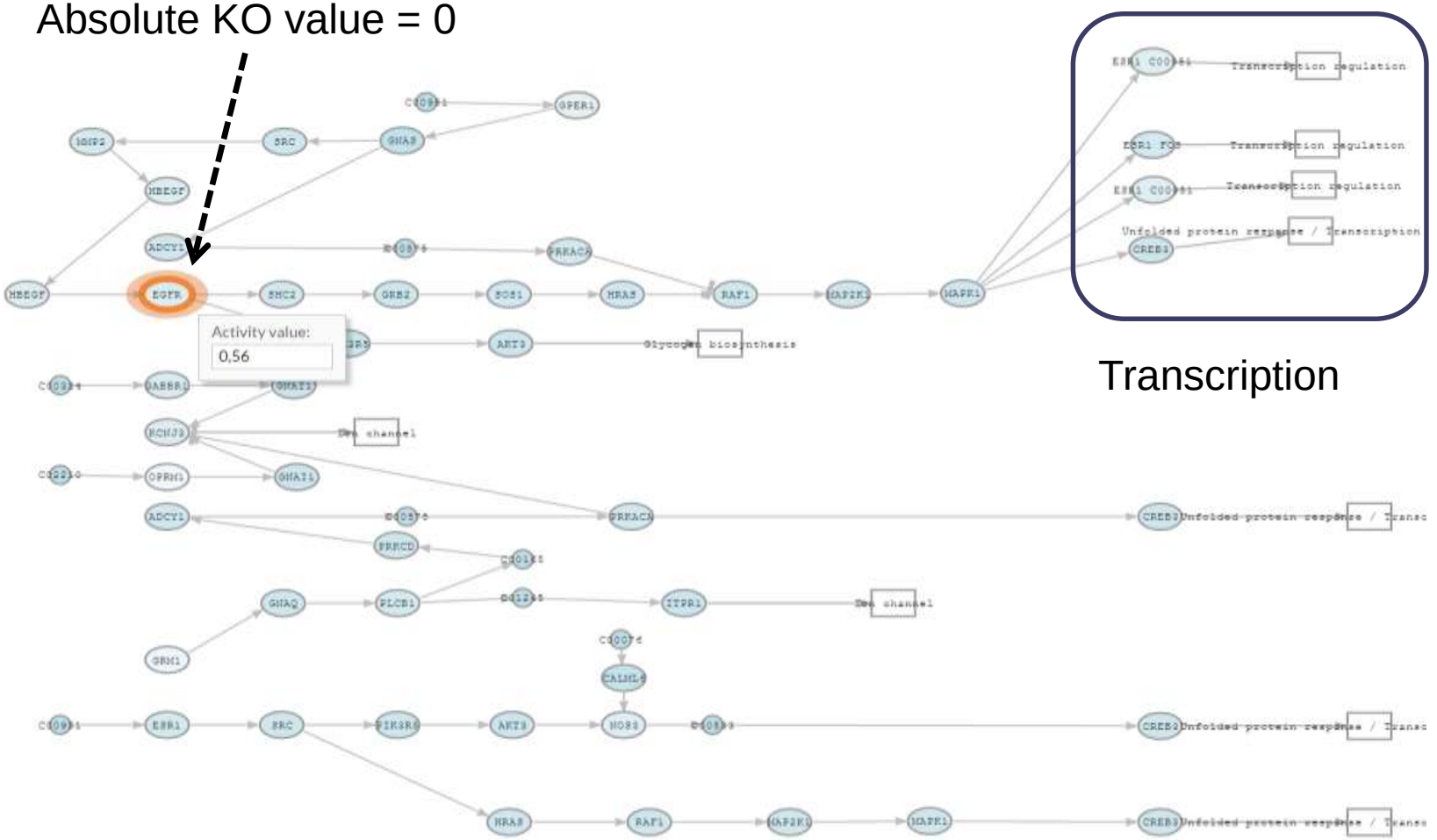
An interesting property of **mechanistic models** is that they can be used to predict the potential consequences that perturbations (KOs, inhibitions, mutations or changes in the expression) of the proteins that compose the pathway can have over the individual circuits that trigger cell actions or the production of metabolites. Thus, changes in cell activity can be linked to changes in gene activity/integrity by means of a **causal relationship**.

By **simulating** changes of gene expression/activity it is easy to:

- Directly study of the consequences of induced gene over-expressions or KOs
- Carry out reverse studies of genes that need to be perturbed to change cell functionalities, such as:
  - Reverting the “normal” functional status of a cell
  - Selectively kill diseased cells without affecting normal cells
  - Studying the effect of mutations
  - Etc.

Causal effect of an intervention

We can inhibit EGFR (target of Afanatib) by reducing its activity value (0.56 in cancer).
Absolute KO value = 0



Transcription

Estrogen signaling pathway

http://pathact.babelomics.org/

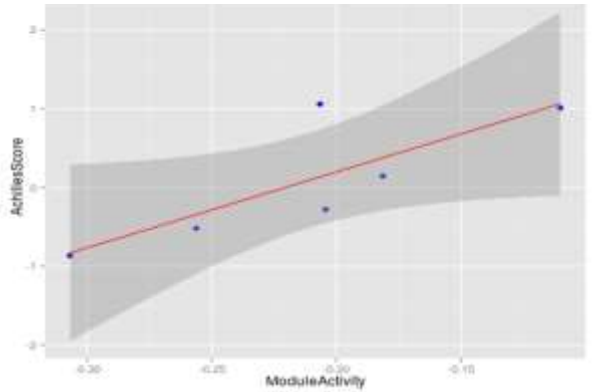
Model validation with massive KOs (1)

The activity of some signaling circuits is correlated with cell survival

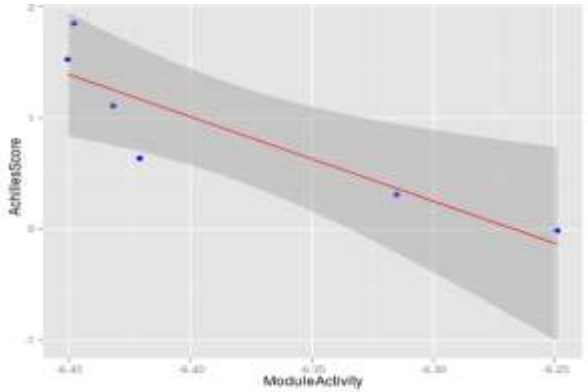
Survival data from Achilles cell line KOs (Broad Institute) can be compared to the change in circuit activities **predicted** by the model

Increased survival

Onco-circuit



Tumor suppressor circuit



Increased circuit activity

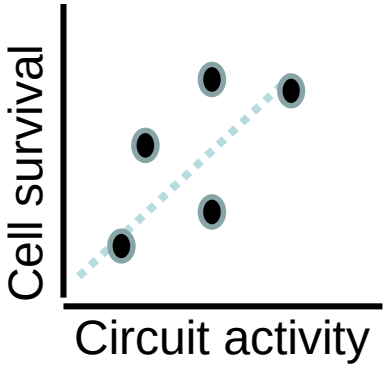
Functions:

- Repression of apoptosis
- Activation of proliferation

- Activation of apoptosis
- Repression of proliferation

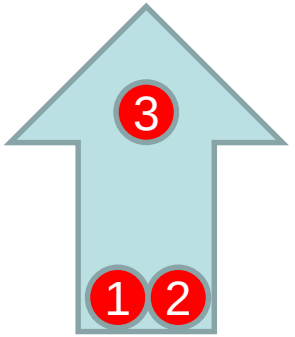
Essential circuits: once found, other ways of deactivating these circuits can be find, opening the door to knowledge-based target discovery

Model validation with massive KOs (2)

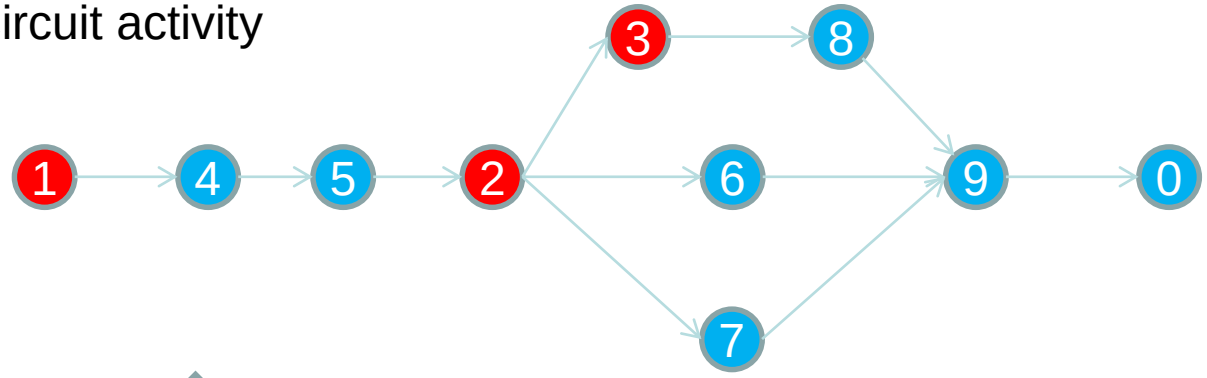


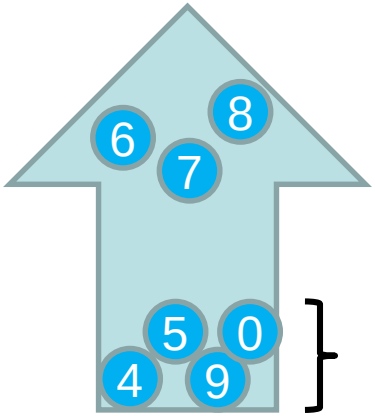
In modules known to be associated to cell survival

Achilles I



Circuit activity with gene KO



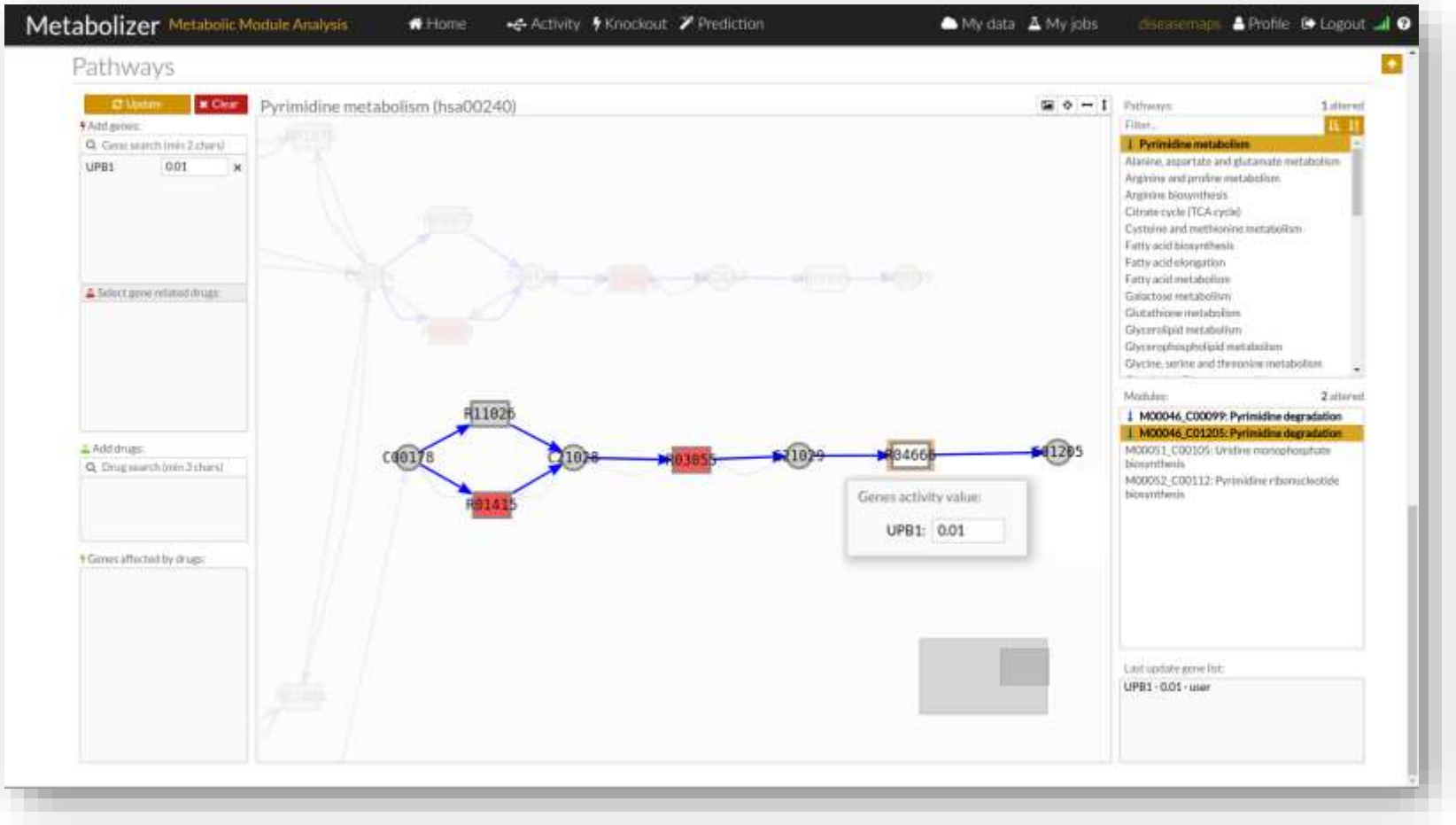


Prediction of other gene targets, whose inhibition (modeled KO) **deactivate** these **circuits** and consequently **decrease** cell **viability**

Potential target (inhibition of circuit activity)

An example of KO prediction

Pyrimidine degradation pathway was predicted to be an onco-module in gastric cancer cell lines. Predicted genes that switch the pathway off are *DPYD*, *DPYS* (confirmed in Achilles) and *UPB1* (not present in Achilles)



Prediction of gene essentiality from metabolic module essentiality

Pyrimidine degradation pathway was predicted to be an onco-module in gastric cancer cell lines. Predicted genes that switch the pathway off are *DPYD*, *DPYS* (confirmed in Achilles) and *UPB1*

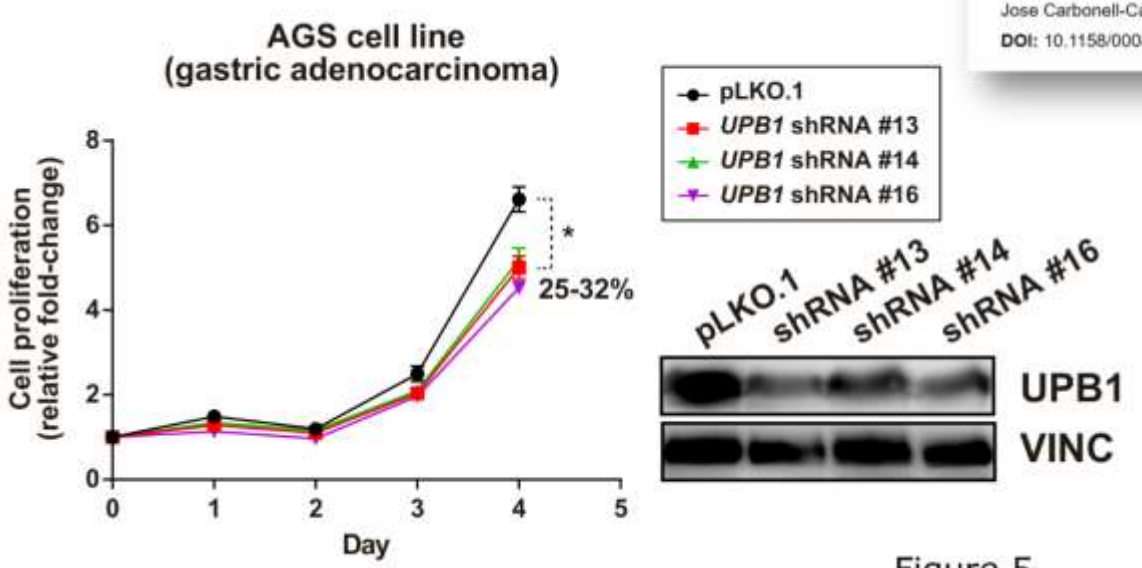


Figure 5

Disecting cell behavior at Single Cell level with mechanistic models

Published online 25 June 2020

NAR Cancer, 2020, Vol. 2, No. 2 **1**
doi: 10.1093/narcan/zcaa011

Mechanistic models of signaling pathways deconvolute the glioblastoma single-cell functional landscape

Matías M. Falco[1,2], María Peña-Chilet[1,2], Carlos Loucera[1], Marta R. Hidalgo[3] and Joaquín Dopazo[1,2,4,5,*]

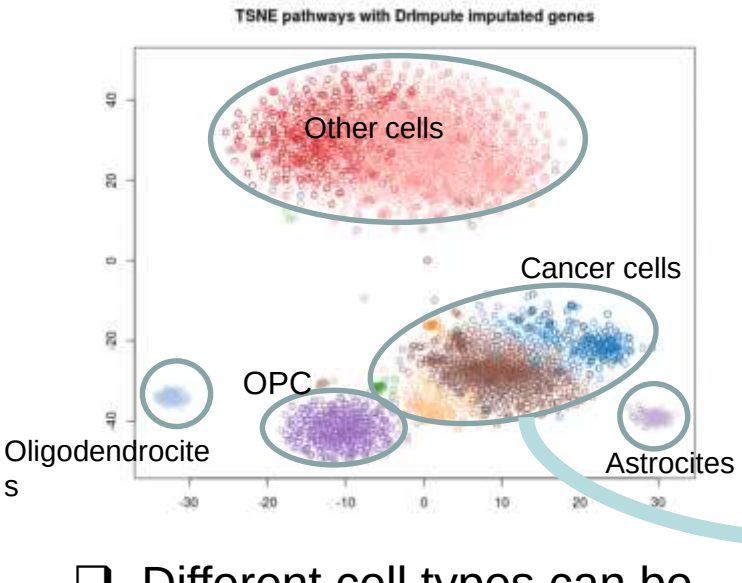
[1]Clinical Bioinformatics Area, Fundación Progreso y Salud (FPS), Hospital Virgen del Rocío, 41013 Sevilla, Spain, [2]Bioinformatics in Rare Diseases (BiER), Centro de Investigaciones Biomédicas en Red en Enfermedades Raras (CIBERER), 41013 Sevilla, Spain, [3]Unidad de Bioinformática y Bioestadística, Centro de Investigación Príncipe Felipe (CIPF), 46012 Valencia, Spain, [4]Functional Genomics Node, FPS/ELIXIR-ES, Hospital Virgen del Rocío, 41013 Sevilla, Spain and [5]Computational Systems Medicine Group, Institute of Biomedicine of Seville (IBIS), Hospital Virgen del Rocío, 41013 Sevilla, Spain

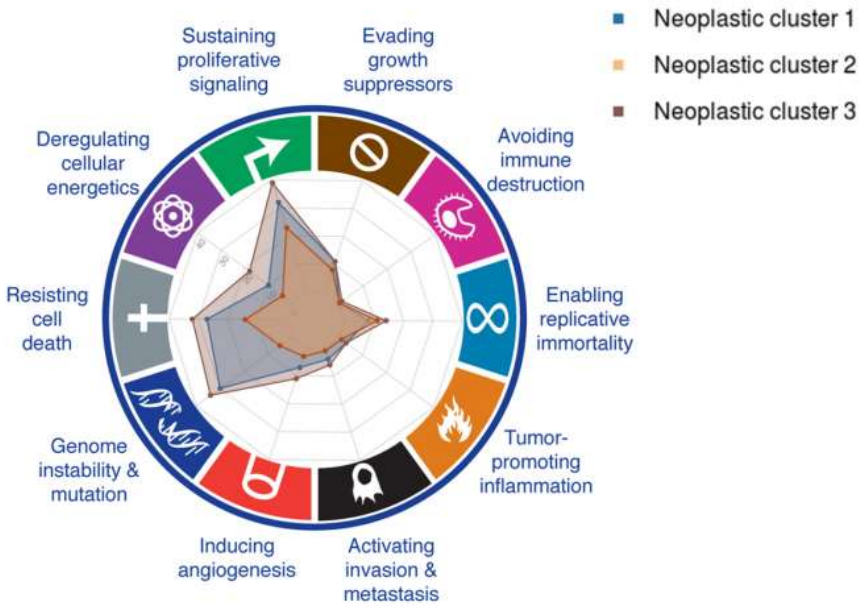
Received March 10, 2020; Revised June 08, 2020; Editorial Decision June 10, 2020; Accepted June 11, 2020

https://academic.oup.com/narcancer/article/2/2/zcaa011/5862620

Disecting cell behavior at Single Cell level with mechanistic models

Single-cell RNA-seq analysis of 3589 individual cells including glioblastoma cells and other surrounding and infiltrating cells .

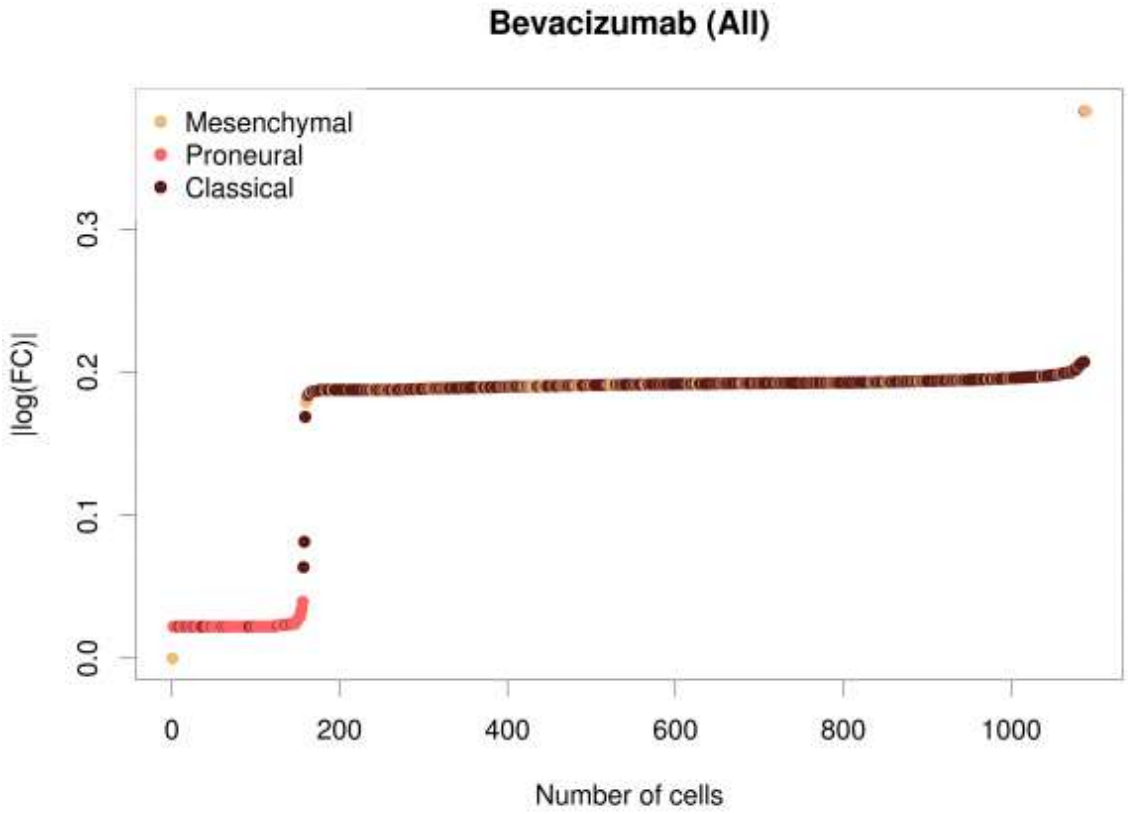




- ❑ Different cell types can be distinghised according to their functional profiles.
- ❑ Cancer cells actually define a heterogeneous subpopulation with 3 clusters

Different functional strategies across different cancer cells can be discovered

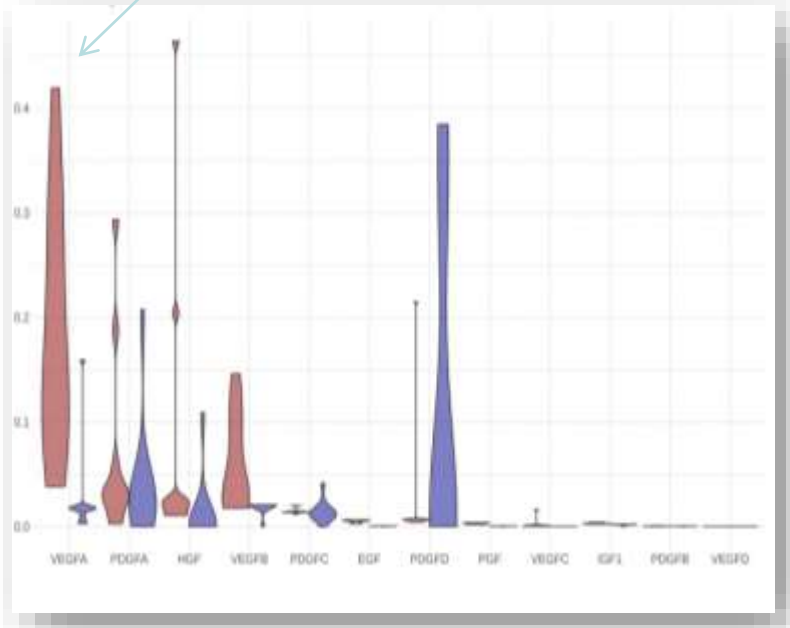
Disecting drug response at single cell level



Drug effect can be simulated and individual cell-level responses enables the study of strategies to evade the drug and allows suggesting complementary drugs

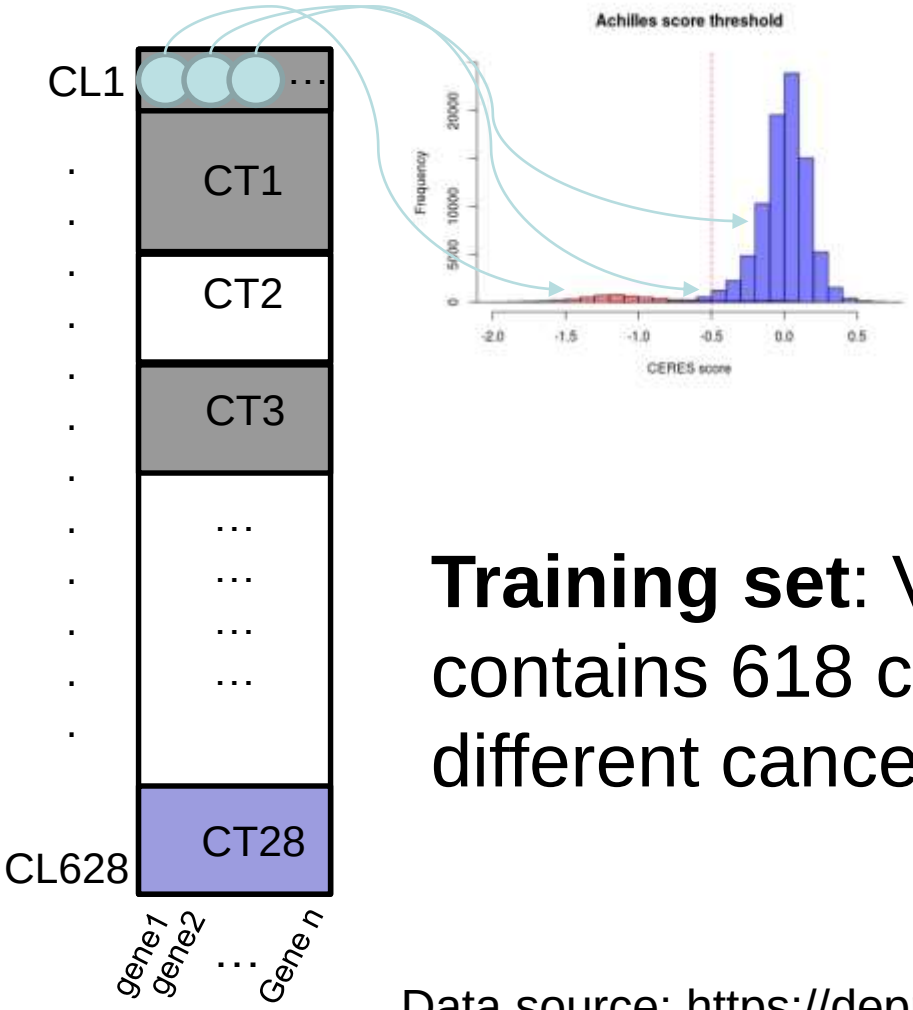
Disecting drug response at single cell level

VEGF+39 more genes



Low-responder cells are characterized by a low level of *VEGF* compensated with a high level of *PDGFD*, which makes these cells virtually insensitive to the inhibition of *VEGF*

ML to predict gene essentiality in cancer cell lines through their causal effect on signaling

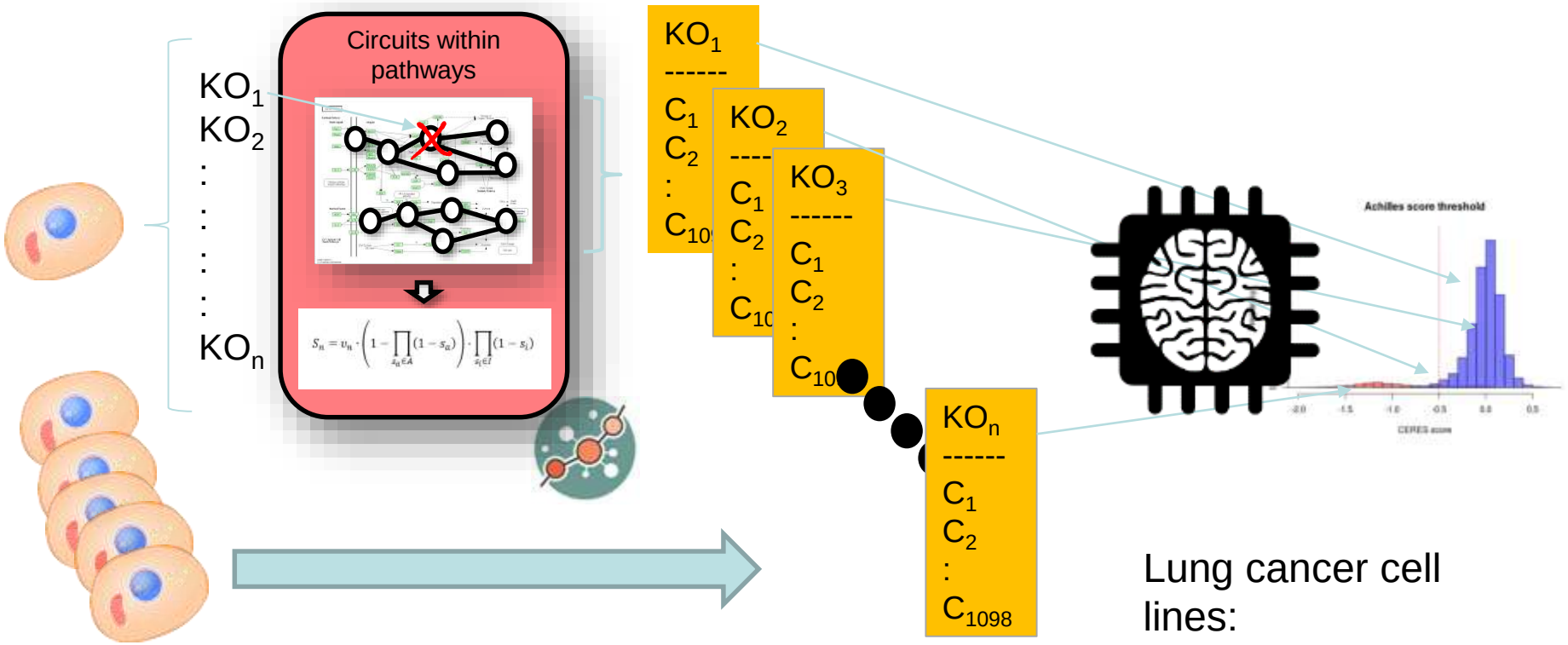


The Achilles experiment: massive KO of cell lines with CRISP-R -> cell proliferation score

Training set: Version 7 (19_Q3) contains 618 cell lines belonging to 28 different cancer types.

Data source: https://depmap.org/portal/download/

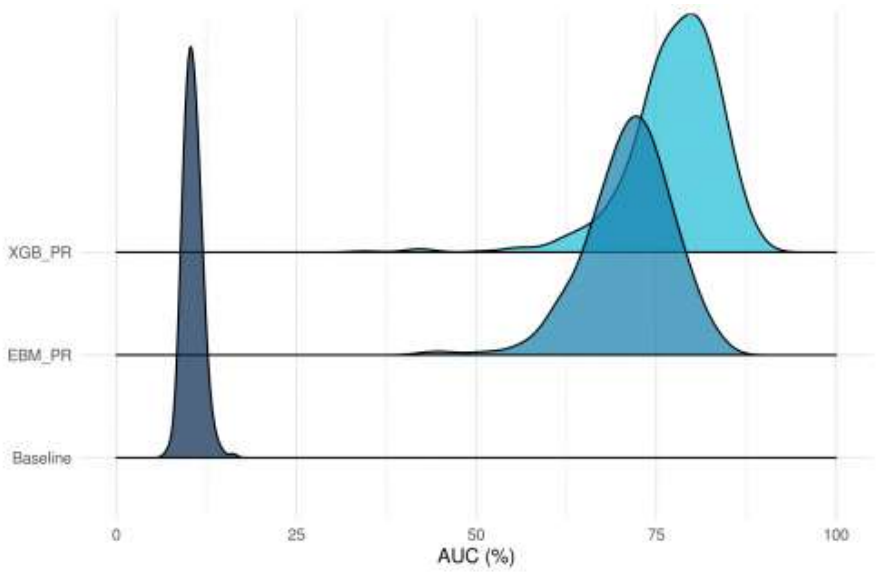
Prediction of the effect of KOs on cell survival

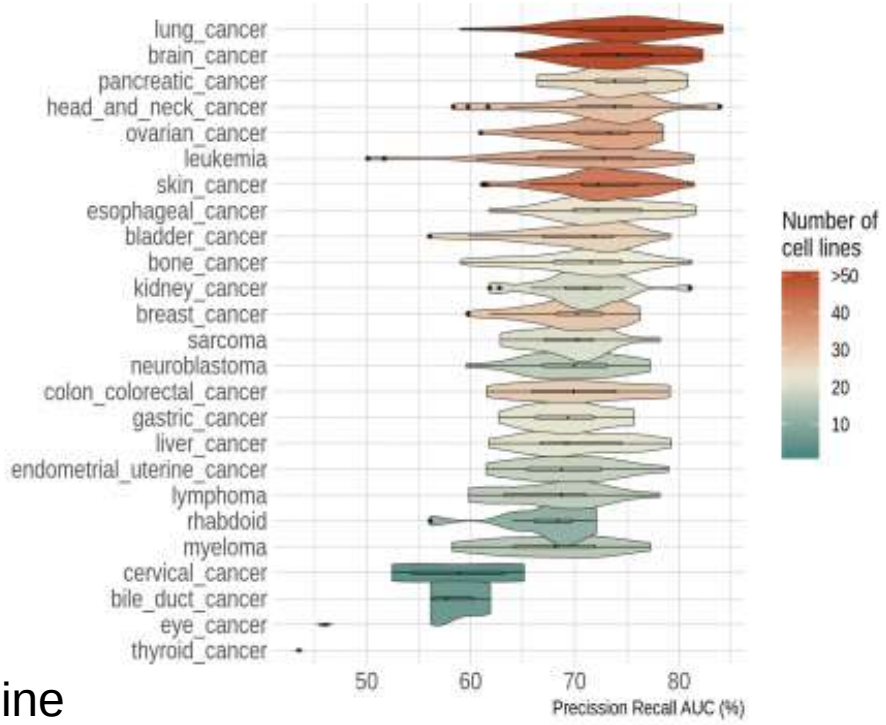


Ensemble and stacking of classifiers (decision trees) with Bayesian hyper-parameter optimization, using geometric mean that capture parameter imbalances

Lung cancer cell lines:
6751 KOs. Vectors of 1098 circuits to predict a binary value (unbalanced: only 10% KOs kill the cell)

General prediction precision

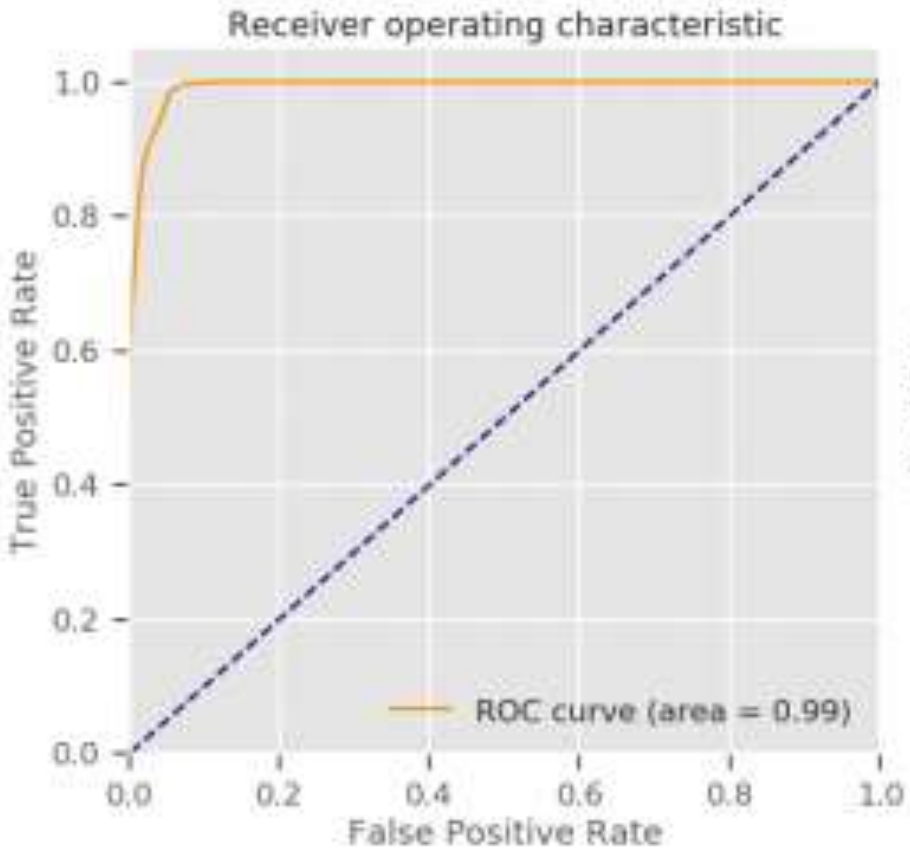


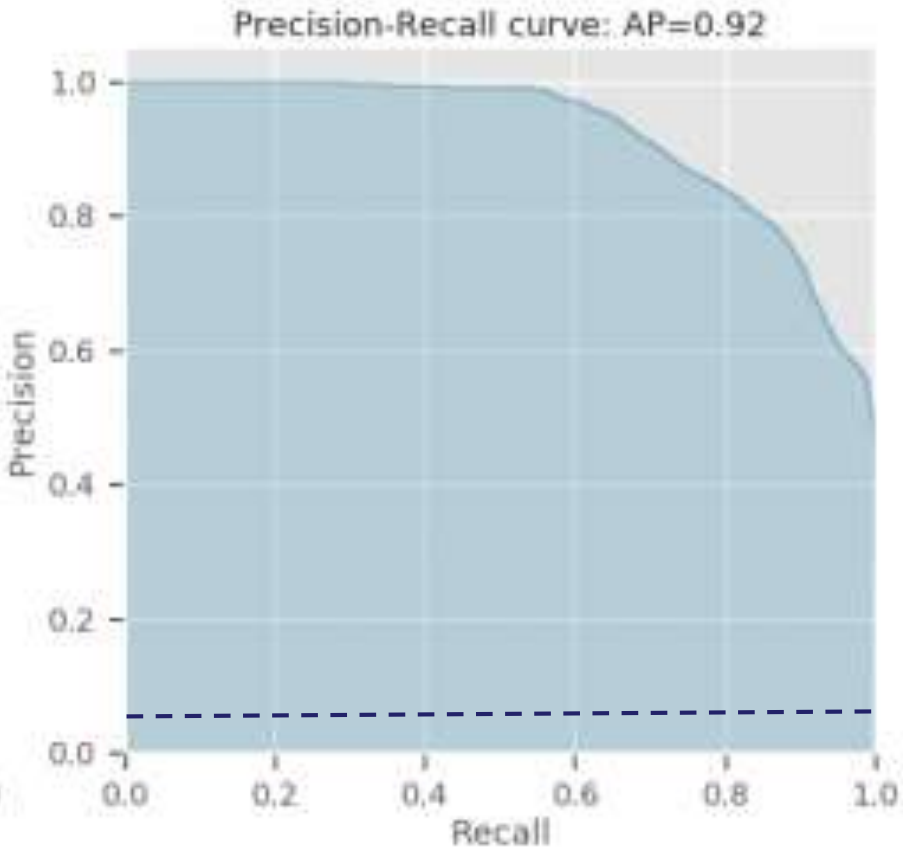


Precision measured by Leaving One Cell line Out (LOCO)
Extreme Gradient Boosting (XGB) is more precise than Explainable Boosting Machine (EBM), but lacks a clear interpretability.

The more samples to train the better the result

Prediction of the effect of a KO in a new cell line of the same cancer type

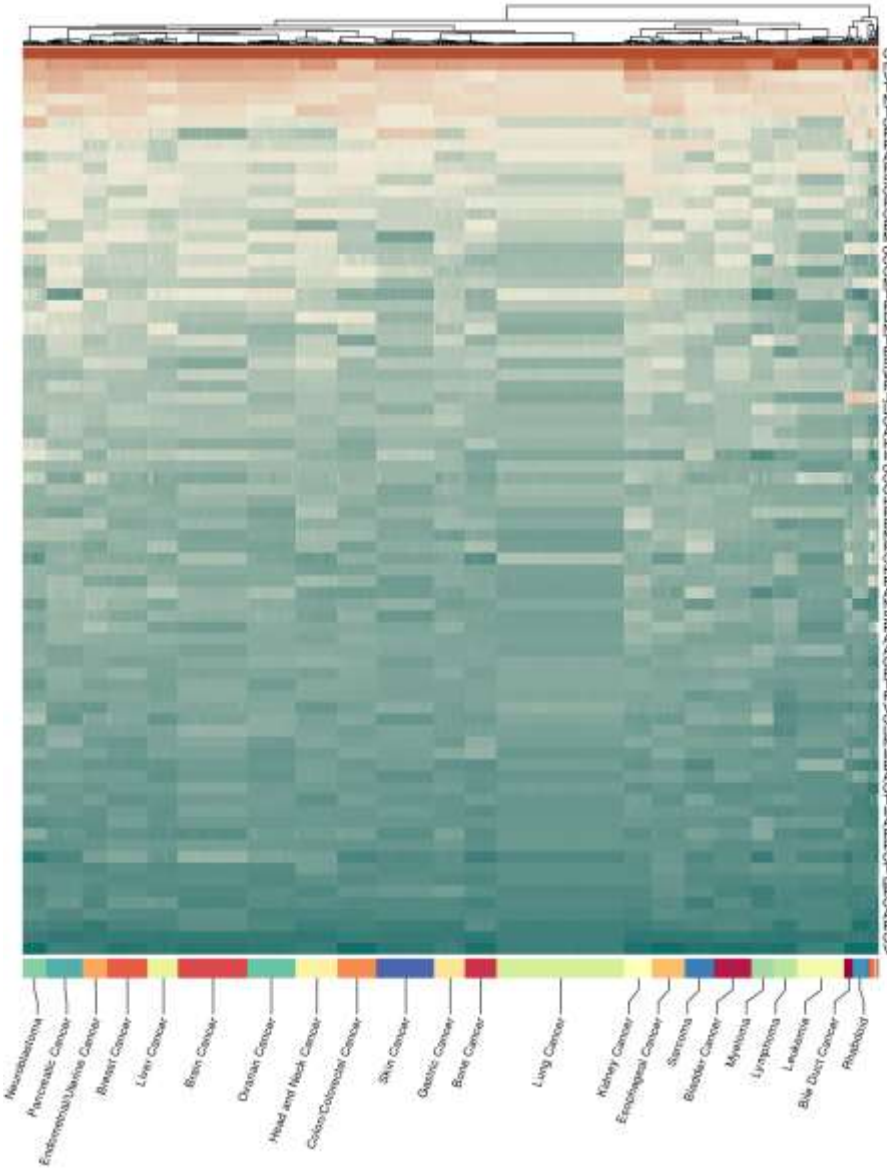




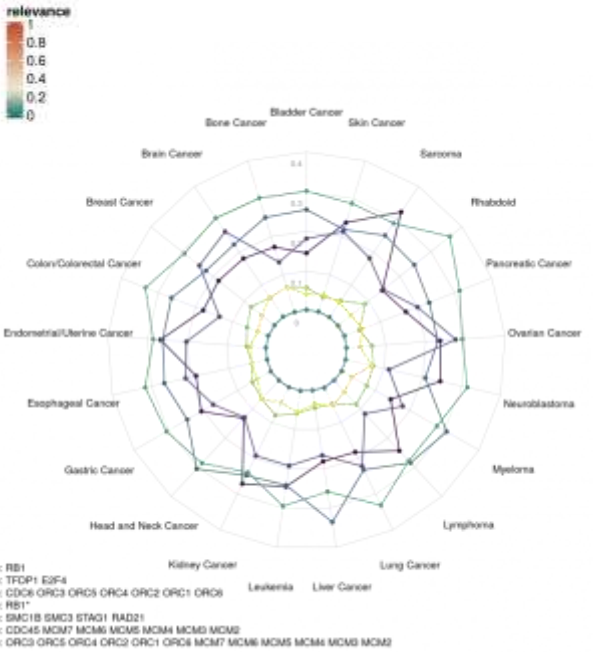
The predictor trained with Achilles v1 cell lines predict the effect or KOs in Achilles v2 cell lines

NB: essential genes are not conserved across cell lines

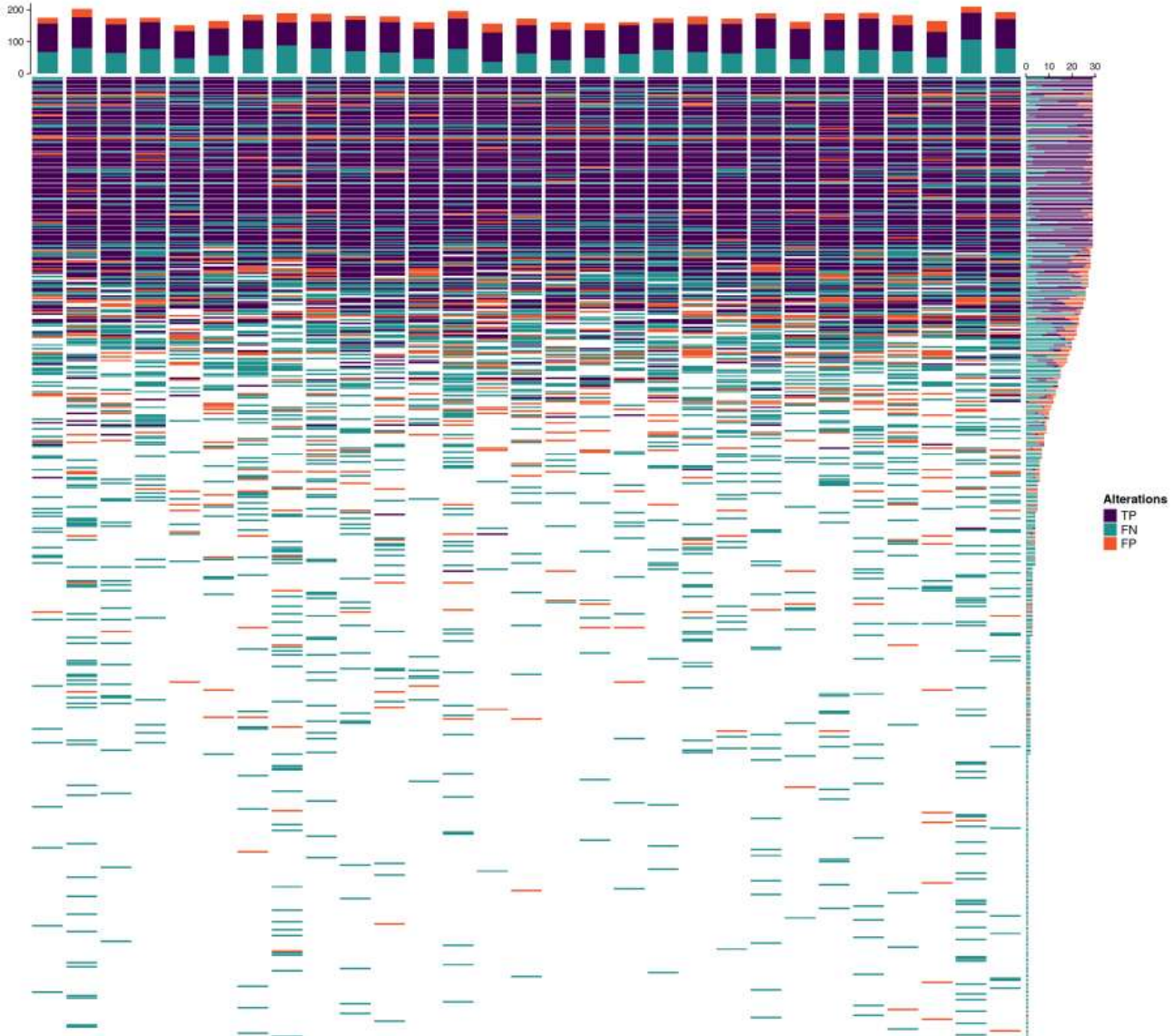
Most relevant circuits for the predictor



SHAP is used for obtaining relevancies.
As expected, the cell cycle pathway is the most relevant in defining cell survival.

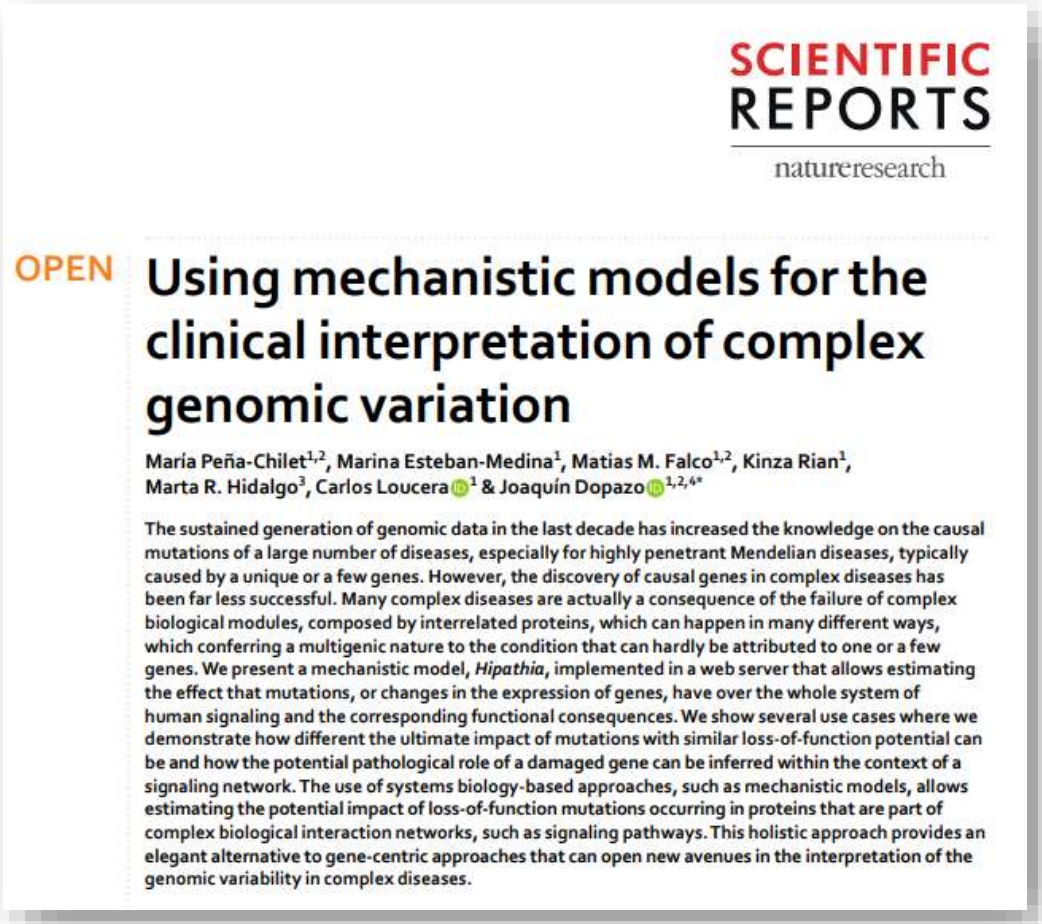


Oncoprint of breast cancer



Different breast cancer cell lines in which true positives (blue), false negatives (green) and false positives (red) are depicted. The number of true positives among the first genes ranked by predicted lethality is considerable.

Mutations as perturbations



SCIENTIFIC REPORTS
nature research

OPEN

Using mechanistic models for the clinical interpretation of complex genomic variation

María Peña-Chilet[1,2], Marina Esteban-Medina[1], Matías M. Falco[1,2], Kinza Rian[1], Marta R. Hidalgo[3], Carlos Loucera[1] & Joaquín Dopazo[1,2,4*]

The sustained generation of genomic data in the last decade has increased the knowledge on the causal mutations of a large number of diseases, especially for highly penetrant Mendelian diseases, typically caused by a unique or a few genes. However, the discovery of causal genes in complex diseases has been far less successful. Many complex diseases are actually a consequence of the failure of complex biological modules, composed by interrelated proteins, which can happen in many different ways, which conferring a multigenic nature to the condition that can hardly be attributed to one or a few genes. We present a mechanistic model, *Hipathia*, implemented in a web server that allows estimating the effect that mutations, or changes in the expression of genes, have over the whole system of human signaling and the corresponding functional consequences. We show several use cases where we demonstrate how different the ultimate impact of mutations with similar loss-of-function potential can be and how the potential pathological role of a damaged gene can be inferred within the context of a signaling network. The use of systems biology-based approaches, such as mechanistic models, allows estimating the potential impact of loss-of-function mutations occurring in proteins that are part of complex biological interaction networks, such as signaling pathways. This holistic approach provides an elegant alternative to gene-centric approaches that can open new avenues in the interpretation of the genomic variability in complex diseases.

Mutations can be understood as perturbations of the system and the consequences can be predicted by the model. Some LoF mutations can have irrelevant or null effects and other can affect to many critical cell functions, depending on the relationship of the mutated protein with other proteins and their activity

Understanding pathologic variation in complex diseases (an example with diabetes)

Cell Metabolism

Resource

A Systems Genetics Approach Identifies Genes and Pathways for Type 2 Diabetes in Human Islets

Jalal Taneera,[1,8,*] Stefan Lang,[1,8] Amitabh Sharma,[1,2,8] Joao Fadista,[1] Yuedan Zhou,[1] Emma Ahlqvist,[1] Anna Jonsson,[1] Valeriya Lyssenko,[1] Petter Vikman,[1] Ola Hansson,[1] Hemang Parikh,[3] Olle Korsgren,[2] Arvind Soni,[3] Ulrika Krus,[1] Emming Zhang,[3] Xing-Jun Jing,[3] Jonathan L.S. Esguerra,[4] Claes B. Wollheim,[1] Albert Salehi,[2] Anders Rosengren,[1,5] Erik Renström,[3] and Leif Groop[6,7]

[1]Lund University Diabetes Center, Department of Clinical Sciences, Diabetes and Endocrinology

[2]Department of Clinical Sciences, Islet Cell Physiology

[3]Department of Clinical Sciences, Islet Pathophysiology

[4]Department of Clinical Sciences, Islet Cell Exocytosis

Skåne University Hospital Malmö, Lund University, Malmö 20502, Sweden

[5]Center for Complex Network Research and Department of Physics, Northeastern University, Boston, MA 02115, USA

[6]Laboratory of Translational Genomics, Division of Cancer Epidemiology and Genetics, National Cancer Institute, National Institutes of Health, Bethesda, MD 20877, USA

[7]Institute of Immunology, Genetics and Pathology, Rudbeckilaboratoriet, Uppsala University, Uppsala 75185, Sweden

[8]These authors contributed equally to this work

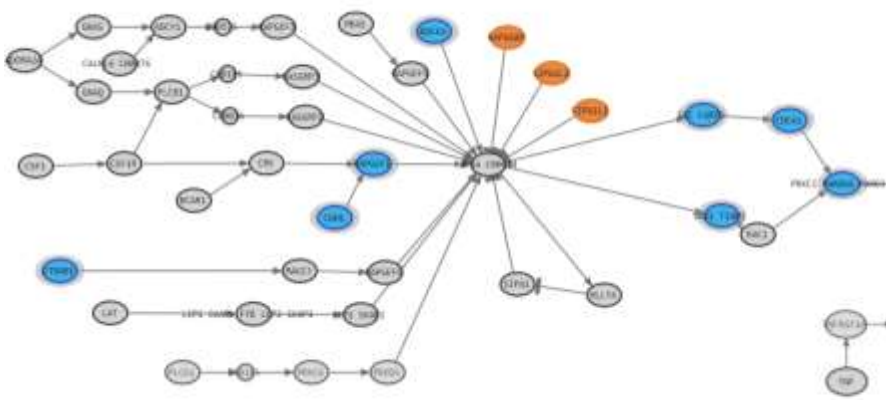
[*]Correspondence: jalal.taneera@med.lu.se (J.T.), leif.groop@med.lu.se (L.G.)

http://dx.doi.org/10.1016/j.cmet.2012.06.006

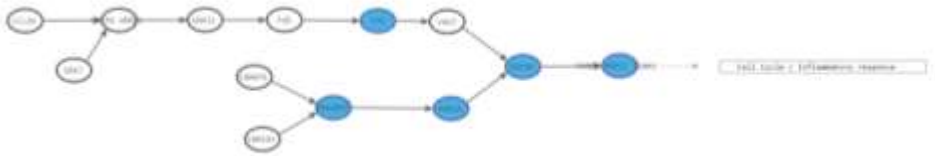
A total of 30 signaling circuits trigger inflammation. We used data on T2D to focus on those circuits specifically related with the disease

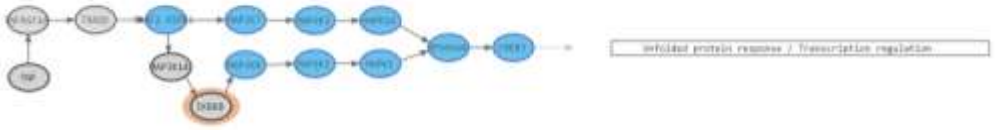
Circuit	Log Fold Change	t	P-value	FDR-adj. p-value	Function (Uniprot)
Rap1 signaling pathway: PRKCI PARD6A PARD3	0.001003	3.83581	0.00029	0.00892	Inflammatory response, Cell cycle, Cell division, Differentiation
NF-kappa B signaling pathway: CCL19	0.01209	3.37450	0.00128	0.01676	Inflammatory response. Chemotaxis
NF-kappa B signaling pathway: CCL21	0.02025	3.28701	0.00167	0.01676	Inflammatory response. Chemotaxis

Rap1 signaling pathway: PRKCI PARD6A PARD3



NF-kappa B signaling pathway: CCL19

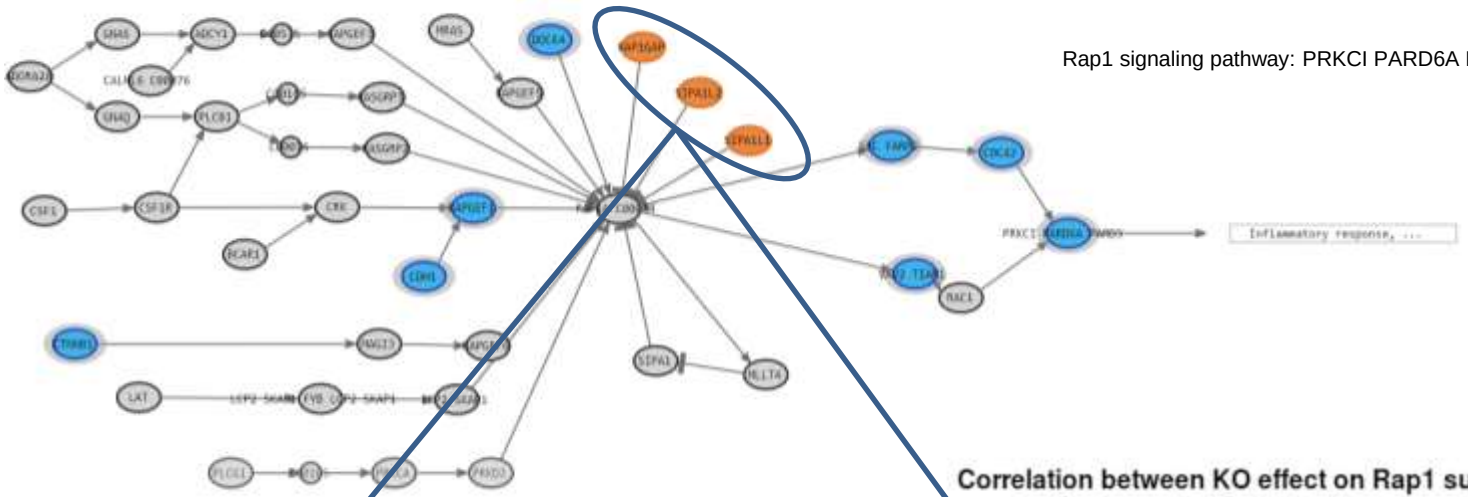


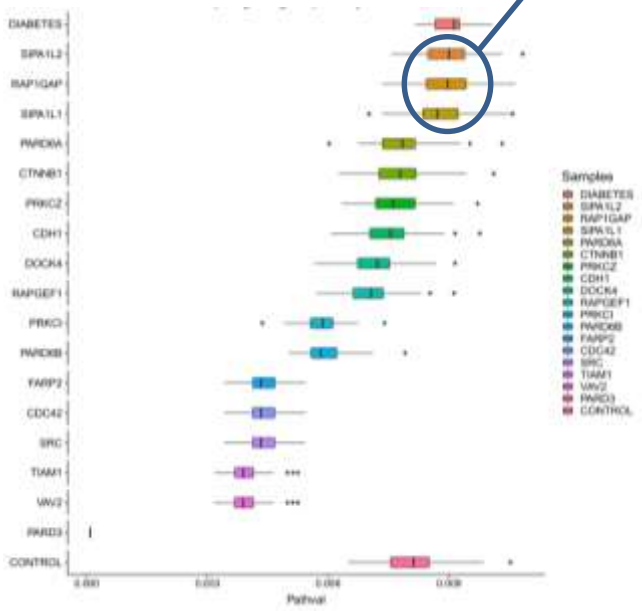


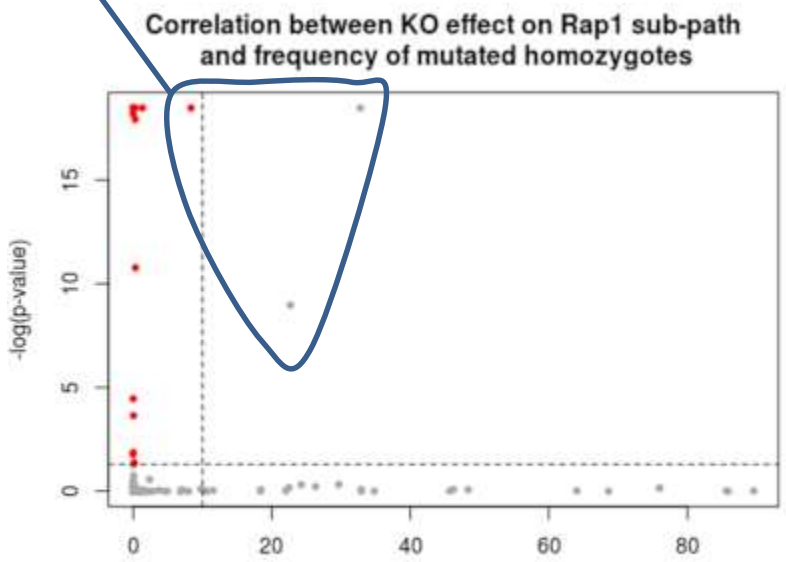
NF-kappa B signaling pathway: CCL21

LoF mutations with an effect on signaling similar to the disease

Rap1 signaling pathway: PRKCI PARD6A PARD3

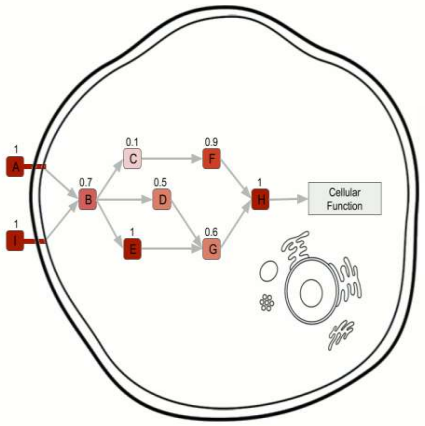


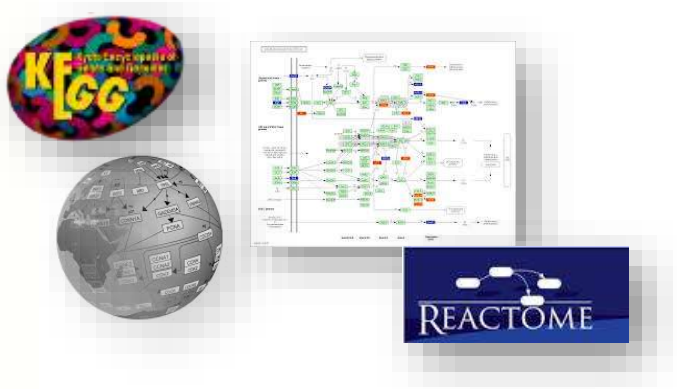




Terra incognita: beyond the modelizable part of the genome

Mechanistic models rely on **biological knowledge**: pathways that describe the functional interplay of molecules in the cell, and how cells



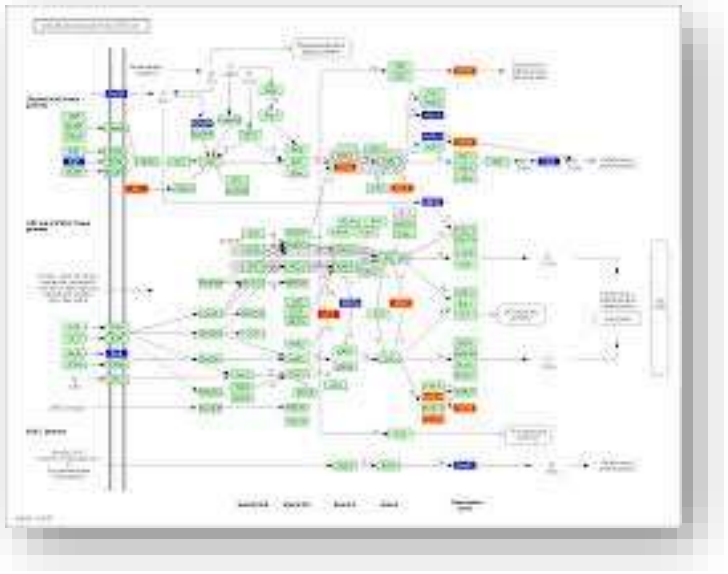


Currently, only one third of the genome can be modelled

Causal relationships gene activity/integrity to effect cannot be modelled and extended beyond all the pathway information available

The obtention of biological knowledge is a slow process

Pathways represent the current biological knowledge with arrows, that account for functional relationships, which connect nodes, that represent molecules (proteins or others)



The generation of these “arrows” requires years of laboratory work by formulating and testing specific hypothesis on particular relationships between molecules.

But… would it be possible to use machine learning to generate biological knowledge from data?

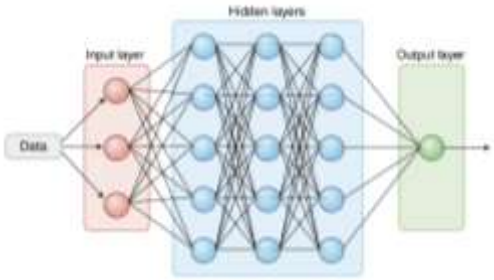
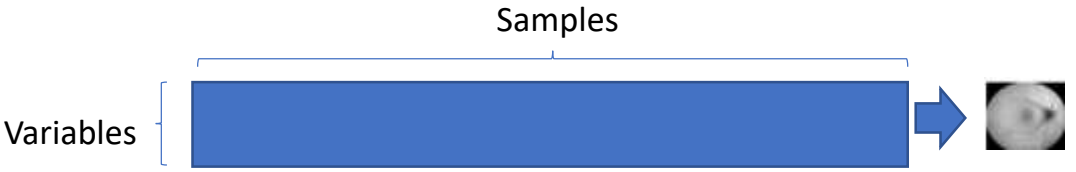
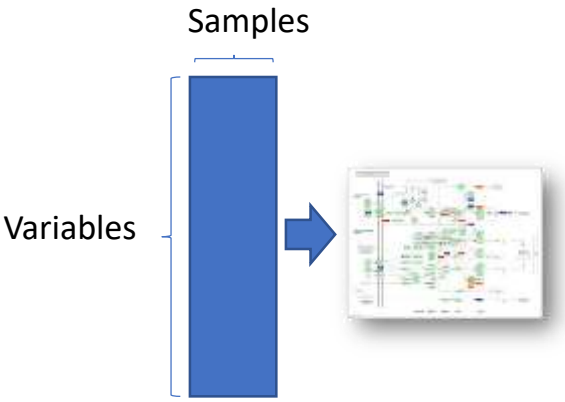


Table 2 | FDA AI approvals are accelerating

Company	FDA Approval	Indication
Apple	September 2018	Atrial fibrillation detection
Aidoc	August 2018	CT brain bleed diagnosis
iCAD	August 2018	Breast density via mammography
Zebra Medical	July 2018	Coronary calcium scoring
Bay Labs	June 2018	Echocardiogram EF determination
Neural Analytics	May 2018	Device for paramedic stroke diagnosis
IDx	April 2018	Diabetic retinopathy diagnosis
Icometrix	April 2018	MRI brain interpretation
Imagen	March 2018	X-ray wrist fracture diagnosis
Viz.ai	February 2018	CT stroke diagnosis
Arterys	February 2018	Liver and lung cancer (MRI, CT) diagnosis
MaxQ-AI	January 2018	CT brain bleed diagnosis
Alivecor	November 2017	Atrial fibrillation detection via Apple Watch
Arterys	January 2017	MRI heart interpretation

Topol, 2019, Nat. Med.

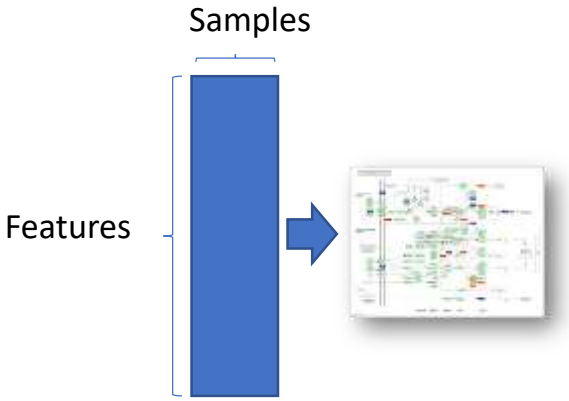




Curse of dimensionality

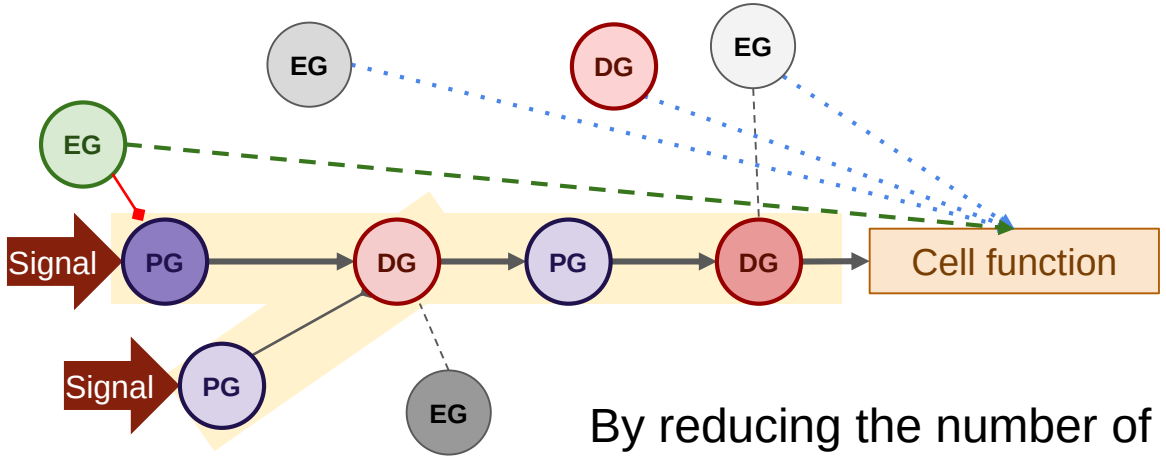
Learning biological knowledge from the data is currently quite complex. New methods for feature selection, dimensionality reduction, multi-view learning and network learning need to be developed.

Learning biological knowledge from the data



Curse of dimensionality
Learning biological knowledge from the data is currently quite complex.
Knowledge-based dimensionality reduction

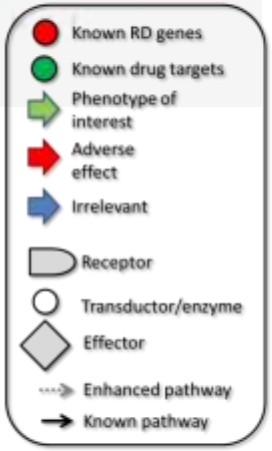
Learning biology from the data is (at least partly) possible

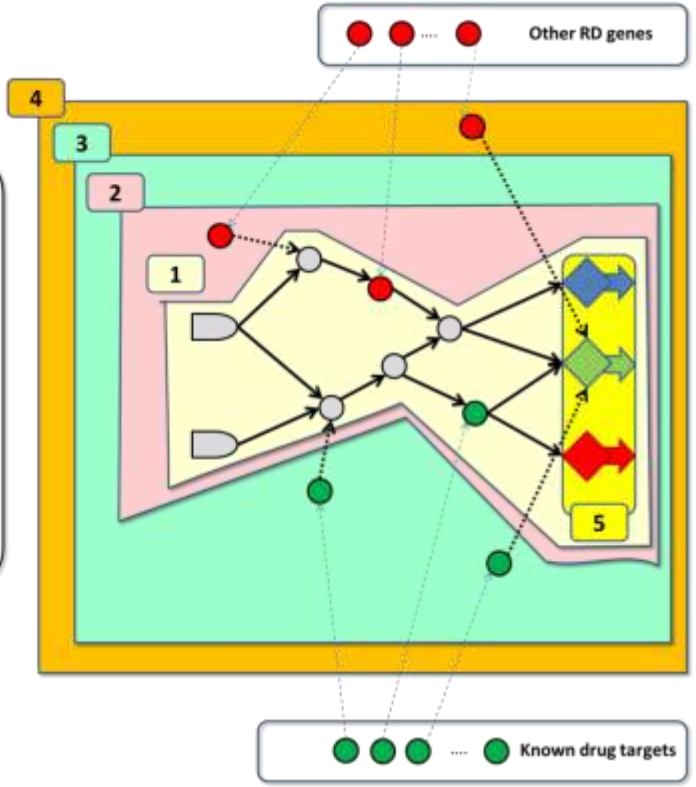


By reducing the number of potential causal relationships other more sophisticated models for causality or simply experimental validations are affordable

Systematic drug repositioning in Rare Diseases (Project funded by FBBVA)

…which can also have complex Disease Mechanisms



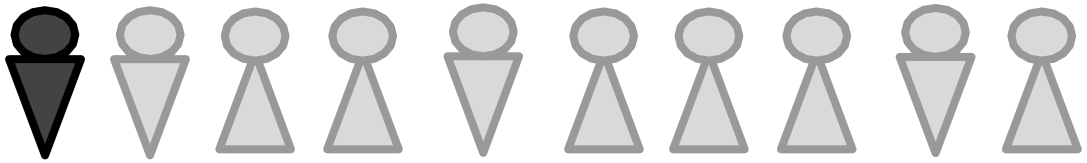


ML (*Sparse Variational Gausian Process* - SVGP) is used to detect the influence of target of known drugs over functions that we identify as hallmarks of the RD and that we can model with mechanistic models

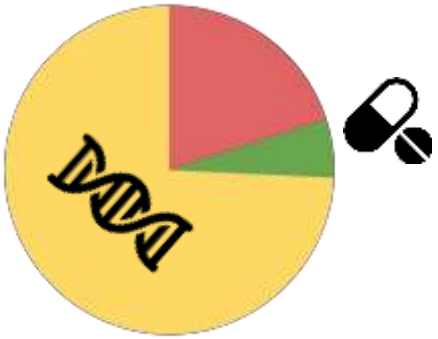
Why repurposing drugs in rare diseases?

RDs affect to less than **1 among 2000** individuals.

There are more than **7000 different RDs**.
Globally, RDs affect to **6-8%** of the European population (1 out of 12).



80% genetic basis.



Only treatments available for **400 RDs**.

Pharma companies do not invest in RDs (very small, heterogeneous and fragmented market where investments are difficult to recover

Advantages of drug repurposing:

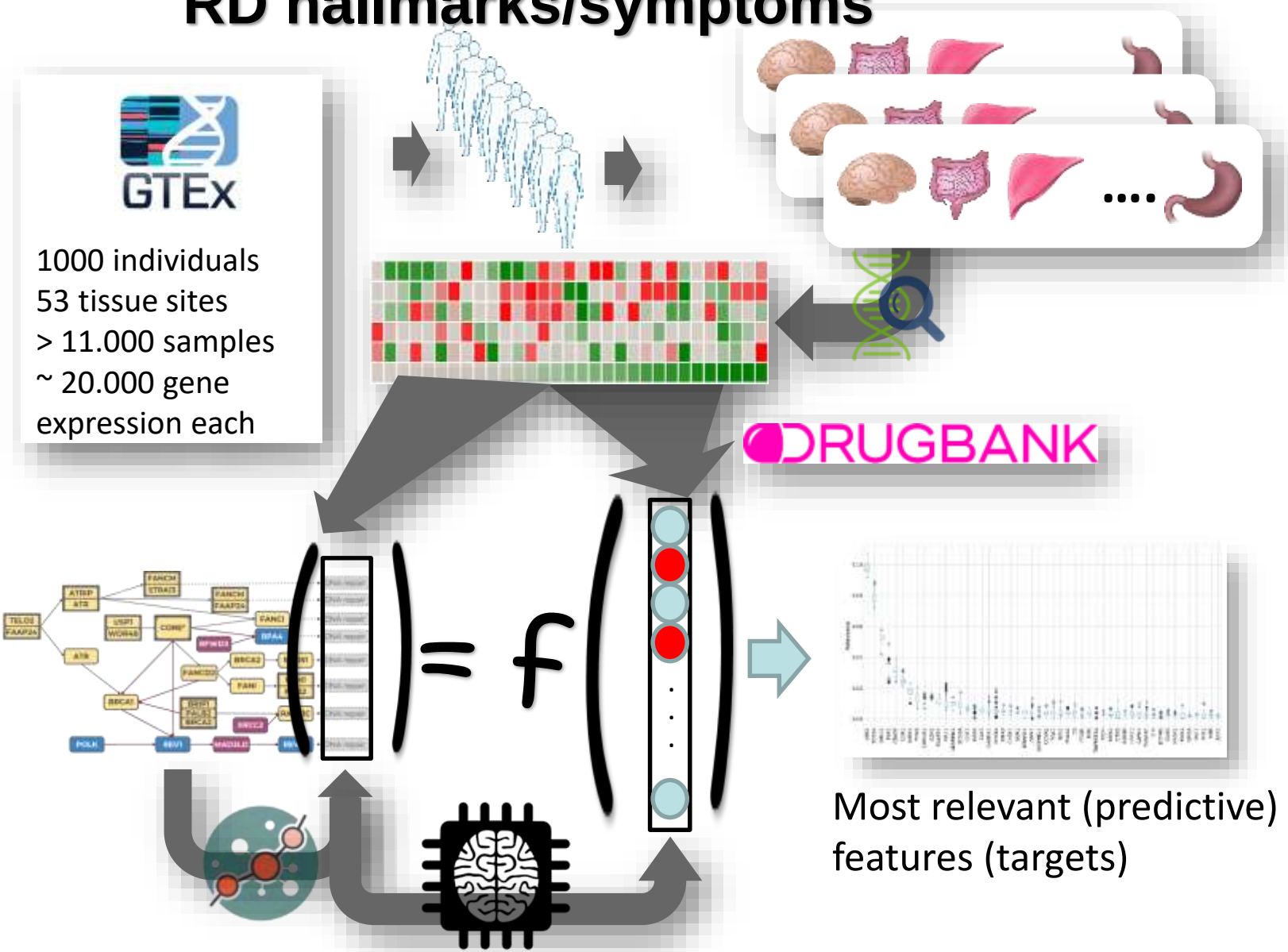
Already registered

Known security profile

Known mechanism of action

Drug repositioning in Rare diseases

Learning from the data what target(s) affect(s) to RD hallmarks/symptoms



Biological knowledge can systematically be learned from data

- Deriving comprehensive disease pathways and building from them mechanistic models for a subset of about 100 RDs that have at least 3 genes within known signaling and metabolic pathways
- Finding therapeutic targets that revert the disease hallmarks modeled to the healthy status or that alleviate disease symptoms and, among these we will pay especial attention to genes that are already targets of drugs in other diseases.

So far Fanconi anemia, Juvenile arthritis and familiar melanoma have been modeled.

Aim:
Demonstrate that ML can help to generate biological knowledge in an "industrial" manner.

Esteban-Medina et al. BMC Bioinformatics (2019) 20:370
https://doi.org/10.1186/s12859-019-2969-0

BMC Bioinformatics

RESEARCH ARTICLE Open Access

Exploring the druggable space around the Fanconi anemia pathway using machine learning and mechanistic models

Marina Esteban-Medina[1], María Peña-Chilet[1,2], Carlos Loucera[1] and Joaquín Dopazo[1,2,3*]

Abstract

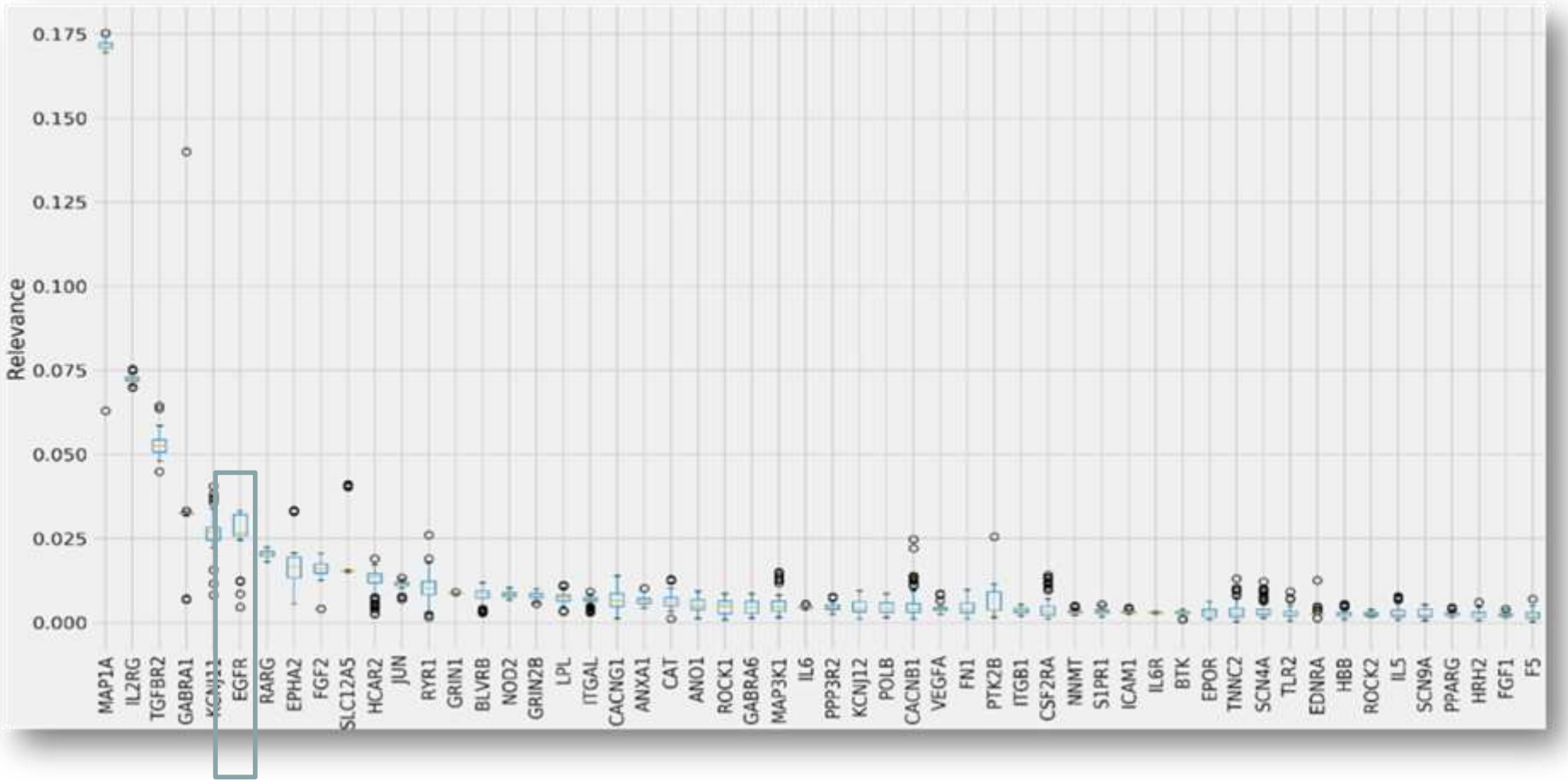
Background: In spite of the abundance of genomic data, predictive models that describe phenotypes as a function of gene expression or mutations are difficult to obtain because they are affected by the curse of dimensionality, given the disbalance between samples and candidate genes. And this is especially dramatic in scenarios in which the availability of samples is difficult, such as the case of rare diseases.

Results: The application of multi-output regression machine learning methodologies to predict the potential effect of external proteins over the signaling circuits that trigger Fanconi anemia related cell functionalities, inferred with a mechanistic model, allowed us to detect over 20 potential therapeutic targets.

Conclusions: The use of artificial intelligence methods for the prediction of potentially causal relationships between proteins of interest and cell activities related with disease-related phenotypes opens promising avenues for the systematic search of new targets in rare diseases.

Keywords: Genomics, Big data, Machine learning, Fanconi anemia, Signaling pathways, Mathematical models

Top therapeutic target predicted



EGFR the epidermal growht factor receptor gene

Two drugs predicted as repurposable were validated

BMC Bioinformatics

RESEARCH ARTICLE Open Access

Exploring the druggable space around the Fanconi anemia pathway using machine learning and mechanistic models

Marina Esteban-Medina[1], María Peña-Chilet[1,2], Carlos Loucera[1] and Joaquín Dopazo[1,2,3*]

Abstract

Background: In spite of the abundance of genomic data, predictive models that describe phenotypes as a function of gene expression or mutations are difficult to obtain because they are affected by the curse of dimensionality, given the disbalance between samples and candidate genes. And this is especially dramatic in scenarios in which the availability of samples is difficult, such as the case of rare diseases.

Results: The application of multi-output regression machine learning methodologies to predict the potential effect of external proteins over the signaling circuits that trigger Fanconi anemia related cell functionalities, inferred with a mechanistic model, allowed us to detect over 20 potential therapeutic targets.

Conclusions: The use of artificial intelligence methods for the prediction of potentially causal relationships between proteins of interest and cell activities related with disease-related phenotypes opens promising avenues for the systematic search of new targets in rare diseases.

Keywords: Genomics, Big data, Machine learning, Fanconi anemia, Signaling pathways, Mathematical models

CLINICAL CANCER RESEARCH

Home About Articles For Authors Alerts News Search Q

Translational Cancer Mechanisms and Therapy

Gefitinib and Afatinib Show Potential Efficacy for Fanconi Anemia–Related Head and Neck Cancer

Helena Montaxay, Águeda Martínez-Barriocanal, José Antonio Casado, Lloreç Rovirosa, María José Ramírez, Rocío Nieto, Carlos Carrascoso-Rubio, Pau Riera, Alan González, Enrique Lerma, Adriana Lasa, Jordi Carreras-Puigvert, Thomas Helleday, Juan A. Bueren, Diego Arango, Jordi Mingullón, and Jordi Surrallés

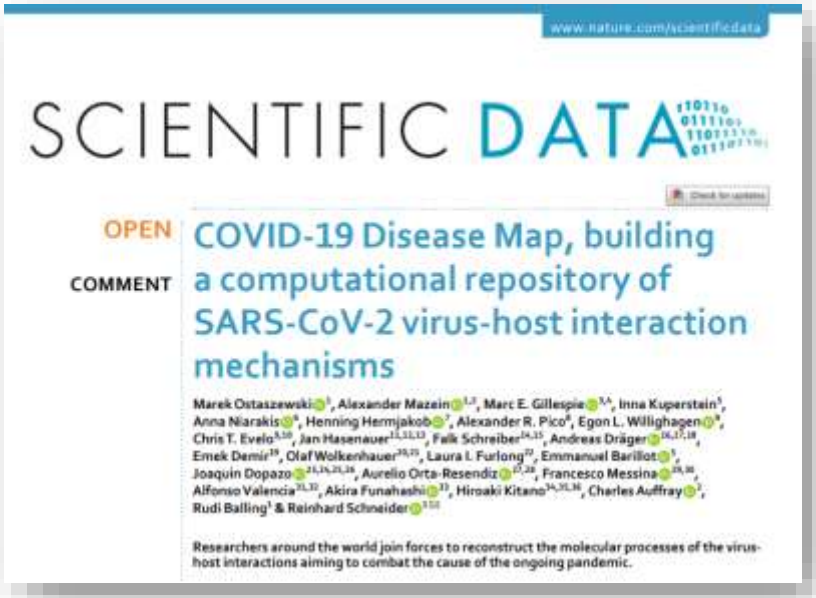
DOI: 10.1158/1078-0432.CCR-19-1625

Article Figures & Data Info & Metrics PDF

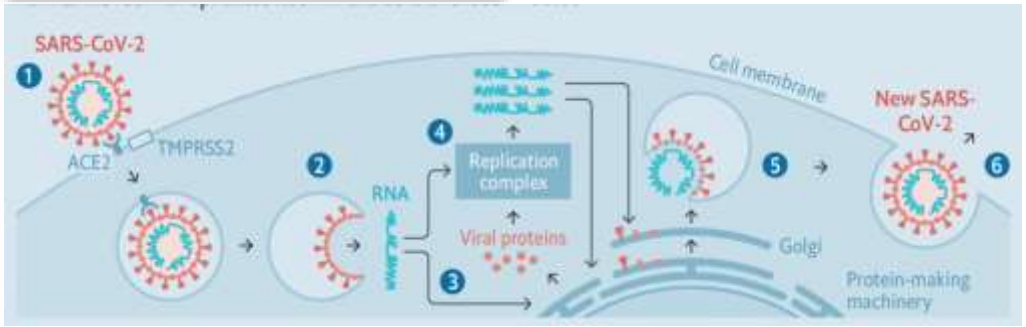
Published OnlineFirst April 8, 2020

DRUG ID	DRUG NAME	SYMBOL	ACTIONS
DB01196	Estramustine	MAP1A	antagonist
DB00041	Aldesleukin	IL2RG	agonist
DB10772	Foreskin keratinocyte (neonatal)	TGFBR2	agonist
DB00186	Lorazepam	GABRA1	positive allosteric modulator
DB00228	Enflurane	GABRA1	positive allosteric modulator
DB00317	Gefitinib	EGFR	antagonist
DB08916	Afatinib	EGFR	antagonist
DB01259	Lapatinib	EGFR	antagonist
	Tazarotene	RARG	agonist
	Dasatinib	EPHA2	antagonist
	Regorafenib	EPHA2	inhibitor
	Sucralfate	FGF2	agonist, inducer
	Niacin	HCAR2	agonist

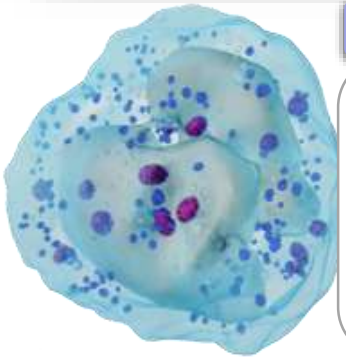
The Disease maps consortium: expanding drug repurposing to COVID19



SARS-CoV-2 ENTRY



IMMUNE RESPONSE



Innate immune response

- Delayed or supressed type I IFN response.
- Hyperinflammatory response and cytokine storm
- Influx of activated neutrophils and inflammatory macrophages.

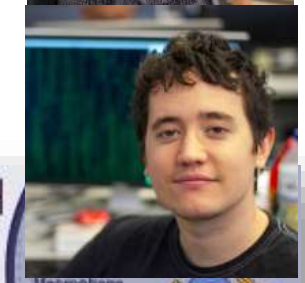
Adaptive immune response

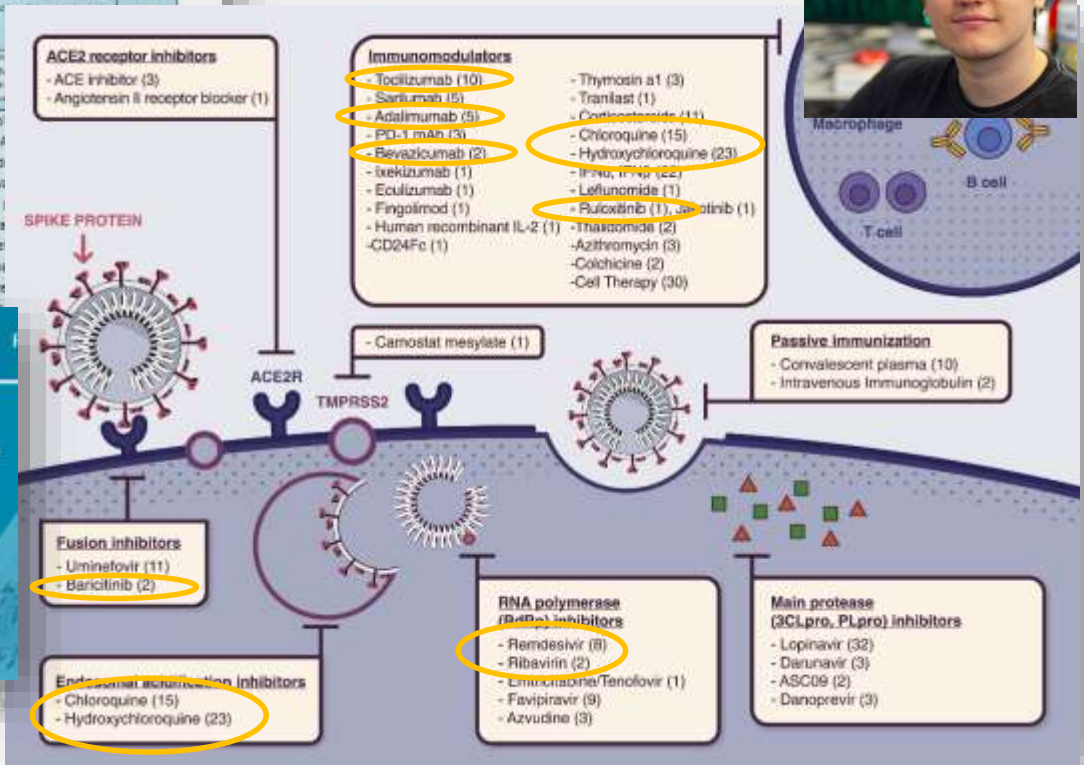
- T helpers cells Th1/Th17 are engage.
- IgA, IgM and IgG are detectable within 2 weeks after infection.
- Lymphopenia may be related to bone marrow suppression.

The COVID-19 druggable space



Marina Esteban-Medina
María Peña-Chilet
Carlos Loucera
Kinza Rian
Matias M Falco

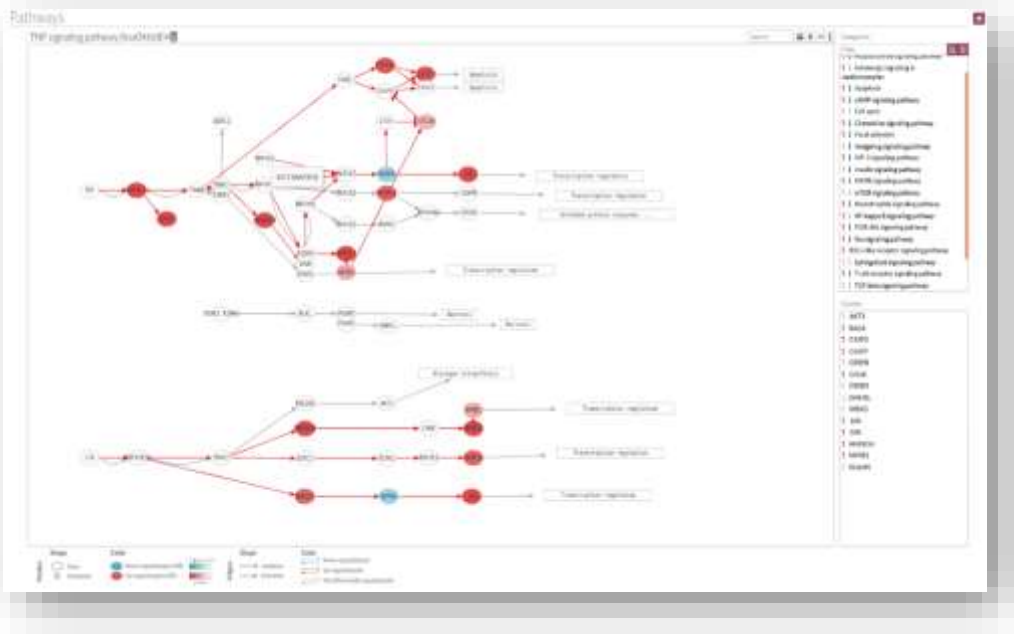


A version of hipathia specific for the COVID-19 Disease map

CoV-HiPathia
Covid19 pathway interpretation and analysis

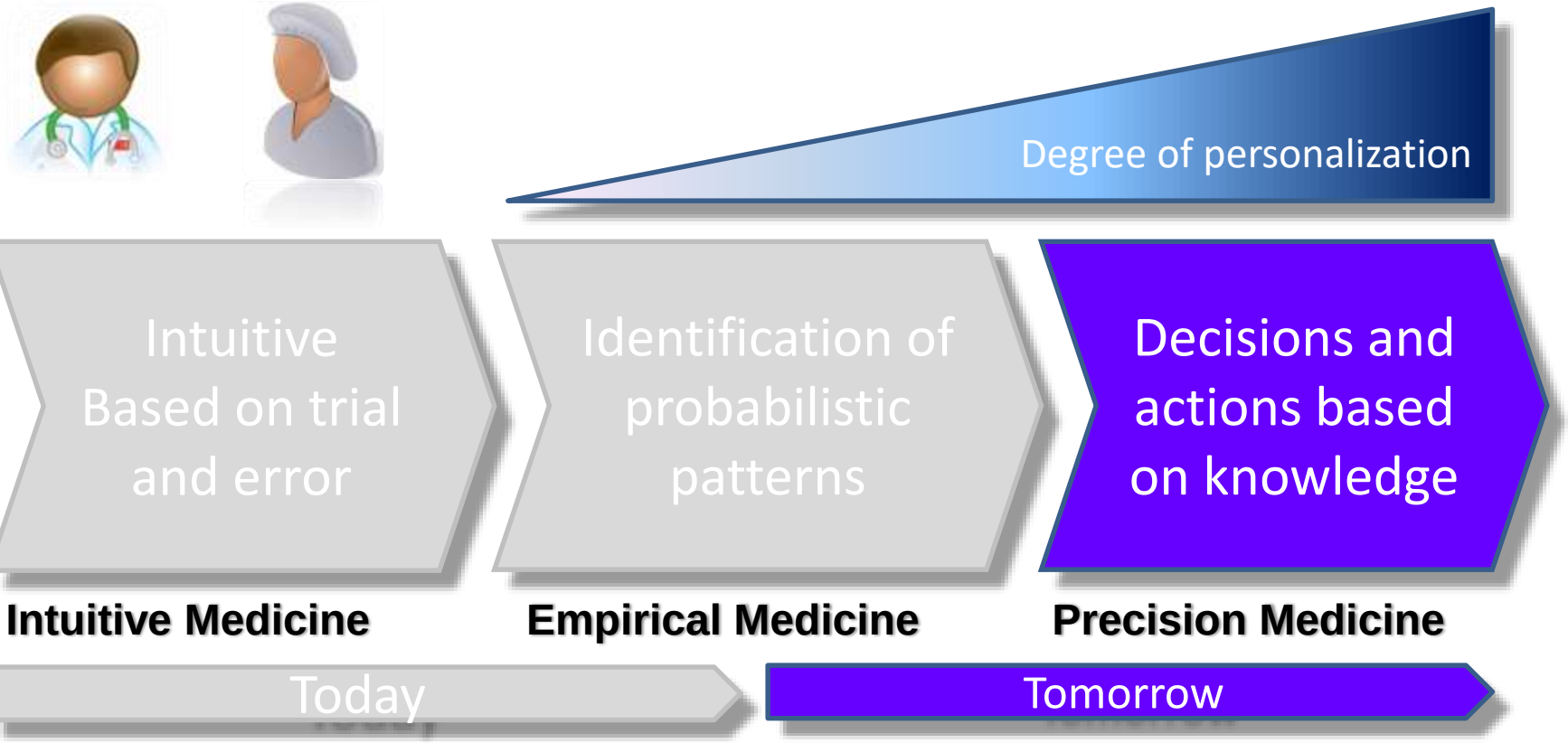
MAIN FEATURES

- An interactive webtool for mechanistic pathway analysis of signaling gene networks related to COVID-19
- High-throughput estimation of functional cell activities
- Accessible and interactive web tool
- *in silico* simulations on transcriptomics data
- Analysis of annotated Activity Flow maps related to COVID-19 to predict clinical outcomes.



http://hipathia.babelomics.org/covid19/

The real transition to precision medicine

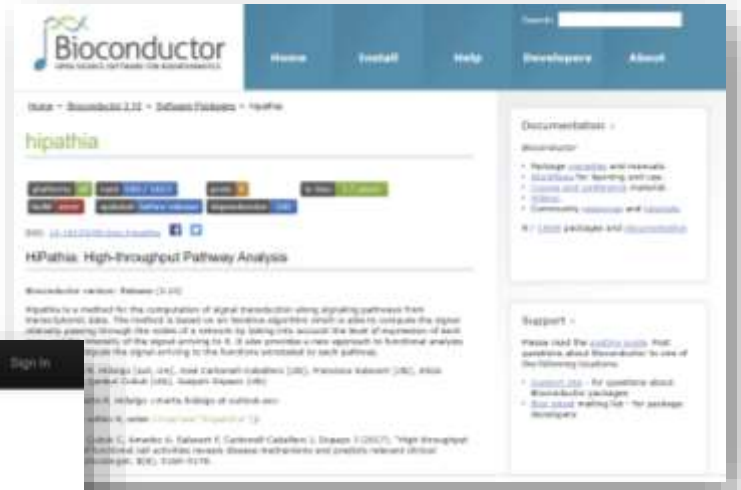


The use of new algorithms that enable the transformation of genomic measurements into cell functionality measurements that account for disease mechanisms and for drug mechanisms of action will ultimately allow the real transition from today's empirical medicine to precision medicine and provide an increasingly personalized medicine

The hipathia algorithm is available as a web interface, as well as bioconductor and cytoscape applications

http://hipathia.babelomics.org/

https://bioconductor.org/packages/release/bioc/html/hipathia.html



http://apps.cytoscape.org/apps/cypathia

Clinical Bioinformatics Area
Fundación Progreso y Salud, Sevilla, Spain, and Instituto de Biomedicina de Sevilla
and the FPS/ELIXIR-ES and the BiER (CIBERER)

https://www.slideshare.net/xdopazo/

Follow us on twitter
@xdopazo
@ClinicalBioinfo