Análisis de genomas, RNA-seq y otras omicas

Joaquín Dopazo

Clinical Bioinformatics Research Area,
Fundación Progreso y Salud,
Functional Genomics Node, (INB-ELIXIR-es),
Bioinformatics in Rare Diseases (BiER-CIBERER),
Sevilla, Spain.

http://www.clinbioinfosspa.es/
http://www.babelomics.org

@xdopazo @ClinicalBioinfo

Máster en Bioinformática Aplicada a Medicina Personalizada y Salud, Madrid, 11 abril, 2018

Motivation

Progress in science depends on new techniques, new discoveries and new ideas, probably in that order.

(Sydney Brenner, Nobel Prize Physiology or Medicine 1980)

The introduction and popularisation of high-throughput techniques has drastically changed the way in which biological problems **can** be addressed and hypotheses **can** be tested. (although not necessarily the way in which we really address or test them)

Omics technologies have a major impact in Medicine

Background:
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** **Precision Medicine**

Today

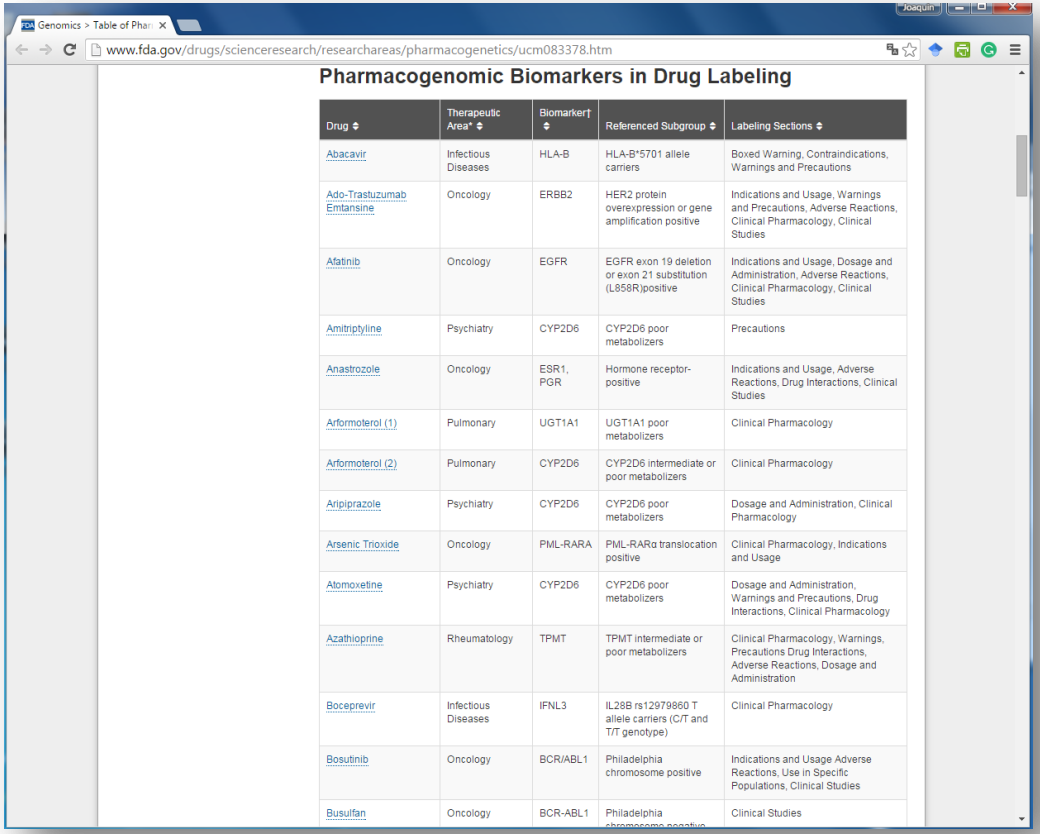
Tomorrow

Molecular biomarkers

Genomic biomarkers

Precision medicine is based on a better knowledge of phenotype-genotype relationships. That is the knowledge od **disease** and **drug action mechanisms**
Requires of a better way of defining diseases by introducing **genomic** technologies in the **diagnostic** procedures and **treatment decisions**

And how do we identify patterns? Using single-gene biomarkers



Pharmacogenomic Biomarkers in Drug Labeling

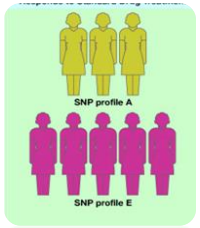
Drug	Therapeutic Area*	Biomarker†	Referenced Subgroup	Labeling Sections
Abacavir	Infectious Diseases	HLA-B	HLA-B*5701 allele carriers	Boxed Warning, Contraindications, Warnings and Precautions
Ado-Trastuzumab Emtansine	Oncology	ERBB2	HER2 protein overexpression or gene amplification positive	Indications and Usage, Warnings and Precautions, Adverse Reactions, Clinical Pharmacology, Clinical Studies
Afatinib	Oncology	EGFR	EGFR exon 19 deletion or exon 21 substitution (L858R)positive	Indications and Usage, Dosage and Administration, Adverse Reactions, Clinical Pharmacology, Clinical Studies
Amitriptyline	Psychiatry	CYP2D6	CYP2D6 poor metabolizers	Precautions
Anastrozole	Oncology	ESR1, PGR	Hormone receptor-positive	Indications and Usage, Adverse Reactions, Drug Interactions, Clinical Studies
Arformoterol (1)	Pulmonary	UGT1A1	UGT1A1 poor metabolizers	Clinical Pharmacology
Arformoterol (2)	Pulmonary	CYP2D6	CYP2D6 intermediate or poor metabolizers	Clinical Pharmacology
Aripiprazole	Psychiatry	CYP2D6	CYP2D6 poor metabolizers	Dosage and Administration, Clinical Pharmacology
Arsenic Trioxide	Oncology	PML-RARA	PML-RARα translocation positive	Clinical Pharmacology, Indications and Usage
Atomoxetine	Psychiatry	CYP2D6	CYP2D6 poor metabolizers	Dosage and Administration, Warnings and Precautions, Drug Interactions, Clinical Pharmacology
Azathioprine	Rheumatology	TPMT	TPMT intermediate or poor metabolizers	Clinical Pharmacology, Warnings, Precautions Drug Interactions, Adverse Reactions, Dosage and Administration
Boceprevir	Infectious Diseases	IFNL3	IL28B rs12979860 T allele carriers (C/T and T/T genotype)	Clinical Pharmacology
Bosutinib	Oncology	BCR/ABL1	Philadelphia chromosome positive	Indications and Usage Adverse Reactions, Use in Specific Populations, Clinical Studies
Busulfan	Oncology	BCR-ABL1	Philadelphia	Clinical Studies

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

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

And.. where biomarkers come from?

Genes in the DNA…

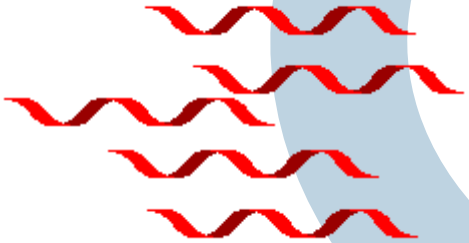


…produces the final phenotype

Integrity

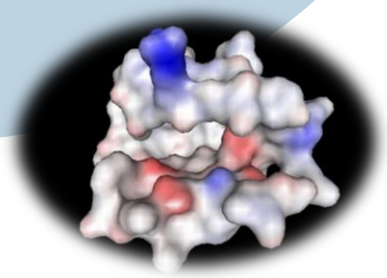
From genotype to phenotype.

…plus the environment…



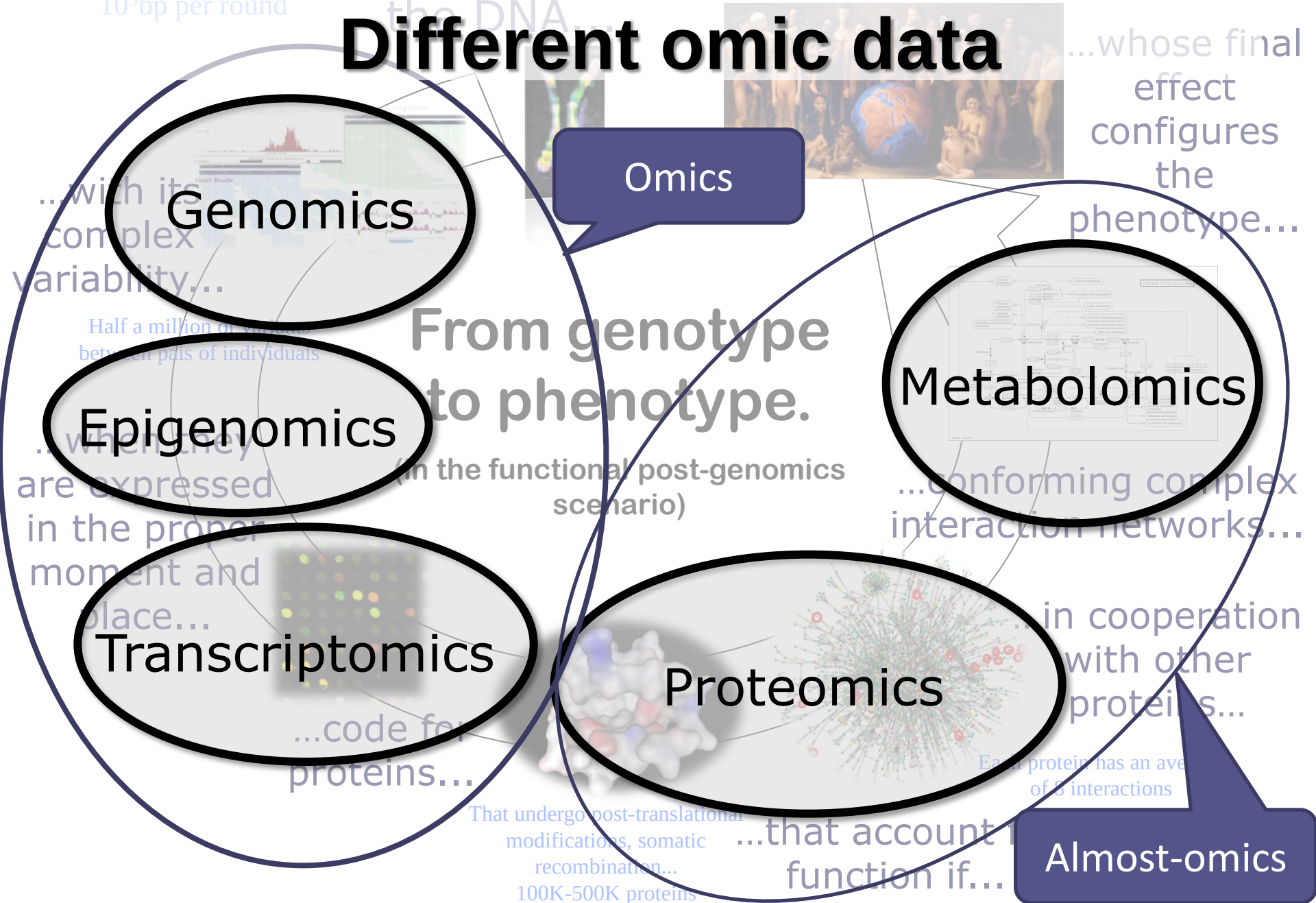
…are transcribed and…

…code for proteins…

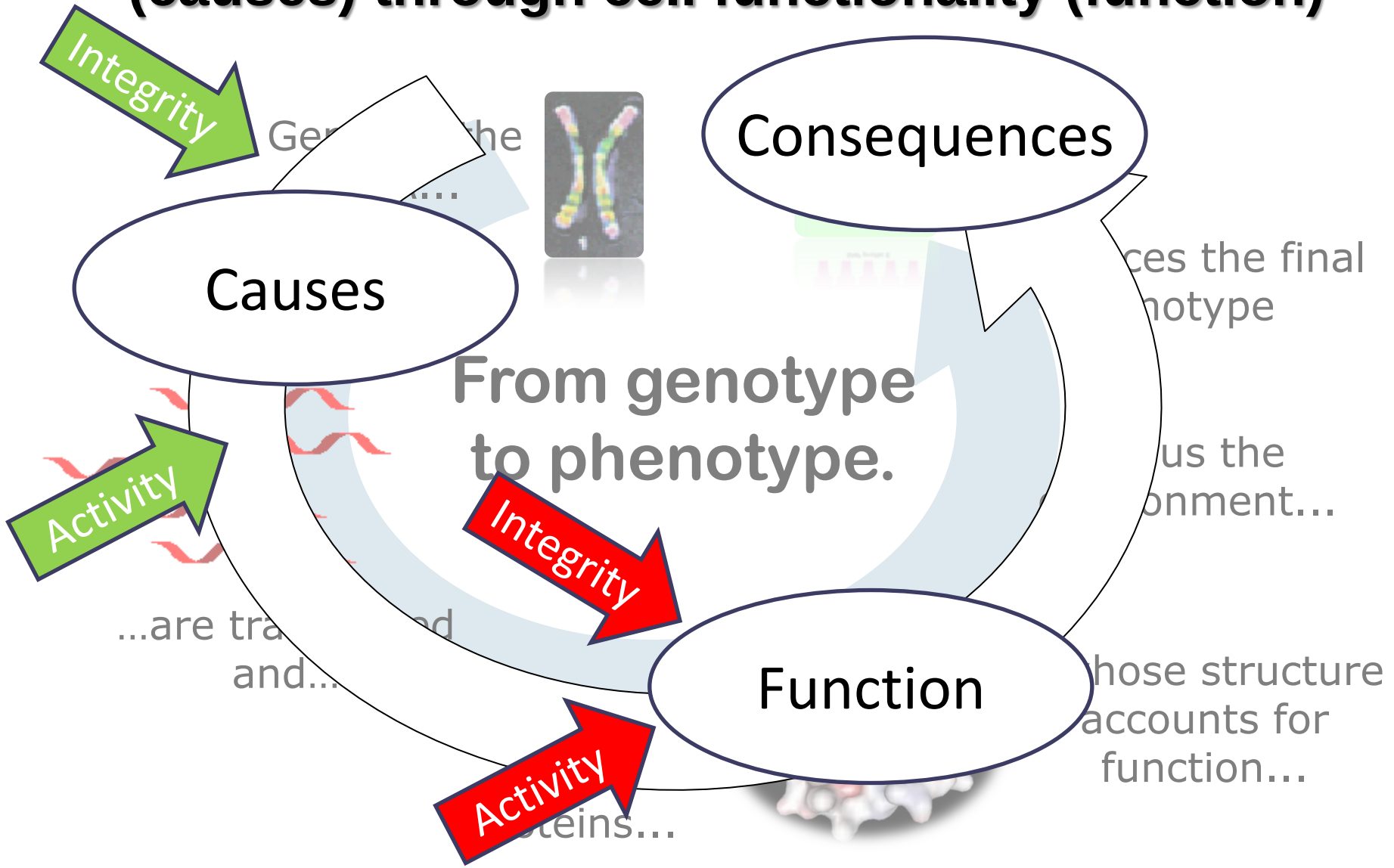


…whose structure accounts for function…

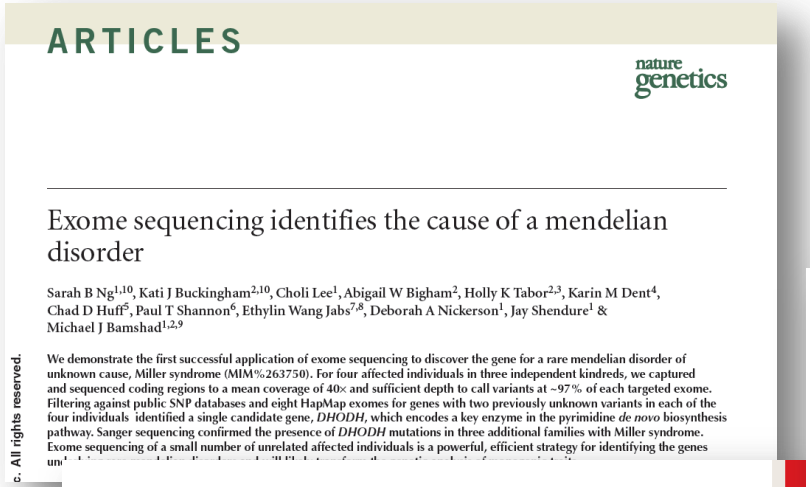
Our observational tools: Different omic data



Reductionist approach to explain phenotype (consequences) from gene integrity and activity (causes) through cell functionality (function)



Exome sequencing has been systematically used to identify Mendelian disease genes



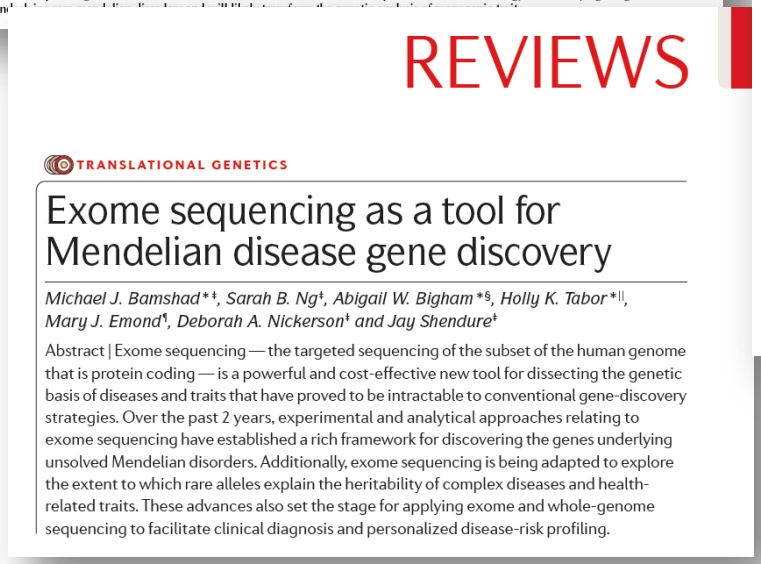
ARTICLES

nature genetics

Exome sequencing identifies the cause of a mendelian disorder

Sarah B Ng[1,10], Kati J Buckingham[2,10], Choli Lee[1], Abigail W Bigham[2], Holly K Tabor[2,3], Karin M Dent[4], Chad D Huff[5], Paul T Shannon[6], Ethylin Wang Jabs[7,8], Deborah A Nickerson[1], Jay Shendure[1] & Michael J Bamshad[1,2,9]

We demonstrate the first successful application of exome sequencing to discover the gene for a rare mendelian disorder of unknown cause, Miller syndrome (MIM%263750). For four affected individuals in three independent kindreds, we captured and sequenced coding regions to a mean coverage of 40× and sufficient depth to call variants at ~97% of each targeted exome. Filtering against public SNP databases and eight HapMap exomes for genes with two previously unknown variants in each of the four individuals identified a single candidate gene, *DHODH*, which encodes a key enzyme in the pyrimidine *de novo* biosynthesis pathway. Sanger sequencing confirmed the presence of *DHODH* mutations in three additional families with Miller syndrome. Exome sequencing of a small number of unrelated affected individuals is a powerful, efficient strategy for identifying the genes



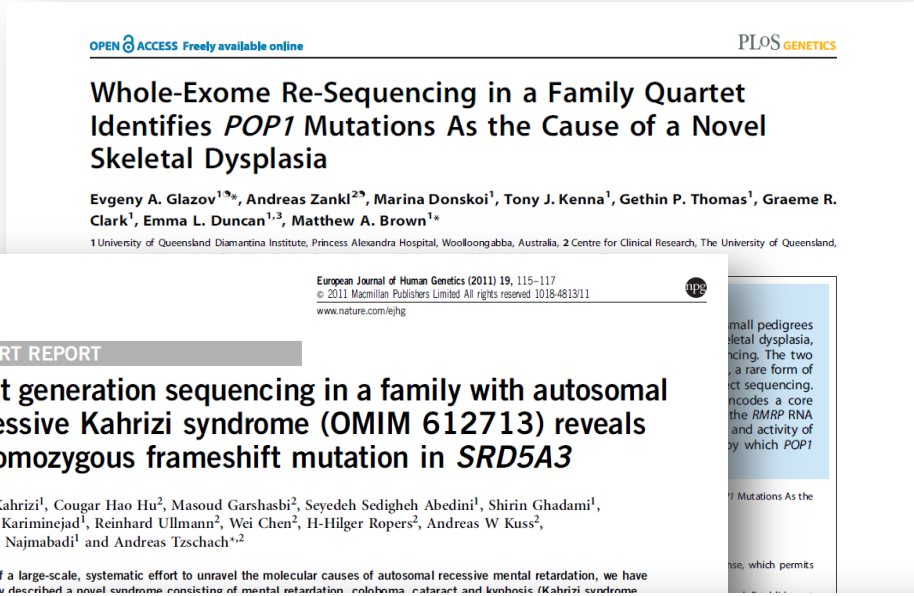
REVIEWS

TRANSLATIONAL GENETICS

Exome sequencing as a tool for Mendelian disease gene discovery

Michael J. Bamshad†‡, Sarah B. Ng‡, Abigail W. Bigham*§, Holly K. Tabor*||, Mary J. Emond¶, Deborah A. Nickerson‡ and Jay Shendure‡*

Abstract | Exome sequencing — the targeted sequencing of the subset of the human genome that is protein coding — is a powerful and cost-effective new tool for dissecting the genetic basis of diseases and traits that have proved to be intractable to conventional gene-discovery strategies. Over the past 2 years, experimental and analytical approaches relating to exome sequencing have established a rich framework for discovering the genes underlying unsolved Mendelian disorders. Additionally, exome sequencing is being adapted to explore the extent to which rare alleles explain the heritability of complex diseases and health-related traits. These advances also set the stage for applying exome and whole-genome sequencing to facilitate clinical diagnosis and personalized disease-risk profiling.



OPEN ACCESS Freely available online

PLoS GENETICS

Whole-Exome Re-Sequencing in a Family Quartet Identifies *POP1* Mutations As the Cause of a Novel Skeletal Dysplasia

Evgeny A. Glazov[1☯]*, Andreas Zankl[2☯], Marina Donskoi[1], Tony J. Kenna[1], Gethin P. Thomas[1], Graeme R. Clark[1], Emma L. Duncan[1,3], Matthew A. Brown[1]*

1 University of Queensland Diamantina Institute, Princess Alexandra Hospital, Woolloongabba, Australia, 2 Centre for Clinical Research, The University of Queensland,

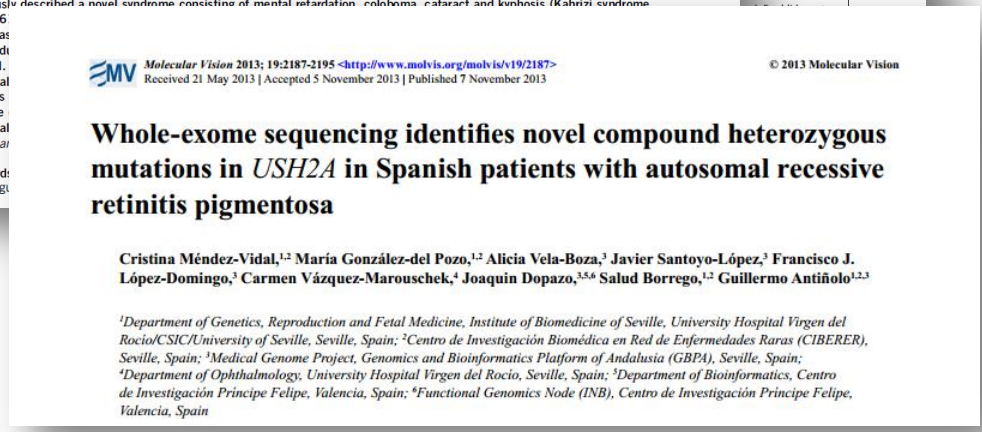
European Journal of Human Genetics (2011) 19, 115–117

www.nature.com/ejhg

SHORT REPORT

Next generation sequencing in a family with autosomal recessive Kahrizi syndrome (OMIM 612713) reveals a homozygous frameshift mutation in *SRD5A3*

Kimia Kahrizi[1], Cougar Hao Hu[2], Masoud Garshasbi[2], Seyedeh Sedigheh Abedini[1], Shirin Ghadami[1], Roxana Kariminejad[1], Reinhard Ullmann[2], Wei Chen[2], H-Hilger Ropers[2], Andreas W Kuss[2], Hossein Najmabadi[1] and Andreas Tzschach*[,2]



Molecular Vision 2013; 19:2187-2195 <http://www.molvis.org/molvis/v19/2187>

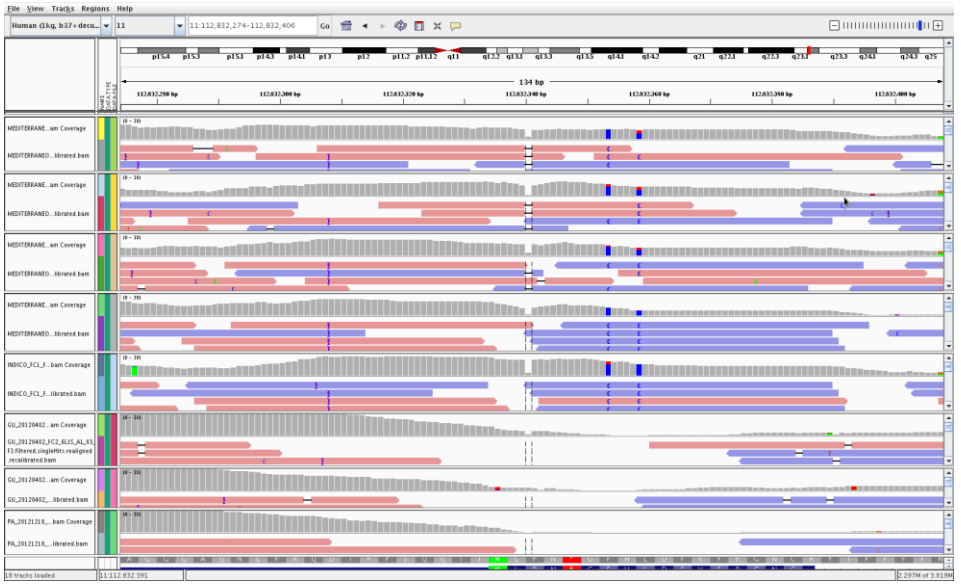
Received 21 May 2013 | Accepted 5 November 2013 | Published 7 November 2013

Whole-exome sequencing identifies novel compound heterozygous mutations in *USH2A* in Spanish patients with autosomal recessive retinitis pigmentosa

Cristina Méndez-Vidal,[1,2] María González-del Pozo,[1,2] Alicia Vela-Boza,[3] Javier Santoyo-López,[3] Francisco J. López-Domingo,[3] Carmen Vázquez-Marouschek,[4] Joaquin Dopazo,[3,5,6] Salud Borrego,[1,2] Guillermo Antiñolo[1,2,3]

[1]*Department of Genetics, Reproduction and Fetal Medicine, Institute of Biomedicine of Seville, University Hospital Virgen del Rocío/CSIC/University of Seville, Seville, Spain;* [2]*Centro de Investigación Biomédica en Red de Enfermedades Raras (CIBERER), Seville, Spain;* [3]*Medical Genome Project, Genomics and Bioinformatics Platform of Andalusia (GBPA), Seville, Spain;* [4]*Department of Ophthalmology, University Hospital Virgen del Rocío, Seville, Spain;* [5]*Department of Bioinformatics, Centro de Investigación Príncipe Felipe, Valencia, Spain;* [6]*Functional Genomics Node (INB), Centro de Investigación Príncipe Felipe, Valencia, Spain*

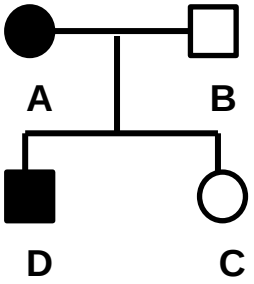
The principle: comparison of patients to reference controls or segregation within families

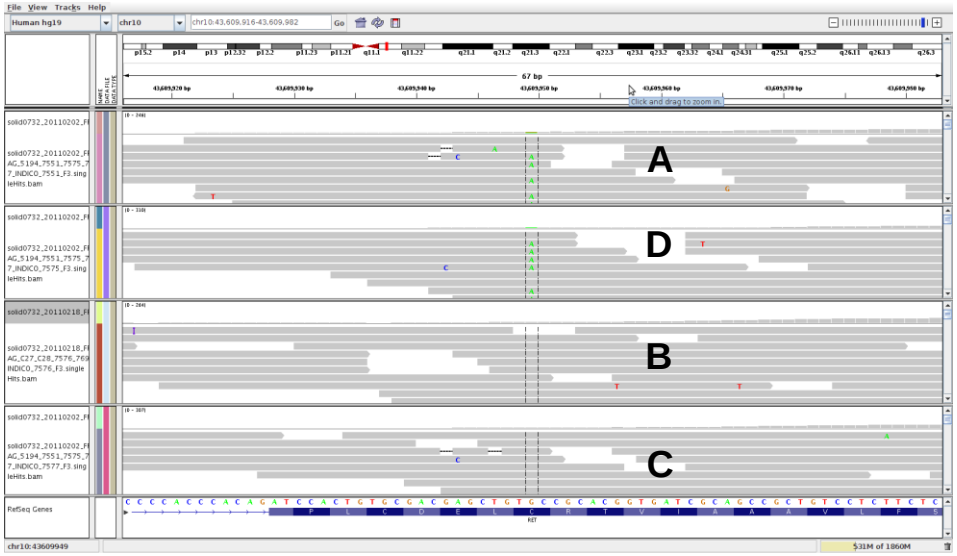


Cases

Controls

Segregation within a pedigree





Pipeline of data analysis

Primary processing	Secondary analysis (Heuristic filtering)	Knowledge-based prioritization
Initial QC FASTQ file	Variant annotation	Proximity to other known disease genes
Mapping BAM file	Filtering by effect	Functional proximity
Variant calling VCF File	Filtering by MAF	Network proximity
	Filtering by family segregation	Burden tests
		Other prioritization methods

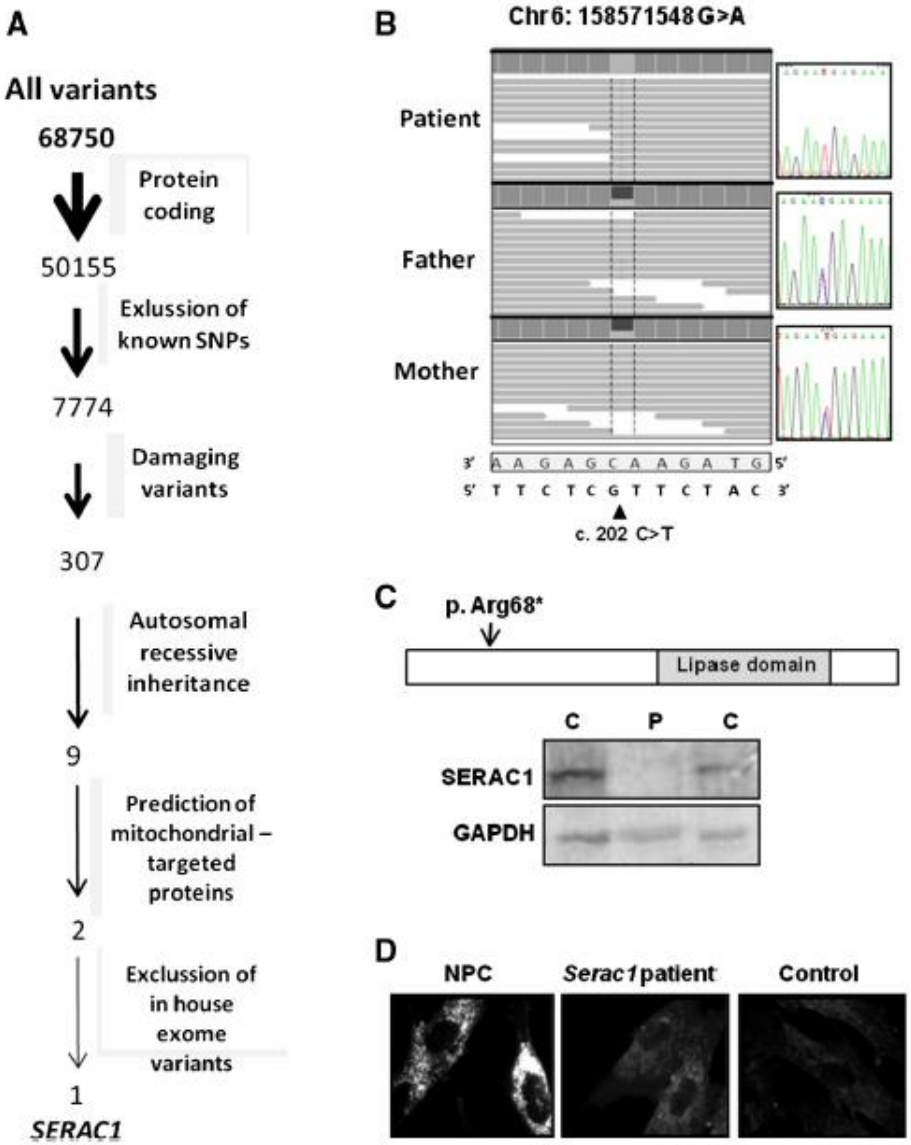
Primary analysis

Gene prioritization

Heuristic Filtering approach

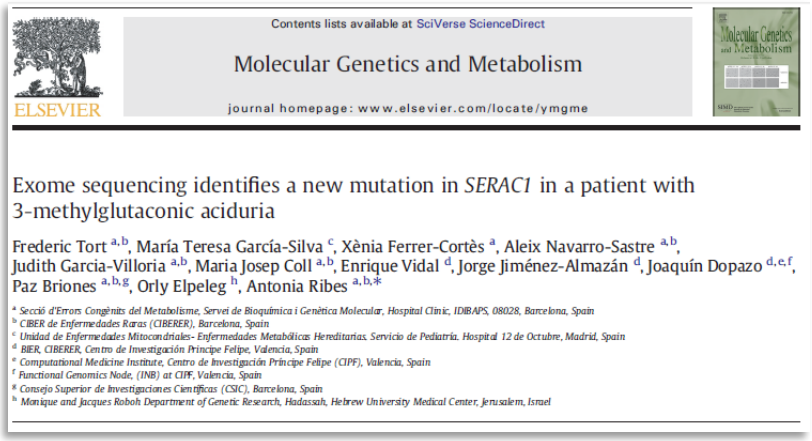
An example with 3-Methylglutaconic aciduria syndrome

F. Tort et al. / Molecular Genetics and Metabolism xxx (2013) xxx–xxx



3-Methylglutaconic aciduria (3-MGA-uria) is a heterogeneous group of syndromes characterized by an increased excretion of 3-methylglutaconic and 3-methylglutaric acids.

WES with a consecutive filter approach is enough to detect the new mutation in this case.



Contents lists available at SciVerse ScienceDirect

Molecular Genetics and Metabolism

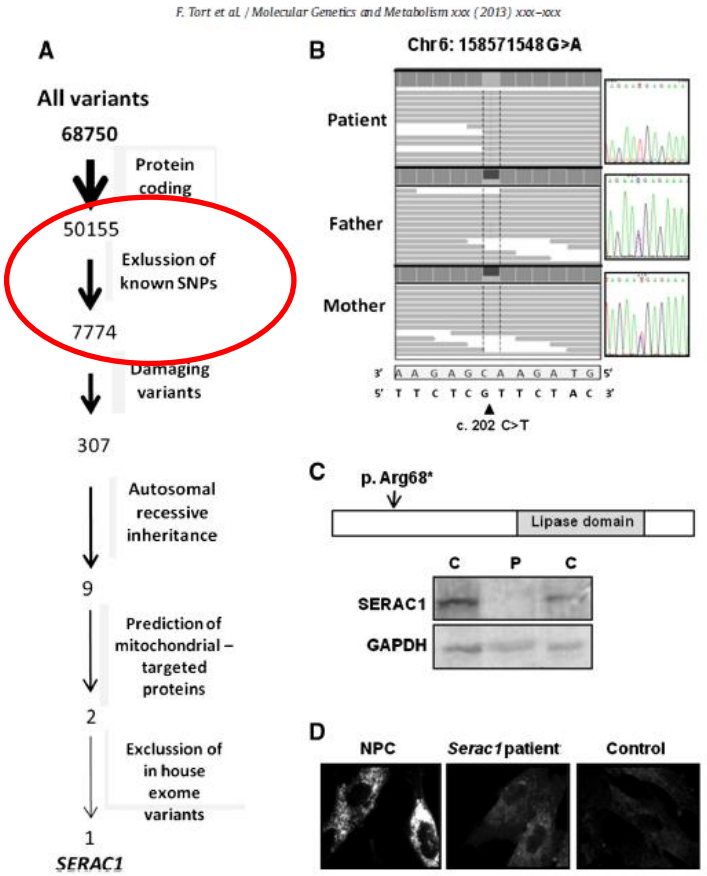
journal homepage: www.elsevier.com/locate/ymgme

Exome sequencing identifies a new mutation in *SERAC1* in a patient with 3-methylglutaconic aciduria

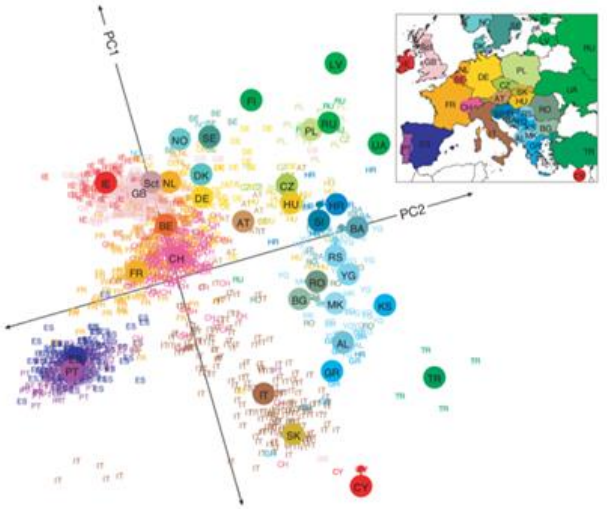
Frederic Tort [a,b], María Teresa García-Silva [c], Xènia Ferrer-Cortès [a], Aleix Navarro-Sastre [a,b], Judith Garcia-Villoria [a,b], Maria Josep Coll [a,b], Enrique Vidal [d], Jorge Jiménez-Almazán [d], Joaquín Dopazo [d,e,f], Paz Briones [a,b,g], Orly Elpeleg [h], Antonia Ribes [a,b,*]

[a] Secció d'Errors Congènits del Metabolisme, Servei de Bioquímica i Genètica Molecular, Hospital Clínic, IDIBAPS, 08028, Barcelona, Spain
[b] CIBER de Enfermedades Raras (CIBERER), Barcelona, Spain
[c] Unidad de Enfermedades Mitocondriales- Enfermedades Metabólicas Hereditarias. Servicio de Pediatría, Hospital 12 de Octubre, Madrid, Spain
[d] BIER, CIBERER, Centro de Investigación Príncipe Felipe, Valencia, Spain
[e] Computational Medicine Institute, Centro de Investigación Príncipe Felipe (CIPF), Valencia, Spain
[f] Functional Genomics Node, (INB) at CIPF, Valencia, Spain
[g] Consejo Superior de Investigaciones Científicas (CSIC), Barcelona, Spain
[h] Monique and Jacques Roboh Department of Genetic Research, Hadassah, Hebrew University Medical Center, Jerusalem, Israel

Lessons learned: the importance of local variability in the prioritization process

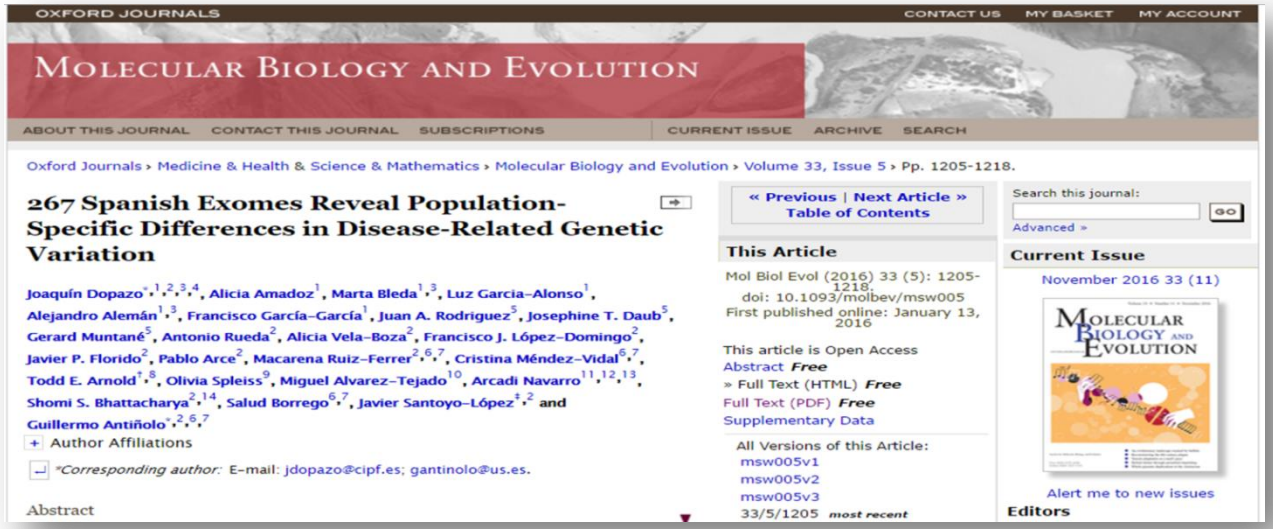


One of the most stringent filtering steps is the exclusion of known population polymorphisms. Public databases (1000 genomes, ESP, ExAC)



It is well known that population is structured, but, to what extent is this structure important in the filtering process?

Lessons learned: the importance of local variability in the prioritization process



OXFORD JOURNALS — CONTACT US | MY BASKET | MY ACCOUNT

MOLECULAR BIOLOGY AND EVOLUTION

ABOUT THIS JOURNAL | CONTACT THIS JOURNAL | SUBSCRIPTIONS | CURRENT ISSUE | ARCHIVE | SEARCH

Oxford Journals › Medicine & Health & Science & Mathematics › Molecular Biology and Evolution › Volume 33, Issue 5 › Pp. 1205-1218.

267 Spanish Exomes Reveal Population-Specific Differences in Disease-Related Genetic Variation

Joaquín Dopazo[*,1,2,3,4], Alicia Amadoz[1], Marta Bleda[1,3], Luz García-Alonso[1], Alejandro Alemán[1,3], Francisco García-García[1], Juan A. Rodriguez[5], Josephine T. Daub[5], Gerard Muntané[5], Antonio Rueda[2], Alicia Vela-Boza[2], Francisco J. López-Domingo[2], Javier P. Florido[2], Pablo Arce[2], Macarena Ruiz-Ferrer[2,6,7], Cristina Méndez-Vidal[6,7], Todd E. Arnold[†,8], Olivia Spleiss[9], Miguel Alvarez-Tejado[10], Arcadi Navarro[11,12,13], Shomi S. Bhattacharya[2,14], Salud Borrego[6,7], Javier Santoyo-López[‡,2] and Guillermo Antiñolo[*,2,6,7]

+ Author Affiliations

*Corresponding author: E-mail: jdopazo@cipf.es; gantinolo@us.es.

Abstract

« Previous | Next Article » Table of Contents

This Article

Mol Biol Evol (2016) 33 (5): 1205-1218.
doi: 10.1093/molbev/msw005
First published online: January 13, 2016

This article is Open Access
Abstract *Free*
» Full Text (HTML) *Free*
Full Text (PDF) *Free*
Supplementary Data

All Versions of this Article:
msw005v1
msw005v2
msw005v3
33/5/1205 *most recent*

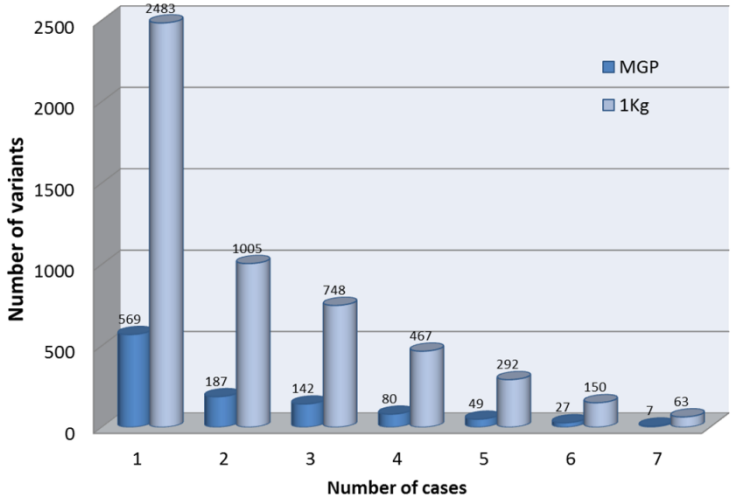
Search this journal: GO
Advanced »

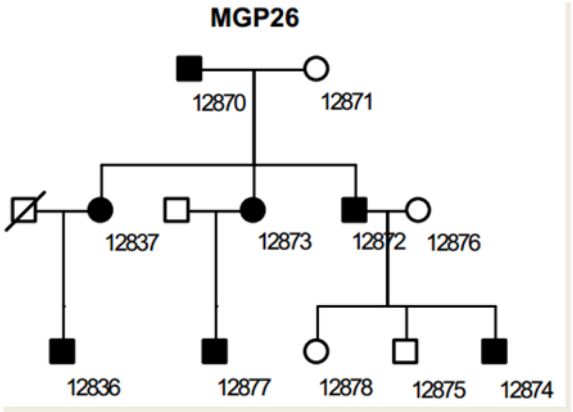
Current Issue
November 2016 33 (11)

Alert me to new issues

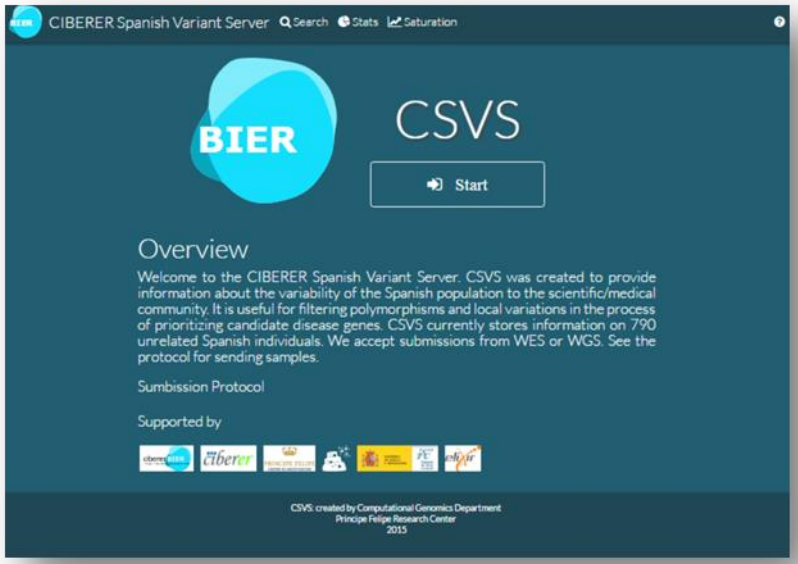
Editors

The filtering efficiency of the local population can be between 5 and 10 times those of a general database, such as the 1000 genomes





The CSVS is a crowdsourcing project

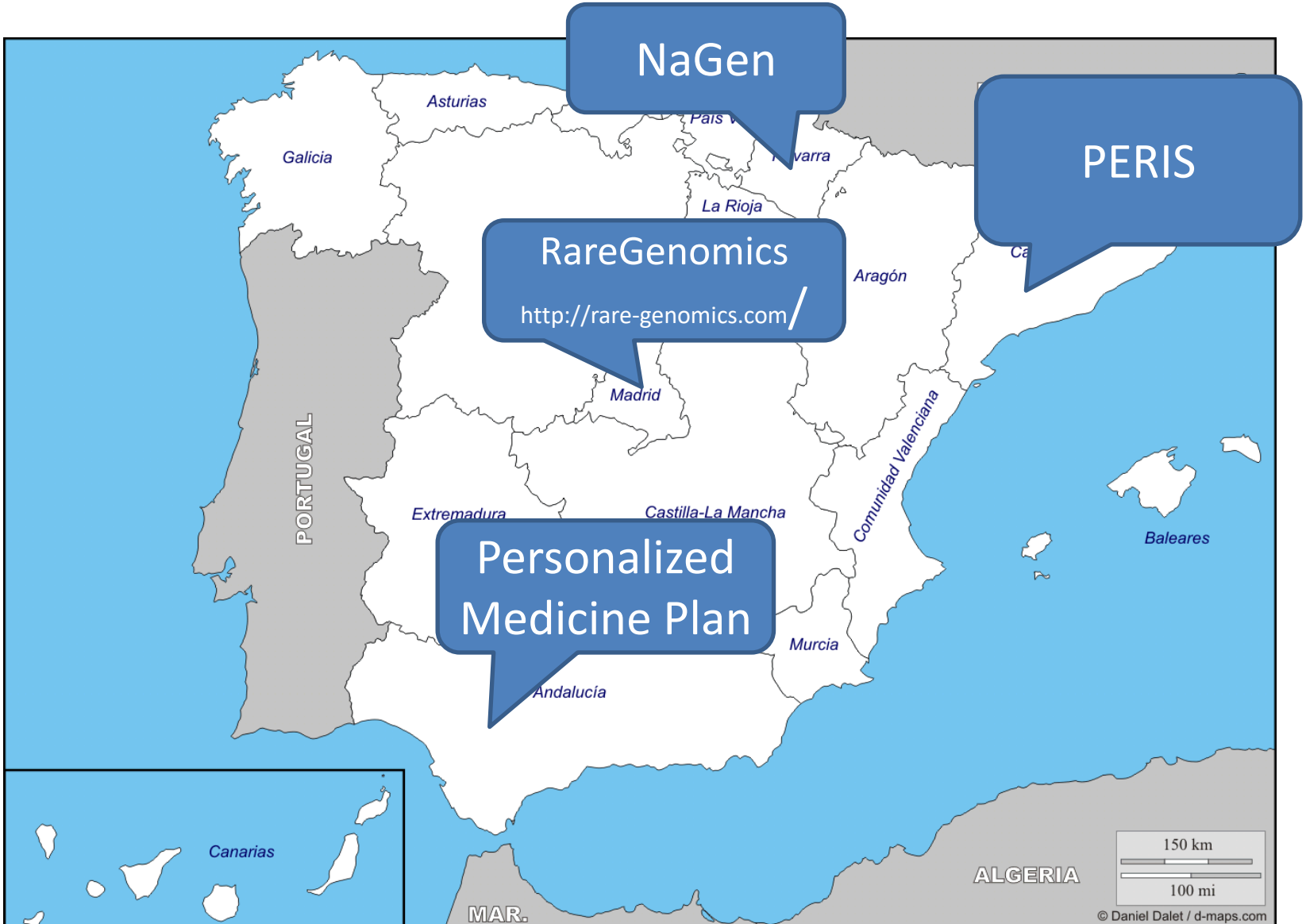


http://ciberer.es/bier/exome-server/

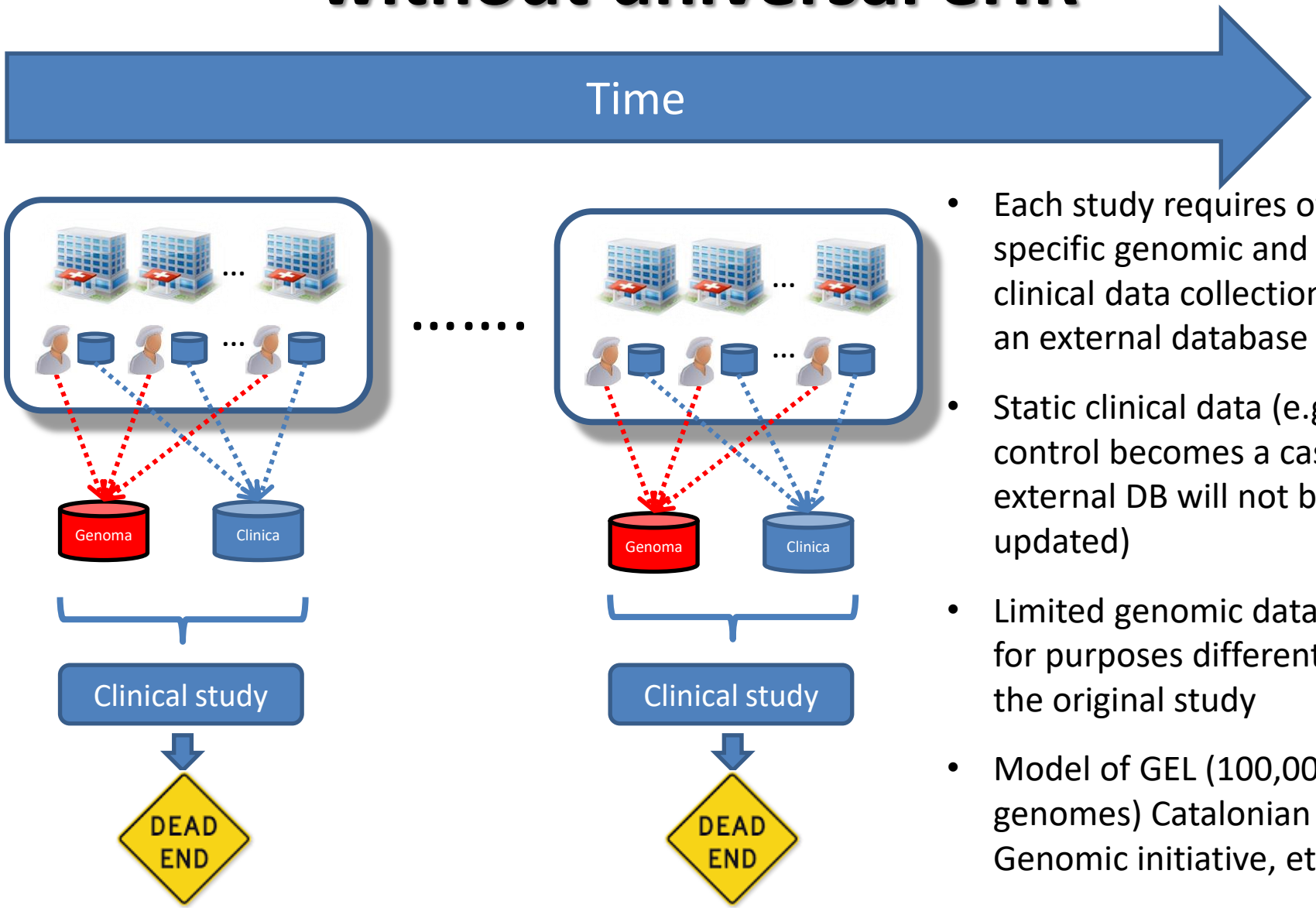
A widely used tool containing over 800 exomes and >2000 in September

Scenario: Sequencing projects of healthy population are expensive and funding bodies are reluctant to fund them
CSVS Aim: To offer increasingly accurate information on variant frequencies characteristic of Spanish population.
CSVS Main use: Frequency-based filtering of candidate variants
Main data source: Sequencing projects of individual researchers (CIBERER and others)
Problem: Most of the contributions correspond to patient exomes
Idea: Patients of disease A can be considered healthy **pseudo-controls** for disease B (providing no common genetic background exist between A and B)
Beacon: CSVS will soon appear in the Beacon server

Genomic initiatives

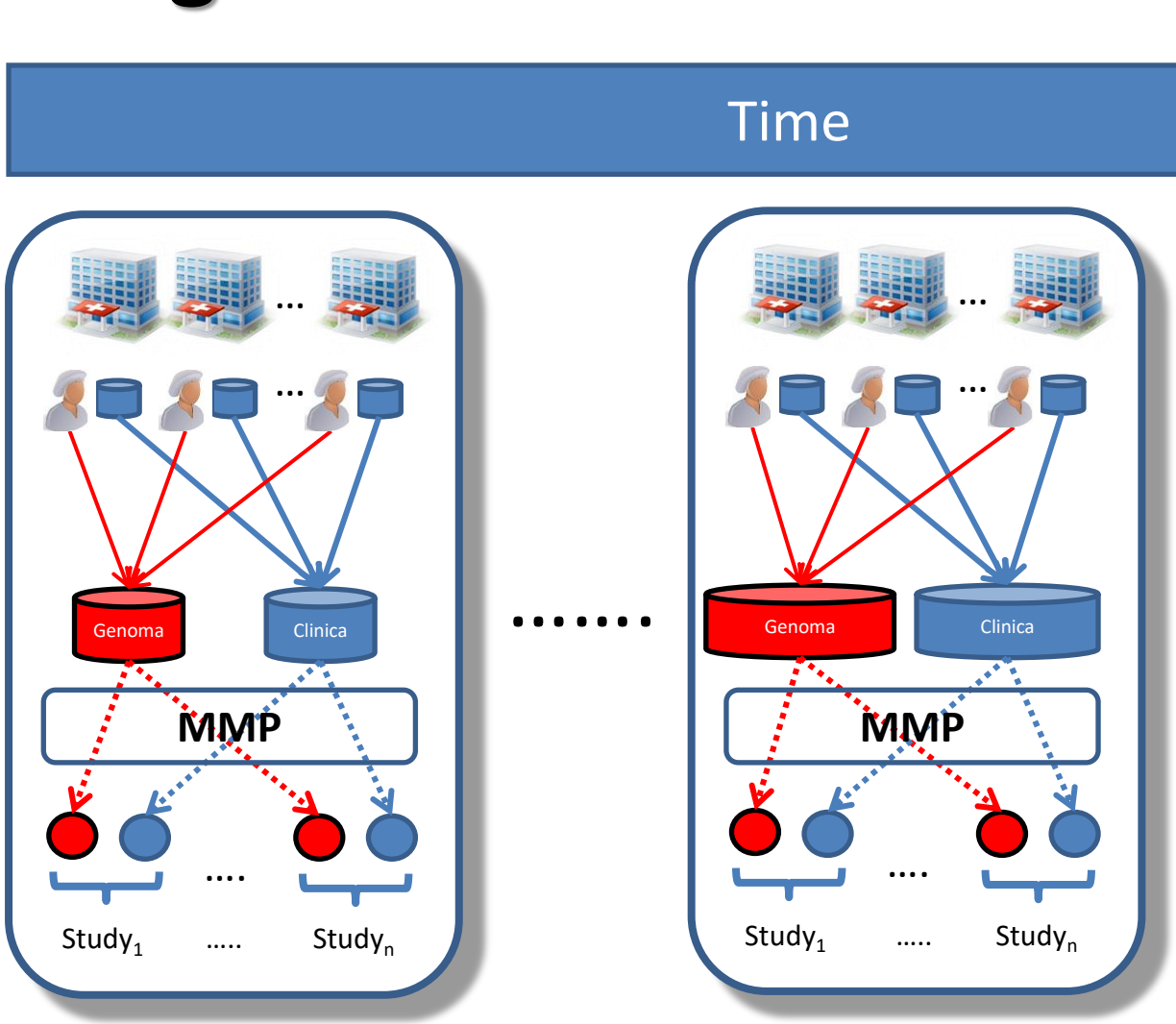


Personalized Medicine Model without universal eHR



- Each study requires of a specific genomic and clinical data collection into an external database
- Static clinical data (e.g. if a control becomes a case the external DB will not be updated)
- Limited genomic data reuse for purposes different from the original study
- Model of GEL (100,000 genomes) Catalonian Genomic initiative, etc.

Advantages of a model that integrates genomic data and universal eHR



- The whole health system becomes a humongous potential prospective study
- Clinical data dynamically associated to patients
- Possibility of many clinical studies by reanalyzing genomic data under diverse perspectives (with no extra investment)
- Growing genomic DB with increasing study possibilities

The (relative) success in rare diseases has not been reproduced in complex diseases

How to explain missing heritability?
Rare Variants, rare CNVs, epigenetics?

Table 1 | Estimates of heritability and number of loci for several complex traits

Disease	Number of loci	Proportion of heritability explained
Age-related macular degeneration[72]	5	50%
Crohn's disease[21]	32	20%
Systemic lupus erythematosus[73]	6	15%
Type 2 diabetes[74]	18	6%
HDL cholesterol[75]	7	5.2%
Height[15]	40	5%
Early onset myocardial infarction[76]	9	2.8%
Fasting glucose[77]	4	1.5%

* Residual is after adjustment for age, gender, diabetes.

Is the heritability missing or are we looking at the wrong place?

How to explain missing heritability?
Rare Variants, rare CNVs, epigenetics or.. **epistatic effects**?

Table 1 | Estimates of heritability and number of loci for several complex traits

Disease	Number of loci	Proportion of heritability explained
Age-related macular degeneration[72]	5	50%
Crohn's disease[21]	32	20%
Systemic lupus erythematosus[73]	6	15%
Type 2 diabetes[74]	18	6%
HDL cholesterol[75]	7	5.2%
Height[15]	40	5%
Early onset myocardial infarction[76]	9	3.8%
Fasting glucose[77]	4	1.5%

* Residual is after adjustment for age, gender, diabetes.

nature genetics

Common SNPs explain a large proportion of the heritability for human height

Jian Yang[1], Beben Benyamin[1], Brian P McEvoy[1], Scott Gordon[1], Anjali K Henders[1], Dale R Nyholt[1], Pamela A Madden[2], Andrew C Heath[2], Nicholas G Martin[1], Grant W Montgomery[1], Michael E Goddard[3] & Peter M Visscher[1]

SNPs discovered by genome-wide association studies (GWASs) account for only a small fraction of the genetic variation of complex traits in human populations. Where is the remaining heritability? We estimated the proportion of variance for human height explained by 294,831 SNPs genotyped on 3,925 unrelated individuals using a linear model analysis, and validated the estimation method with simulations based on the observed genotype data. We show that 45% of variance can be explained by considering all SNPs simultaneously. Thus, most of the heritability is not missing but has not previously been detected because the individual effects are too small to pass stringent significance tests. We provide evidence that the remaining heritability is due to incomplete linkage disequilibrium between causal variants and genotyped SNPs, exacerbated by causal variants having lower minor allele frequency than the SNPs explored to date.

GWASs in human populations have discovered hundreds of SNPs

of variation that their effects do not reach stringent significance thresholds and/or the causal variants are not in complete linkage disequilibrium with the SNPs that have been genotyped [and] of complete [illegible], or [illegible] minor allele frequency (MAF) than genotyped SNPs. Here we test these two hypotheses and estimate the contribution of the heritability of height that is [illegible]

Height in humans is a classical quantitative trait, easy to measure and studied for well over a century as a model for investigating the genetic basis [illegible]. The heritability of height has been estimated to be ~0.8 (refs. 9,11–13). Rare mutations that cause extreme short or tall stature have been found[14,15], but these do not explain much of the variation in the general population. Recent GWASs on tens of thousands of individuals have detected ~50 variants that are associated with height in the population, but these in total account for only ~5% of phenotypic variance[16–19].

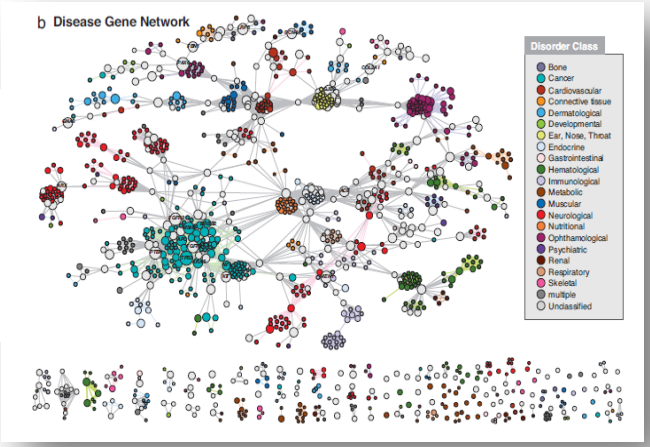
Data from a GWAS that are collected to detect statistical associations between SNPs and complex traits are usually analyzed by testing each

At the end, most of the heritability was there…

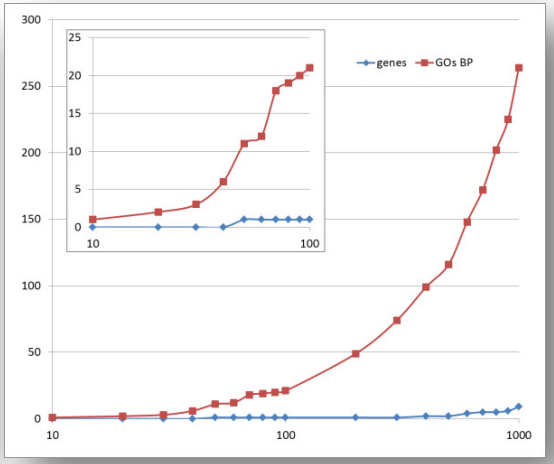
Human genetic diseases have a modular nature

- With the development of **systems biology**, studies have shown that phenotypically **similar diseases** are often caused by **functionally related genes**, being referred to as the **modular nature of human genetic diseases** (Oti and Brunner, 2007; Oti et al, 2008).
- This modularity suggests that **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



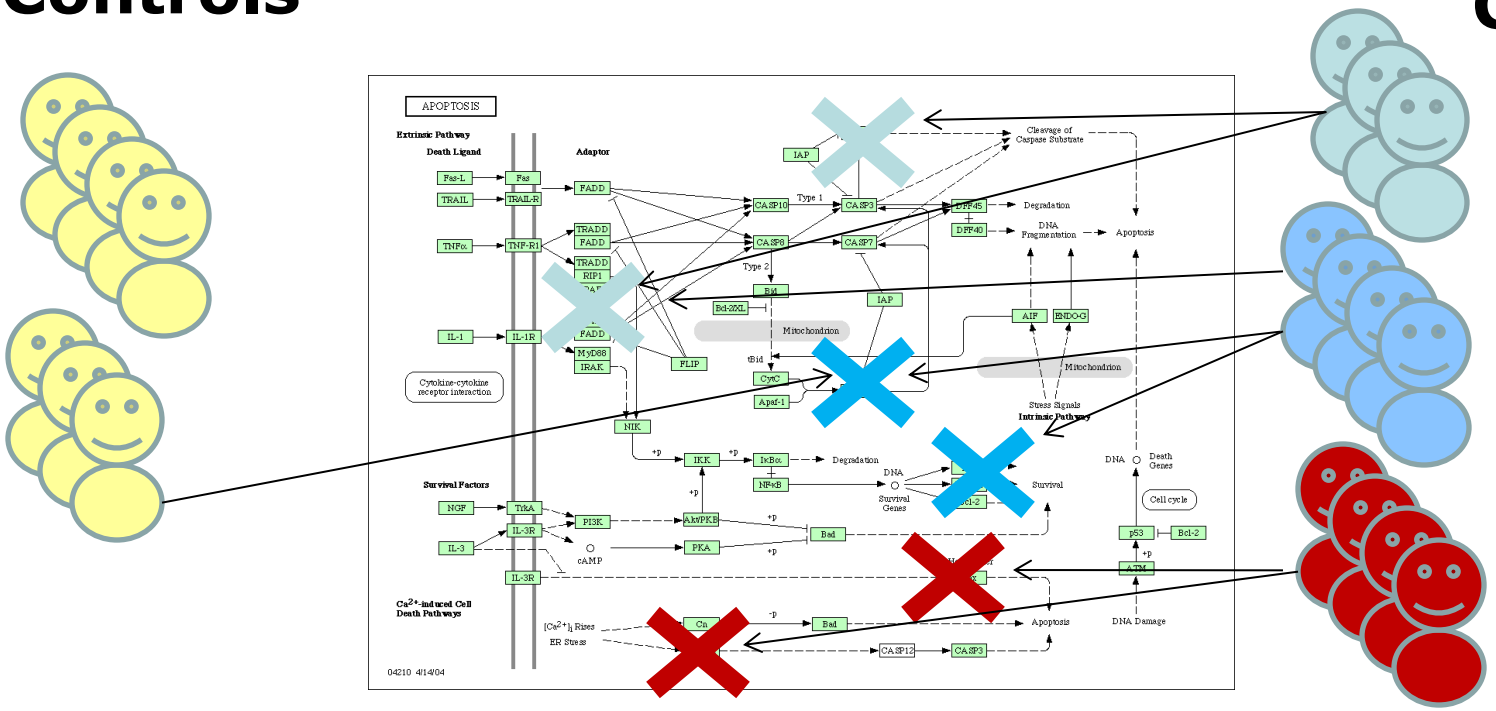
Fernandez, 2013, Orphanet J Rare Dis.

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

The modular nature of human diseases

Controls

Cases



Affected **cases** in complex diseases will be a **heterogeneous** population with different mutations (or combinations).

Many cases and controls are needed to obtain significant associations.

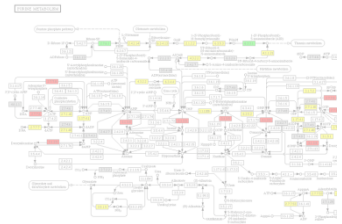
The only **common element** is the (know or unknown) **module affected**.

Disease understood as the failure of a functional module

Two problems: defining functional modules and modeling their behavior

Gene ontology: descriptive; unstructured functional labels

Interactome: relationships among components but unknown function



Pathways: relationships among components and their functional roles

Models

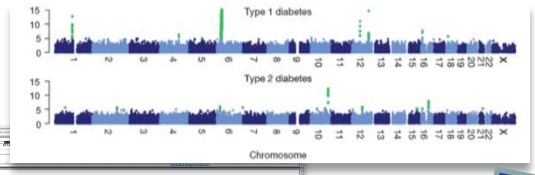
Enrichment methods. GO, etc. (simple statistical tests)

Connectivity models. Protein-protein, protein-DNA and protein-small molecule interactions (tests on network properties)

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

Detailed models. Kinetic models including stoichiometry, balancing reactions, etc. (mathematical models)

From gene-based to function-based perspective



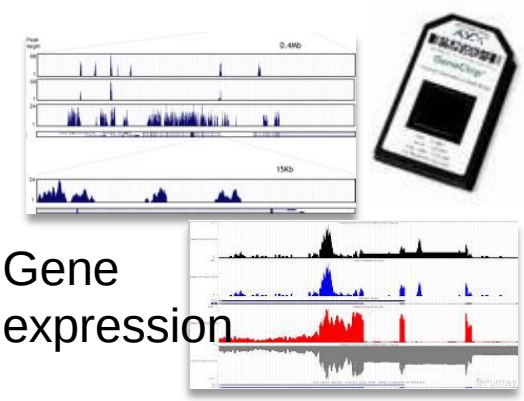
SNPs

WES/WGS

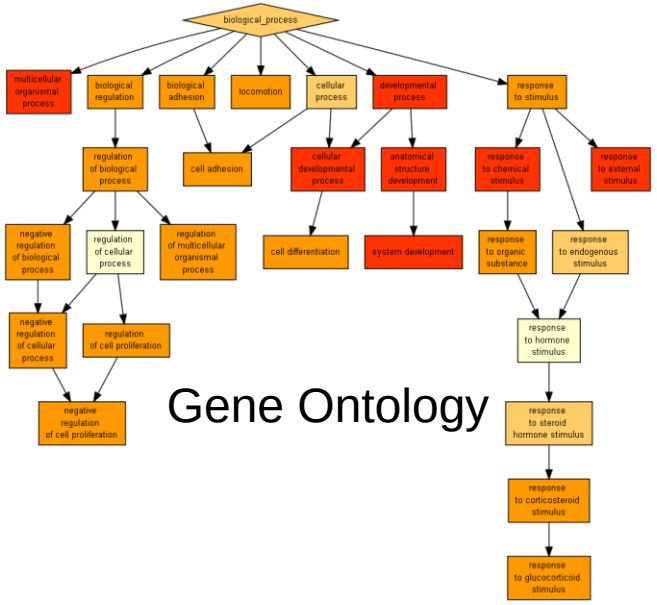
Gene integrity

AND/OR

Gene activity



Gene expression



Gene Ontology

Gene Ontology are **labels** to genes that describe, by means of a controlled vocabulary (ontology), the **functional role(s)** played by the genes in the cell.
A set of genes **sharing** a **GO** annotation can be considered a **functional module**.

An example of GWAS

GWAS in Breast Cancer.

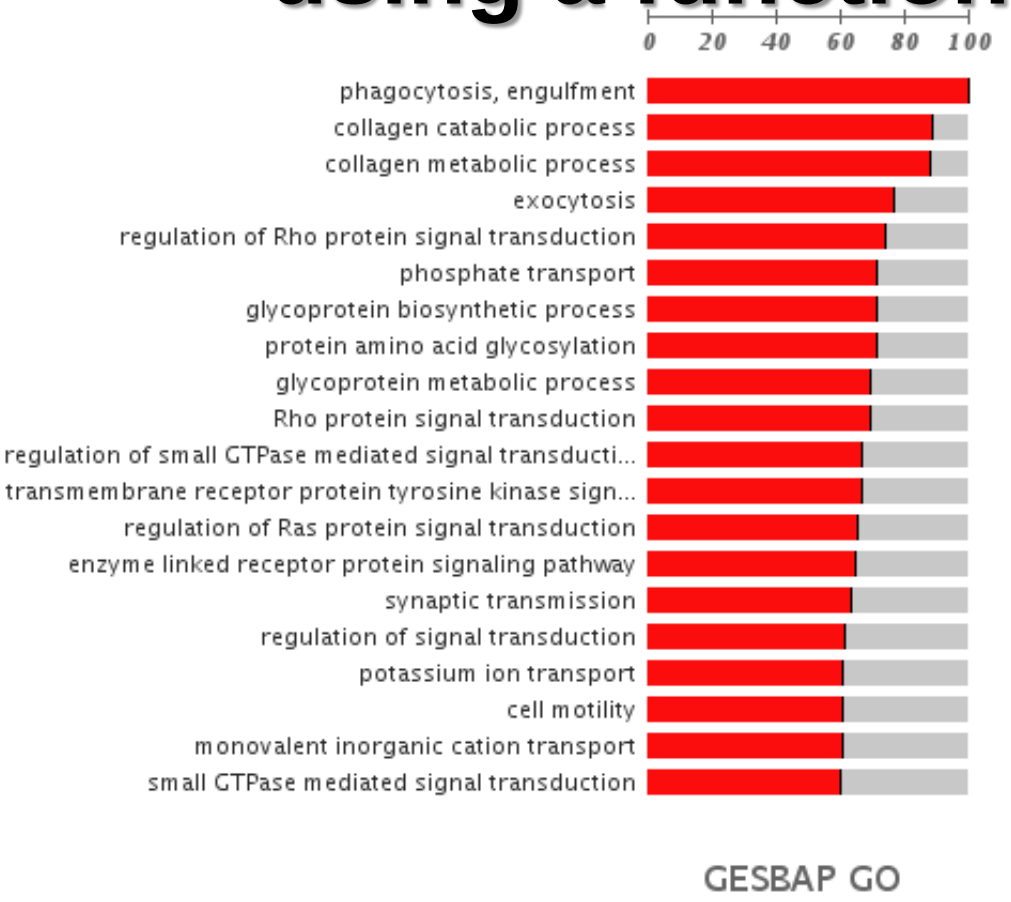
The CGEMS initiative. (Hunter et al. Nat Genet 2007)

1145 cases 1142 controls. Affy 500K

Conventional association test reports only 4 SNPs significantly mapping only on one gene: FGFR2

Conclusions: **conventional SNP-based** or **gene-based tests** are not providing much resolution.

The same GWAS data re-analyzed using a function-based test



Breast Cancer

CGEMS initiative.
(Hunter et al. Nat Genet 2007)

1145 cases 1142 controls. Affy 500K

Only 4 SNPs were significantly associated, mapping only in one gene: FGFR2

PBA reveals 19 GO categories including *regulation of signal transduction* (FDR-adjusted p-value=4.45×10^{-03}) in which FGFR2 is included.

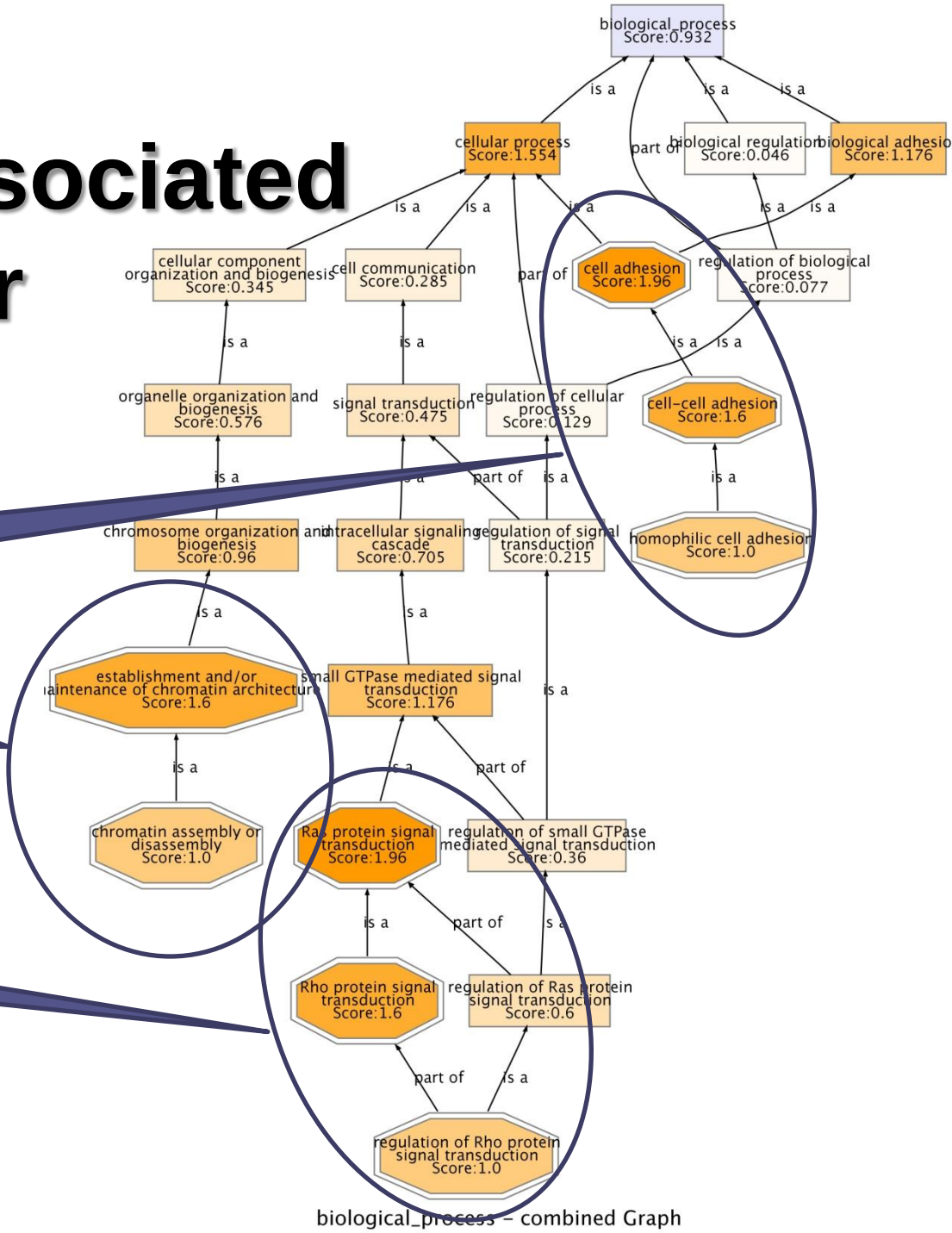
Bonifaci et al., BMC Medical Genomics 2008; Medina et al., 2009 NAR

GO processes significantly associated to breast cancer

Metastasis

Chromosomal instability

Rho pathway

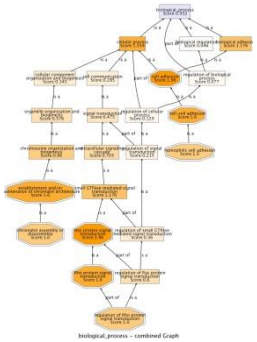


From gene-based to function-based perspective

SNPs,
Gene expression

$Gene_1$
$Gene_2$
$Gene_3$
$Gene_4$
:
:
:
$Gene_{22000}$

Gene Ontology

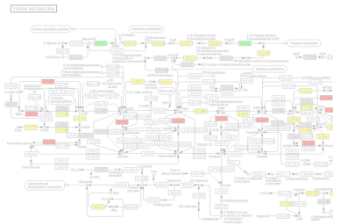


	SNPs, gene exp.	GO
Detection power	Low (only very prevalent genes)	high
Annotations available	many	many
Use	Biomarker	Illustrative, give hints

Two problems: defining functional modules and modeling their behavior

Gene ontology: descriptive; unstructured functional labels

Interactome: relationships among components but unknown function



Pathways: relationships among components and their functional roles

Models

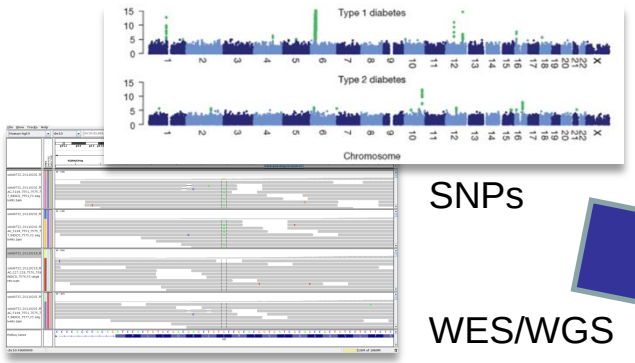
Enrichment methods. GO, etc. (simple statistical tests)

Connectivity models. Protein-protein, protein-DNA and protein-small molecule interactions (tests on network properties)

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

Detailed models. Kinetic models including stoichiometry, balancing reactions, etc. (mathematical models)

From gene-based to function-based perspective

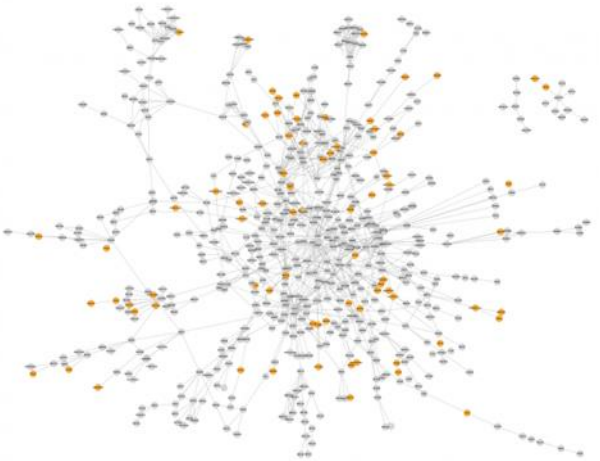


SNPs

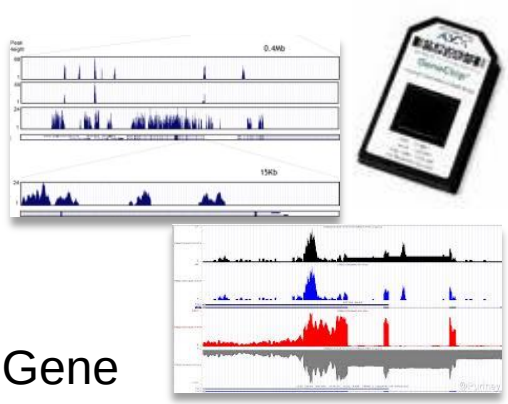
WES/WGS

Gene integrity

Using protein interaction networks as an scaffold to interpret the genomic data in a functionally-derived context



AND/OR



Gene expression

Gene activity

What part of the interactome is active and/or is damaged

Network analysis helps to find disease genes in complex diseases

Research

Open Access

Four new loci associations discovered by pathway-based and network analyses of the genome-wide variability profile of Hirschsprung's disease

Raquel Ma Fernández[1,2], Marta Bleda[2,3], Rocío Núñez-Torres[1,2], Ignacio Medina[3,4], Berta Luzón-Toro[1,2], Luz García-Alonso[3], Ana Torroglosa[1,2], Martina Marbà[3,4], Ma Valle Enguix-Riego[1,2], David Montaner[3], Guillermo Antiñolo[1,2], Joaquín Dopazo[2,3,4]* and Salud Borrego[1,2]*

* Corresponding authors: Joaquín Dopazo jdopazo@cipf.es - Salud Borrego salud.borrego.sspa@juntadeandalucia.es

▸ Author Affiliations

For all author emails, please log on.

Orphanet Journal of Rare Diseases 2012, **7**:103 doi:10.1186/1750-1172-7-103

Published: 28 December 2012

Nucleic Acids Research Advance Access published May 19, 2009

Nucleic Acids Research, 2009, 1–6
doi:10.1093/nar/gkp402

SNOW, a web-based tool for the statistical analysis of protein–protein interaction networks

Pablo Minguez[1], Stefan Götz[1,2], David Montaner[1], Fatima Al-Shahrour[1] and Joaquin Dopazo[1,2,3,*]

[1]Department of Bioinformatics and Genomics, Centro de Investigación Príncipe Felipe (CIPF), [2]CIBER de Enfermedades Raras (CIBERER) and [3]Functional Genomics Node (INB) at CIPF, Valencia, Spain

Received January 21, 2009; Revised April 22, 2009; Accepted May 2, 2009

Published online 27 July 2012 *Nucleic Acids Research, 2012, Vol. 40, No. 20 e158*
doi:10.1093/nar/gks699

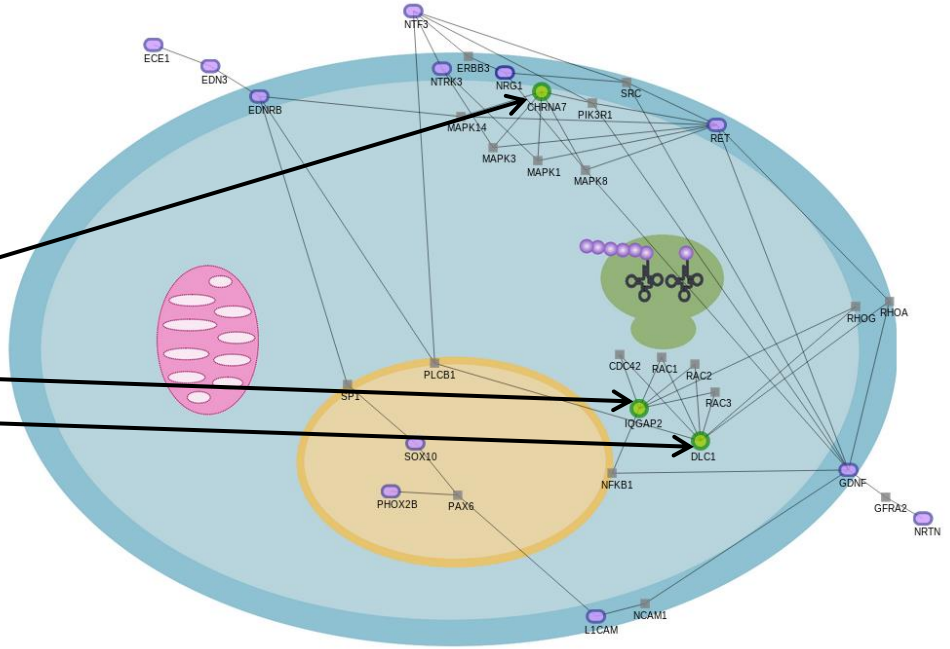
Discovering the hidden sub-network component in a ranked list of genes or proteins derived from genomic experiments

Luz García-Alonso[1], Roberto Alonso[1], Enrique Vidal[1], Alicia Amadoz[1], Alejandro de María[1], Pablo Minguez[2], Ignacio Medina[1,3] and Joaquín Dopazo[1,3,4,*]

[1]Department of Bioinformatics, Centro de Investigación Príncipe Felipe (CIPF), Valencia, Spain, [2]European Molecular Biology Laboratory, Meyerhofstrasse 1, 69117 Heidelberg, Germany, [3]Functional Genomics Node (INB) at CIPF, Valencia and [4]CIBER de Enfermedades Raras (CIBERER), Valencia, Spain

Received March 14, 2012; Revised June 1, 2012; Accepted June 26, 2012

CHRNA7 (rs2175886 p = 0.000607)
IQGAP2 (rs950643 p = 0.0003585)
DLC1 (rs1454947 p = 0.007526)



SNPs validated in independent cohorts

From gene-based to function-based perspective

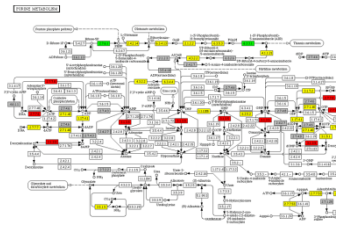
	SNPs, gene expression, etc.	GO	Protein interaction networks
Detection power	Low (only very prevalent genes)	High	High
Information coverage	Almost all	Almost all	Less (~9000 genes in human)
Use	Biomarker	Illustrative, give hints	Biomarker*

**Need of extra information (e.g. GO) to provide functional insights in the findings*

Two problems: defining functional modules and modeling their behavior

Gene ontology: descriptive; unstructured functional labels

Interactome: relationships among components but unknown function



Pathways: relationships among components and their functional roles

Models

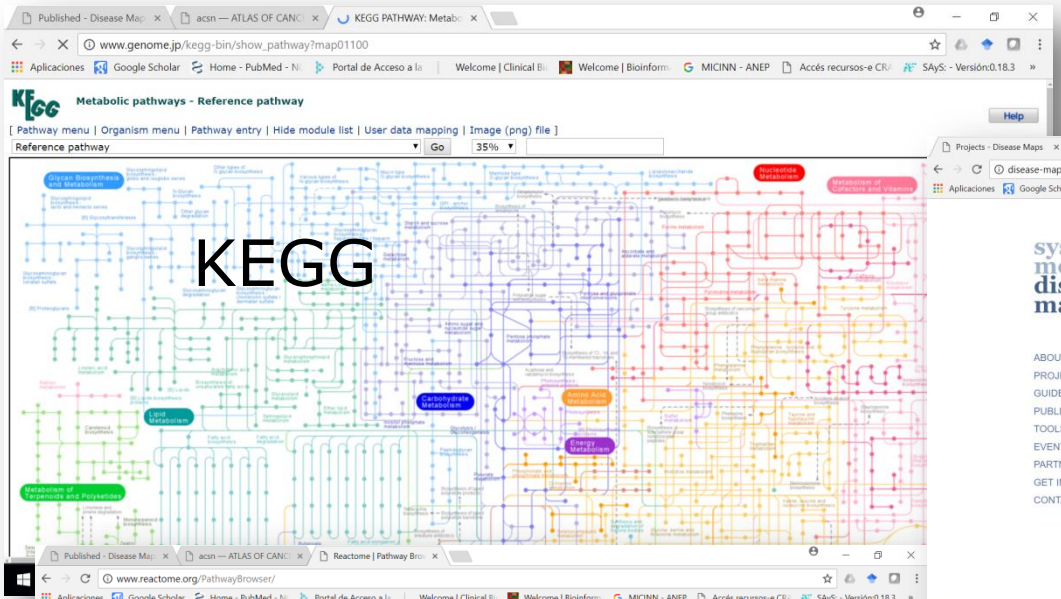
Enrichment methods. GO, etc. (simple statistical tests)

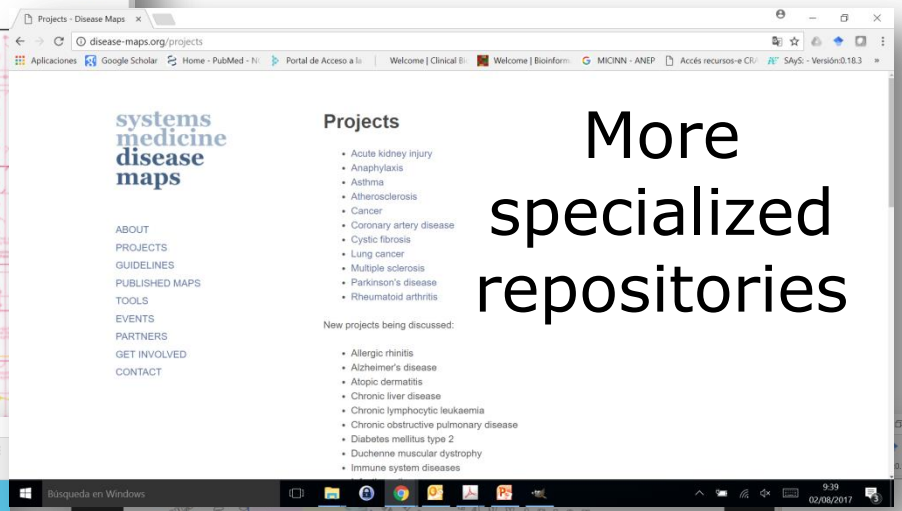
Connectivity models. Protein-protein, protein-DNA and protein-small molecule interactions (tests on network properties)

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

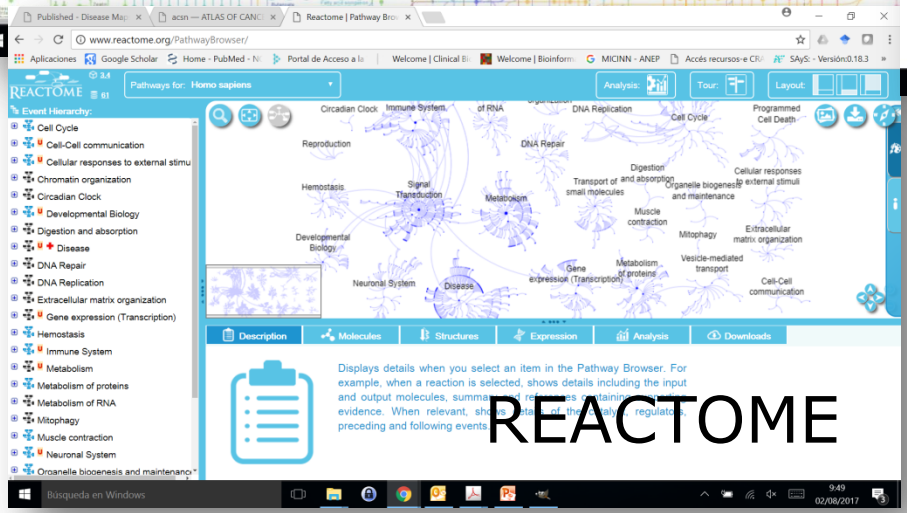
Mathematical models. Kinetic models including stoichiometry, balancing reactions, etc. (mathematical models)

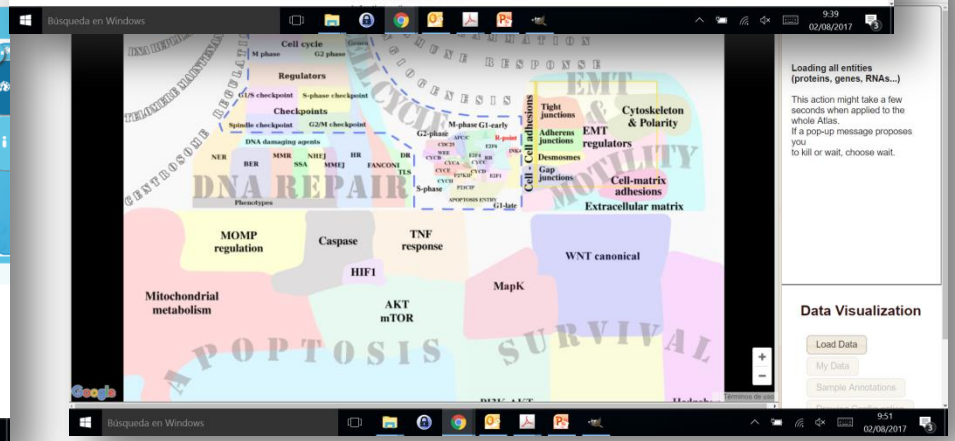
Where the cell activity maps come from?





More specialized repositories





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 alternative 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

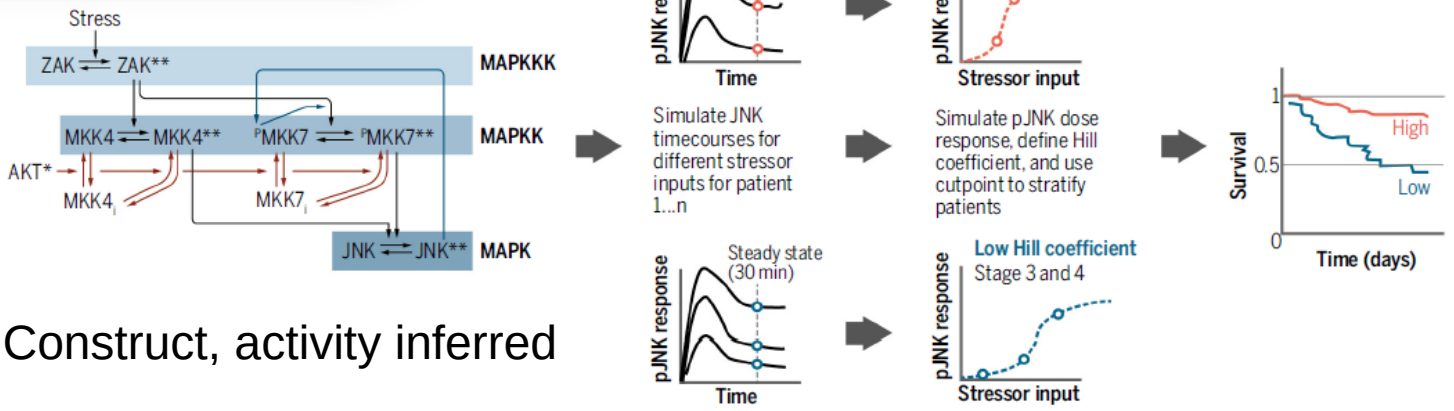


Fig. 1. Using network descriptors of signaling pathway activation potential to predict patient response. After construction of a computational model based on the validated network topology and that reproduces the signaling pathway dynamics, the model can be used to identify network descriptors, such 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.

From gene-based to function-based perspective

	SNPs, gene expression, etc.	GO	Protein interaction networks	Models of cellular functions
Detection power	Low (only very prevalent genes)	High	High	Very high
Information coverage	Almost all	Almost all	Low (~9000 genes in human)	Low (~6700 human genes)*
Use	Biomarker	Illustrative, give hints	Biomarker	Biomarker that explain disease mechanism

*Only ~1000 genes in human signaling pathways

The real 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** **Precision Medicine**

Today

Tomorrow

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