

A NGS data analysis course

Introduction

Computational Genomics Department,
Centro de Investigación Príncipe Felipe (CIPF),
Functional Genomics Node, (INB),
Bioinformatics Group (CIBERER)
Spain.

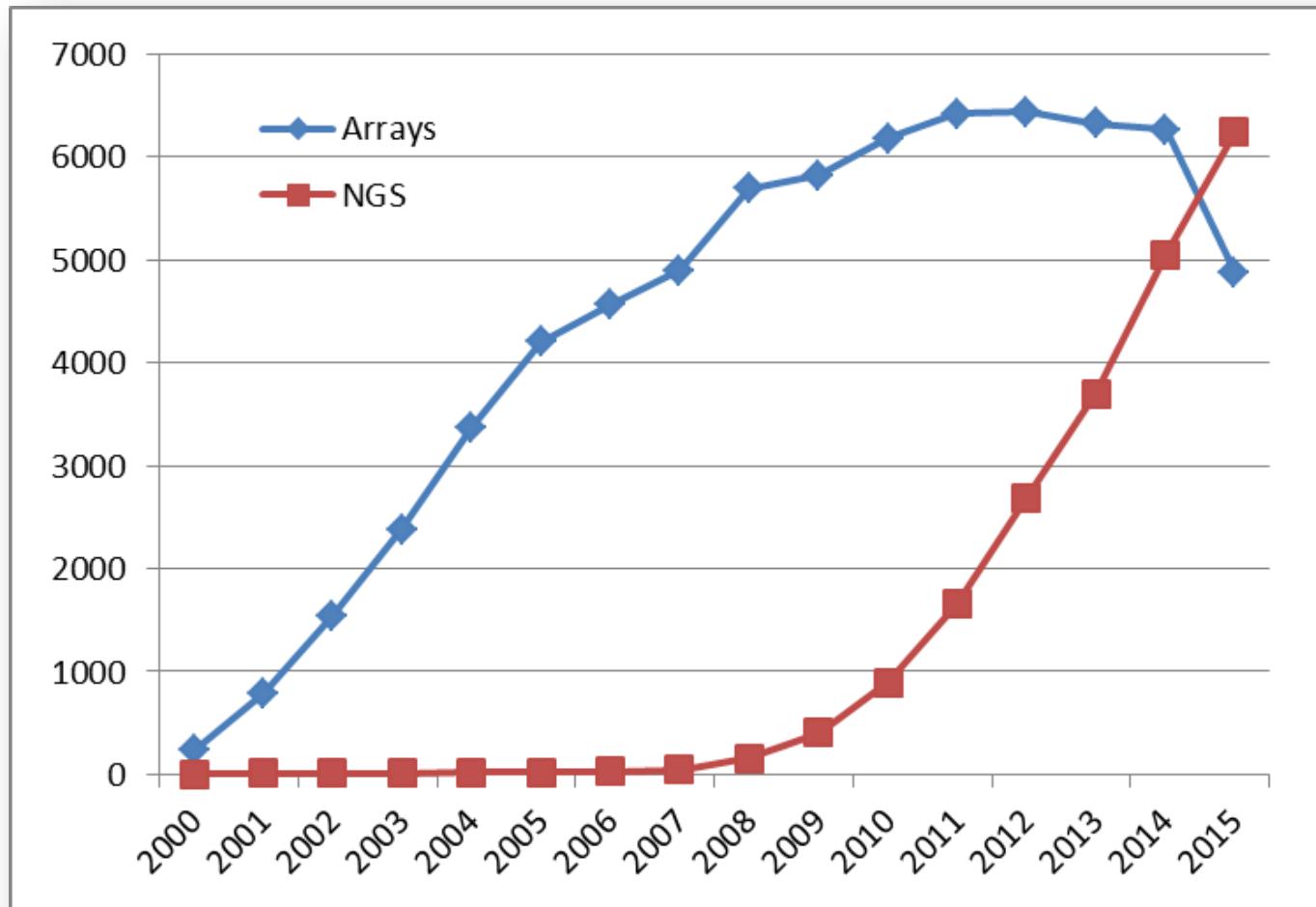
<http://bioinfo.cipf.es>
<http://www.babelomics.org>
<http://www.hpc4g.org>
 @xdopazo



University of Cambridge, 17-19 June 2015

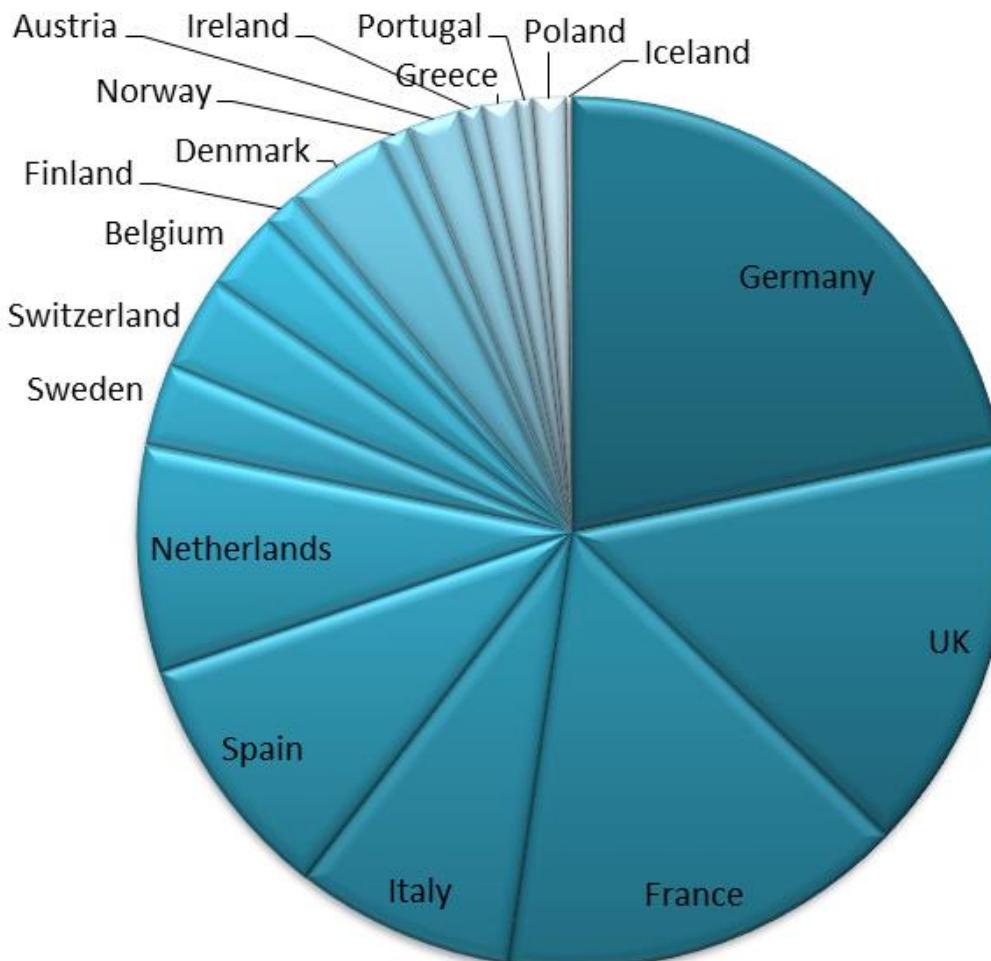


Evolution of the papers published in microarray and NGS technologies



Source Pubmed. Query: "high-throughput sequencing"[Title/Abstract] OR "next generation sequencing"[Title/Abstract] OR "rna seq"[Title/Abstract]) AND year[Publication Date]

Some bibliographic data: NGS publications



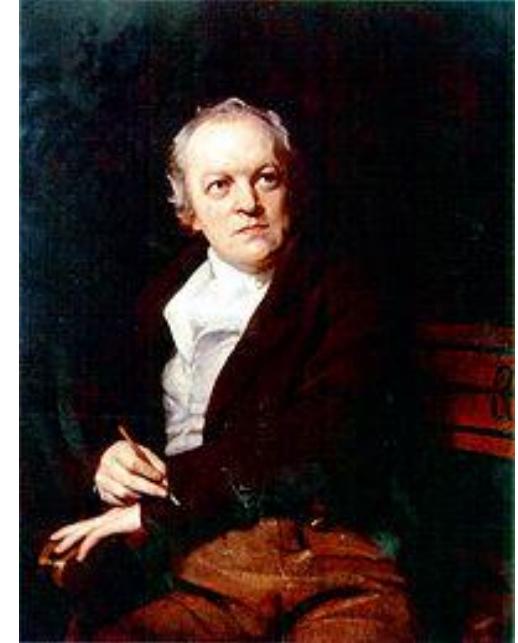
2015 Europe

Source Pubmed. Query:
("high-throughput sequencing"[Title/Abstract] OR "next generation sequencing"[Title/Abstract] OR "rna seq"[Title/Abstract]) AND
"2015"[Publication Date] AND country[Affiliation]

Background

**The road of excess leads to
the palace of wisdom**

(*William Blake, 28 November 1757 – 12 August 1827, poet, painter, and printmaker*)

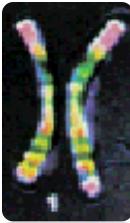


The introduction and popularisation of high-throughput techniques has drastically changed the way in which biological problems **can** be addressed and hypotheses **can** be tested.

But not necessarily the way in which we really address or test them...

Where do we come from? The pre-genomics paradigm

Genes in the DNA...



...code for
proteins...

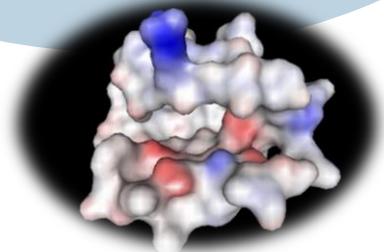
>protein kinase
acctgttgatggcagacagggactgtatgct
atctatgctgtatgcatgcatgctgactactga
tgtggggctattgactgtatgtctatc....

...produces the final phenotype

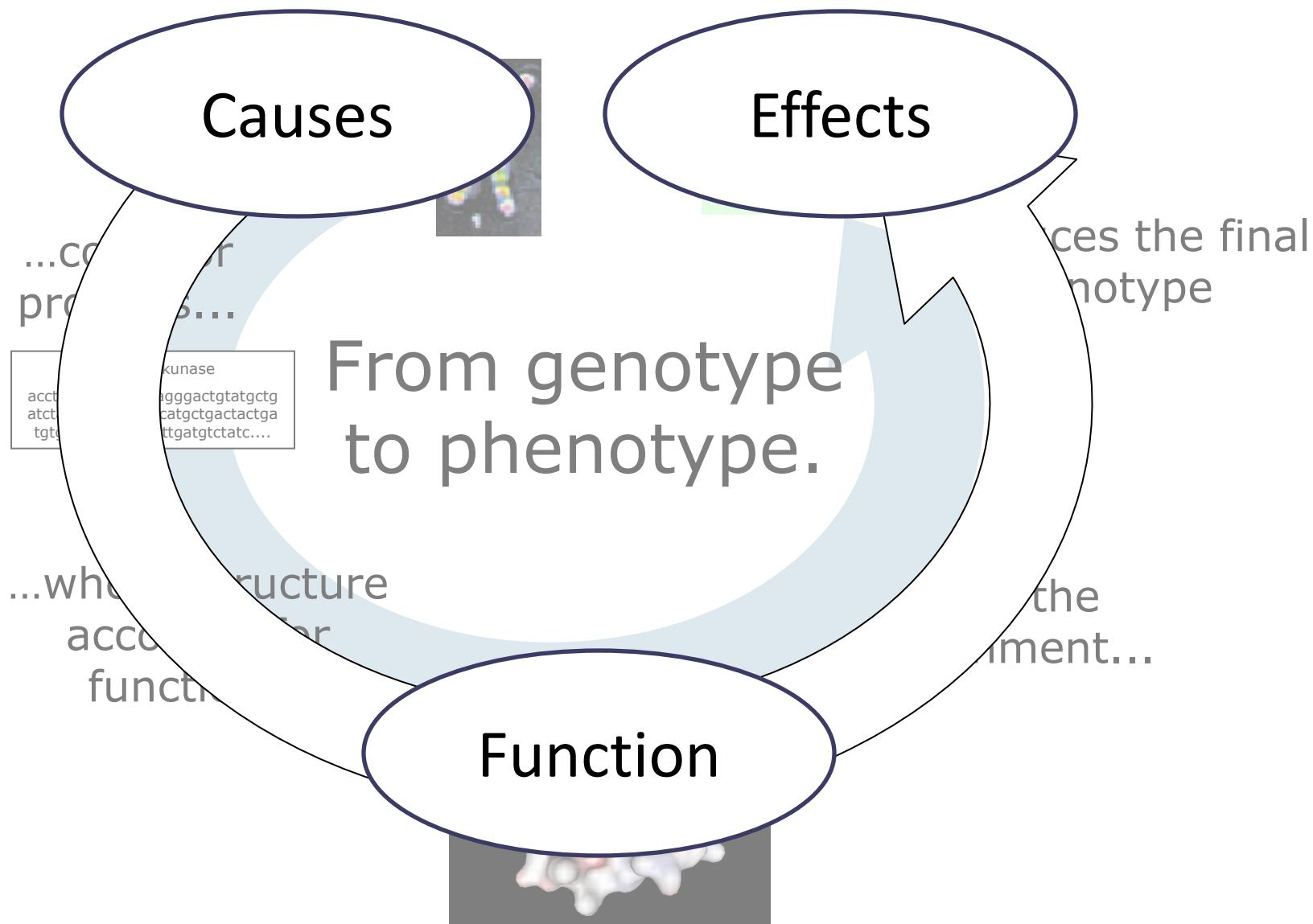
From genotype to phenotype.

...whose structure
accounts for
function...

...plus the
environment...

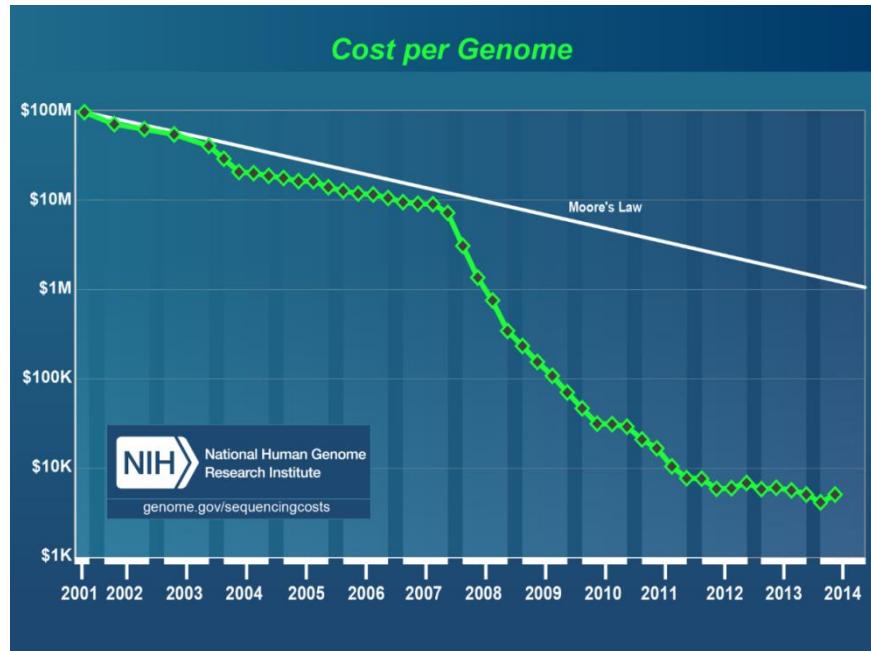


Reductionistic approach to link causes (genome) to effects (phenotype) through actions (function)

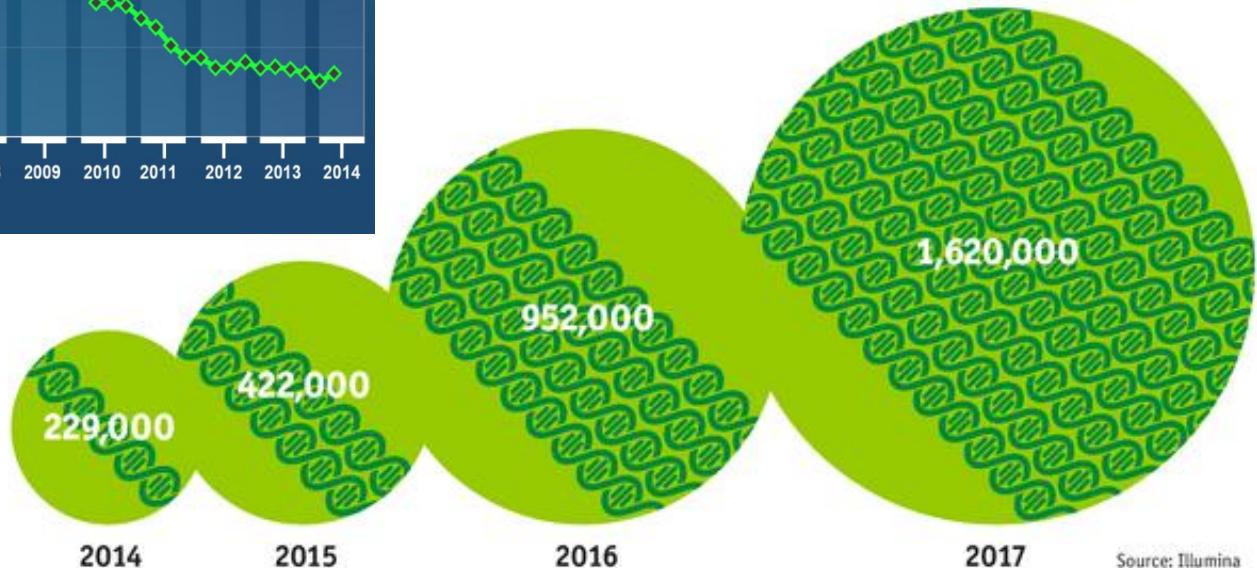


The genome sequencing pace

<http://www.genome.gov/sequencingcosts/>



NGS is matching the cost of many conventional clinical tests

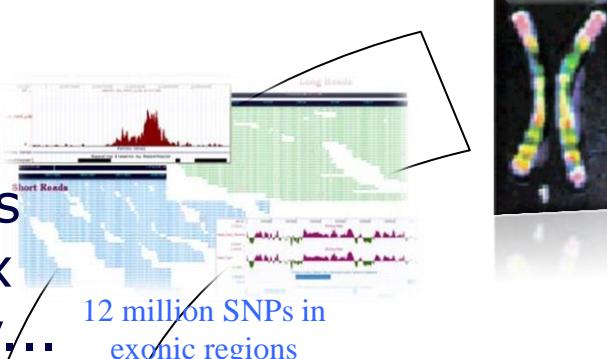


<http://www.economist.com/news/21631808-so-much-genetic-data-so-many-uses-genes-unzipped>

Next Generation Sequencing

10^9 bp per round

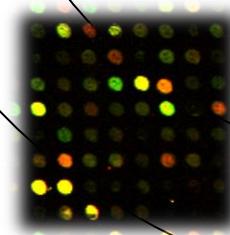
Genes in the DNA...



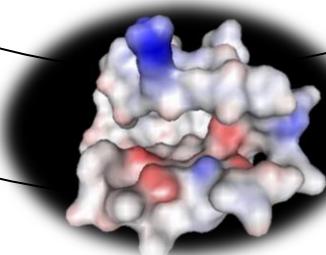
...whose final effect configures the phenotype...

From genotype to phenotype.

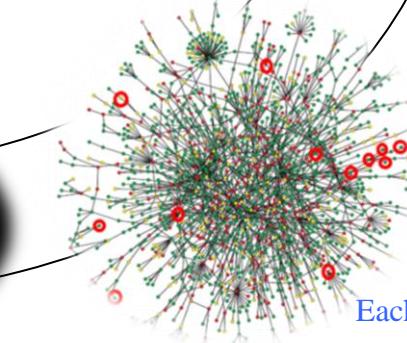
(in the post-genomics scenario)



...code for
proteins...

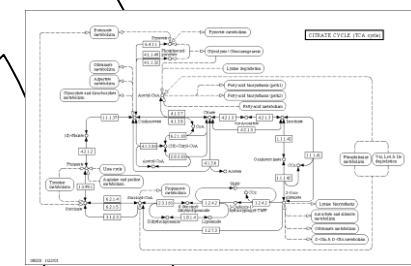


That undergo post-translational modifications, somatic recombination...
100K-500K proteins



Each protein has an average of 8 interactions

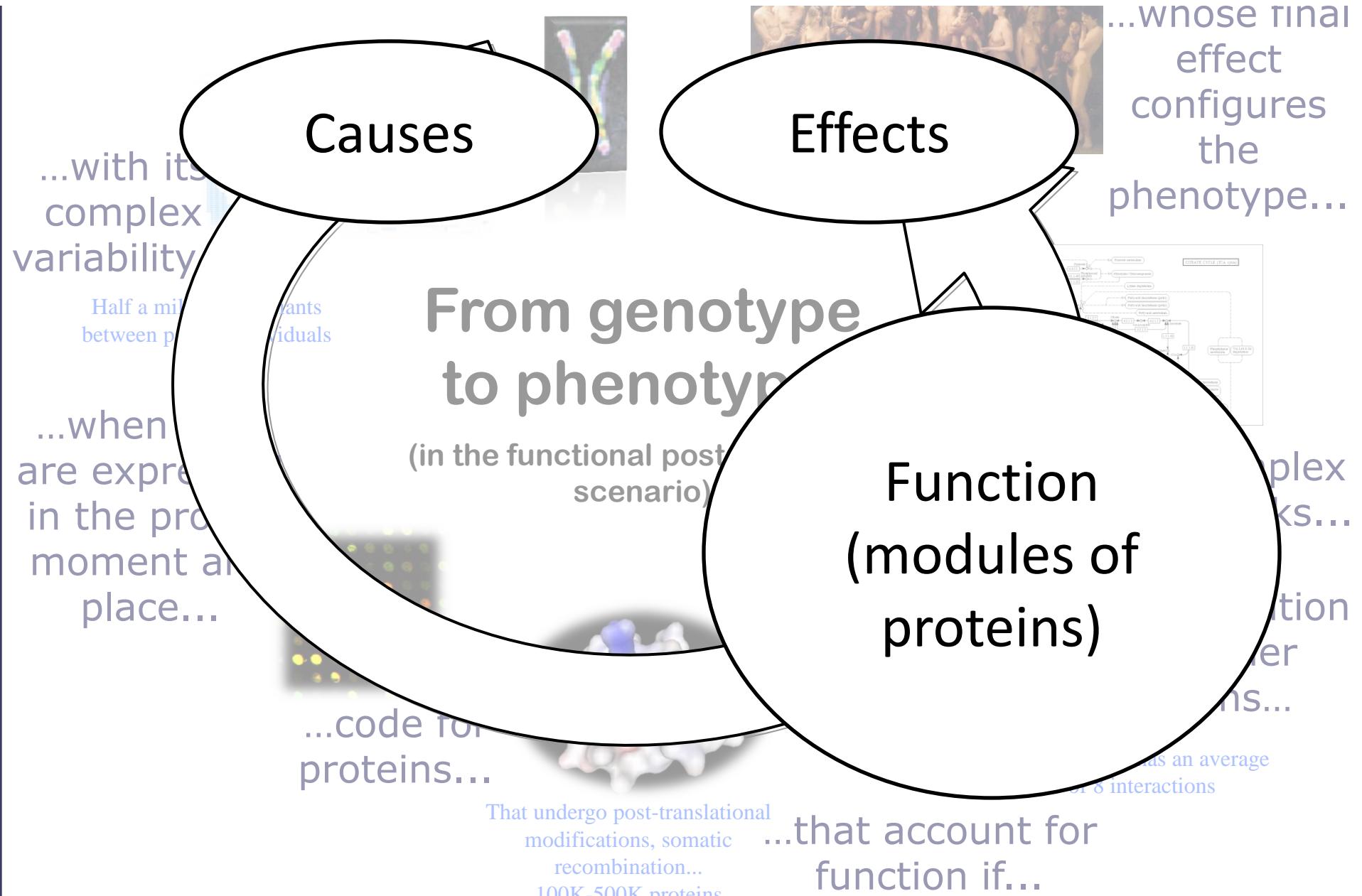
...that account for function if...



...conforming complex interaction networks...

...in cooperation with other proteins...

Holistic approach. Causes and effects remain essentially the same. The concept of function has changed



High-throughput data for functional genomics

Next Generation Sequencing
10⁹bp per round

the DNA

...whose final effect configures the phenotype...

Genotyping

...with its complex variability...

Half a million of variants between pairs of individuals

Genome wide

From genotype to phenotype.

(in the functional post-genomics scenario)

Metabolomics

...conforming complex interaction networks...

Transcriptomics



...code for proteins...

Proteomics

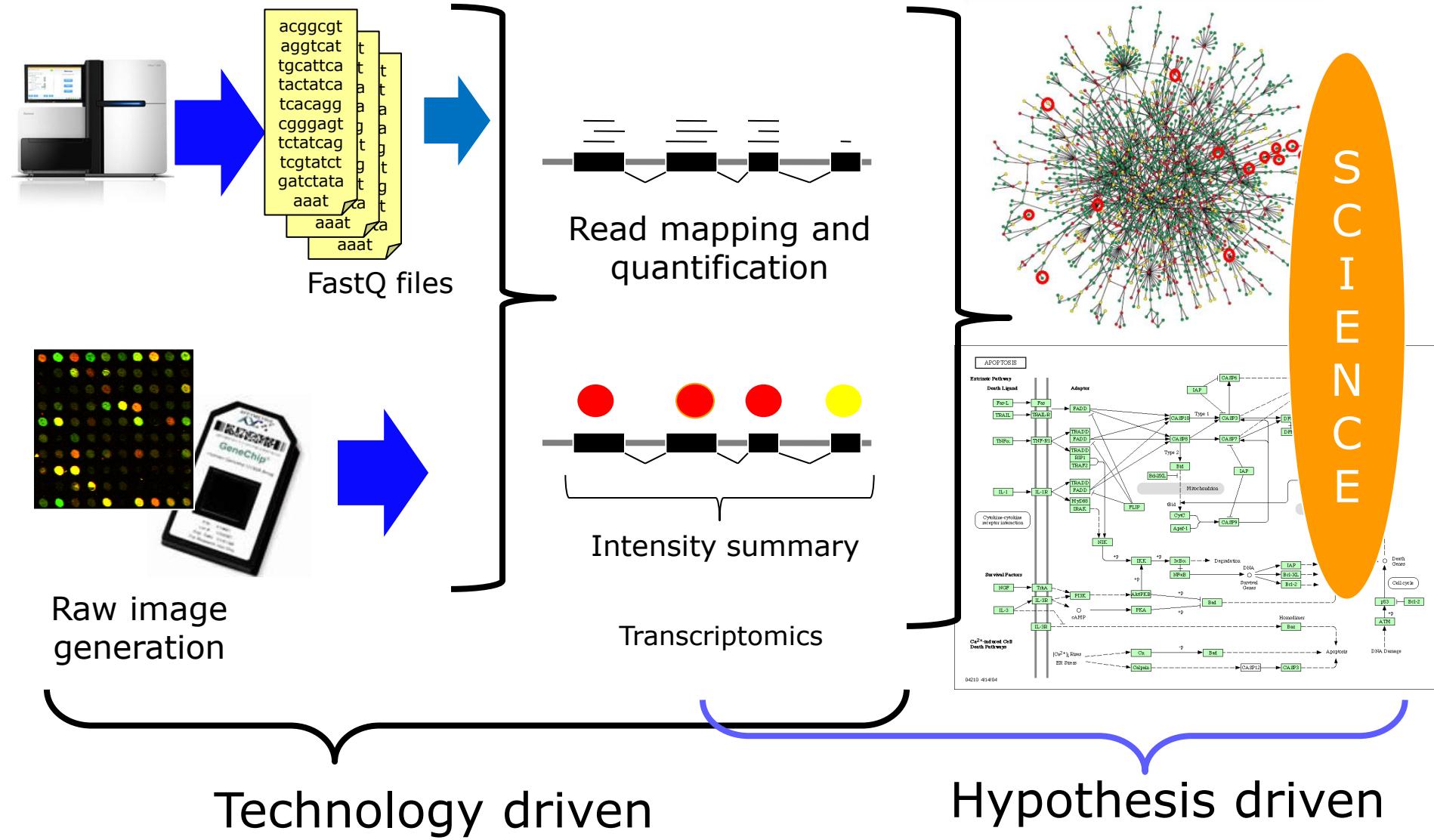
That undergo post-translational modifications, somatic recombination...
100K-500K proteins

...that account for function if...

Almost-omics

Transcriptomics

The double challenge:
Data processing and interpretation



Before analysing your data you must know what is your question.

What is the aim? Class discovery? sample

classification? gene selection? ...

Can we find groups
of experiments with
similar gene
expression profiles?

Molecular
classification of
samples

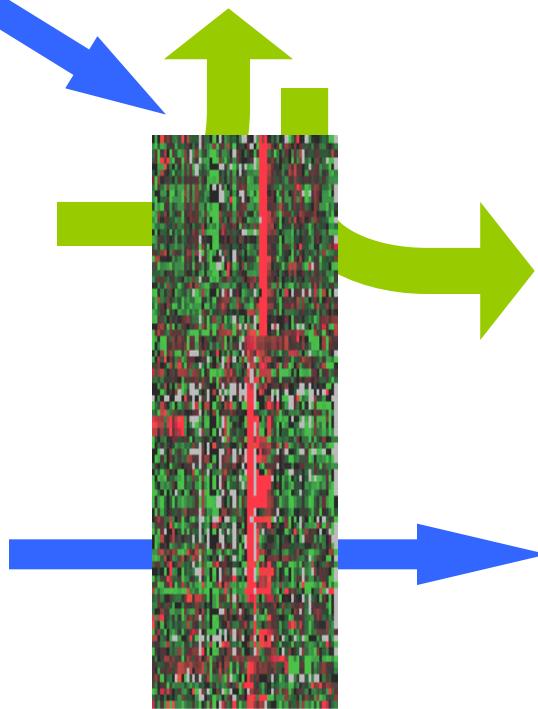
Co-expressing genes...

Different classes...

- Unsupervised
- Supervised

What genes are
responsible for?

What do they
have in
common?



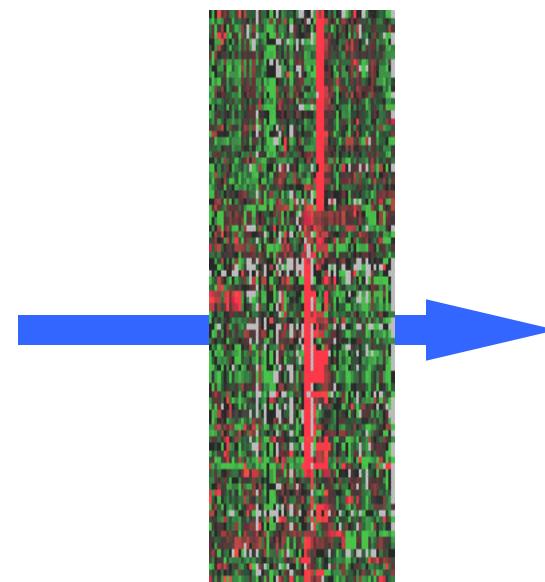
Unsupervised problem: class discovery

Our interest is in discovering clusters of items (genes or experiments) which we do not know beforehand

Can we find groups of experiments
with similar gene expression
profiles?



Co-expressing
genes...



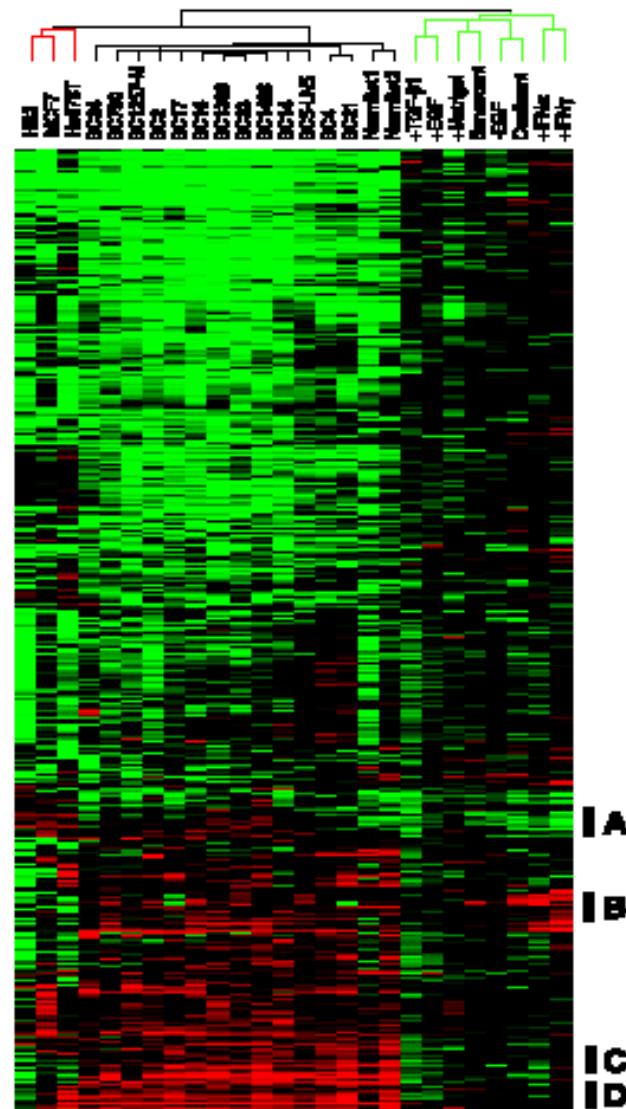
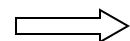
- What genes co-express?
- How many different expression patterns do we have?
- What do they have in common?
- Etc.

Clustering of experiments: The rationale

If enough genes have their expression levels altered in the different experiments, we might be able of finding these classes by comparing gene expression profiles.

Distinctive gene expression patterns in human mammary epithelial cells and breast cancers

Overview of the combined *in vitro* and breast tissue specimen cluster diagram. A scaled-down representation of the 1,247-gene cluster diagram. The black bars show the positions of the clusters discussed in the text: (A) proliferation-associated, (B) IFNregulated, (C) B lymphocytes, and (D) stromal cells.



Perou et al., PNAS (1999)

Supervised problems.

Differential gene expression

Can we find groups
of experiments with
similar gene
expression profiles?

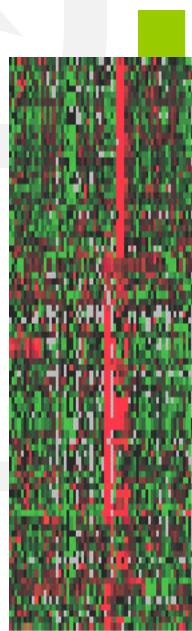
Different classes...

Molecular
classification of
samples

**What genes are
responsible for?**

Co-expressing genes...

What do they
have in
common?



Differential gene expression

The simplest way: univariant gene-by-gene.
Other multivariant approaches can be used

- **Two classes**

- T-test
- Limma
- Fold-change

- **Continuous variable (e.g. level of a metabolite)**

- Pearson
- Spearman
- Regression

- **Multiclass**

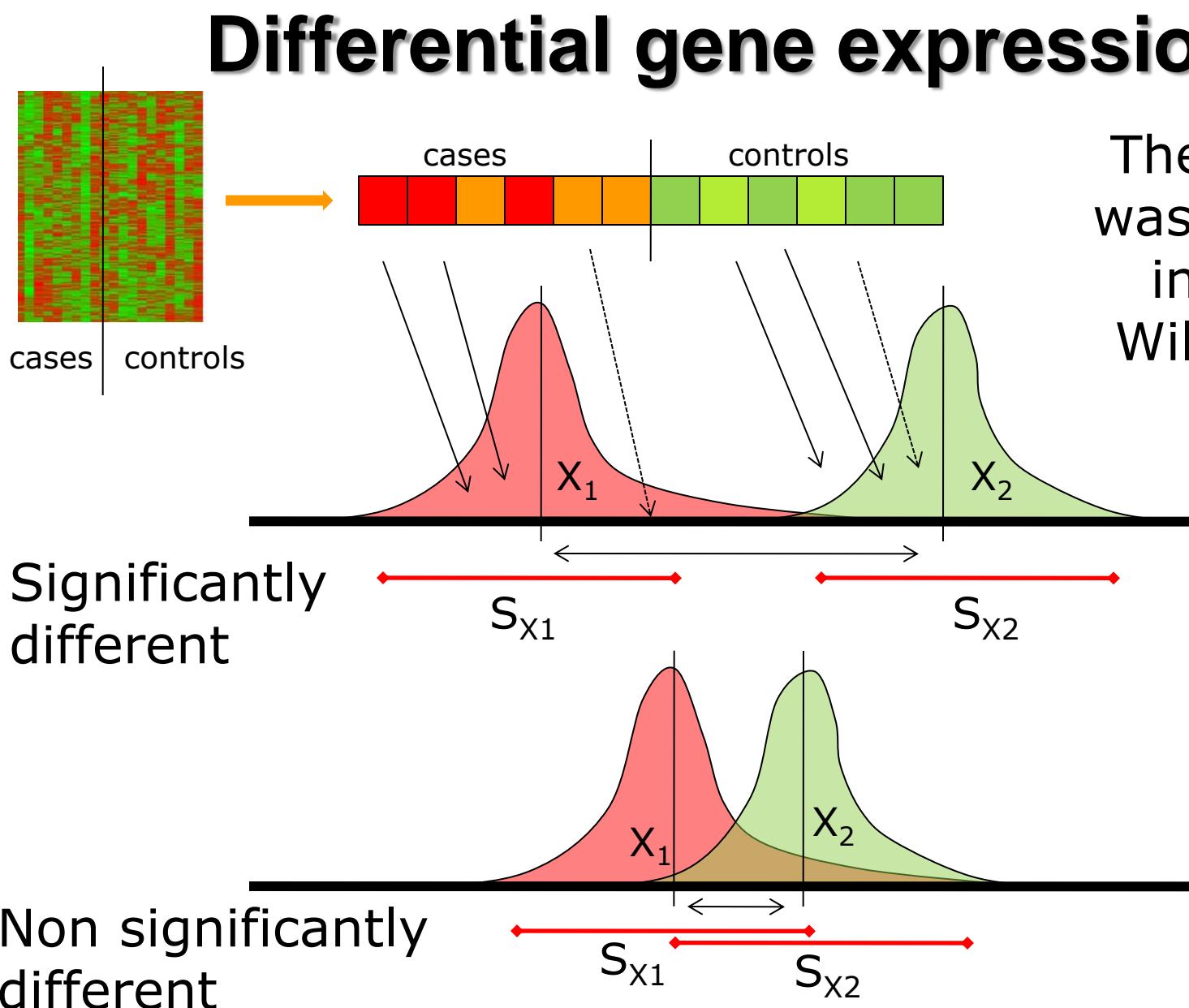
- Anova
- Limma

- **Survival**

- Cox model

- **Time Course**

Differential gene expression



The t-statistic was introduced in 1908 by William Sealy Gosset

$$t = \frac{\bar{X}_1 - \bar{X}_2}{S_{X_1 X_2} \cdot \sqrt{\frac{2}{n}}} \quad \text{being} \quad S_{X_1 X_2} = \sqrt{\frac{S_{X_1}^2 + S_{X_2}^2}{2}}.$$

Supervised problems.

sample classification

Can we find groups
of experiments with
similar gene
expression profiles?

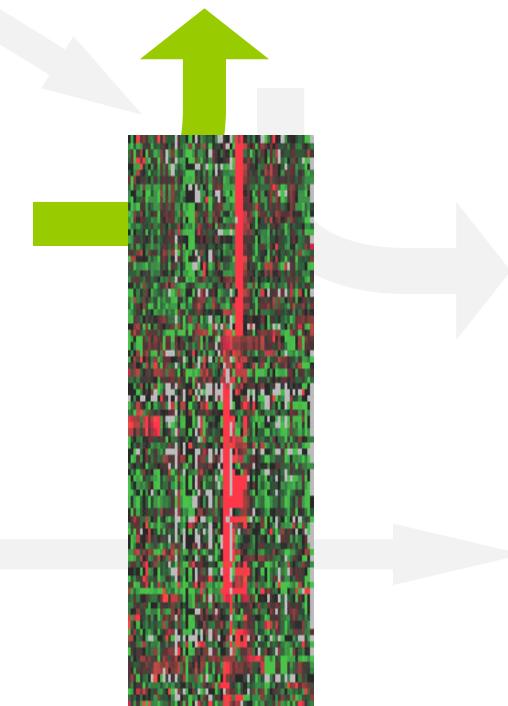
Different classes...

Molecular
classification of
samples

What genes are
responsible for?

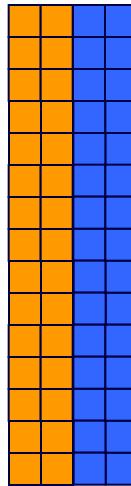
Co-expressing genes...

What do they
have in
common?



Predictors and molecular signatures

A B X



Is X,
A
or B?

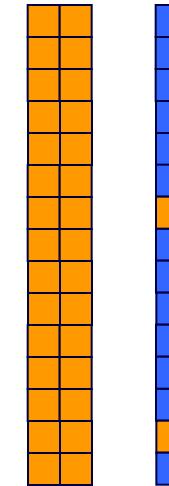


Diff (B, X) = 2



What is a predictor?

Intuitive notion:



Diff (A, X) = 13

Most probably X belongs to class B

Algorithms: DLDA, KNN, SVM, random forests, PAM, etc.

Predictor of clinical outcome in breast cancer



Genes are arranged to their correlation with the prognostic groups

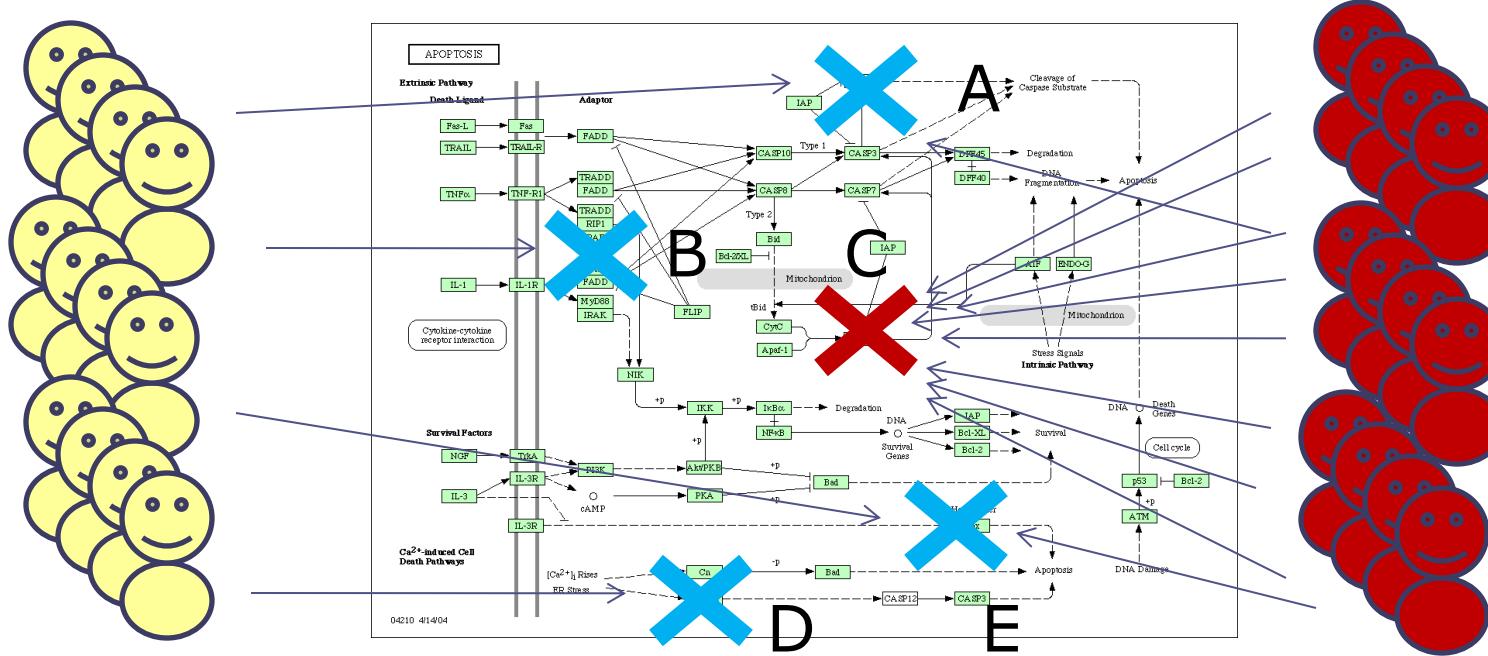
Pronostic classifier with optimal accuracy

van't Veer et al.,
Nature, 2002

Genotyping/Resequencing:

Finding mutations associated to diseases

The simplest case: monogenic disease



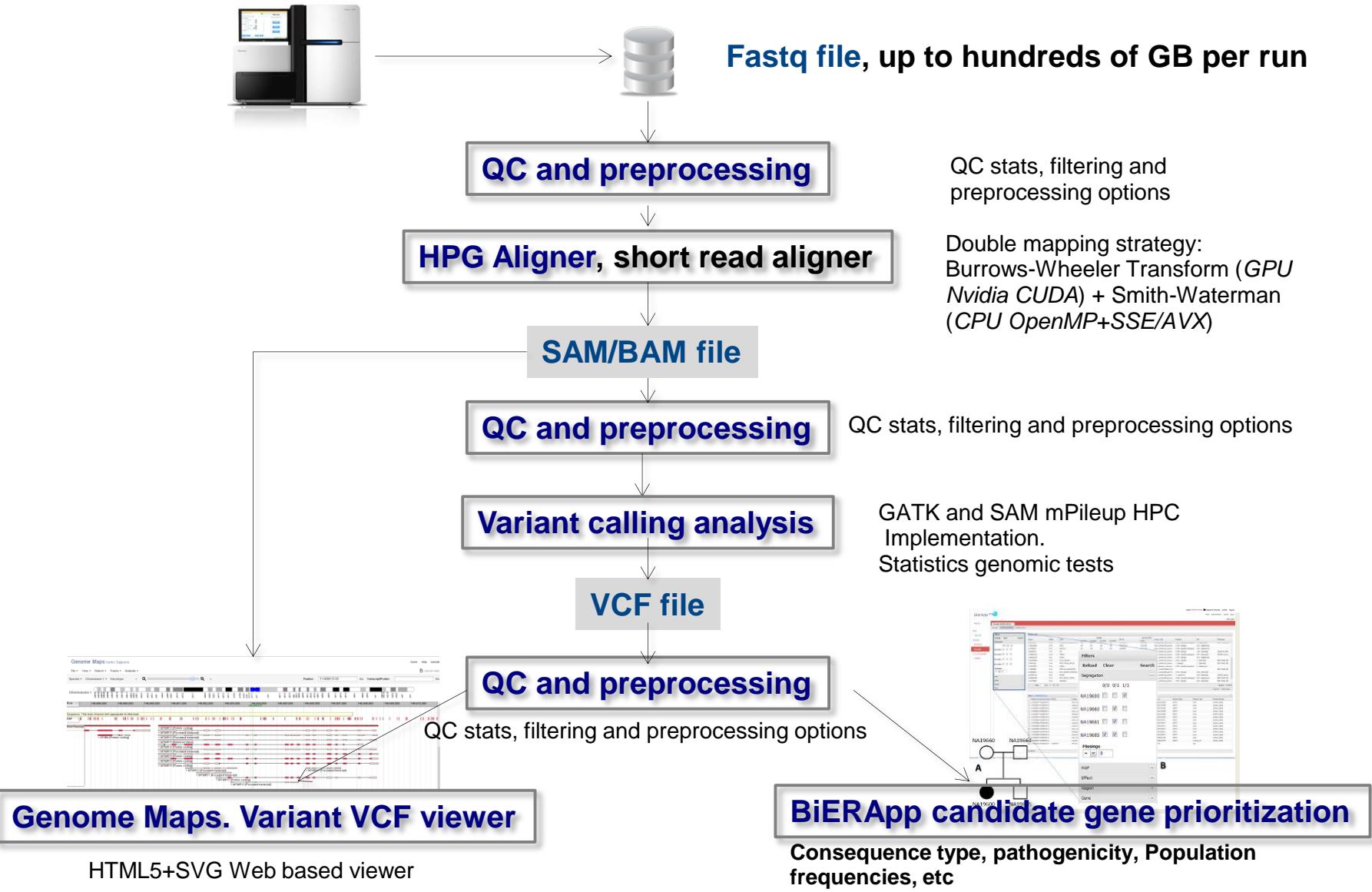
Controls

Cases

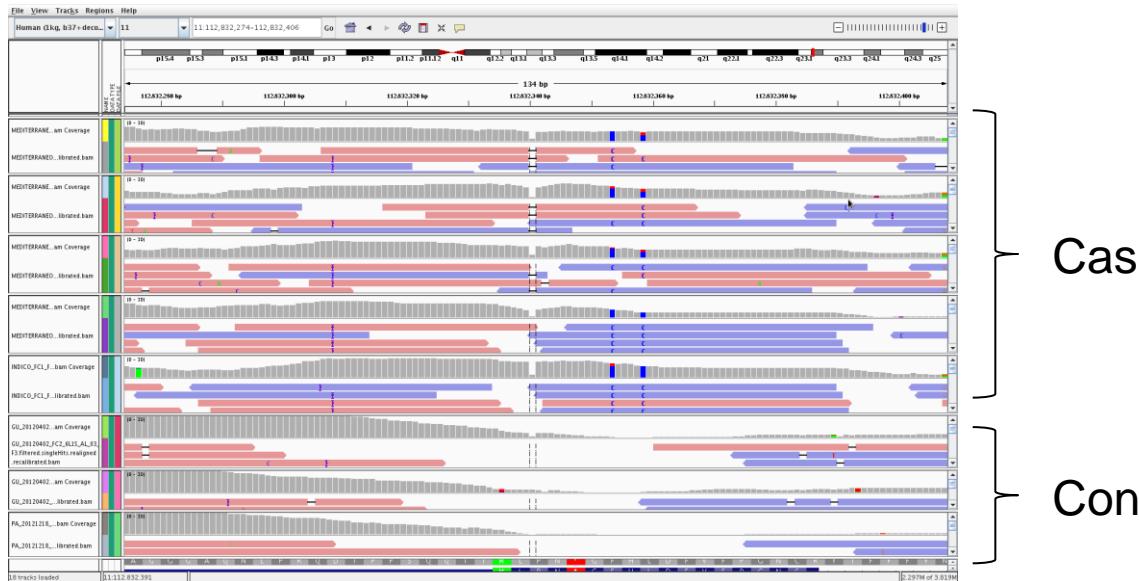
Gene A	1 0 0 0 0 0 0 0 0 0 0 0	0 0 0 0 1 0 0 0 0 0 0 0
Gene B	0 0 0 1 0 0 0 0 0 0 0 0	0 0 0 0 0 0 0 0 0 0 0 0
Gene C	0 0 0 0 0 0 0 0 0 0 0 0	1 1 1 1 1 1 1 1 1 1 1 1
Gene D	0 0 0 0 0 0 0 0 0 0 1 0	0 0 0 0 0 0 0 0 0 0 0 0
Gene E	0 0 0 0 0 1 0 0 0 0 0 0	0 0 0 0 0 0 1 0 0 0 0 0



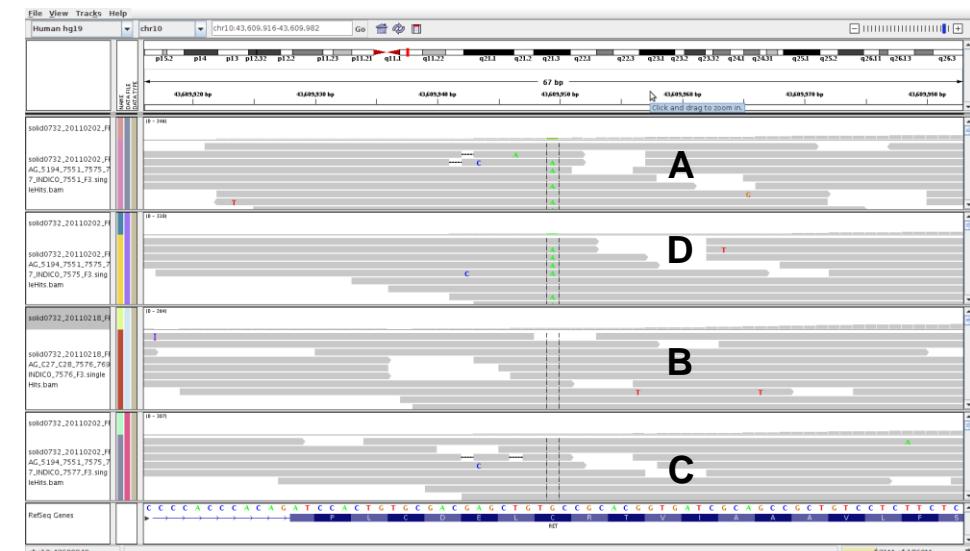
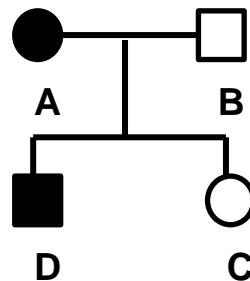
Primary data analysis tools



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



Segregation
within a
pedigree



Variant/gene prioritization by successive filtering



Variant level

Potential impact of the variant

Population frequencies

Experimental design level

Family(es)
Trios
Case / control

Functional (system) level

Gene set
Network analysis
Pathway analysis

Control of sequencing errors (missing values)

Testing strategies



Pipeline of data analysis

Initial QC

Sequence cleansing
Base quality
Remove adapters
Remove duplicates

FASTQ file

Mapping + QC

Mapping (HPG)
Remove multiple mapping reads
Remove low quality mapping reads
Realigning
Base quality recalibrating

BAM file

Variant calling + QC

Calling and labeling of missing values
Calling SNVs and indels (GATK) using 6 statistics based on QC, strand bias, consistency (poor QC callings are converted to missing values as well)
Create multiple VCF with missing, SNVs and indels

VCF file

Variant and gene prioritization + QC

Counts of sites with variants
Variant annotation (function, putative effect, conservation, etc.)
Inheritance analysis (including compound heterozygotes in recessive inheritance)
Filtering by frequency with external controls (**Spanish controls**, dbSNP, 1000g, 5500g) and annotation
Multi-family intersection of genes and variants
Network-based prioritization
Report

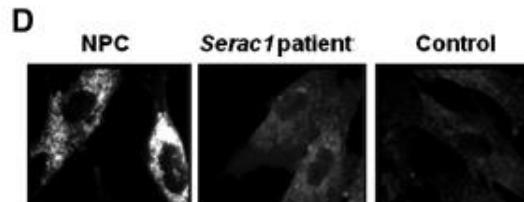
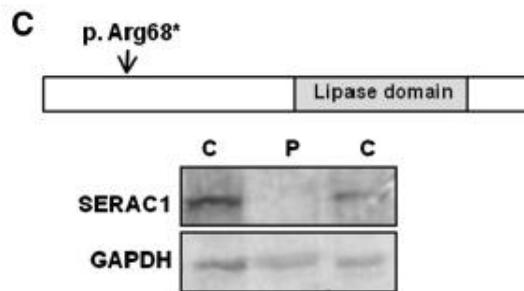
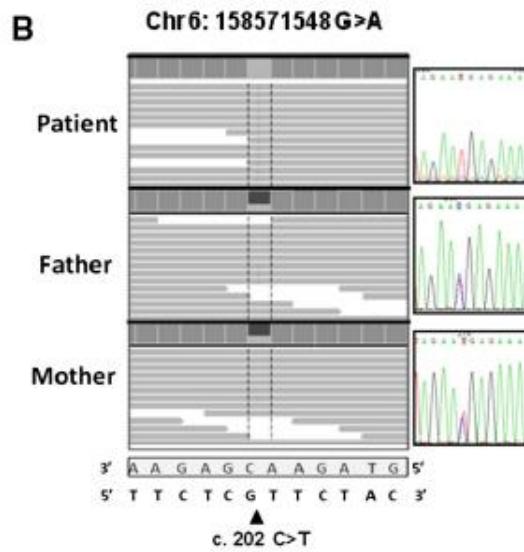
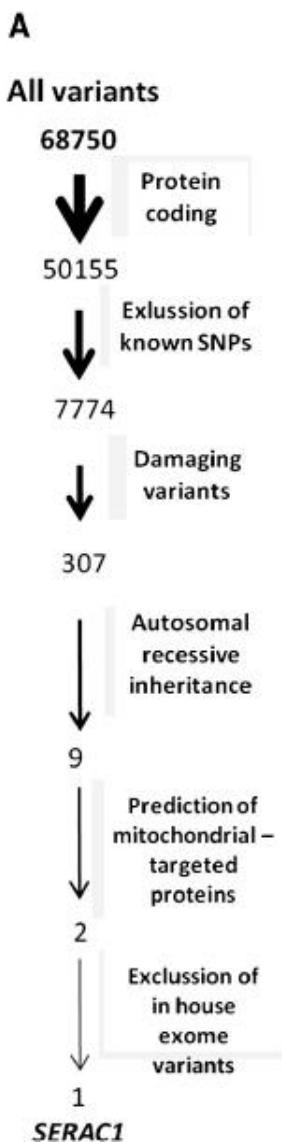
Primary analysis

Gene prioritization

Successive 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-MGAuria) 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.



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 García-Villoria ^{a,b}, María 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 Mitochondriales- Enfermedades Metabólicas Hereditarias, Servicio de Pediatría, Hospital 12 de Octubre, Madrid, Spain

^d BIER, CIBERER, Centro de Investigación Príncipe Felipe (CIPF), Valencia, Spain

^e Computational Medicinal 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

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 Lee¹, Abigail W Bigham², Holly K Tabor^{2,3}, Karin M Dent⁴, Chad D Huff⁵, Paul T Shannon⁶, Ethylin Wang Jabs^{7,8}, Deborah A Nickerson¹, Jay Shendure¹ & 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 40x, and sufficient depth to call variants at ~97% of each targeted exon. 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.

IN THIS ISSUE | May 2011 | Volume 19 | Issue 5

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^{2,3}, Marina Donskoi¹, Tony J. Kenna¹, Gethin P. Thomas¹, Graeme R. Clark¹, 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, Herston, Australia, ³ School of Medicine, Faculty of Health Sciences, The University of Queensland, Herston, Australia

Abstract

Recent advances in DNA sequencing have enabled mapping of genes for monogenic traits in families with small pedigrees and even in unrelated cases. We report the identification of disease-causing mutations in a rare, severe, skeletal dysplasia,

European Journal of Human Genetics (2011) 19, 115–117
© 2011 Macmillan Publishers Limited All rights reserved 108-4813/11
www.nature.com/ejhg



. The two
re form of
sequencing
as a core
lMPR RNA
activity of
which *POP1*

lations As the
which permits

establishment
the Rebecca
by a National
The funders

SHORT REPORT

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

Kimia Kahrizi¹, Cougar Hao Hu², Masoud Garshabi², Seyedeh Sedigheh Abedini¹, Shirin Ghadami¹, Roxane Kariminejad³, Reinhard Ullmann⁴, Wei Chen², H-Hilger Ropers², Andreas W Kuss², Hossein Najmabadi¹ and Andreas Tschach^{*2,5}

As part of a large-scale, systematic effort to unravel the molecular causes of autosomal recessive mental retardation, we have previously described a novel syndrome consisting of mental retardation, coloboma, cataract and kyphosis (Kahrizi syndrome)

OMIM 612713

array-based

(c.203kb)

interval.

essential

families

and eye

potential

European

Keywords:

consanguinity

MV 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

© 2013 Molecular Vision

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

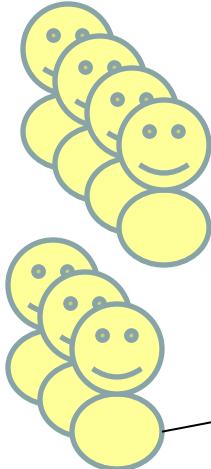
Cristina Méndez-Vidal,^{1,2} María González-del Pozo,^{1,2} Alicia Vela-Boza,³ Javier Santoyo-López,³ Francisco J. López-Domínguez,³ Carmen Vázquez-Marouschek,⁴ Joaquín Dopazo,^{3,5,6} Salud Borrego,^{1,2} Guillermo António,^{1,2,3}

¹Department of Genetics, Reproduction and Fetal Medicine, Institute of Biomedicine of Seville, University Hospital Virgen del Rocío/CSIC/University of Seville, Seville, Spain; ²Centro de Investigación Biomédica en Red de Enfermedades Raras (CIBERER), Seville, Spain; ³Medical Genome Project, Genomics and Bioinformatics Platform of Andalucía (GBPA), Seville, Spain;

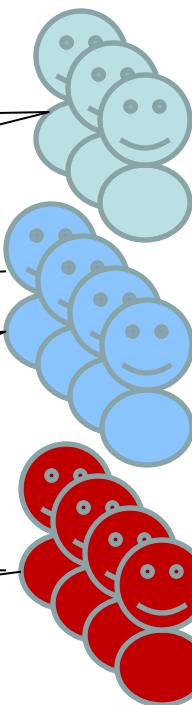
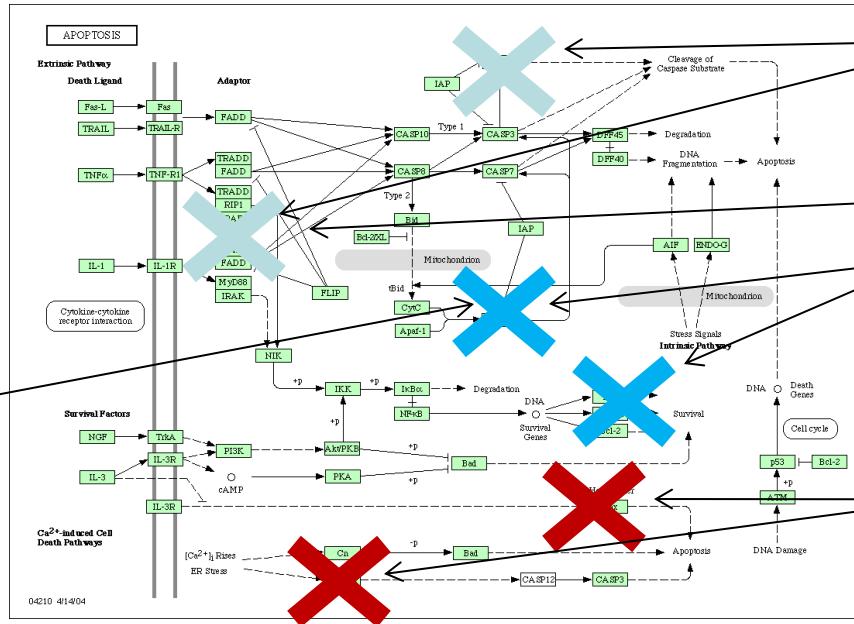
⁴Department of Ophthalmology, University Hospital Virgen del Rocío, Seville, Spain; ⁵Department of Bioinformatics, Centro de Investigación Príncipe Felipe, Valencia, Spain; ⁶Functional Genomics Node (INB), Centro de Investigación Príncipe Felipe, Valencia, Spain

An approach inspired on systems biology can help in detecting causal genes

Controls



Cases



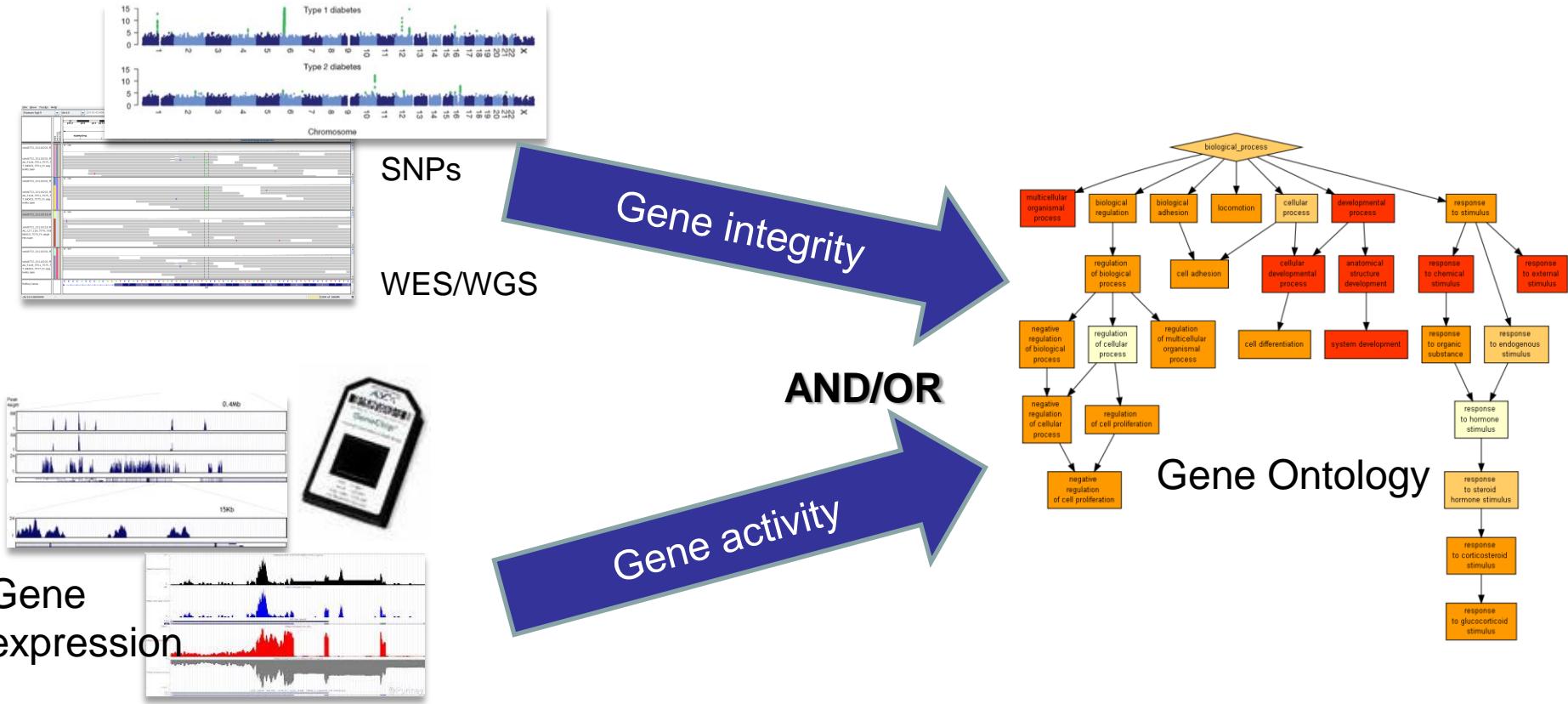
Affected **cases** in complex diseases will be a **heterogeneous** population with different mutations (or combinations).

Many cases and controls are needed to obtain significant associations.

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

Disease understood as the failure of a functional module

From gene-based to function-based perspective



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.

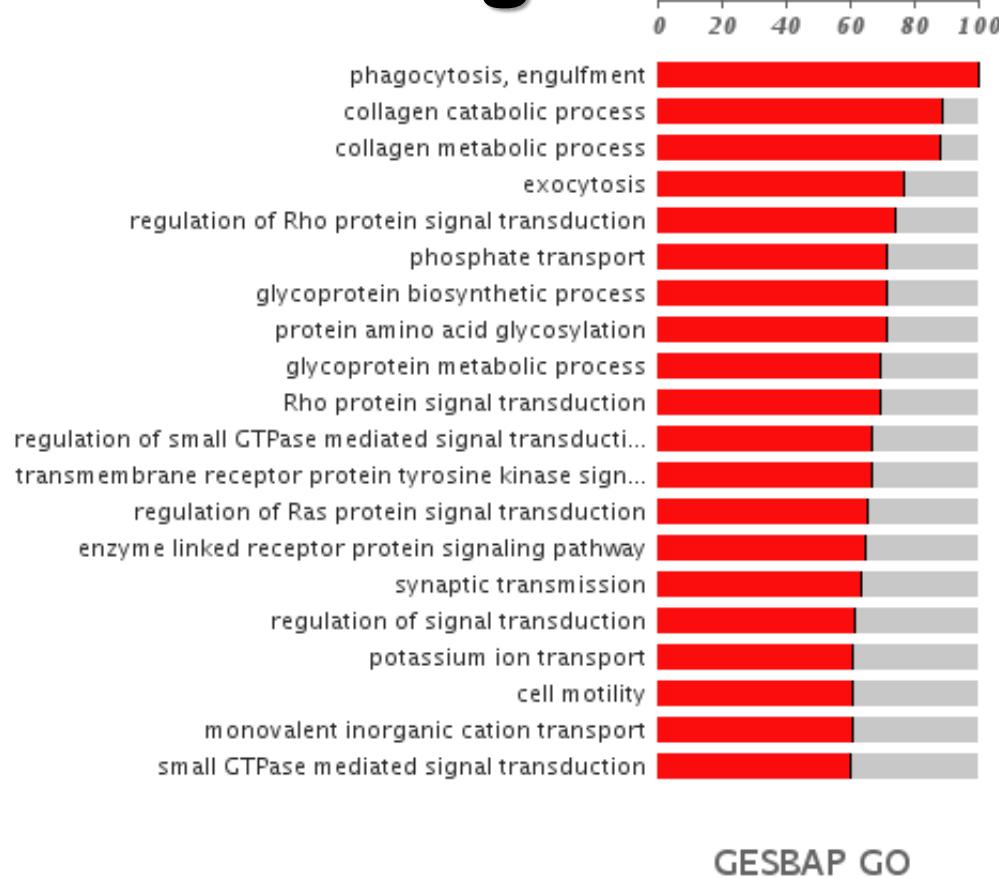
The CGEMS initiative. (Hunter et al. Nat Genet 2007)

1145 cases 1142 controls. Affy 500K

Conventional association test reports only 4 SNPs
significantly mapping 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

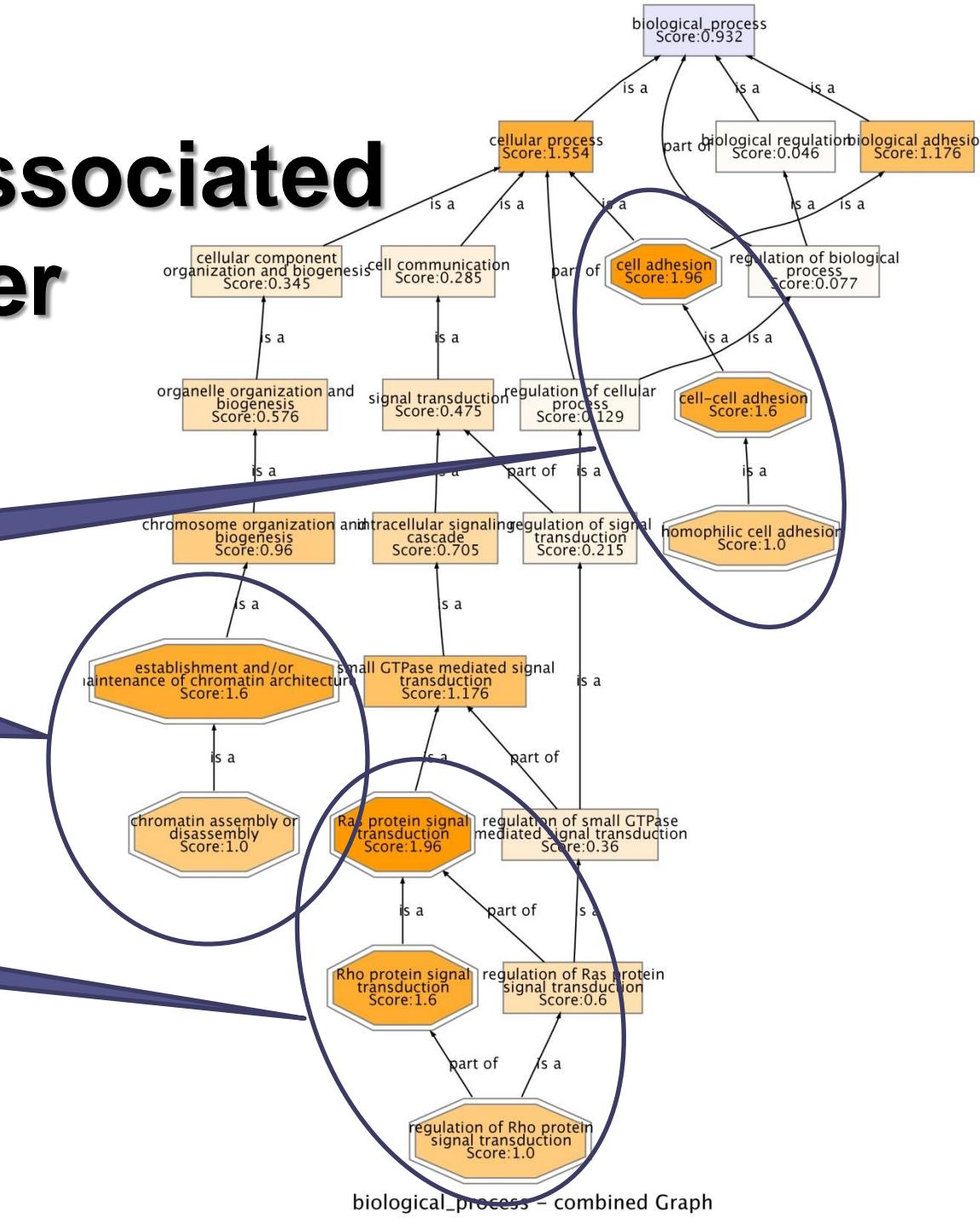
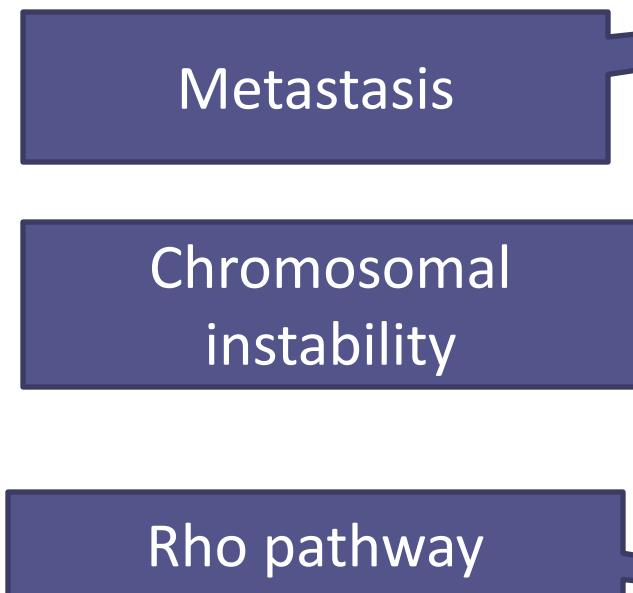
CGEMS initiative.
(Hunter et al. Nat Genet 2007)

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=4.45x10⁻³) in which FGFR2 is included.

GO processes significantly associated to breast cancer



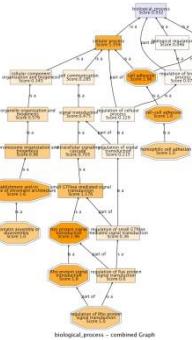
From gene-based to function-based perspective

SNPs,
Gene expression

Gene₁
Gene₂
Gene₃
Gene₄
:
:
:
Gene₂₂₀₀₀

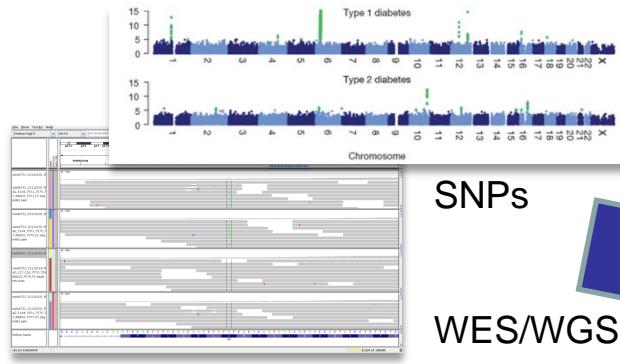


Gene
Ontology



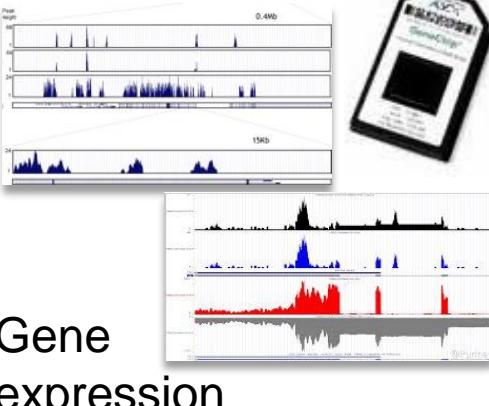
	SNPs, gene exp.	GO
Detection power	Low (only very prevalent genes)	high
Annotations available	many	many
Use	Biomarker	Illustrative, give hints

From gene-based to function-based perspective



SNPs

WES/WGS



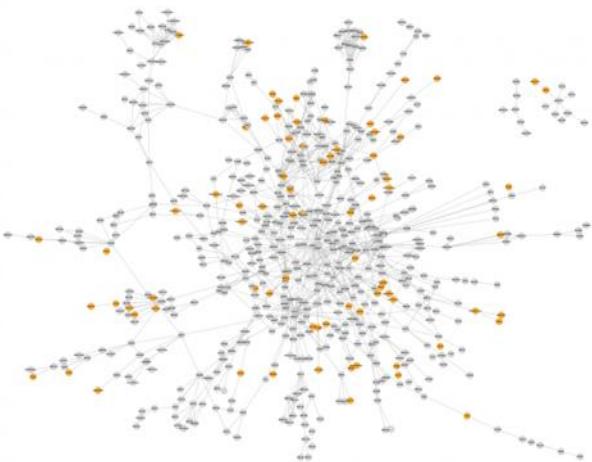
Gene expression

Gene integrity

AND/OR

Gene activity

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



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-Alonso³, Ana Torroglosa^{1,2}, Martina Marbà^{3,4}, Ma Valle Enguix-Riego^{1,2}, David Montaner³, Guillermo Antiñolo^{1,2}, Joaquín Dopazo^{2,3,4*} and Salud Borrego^{1,2*}

* Corresponding authors: Joaquín Dopazo idopazo@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

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-Alonso¹, Roberto Alonso¹, Enrique Vidal¹, Alicia Amadoz¹, Alejandro de María¹, Pablo Minguez², Ignacio Medina^{1,3} and Joaquín Dopazo^{1,3,4,*}

¹Department of Bioinformatics, Centro de Investigación Príncipe Felipe (CIPF), Valencia, Spain, ²European Molecular Biology Laboratory, Meyerhofstrasse 1, 69117 Heidelberg, Germany, ³Functional Genomics Node (INB) at CIPF, Valencia and ⁴CIBER 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)
RASGEF1A* (rs1254964 p = 3.856x10⁻⁰⁵)

*no interactions known (yet)

SNPs validated in independent cohorts

Nucleic Acids Research Advance Access published May 19, 2009

Nucleic Acids Research, 2009, 37–6
doi:10.1093/nar/gkp402

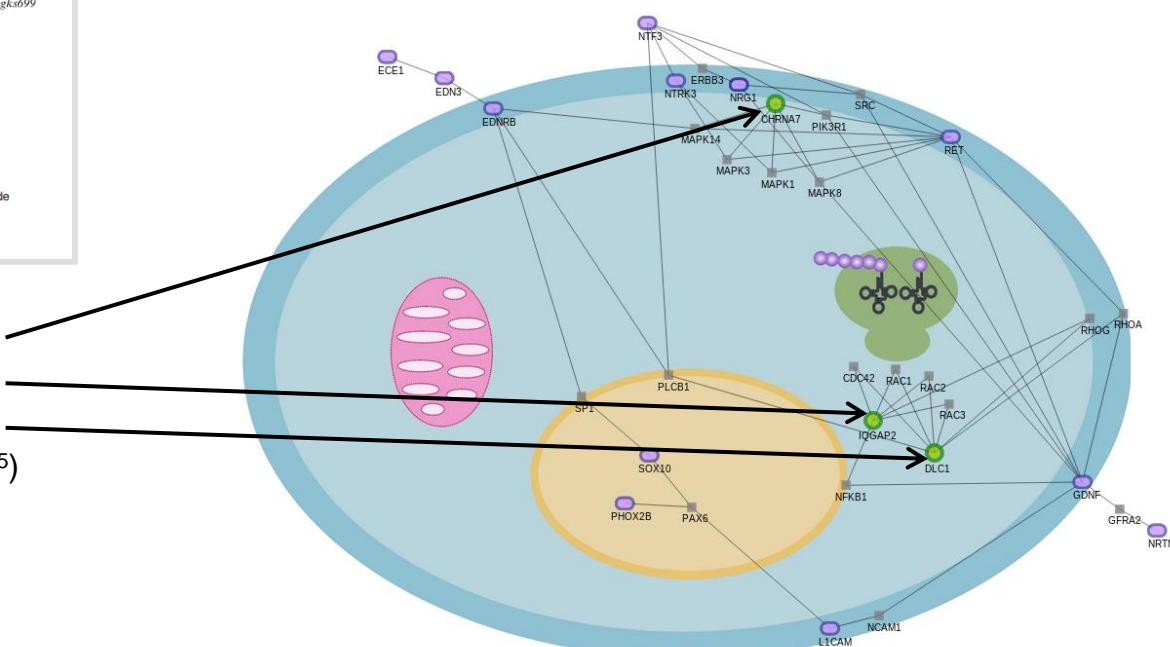
SNOW, a web-based tool for the statistical analysis of protein-protein interaction networks

Pablo Minguez¹, Stefan Götz^{1,2}, David Montaner¹, Fatima Al-Shahrour¹ and Joaquin Dopazo^{1,2,3,*}

¹Department of Bioinformatics and Genomics, Centro de Investigación Príncipe Felipe (CIPF),

²CIBER de Enfermedades Raras (CIBERER) and ³Functional Genomics Node (INB) at CIPF, Valencia, Spain

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



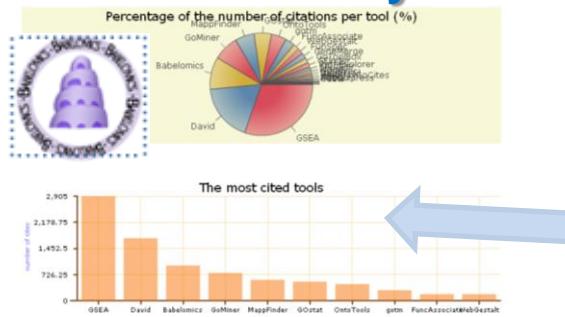
From gene-based to function-based perspective

	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

Software development

Functional analysis



Babelomics is the third most cited tool for functional analysis. Includes more than 30 tools for advanced, systems-biology based data analysis



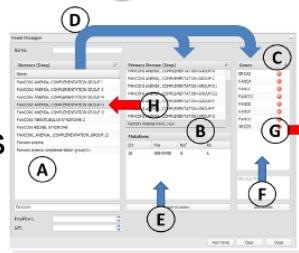
Mapping

HPC on CPU, SSE4,
GPUs on NGS data
processing
Speedups up to 40X

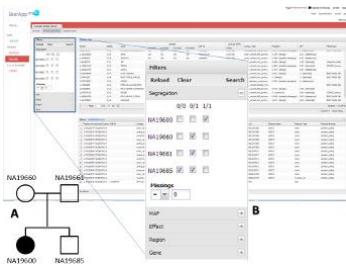
Visualization



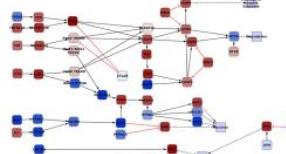
Diagnostic



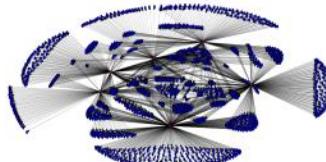
Variant prioritization



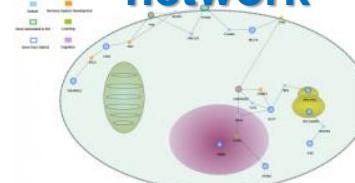
Signaling network



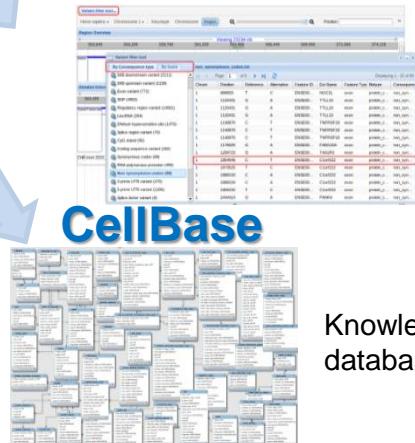
Regulatory network



Interaction network



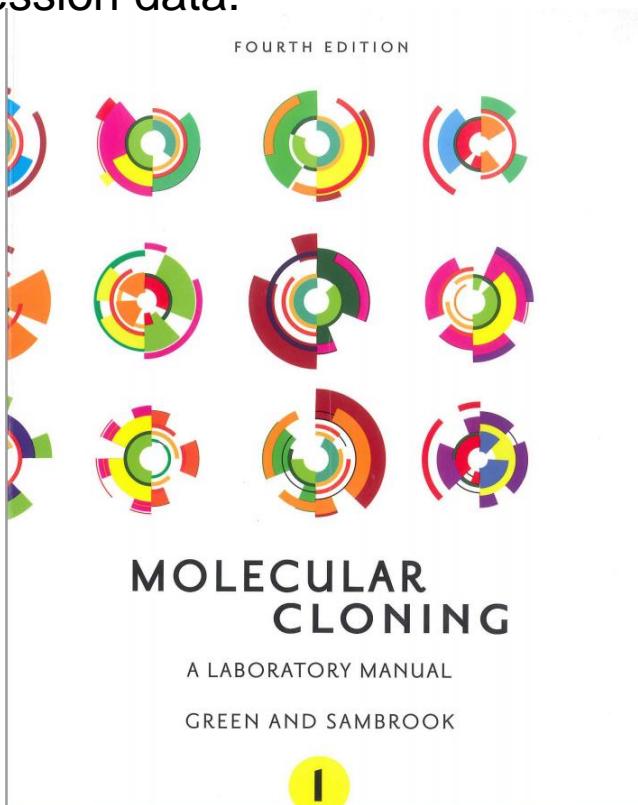
Variant annotation



More than 150.000 experiments were analyzed in our tools during the last year

Babelomics in the Maniatis

The Babelomics suite of programs becomes a classic. Now is cited as a method in the last edition of **Molecular Cloning**, the popular **Maniatis**. The protocol 4 of chapter 8, Expression Profiling by Microarray and RNA-seq, contains a description on how to use Babelomics to analyze expression data.



High impact developments

A screenshot of the Babelomics 4 web interface. The top navigation bar includes links for Home, Babel, Utilities, and Examples. The page title is "578 / Chapter 8". The main content area is titled "BABELOMICS 4 gene expression and functional profiling analysis suite". It shows a "Upload data" form with a "Browse" button highlighted with a red oval. To the right, a "Select format" panel is open, showing a tree view of data types: "Microarray - Expression - One-channel - Affymetrix" is selected. Other options like "Two-channel", "Gene", "Protein", and "Metabolite" are also listed. A green arrow points from the "One-channel" link in the tree to the "Select format" panel. The "Data list" panel on the right shows a single entry: "cMyc_raw_data". At the bottom right of the upload form, there is an "Accept" button highlighted with a red oval.

FIGURE 1. Babelomics data uploading form. Click "Browse" to upload the data file named cMyc.zip. Select "Affymetrix" as the format, click "Accept" in the pop-up "Select format" panel, and assign the name as "cMyc_raw_data." Click "Upload" to submit the file.

- iii. Assign "cMyc_raw_data" as the data name.
- iv. Click "Upload" to submit the files and wait for the validation to complete. All submitted data are listed in the "Data list" panel.
3. When the data submission is finished, its status in the "Data list" panel changes to "valid."
 - i. To check the expression intensity of the raw data before normalization, click the "Microarray raw-data plot" link in the "Utilities" tab.
 - ii. In the page followed by the link, click "browse server," select "Uploaded data" → "cMyc_raw_Data," and click "Accept."
 - iii. Set the job name as "CMyc_original_boxplot" and click "Run."
 - iv. After the job is finished, click it in the "Job list" panel and the "Box-plots" link to view the box plots as shown in Figure 2.

Each box plot displays summary statistics of a sample, with the box containing the middle 50% of the data, the upper (lower) edge of the box indicating 75th (25th) percentile of the data, and the vertical lines (whiskers) indicating maximum and minimum values. We can see that the eight data sets in our example have systematically different distributions of intensities.

SOCIAL:

MDA group in LinkedIn

Babelomics group in Facebook and twitter

Facebook | Babelomics fans - Mozilla Firefox

Archivo Editar Ver Historial Marcadores Herramientas Ayuda

Más visitados Spanish English diction... ScholarOne Manuscripts Google Académico Google Reader (116) PubMed Publicaciones electrónicas WOK Projects - Ximo...

SEBBM - So... Sage Biometr... Pathway Se... Supplemen... Transcrip... XXXIII Congr... used_pathw... http://s.aspx Module netw... PLoS ONE... BioMed Cent... Faceb...

facebook Search

Babelomics fans

Wall Info Discussions Photos Video Events

Write something...

Attach:

Joaquin Dopazo The new version Babelomics 4.0 has officially released. Enjoy it! A few seconds ago Comment Like Report

Message all members Promote Group with an Advert Edit group settings Edit members Invite people to join Create group event Leave Group

Write something about Babelomics fans.

Information

Category: Internet & Technology - Software Description: Babelomics is the integrative platform for the analysis of transcriptomics, ... Transfiendo datos desde static.fbcdn.net...

Inicio 3 Micro... 3 Firefox 10 Expl... 5 Micro... Skype™ 2 Micro...

LinkedIn: MDA Group - Mozilla Firefox

Archivo Editar Ver Historial Marcadores Herramientas Ayuda

Más visitados MailSite Express Google Académico PubMed Home FireFTP Google Maps

LinkedIn: MDA Group

LinkedIn People Jobs Answers Companies

Explore People Search: Director at Apple - Internet (Silicon Valley) - VP Operations

Search People

Account & Settings Help Sign Out Language

Groups

MDA

Computational Biology Genomics: Next Gen Sequencing... See all »

Profile Edit My Profile View My Profile Recommendations

Contacts Connections Imported Contacts Network Statistics

Inbox (6) Compose Message Received (6) Sent Archived

Applications Reading List by Amazon

Overview Discussions News Jobs Updates Members Settings Manage Group Profile

Activity: Last 7 Days

5 New members 1 New discussion

Announcement from MDA By Joaquin Dopazo 1 day ago Following Stop Add comment >

Recent Activity Recent Discussions Most Comments

Discussions Start a Discussion

My discussions | My comments | See all discussions >

News

Submit News for Discussion

LinkedIn

Dolphin-Assisted Therapy May 06 2009, This is a virtual ... Learn about Dolphin-Assisted Th... events.linkedin.com

From: Alison Bodor

Bilingual Italian Books Contemporary Italian books with English glossaries for easy reading

@babelomics



PRINCIPE FELIPE
CENTRO DE INVESTIGACION

@xdopazo
@bioinfocipf

The Computational Genomics Department at the Centro de Investigación Príncipe Felipe (CIPF), Valencia, Spain, and...



...the INB, National Institute of
Bioinformatics (Functional
Genomics Node)
and the BiER (CIBERER Network of
Centers for Rare Diseases), and...



... Ignacio Medina
Head of Computational Biology
Lab, HPC Services
University of Cambridge, UK

EMBL-EBI Scientific Collaborator

 Fundació
La Marató de TV3

 MINISTERIO
DE CIENCIA E
INNOVACIÓN

 F.E.D.E.R.

 Proyectos
PROMETEO
Generalitat Valenciana

 Machine
Learning
for
Personalized
Medicine

 INB

 ciberer BIER
PLATAFORMA DE BIOINFORMATICA PARA LA ENFERMEDADES RARAS

 er

 ciberer
CENTRO DE INVESTIGACIÓN BIOMÉDICA EN RED
DE ENFERMEDADES RARAS

 BULL

 HPC4G

 iSCB
INTERNATIONAL
SOCIETY FOR
COMPUTATIONAL
BIOLOGY
AFFILIATED REGIONAL INSTITUTE