

Exome-Seq Data Analysis Workshop

El Escorial, 10th March 2015



PRINCIPE FELIPE
CENTRO DE INVESTIGACION

Computational • Genomics

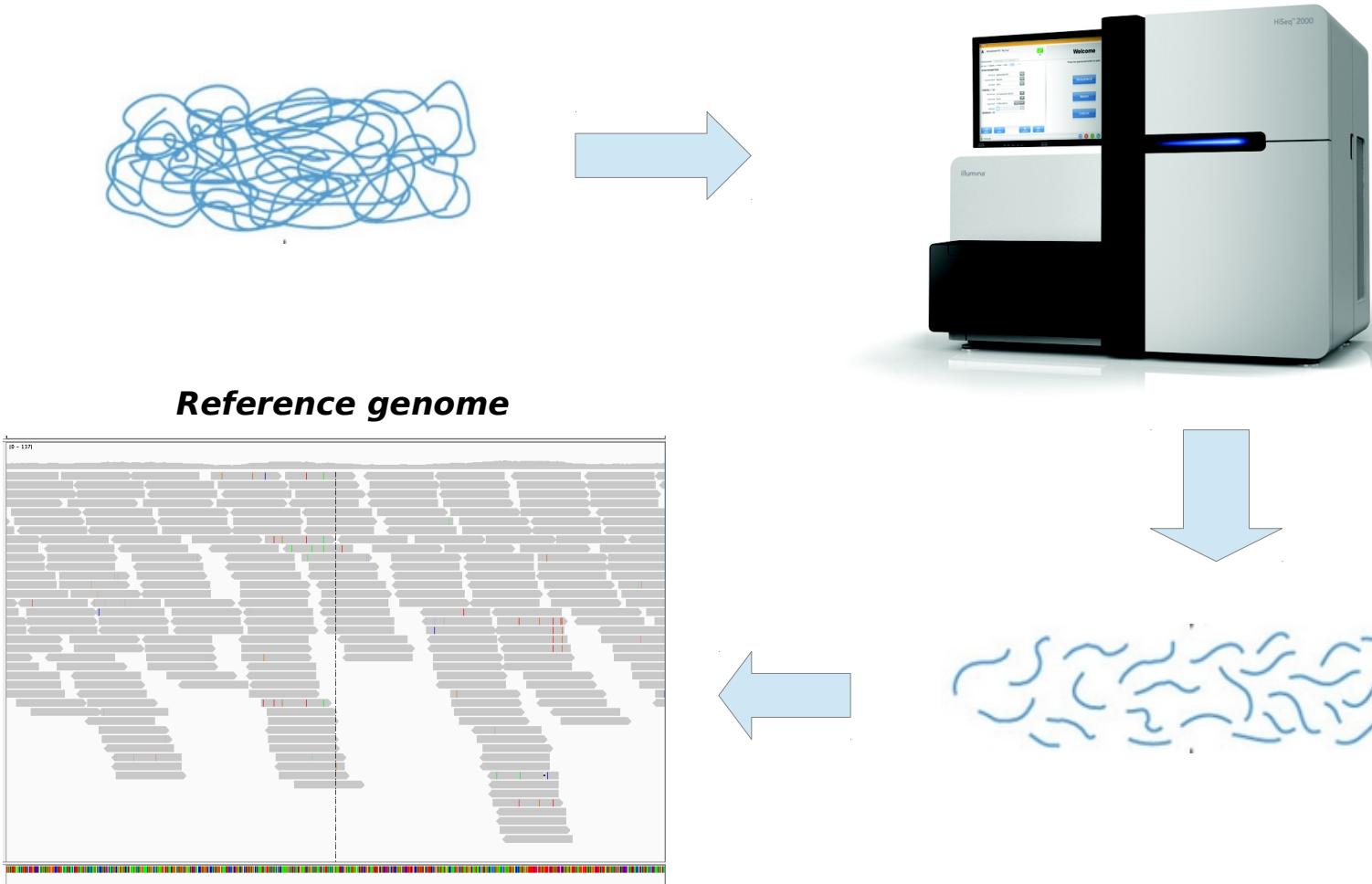


Outline

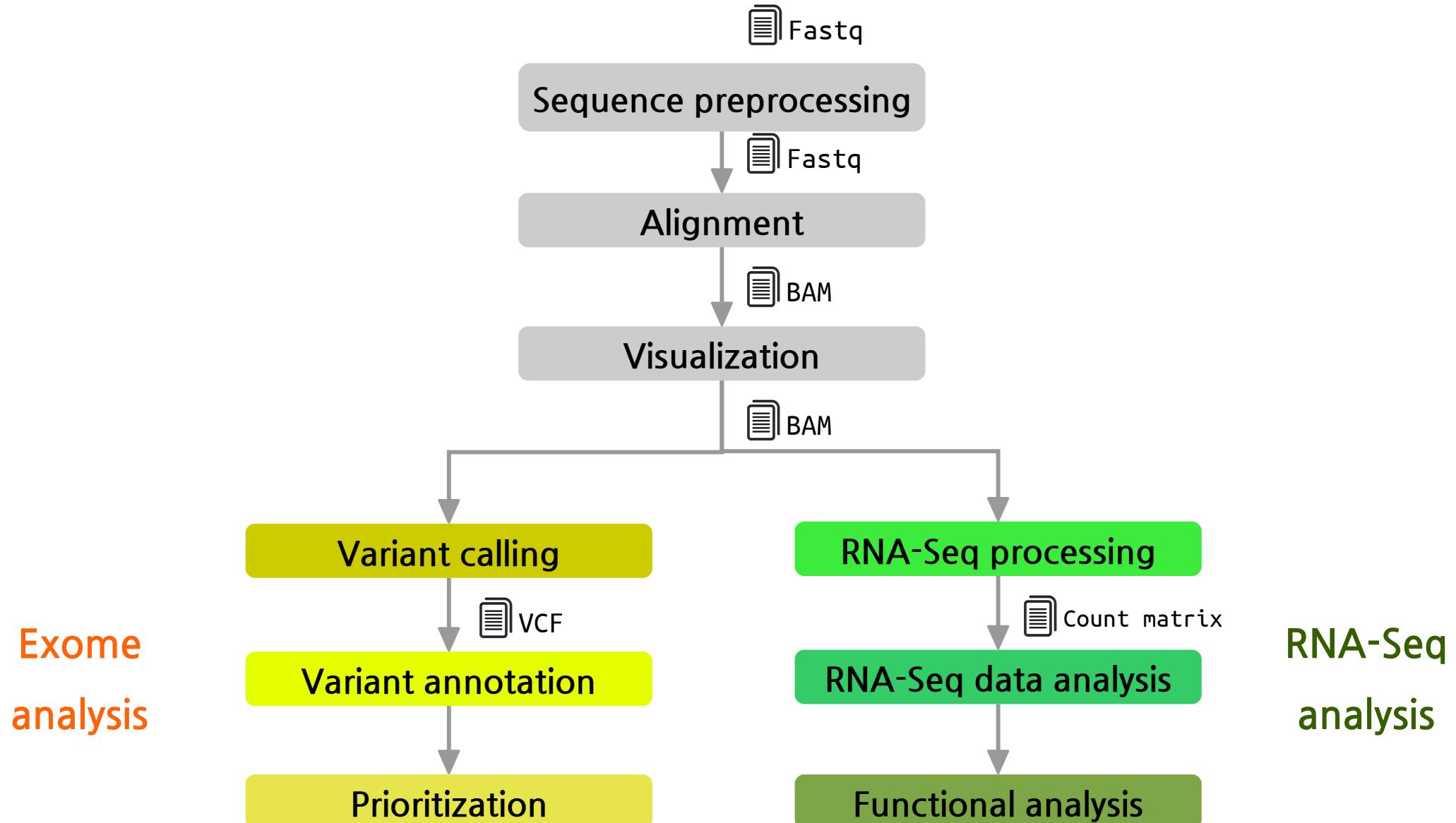
- 1) Introduction to NGS data analysis**
- 2) Web tools to analyze Genomics Data**
- 3) Let's practise!**

NGS technologies

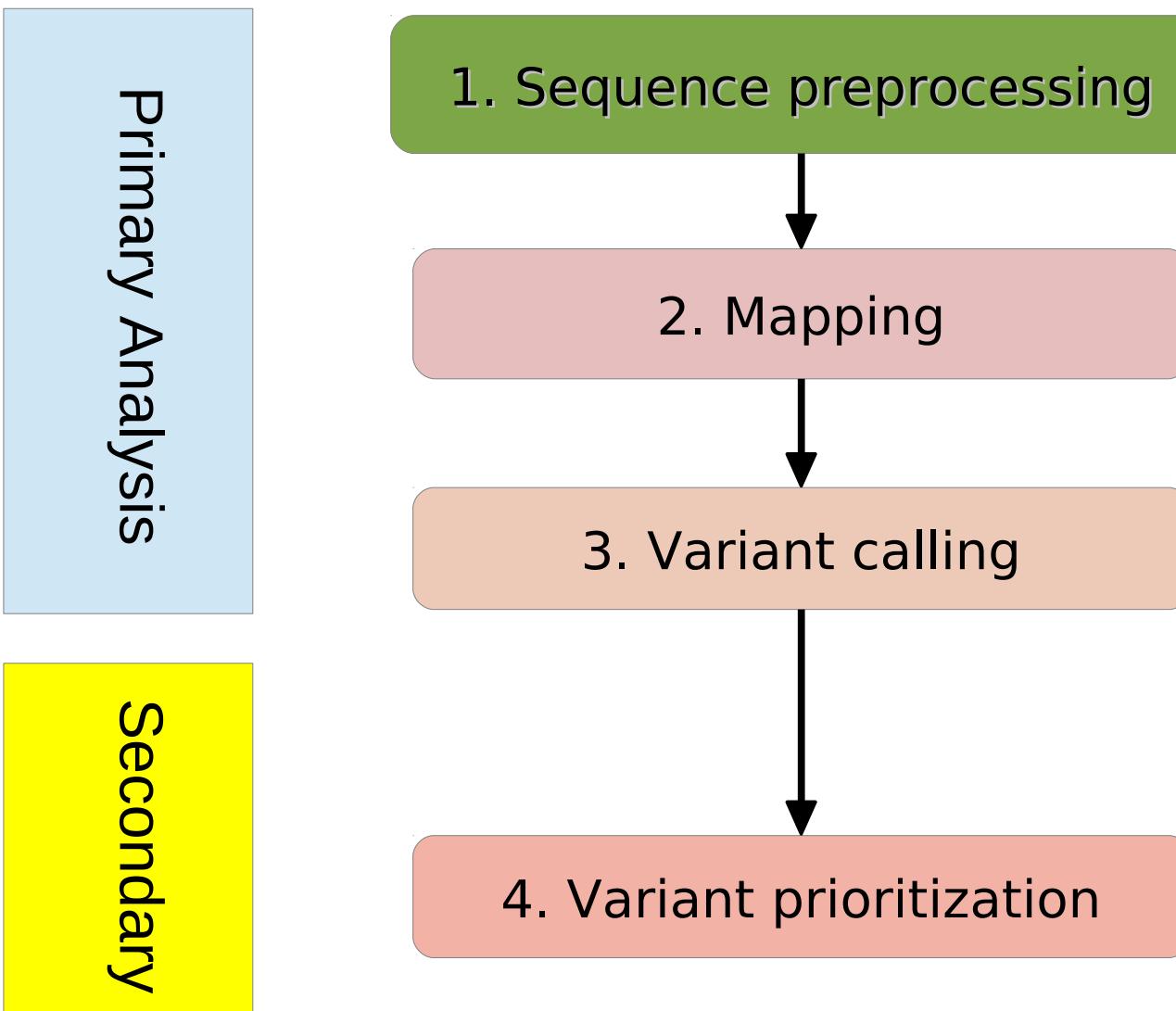
How do these technologies work ?



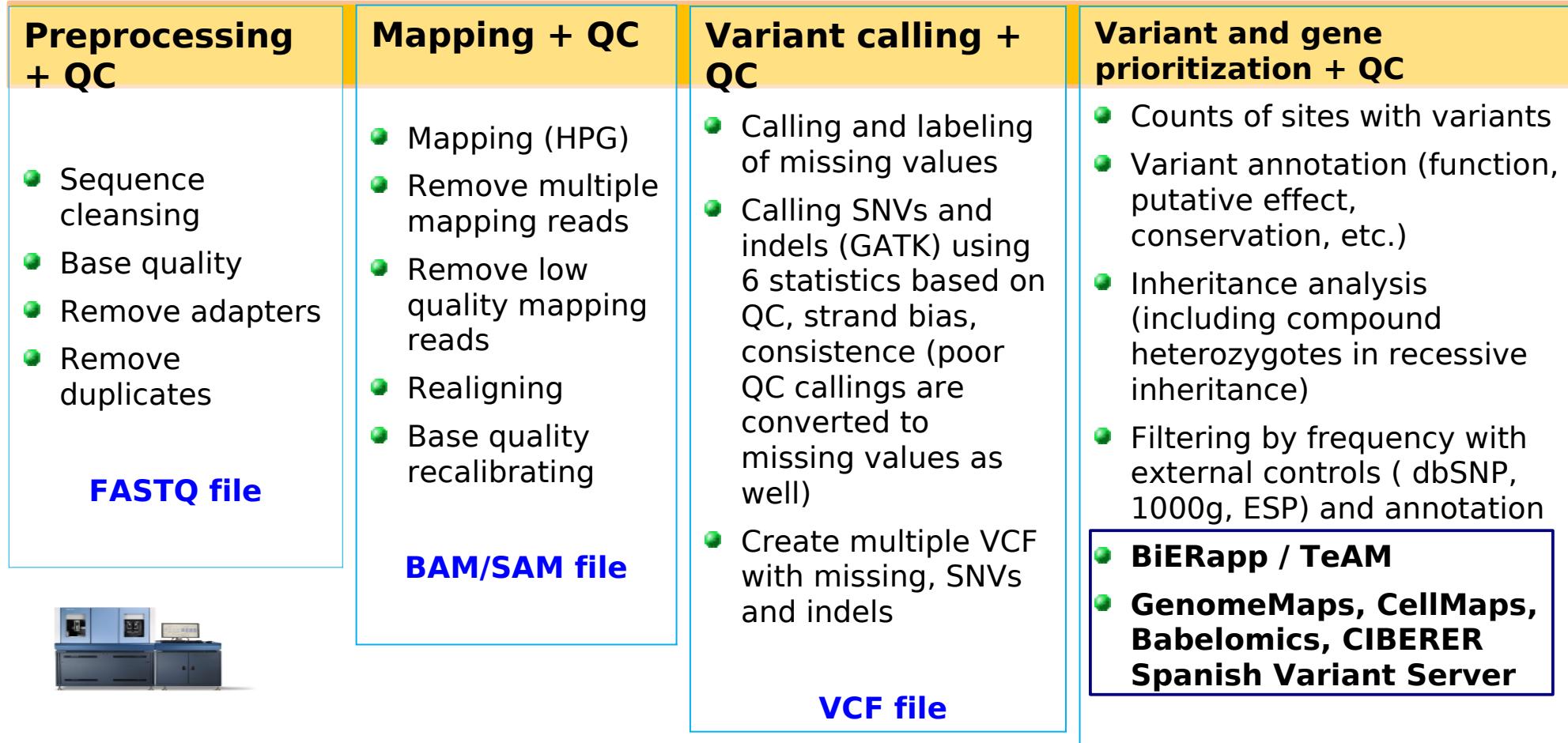
NGS Data Analysis Pipeline



Genomics Data Analysis Pipeline (1)



Genomics Data Analysis Pipeline (2)



Primary analysis

Gene prioritization

Introduction to NGS data analysis

Fastq format

- We could say “it is a fasta with **qualities**”:
 - 1. Header (like the fasta but starting with “@”)
 - 2. Sequence (string of nt)
 - 3. “+” and sequence ID (optional)
 - 4. Encoded quality of the sequence

```
@SEQ_ID
GATTTGGGGTTCAAAGCAGTATCGATCAAATAGTAAATCCATTGTTCAACTCACAGTTT
+
! ' ' * ( ( ( ( * * * + ) ) % % % + + ) ( % % % ) . 1 * * * - + * ' ' ) ) * * 55CCF>>>>CCCCCCCC65
```

BAM/SAM format

```
@PG ID:HPG-Aligner VN:1.0  
@SQ SN:20 LN:63025520
```

```
HWI-ST700660_138:2:2105:7292:79900#2@0/1 16 20 76703 254 76= * 0 0  
GTTTAGATACTGAAAGGTACATACTTCTTGAGGAACAAGCTATCATGCTGCATTCTATAATATCACATGAATA  
GIJGJLGGFLILGGIEIFEKEDELIGLJHJFIKKFELFIKLFFGLGHKKGJLFIIGKFFEFGKCKFHHCCCF AS:i:254 NH:i:1 NM:i:0  
  
HWI-ST700660_138:2:2208:6911:12246#2@0/1 16 20 76703 254 76= * 0 0  
GTTTAGATACTGAAAGGTACATACTTCTTGAGGAACAAGCTATCATGCTGCATTCTATAATATCACATGAATA  
HHJFHLLGFFILEGIKIEEMGEDLIGLHIIHJFIKKFELFIKLEFGKGHEKHJLFHIGKFFDFEEFGKDKFHHCCCF AS:i:254 NH:i:1 NM:i:0  
  
HWI-ST700660_138:2:1201:2973:62218#2@0/1 0 20 76655 254 76M * 0 0  
AACCCCCAAAATGTTGGAAGAATAATGTAGGACATTGCAGAAGACGATGTTAGATACTGAAAGGGACATACTTCT  
FEFFGHHGGHKCCJKFHIGIFFLDEJKGJGGFKIHLFIJGIEGFLDEDLFGEIIMHHIKL$BBGFFJIEHE AS:i:254 NH:i:1 NM:i:1  
  
HWI-ST700660_138:2:1203:21395:164917#2@0/1 256 20 68253 254 4M1D72M * 0 0  
NCACCCATGATAGACCAGTAAAGGTGACCACTTAAATTCTTGCTGTGCAGTGTCTGTATTCTCAGGACACAGA  
#4@ADEHFJFFJDHGKEFIHGBGFHHFIICEIFFKKIFHEGJEHHGLELEGKJMFGGGLEIKHLFGKIKHDG AS:i:254 NH:i:3 NM:i:1  
  
HWI-ST700660_138:2:1105:16101:50526#6@0/1 16 20 126103 246 53M4D23M * 0 0  
AAGAAGTGCAAACCTGAAGAGATGCATGTAAGAATGGTTGGGCAATGTGCGGCAAAGGGACTGCTGTGTTCCAGC  
FEHIGGHIGIGJI6FCFHJIFFLJJCJGJHGFKKKKGIJKHFFKIFFFKHFLKHGKJLJGKILLEFFLIHJIEIIB AS:i:368 NH:i:1 NM:i:4
```

SAM Specification:

<http://samtools.sourceforge.net/SAM1.pdf>

Introduction to NGS data analysis

VCF format

```

##fileformat=VCFv4.1
##fileDate=20090805
##source=myImputationProgramV3.1
##reference=file:///seq/references/1000GenomesPilot-NCBI36.fasta
##contig=<ID=20,length=62435964,assembly=B36,md5=f126cdf8a6e0c7f379d618ff66beb2da,species="Homo sapiens",taxonomy=x>
##phasing=partial
##INFO=<ID=NS,Number=1,Type=Integer,Description="Number of Samples With Data">
##INFO=<ID=DP,Number=1,Type=Integer,Description="Total Depth">
##INFO=<ID=AF,Number=A,Type=Float,Description="Allele Frequency">
##INFO=<ID=AA,Number=1,Type=String,Description="Ancestral Allele">
##INFO=<ID=DB,Number=0,Type=Flag,Description="dbSNP membership, build 129">
##INFO=<ID=H2,Number=0,Type=Flag,Description="HapMap2 membership">
##FILTER=<ID=q10,Description="Quality below 10">
##FILTER=<ID=s50,Description="Less than 50% of samples have data">
##FORMAT=<ID=GT,Number=1,Type=String,Description="Genotype">
##FORMAT=<ID=GQ,Number=1,Type=Integer,Description="Genotype Quality">
##FORMAT=<ID=DP,Number=1,Type=Integer,Description="Read Depth">
##FORMAT=<ID=HQ,Number=2,Type=Integer,Description="Haplotype Quality">
#CHROM POS ID REF ALT QUAL FILTER INFO FORMAT NA00001 NA00002 NA00003
20 14370 rs6054257 G A 29 PASS NS=3;DP=14;AF=0.5;DB;H2 GT:GQ:DP:HQ 0|0:48:1:51,51 1|0:48:8:51,51 1/1:43:5:...
20 17330 . T A 3 q10 NS=3;DP=11;AF=0.017 GT:GQ:DP:HQ 0|0:49:3:58,50 0|1:3:5:65,3 0/0:41:3
20 1110696 rs6040355 A G,T 67 PASS NS=2;DP=10;AF=0.333,0.667;AA=T;DB GT:GQ:DP:HQ 1|2:21:6:23,27 2|1:2:0:18,2 2/2:35:4
20 1230237 . T . 47 PASS NS=3;DP=13;AA=T GT:GQ:DP:HQ 0|0:54:7:56,60 0|0:48:4:51,51 0/0:61:2
20 1234567 microsat1 GTC G GTCT 50 PASS NS=3;DP=9;AA=G GT:GQ:DP:HQ 0|1:35:4 0|2:17:2 1/1:40:3

```

<http://www.1000genomes.org/>

Outline

- 1) Introduction to NGS data analysis
- 2) Web tools to analyze Genomics Data
- 3) Let's practise!

BiERapp:

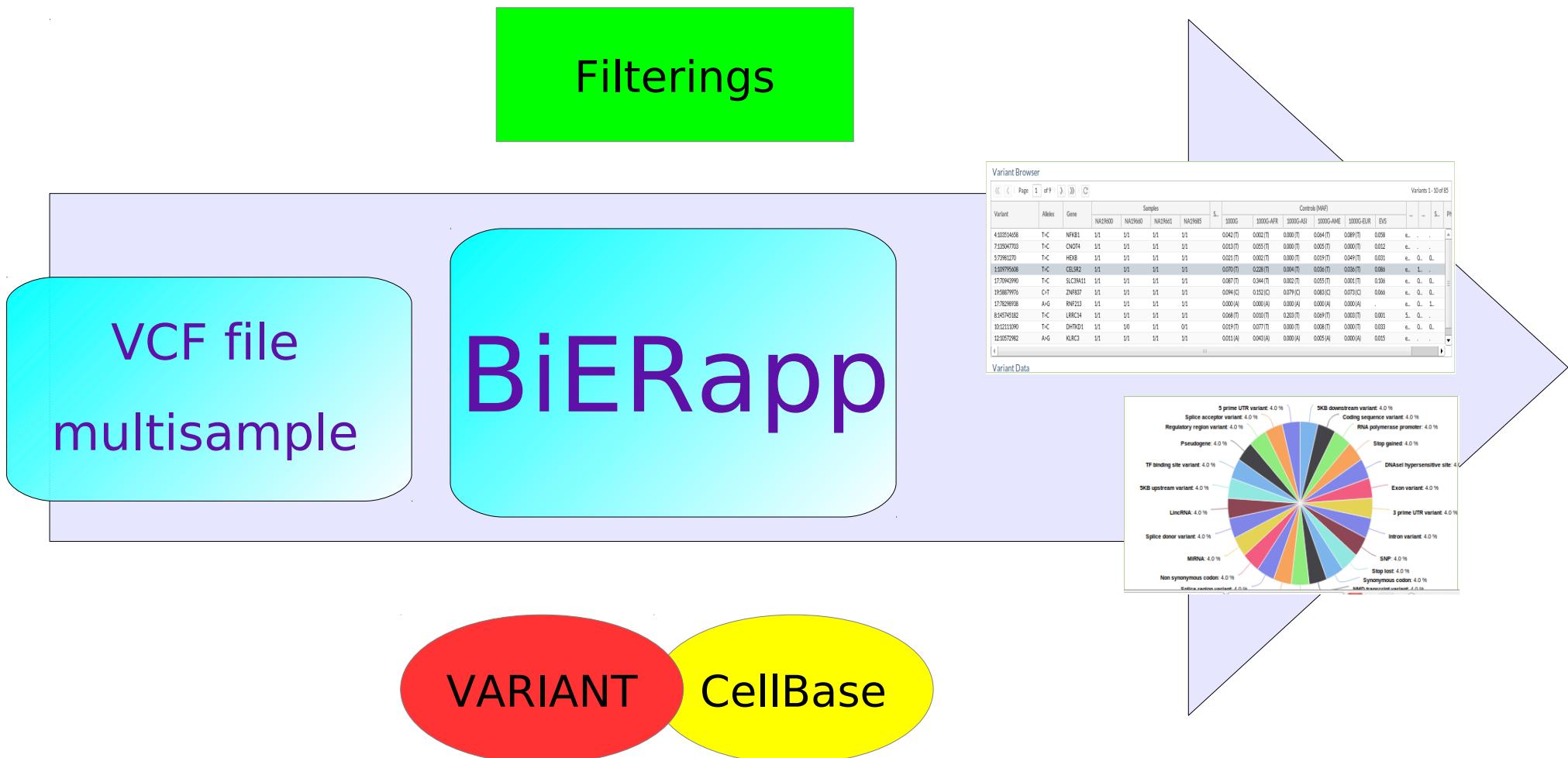
Una herramienta web para la priorización de variantes

<http://ciberer.es/bier/bierapp>

Introduction

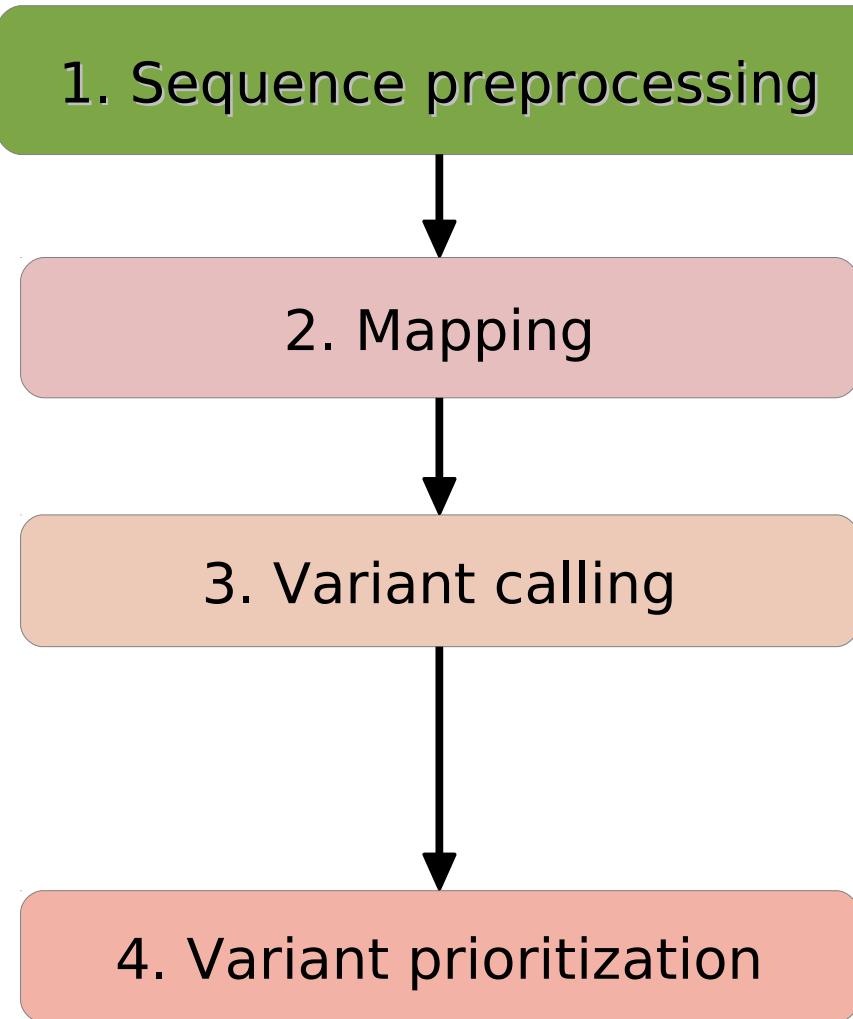
- Whole-exome sequencing has become a fundamental tool for the discovery of disease-related genes of familial diseases but there are difficulties to **find the causal mutation among the enormous background**
- There are different scenarios, so we need **different and immediate strategies of prioritization**
- Vast amount of **biological knowledge available** in many databases
- We need a tool to **integrate this information and filter immediately** to select candidate variants related to the disease

How does BiERapp work?



Input: VCF file

Primary Analysis



VCF files

Secondary



Input: VCF multisample

```
##fileformat=VCFv4.1
##fileDate=20090805
##source=myImputationProgramV3.1
##reference=file:///seq/references/1000GenomesPilot-NCBI36.fasta
##contig=<ID=20,length=62435964,assembly=B36,md5=f126cdf8a6e0c7f379d618ff66beb2da,species="Homo sapiens",taxonomy=x>
##phasing=partial
##INFO=<ID=NS,Number=1,Type=Integer,Description="Number of Samples With Data">
##INFO=<ID=DP,Number=1,Type=Integer,Description="Total Depth">
##INFO=<ID=AF,Number=A,Type=Float,Description="Allele Frequency">
##INFO=<ID=AA,Number=1,Type=String,Description="Ancestral Allele">
##INFO=<ID=DB,Number=0,Type=Flag,Description="dbSNP membership, build 129">
##INFO=<ID=H2,Number=0,Type=Flag,Description="HapMap2 membership">
##FILTER=<ID=q10,Description="Quality below 10">
##FILTER=<ID=s50,Description="Less than 50% of samples have data">
##FORMAT=<ID=GT,Number=1,Type=String,Description="Genotype">
##FORMAT=<ID=GQ,Number=1,Type=Integer,Description="Genotype Quality">
##FORMAT=<ID=DP,Number=1,Type=Integer,Description="Read Depth">
##FORMAT=<ID=HQ,Number=2,Type=Integer,Description="Haplotype Quality">
#CHROM POS ID REF ALT QUAL FILTER INFO FORMAT NA00001 NA00002 NA00003
20 14370 rs6054257 G A 29 PASS NS=3;DP=14;AF=0.5;DB;H2 GT:GQ:DP:HQ 0|0:48:1:51,51 1|0:48:8:51,51 1/1:43:5:..
20 17330 . T A 3 q10 NS=3;DP=11;AF=0.017 GT:GQ:DP:HQ 0|0:49:3:58,50 0|1:3:5:65,3 0/0:41:3
20 1110696 rs6040355 A G,T 67 PASS NS=2;DP=10;AF=0.333,0.667;AA=T;DB GT:GQ:DP:HQ 1|2:21:6:23,27 2|1:2:0:18,2 2/2:35:4
20 1230237 . T . 47 PASS NS=3;DP=13;AA=T GT:GQ:DP:HQ 0|0:54:7:56,60 0|0:48:4:51,51 0/0:61:2
20 1234567 microsat1 GTC G,GTCT 50 PASS NS=3;DP=9;AA=G GT:GQ:DP 0/1:35:4 0/2:17:2 1/1:40:3
```

**One VCF (Variant Calling Format) file
for family or group**

Getting information

□ SIFT

- SIFT predicts whether an amino acid substitution affects protein function
- **Interpretation:** 1 (tolerated) to 0 (not tolerated)

<http://sift.jcvi.org/>

J. Craig VenterTM
INSTITUTE

SIFT

□ PolyPhen

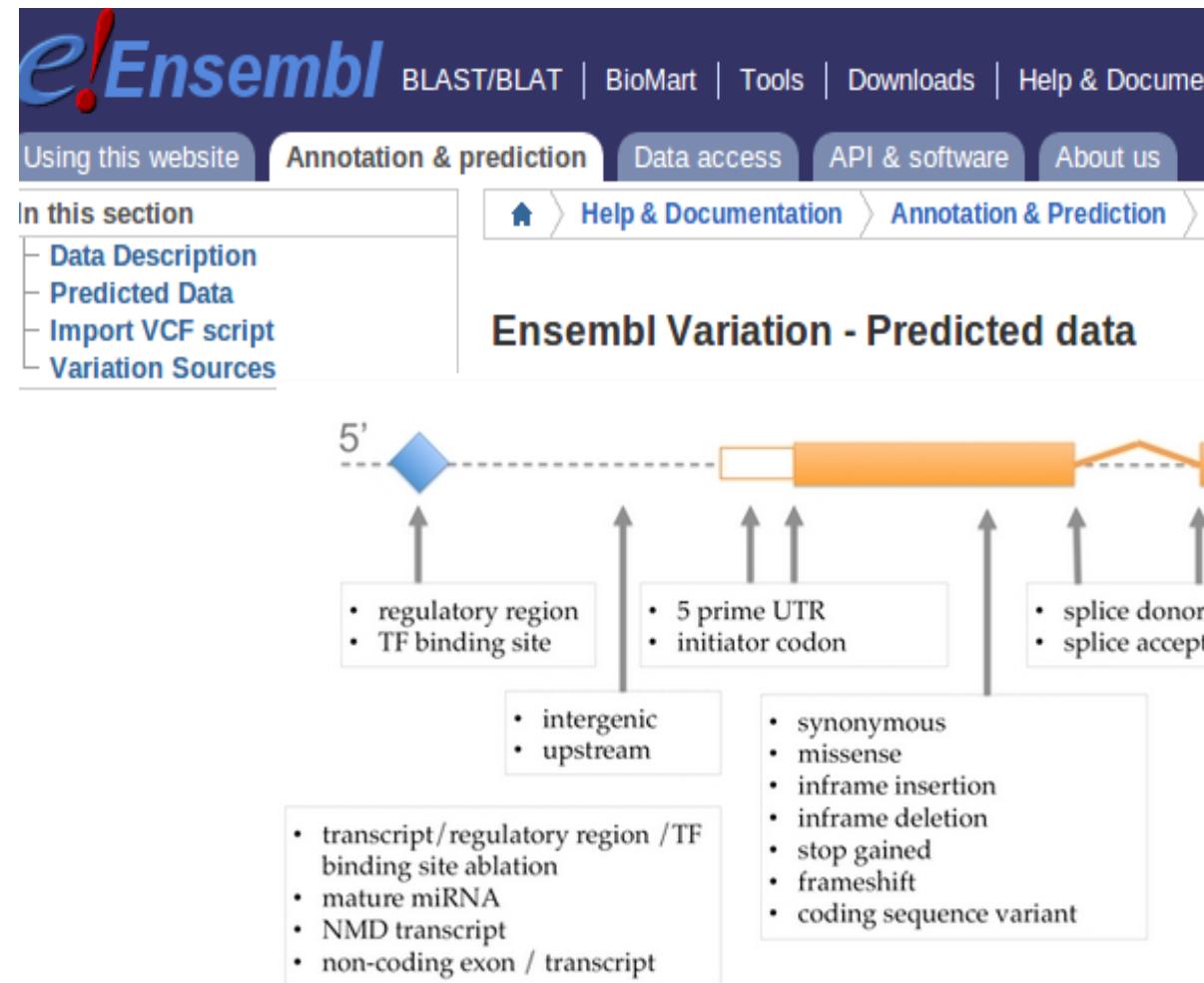
- Polymorphism Phenotyping is a tool which predicts possible impact of an amino acid substitution on the structure and function of a human protein
- **Interpretation:** 1 (probably damage) to 0 (benign)

<http://genetics.bwh.harvard.edu/pph2/index.shtml>



PolyPhen-

Getting information



Consequence type or effect

http://www.ensembl.org/info/genome/variation/predicted_data.html

Tool interface

<http://ciberer.es/bier/bierapp>

Menu BierApp ciber^{BIER} Home

Overview

Welcome to the gene/variant prioritization tool of the BIER (the Team of Bioinformatic for Rare Diseases). This interactive tool allows finding genes affected by deleterious variants that segregate along family pedigrees , case-controls or sporadic samples.

Try an Example

Here you can try all the filtering options and discover the gene affected in a test family.

Analyze your own families or case-control data

Here you can upload your VCF file containing the exomes to be analyzed. Define the thresholds of allele frequencies, pathogenicity, conservation; the type of variants sought; and define the type of inheritance and the segregation schema along the family.

Supported by

logout upload & manage profile jobs support

Tool interface

BierApp  Home

Example 1000G(Short)

Variant Browser

Variant Alleles Gene Samples ... Controls (MAF) Variants 1-10 of 85

Variant	Alleles	Gene	NA19600	NA19660	NA19661	NA19685	...	1000G	1000G-AFR	1000G-ASJ	1000G-AME	1000G-EUR	EVS	P
4:103514658	T>C	NFKB1	1/1	1/1	1/1	1/1	...	0.042(T)	0.002(T)	0.000(T)	0.064(T)	0.089(T)	0.058	e...	.	.	
7:135047703	T>C	CNOT4	1/1	1/1	1/1	1/1	...	0.013(T)	0.055(T)	0.000(T)	0.005(T)	0.000(T)	0.012	e...	.	.	
5:73981270	T>C	HEXB	1/1	1/1	1/1	1/1	...	0.021(T)	0.002(T)	0.000(T)	0.019(T)	0.049(T)	0.031	e...	0...	0...	
1:109795608	T>C	CELSR2	1/1	1/1	1/1	1/1	...	0.070(T)	0.228(T)	0.004(T)	0.036(T)	0.036(T)	0.086	e...	1...	.	
17:70943990	T>C	SLC39A11	1/1	1/1	1/1	1/1	...	0.087(T)	0.344(T)	0.002(T)	0.055(T)	0.001(T)	0.106	e...	0...	0...	
19:58879976	C>T	ZNF837	1/1	1/1	1/1	1/1	...	0.094(C)	0.152(C)	0.079(C)	0.083(C)	0.073(C)	0.066	e...	0...	0...	
17:78298938	A>G	RNF213	1/1	1/1	1/1	1/1	...	0.000(A)	0.000(A)	0.000(A)	0.000(A)	0.000(A)	.	e...	0...	1...	
8:145745182	T>C	LRRC14	1/1	1/1	1/1	1/1	...	0.068(T)	0.010(T)	0.203(T)	0.069(T)	0.003(T)	0.001	5...	0...	.	
10:12111090	T>C	DHTKD1	1/1	1/0	1/1	0/1	...	0.019(T)	0.077(T)	0.000(T)	0.008(T)	0.000(T)	0.033	e...	0...	0...	

Variant Data

Genomic Context Effect & Annotation Study Summary

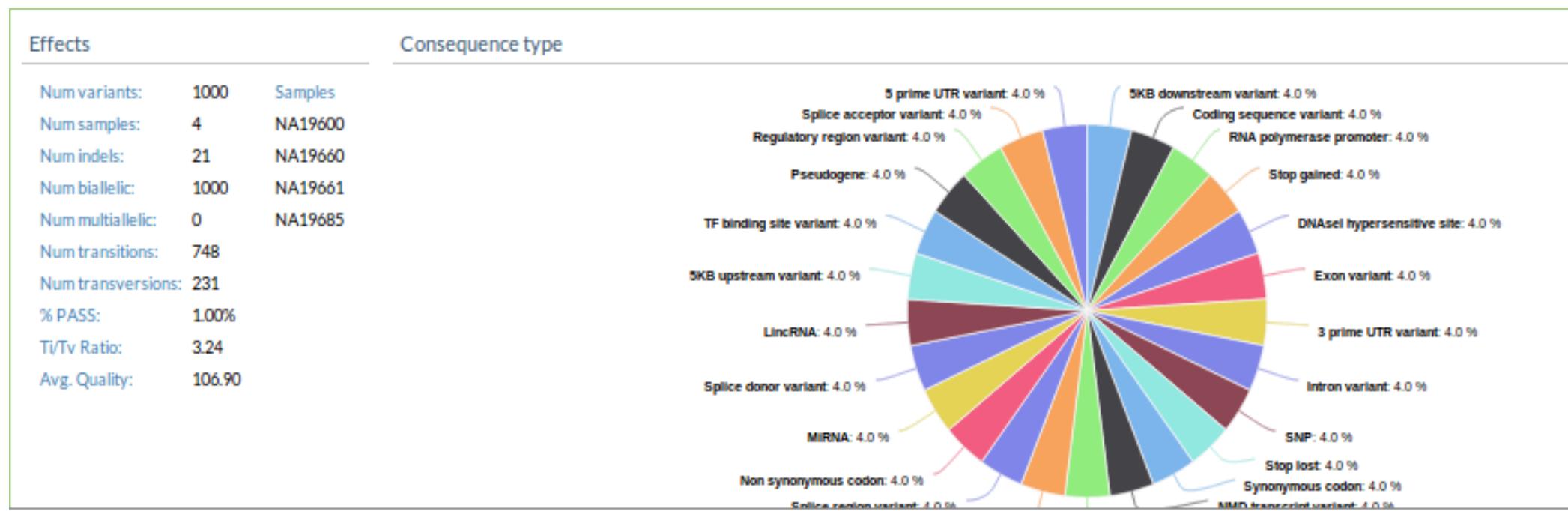
Effects Consequence type

Effects	Consequence type
Num variants: 1000	Samples
Num samples: 4	NA19600
Num indels: 21	NA19660
Num biallelic: 1000	NA19661
Num multiallelic: 0	NA19685
Num transitions: 748	
Num transversions: 231	
% PASS: 100%	
Ti/Tv Ratio: 3.24	
Avg. Quality: 106.90	

BiERapp: discovering variants

Results

1. Summary. Description about number of variants, INDELs... Also a distribution of consequences types.



Results

2. List of candidate variants.

We can order this list by several criteria.

Variant Browser

Page 1 of 9 | C Variants 1 - 10 of 85

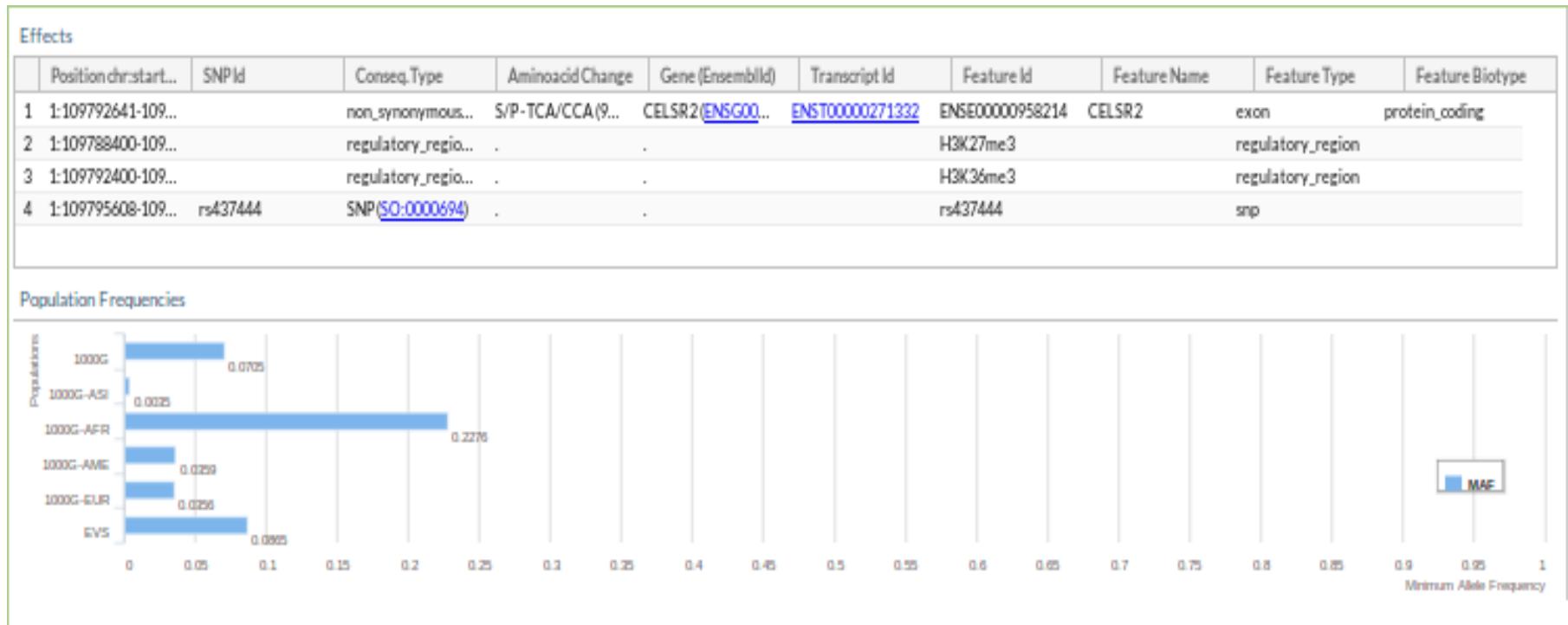
Variant	Alleles	Gene	Samples				S..	Controls (MAF)						... S..	Ph
			NA19600	NA19660	NA19661	NA19685		1000G	1000G-AFR	1000G-ASI	1000G-AME	1000G-EUR	EVS		
4:103514658	T>C	NFKB1	1/1	1/1	1/1	1/1	0.042 (T)	0.002 (T)	0.000 (T)	0.064 (T)	0.089 (T)	0.058	e.. . .		
7:135047703	T>C	CNOT4	1/1	1/1	1/1	1/1	0.013 (T)	0.055 (T)	0.000 (T)	0.005 (T)	0.000 (T)	0.012	e.. . .		
5:73981270	T>C	HEXB	1/1	1/1	1/1	1/1	0.021 (T)	0.002 (T)	0.000 (T)	0.019 (T)	0.049 (T)	0.031	e.. 0.. 0..		
1:109795608	T>C	CELSR2	1/1	1/1	1/1	1/1	0.070 (T)	0.228 (T)	0.004 (T)	0.036 (T)	0.036 (T)	0.086	e.. 1.. .		
17:70943990	T>C	SLC39A11	1/1	1/1	1/1	1/1	0.087 (T)	0.344 (T)	0.002 (T)	0.055 (T)	0.001 (T)	0.106	e.. 0.. 0..		
19:58879976	C>T	ZNF837	1/1	1/1	1/1	1/1	0.094 (C)	0.152 (C)	0.079 (C)	0.083 (C)	0.073 (C)	0.066	e.. 0.. 0..		
17:78298938	A>G	RNF213	1/1	1/1	1/1	1/1	0.000 (A)	0.000 (A)	0.000 (A)	0.000 (A)	0.000 (A)	.	e.. 0.. 1..		
8:145745182	T>C	LRRC14	1/1	1/1	1/1	1/1	0.068 (T)	0.010 (T)	0.203 (T)	0.069 (T)	0.003 (T)	0.001	5.. 0.. .		
10:12111090	T>C	DHTKD1	1/1	1/0	1/1	0/1	0.019 (T)	0.077 (T)	0.000 (T)	0.008 (T)	0.000 (T)	0.033	e.. 0.. 0..		
12:10572982	A>G	KLRC3	1/1	1/1	1/1	1/1	0.011 (A)	0.043 (A)	0.000 (A)	0.005 (A)	0.000 (A)	0.015	e.. . .		

Variant Data

Results

3. Effects for each transcript where we detected a candidate variant.

The plot shows MAFs for different groups (1000 Genomes, Exome Variant Server)



Results

4. Visualization of candidate variants from GenomeMaps



Remarks

- The proposed web-based interactive framework has **great potential to detect disease-related variants** in familial diseases as demonstrated by its successful use in several studies
- **The use of the filters is interactive** and the results are almost instantaneously displayed in a panel that includes the genes affected, the variants and specific information for them
- Candidate variants are **new knowledge useful for future diagnostic**

TEAM:

Una **herramienta web** para el diseño y
gestión de **paneles de genes** en
secuenciación dirigida
con aplicaciones clínicas

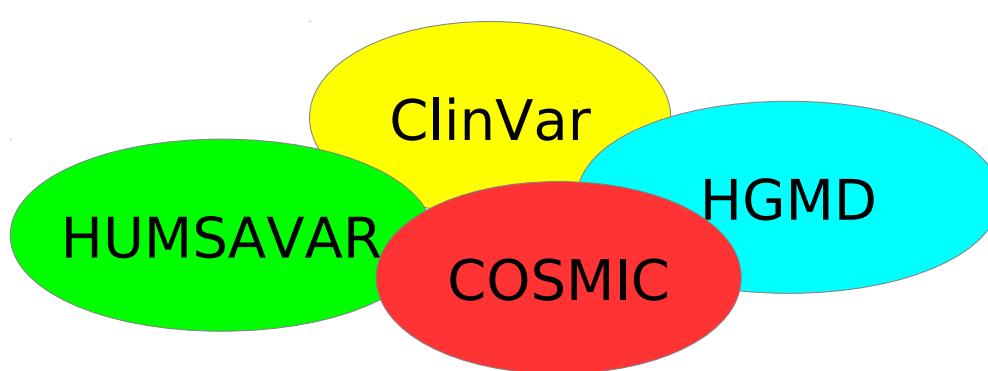
<http://ciberer.es/bier/team>

Introduction

- **Development of high throughput sequencing technologies:**
 - ✚ Rapid and economical genome sequencing.
 - ✚ Disease targeted sequencing: powerful and cost-effective application.
- **Vast amount of biological knowledge available:**
 - ✚ HGMD-public, HUMSAVAR, ClinVar, COSMIC.
- We need a tool to connect **sequencing data and biological knowledge for diagnostic:**
 - ✚ **TEAM** (Targeted Enrichment Analysis and Management).

TEAM: Targeted Enrichment Analysis and Management

How does TEAM work?



How does TEAM work?

<http://ciberer.es/bier/team>

1. Defining panel

The screenshot shows the 'Panels' section of the TEAM interface. At the top, there are buttons for 'New Panel', 'Import Panels', 'Save Panels', and 'Clear Panels'. Below these, there are two tabs: 'User-defined' (which is selected) and 'Examples'. A text input field labeled 'name' contains the value 'RETINITIS_panel10'. To the right of the input field are edit and delete icons.

2. Uploading input data

The screenshot shows the 'Example Data' section of the TEAM interface. It features a search bar at the top. Below it, there are fields for 'Panel:' (set to 'Panel Retinitis Pigmentosa') and 'VCF File:' (containing the path 'C:\fakepath\patient1_R.vcf'). There is also a 'Browse...' button. At the bottom are 'Run' and 'Reset' buttons.

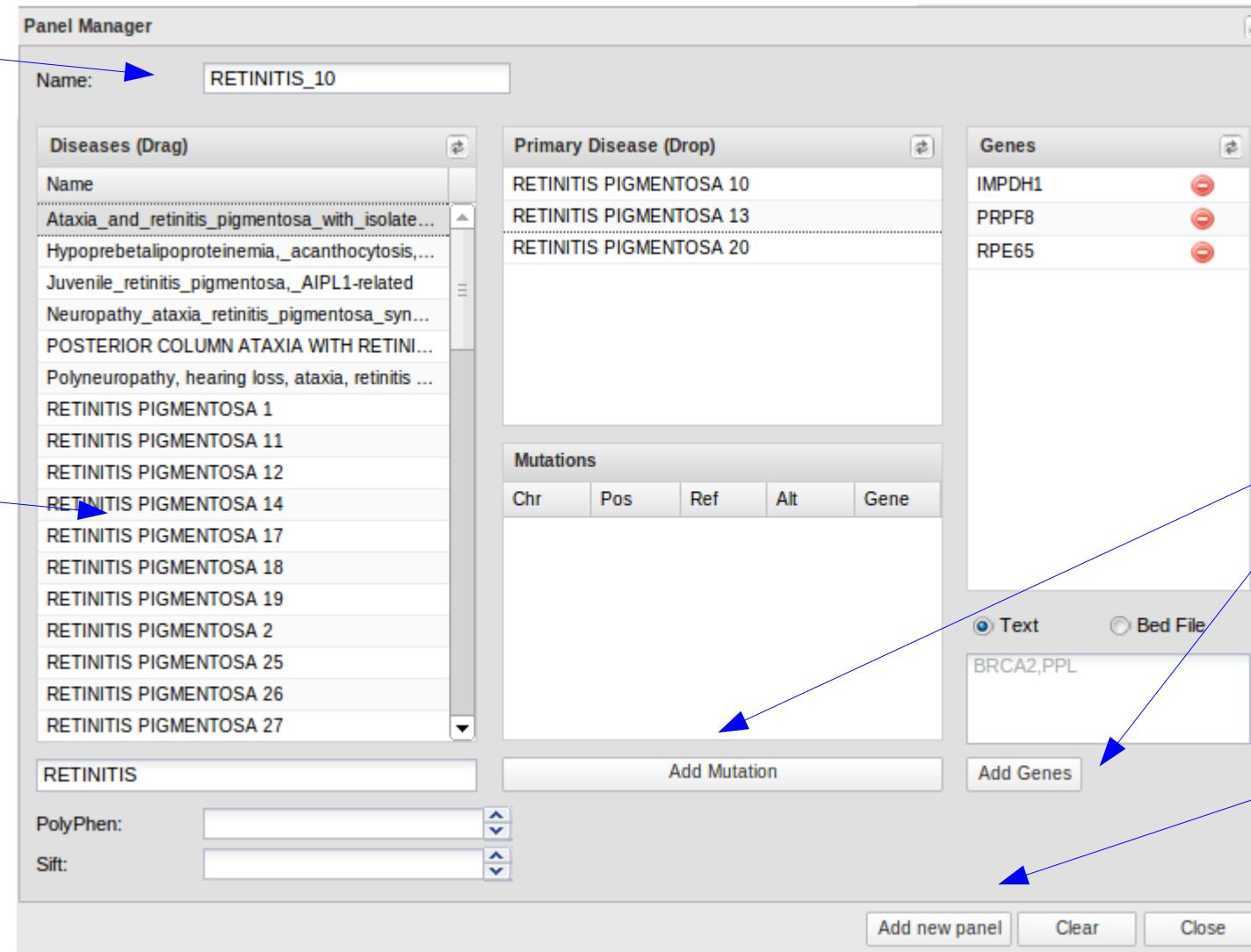
3. Getting results

The screenshot shows the 'Results' section of the TEAM interface. It has tabs for 'Diagnostic' and 'Secondary findings'. The 'Diagnostic' tab is selected. The table has columns: Chromosome, Position, SNP Id, Ref, Alt, Gene, Conseq., and Phenotype. There are four rows of data:

Chromosome	Position	SNP Id	Ref	Alt	Gene	Conseq.	Phenotype
gene: (1 Item)							
1	3	129247734	.	T	C	.	exon_... RETINITIS PIGMENTOSA 4
gene: RHO (3 Items)							
2	3	129247734	.	T	C	RHO	exon_... RETINITIS PIGMENTOSA 4
3	3	129247734	.	T	C	RHO	exon_... RETINITIS PIGMENTOSA 4
4	3	129247734	.	T	C	RHO	Retinitis pigmentosa type 4

How to define a panel?

1. Name
of panel

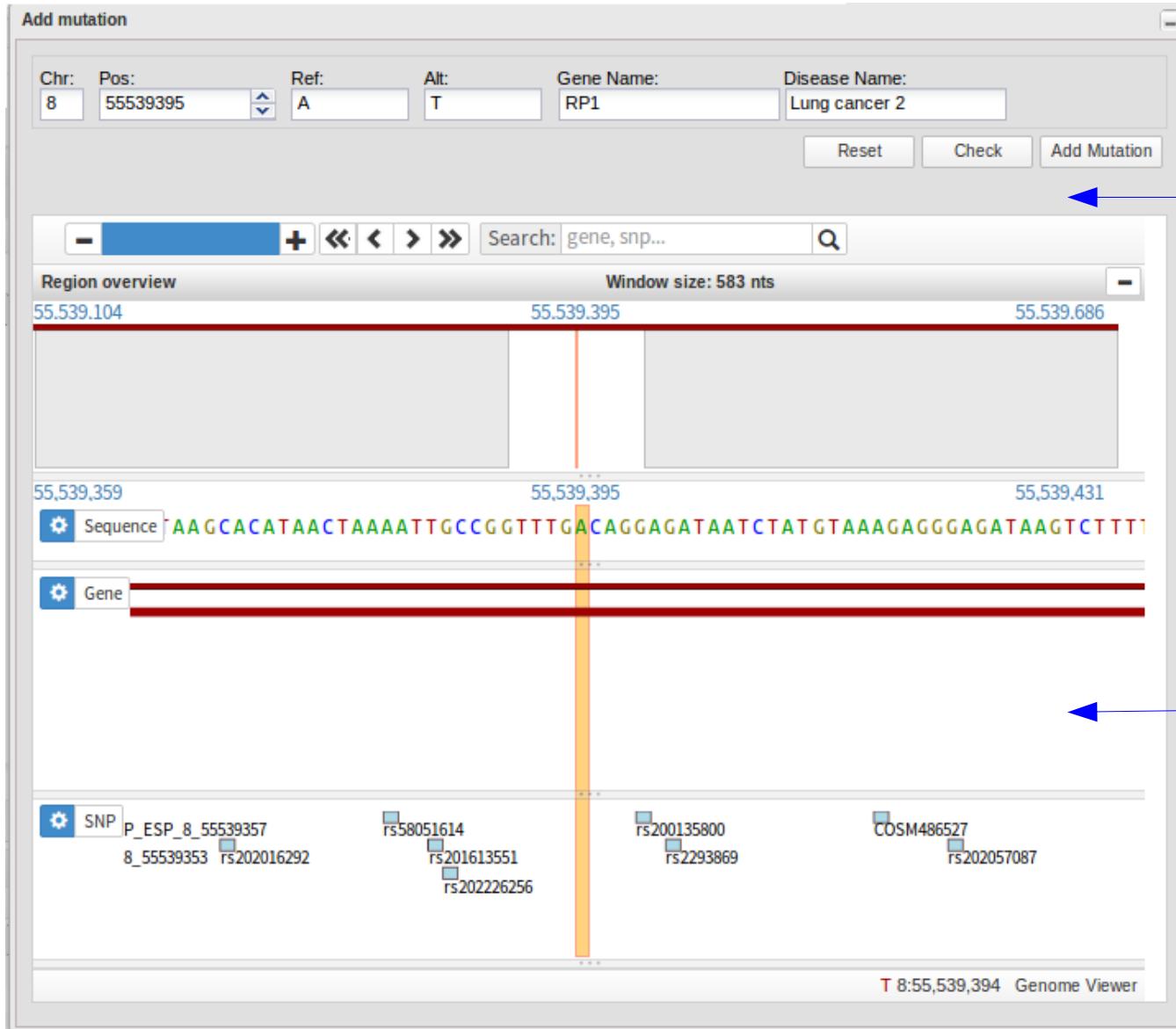


2. Diseases

3. Adding:
- more genes
- mutations

4. Save panel

How to define a panel?



Adding
new mutations

Checking
mutations from
Genome Viewer

Results

Results								
Diagnostic		Secondary findings						
	Chromosome	Position	SNP Id	Ref	Alt	Gene	Conseq. Type	Phenotype
gene: (1 Item)								
1	3	129247734	.	T	C	.	exon_vari...	RETINITIS PIGMENTOSA 4
gene: RHO (3 Items)								
2	Variant Effect - 3:129247734 T>C							
3	Position chr:start:end (strand)		SNP Id		Conseq. Type		Aminoacid Change	
4	1	3:129247734-129247734 (+)	CM920608	SNP (SO:0000694)		.		.
2	3:129247483-129247937 (+)	synonymous_codon (SO:00...		P/P - CCC/CCC (53)				
3	3:129245550-129248350	regulatory_region_variant (...)		.		.		
4	3:129247734-129247734 (+)	rs28933395	SNP (SO:0000694)		.		.	

A. Web results

B. PDF report

Diagnostic: Retinitis

	Chromosome	Position	SNP Id	Ref	Alt	Gene	Conseq. Type
1	3	129247734	.	T	C	.	exon_vari...
2	3	129247734	.	T	C	RHO	exon_vari...
3	3	129247734	.	T	C	RHO	exon_vari...
4	3	129247734	.	T	C	RHO	exon_vari...

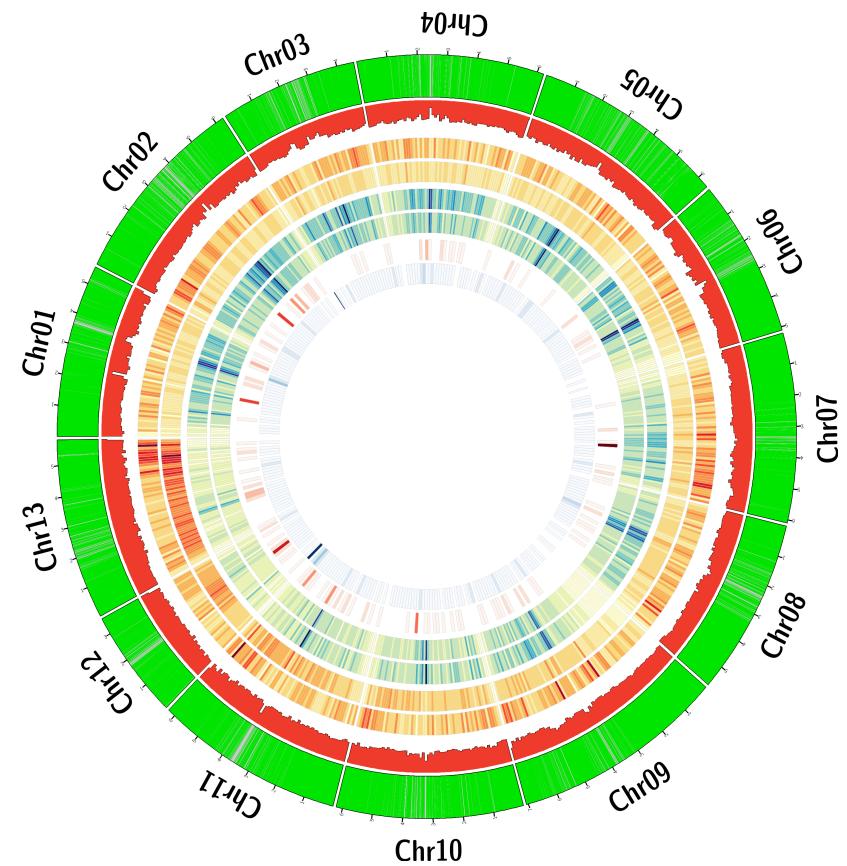
TEAM: Targeted Enrichment Analysis and Management

Remarks

- TEAM is an **free and easy-to-use web tool** that fills the gap between the enormous amounts of data in targeted enrichment sequencing analysis and the **biological knowledge** available.
- TEAM **provides an intuitive environment for the clinician** in which unprocessed data on patient's genomic variation can easily be transformed in a **diagnostic**.
- The entire patient's sequencing information is managed locally thus avoiding any problem of data **privacy or confidentiality**.

Next improvements:

- Inclusion of a **database with public panels genes** of various diseases.
- **Comparative Analysis** for groups of panels.
- **Visualization results.**



CSVS: CIBERER Spanish Variant Server

Repositorio de frecuencias de variantes
en la población española

<http://bioinfo.cipf.es/apps-beta/spvs/1.0.0/>

Two initial repositories

- 1) <http://www.ciberer.es/bier/exome-server/>
- 2) <http://bioinfo.cipf.es/apps-beta/spv/1.0.1/>

Spanish Population Variability

Filters

Variant Info

Variant	Alleles	SNP Id	Gene	SPV				MAF	
				Genotypes					
				0/0	0/1	1/1	.		
2:14004	G>A			266	1	.	.	0.002	
2:14190	C>T			266	1	.	.	0.002	
2:14238	G>A			266	1	.	.	0.002	
2:14296	G>A			266	1	.	.	0.002	
2:14309	G>A			266	1	.	.	0.002	
2:14485	T>C			240	27	.	.	0.051	
2:14489	A>G			265	2	.	.	0.004	
2:41366	C>T	FAM110C		259	6	2	.	0.019	

Tool interface

Spanish Population Variant Server **beta** Search Studies Stats [?](#)

Position

Chromosomal Location: **1:1-100000**

Gene: **BRCA2, PPL**

Studies

Mgp
 Virginia Nunes
 Miguel Angel Moreno
 Aurora Pujol
 Francesc Palau

Diseases

Healthy Population

1000G

Chr	Position	Alleles	Id	MAF	1000G					EVS										
					Genotypes	Freq.	Genotypes	Freq.	Genotypes	Freq.	MAF									
0/0	0/1	1/1	0 freq	1 freq	0/0	0/1	1/1	0 freq	1 freq	MAF										
1	17483	C>T		403	1	.	0.917	0.083	0.083	
1	18422	T>C		397	6	1	0.733	0.267	0.267
1	18256	T>G		403	1	.	0.633	0.033	0.033
1	18256	T>C		394	10	.	0.633	0.333	0.333
1	18094	C>T		401	3	.	0.900	0.100	0.100
1	17398	C>A		399	5	.	0.833	0.167	0.167
1	16974	C>T		394	10	.	0.667	0.333	0.333
1	16809	C>G		393	9	2	0.567	0.433	0.433
1	16794	G>A		403	1	.	0.967	0.033	0.033
1	16619	C>T		402	.	2	0.867	0.133	0.133

Genomic Context **Effect** **Frequencies** **Phenotype**

Gene Name Ensembl Gene Id Ensembl Transcript Id Conseq. type Relative Position Codon Strand

« < Page 0 of 1 > »

Variants per Study

Ester Lopez: 83899
M. Jesus Melia: 87022
Jordi Surralles: 302244
Placido Navas: 76094
Carmen Espinosa: 62509
Daniel Grinberg: 280912
Rafael Artuch: 446798
Aitor Delmiro: 377232
Jose Maria Millan: 161126
Carmen Ayuso: 232782
Mgp: 710993
Virginia Nunes: 123820
Miguel Angel Moreno: 125208
Aurora Pujol: 127164
Francesc Palau: 150932
Roser Gonzalez: 117166
Magdalena Ugarté: 225402
Antonia Ribes: 242072

Variants

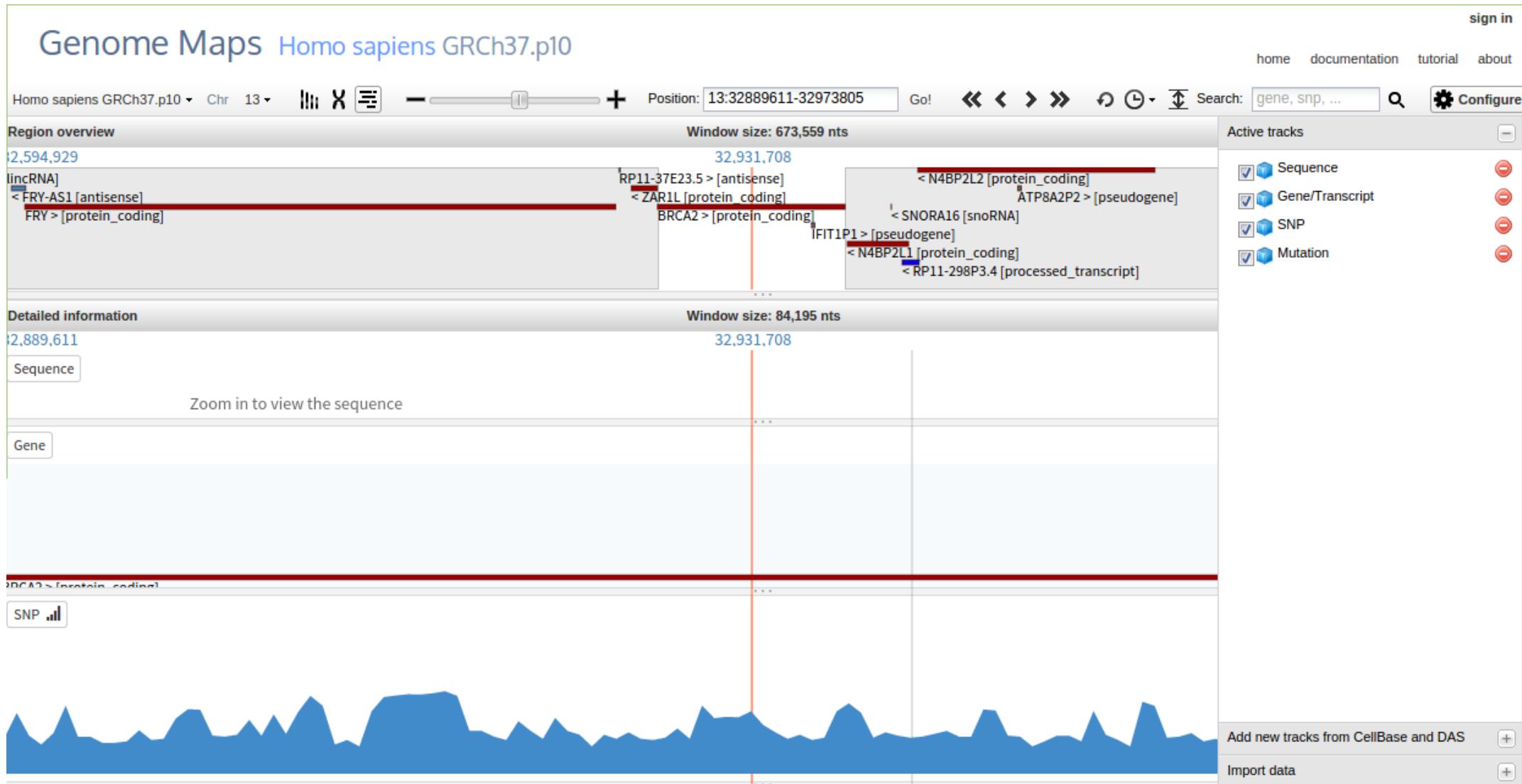
<http://bioinfo.cipf.es/apps-beta/spvs/1.0.0/>

Genome Maps

Visualizador genómico que interactúa
con bases de datos funcionales

<http://genomemaps.org/>

Tool interface



Cell Maps

Herramienta de modelización y
visualización de redes biológicas

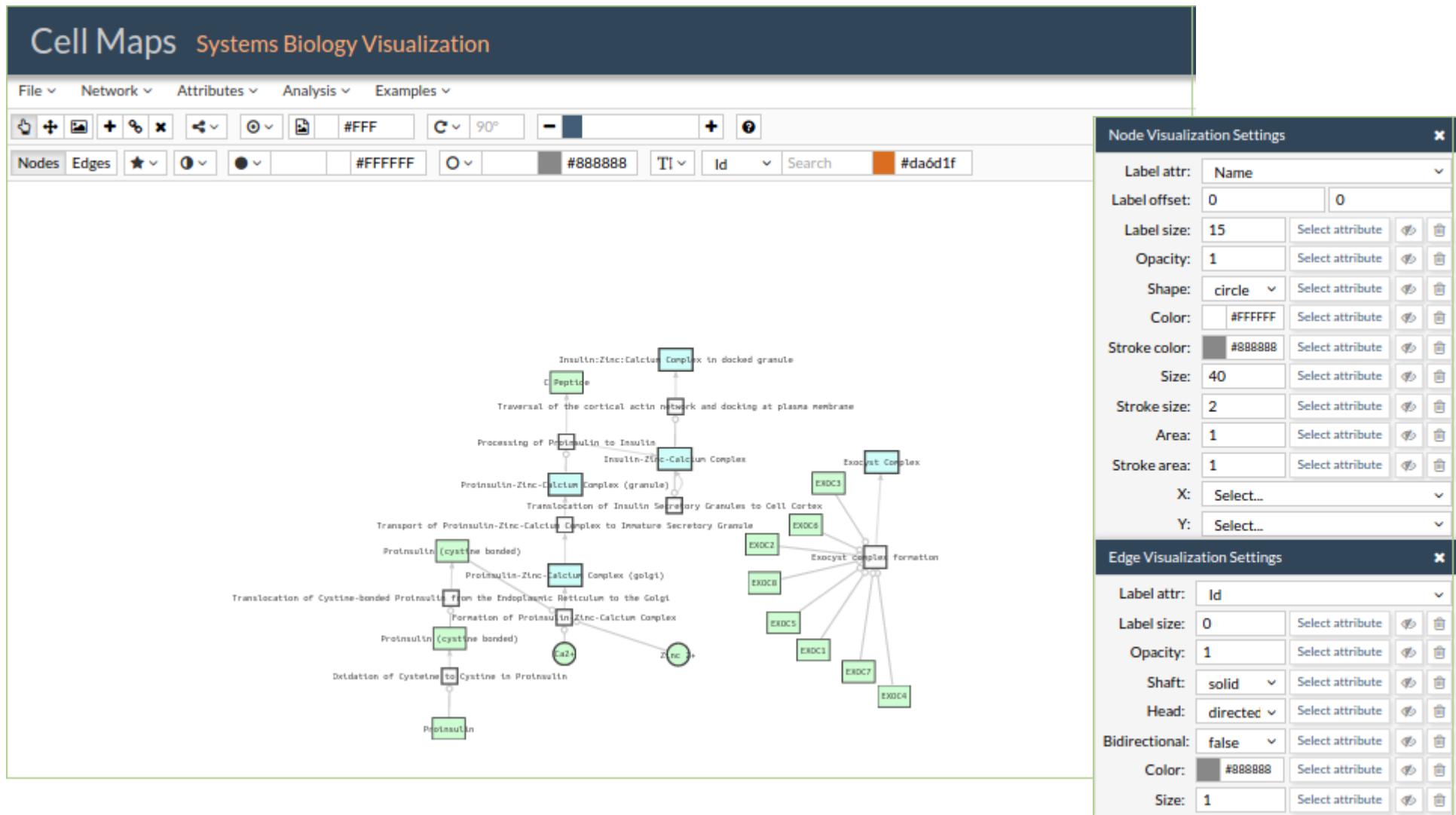
<http://cellmaps.babelomics.org/>

Cell Maps

- 1) Es una herramienta que permite la integración, visualización y el análisis de redes biológicas.
- 2) El **input** es un fichero donde indicamos las relaciones entre los nodos de nuestra red. Opcionalmente podemos incluir un fichero con los atributos de cada nodo.
- 3) El **output gráfico** es una red en la que se muestran las relaciones de los distintos nodos que la integran.

Tutorial: <https://github.com/opencb/cell-maps/wiki>

Tool interface

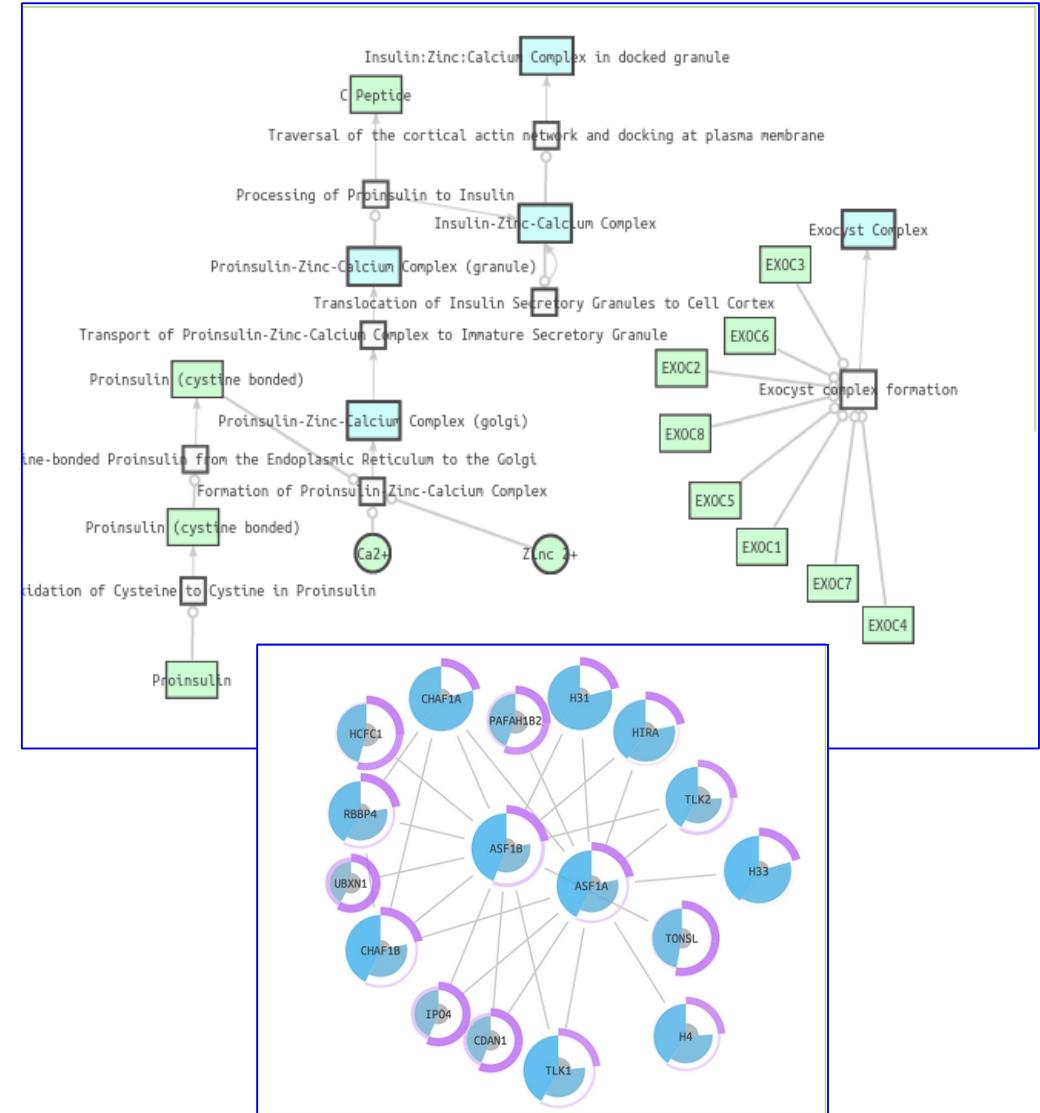
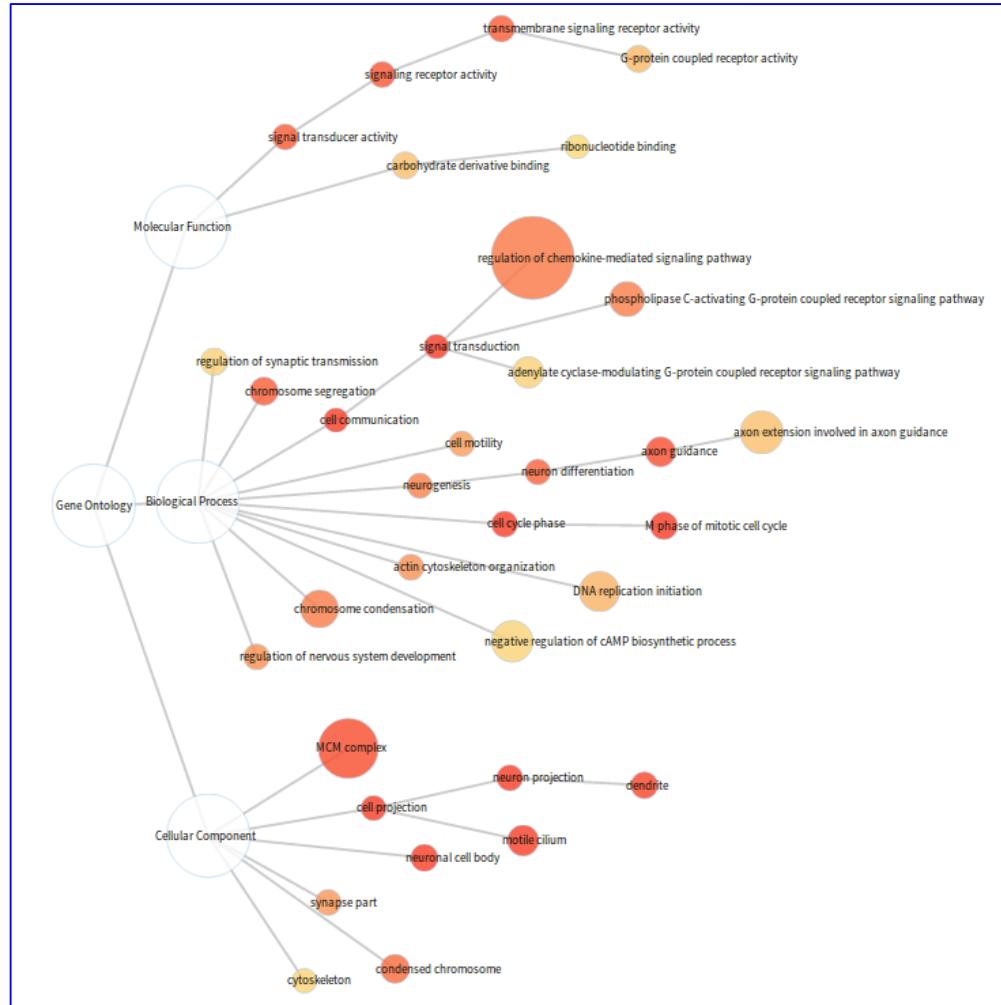


Cell Maps: inputs

```
GO:0000001» pp» GO:0003674
GO:0000001» pp» GO:0005575
GO:0000001» pp» GO:0008150
GO:0003674» pp» GO:0004871
GO:0004871» pp» GO:0038023
GO:0038023» pp» GO:0004888
GO:0004888» pp» GO:0004930
GO:0003674» pp» GO:0097367
GO:0097367» pp»
GO:0005575» pp»
GO:0005575» pp»
GO:0005575» pp»
GO:0005575» pp»
GO:0042995» pp»
GO:0043005» pp»
GO:0042995» pp»
GO:0005575» pp»
```

ID	pvalor	indi2	descriptor
GO:0031514	0.001	0.16	motile cilium
GO:0000793	0.013	0.129	condensed chromosome
GO:0043025	0.001	0.1	neuronal cell body
GO:0030425	0.003	0.094	dendrite
GO:0044456	0.026	0.086	synapse part
GO:0043005	0.000	0.08	neuron projection
GO:0042995	0.001	0.067	cell projection
GO:0005856	0.044	0.059	cytoskeleton

Cell Maps: outputs





Babelomics 5

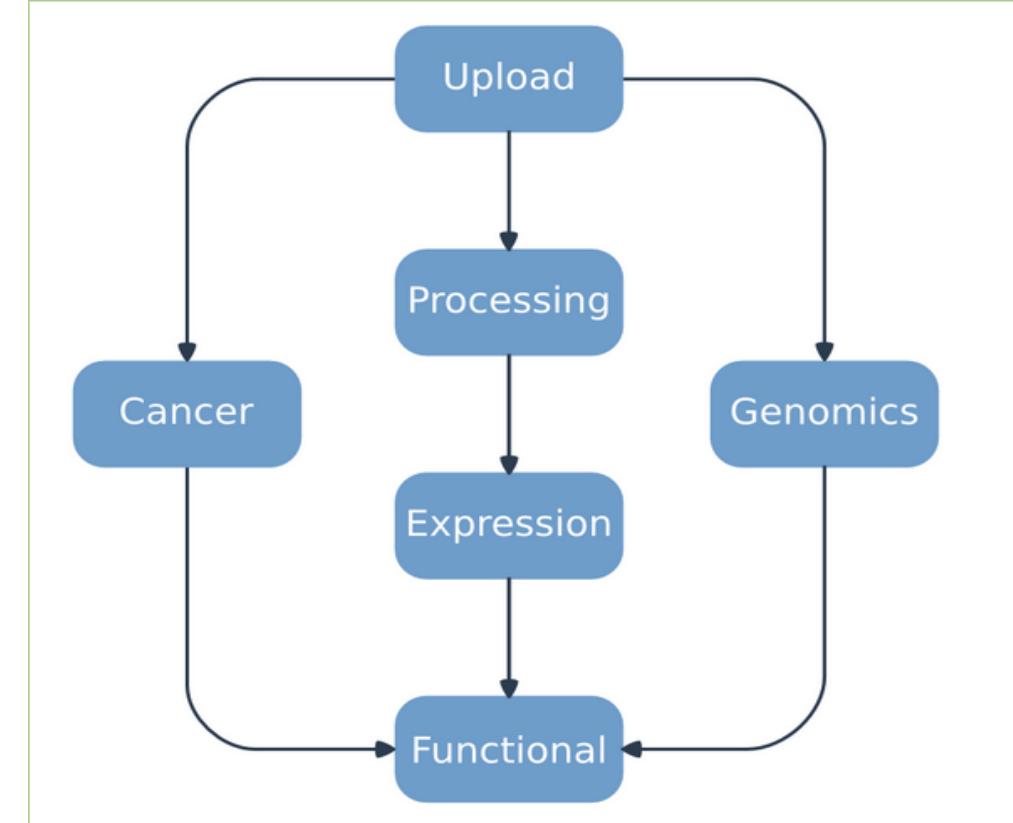
Plataforma de análisis de
datos de Transcriptómica, Proteómica
y Genómica con diferentes abordajes
funcionales

<http://babelomics.bioinfo.cipf.es/>

Tool interface

Babelomics 5

GENE EXPRESSION, GENOME
VARIATION AND FUNCTIONAL
PROFILING ANALYSIS SUITE



More info + questions



Tutorial: web tools



Nucleic Acids Research Advance Access published May 26, 2014

Nucleic Acids Research, 2014 1
doi: 10.1093/nar/gku472

A web tool for the design and management of panels of genes for targeted enrichment and massive sequencing for clinical applications

Alejandro Alemán^{1,2}, Francisco Garcia-Garcia¹, Ignacio Medina¹ and Joaquín Dopazo^{1,2,3,*}

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

²Bioinformatics of Rare Diseases (BIER), CIBER de Enfermedades Raras (CIBERER), Valencia, 46012, Spain and

³Functional Genomics Node, (INB) at CIPF, Valencia, 46012, Spain

Nucleic Acids Research Advance Access published May 6, 2014

Nucleic Acids Research, 2014 1
doi: 10.1093/nar/gku407

A web-based interactive framework to assist in the prioritization of disease candidate genes in whole-exome sequencing studies

Alejandro Alemán^{1,2}, Francisco Garcia-Garcia¹, Francisco Salavert^{1,2}, Ignacio Medina¹ and Joaquín Dopazo^{1,2,3,*}

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

²Bioinformatics of Rare Diseases (BIER), CIBER de Enfermedades Raras (CIBERER), Valencia 46012, Spain and

Published online 8 June 2013

Nucleic Acids Research, 2013, Vol. 41, Web Server issue W41-W46
doi:10.1093/nar/gkt530

Genome Maps, a new generation genome browser

Ignacio Medina^{1,*}, Francisco Salavert^{1,2}, Rubén Sanchez³, Alejandro de María¹, Roberto Alonso¹, Pablo Escobar¹, Marta Bleda^{1,2} and Joaquín Dopazo^{1,2,4,**}

¹Department of Computational Genomics, Centro de Investigación Príncipe Felipe (CIPF), Valencia 46012, Spain, ²CIBER de Enfermedades Raras (CIBERER), Valencia 46012, Spain, ³Genometra S.L., Valencia, Spain

and ⁴Functional Genomics Node (INB) at CIPF, Valencia 46012, Spain

Outline

- 1) Introduction to NGS data analysis**

- 2) Web tools to analyze Genomics Data**

- 3) Let's practise!**

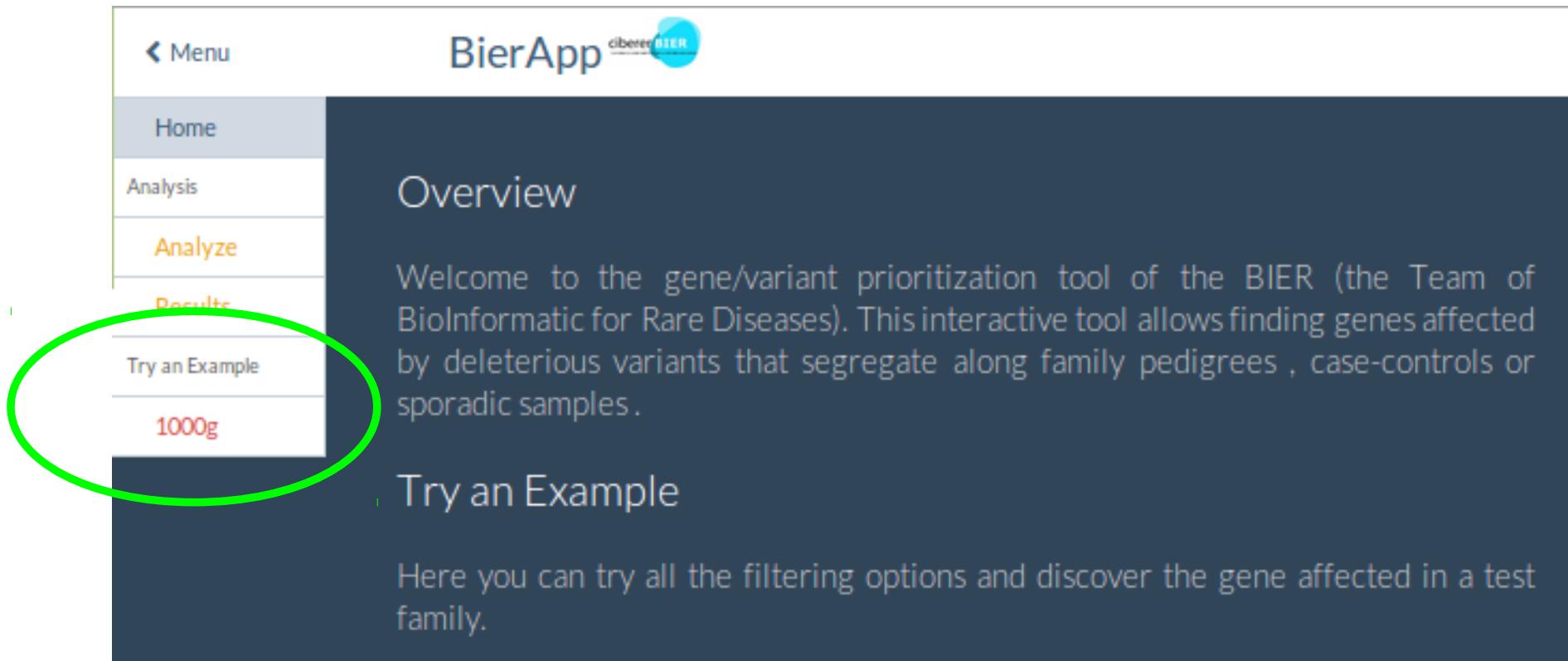
BiERapp:

Una herramienta web para la priorización de variantes

<http://ciberer.es/bier/bierapp>

Cases

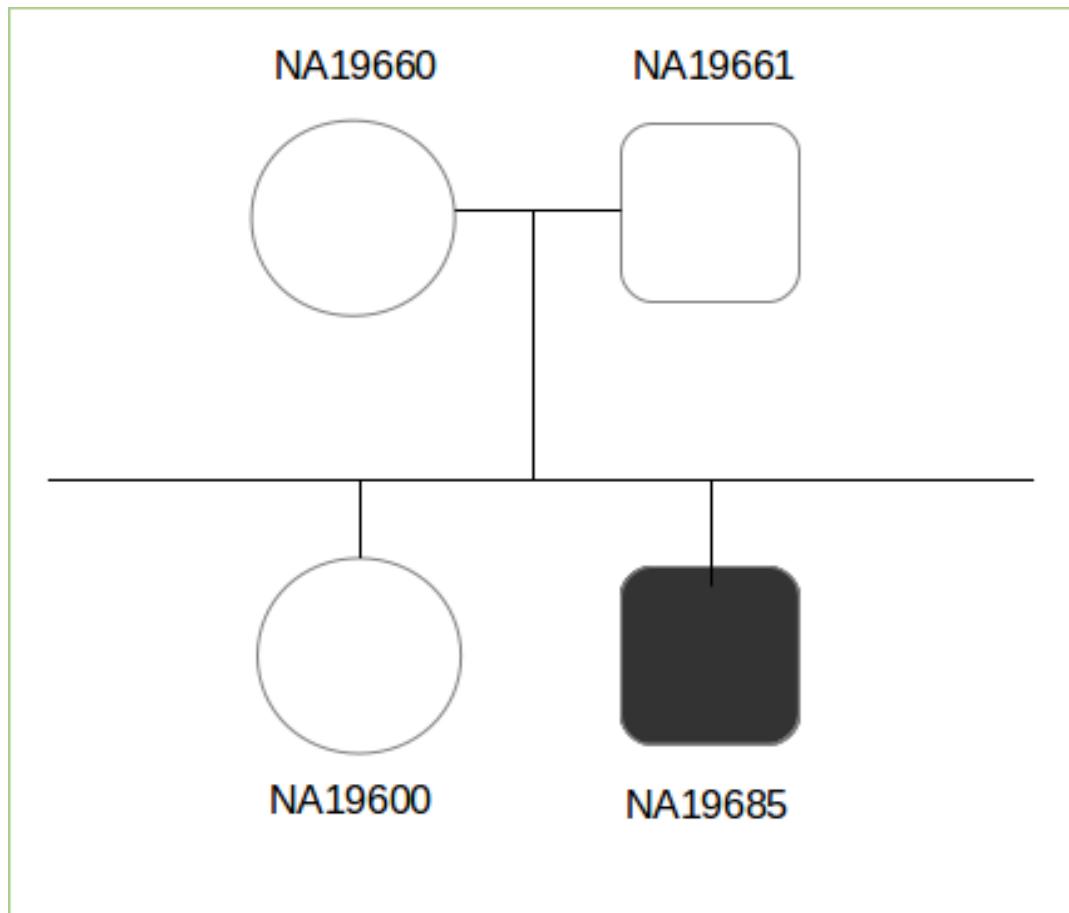
<http://bioinfo.cipf.es/apps-beta/cibererapp/beta/>



The screenshot shows the 'BierApp' interface. On the left, there is a vertical navigation menu with the following items: 'Menu' (with a back arrow), 'Home', 'Analysis', 'Analyze', 'Results' (which is highlighted with a green oval), 'Try an Example', and '1000g'. The main content area has a dark blue background. At the top, it says 'Overview'. Below that, there is a paragraph of text: 'Welcome to the gene/variant prioritization tool of the BIER (the Team of Bioinformatic for Rare Diseases). This interactive tool allows finding genes affected by deleterious variants that segregate along family pedigrees , case-controls or sporadic samples.' Underneath this, there is another section titled 'Try an Example' with the sub-instruction: 'Here you can try all the filtering options and discover the gene affected in a test family.'

Cases

Pedigree



Cases

Case 1.

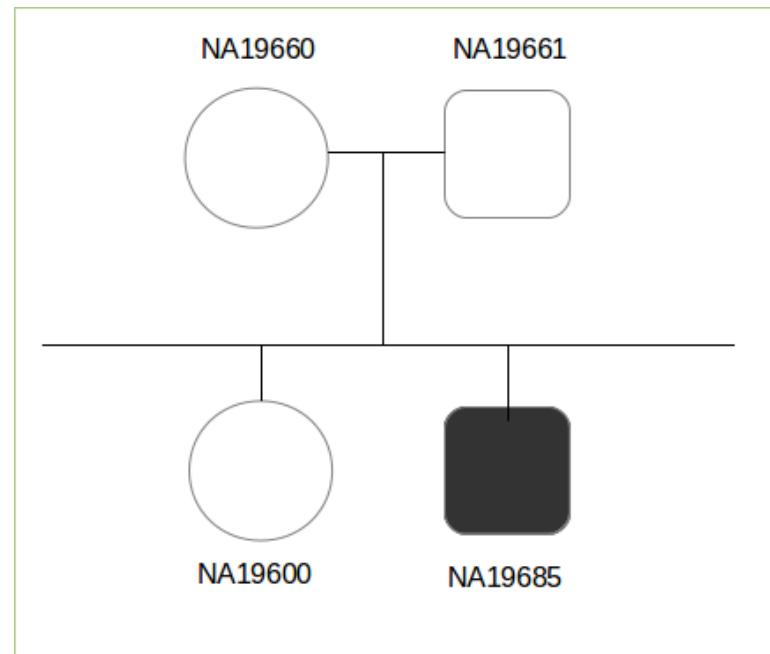
- Dominant heritage

How many variants? **14**

Case 2.

- Recessive heritage

How many variants? **3**



Cases

Case 3.

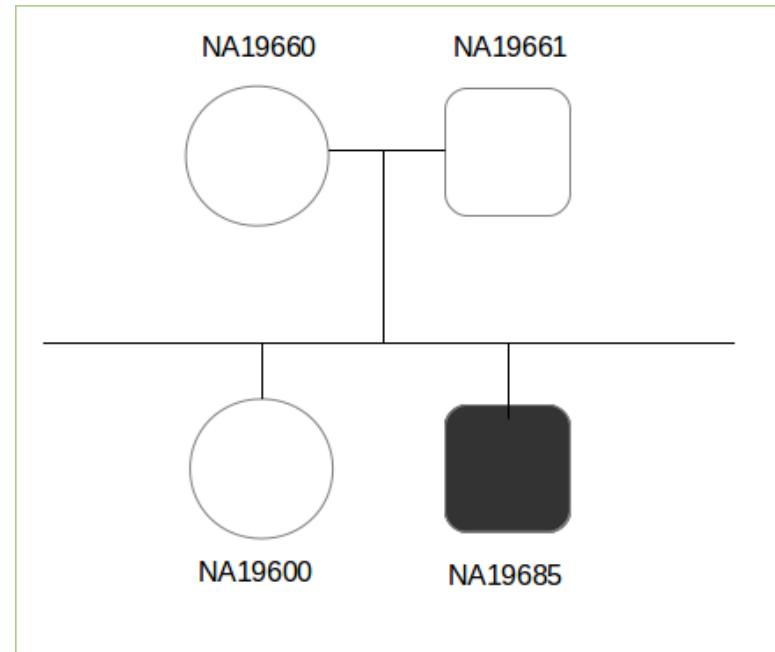
- Dominant heritage
- Rare disease ($MAF < 0.1$)

How many variants? **7**

Case 4.

- Variants in mother and daughter at the same time

How many variants? **85**



Cases

Case 5.

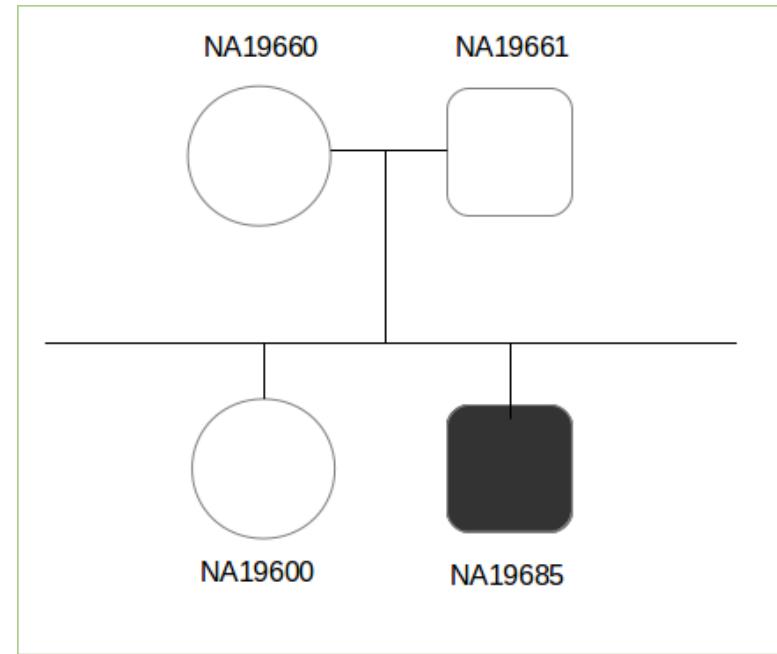
- Variants in mother and daughter at the same time
- Only in chromosome 4

How many variants?

Case 6.

- Variants in mother and daughter at the same time
- Only in these genes:
HEXB,NFKB1,KLRC3

How many variants?



TEAM:

Una **herramienta web** para el diseño y
gestión de **paneles de genes** en
secuenciación dirigida
con aplicaciones clínicas

<http://ciberer.es/bier/team>

Examples

<http://ciberer.es/bier/team>

- 1) Download **example data** from TEAM (3 VCF files).
- 2) **Select the panel** for Retinitis Pigmentosa and **evaluate all three samples**. Do you have variants related to Retinitis for each of the three patients?
- 3) **Generate a PDF report** for each patient including variants related to diagnostic and secondary findings.
- 4) **Design a new panel** for Usher disease.

CSVS: CIBERER Spanish Variant Server

Repositorio de frecuencias de variantes
en la población española

<http://bioinfo.cipf.es/apps-beta/spvs/1.0.0/>

Examples

<http://bioinfo.cipf.es/apps-beta/spvs/1.0.0/>

- 1) How many variants do you find in region:
1:24400-70000? (33 variants)

- 2) What information does SPVS give us for
this position 1:24536? (Effect, phenotype...)



Babelomics 5

Plataforma de análisis de
datos de Transcriptómica, Proteómica
y Genómica con diferentes abordajes
funcionales

<http://babelomics.bioinfo.cipf.es/>

Examples

<http://babelomics.bioinfo.cipf.es/>

- 1) We are searching new functional candidates.
This is the starting point:
BEST1 C2orf71 CA4 CERKL CNGA1 CNGB1 CRB1 CRX EYS
GUCA1B IDH3B USH2A

- 2) From Babelomics explore these new candidates using two approaches: Single Enrichment and Network Enrichment

Genome Maps

Visualizador genómico que interactúa
con bases de datos funcionales

<http://genomemaps.org/>

Examples

<http://genomemaps.org/>

- 1) Visualize this region: 1:100000-200000
- 2) Visualize this gene: LIN28A
- 3) Add new traks: miRNA, TFBS

Más?

- Curso CIBERER de análisis de datos genómico, después del verano.
- Colaboraciones entre grupos CIBERER: ayudas de movilidad.
- <http://bioinfo.cipf.es/>