TEAM: Targeted Enrichment Analysis and Management

Alejandro Alemán
aaleman@cipf.es

TEAM: Targeted Enrichment Analysis and Management

Outline

1) Introduction

2) How does TEAM work?

3) Results

4) Conclusions

5) Examples

Outline

1) Introduction

2) How does TEAM work?

3) Results

4) Conclusions

5) Examples

Introduction

- **Development of high throughput sequencing technologies:**
  - Rapid and economical genome sequencing
  - Disease targeted sequencing: powerful and cost-effective application.

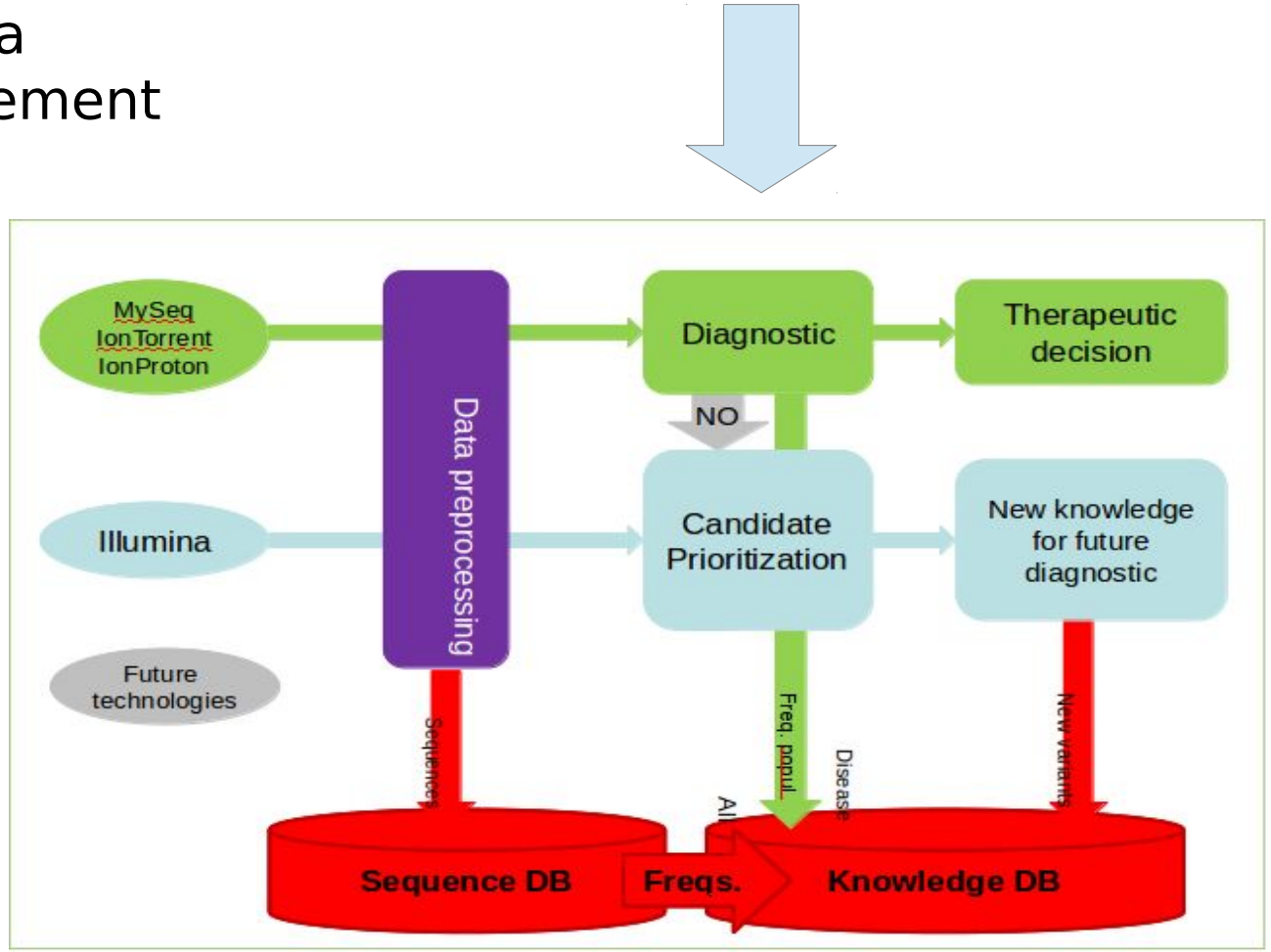
- **Vast amount of biological knowledge available:**
  - HGMD-public, HUMSAVAR, ClinVar, COSMIC

- We need a tool to connect **sequencing data and biological knowledge for diagnostic**:
  - **TEAM** (**T**argeted **E**nrichment **A**nalysis and **M**anagement)

TEAM | **TEAM: Targeted Enrichment Analysis and Management**

Introduction

- **Confidentiality**
  - Sensitive data
  - Local management



Outline

1) Introduction

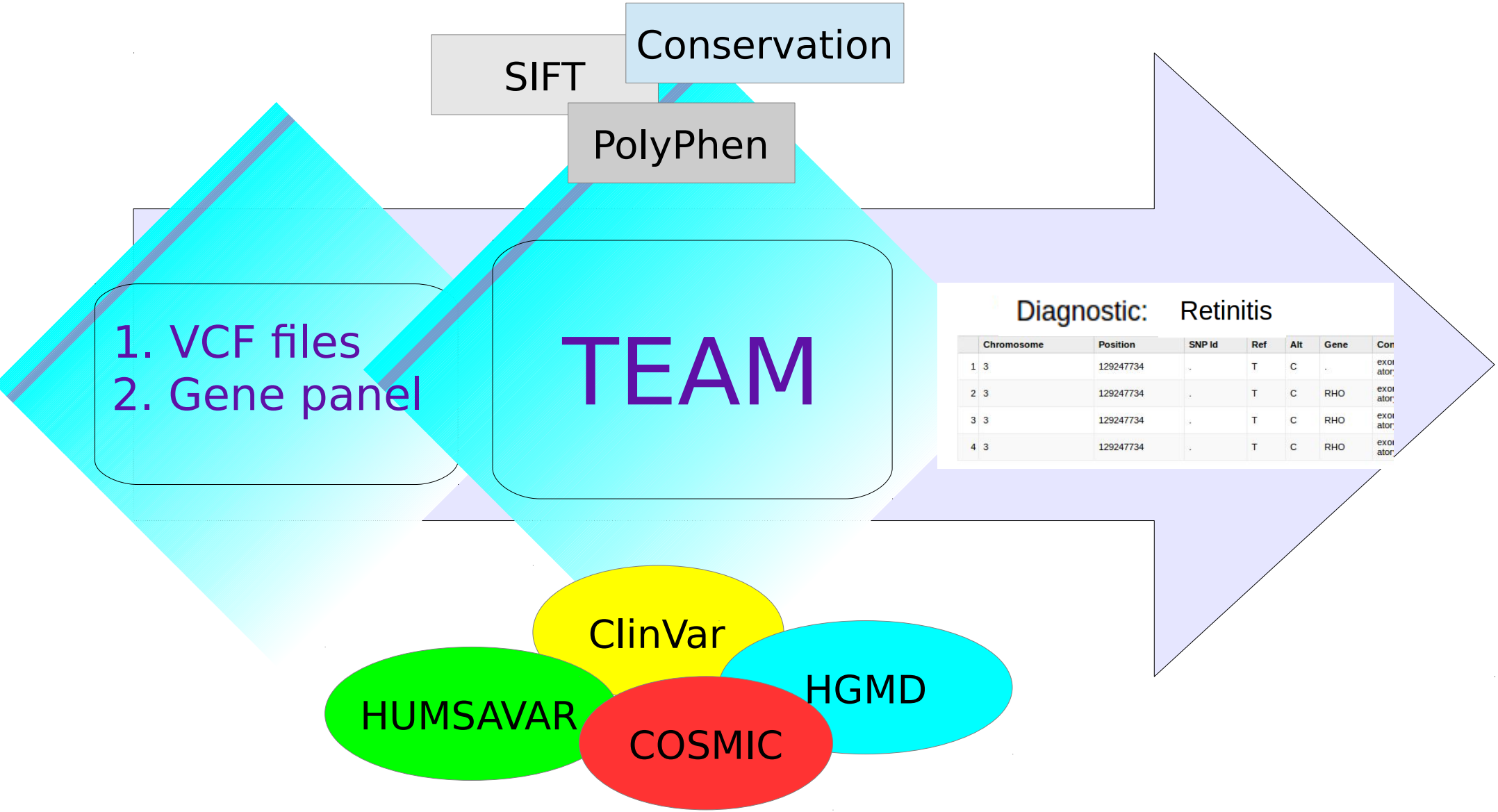
2) How does TEAM work?

3) Results

4) Conclusions

5) Examples

How does TEAM work?

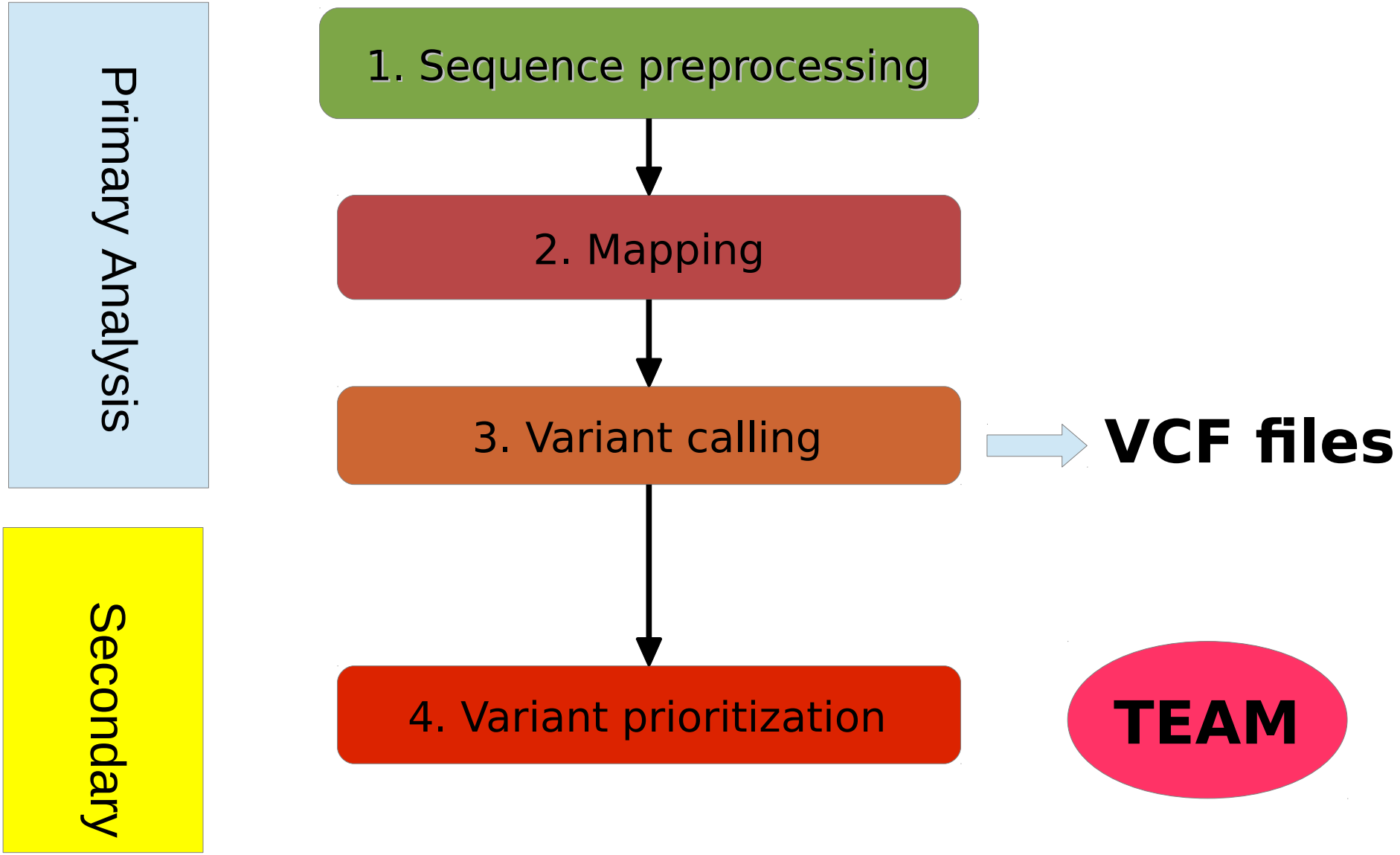


	Chromosome	Position	SNP Id	Ref	Alt	Gene	Cor
1	3	129247734	.	T	C	.	exo ator
2	3	129247734	.	T	C	RHO	exo ator
3	3	129247734	.	T	C	RHO	exo ator
4	3	129247734	.	T	C	RHO	exo ator

TEAM

TEAM: Targeted Enrichment Analysis and Management

Inputs: VCF file



TEAM

TEAM: Targeted Enrichment Analysis and Management

Inputs: VCF file

#CHROM	POS	ID	REF	ALT	QUAL	FILTER	INFO	FORMAT	D053
1	94471075		.	A	G	3331.06	PASS	AC=2;AF=1.00;AN=2;[	
1	94471154		.	C	T	1015.91	PASS	AC=1;AF=0.50;AN=2;[	
1	94473845		.	T	C	431.58	PASS	AC=1;AF=0.50;AN=2;[	
1	94473864		.	C	T	361.62	PASS	AC=1;AF=0.50;AN=2;[	
1	94474328		.	T	C	2760.20	PASS	AC=1;AF=0.50;AN=2;[	
1	94474452		.	T	G	1126.22	PASS	AC=1;AF=0.50;AN=2;[	
1	94476388		.	C	G	694.59	PASS	AC=1;AF=0.50;AN=2;[	
1	94480037		.	C	T	2312.14	PASS	AC=1;AF=0.50;AN=2;[	
1	94544234		.	T	C	2562.62	PASS	AC=1;AF=0.50;AN=2;[	
1	94544276		.	G	A	1680.20	PASS	AC=1;AF=0.50;AN=2;[	

One VCF (Variant Calling Format) **file for each patient**

GT:AD:DP:GQ:PL	0/1:26,32:58:99:1046,0,814
GT:AD:DP:GQ:PL	0/1:15,16:31:99:462,0,501
GT:AD:DP:GQ:PL	0/1:16,12:28:99:392,0,490
GT:AD:DP:GQ:PL	0/1:67,80:147:99:2790,0,2114
GT:AD:DP:GQ:PL	0/1:48,37:85:99:1156,0,1596
GT:AD:DP:GQ:PL	0/1:26,23:49:99:725,0,877
GT:AD:DP:GQ:PL	0/1:64,73:137:99:2342,0,1923
GT:AD:DP:GQ:PL	0/1:74,84:158:99:2593,0,2334
GT:AD:DP:GQ:PL	0/1:88,56:144:99:1710,0,3060
GT:AD:DP:GQ:PL	0/1:49,35:84:99:1080,0,1641

Inputs: Gene Panel

Gene Panels:

Cancer Disease	Gene Panel Description
Hereditary Breast and Ovarian Cancer	**13 genes panel:** BRCA1, BRCA2, RAD51C, CDH1, TP53, PTEN, STK11, PALB2, RAD51D, BRIP1, XRCC2, ERCC4 , ATM.
Hereditary Nonpolyposis Colon Cancer	**4 genes panel:** MLH1, MSH2, MSH6, PMS1
Hereditary Colorectal Adenomatous Polyposis	**2 genes panel:** APC, MUTYH
Juvenile Polyposis Syndrome	**3 genes panel:** SMAD14, BMPR1A, PTEN
PTEN Hamartoma Tumor Syndrome (PHTS)	**1 gene panel:** PTEN

List of genes

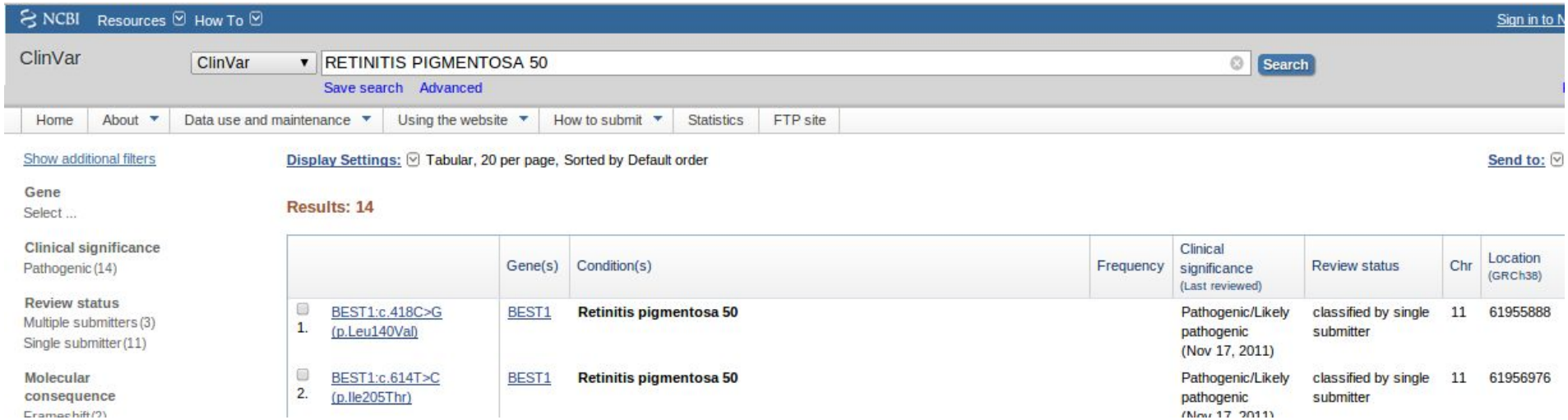
BRCA1
BRCA2
RAD51C
CDH1
TP53
PTEN
STK11

BED file including genes

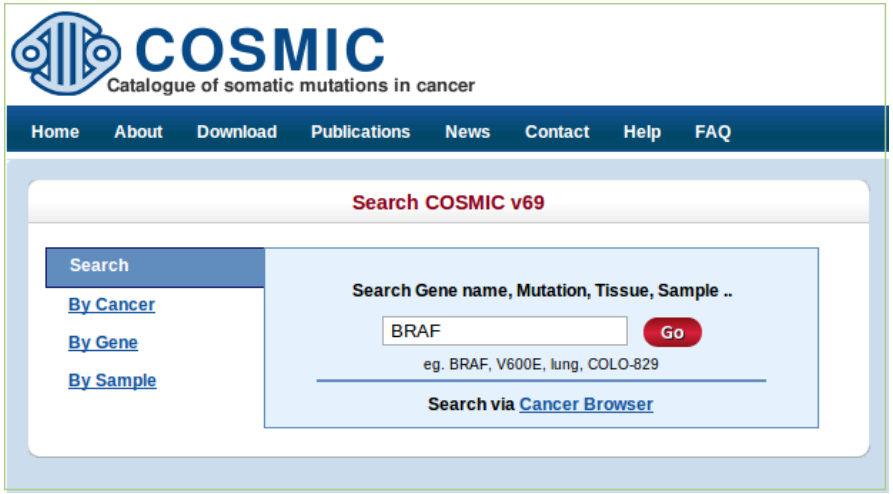
chr17	41196312	41322290	BRCA1
chr13	32889611	32973805	BRCA2
chr17	56769934	56811703	RAD51C
.....			

Getting information

https://www.ncbi.nlm.nih.gov/clinvar

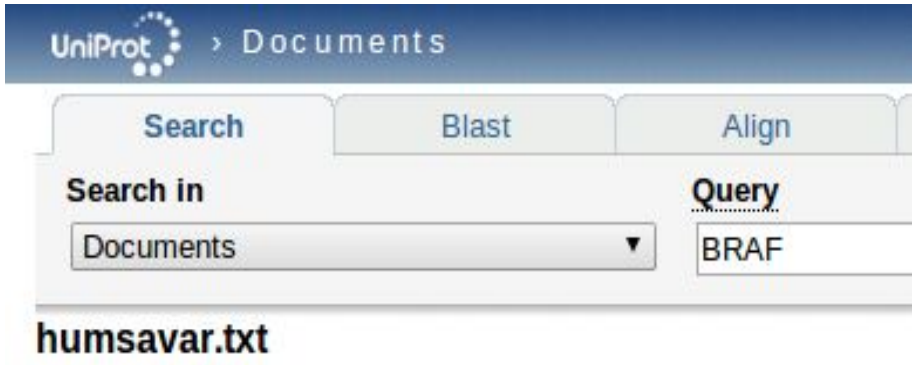


http://cancer.sanger.ac.uk/cancergenome/projects/cosmic/



Getting information

http://www.hgmd.org/



http://www.uniprot.org/docs/humsavar

Getting information

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

http://sift.jcvi.org/

- **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)

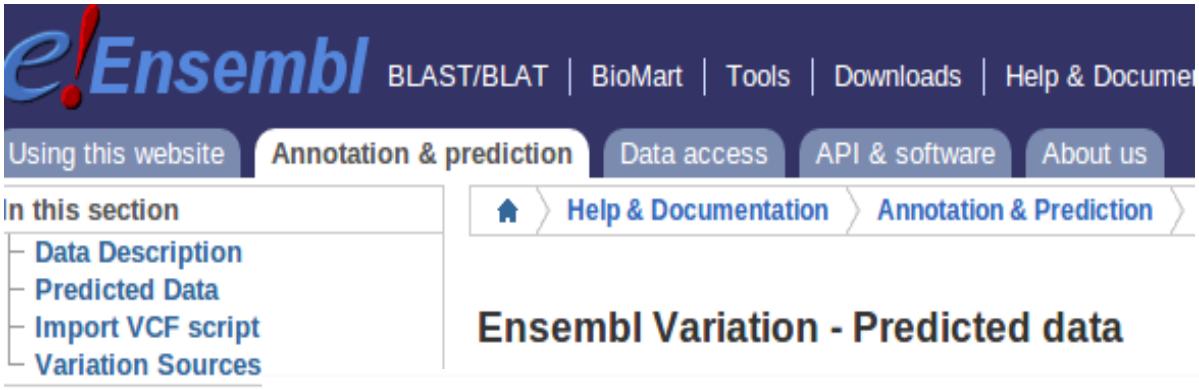
http://genetics.bwh.harvard.edu/pph2/index.shtml

Getting information

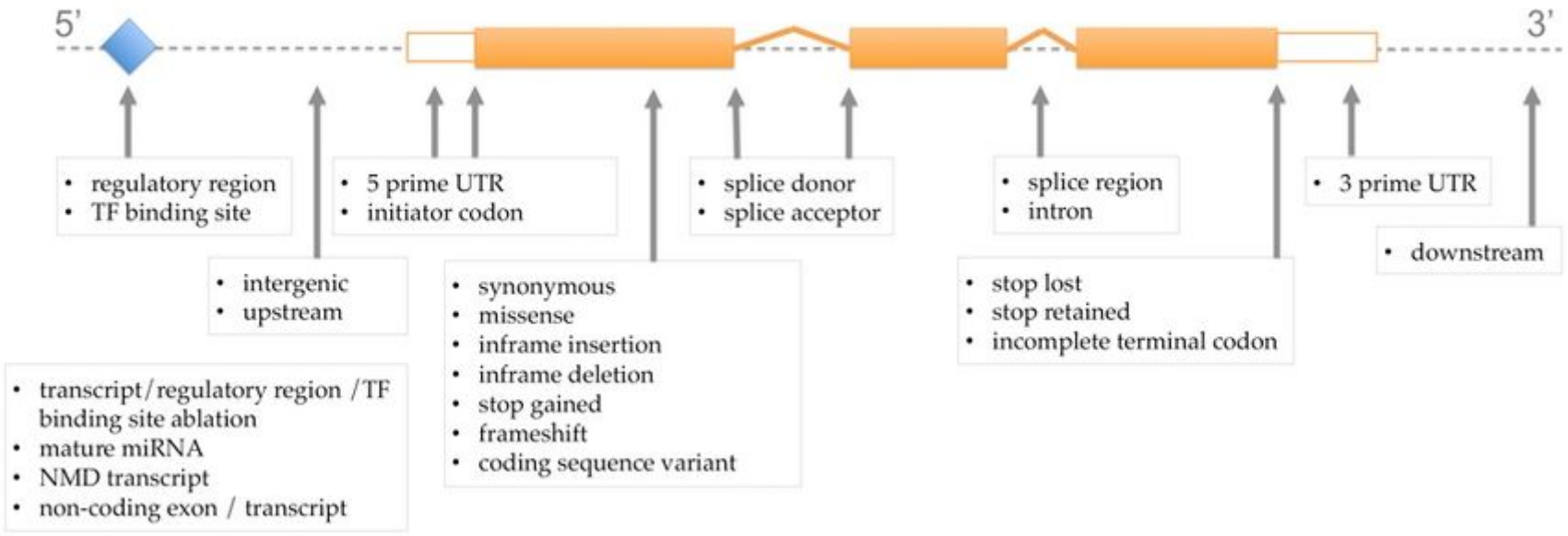
- **PhastCons HOWTO**
  - **Conservation Index:** multi-species conservation for deciding on potential pathogenicity of novel variants
  - **Interpretation**: 1 (high conservation) to 0 (low conservation)

http://compgen.bscb.cornell.edu/phast/phastCons-HOWTO.html

Getting information



Consequence type or effect

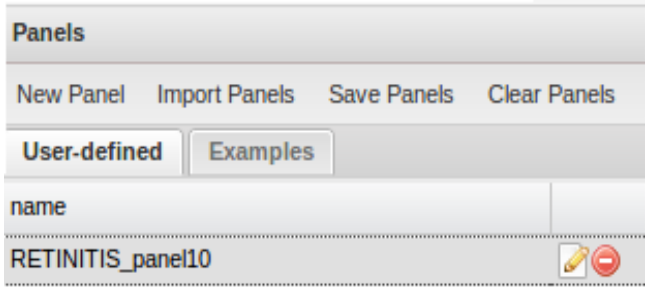


http://www.ensembl.org/info/genome/variation/predicted_data.html

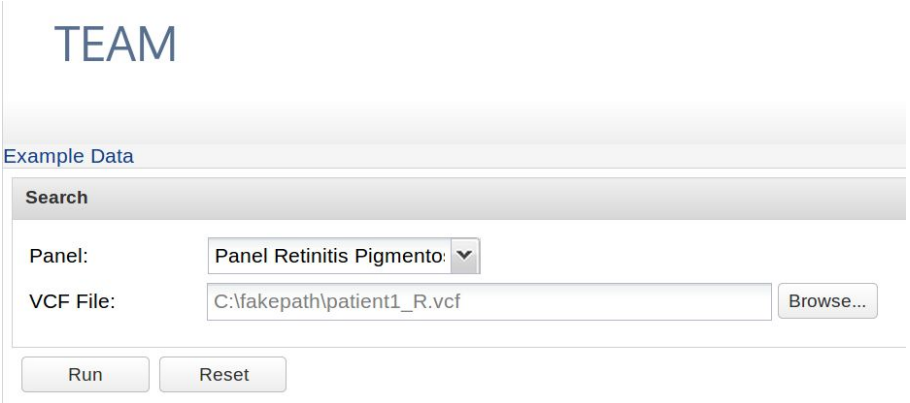
Tool interface

http://team.babelomics.org/

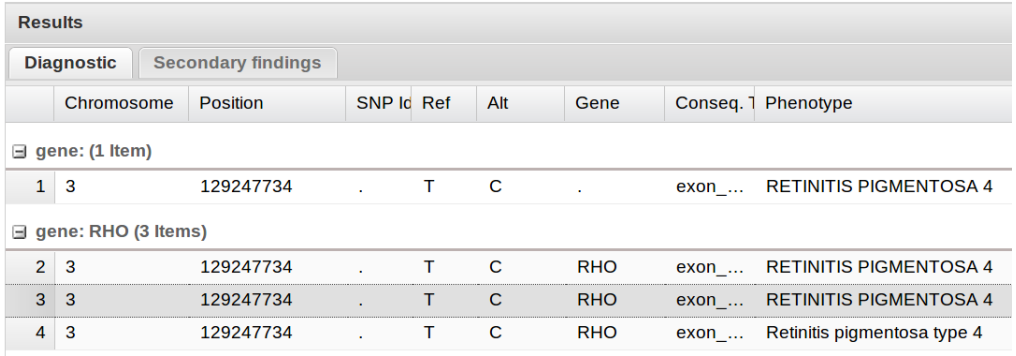
1. Defining panel



2. Uploading input data



3. Getting results



	Chromosome	Position	SNP Id	Ref	Alt	Gene	Conseq. T	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	exon_...	Retinitis pigmentosa type 4

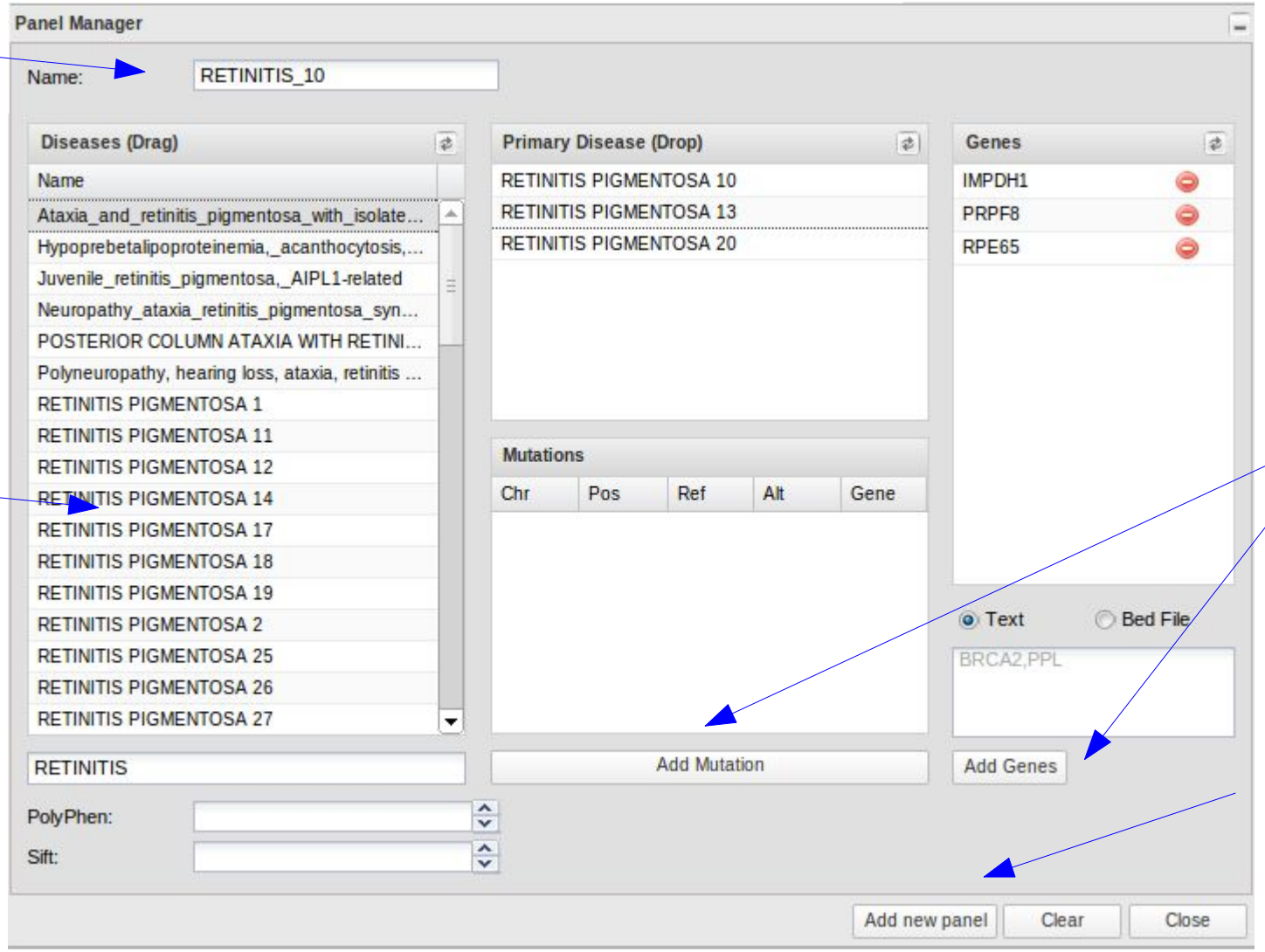
How to define a panel?

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

Outline

1) Introduction

2) How does TEAM work?

3) Results

4) Conclusions

5) Examples

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)								

Variant Effect - 3:129247734 T>C

	Position chr:start:end (strand)	SNP Id	Conseq. Type	Aminoacid Change
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	Con
1	3	129247734	.	T	C	.	exon ator
2	3	129247734	.	T	C	RHO	exon ator
3	3	129247734	.	T	C	RHO	exon ator
4	3	129247734	.	T	C	RHO	exon ator

Outline

1) Introduction

2) How does TEAM work?

3) Results

4) Conclusions

5) Examples

Conclusions

- 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**

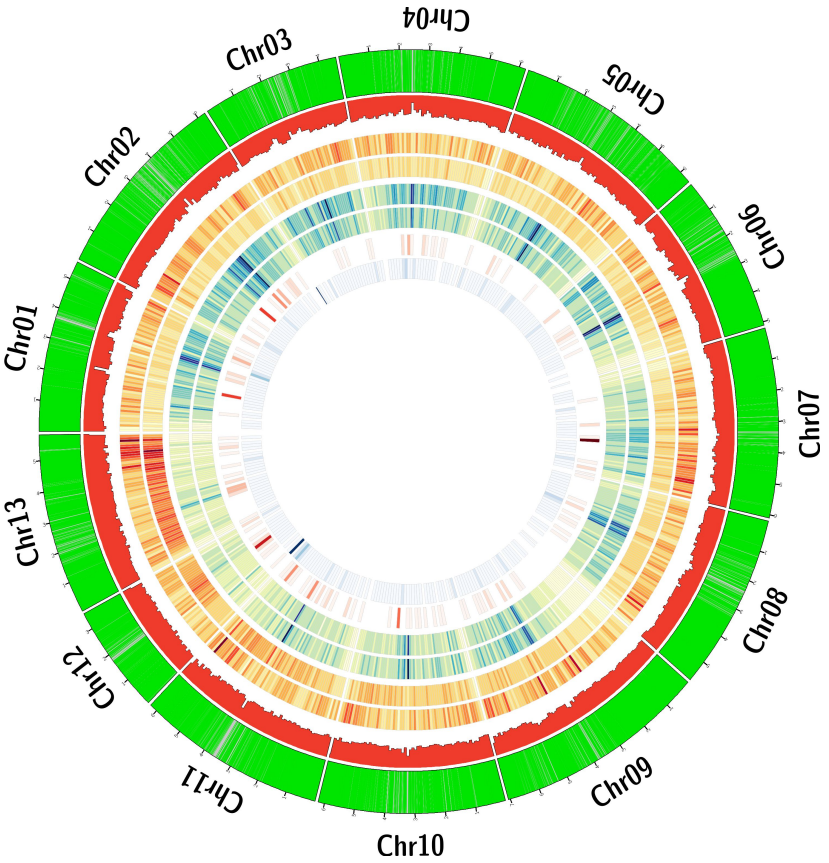
- The entire patient's sequencing information is managed locally thus avoiding any problem of data **privacy or confidentiality**

Next improvements:

- Inclusion of a **database with public panels genes** of various diseases

- **Comparative Analysis** for groups of panels

- **Visualization results**



More info + questions

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-Garcia[1], Ignacio Medina[1] and Joaquín Dopazo[1,2,3,*]

[1]Computational Genomics Department, Centro de Investigación Príncipe Felipe (CIPF), Valencia, 46012, Spain,
[2]Bioinformatics of Rare Diseases (BIER), CIBER de Enfermedades Raras (CIBERER), Valencia, 46012, Spain and
[3]Functional Genomics Node, (INB) at CIPF, Valencia, 46012, Spain

Outline

1) Introduction

2) How does TEAM work?

3) Results

4) Conclusions

5) Examples

Examples

http://team.babelomics.org