

TEAM: Targeted Enrichment Analysis and Management



PRINCIPE FELIPE
CENTRO DE INVESTIGACION

Computational · Genomics



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

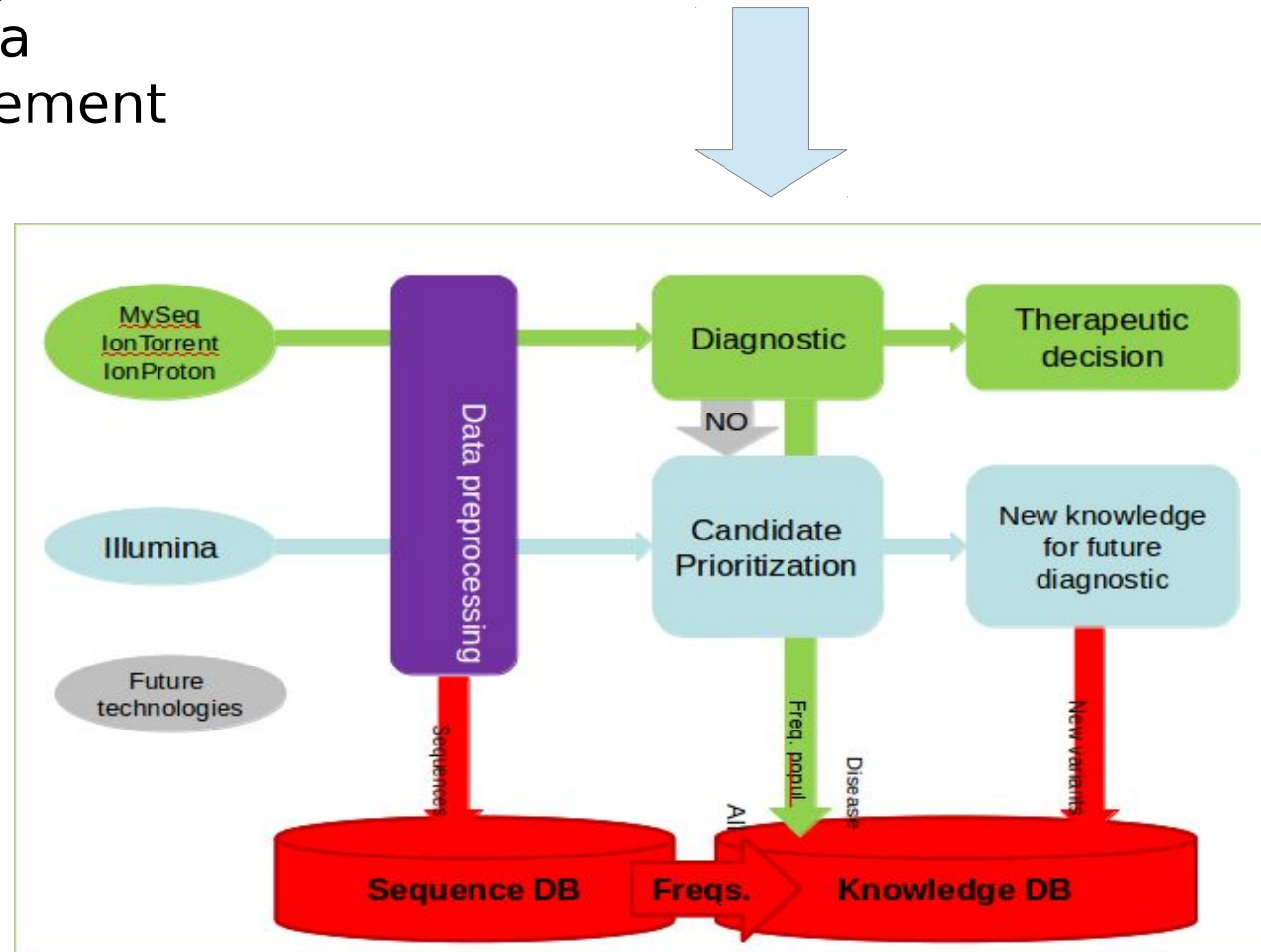
- **Development of high throughput sequencing technologies:**
 - Rapid and economical genome sequencing
 - Disease targeted sequencing: powerful and cost-effective application.
- **Vast amount of biological knowledge available:**
 - HGMD-public, HUMSAVAR, ClinVar, COSMIC
- We need a tool to connect **sequencing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

TEAM

TEAM: Targeted Enrichment Analysis and Management

Introduction

- Confidentiality
 - Sensitive data
 - Local management



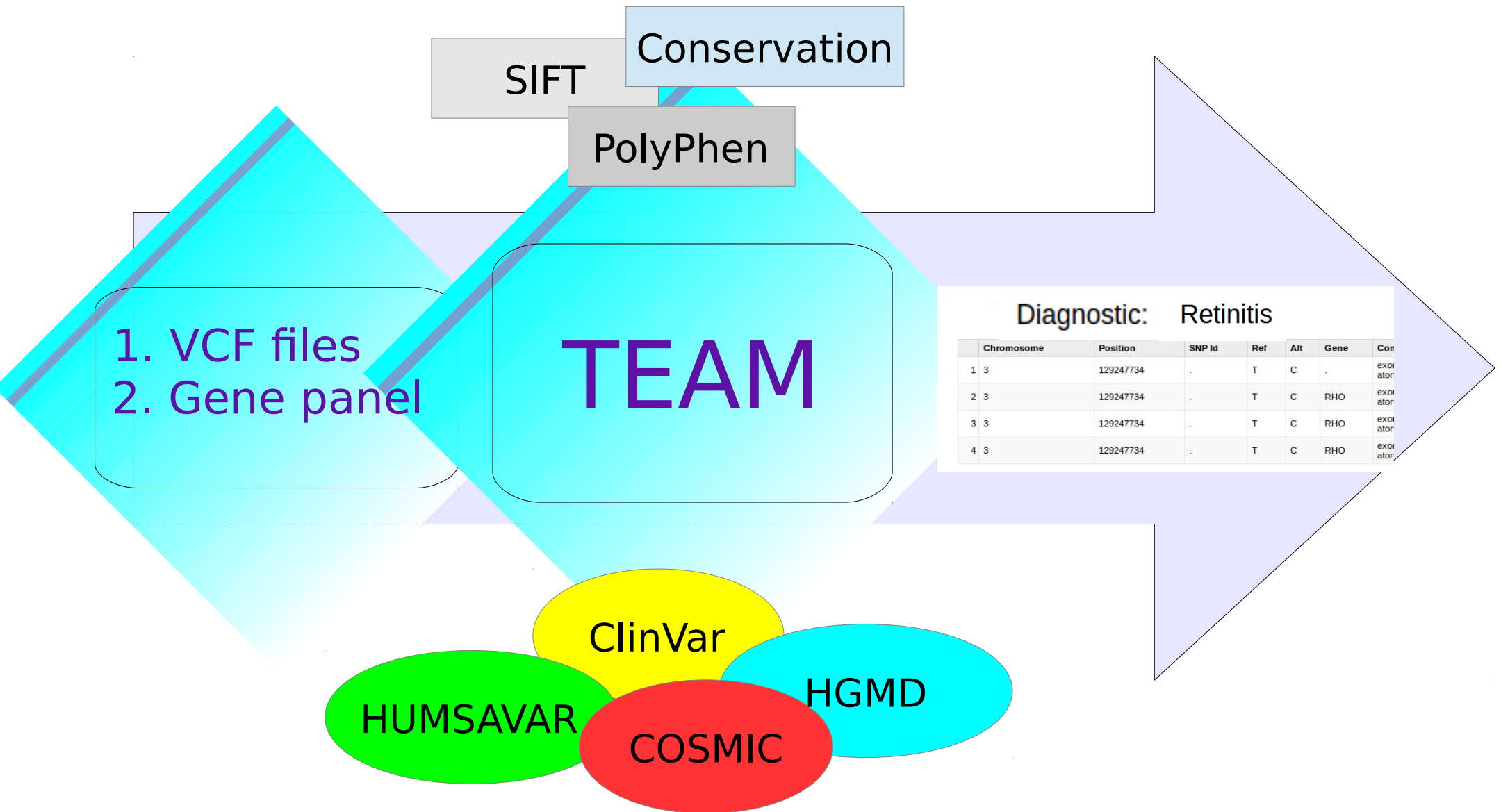
TEAM

TEAM: Targeted Enrichment Analysis and Management

Outline

- 1) Introduction
- 2) How does TEAM work?
- 3) Results
- 4) Conclusions
- 5) Examples

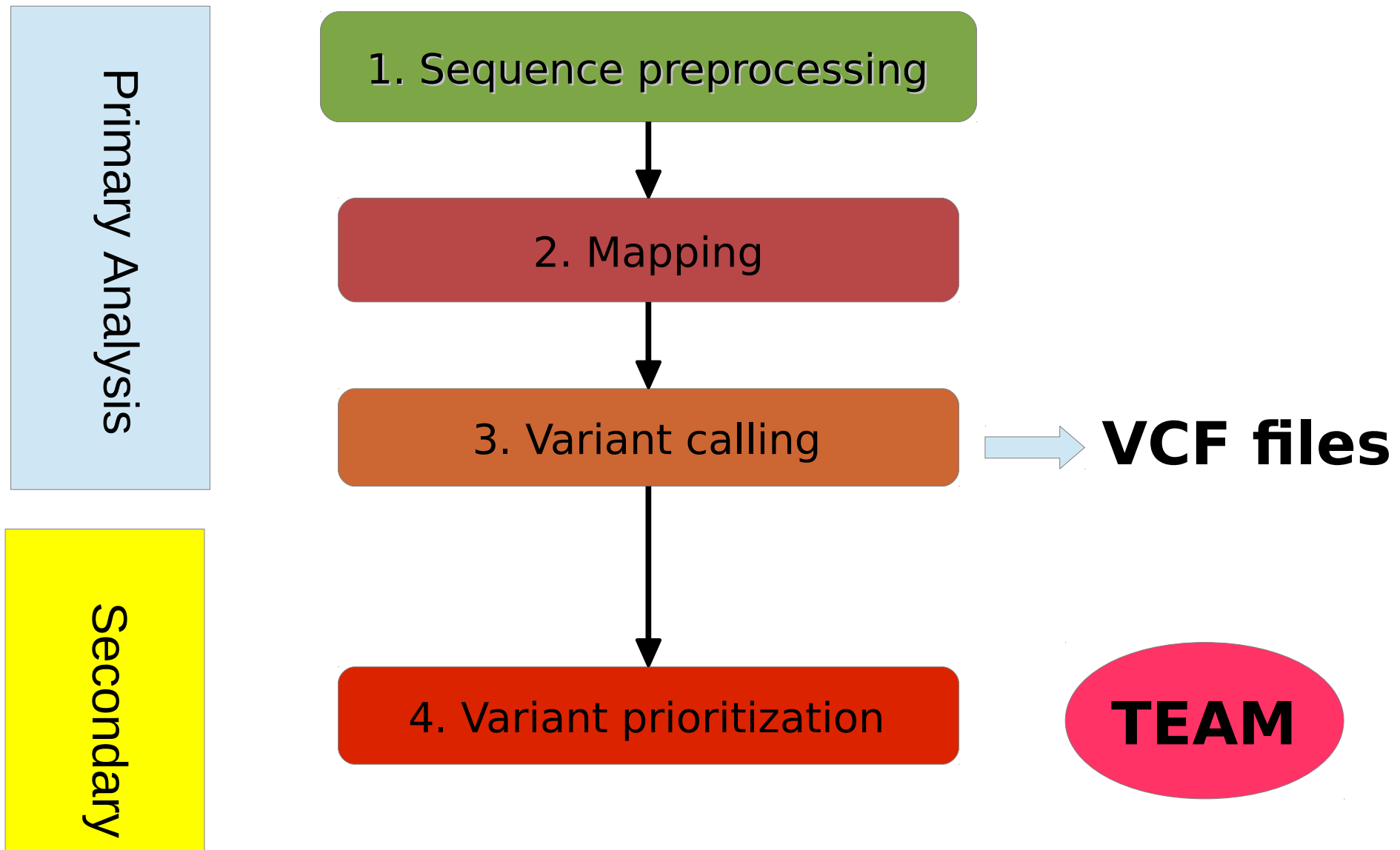
How does TEAM work?



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;
```

```
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
```

One VCF (Variant Calling Format) file for each patient

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  
.....
```

TEAM

TEAM: Targeted Enrichment Analysis and Management

Getting information

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

The screenshot shows the NCBI ClinVar search interface. The search term is "RETINITIS PIGMENTOSA 50". The results are displayed in a table with 7 columns: Gene(s), Condition(s), Frequency, Clinical significance (Last reviewed), Review status, Chr, and Location (GRCh38). Two results are shown, both for the gene BEST1 and condition Retinitis pigmentosa 50. The first result is a missense mutation (p.Leu140Val) classified as Pathogenic/Likely pathogenic (Nov 17, 2011). The second result is another missense mutation (p.Ile205Thr) also classified as Pathogenic/Likely pathogenic (Nov 17, 2011).

	Gene(s)	Condition(s)	Frequency	Clinical significance (Last reviewed)	Review status	Chr	Location (GRCh38)
1. <input type="checkbox"/> BEST1:c.418C>G (p.Leu140Val)	BEST1	Retinitis pigmentosa 50		Pathogenic/Likely pathogenic (Nov 17, 2011)	classified by single submitter	11	61955888
2. <input type="checkbox"/> BEST1:c.614T>C (p.Ile205Thr)	BEST1	Retinitis pigmentosa 50		Pathogenic/Likely pathogenic (Nov 17, 2011)	classified by single submitter	11	61956976

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

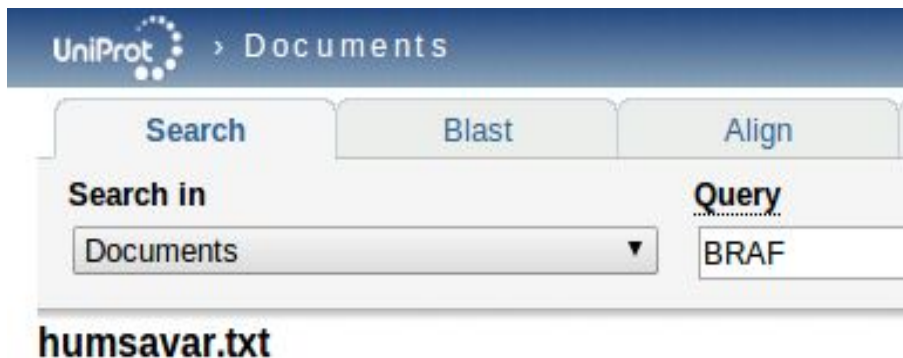
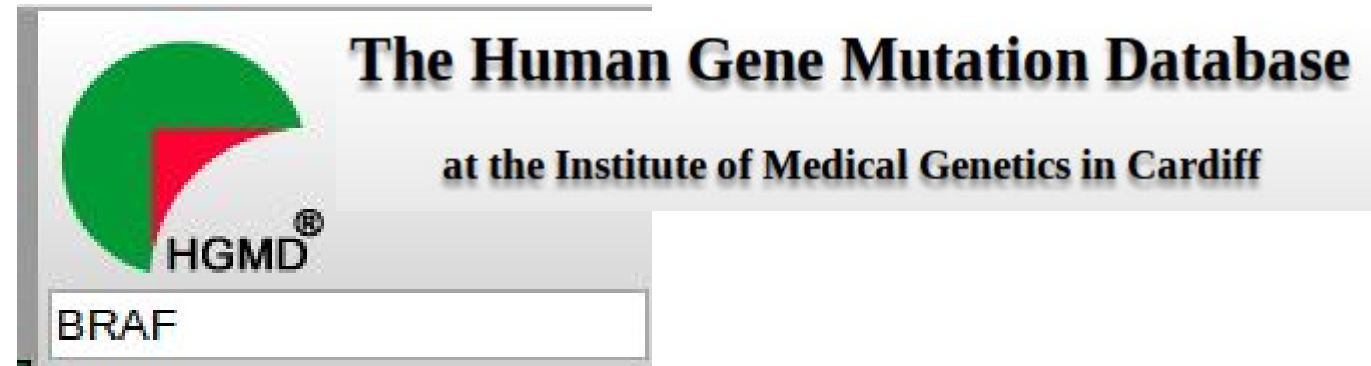
The screenshot shows the COSMIC (Catalogue of somatic mutations in cancer) search interface. The search term is "BRAF". The interface includes a search bar, a "Go" button, and a "Search via Cancer Browser" link. The search results are not visible in this view.

TEAM

TEAM: Targeted Enrichment Analysis and Management

Getting information

<http://www.hgmd.org/>



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

TEAM

TEAM: Targeted Enrichment Analysis and Management

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 Venter™
I N S T I T U T E

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>



TEAM

TEAM: Targeted Enrichment Analysis and Management

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)

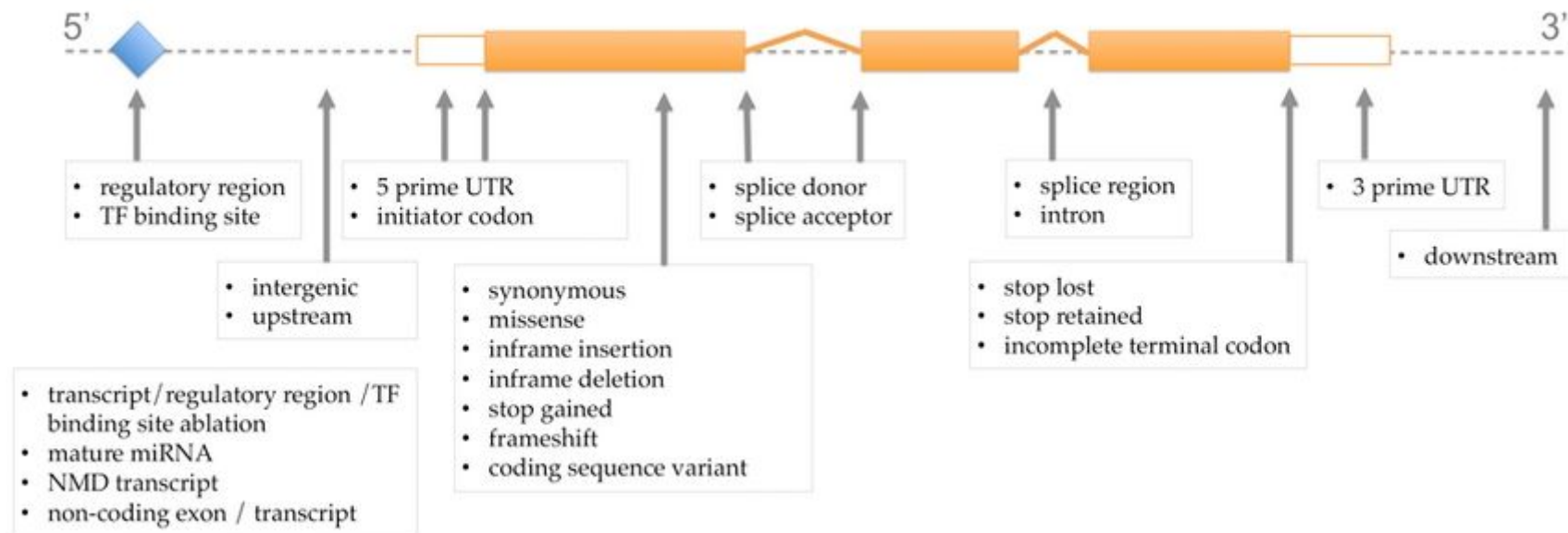
<http://compgen.bscb.cornell.edu/phast/phastCons-HOWTO.html>

Getting information

Consequence type or effect

The screenshot shows the Ensembl website header with navigation links: BLAST/BLAT, BioMart, Tools, Downloads, and Help & Documentation. Below this is a secondary menu with 'Using this website', 'Annotation & prediction', 'Data access', 'API & software', and 'About us'. A 'In this section' sidebar lists 'Data Description', 'Predicted Data', 'Import VCF script', and 'Variation Sources'. The main navigation path is 'Home > Help & Documentation > Annotation & Prediction'.

Ensembl Variation - Predicted data



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

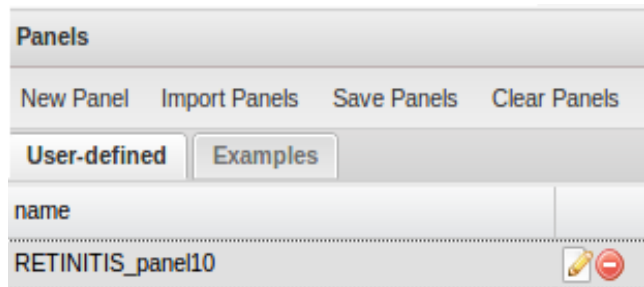
TEAM

TEAM: Targeted Enrichment Analysis and Management

Tool interface

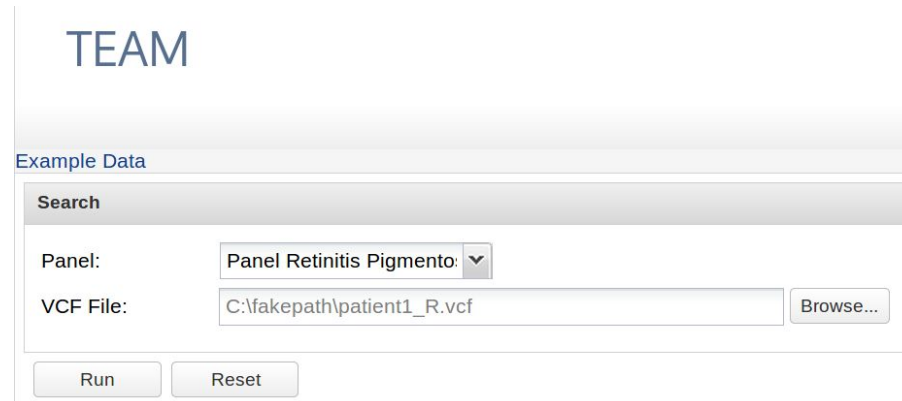
<http://team.babelomics.org/>

1. Defining panel



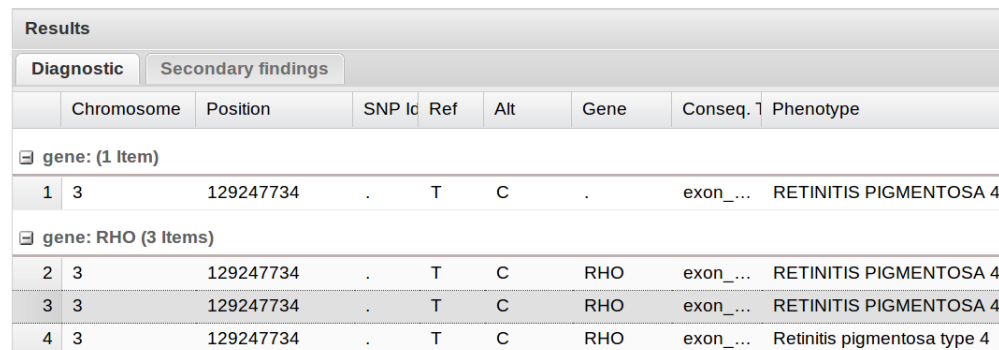
The 'Panels' interface includes buttons for 'New Panel', 'Import Panels', 'Save Panels', and 'Clear Panels'. It has two tabs: 'User-defined' and 'Examples'. A form below has a 'name' field containing 'RETINITIS_panel10' and edit/delete icons.

2. Uploading input data



The 'TEAM' interface shows 'Example Data' with a 'Search' section. It includes a 'Panel' dropdown menu set to 'Panel Retinitis Pigmento...', a 'VCF File' text input field containing 'C:\fakepath\patient1_R.vcf', and a 'Browse...' button. 'Run' and 'Reset' buttons are at the bottom.

3. Getting results



The 'Results' interface has tabs for 'Diagnostic' and 'Secondary findings'. The 'Diagnostic' tab shows a table with columns: Chromosome, Position, SNP Id, Ref, Alt, Gene, Conseq. 1, and Phenotype. The table lists findings for gene: (1 Item) and gene: RHO (3 Items).

	Chromosome	Position	SNP Id	Ref	Alt	Gene	Conseq. 1	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

TEAM

TEAM: Targeted Enrichment Analysis and Management

How to define a panel?

1. Name of panel

The screenshot shows the 'Panel Manager' window with the following components:

- Name:** A text field containing 'RETINITIS_10'.
- Diseases (Drag):** A list of diseases including 'RETINITIS PIGMENTOSA 1' through '27'. 'RETINITIS PIGMENTOSA 14' is highlighted.
- Primary Disease (Drop):** A list containing 'RETINITIS PIGMENTOSA 10', '13', and '20'.
- Genes:** A list containing 'IMPDH1', 'PRPF8', and 'RPE65', each with a red minus sign.
- Mutations:** A table with columns 'Chr', 'Pos', 'Ref', 'Alt', and 'Gene'.
- Text/Bed File:** Radio buttons for 'Text' (selected) and 'Bed File'. A text field contains 'BRCA2,PPL'.
- Buttons:** 'Add Mutation', 'Add Genes', 'Add new panel', 'Clear', and 'Close'.

2. Diseases

3. Adding:
- more genes
- mutations

4. Save panel

TEAM

TEAM: Targeted Enrichment Analysis and Management

How to define a panel?

Add mutation

Chr: 8 Pos: 55539395 Ref: A Alt: T Gene Name: RP1 Disease Name: Lung cancer 2

Reset Check Add Mutation

Region overview Window size: 583 nts

55,539,104 55,539,395 55,539,686

Sequence AAGCACATAACTAAAATTGCCGGTTTGACAGGAGATAATCTATGTAAGAGGGAGATAAGTCTTT

Gene

SNP P_ESP_8_55539357 rs58051614 rs200135800 COSM486527
8_55539353 rs202016292 rs201613551 rs2293869 rs202057087
rs202226256

T 8:55,539,394 Genome Viewer

Adding
new mutations

Checking
mutations from
Genome Viewer

TEAM

TEAM: Targeted Enrichment Analysis and Management

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								
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	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

TEAM

TEAM: Targeted Enrichment Analysis and Management

Outline

- 1) Introduction
- 2) How does TEAM work?
- 3) Results
- 4) **Conclusions**
- 5) Examples

Conclusions

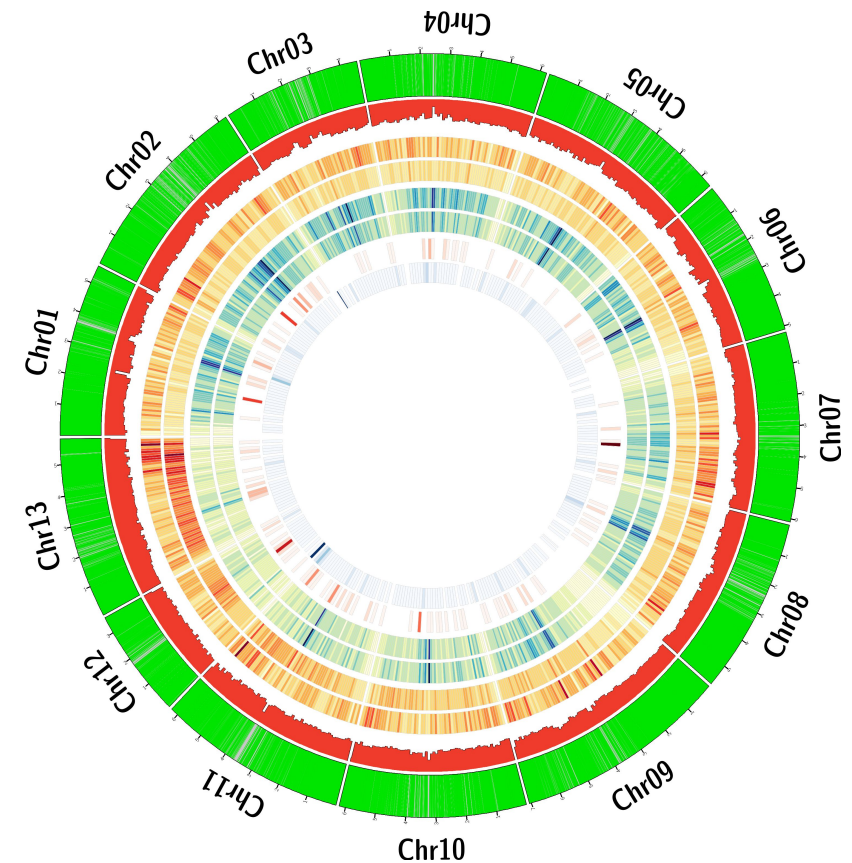
- TEAM is an **free and easy-to-use web tool** that fills the gap between the enormous amounts of data in targeted enrichment sequencing analysis and the **biological knowledge** available
- TEAM **provides an intuitive environment for the clinician** in which unprocessed data on patient's genomic variation 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**

TEAM

TEAM: Targeted Enrichment Analysis and Management

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-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



TEAM

TEAM: Targeted Enrichment Analysis and Management

Outline

- 1) Introduction
- 2) How does TEAM work?
- 3) Results
- 4) Conclusions
- 5) Examples

Examples

<http://team.babelomics.org>

TEAM

home documentation **tutorial** about

Show Panels

Example Data

Search

Panel:

VCF File:

Results

Chromosome	Position	SNP Id	Ref	Alt	Gene	Conseq. Type	Phenotype	Source	SIFT	PolyPhen	Conservation
------------	----------	--------	-----	-----	------	--------------	-----------	--------	------	----------	--------------

TEAM

TEAM: Targeted Enrichment Analysis and Management