Panel of genes: design and analysis for clinical applications. TEAM

Mercedes Medina

September 29th, 2016

GDA

International Course on

Genomic Data Analysis

PRINCIPE FELIPE

CENTRO DE INVESTIGACION

Introduction

- **Development of high throughput sequencing technologies**:
  - Fast 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)

Introduction

Sequencing data

Biological knowledge

GWAS

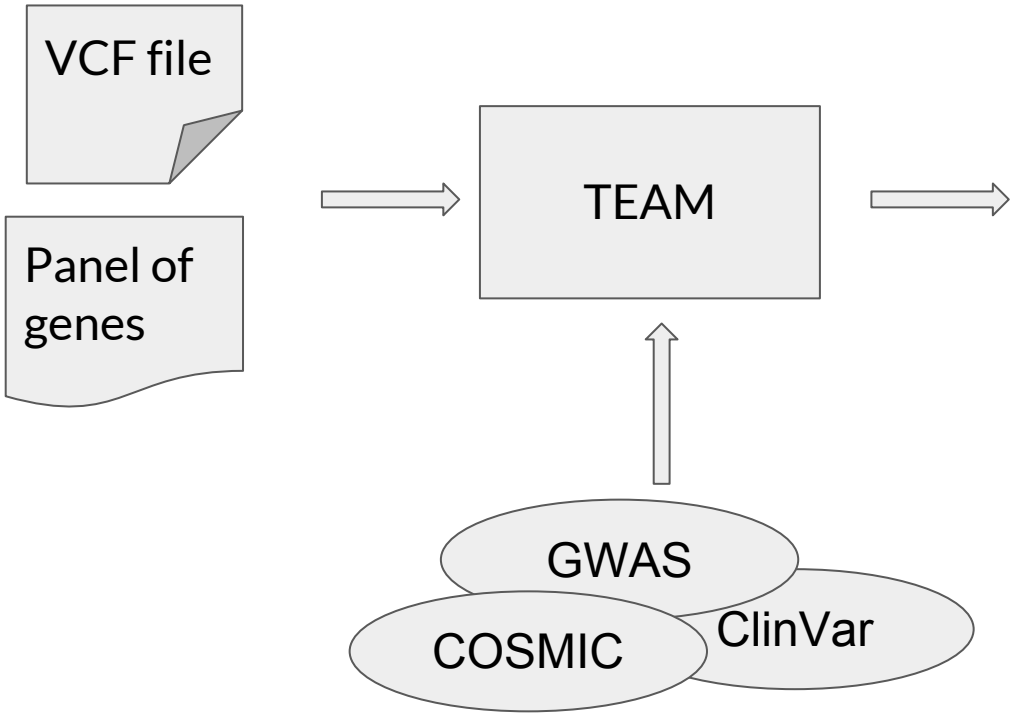
COSMIC

ClinVar

TEAM

Diagnostic

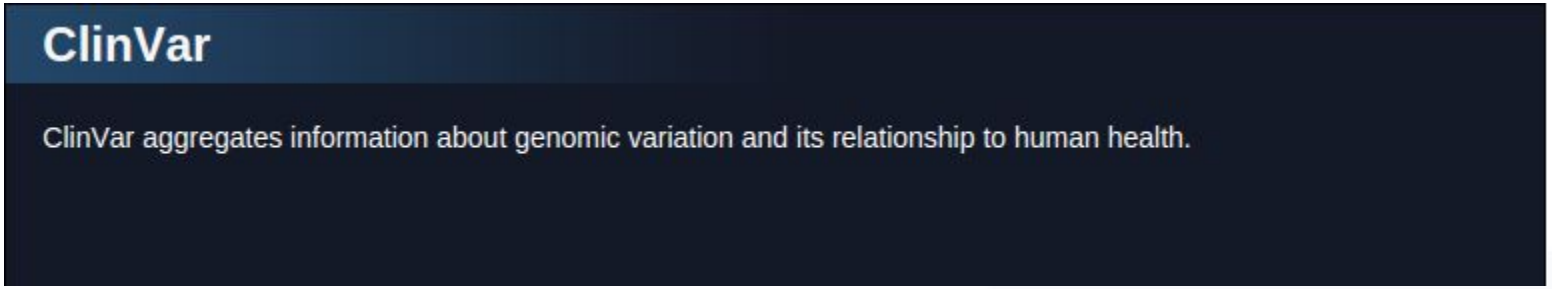
How does TEAM work?



Diagnostic

Chr	Pos	Ref	Alt	SNP Id
9	135786112	T	C	rs6597586
9	135786904	A	G	rs1073123
19	55453134	G	A	rs775886

Getting information



COSMIC
Catalogue of somatic mutations in cancer

GWAS Catalog

The NHGRI-EBI Catalog of published genome-wide association studies

Getting information: SIFT & PolyPhen

- **SIFT**
  - SIFT predicts whether an amino acid substitution affects protein function
  - **Interpretation:** 1 (tolerated) to 0 (deleterious)

http://sift.jcvi.org/

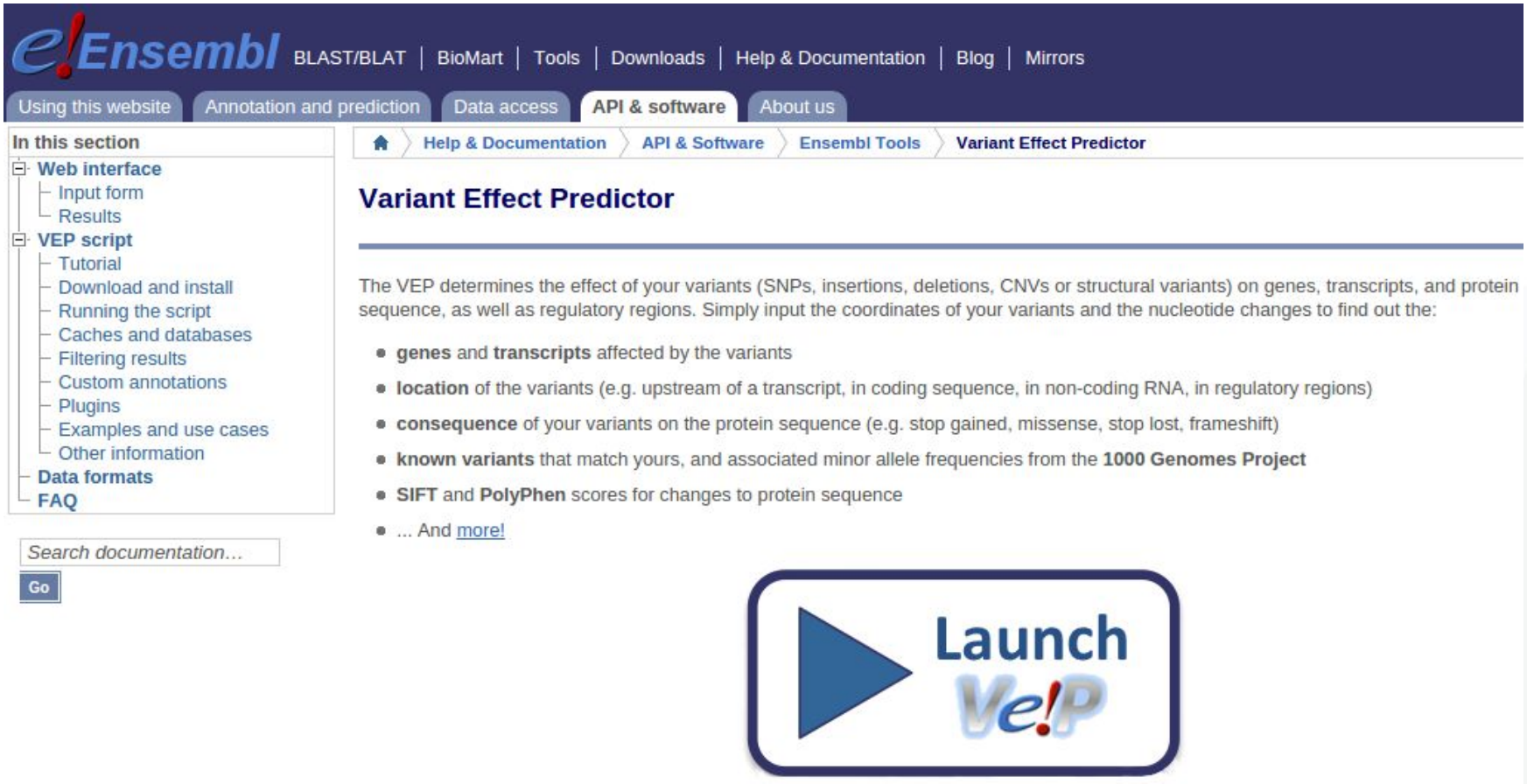
- **PolyPhen**
  - Polymorphism Phenotyping is a tool which predicts possible impact of an amino acid substitution on the structure and function of a human protein.
  - **Interpretation:** 1 (probably damage) to 0 (bening)

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

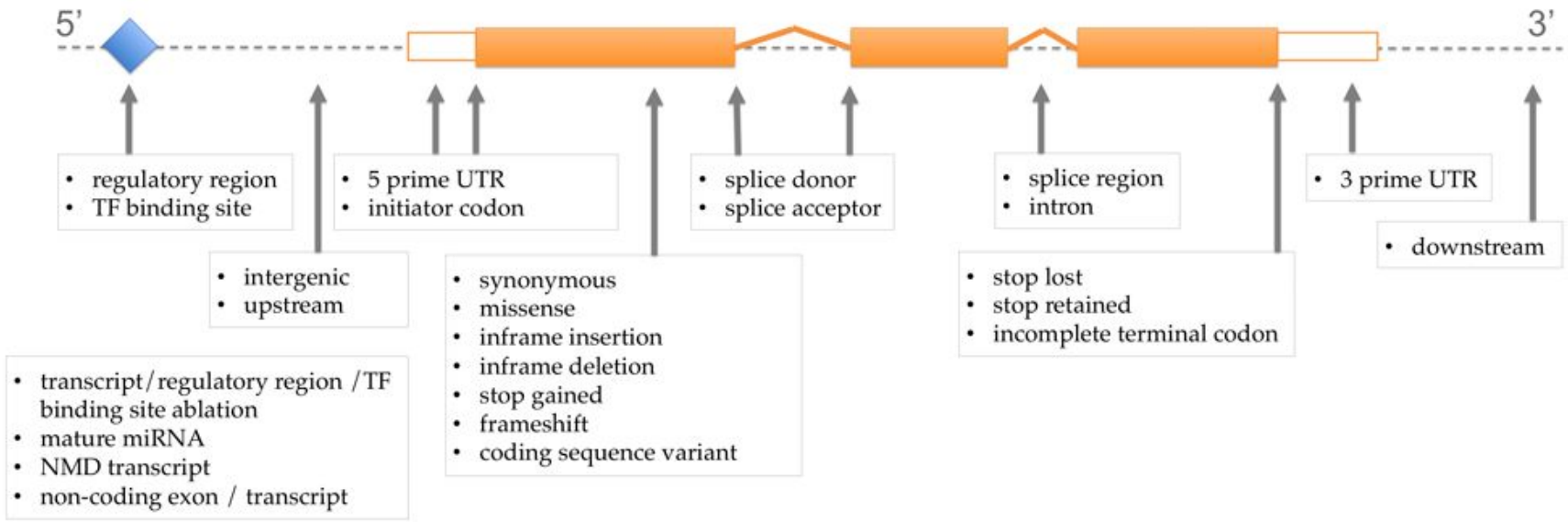
Getting information: Conservation

- **Phylop**
  - PhyloP scores measure evolutionary conservation at individual alignment sites. The scores are interpreted as follows compared to the evolution expected under neutral drift:
    - Positive scores -- Measure conservation, which is slower evolution than expected, at sites that are predicted to be conserved.
    - Negative scores -- Measure acceleration, which is faster evolution than expected, at sites that are predicted to be fast-evolving.
- **PhastCons**
  - PhastCons is a program for identifying evolutionarily conserved elements in a multiple alignment, given a phylogenetic tree.
  - PhastCons essentially does three things:
    - It produces base-by-base conservation scores (as displayed in the conservation tracks in the UCSC browser)
    - It produces predictions of discrete conserved elements (as displayed in the "most conserved" tracks in the browser)
    - It estimates free parameters.

Getting information: Effect



Getting information: Effect



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

Inputs: Panel + VCF

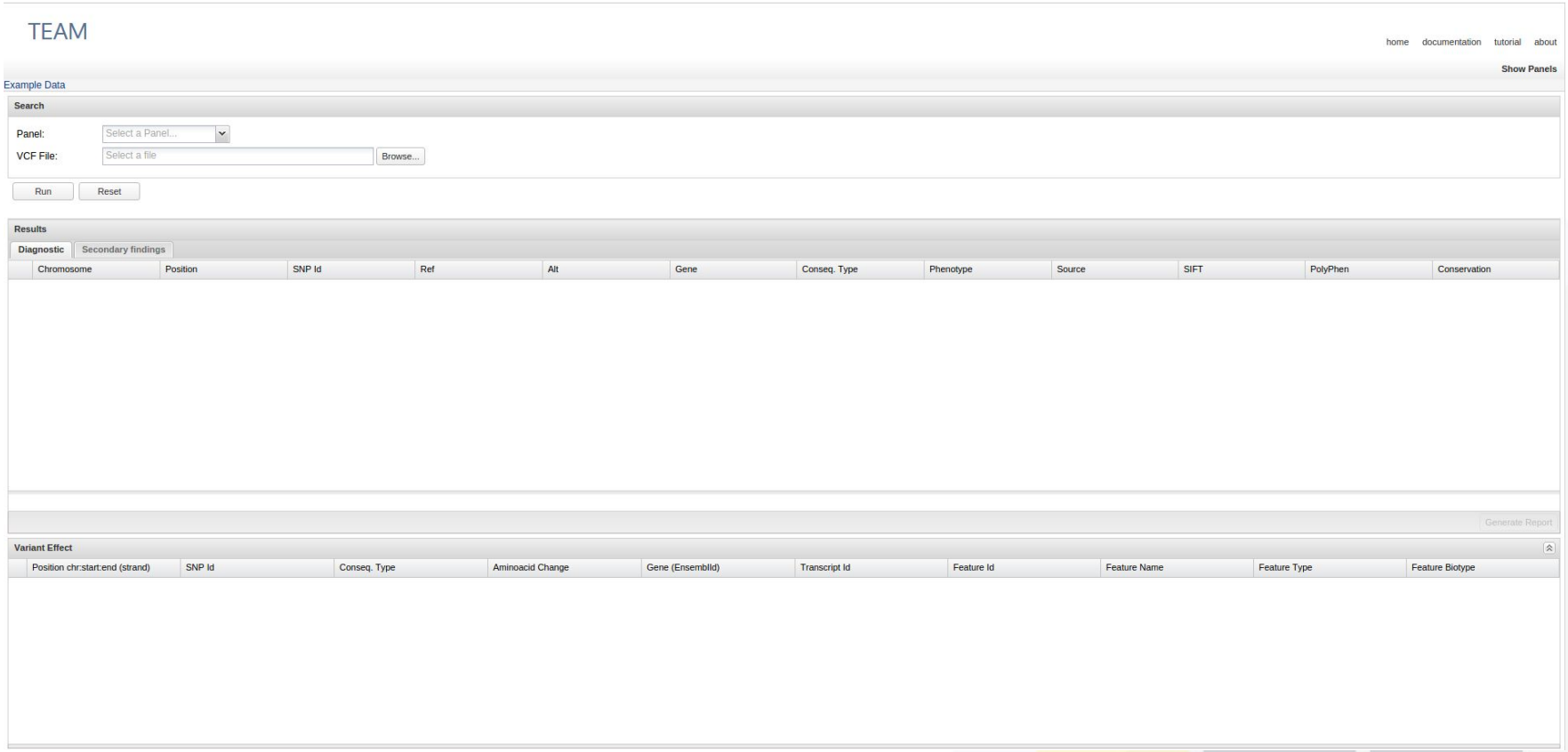
TEAM needs a:

1. **VCF file**:
   - The VCF file needs to be stored/indexed in our database.
   - This file could be already added if we used BiERapp (both tools are compatible)

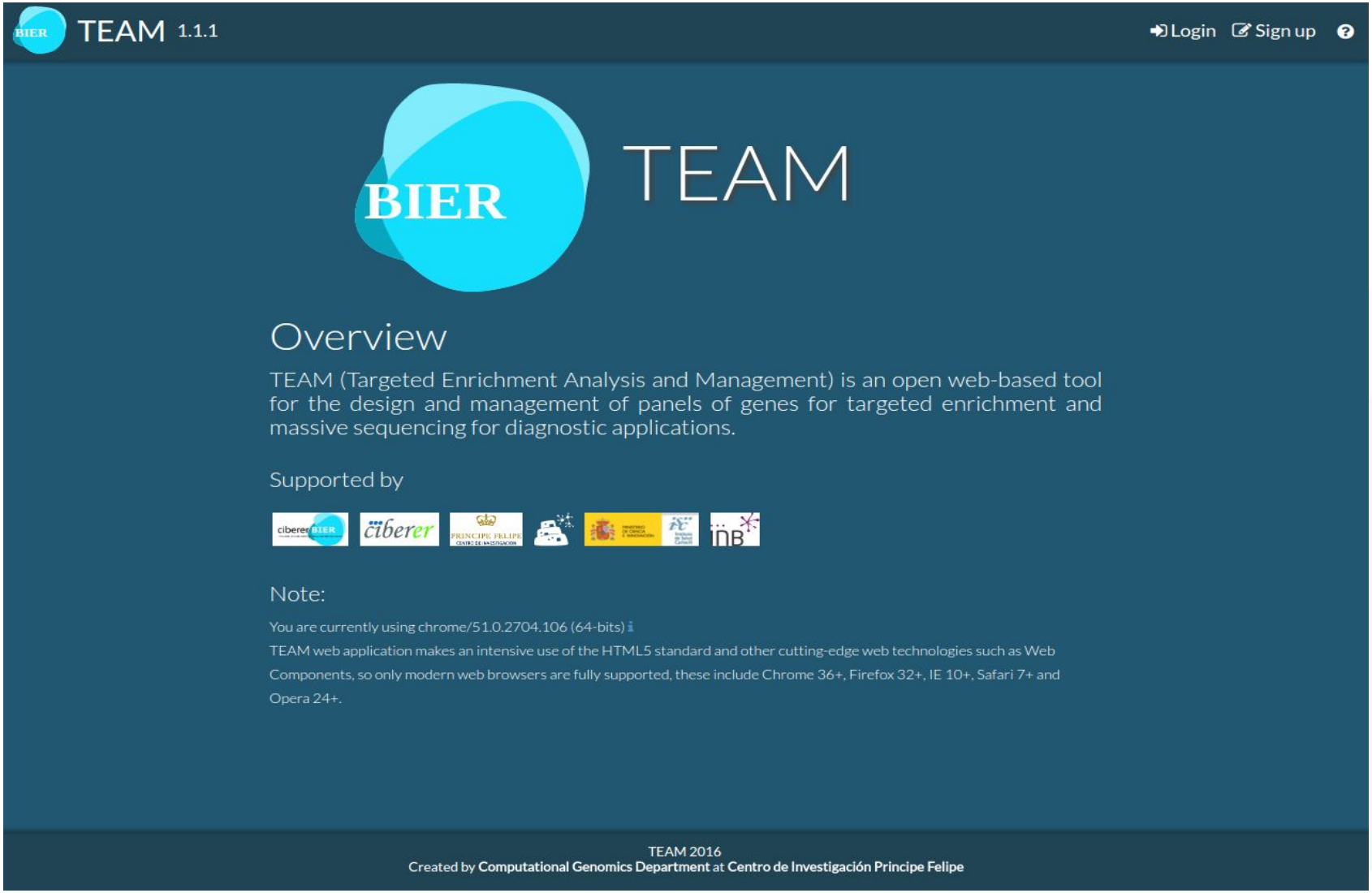
2. **Panel of genes**:
   - TEAM works with virtual panel of genes.
   - You can design/create/manage these panels using TEAM.
   - These panels contains:
     - Phenotypes
     - Genes
     - Mutations

Tool interface: Official release

http://team.babelomics.org/



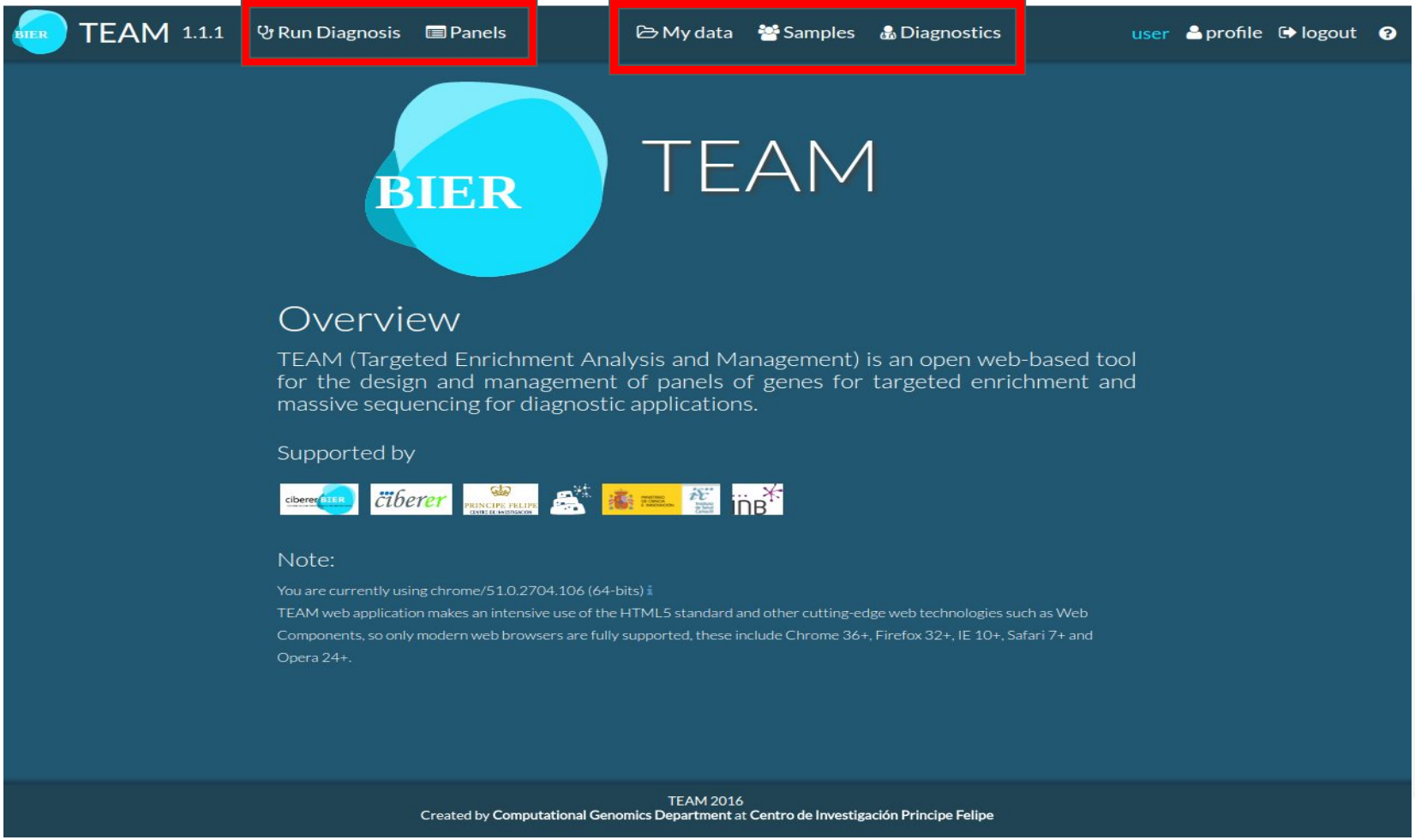
Tool interface: Beta



Tool interface: Beta

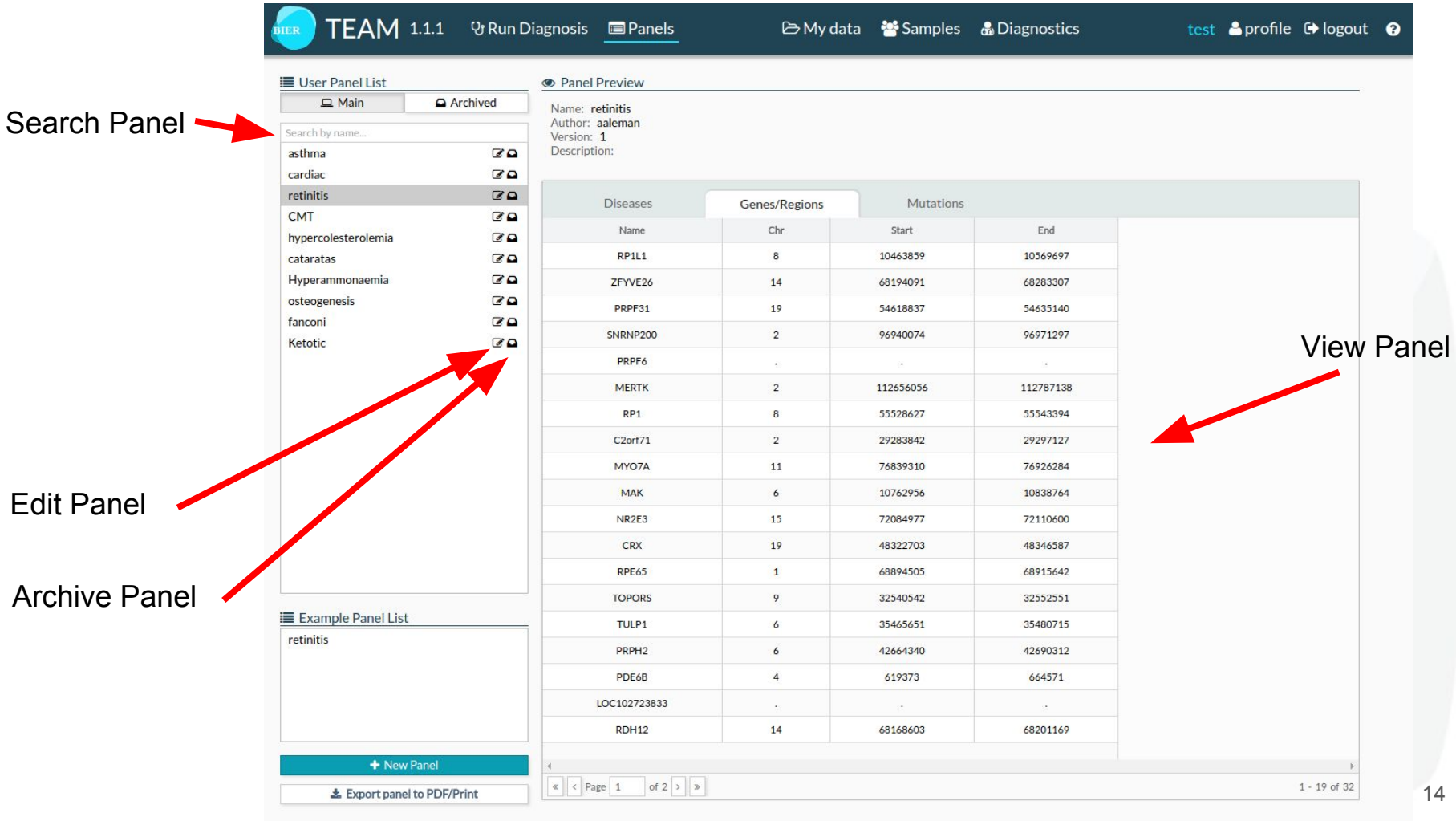
Design/Manage Panels

Data & Diagnoses

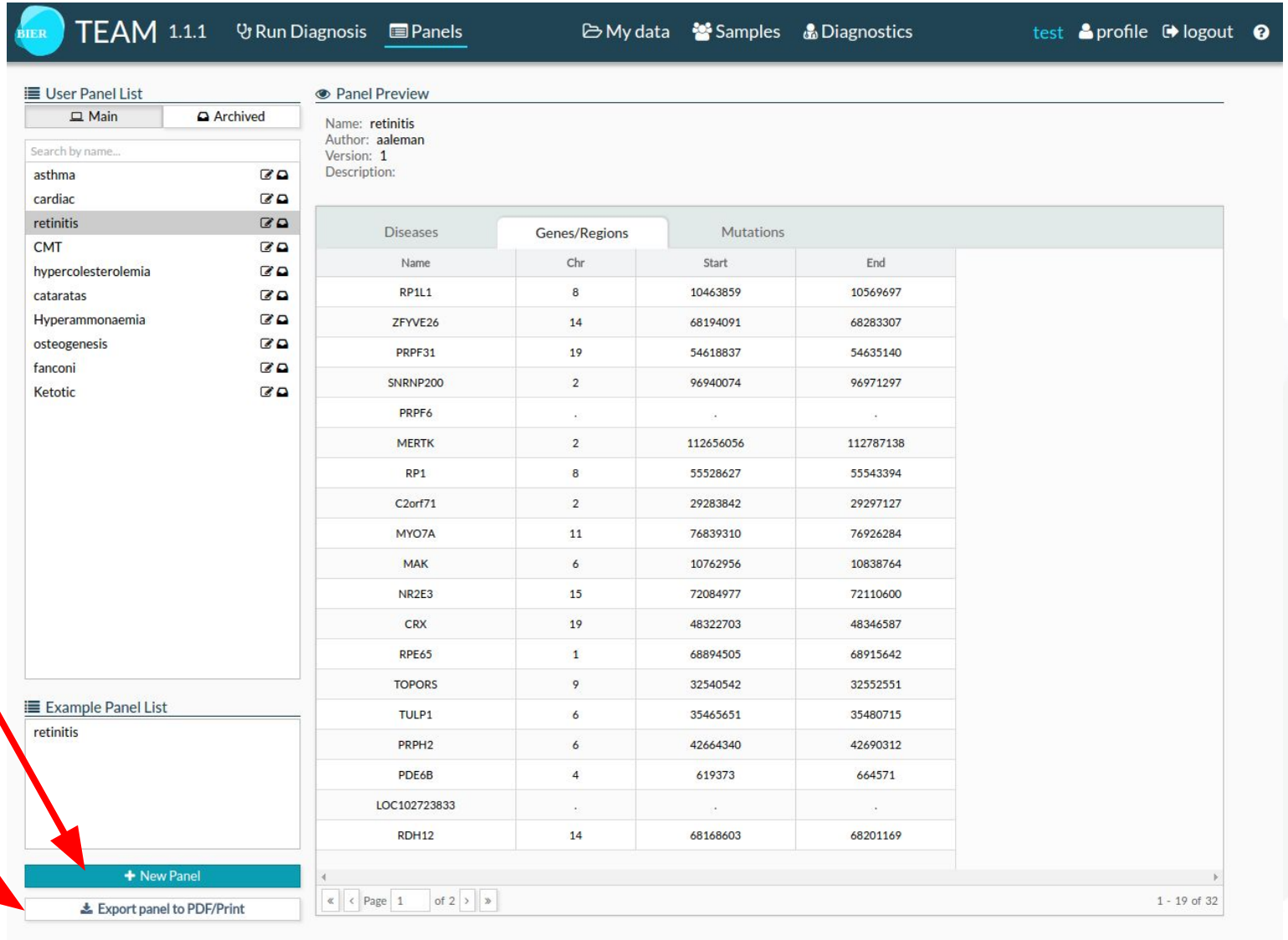


Panel Manager

You can manage your panels using “Panels”



Panel Manager: Create a Panel

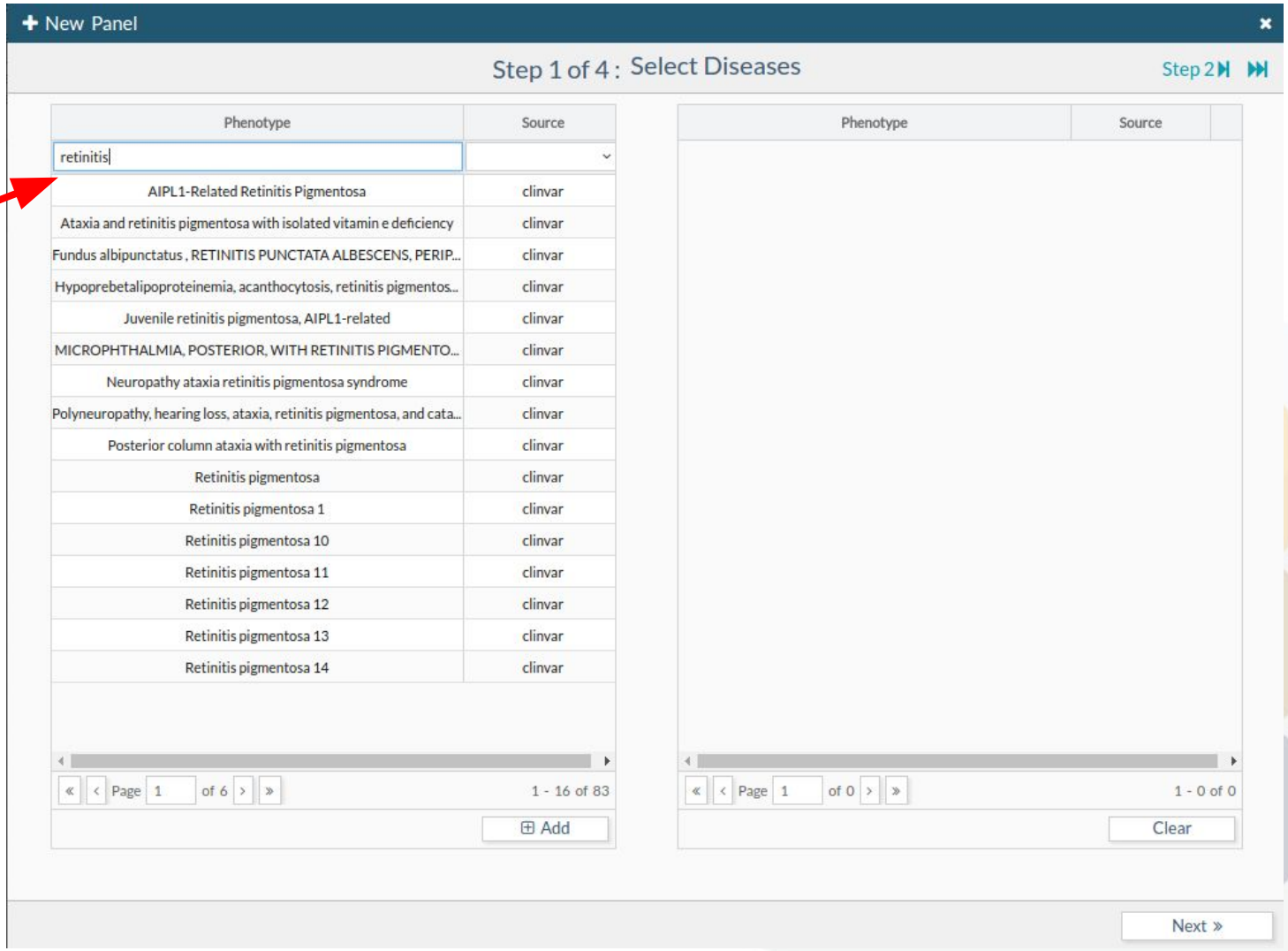


Create a New Panel

Export or Print a Panel

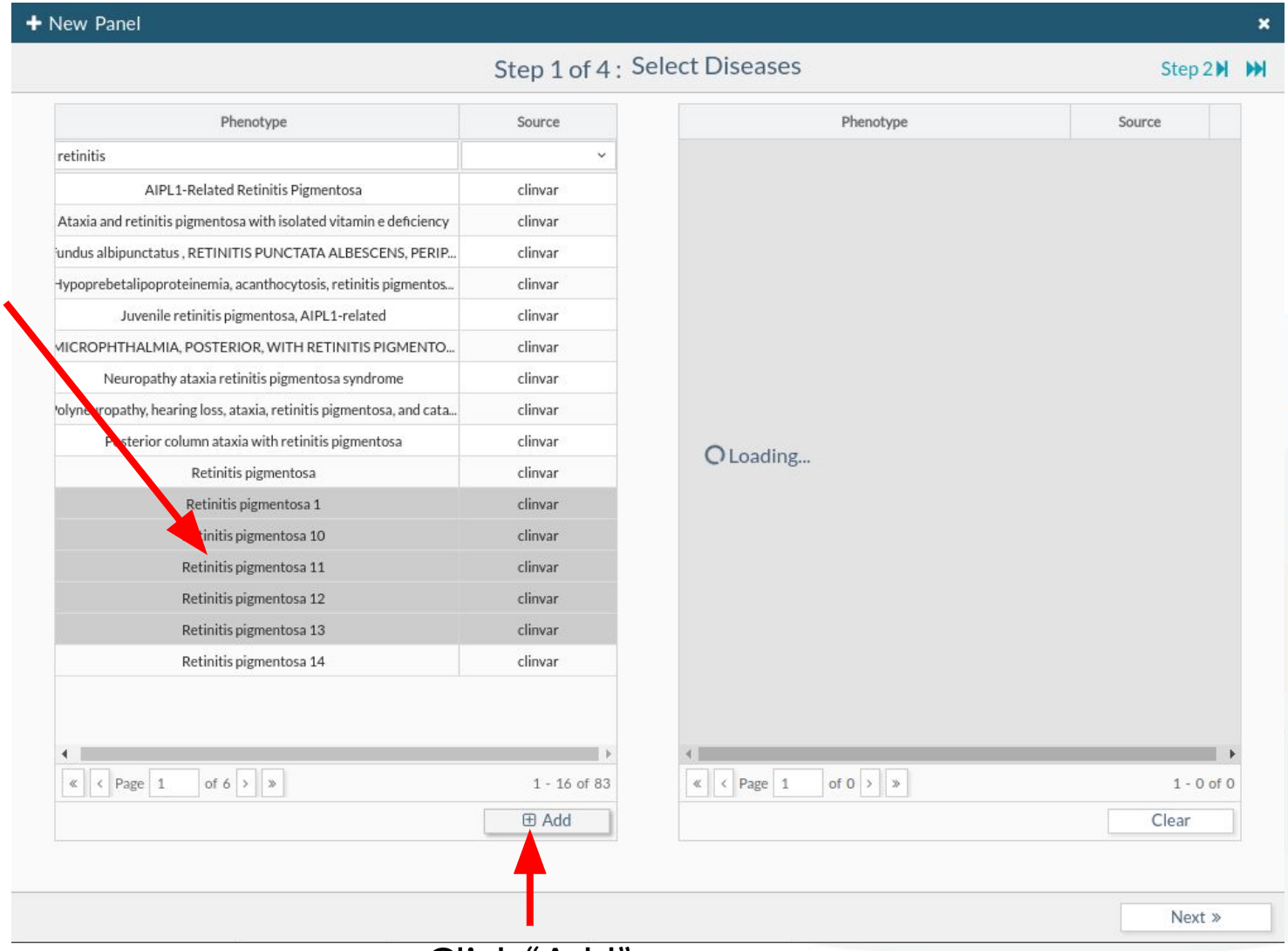
Panel Designer: Diseases

Write the phenotypes you are interested in.



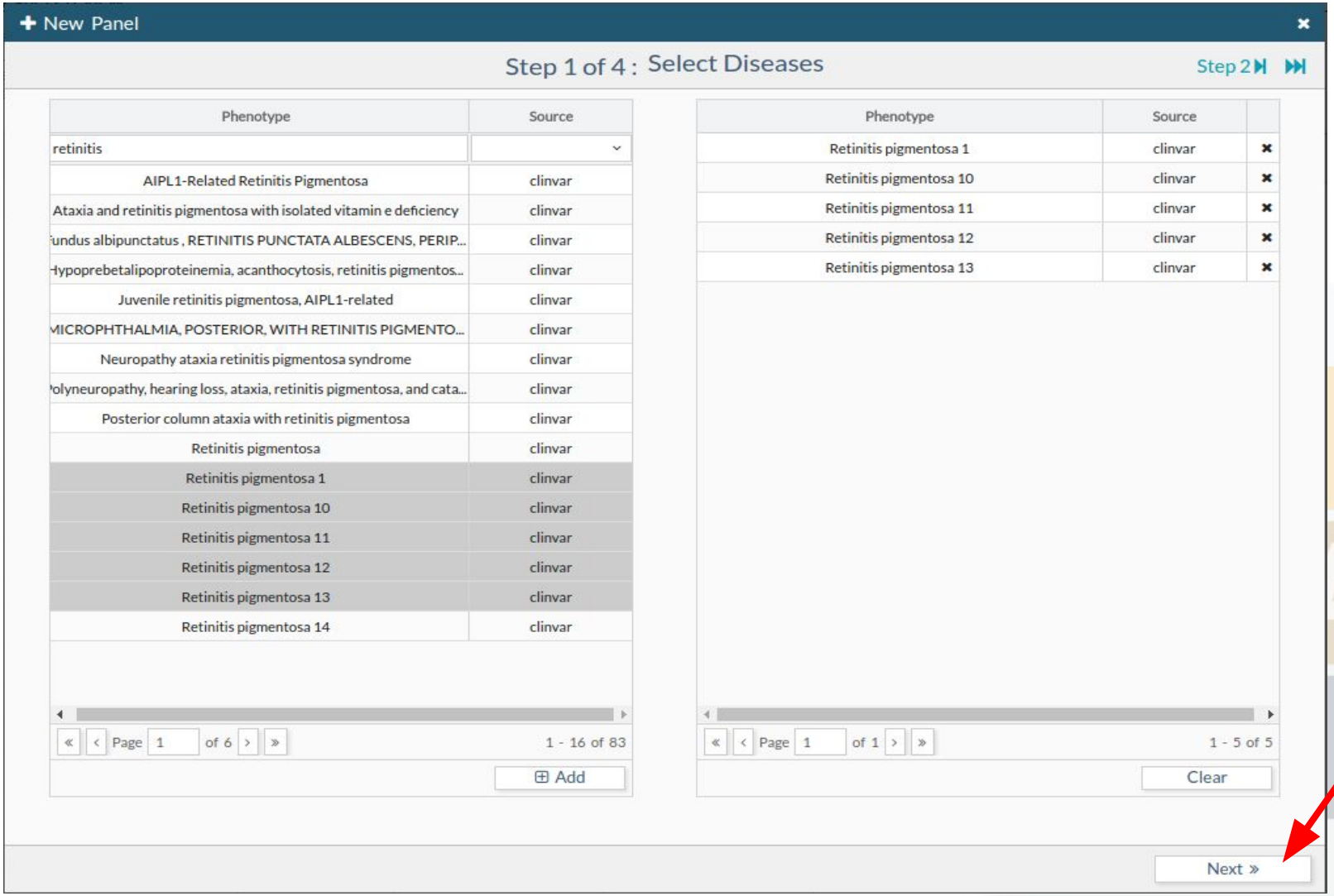
Panel Designer: Diseases

Select the phenotypes you want to add to the virtual panel. That will add the associated genes and mutations



Click “Add”

Panel Designer: Diseases

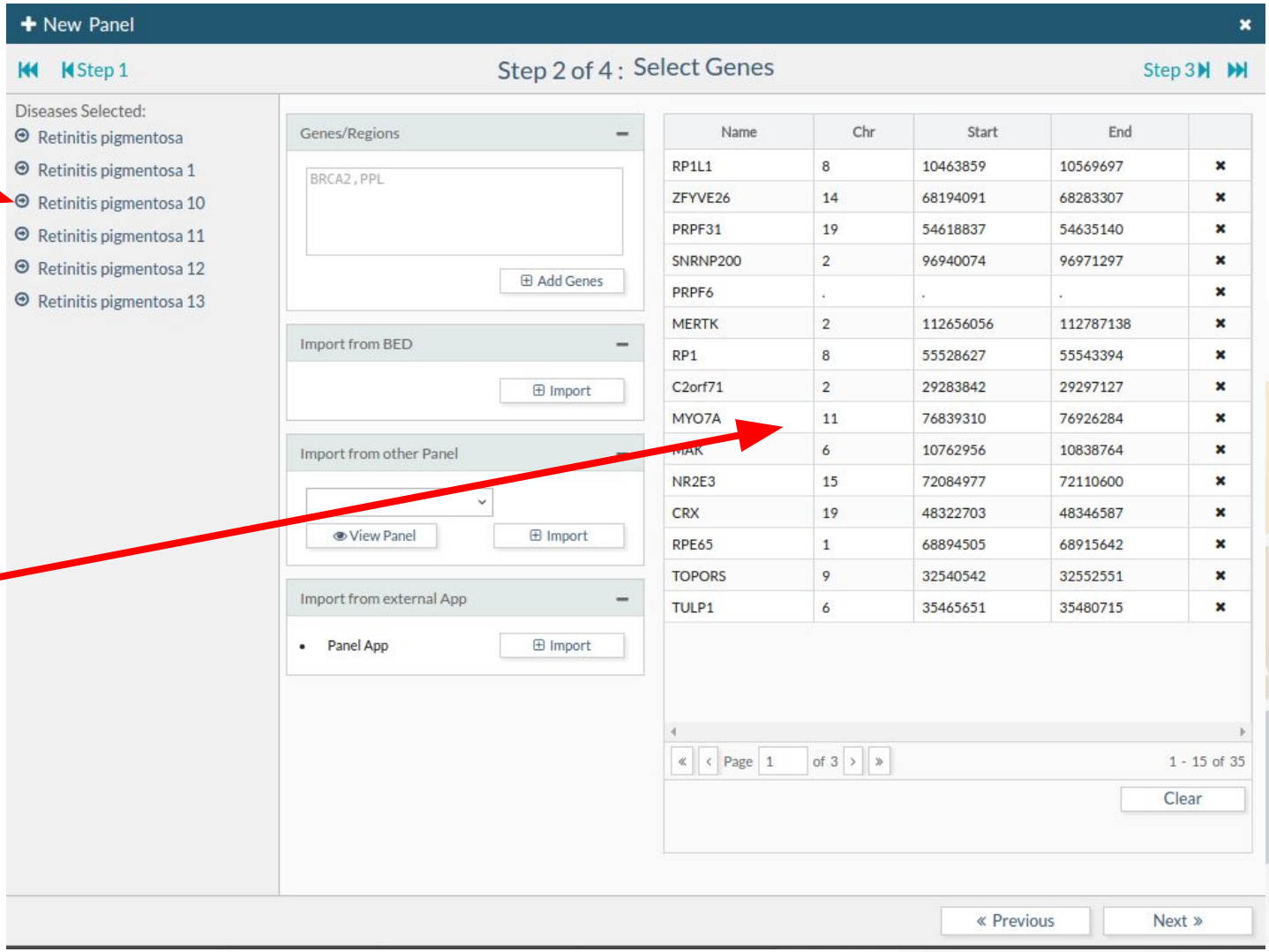


Next Step

Panel Designer: Genes

Selected Diseases in previous step.

These genes have been added by previous step.
These genes are related to the selected diseases

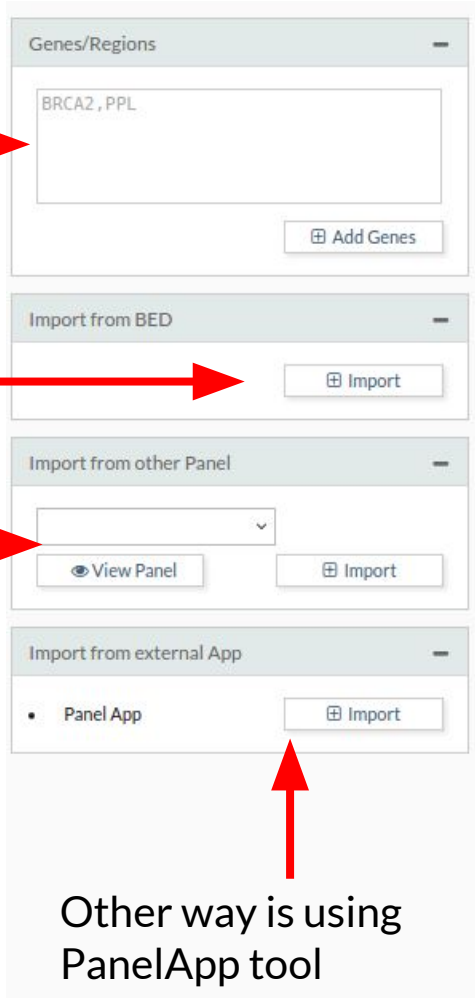


Panel Designer: Genes

We can add new genes typing the name or the region.

Adding regions through a BED file is also supported

We can import genes from other virtual panels already created.



Name	Chr	Start	End	
RP1L1	8	10463859	10569697	✖
ZFYVE26	14	68194091	68283307	✖
PRPF31	19	54618837	54635140	✖
SNRNP200	2	96940074	96971297	✖
PRPF6	.	.	.	✖
MERTK	2	112656056	112787138	✖
RP1	8	55528627	55543394	✖
C2orf71	2	29283842	29297127	✖
MYO7A	11	76839310	76926284	✖
MAK	6	10762956	10838764	✖
NR2E3	15	72084977	72110600	✖
CRX	19	48322703	48346587	✖
RPE65	1	68894505	68915642	✖
TOPORS	9	32540542	32552551	✖
TULP1	6	35465651	35480715	✖

« ‹ Page 1 of 3 › » 1 - 15 of 35

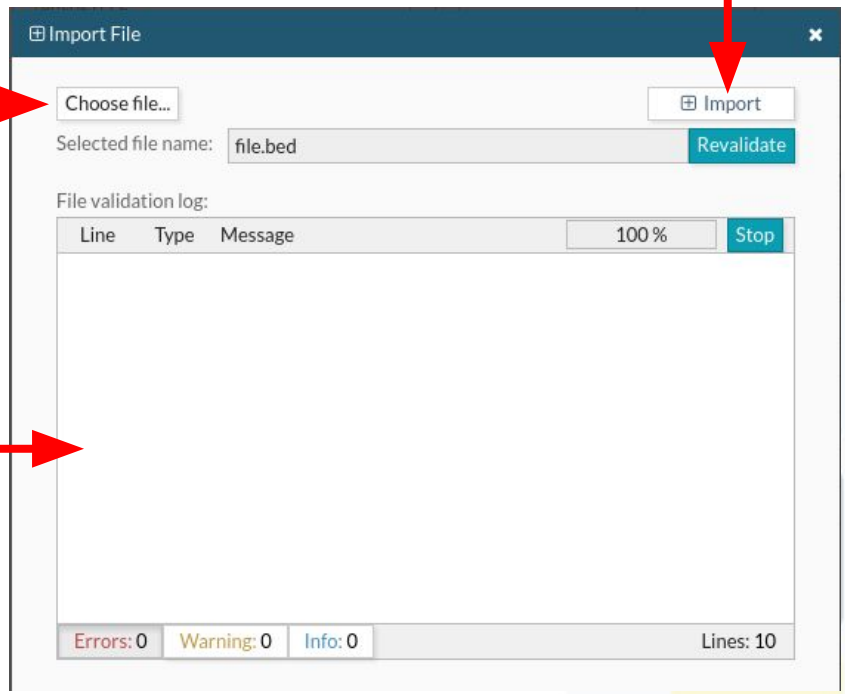
Clear

Other way is using PanelApp tool

Panel Designer: Genes (BED file)

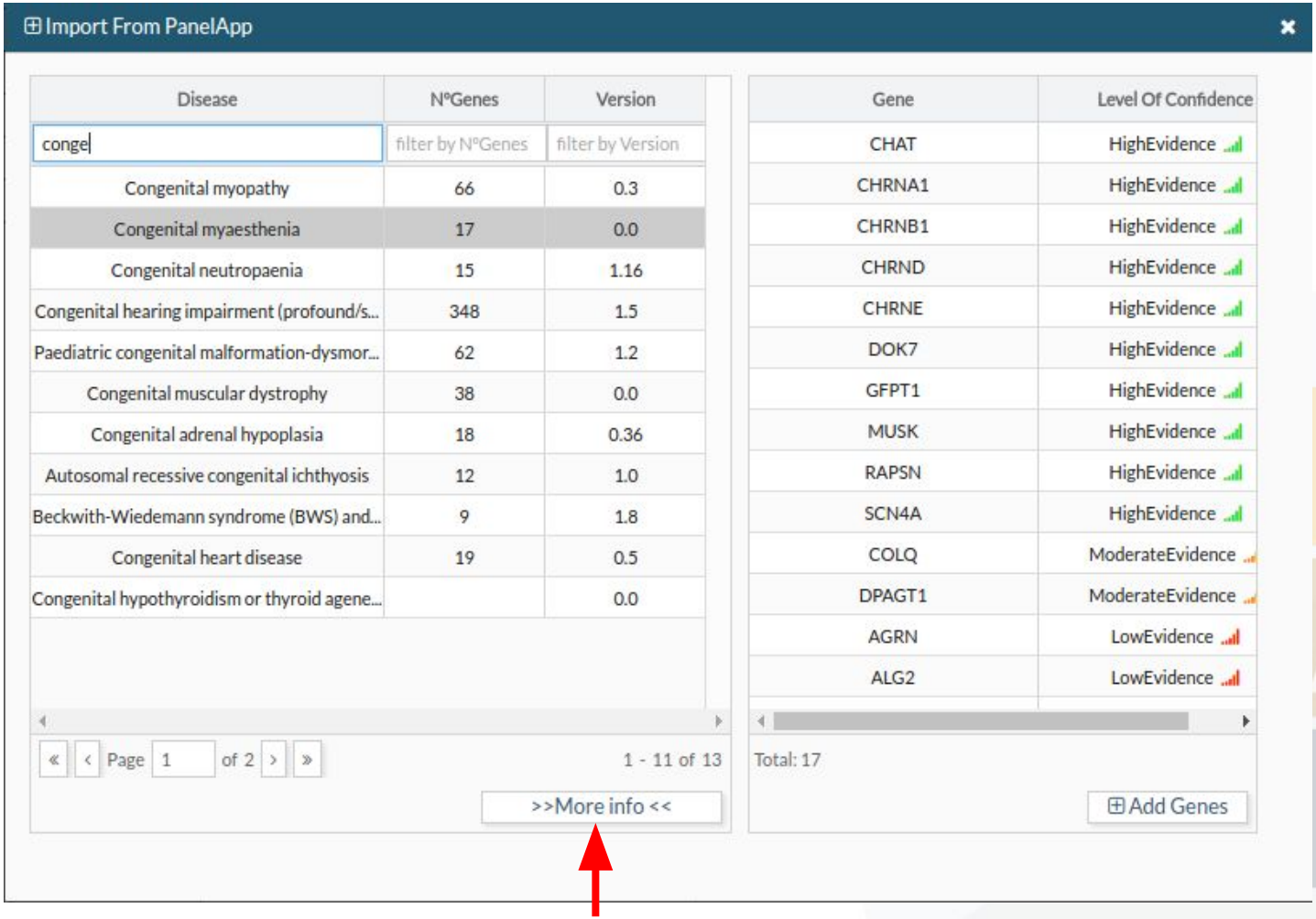
Import genes

Choose a BED file



Check errors or warnings

Panel Designer: Genes (PanelApp)



https://bioinfo.extge.co.uk/crowdsourcing/PanelApp/

Panel Designer: Genes (PanelApp)

Genomics England PanelApp

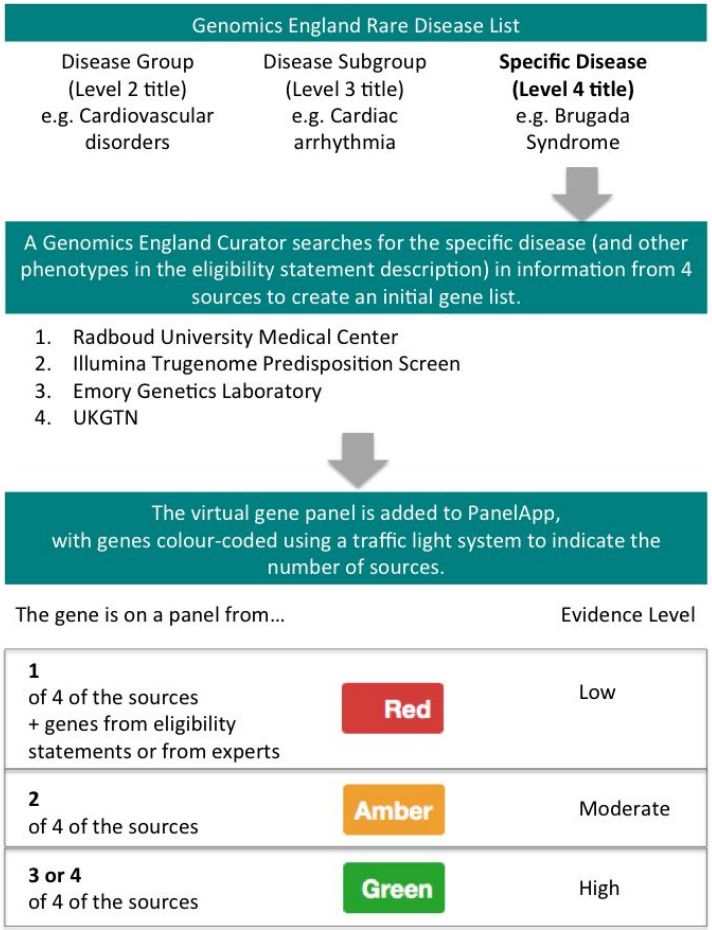
A crowdsourcing tool to allow gene panels to be shared, downloaded, viewed and evaluated by the Scientific Community

Reviews are assessed by Genomics England Curators to establish a final virtual gene panel

Each gene is either…	Evidence Level
Green	High, diagnostic-grade gene that will be used to Interpret genomes.
Red	Low/moderate evidence, genes that currently cannot be used to report clinically, more evidence may arise in the future.

https://bioinfo.extge.co.uk/crowdsourcing/PanelApp/

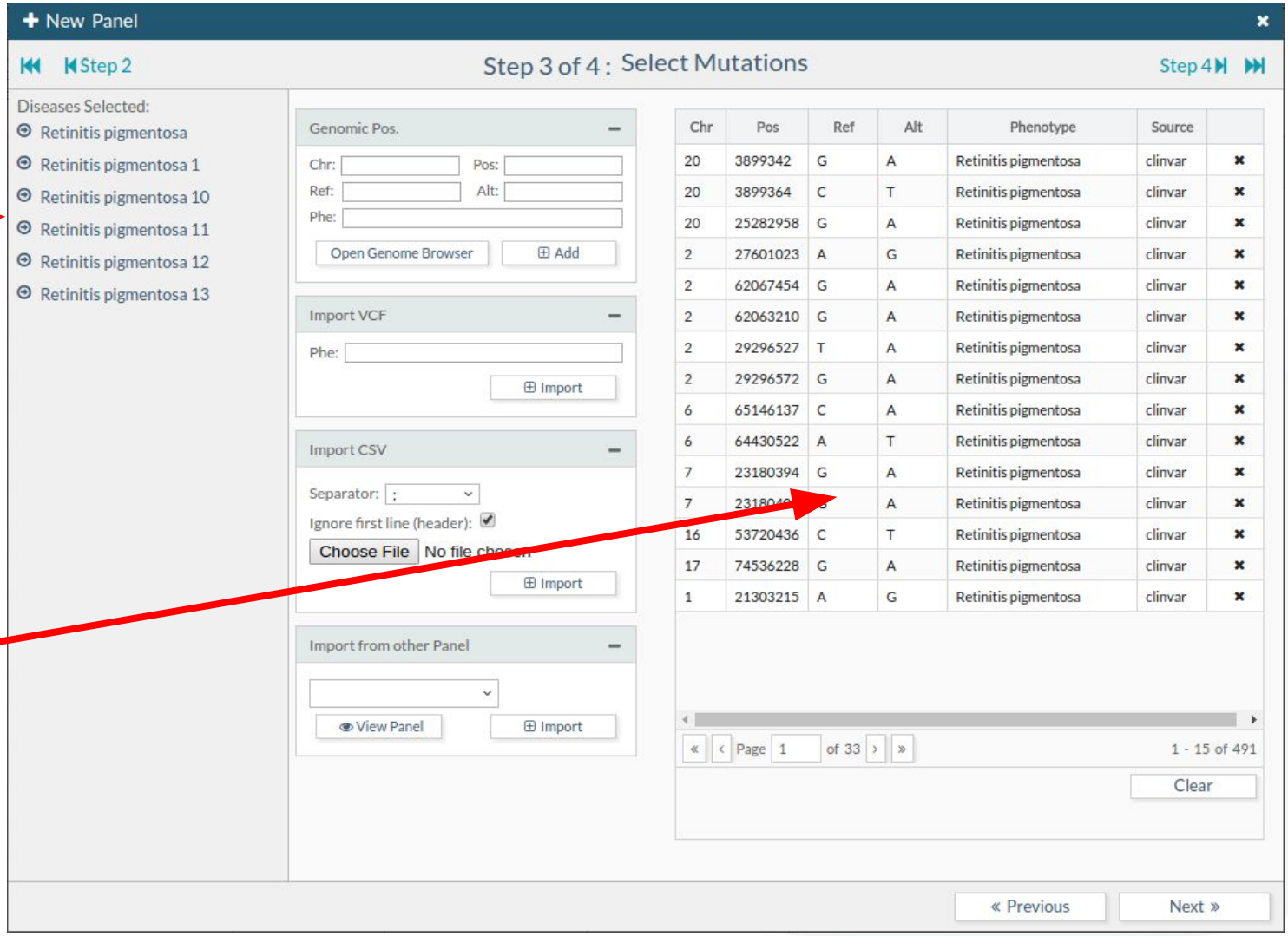
Figure 1: The curation process for the initial establishment of gene panels for each rare disease category



Panel Designer: Mutations

Selected Diseases in previous step.

These mutations have been added by the first step. They are related to the selected diseases

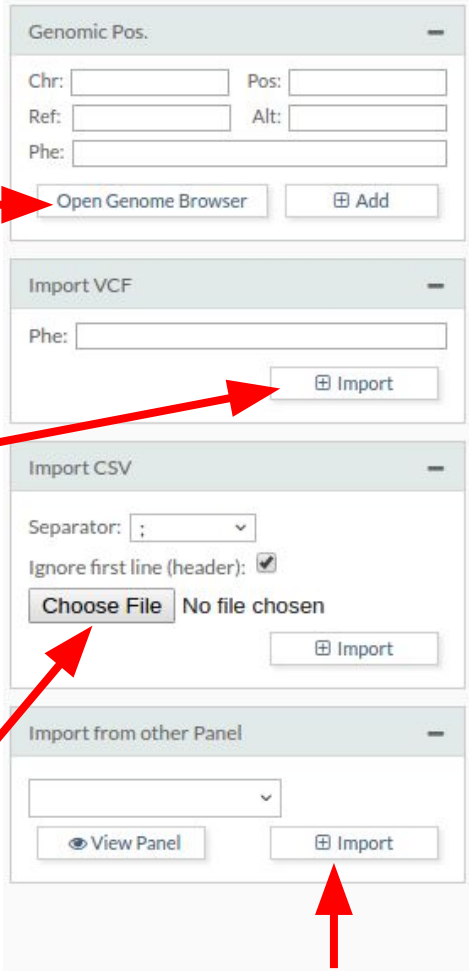


Panel Designer: Mutations

We add our custom mutations by writing the genomic position and the change (ref-alt).
Or we can use the Genome Browser to find the specific position

You can import mutations from a VCF file

Or you can import them from a CSV file with the next format:
CHR POS REF ALT PHE



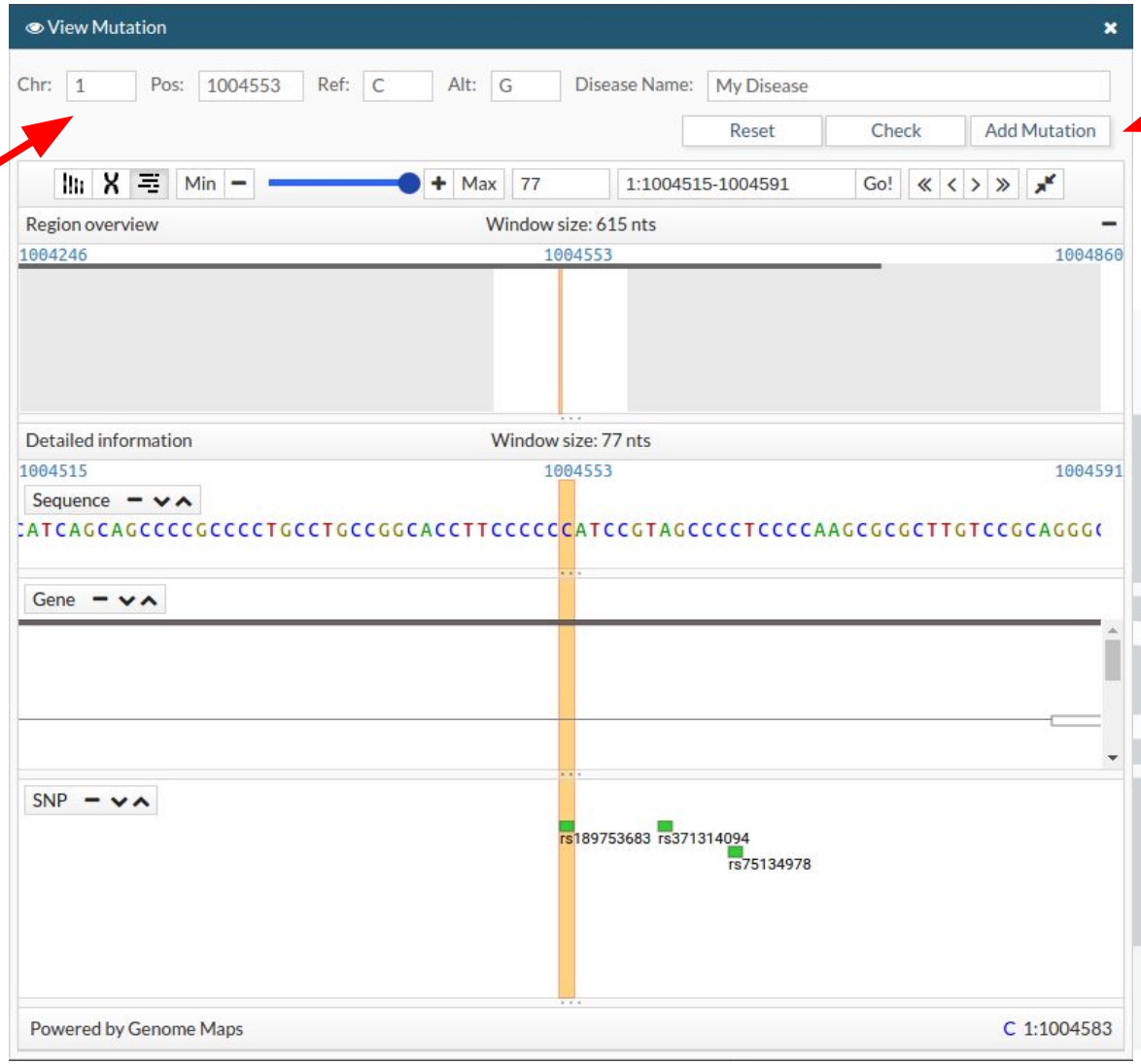
Chr	Pos	Ref	Alt	Phenotype	Source	
20	3899342	G	A	Retinitis pigmentosa	clinvar	✖
20	3899364	C	T	Retinitis pigmentosa	clinvar	✖
20	25282958	G	A	Retinitis pigmentosa	clinvar	✖
2	27601023	A	G	Retinitis pigmentosa	clinvar	✖
2	62067454	G	A	Retinitis pigmentosa	clinvar	✖
2	62063210	G	A	Retinitis pigmentosa	clinvar	✖
2	29296527	T	A	Retinitis pigmentosa	clinvar	✖
2	29296572	G	A	Retinitis pigmentosa	clinvar	✖
6	65146137	C	A	Retinitis pigmentosa	clinvar	✖
6	64430522	A	T	Retinitis pigmentosa	clinvar	✖
7	23180394	G	A	Retinitis pigmentosa	clinvar	✖
7	23180402	G	A	Retinitis pigmentosa	clinvar	✖
16	53720436	C	T	Retinitis pigmentosa	clinvar	✖
17	74536228	G	A	Retinitis pigmentosa	clinvar	✖
1	21303215	A	G	Retinitis pigmentosa	clinvar	✖

« ‹ Page 1 of 33 › » 1 - 15 of 491

Clear

Importing mutations from other virtual panels is supported too.

Panel Designer: Mutations



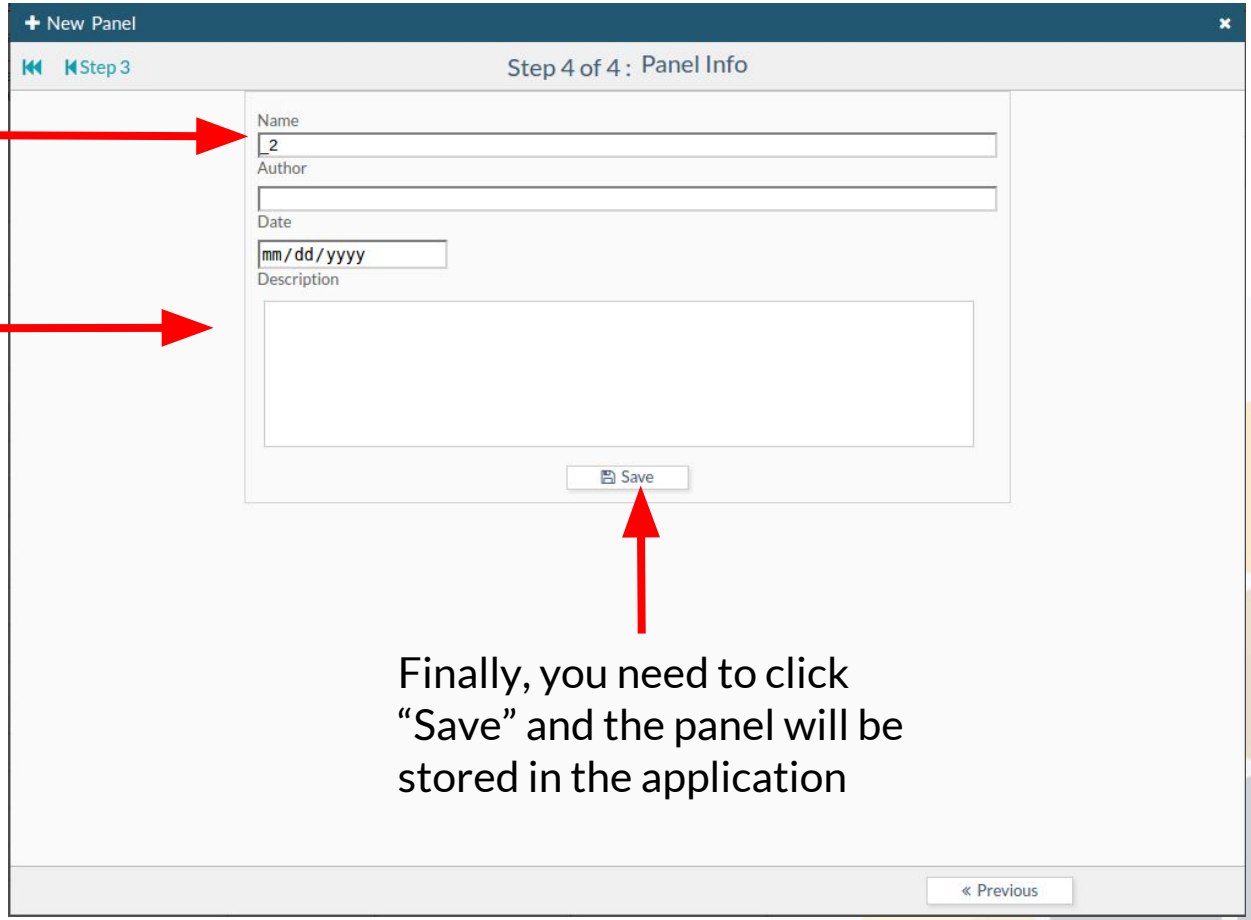
Import mutation

Search mutation

Panel Designer: Panel info

The name of the Panel

The author/ date/ description

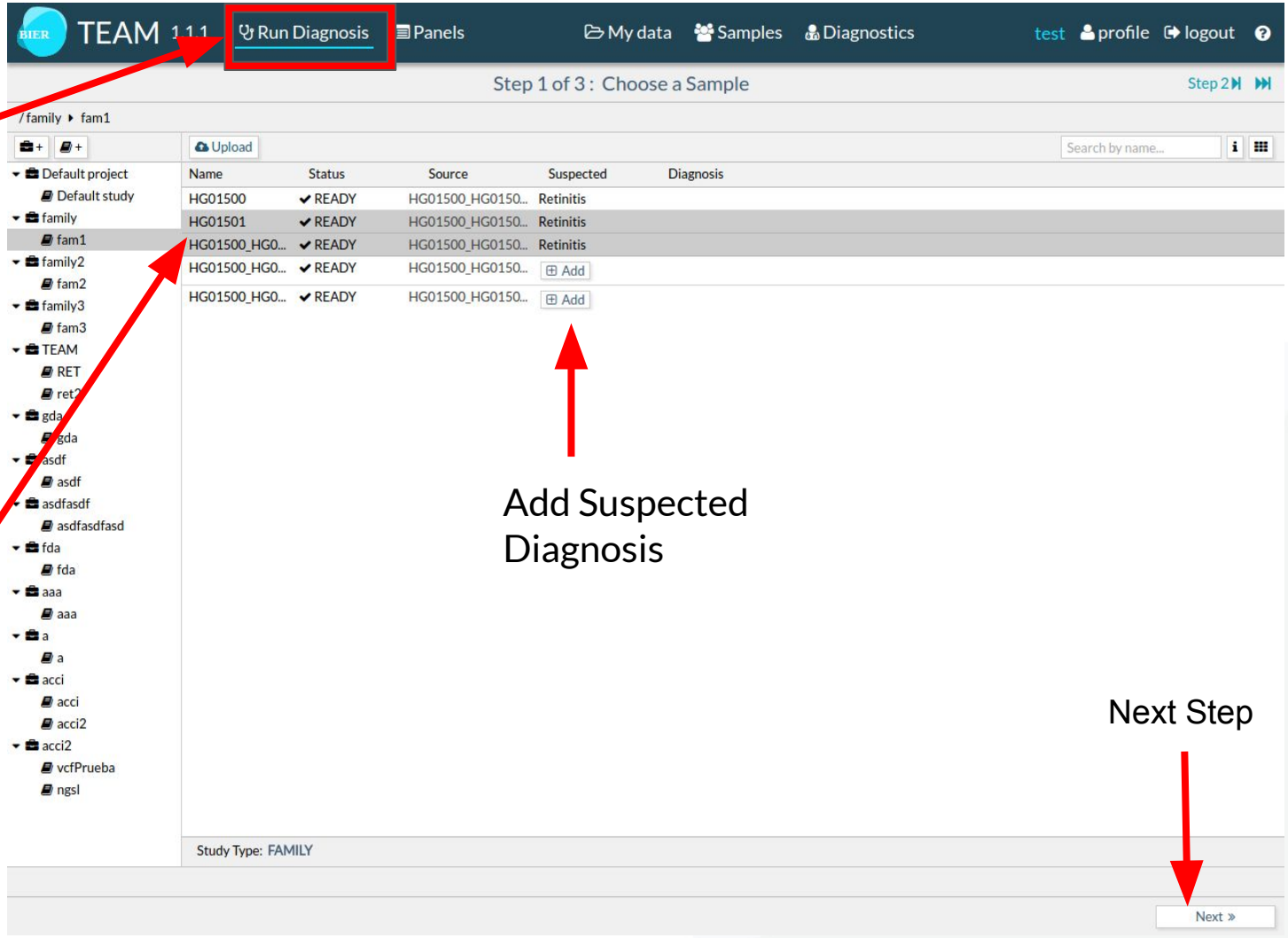


Finally, you need to click "Save" and the panel will be stored in the application

New Diagnosis

Click “Run Diagnosis”

Choose a Study. And then a sample/s in **Ready** status and with **Suspected diagnosis** added.

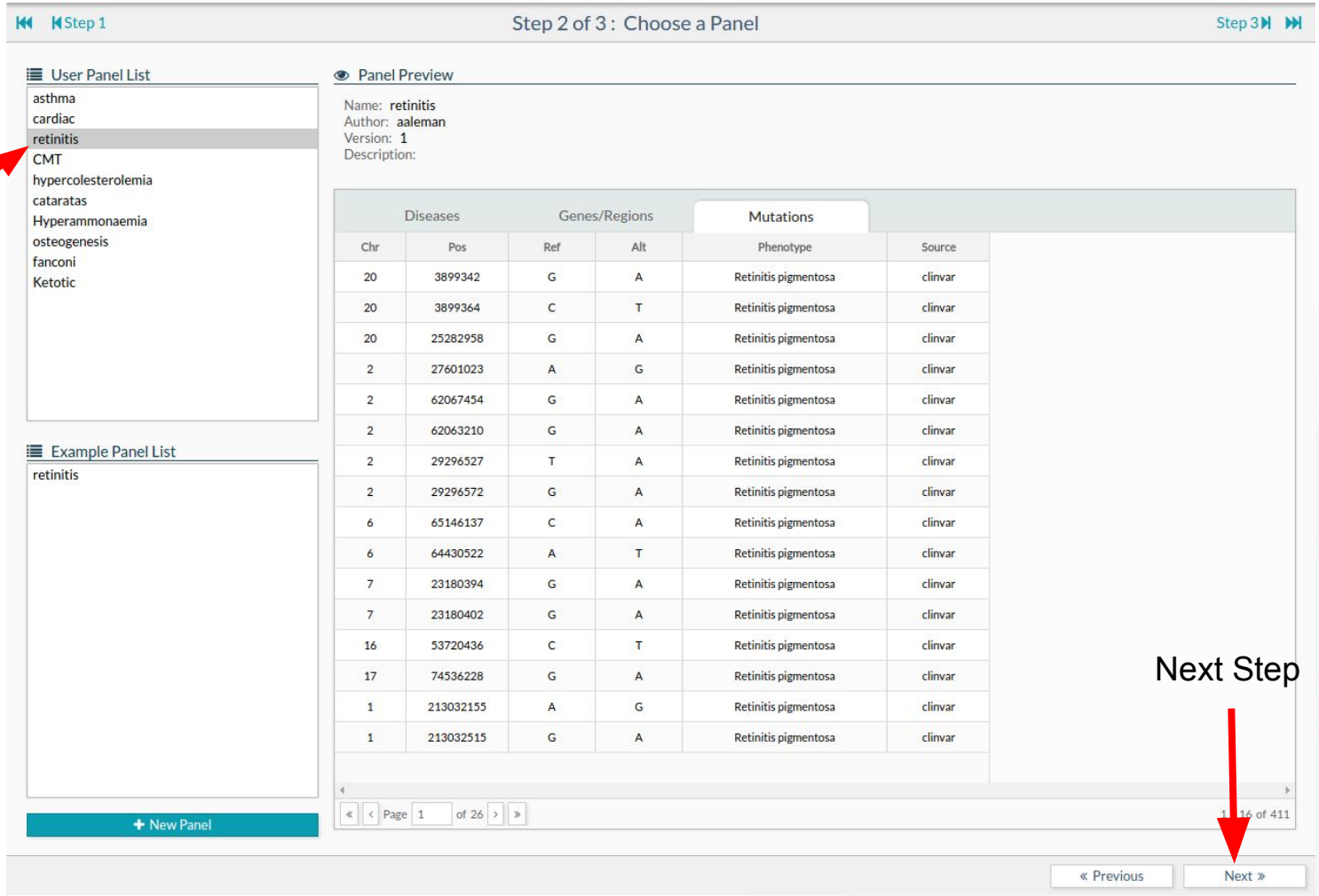


Add Suspected Diagnosis

Next Step

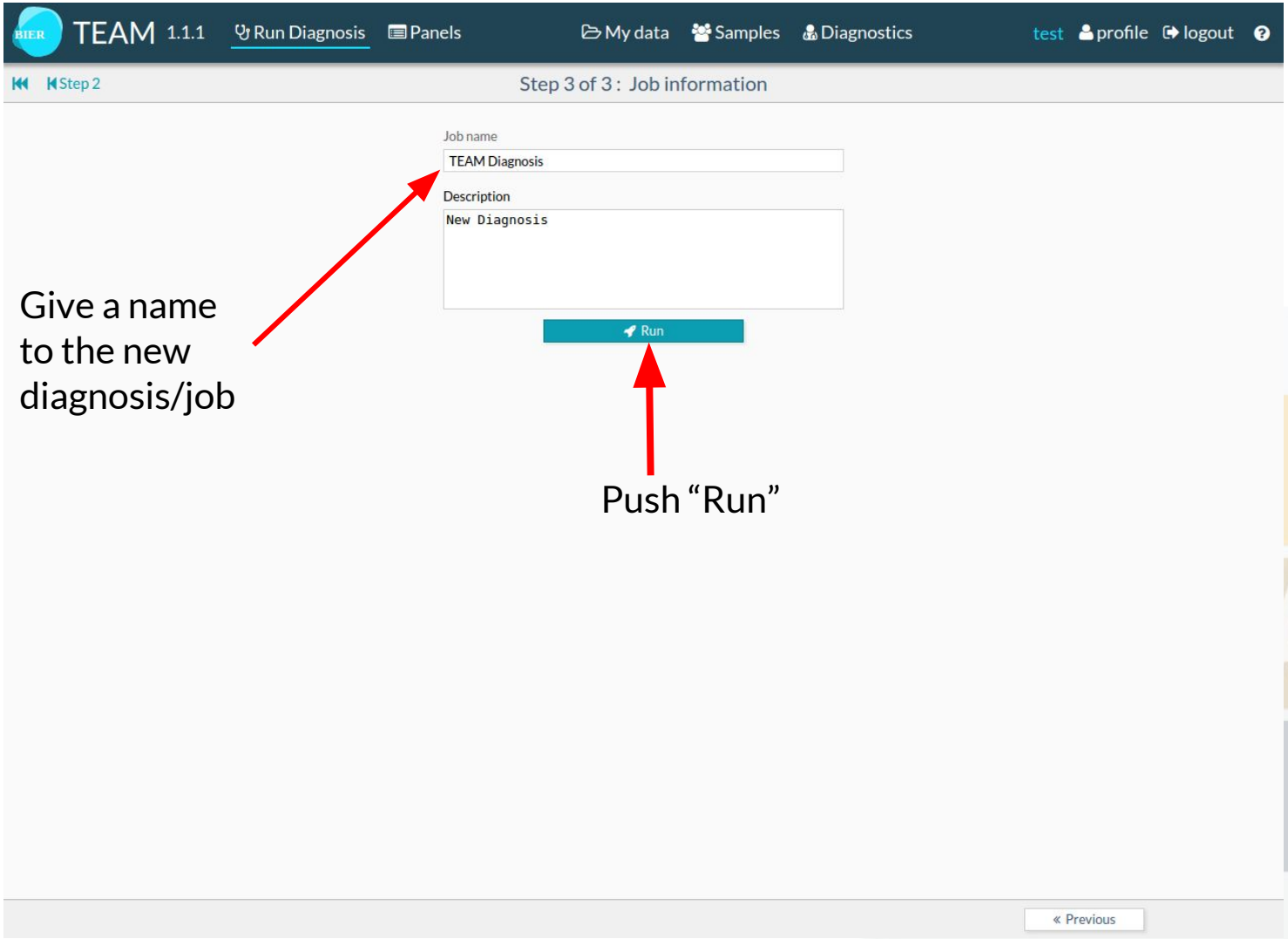
New Diagnosis

Select the panel to be used.



Next Step

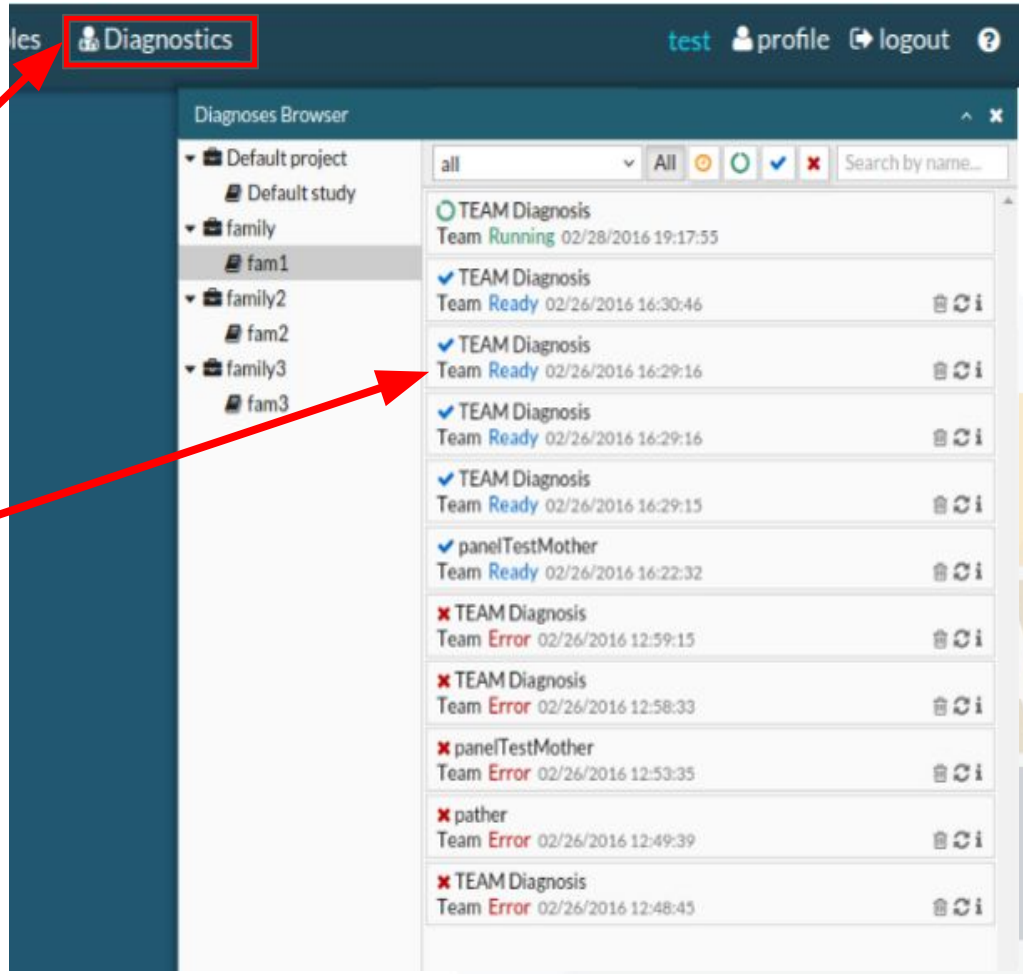
New Diagnosis



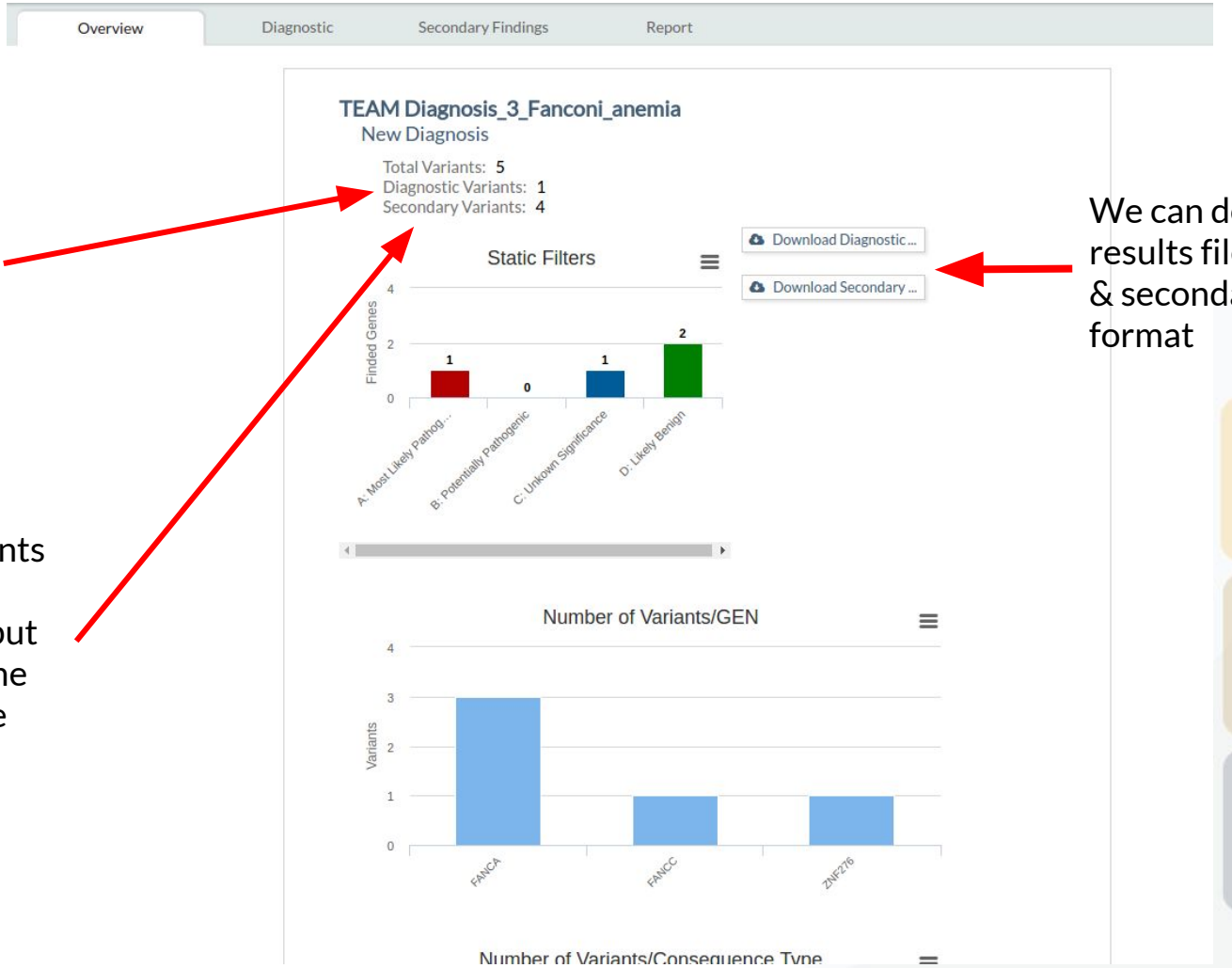
Diagnoses

This view shows the status of the current/past diagnoses.
The different status are: QUEUED, RUNNING, READY, ERROR

If you select a specific Diagnoses you will access to the results of that diagnoses.



Results: Overview

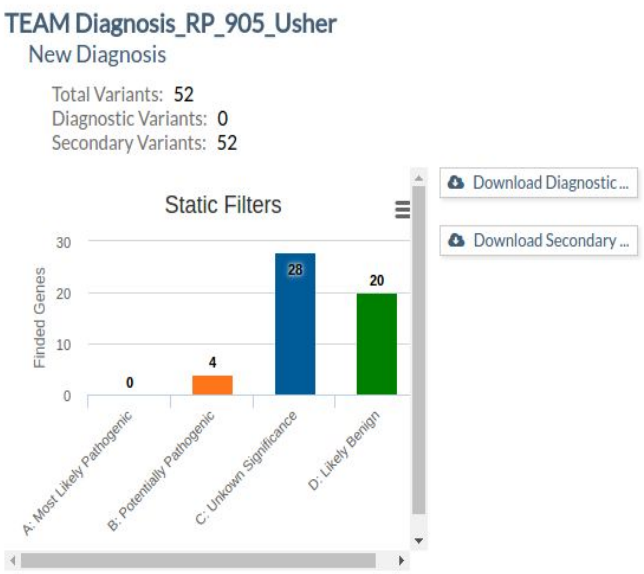


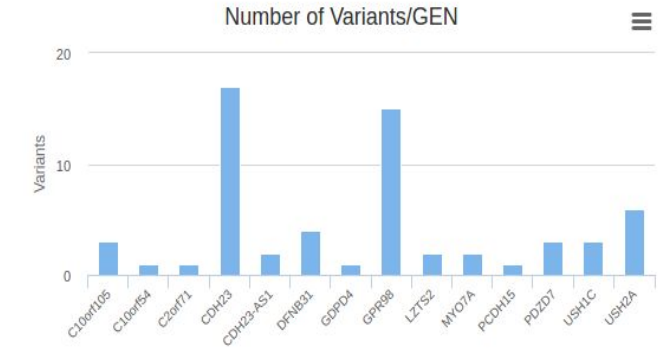
We got one diagnostic variant

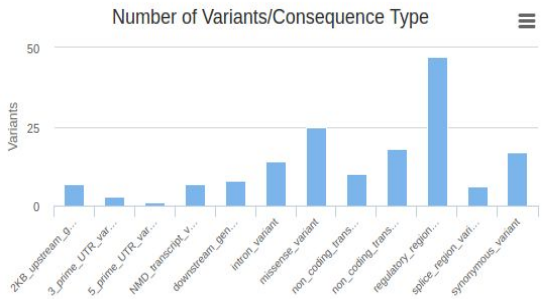
Secondary findings: 4
These variants are not diagnostic but are inside the genes of the panel

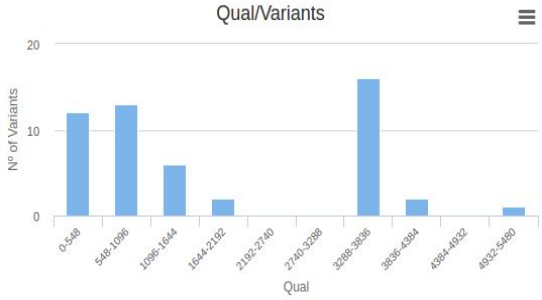
We can download the results files (diagnostic & secondary) in CSV format

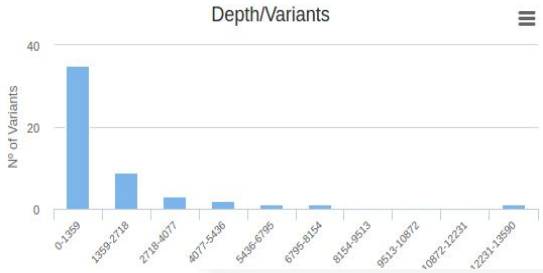
Results: Overview



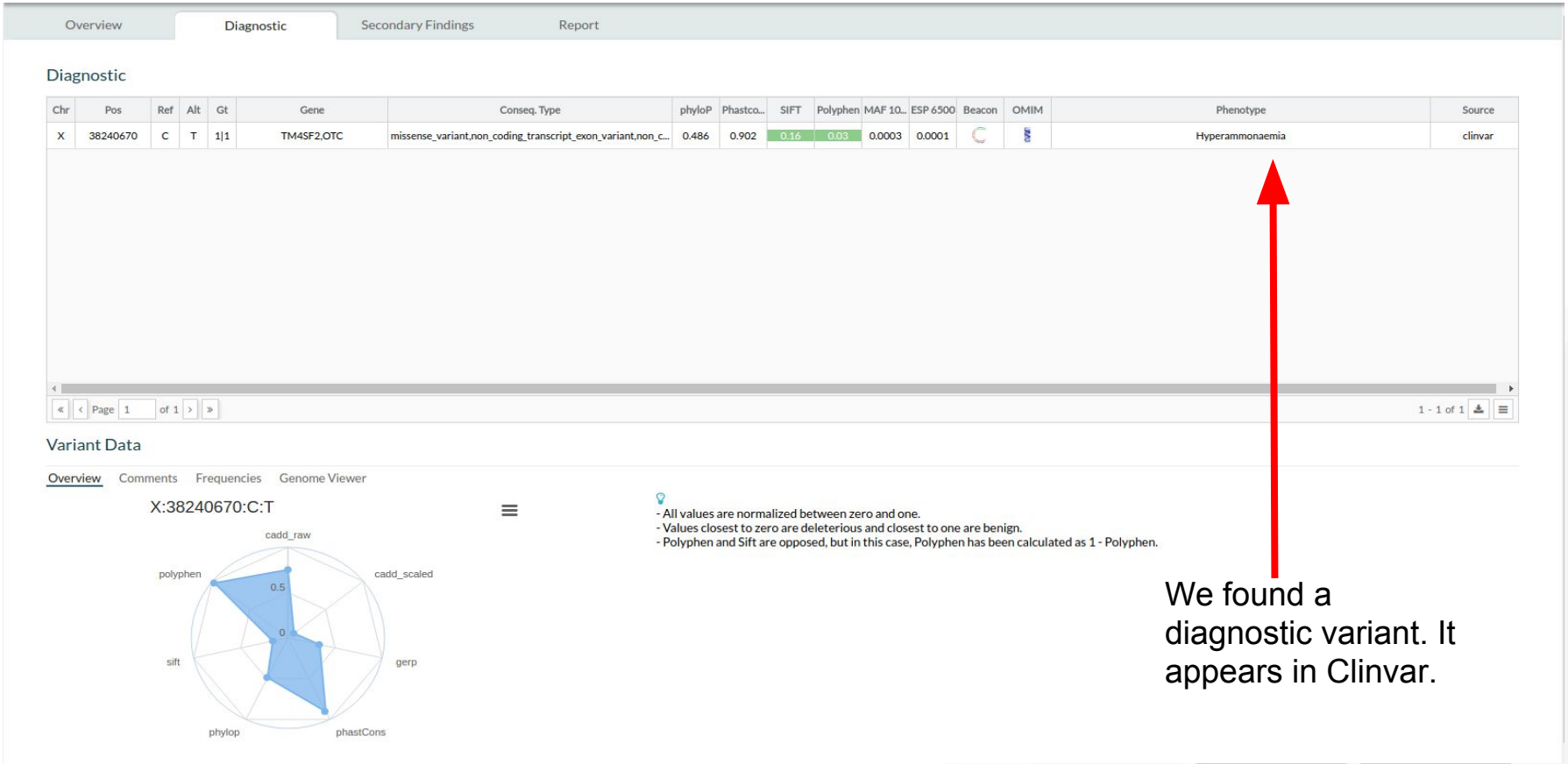






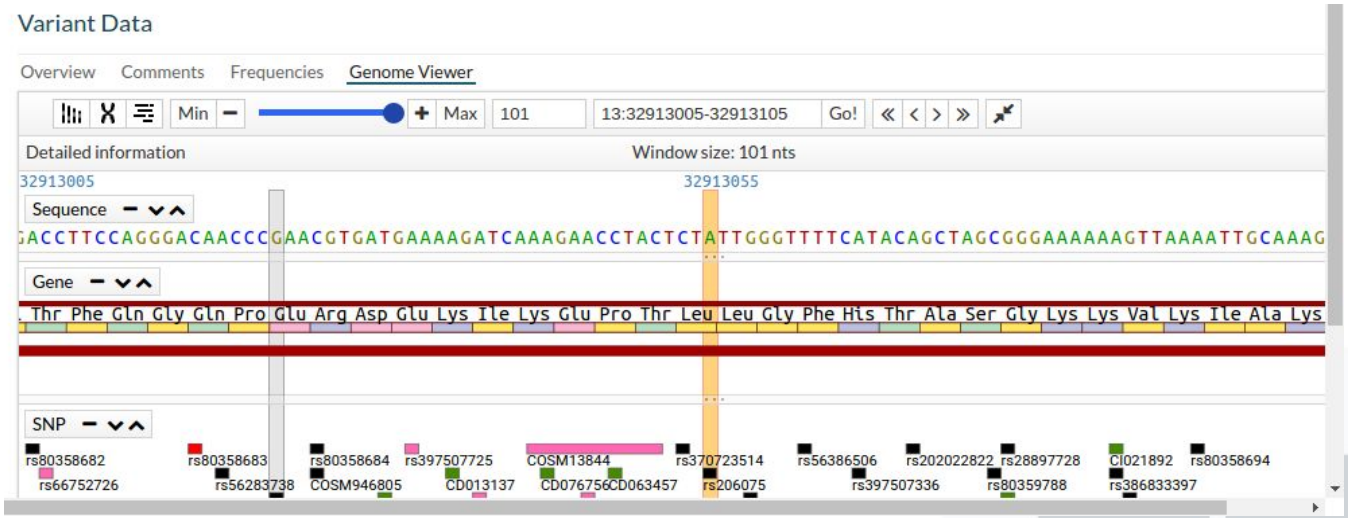


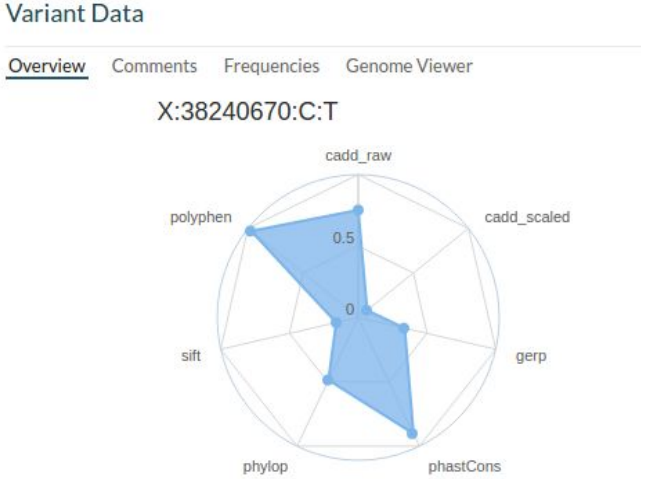
Results: Diagnostic Variants

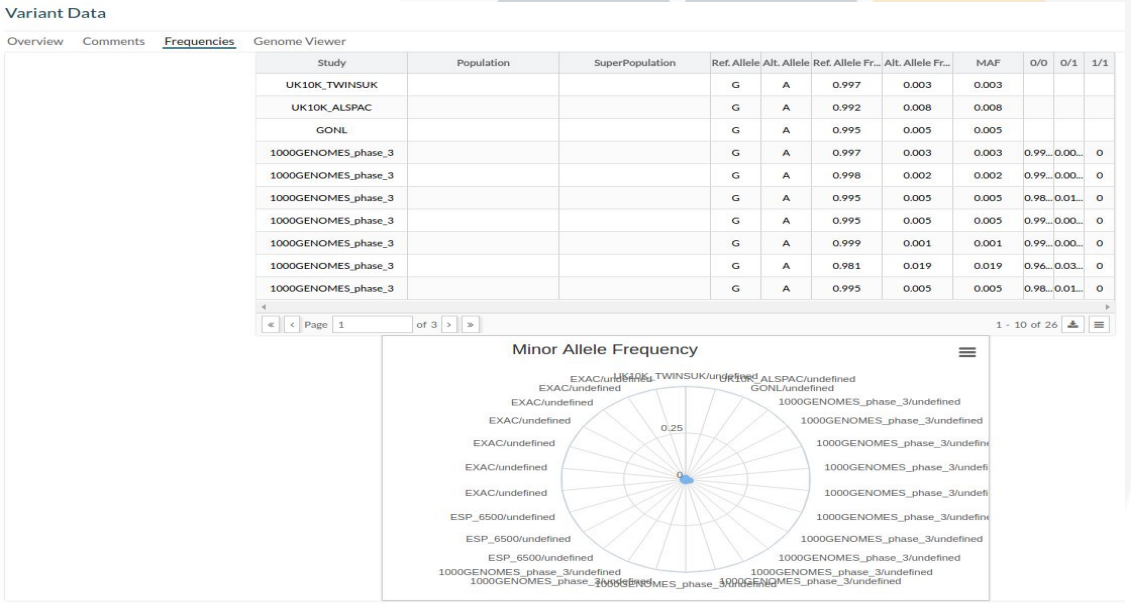


We found a diagnostic variant. It appears in Clinvar.

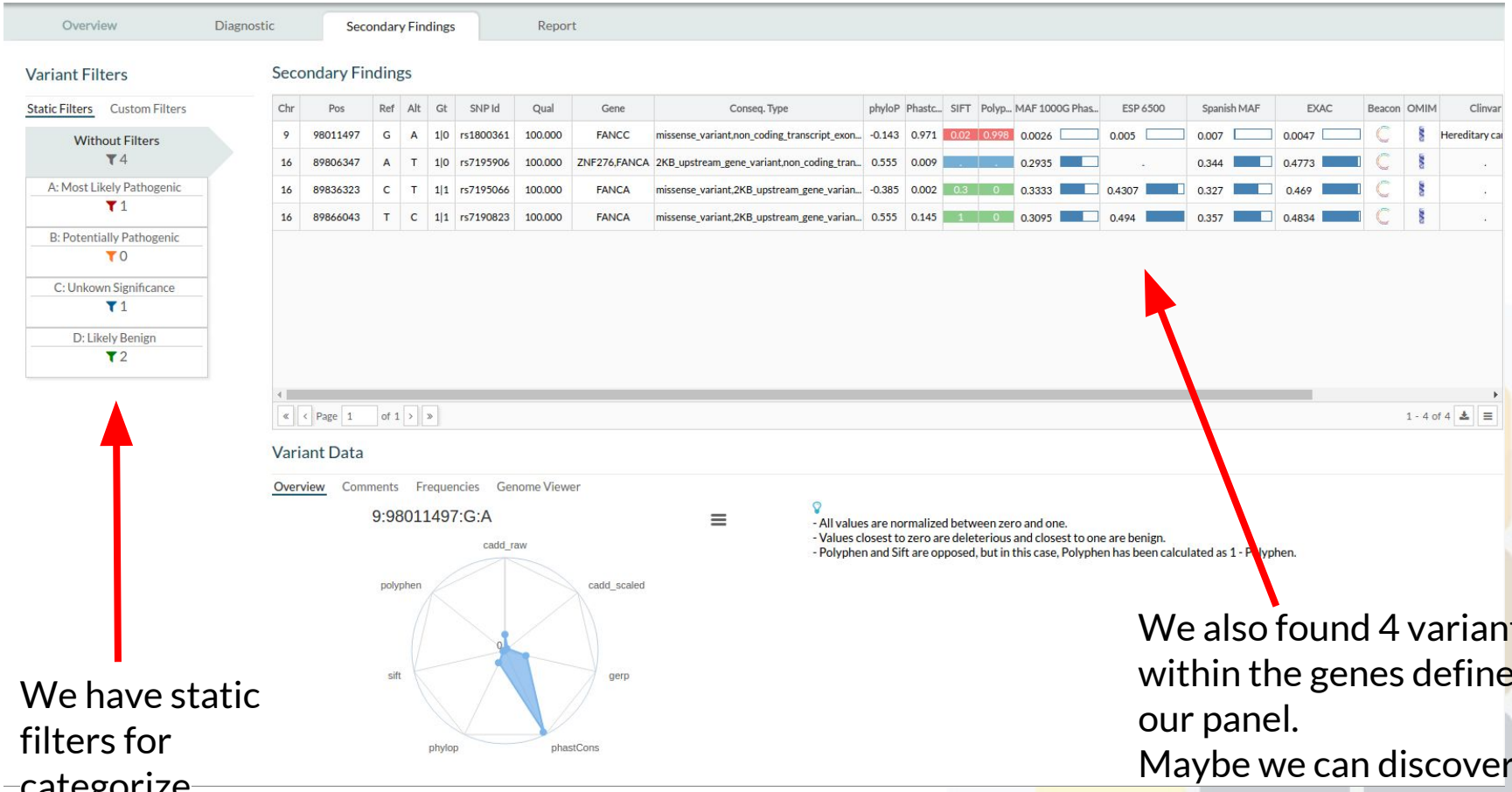
Results: Variant Data







Results: Secondary Findings



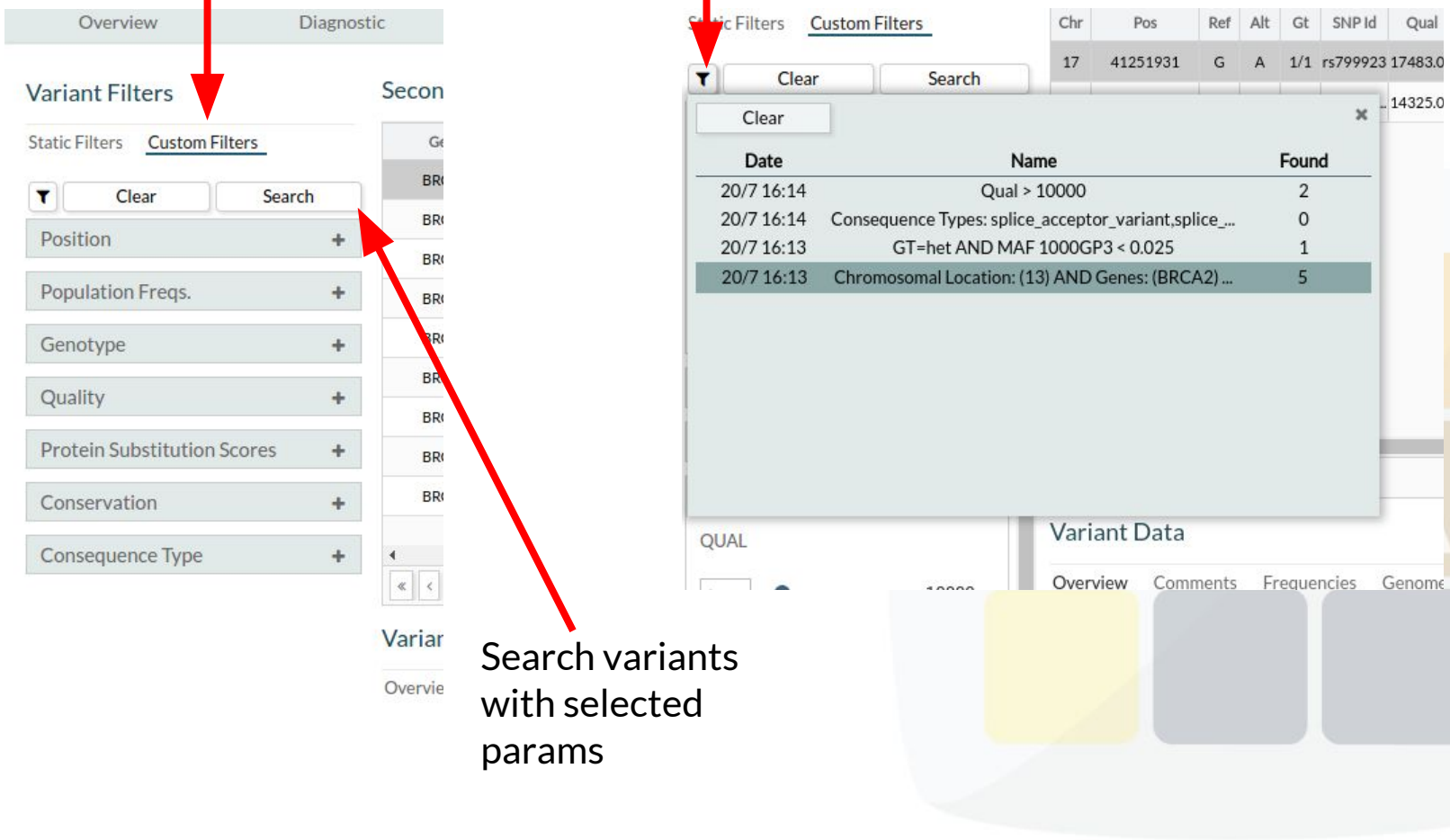
We have static filters for categorize Secondary findings variants in four levels.

We also found 4 variants within the genes defined in our panel.
Maybe we can discover an interesting variant here.

Results: Secondary Findings

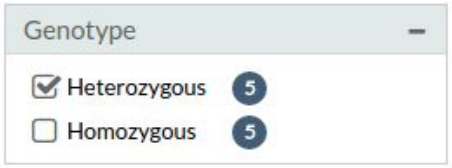
Custom filters

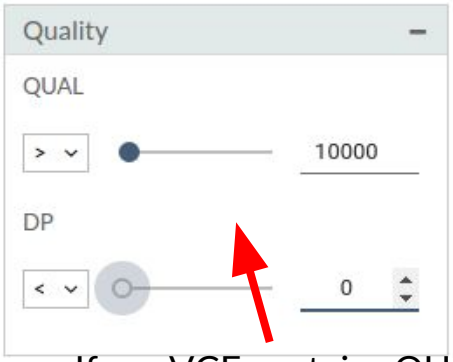
You can see filters used previously in Filters History



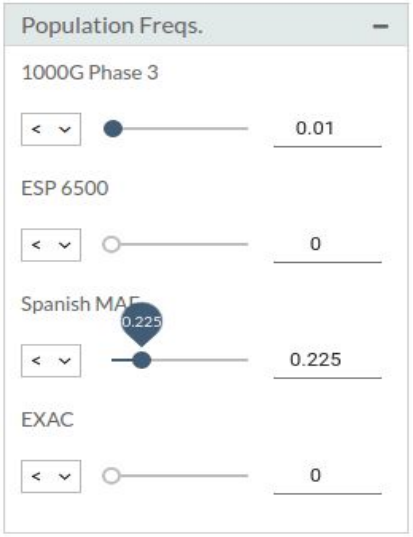
Search variants with selected params

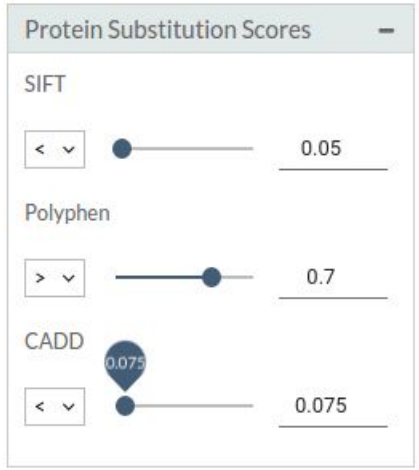
Results: Secondary Findings (Filters)

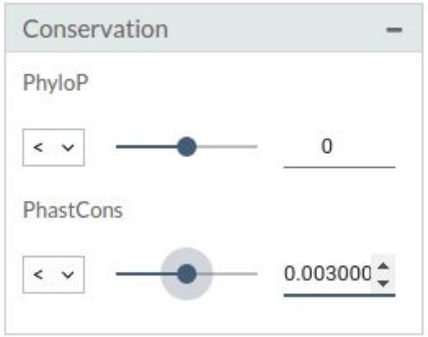


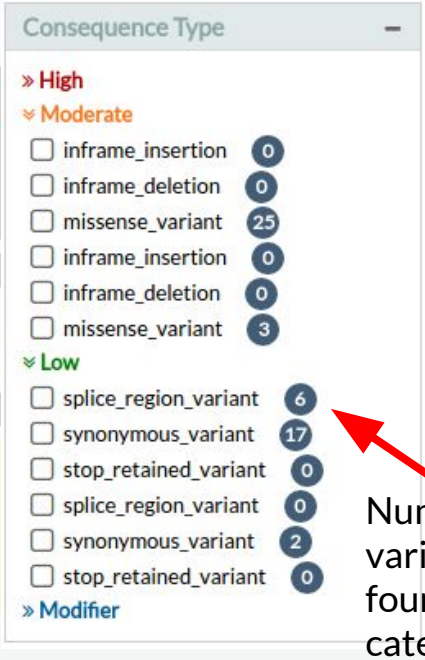


If our VCF contains QUAL & DP we can filter using them.









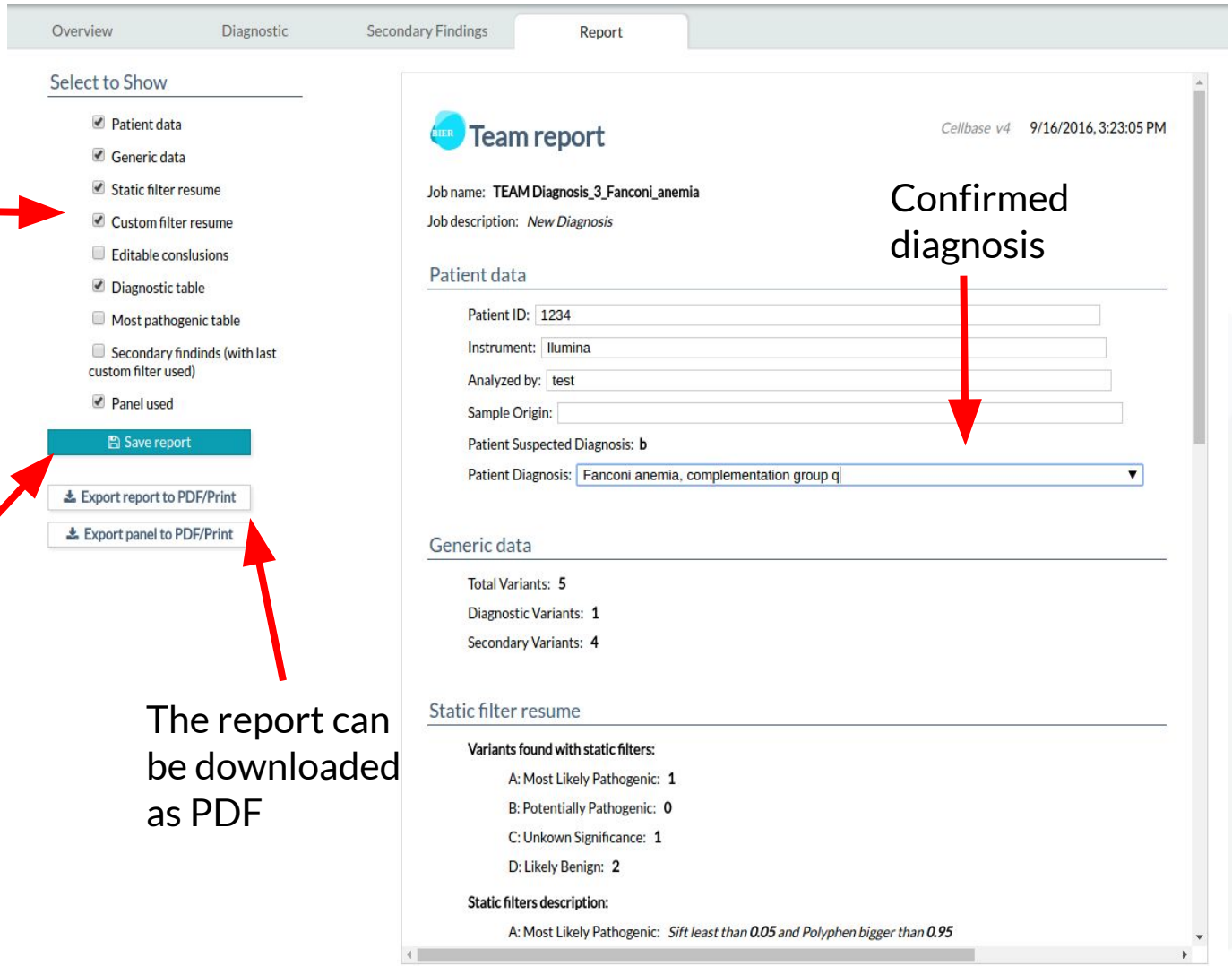
Number of variants found in each category

Results: Report

We can choose which sections we want to print

Complete the information about patient and save the report for a next view

The report can be downloaded as PDF



Who is using TEAM?

ciberer isciii | Centro de Investigación Biomédica en Red Enfermedades Raras

IT4Innovations
national!$11€0
supercomputing
center1001$1$0

Conclusions

- TEAM is a **free and easy-to-use web tool** that fills the gap between the enormous amount of data in targeted enrichment sequencing analysis and the **biological knowledge** available.

- TEAM **provides an intuitive environment for the clinicians** in which unprocessed data on patient's genomic variation can easily be transformed in a **diagnostic**.

- All data is stored in a Server so you can access to you diagnostics in anywhere you want.

More info: publication

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

More info: TEAM behind the scenes