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

Álex Alemán

March 2nd, 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

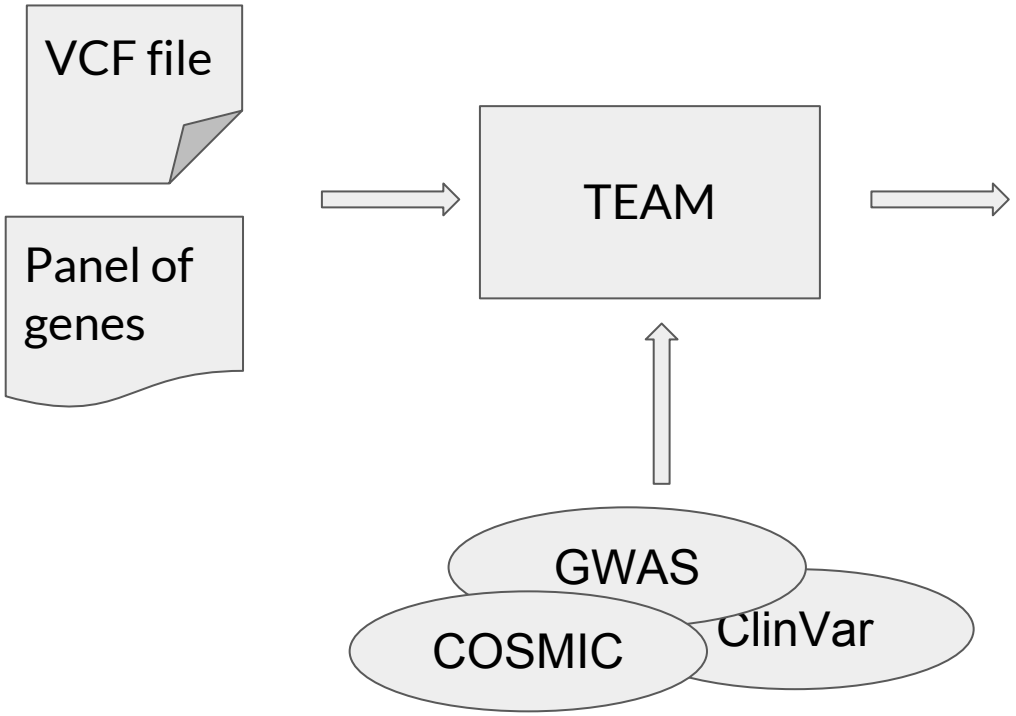
COSMIC

ClinVar

TEAM

Diagnostic

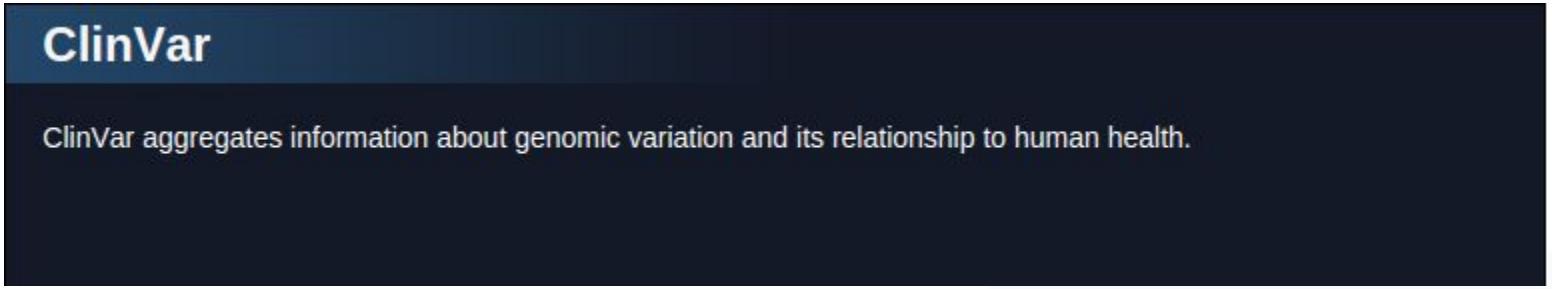
How does TEAM work?



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/

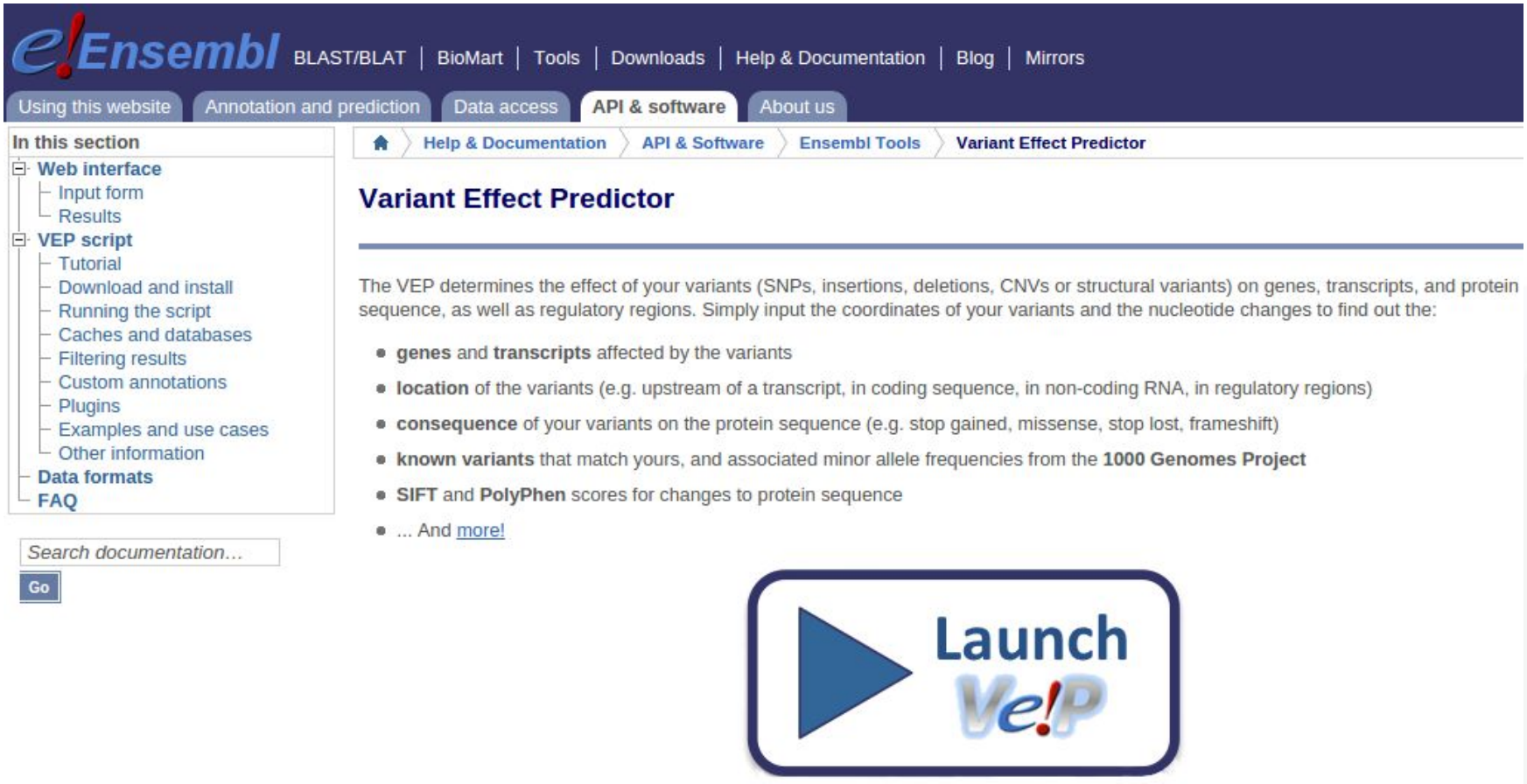
- **PolyPhen**
  - Polymorphism Phenotyping is a tool which preditcs possible impact of an amino acid substitution on the structure and function of a human protein.
  - **Interpretation:** 1 (probably damage) to 0 (bening)

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

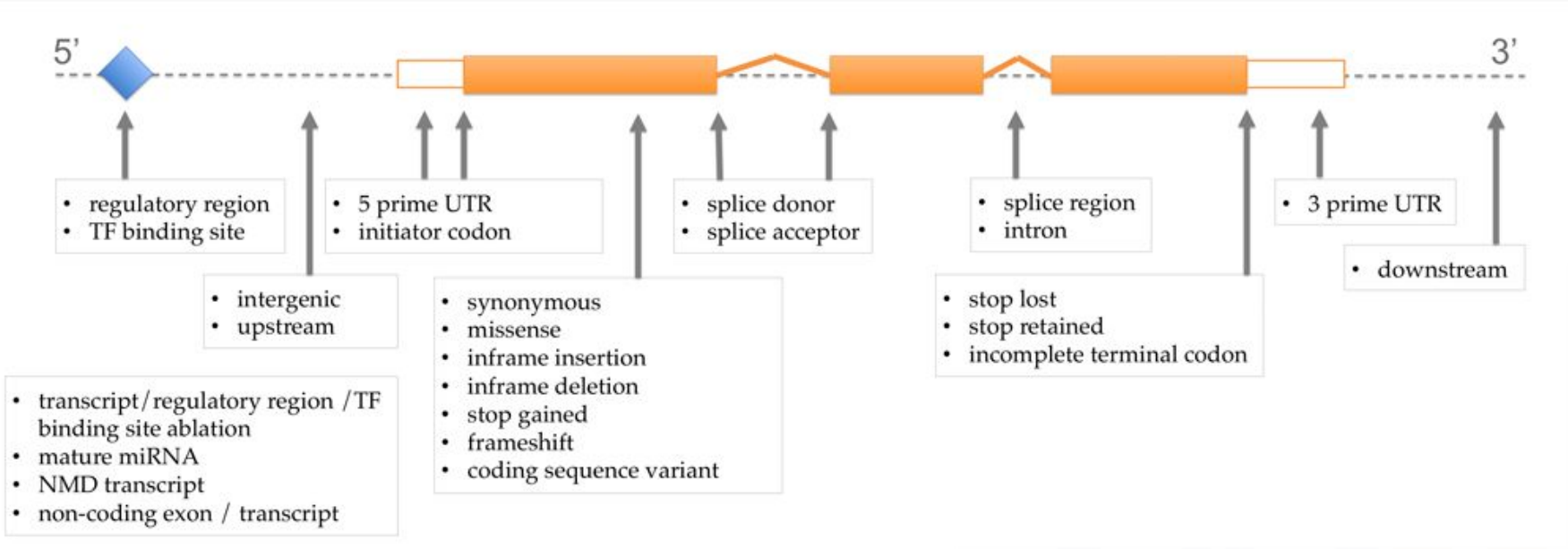
Getting information: Conservation

- **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



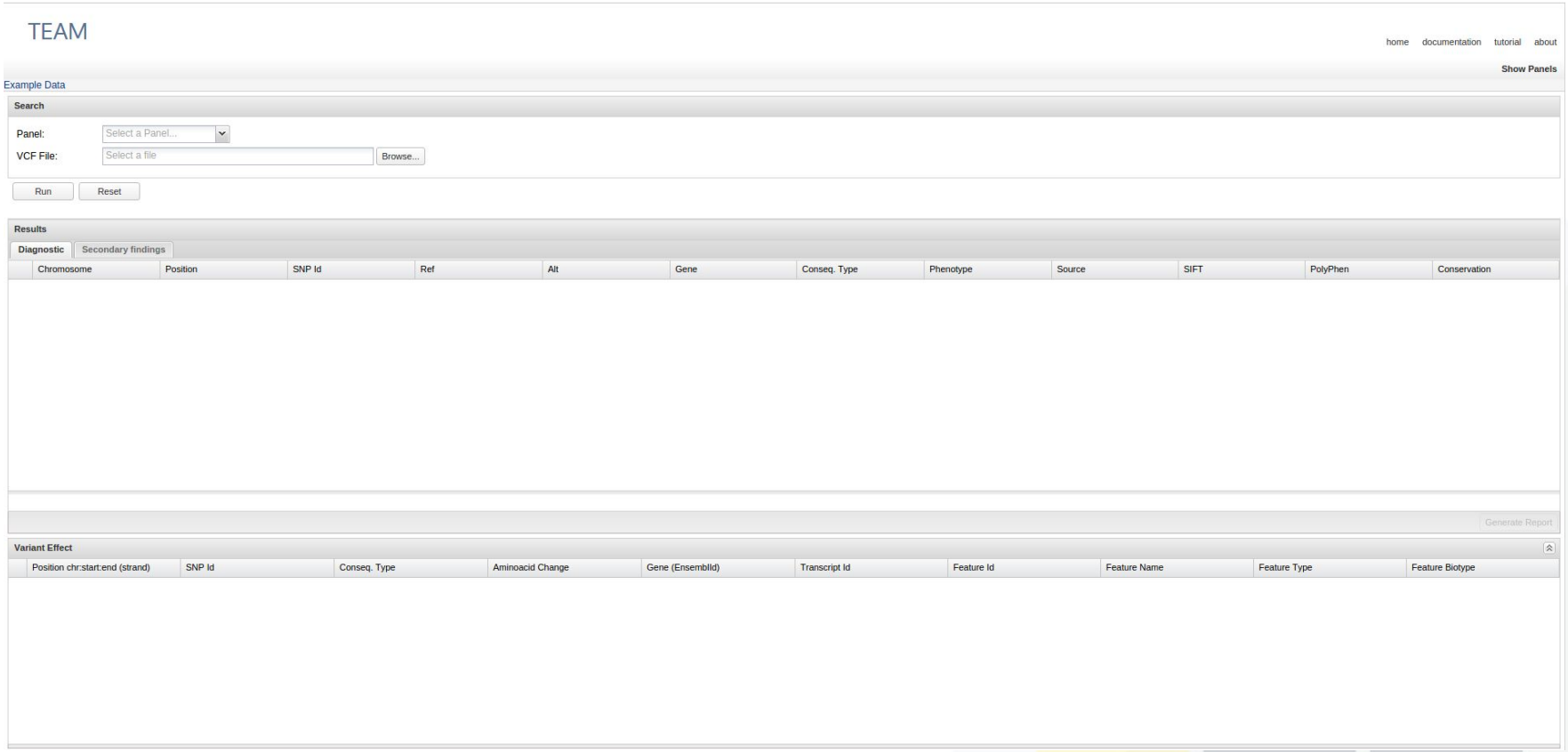
http://www.ensembl.org/info/genome/variation/predicted_data.html

Inputs: Panel + VCF

1. TEAM needs a Panel of genes and a VCF file:
   a. The VCF file needs to be stored/indexed in our database.
   b. This file could be already added if we used BiERapp (both tools are compatible
2. What is a Panel of genes:.
   a. TEAM works with virtual panel of genes.
   b. You can design/create/manage these panels using TEAM.
   c. These panels contains:
      i. Phenotypes
      ii. Genes
      iii. Mutations

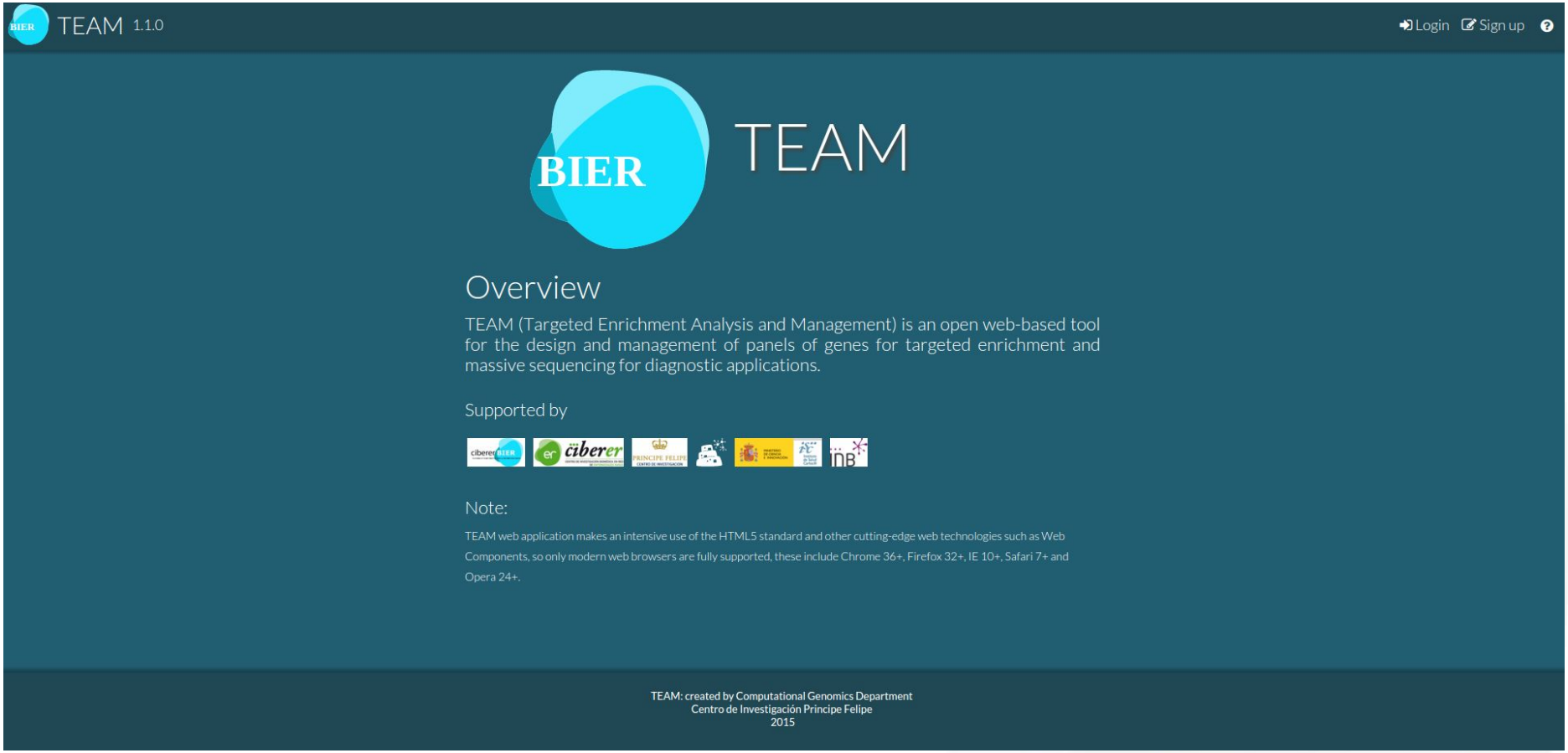
Tool interface: Official release

http://team.babelomics.org/



Tool interface: Beta

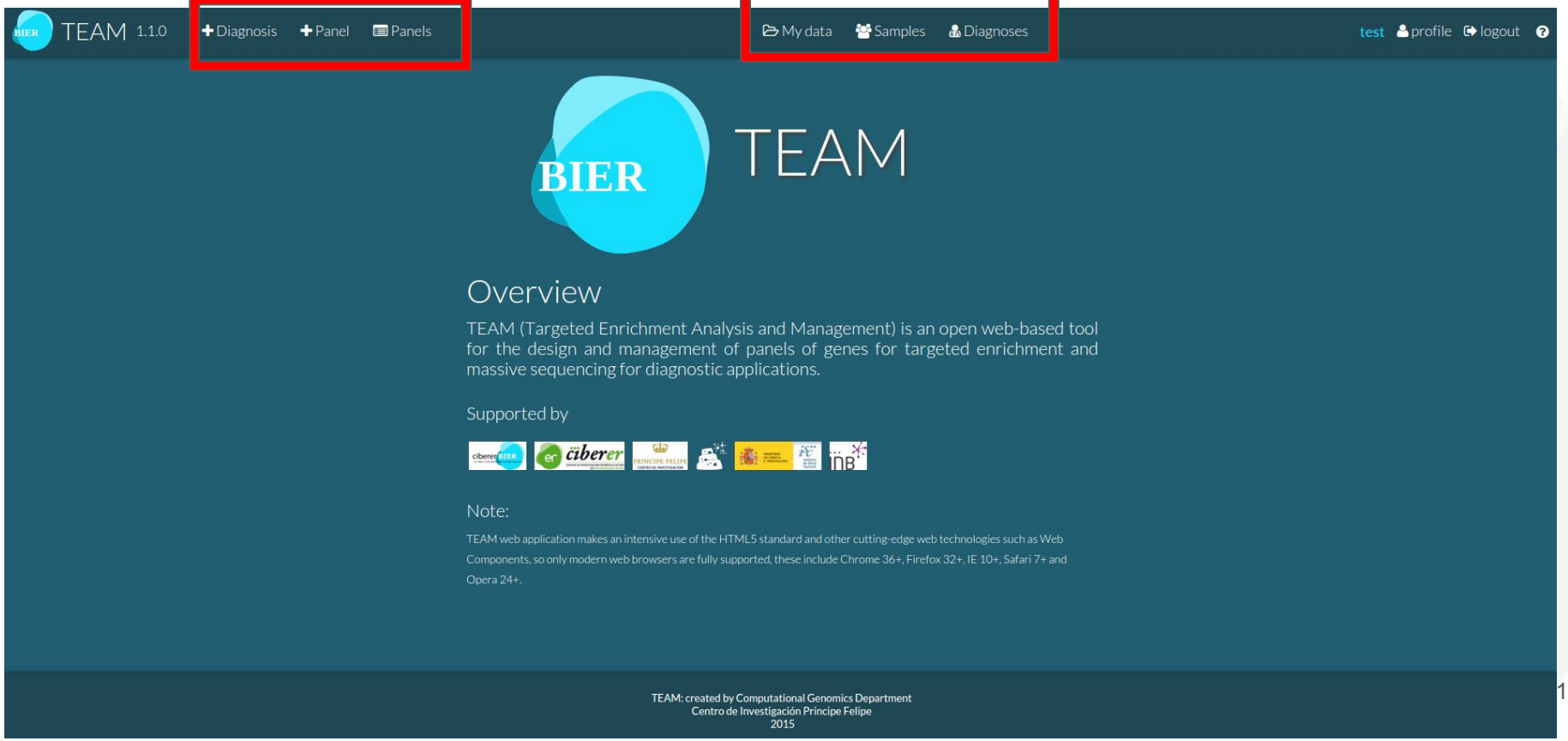
http://team.babelomics.org/beta



Tool interface: Beta

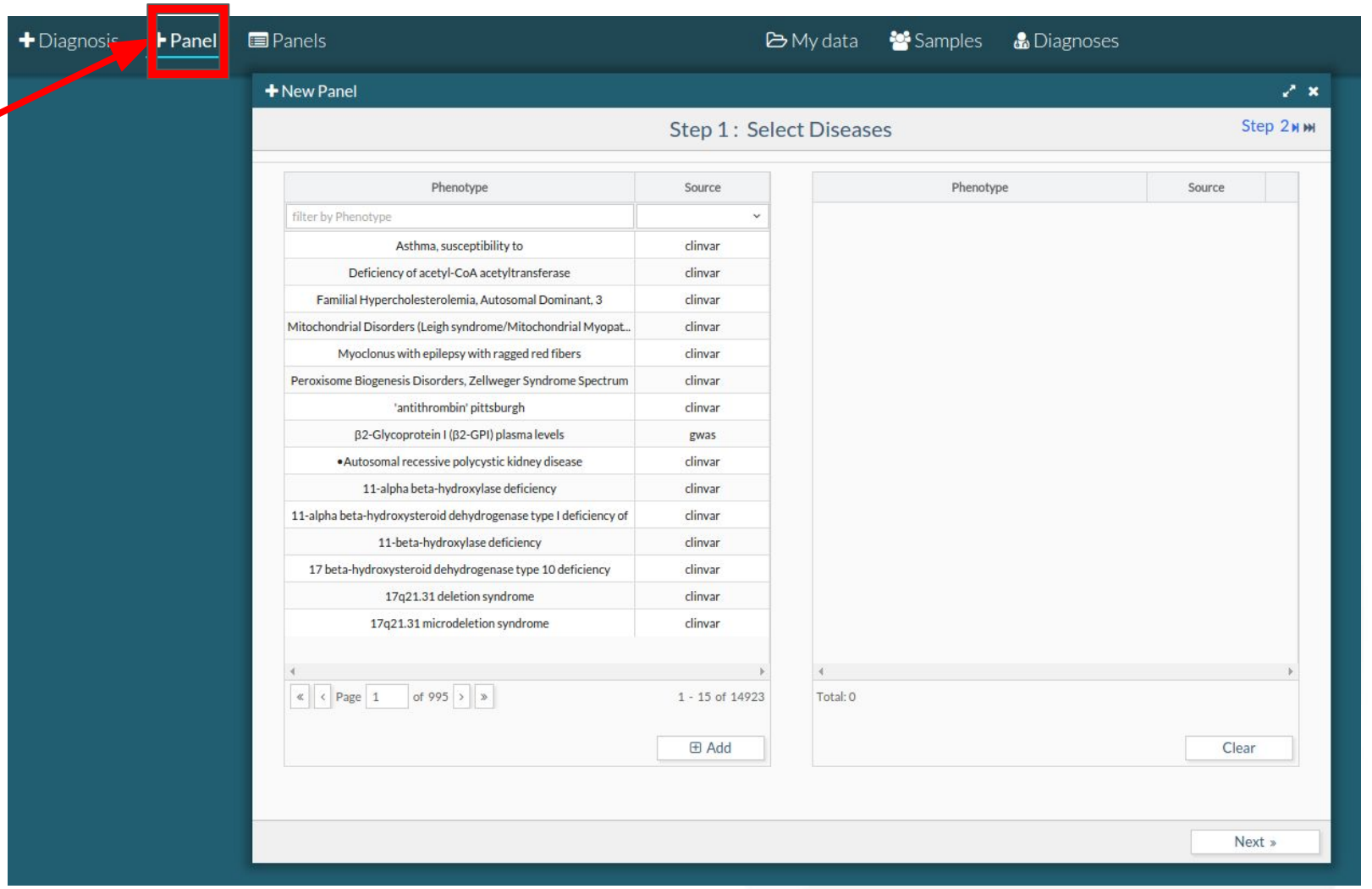
Design/Manage Panels

Data & Diagnoses



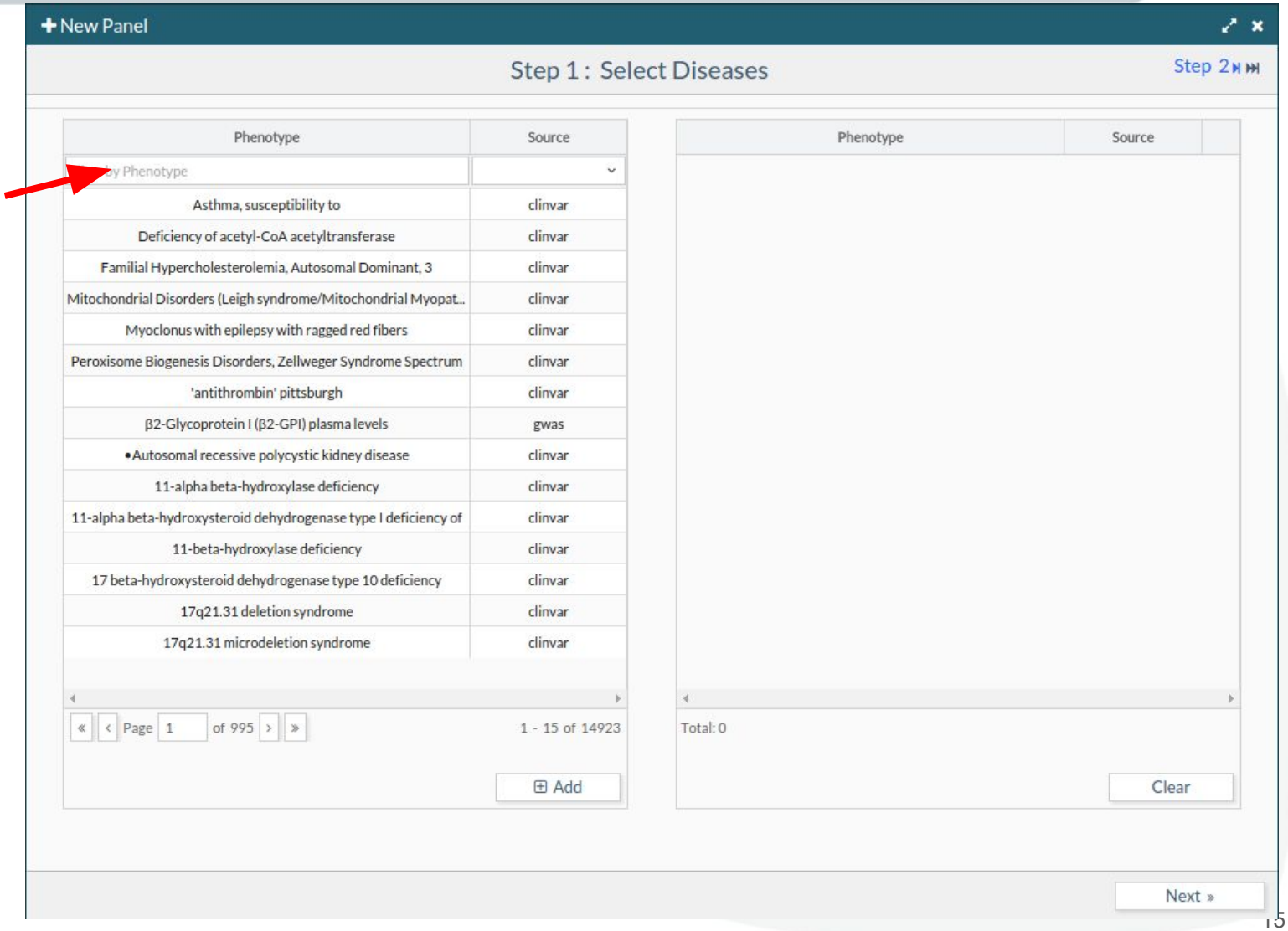
Panel Designer

Select
"+ Panel"



Panel Designer: Diseases

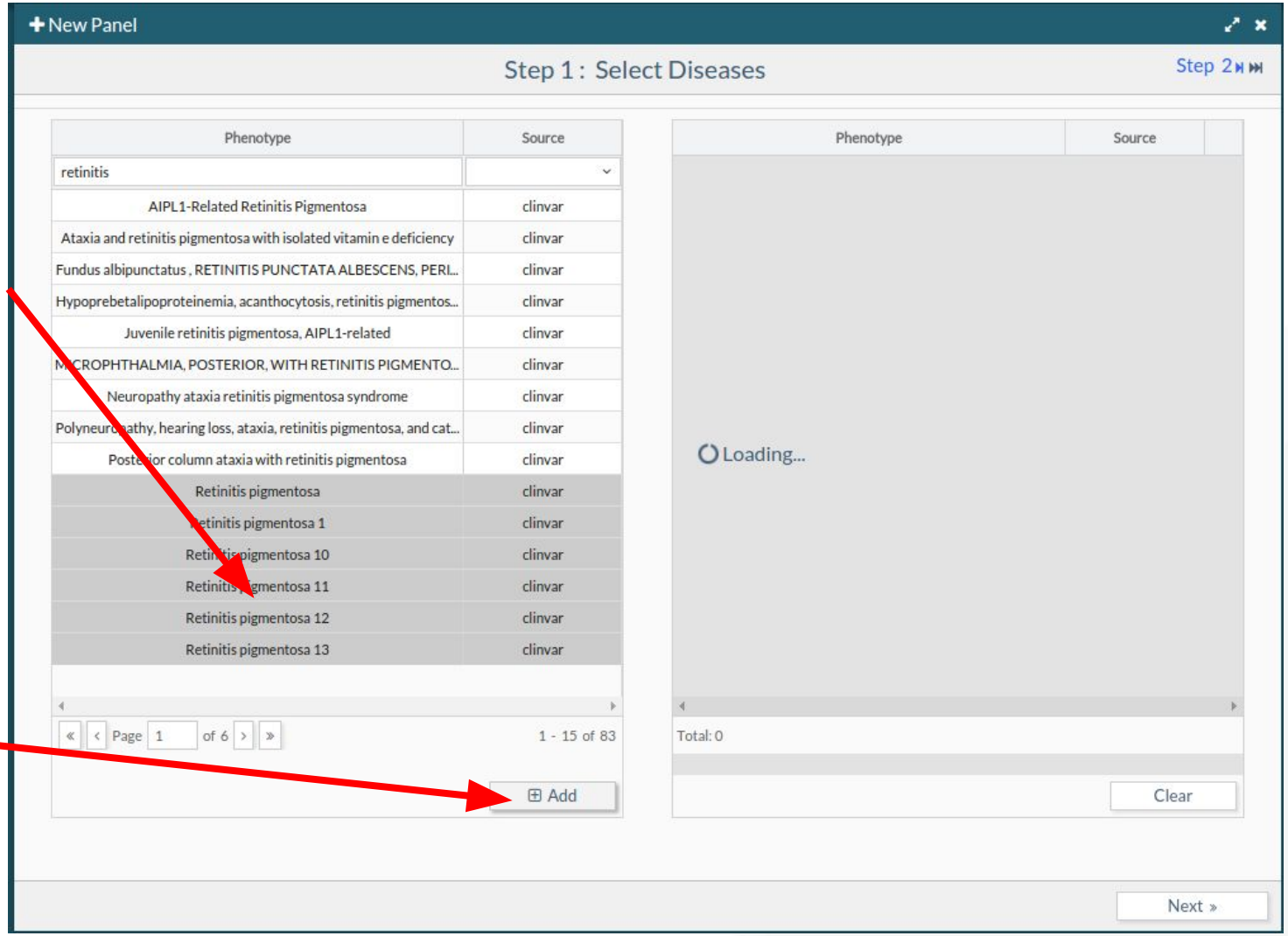
Write the phenotypes you are interested in.



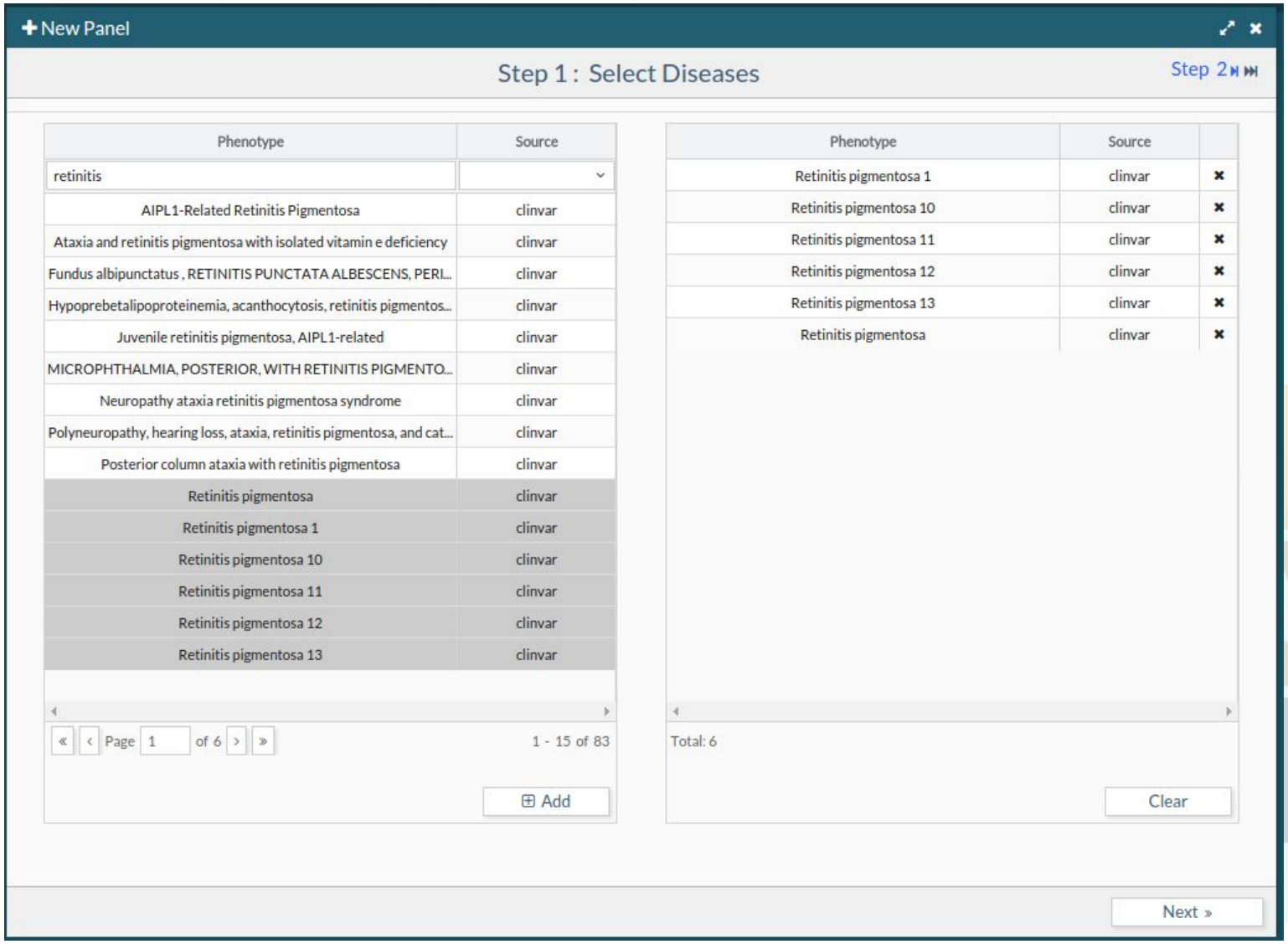
Panel Designer: Diseases

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

Click "Add"



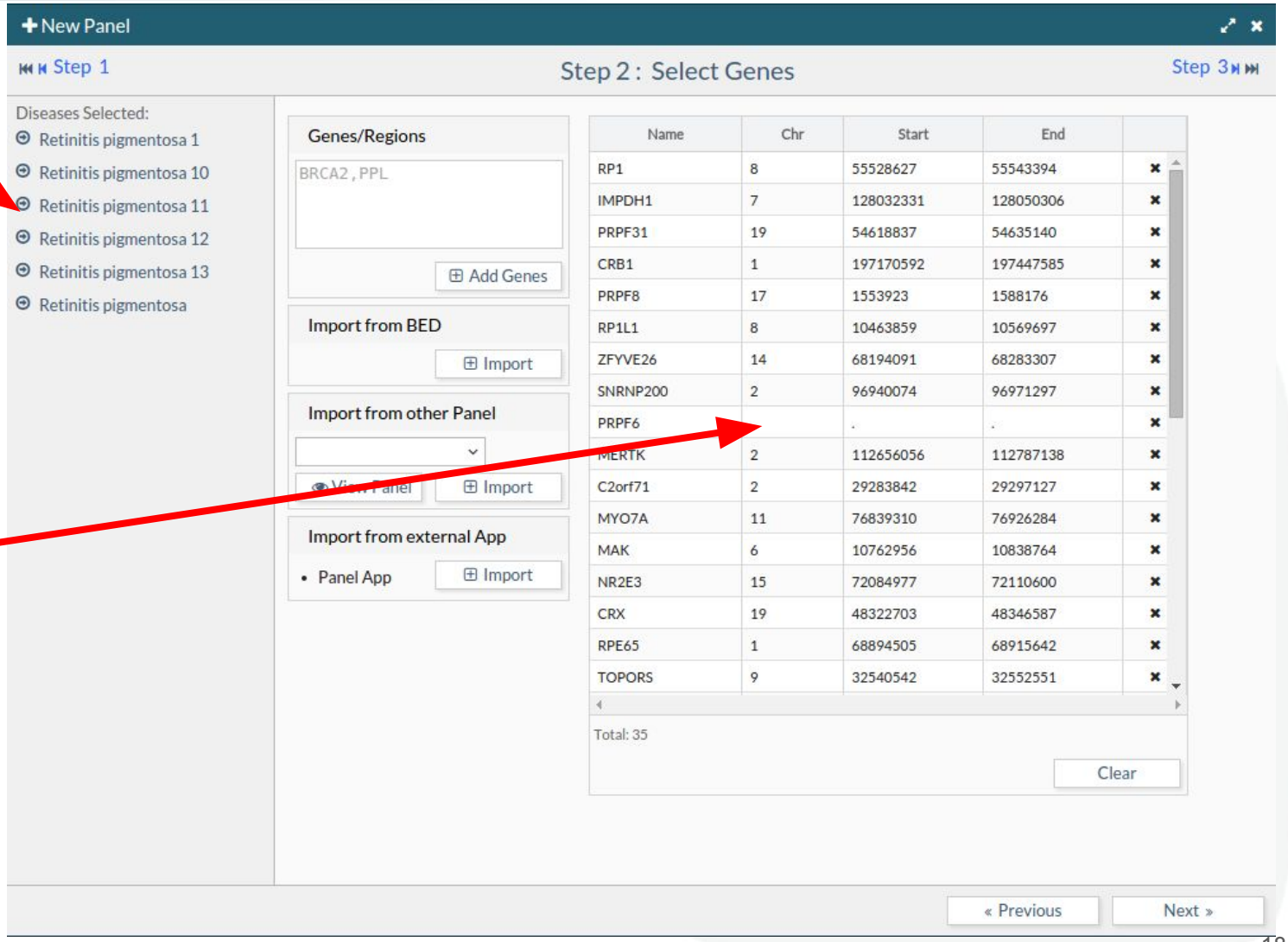
Panel Designer: Diseases



Panel Designer: Genes

Selected Diseases in previous step.

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



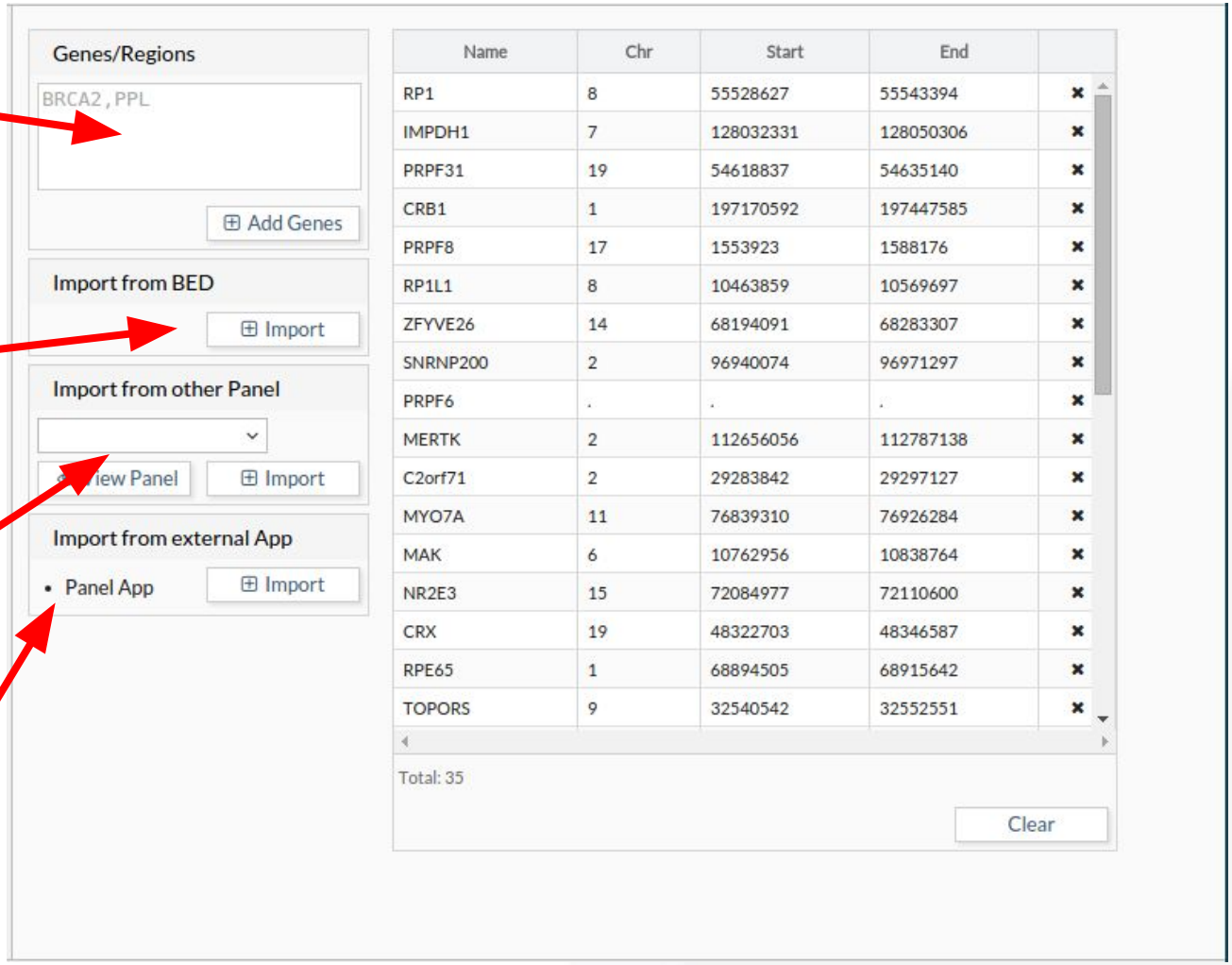
Panel Designer: Genes

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

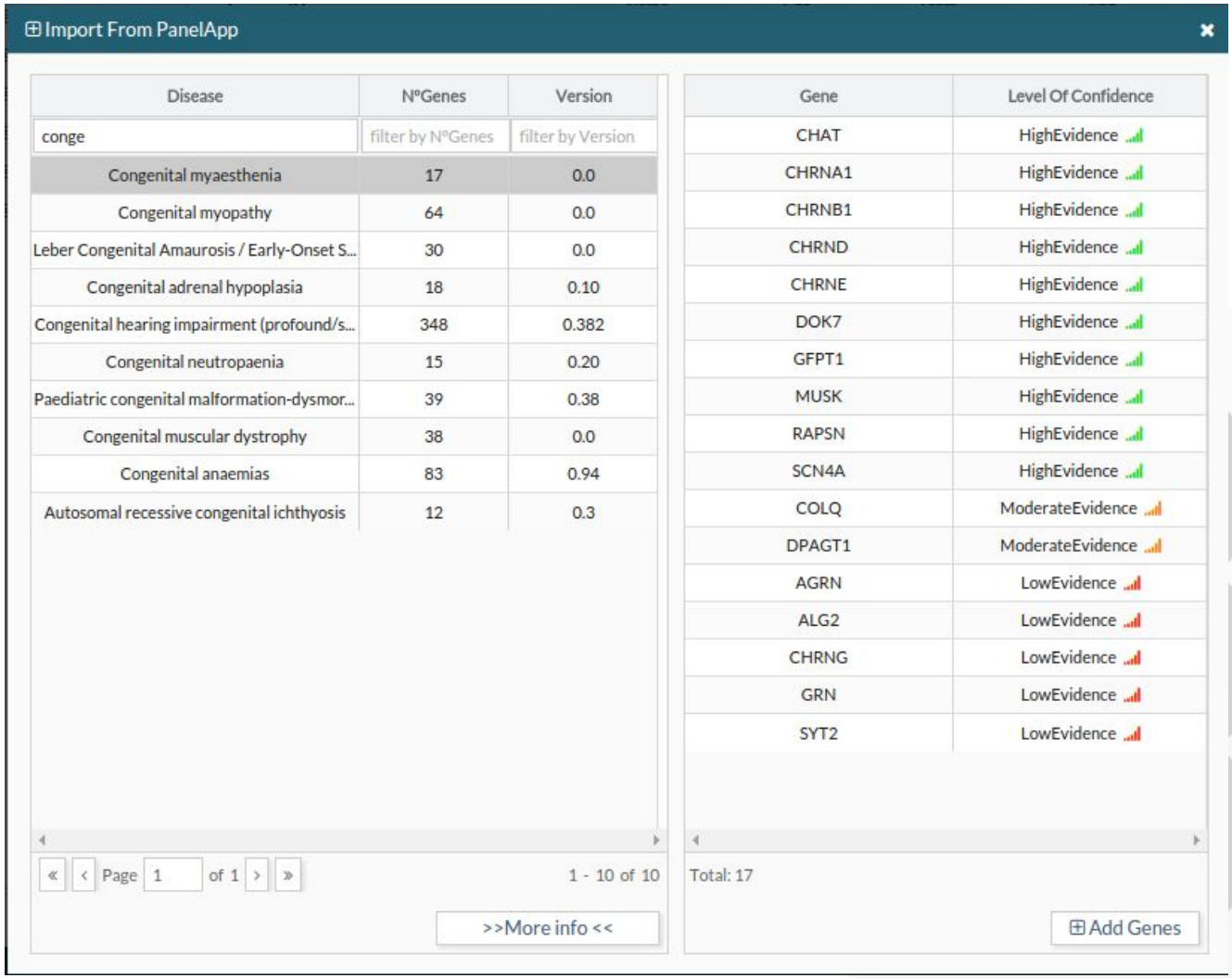
Adding regions through a BED file is also supported

We can import genes from other virtual panels already created.

Other way is using PanelApp tool



Panel Designer: Genes (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

Genomics England Rare Disease List

Disease Group (Level 2 title) e.g. Cardiovascular disorders	Disease Subgroup (Level 3 title) e.g. Cardiac arrhythmia	**Specific Disease (Level 4 title)** e.g. Brugada Syndrome

A Genomics England Curator searches for the specific disease (and other phenotypes in the eligibility statement description) in information from 4 sources to create an initial gene list.

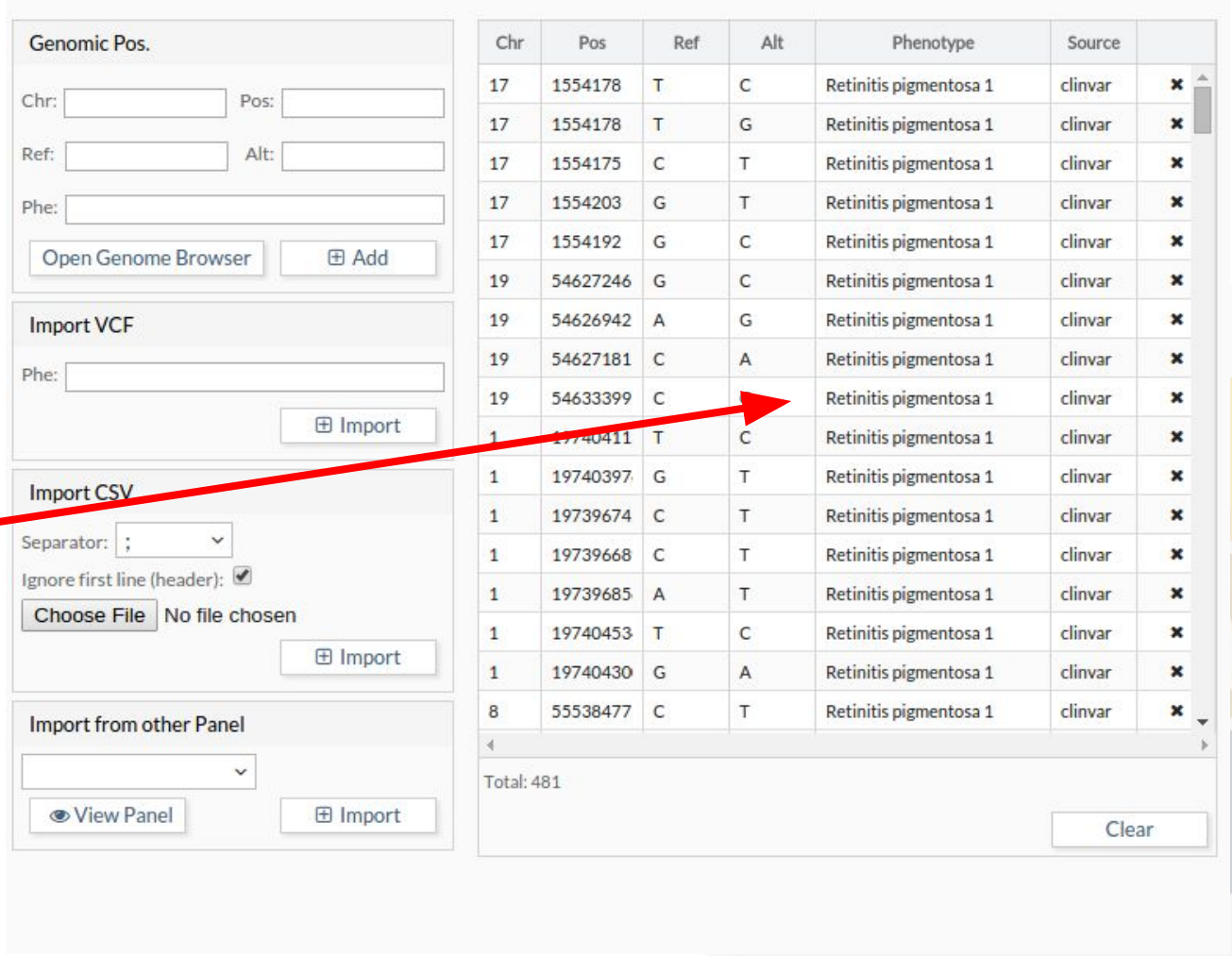
1. Radboud University Medical Center
2. Illumina Trugenome Predisposition Screen
3. Emory Genetics Laboratory
4. UKGTN

The virtual gene panel is added to PanelApp, with genes colour-coded using a traffic light system to indicate the number of sources.

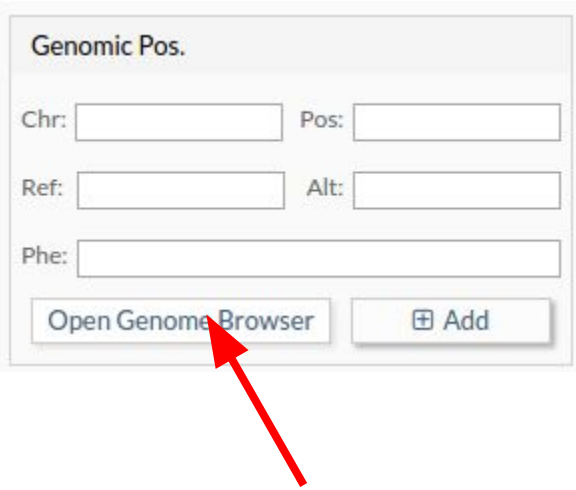
The gene is on a panel from…		Evidence Level
1 of 4 of the sources + genes from eligibility statements or from experts	Red	Low
2 of 4 of the sources	Amber	Moderate
3 or 4 of 4 of the sources	Green	High

Panel Designer: Mutations

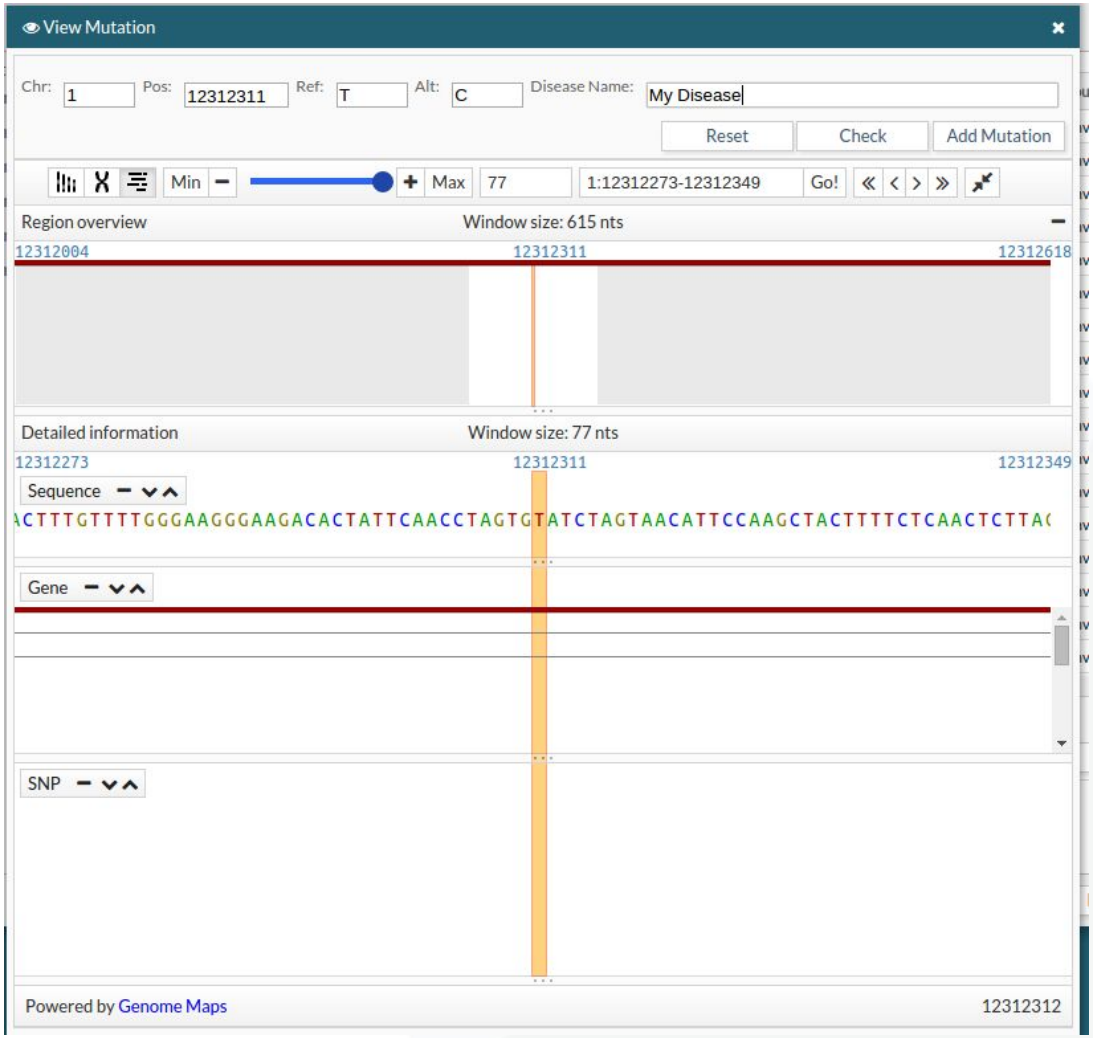
These mutations have been added by the first step. They are related to the selected diseases



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

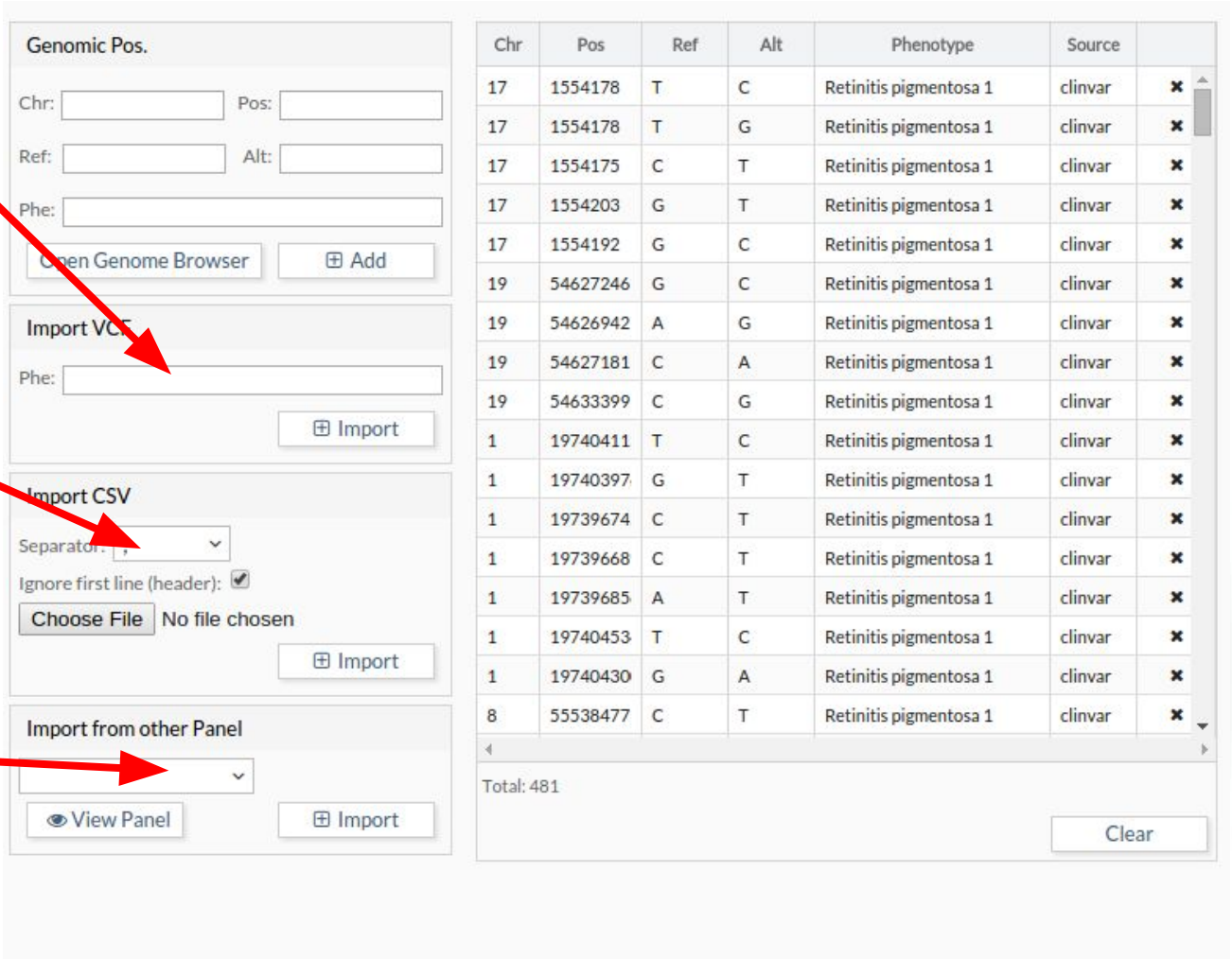


Panel Designer: Mutations

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

Importing mutations from other virtual panels is supported too.

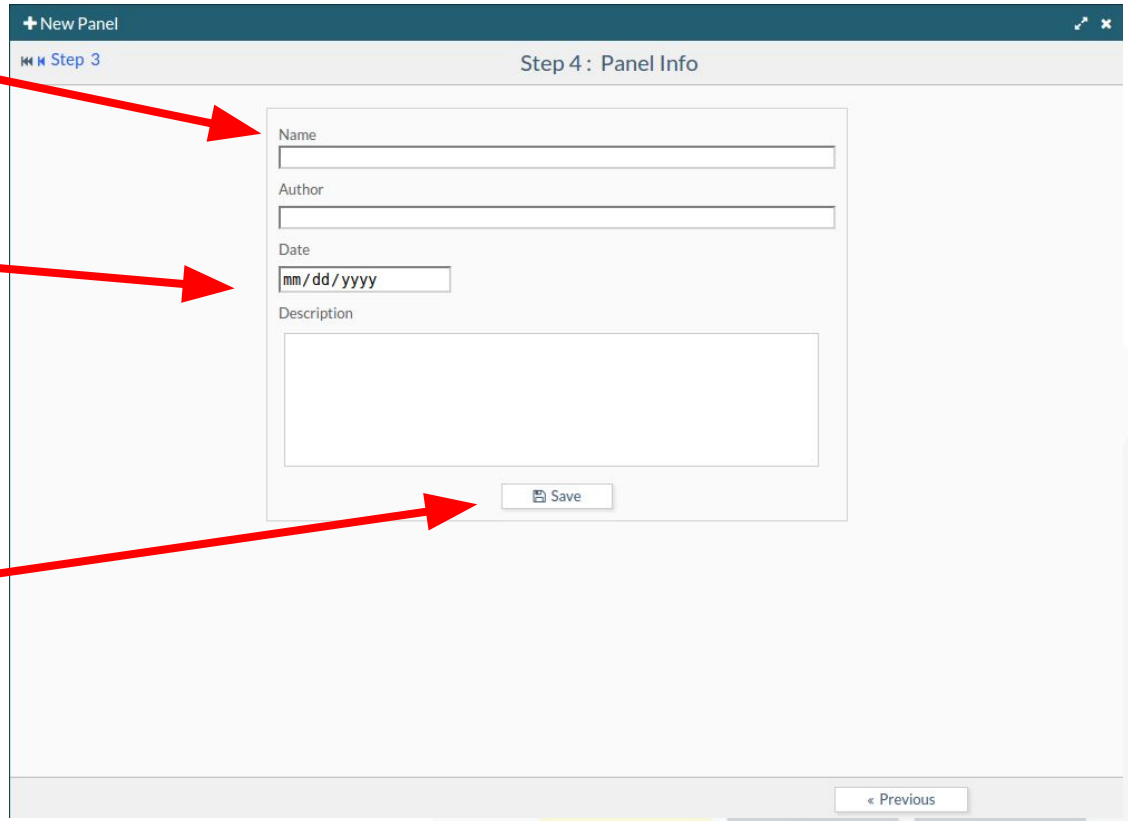


Panel Designer: Panel info

The name of the Panel

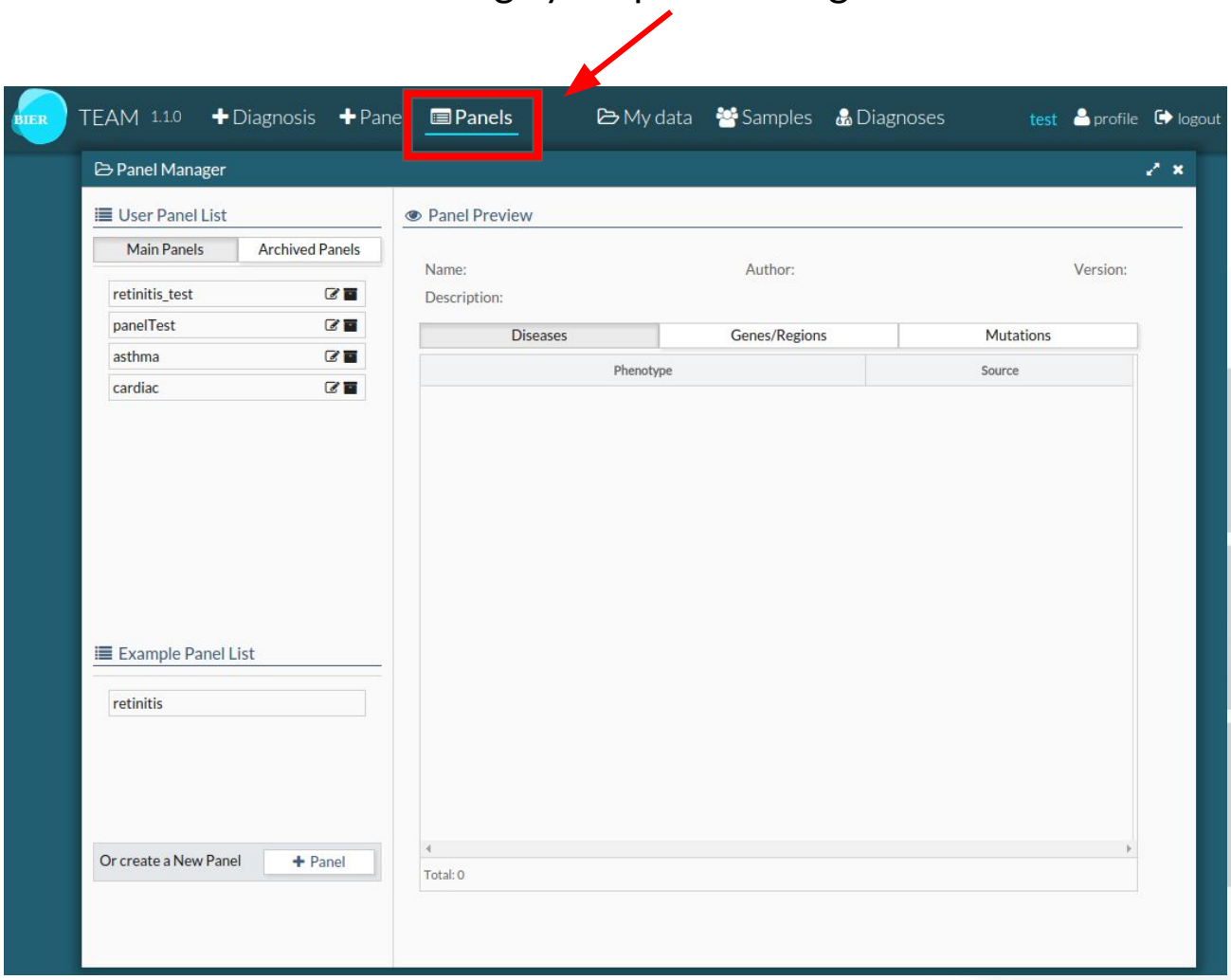
The author/date/description

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



Panel Manager

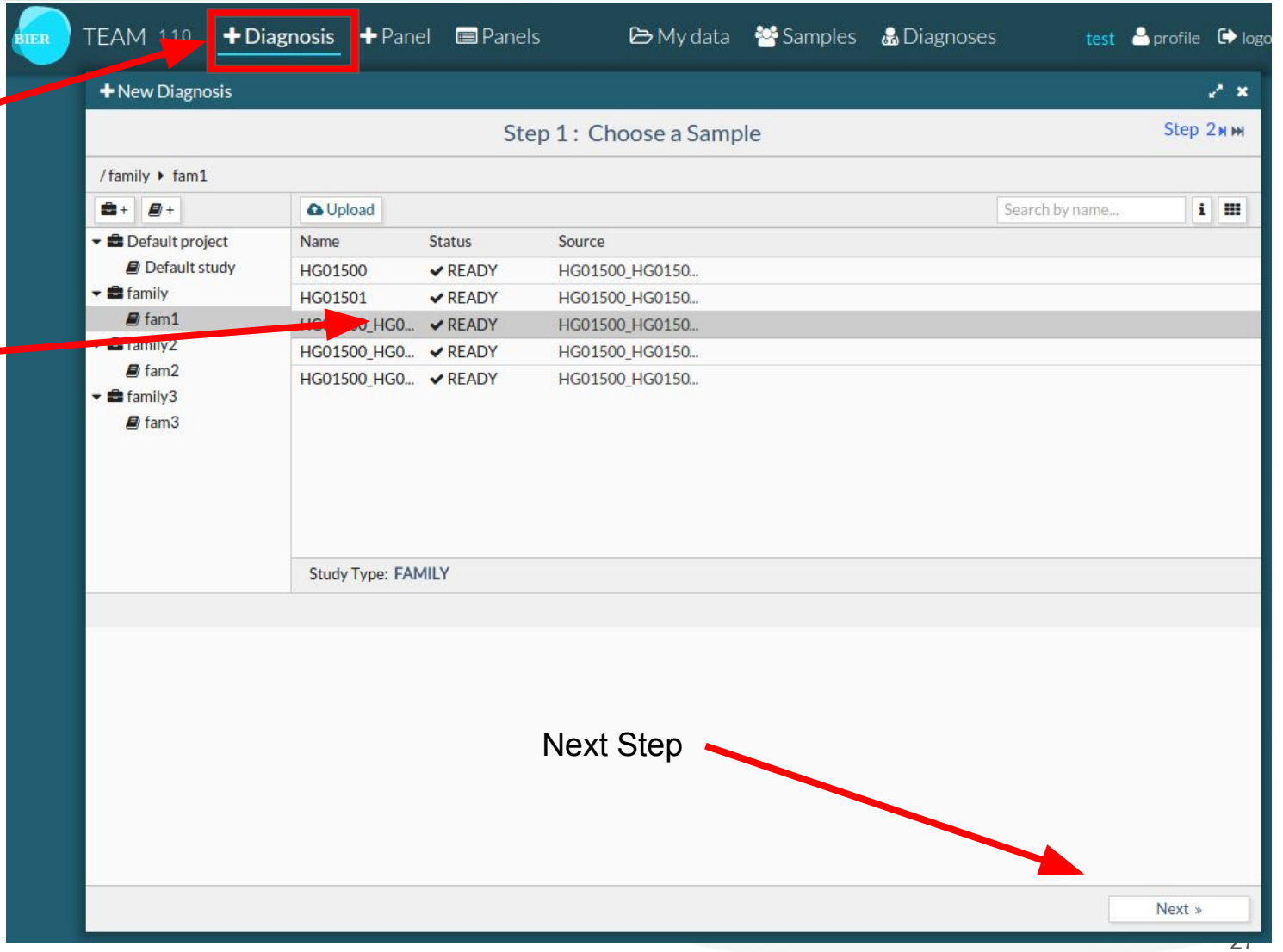
You can manage your panels using “Panels”



New Diagnosis

Click "+ Diagnosis"

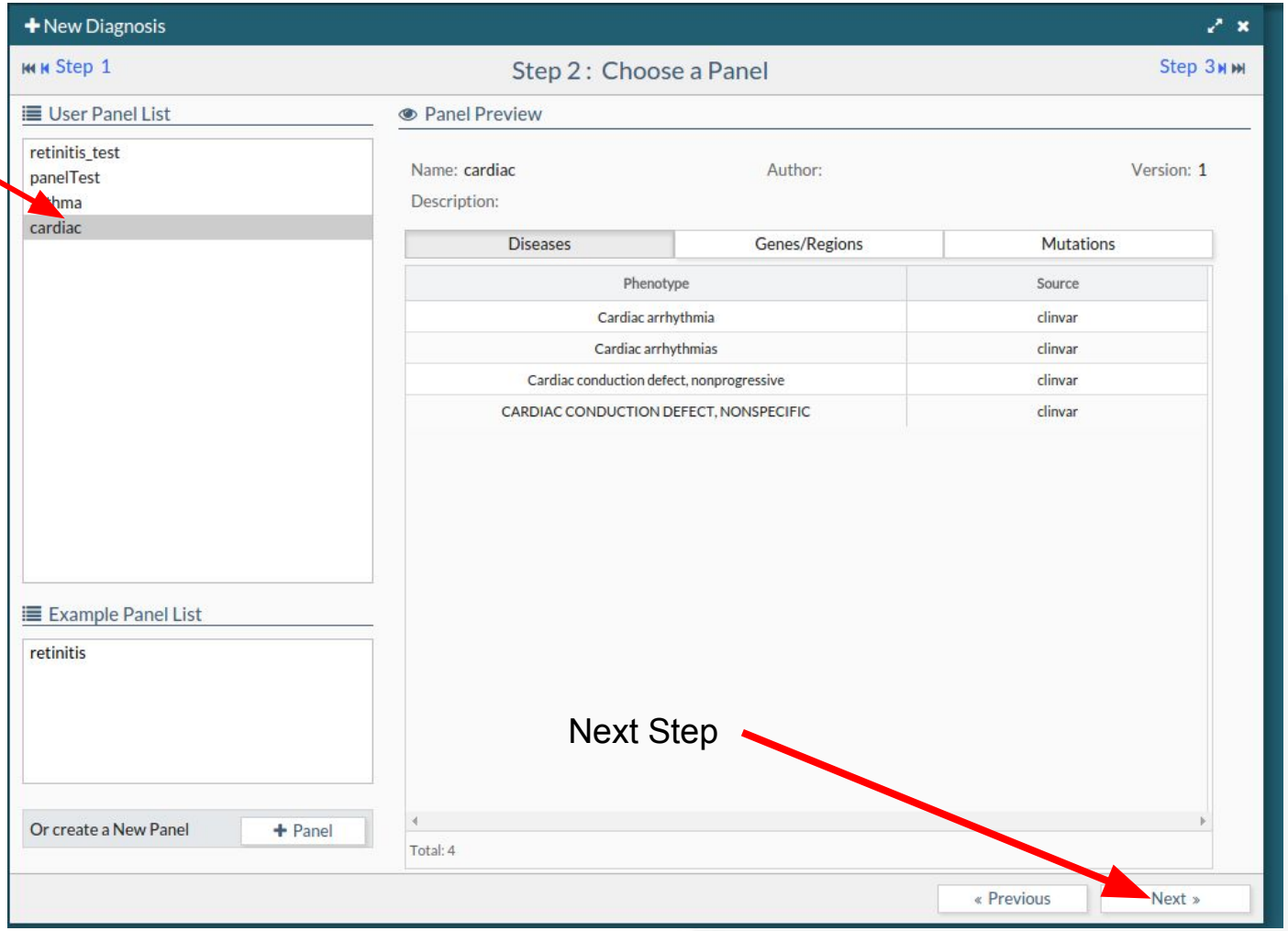
Choose a Study. And then a patient/sample



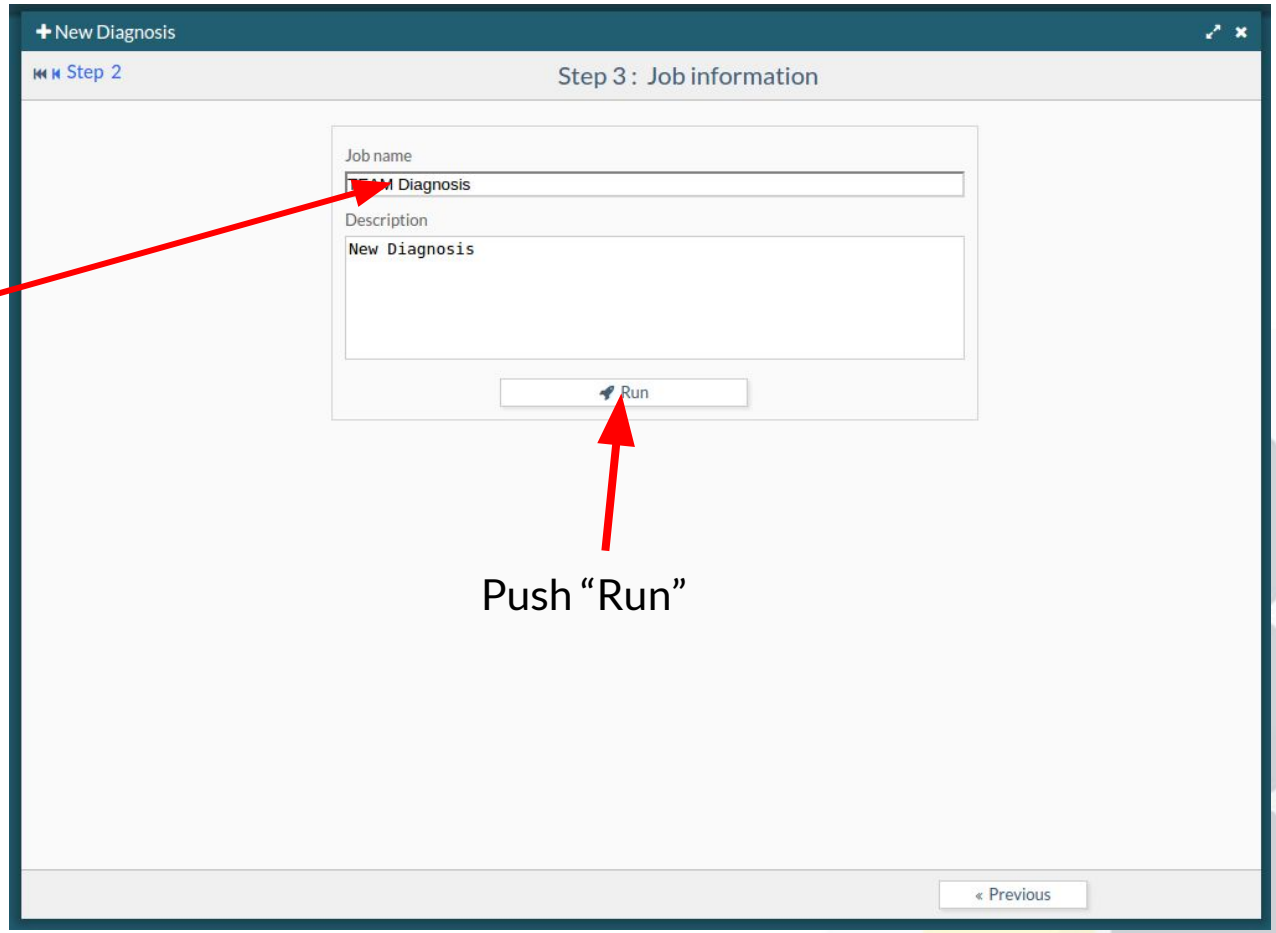
Next Step

New Diagnosis

Select the panel to be used.



New Diagnosis



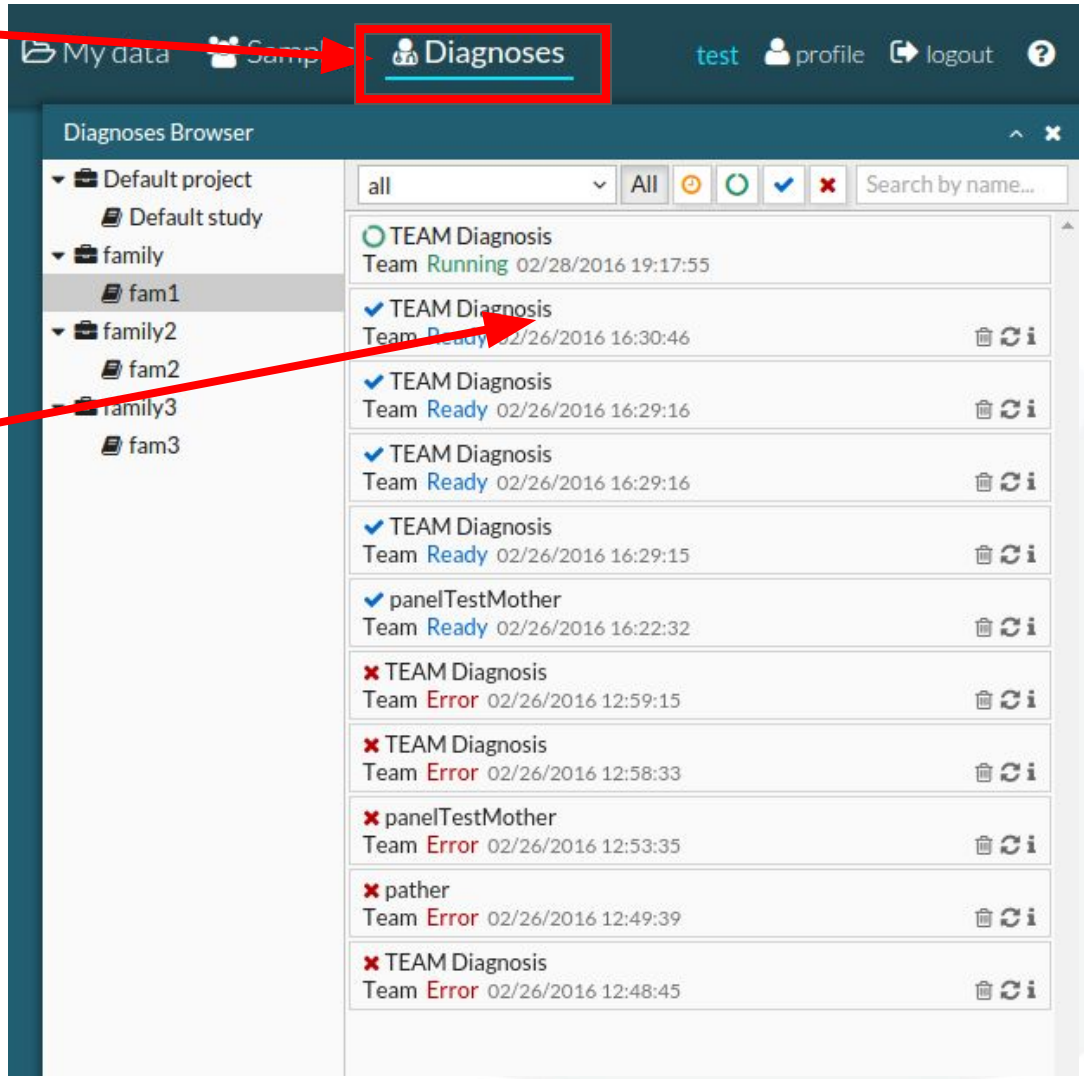
Give a name to the new diagnosis/job

Push “Run”

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.

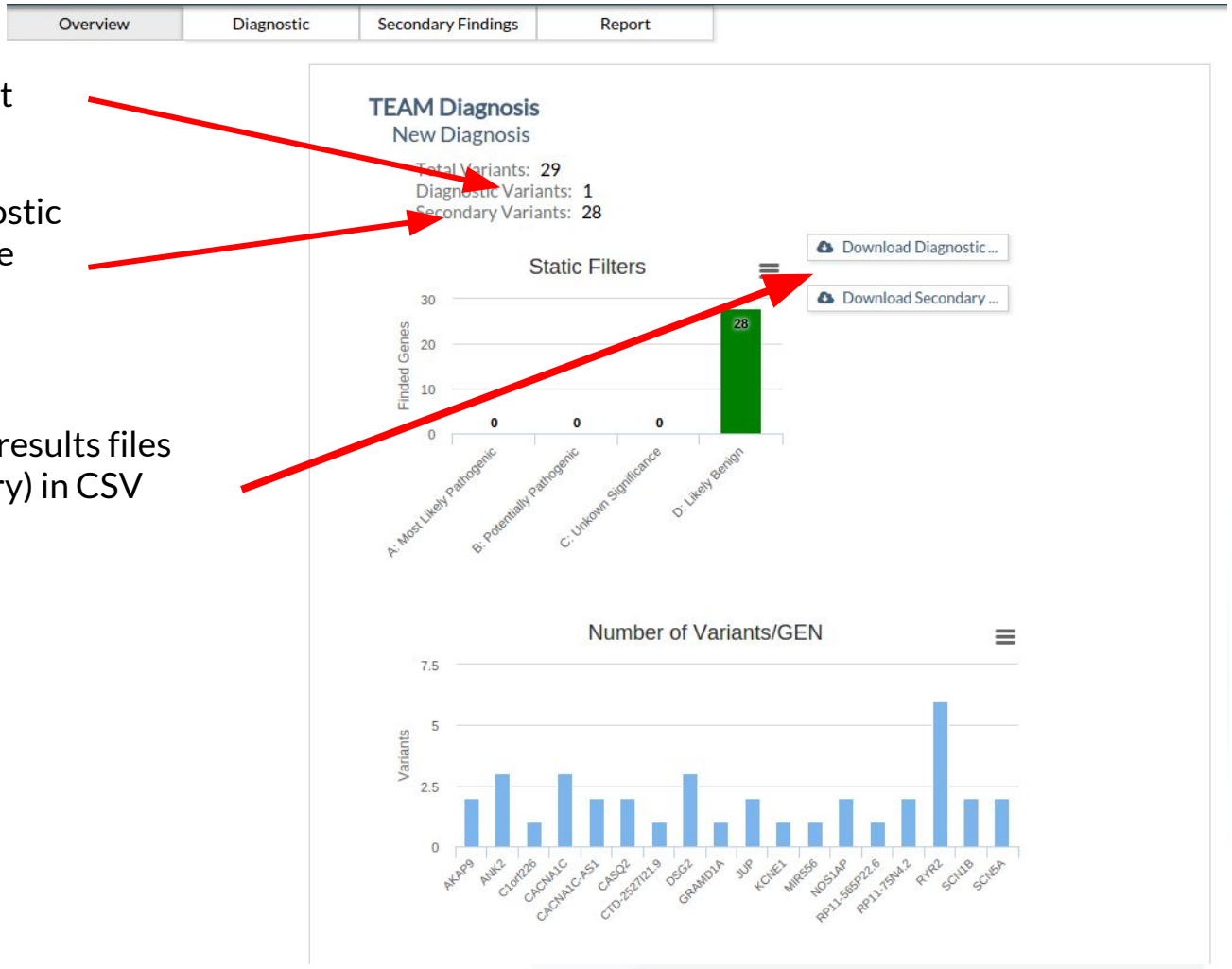


Results: Overview

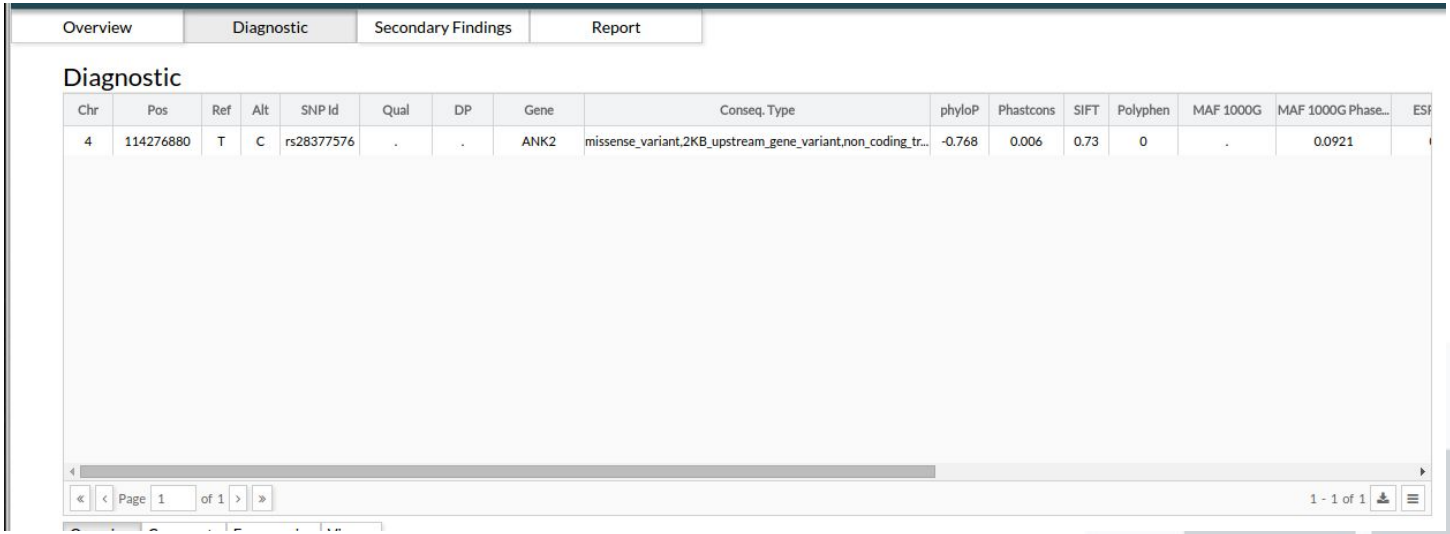
We got one diagnostic variant

Secondary findings: 28
These variants are not diagnostic but are inside the genes of the panel

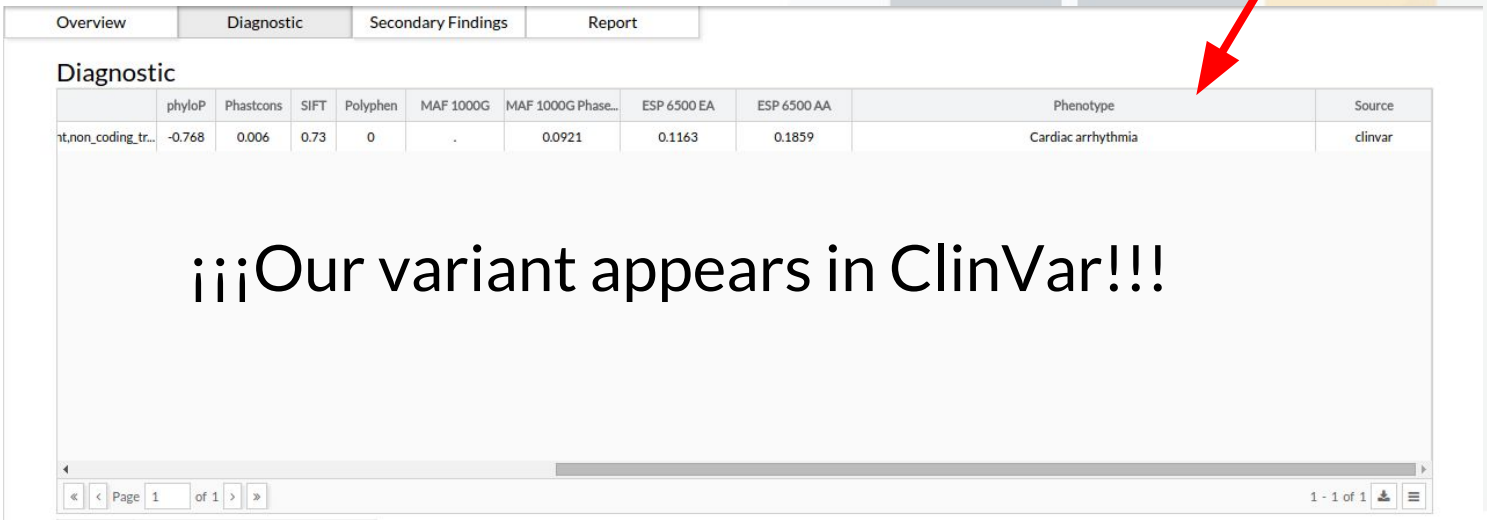
We can download the results files (diagnostic & secondary) in CSV format



Results: Diagnostic Variants

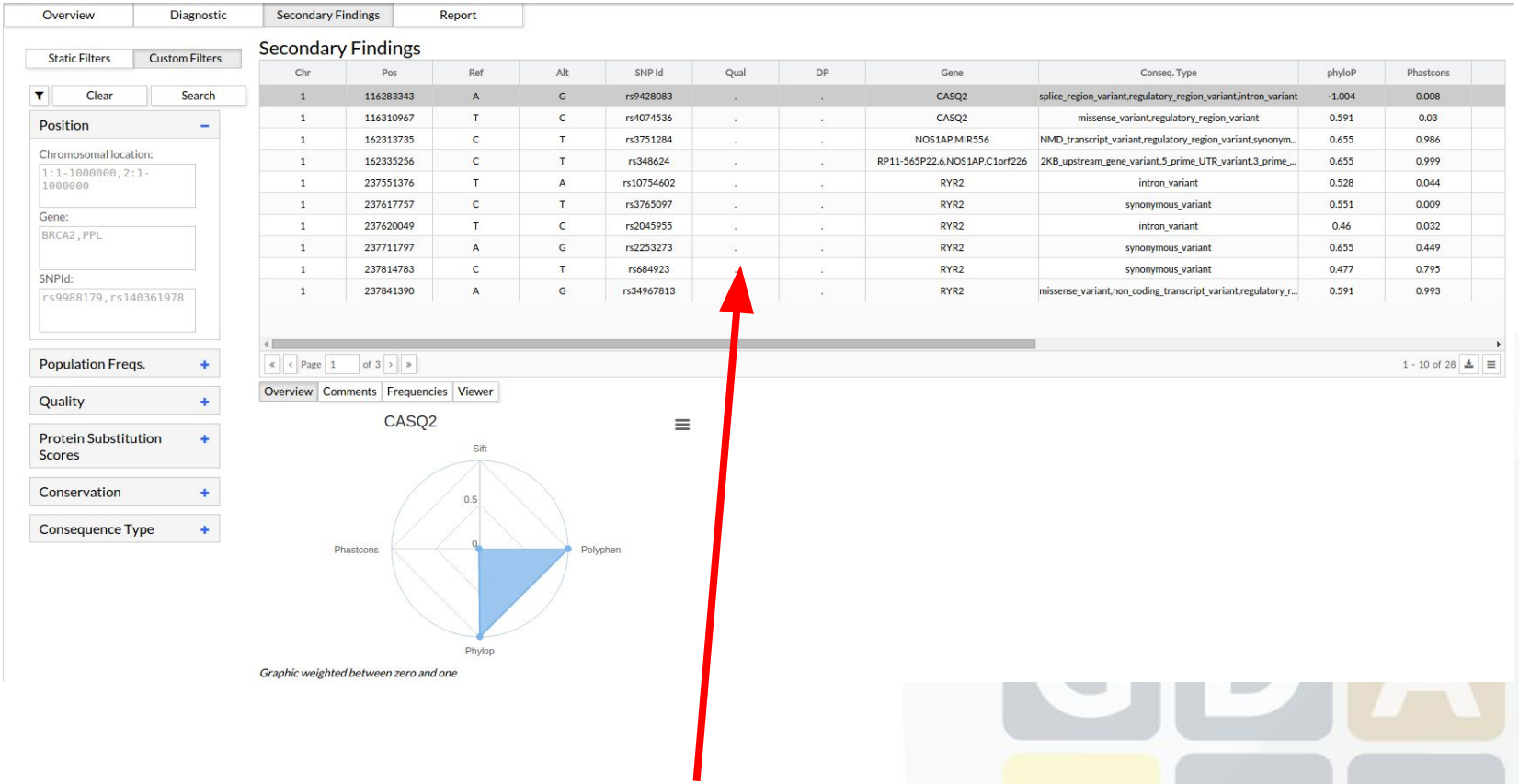


We found a diagnostic variant



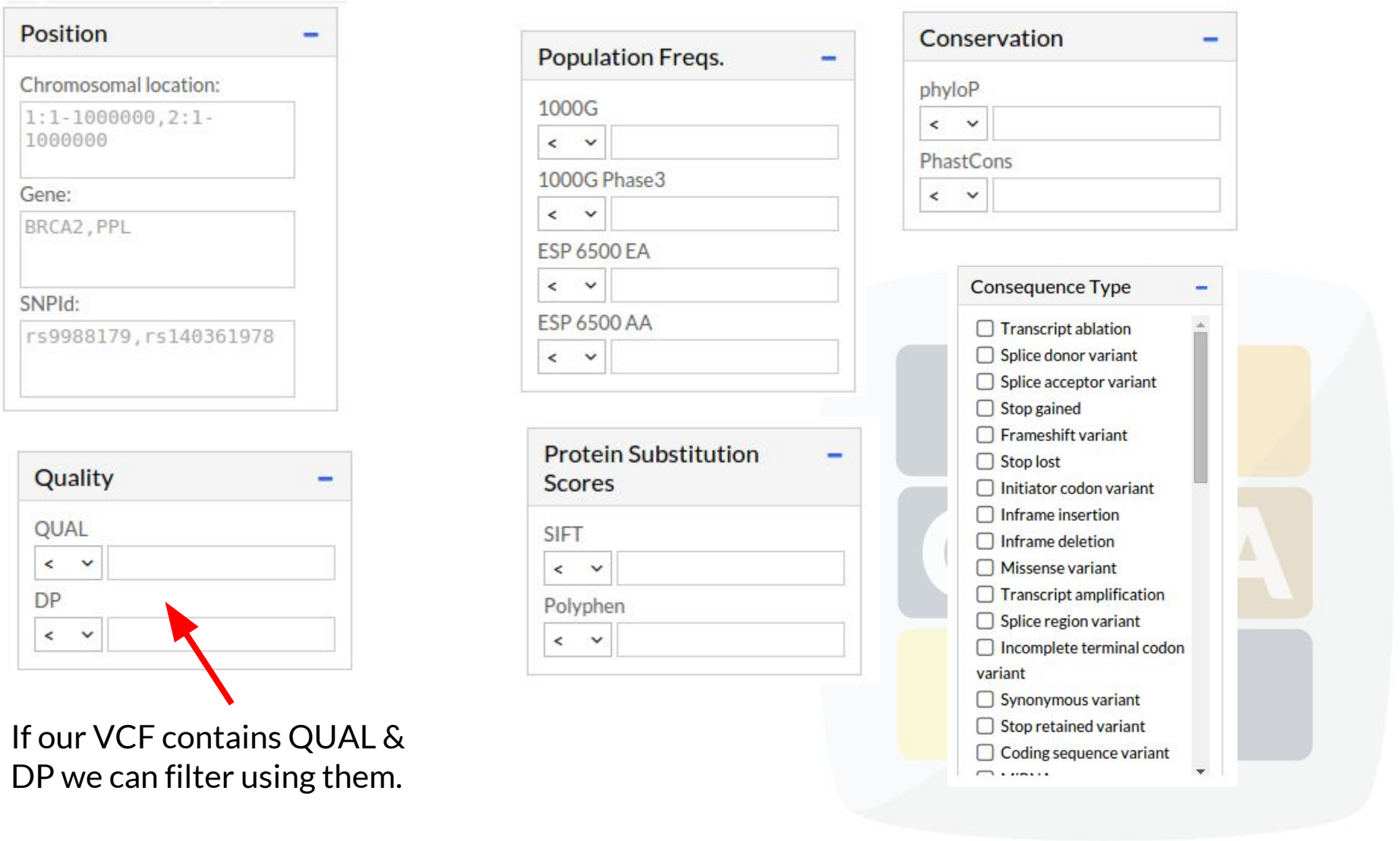
¡¡¡Our variant appears in ClinVar!!!

Results: Secondary Findings



We also found 28 variants within the genes defined in our panel.
Maybe we can discover an interesting variant here.

Results: Secondary Findings (Filters)

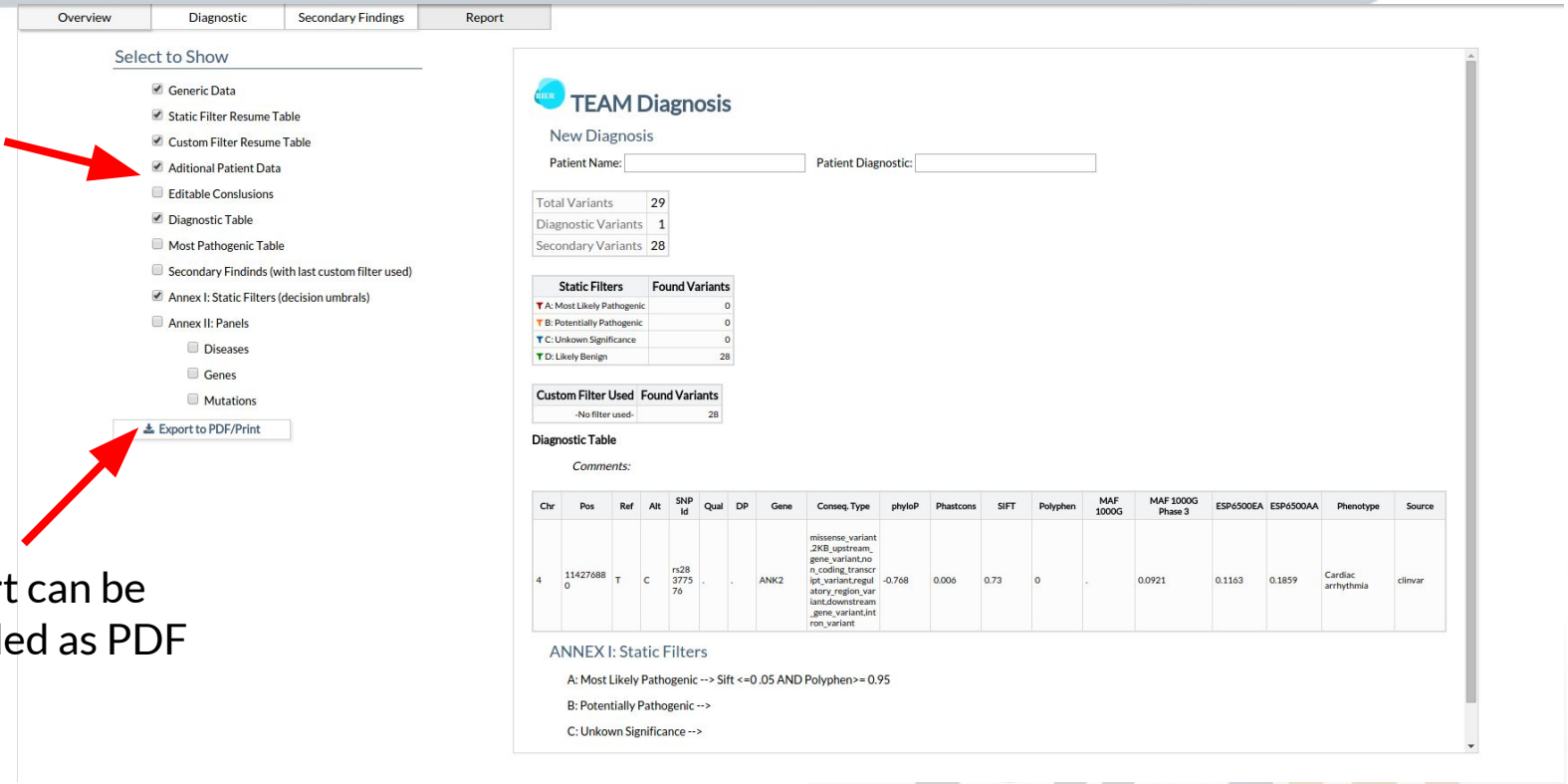


If our VCF contains QUAL & DP we can filter using them.

Results: Report

We can choose which sections we want to print

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