

Variant prioritization

University of Cambridge

Cambridge, UK

10th June 2014

Marta Bleda Latorre

mbleda@cipf.es

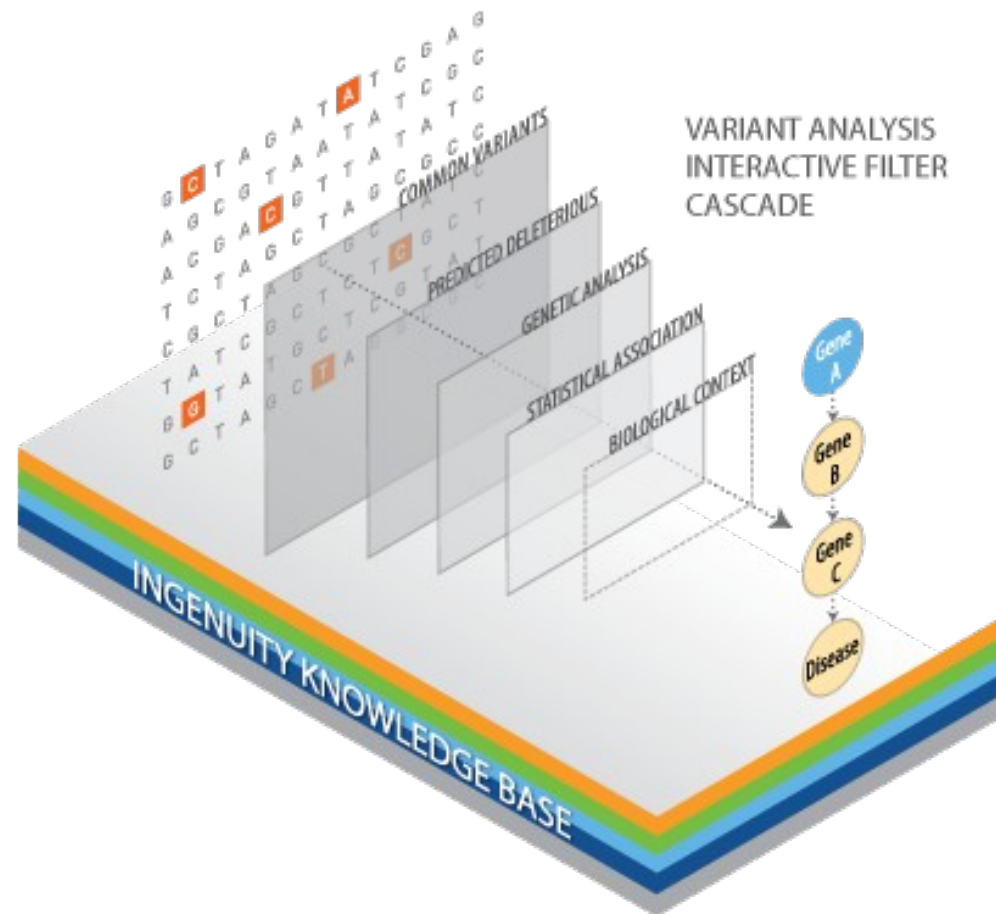
PhD Student at the Computational Genomics Institute

Centro de Investigación Príncipe Felipe (CIPF)

Valencia, Spain

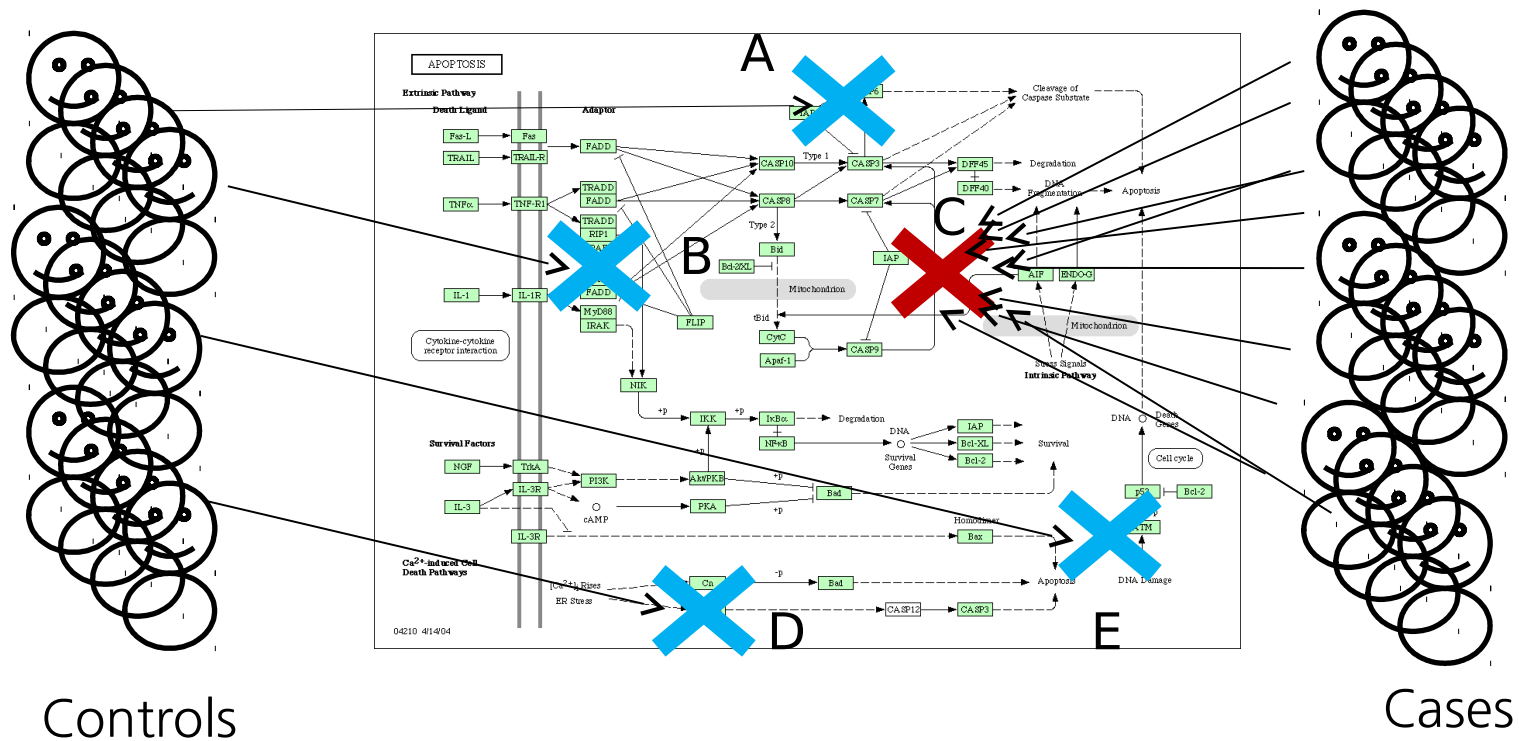


The objective



Finding the mutations causative of diseases

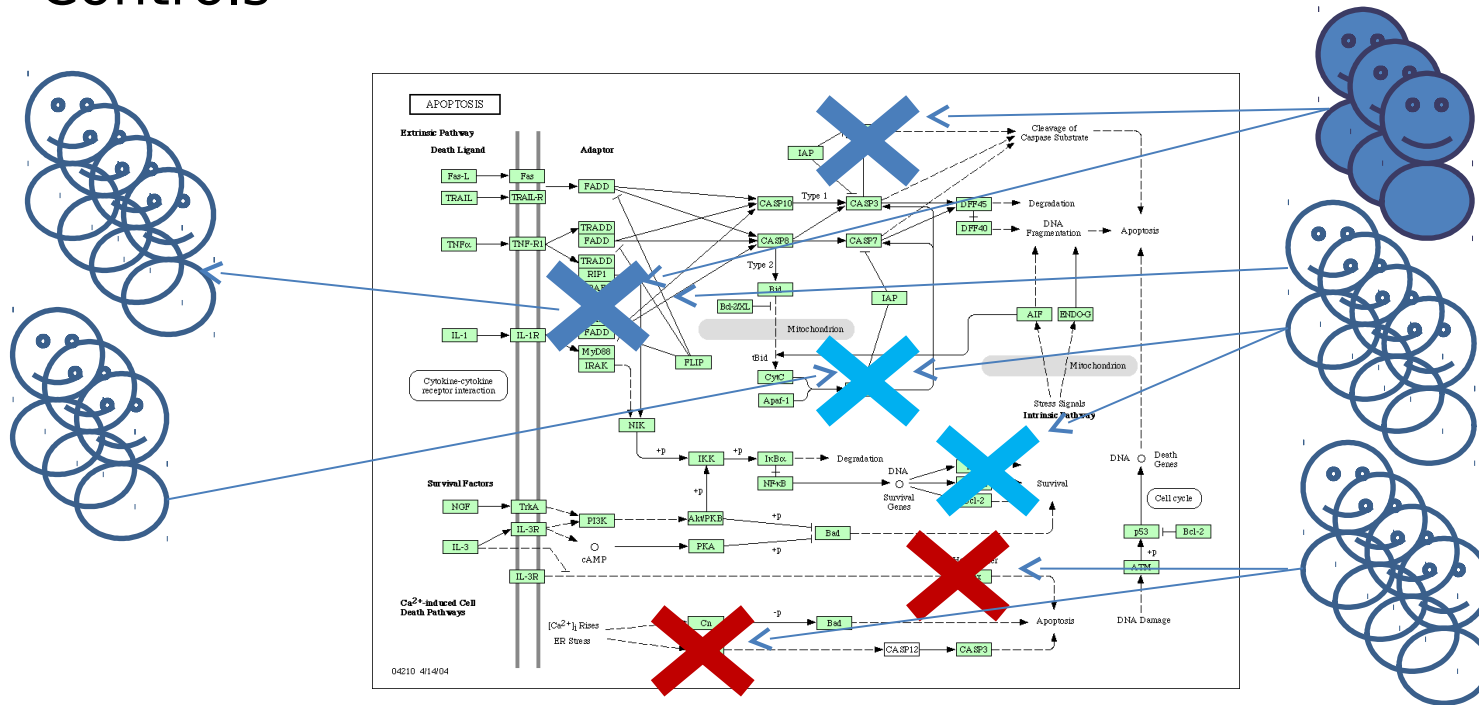
The **simplest case**: monogenic disease due to a single gene



Finding the mutations causative of diseases

Controls

Cases



Clear individual **gene associations** are difficult to find in some diseases

Same phenotype can be due to **different mutations and different genes** (or combinations)

Many cases have to be used to obtain significant associations to many markers

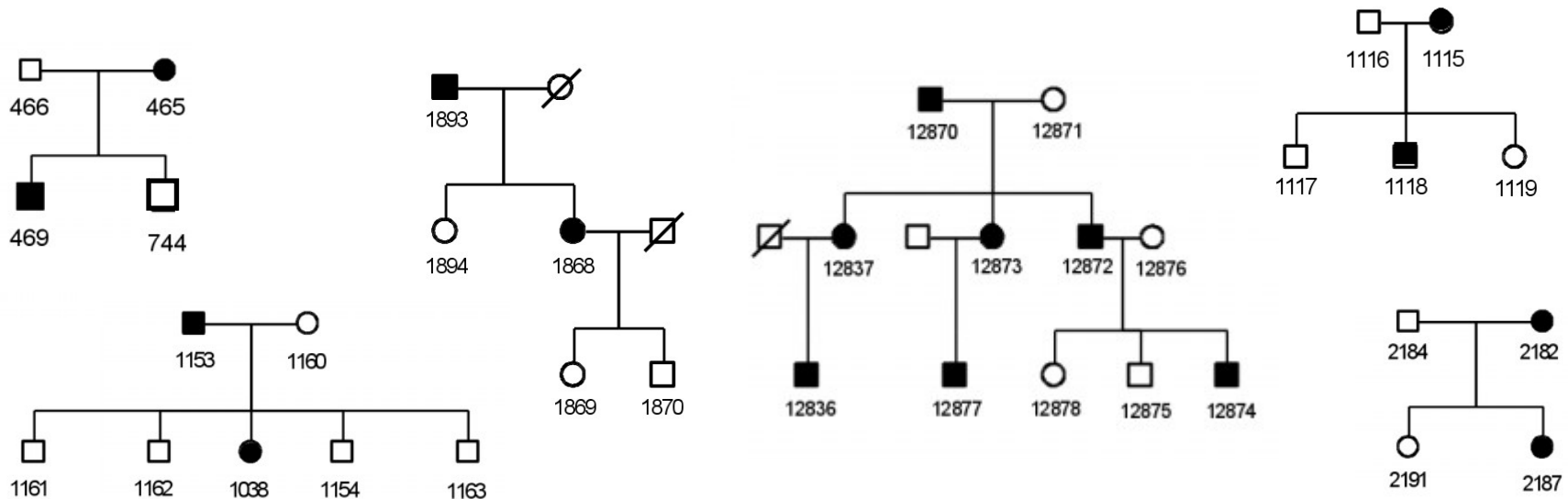
The only common element is the **pathway** (yet unknown) affected

Strategies

- Filtering using **family information**
- **Network** (Systems biology) approaches
 - PPIs
 - Gene regulatory elements (miRNAs, Tfs)
 - GO terms
- GWAS

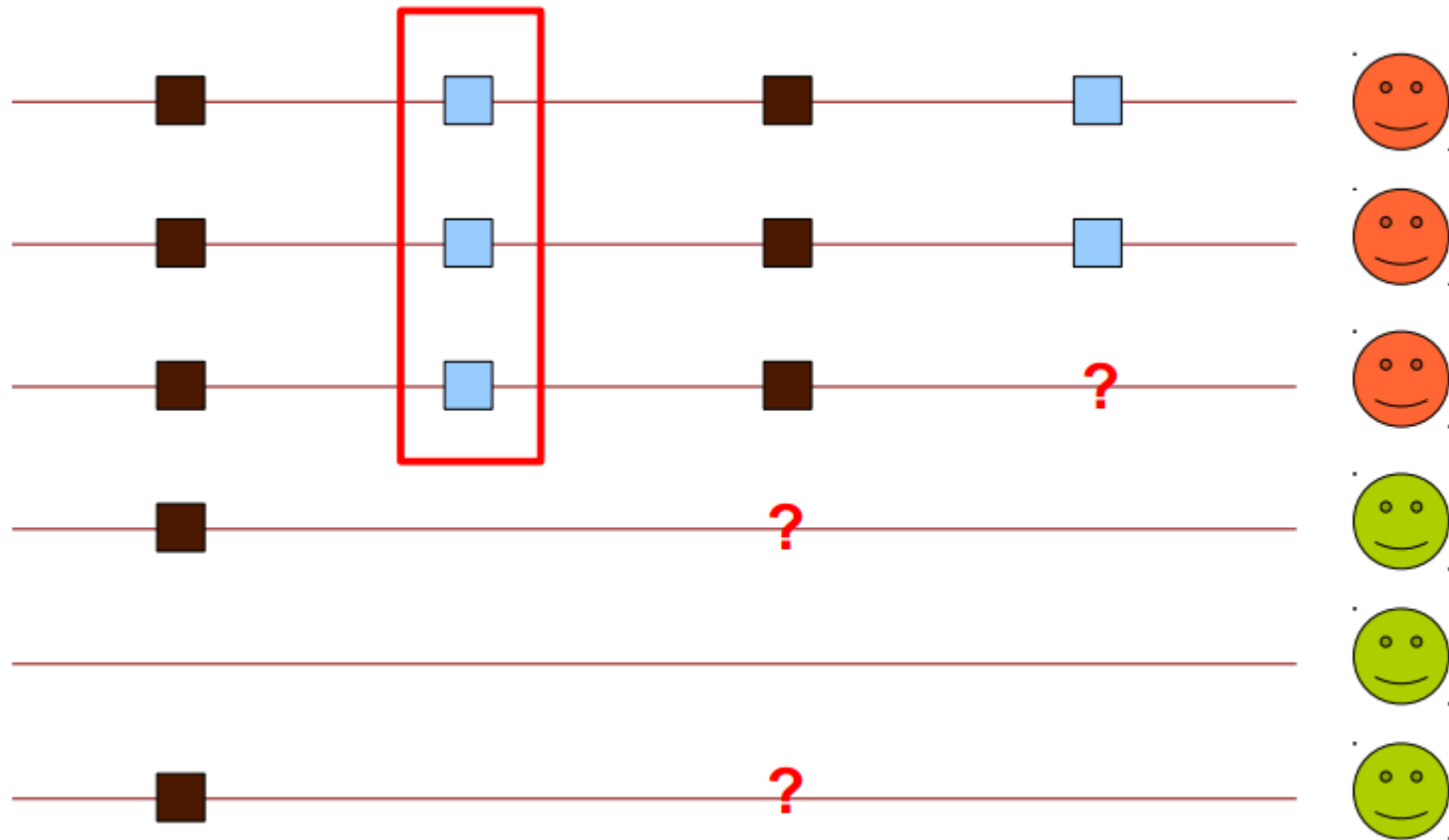
Using family information

- Families containing **control and disease** individuals can help us to **reduce** the number of variants obtained
- Individuals from the same family → **less variability**
- Filter** variants present in healthy people



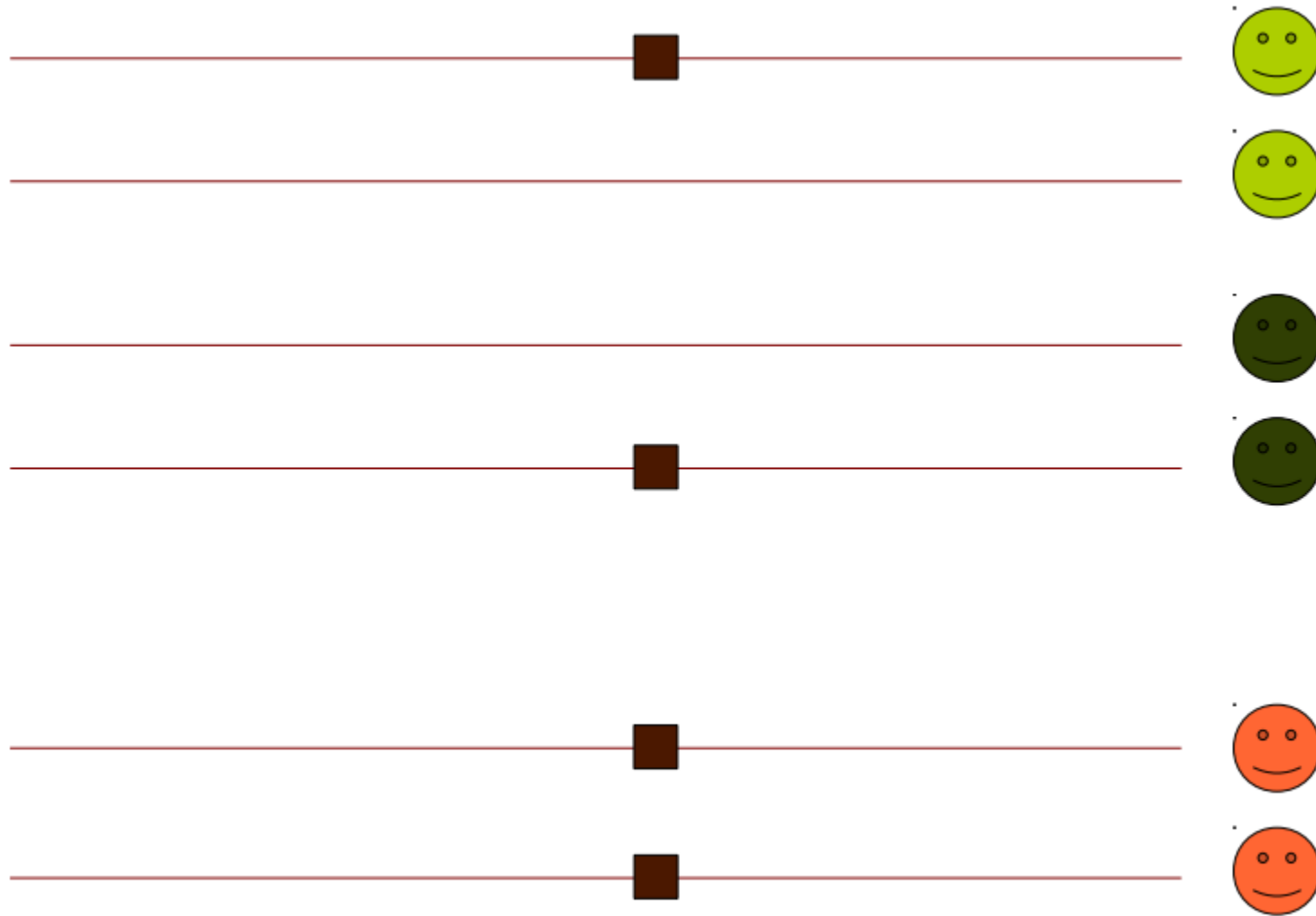
Using family information

Dominant inheritance



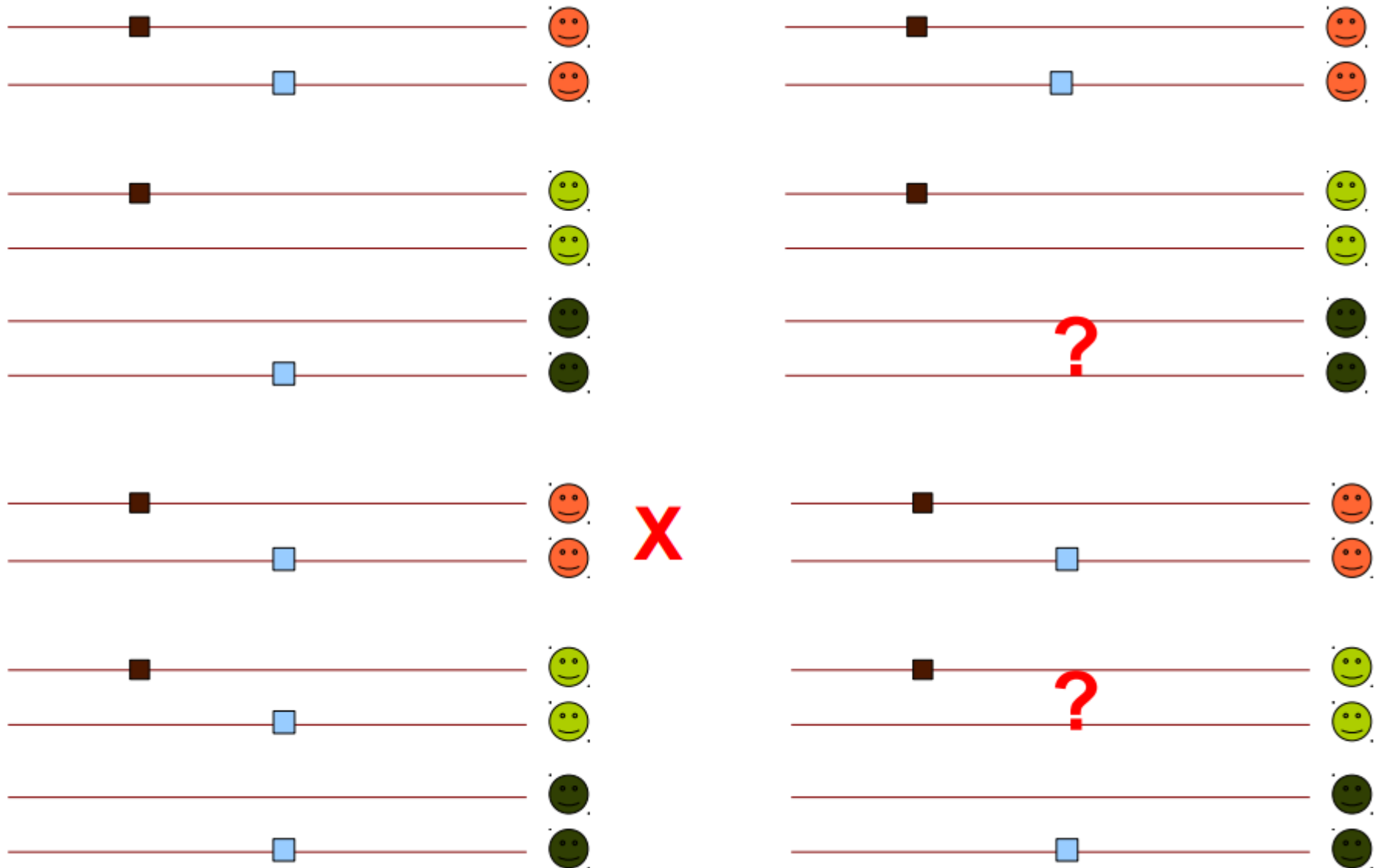
Using family information

Recessive homozygous



Using family information

Recessive - Compound heterozygosity



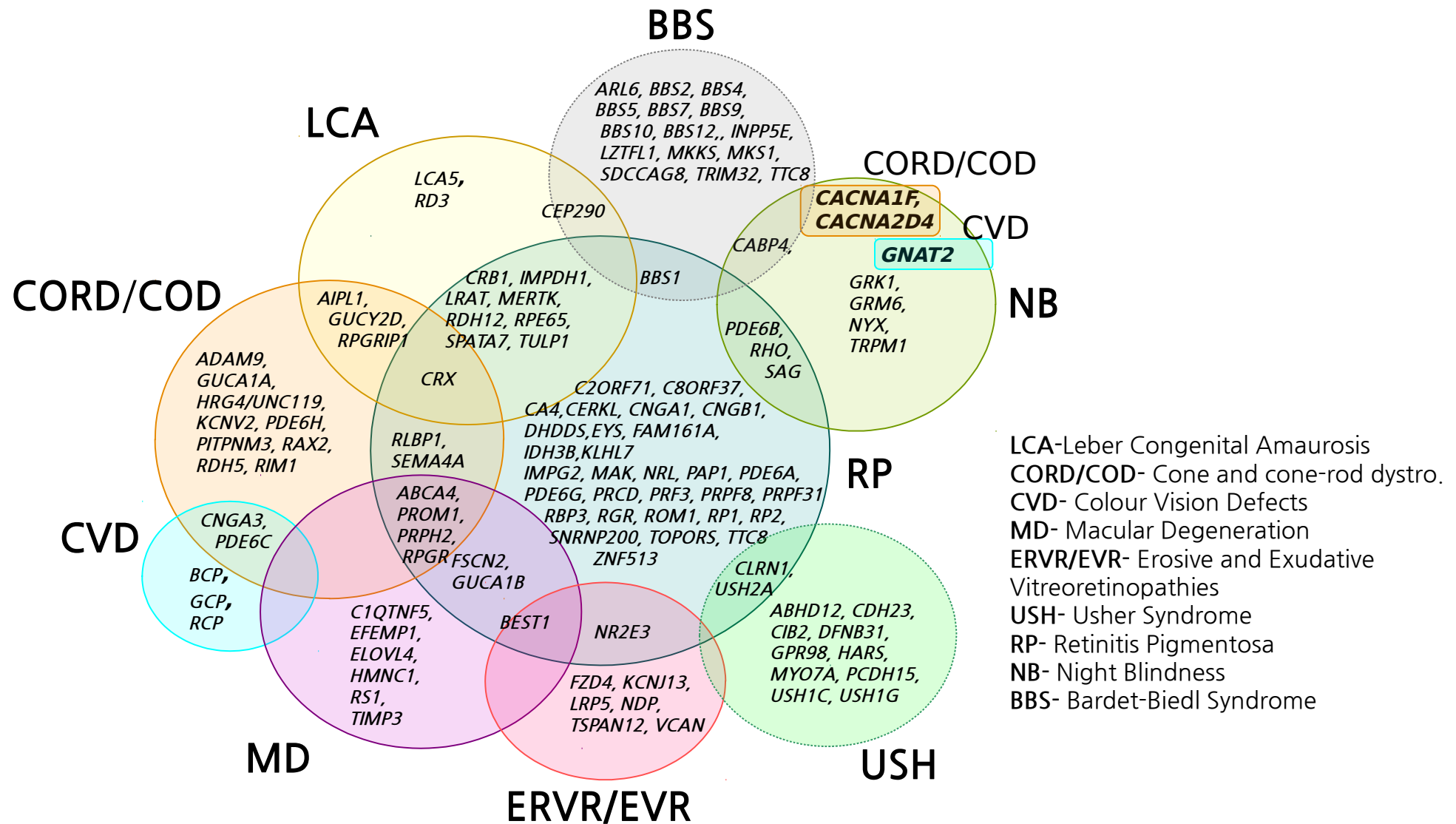


Example with Inherited Retinal Dystrophies (IRD)

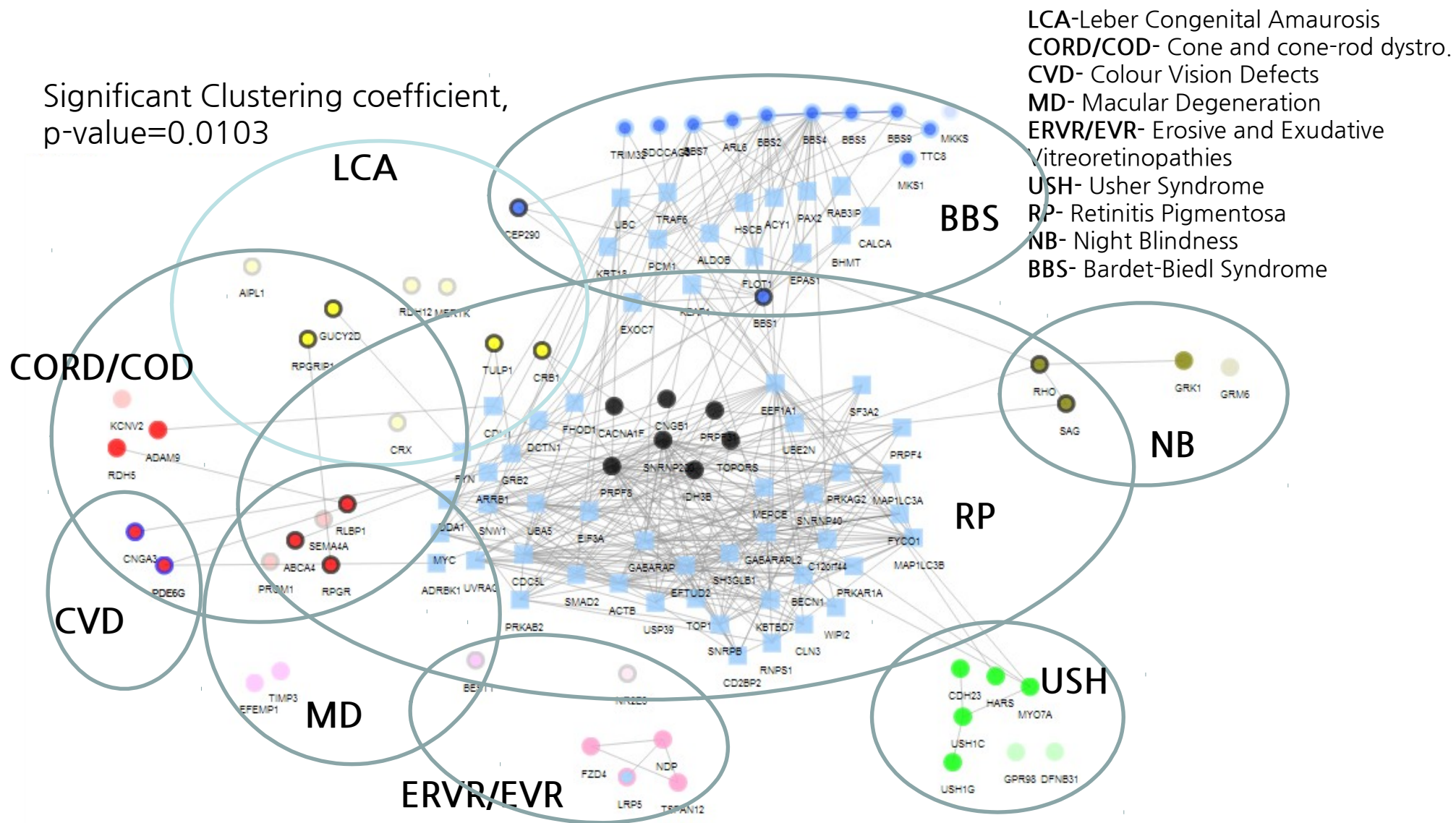
- Prevalence 1 in 3000
- Clinically and genetically very **heterogeneous**
- 190 GENES account for aprox. 50% of IRDs.

Is genetic overlapping among IRDs related to protein interaction?

Example with Inherited Retinal Dystrophies (IRD)



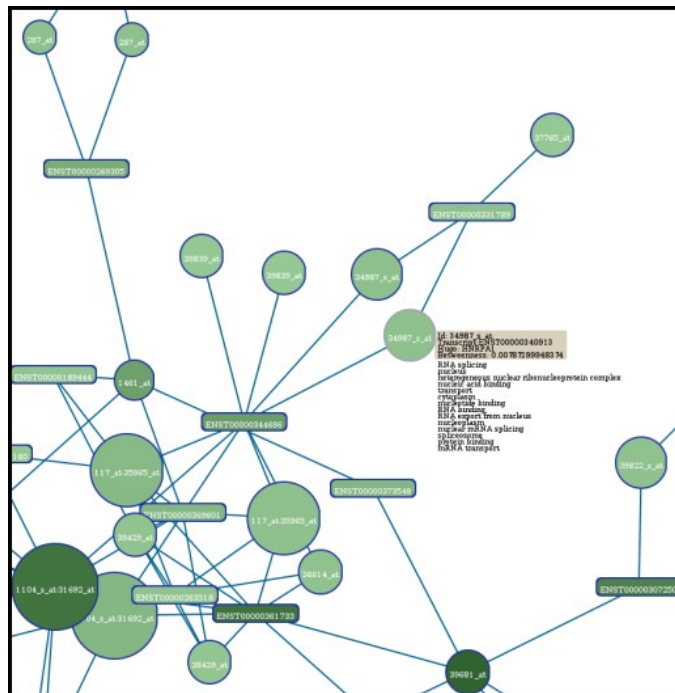
Example with Inherited Retinal Dystrophies (IRD)



SNOW Tool. Minguez et al., NAR 2009 Implemented in Babelomics (<http://www.babelomics.org>)

SNOW

- The SNOW tool introduces **protein-protein interaction data** into the functional profiling of genomic data
 - Evaluates **role of the list within the interactome**: identifies hubs in the list of proteins/genes (nodes) and evaluates the topological parameters of the within the interactome
 - Evaluates the list's cooperative behaviour as a **functional module**



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

NetworkMiner

Prioritizing disease candidate genes

Scenario

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

You have:

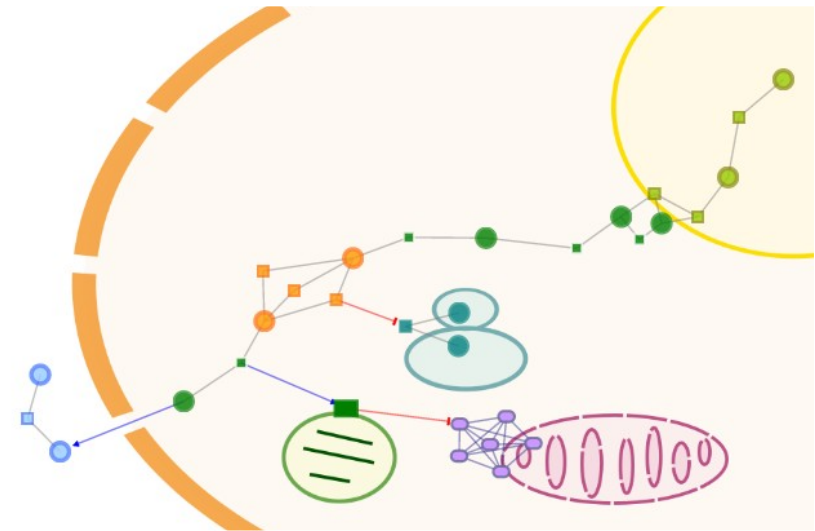
1. a list of disease candidates (ranked by their populational frequency)
2. a list of genes that are known to be associated to the disease

You want to see:

which of your candidates are functionally related or interacting with the known disease genes

NetworkMiner Study

Tests whether any of the candidates is significantly located in the neighborhood of the known disease genes



NetworkMiner

Prioritizing disease candidate genes

Example: Genome-Wide Association Study in Bipolar Disorder

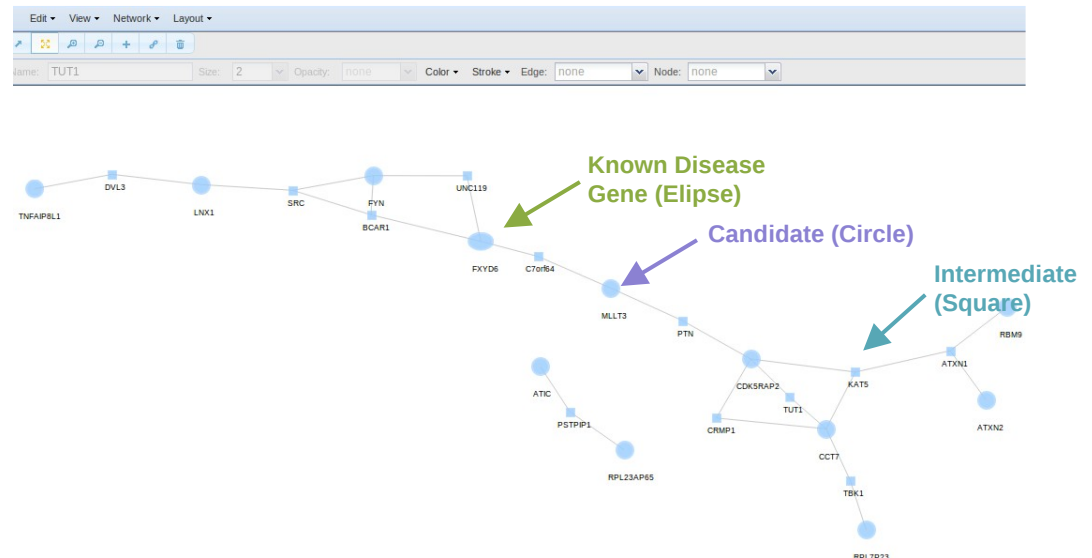
Seed list: Genes associated to Bipolar Disorder

Ranked list: Genes ranked according the association degree in a Case-Control Association Study

Network Miner

▼ Online examples (test the form with example data)

- Example 1: **Genome-Wide Association Study in Bipolar Disorder**
- Example 2: Genes Down-regulated in Fanconi Anemia
- Example 3: Genes Up-regulated in Fanconi Anemia
- Example 4: Essential genes in cancer cell line K562
- Example 5: Essential genes in cancer cell line JURKAT



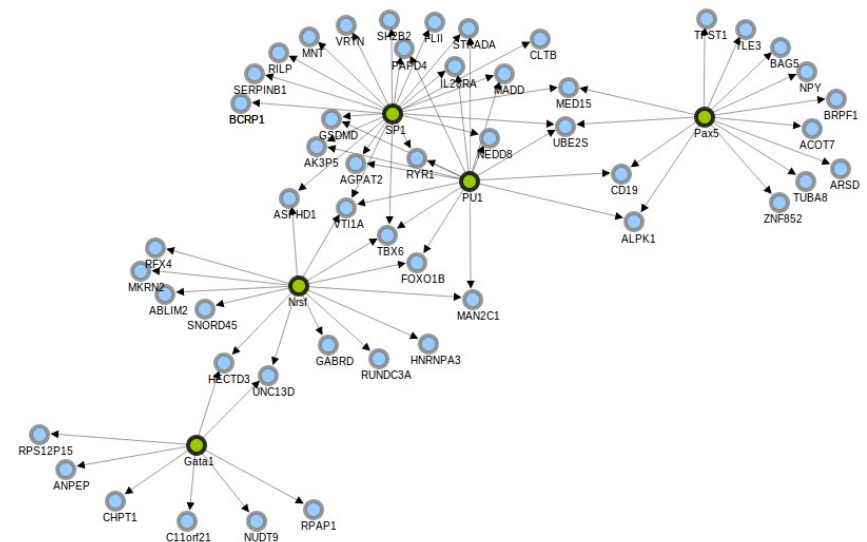
RENATO (REgulatory Network Analysis TOol)

Identifying common regulatory elements

- Sometimes, the problem is not in the gene
- Example: Mutation in the DNA binding domain of a TF
- Tool for the **interpretation and visualization** of transcriptional (TFs) and post-transcriptional (miRNAs) **regulatory information**
- Designed to identify **common regulatory elements** in a list of genes
- RENATO maps these genes to the regulatory network, extracts the corresponding regulatory connections and evaluate each regulator for **significant over-representation** in the list.

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

Term	List1 annotated	List1 unannotated	List2 annotated	List2 unannotated	Odds ratio (log e)	pvalue	Adjusted pvalue
SP1	22	39	1178	17240	2.1108949872	1.19202e-11	3.09925e-10
Pax5	13	48	343	18075	2.6583029464	1.15497e-10	1.50146e-9
Nrsf	14	47	641	17777	2.1115410369	2.07537e-8	1.79865e-7
Gata1	8	53	186	18232	2.6943365253	2.33672e-7	0.00000151887
PU1	16	45	2230	16188	0.94819487401	0.00204712	0.010645



THANK YOU.