

# Variant prioritization

**University of Cambridge**

Cambridge, UK

30<sup>th</sup> September 2014



**Marta Bleda Latorre**

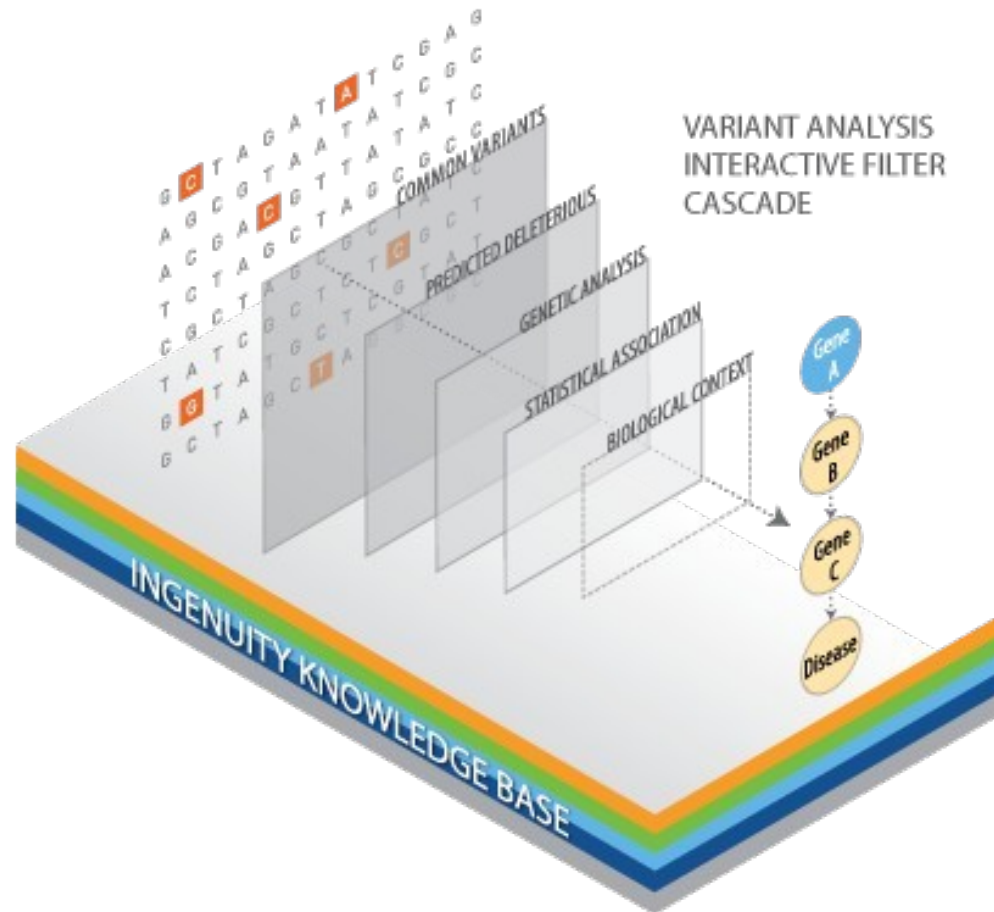
[mb2033@cam.ac.uk](mailto:mb2033@cam.ac.uk)

Research Assistant at the Department of Medicine

University of Cambridge

Cambridge, UK

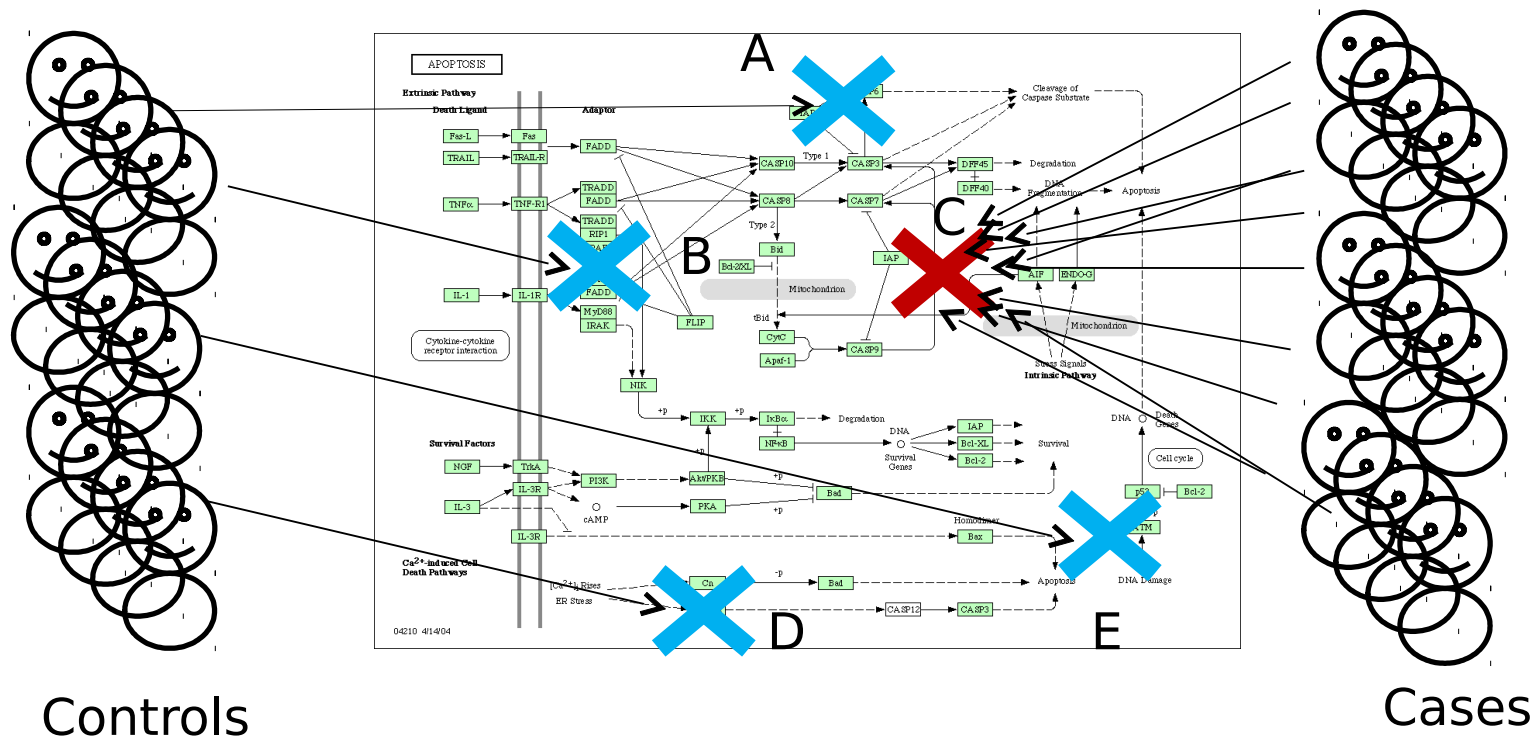
# The objective



# And now what?

## Finding the mutations causative of diseases

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

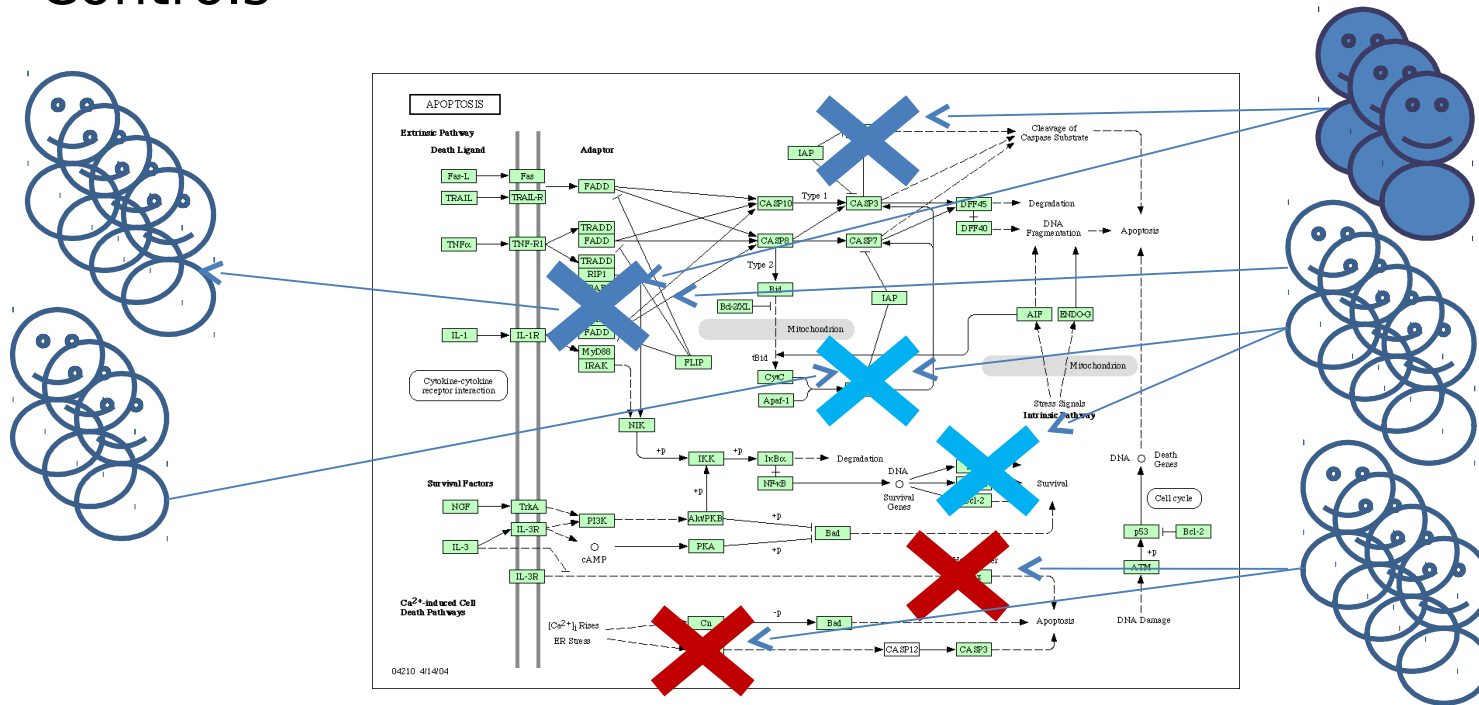


# And now what?

## 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

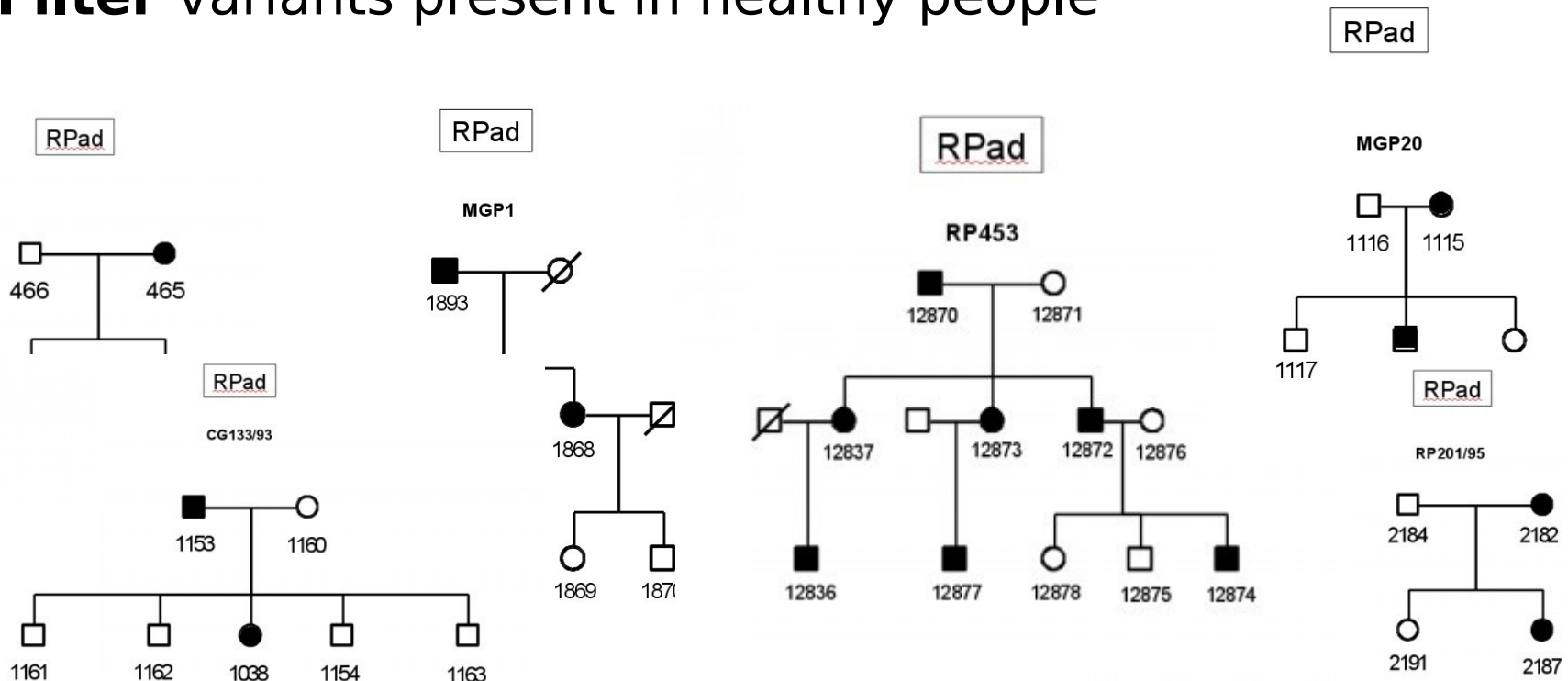
# Strategies

---

- Filtering using **family information**
- **Network** (Systems biology) approaches
  - PPIs
  - Gene regulatory elements (miRNAs, Tfs)
  - GO terms
- GWAS
- Burden tests for rare variants....

# Using family information

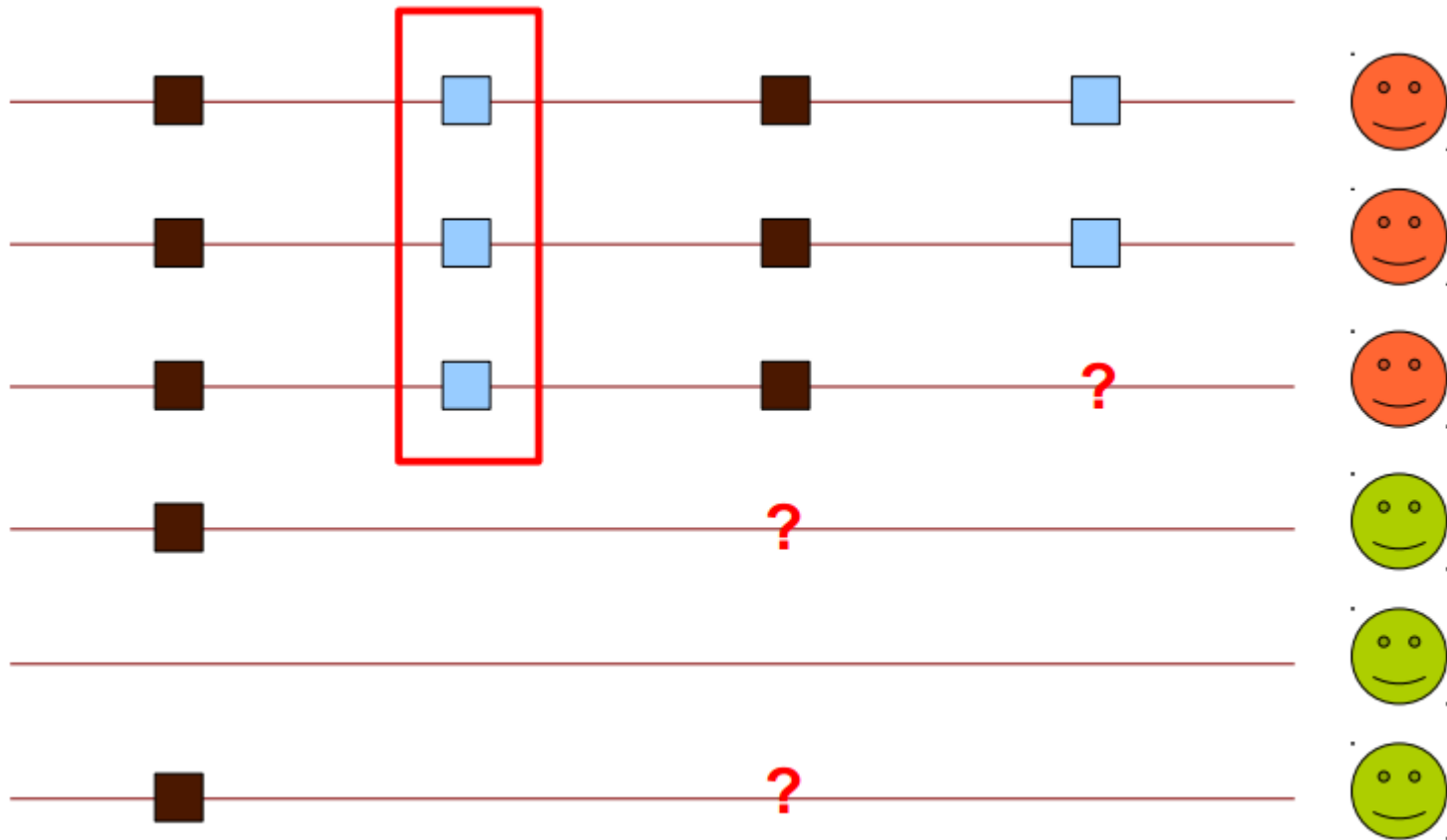
- Families containing **control and disease** individuals can help us **reduce** the number of variants obtained
- Individuals from the same family → **less variability**
- Filter** variants present in healthy people



# Using family information

## Dominant inheritance

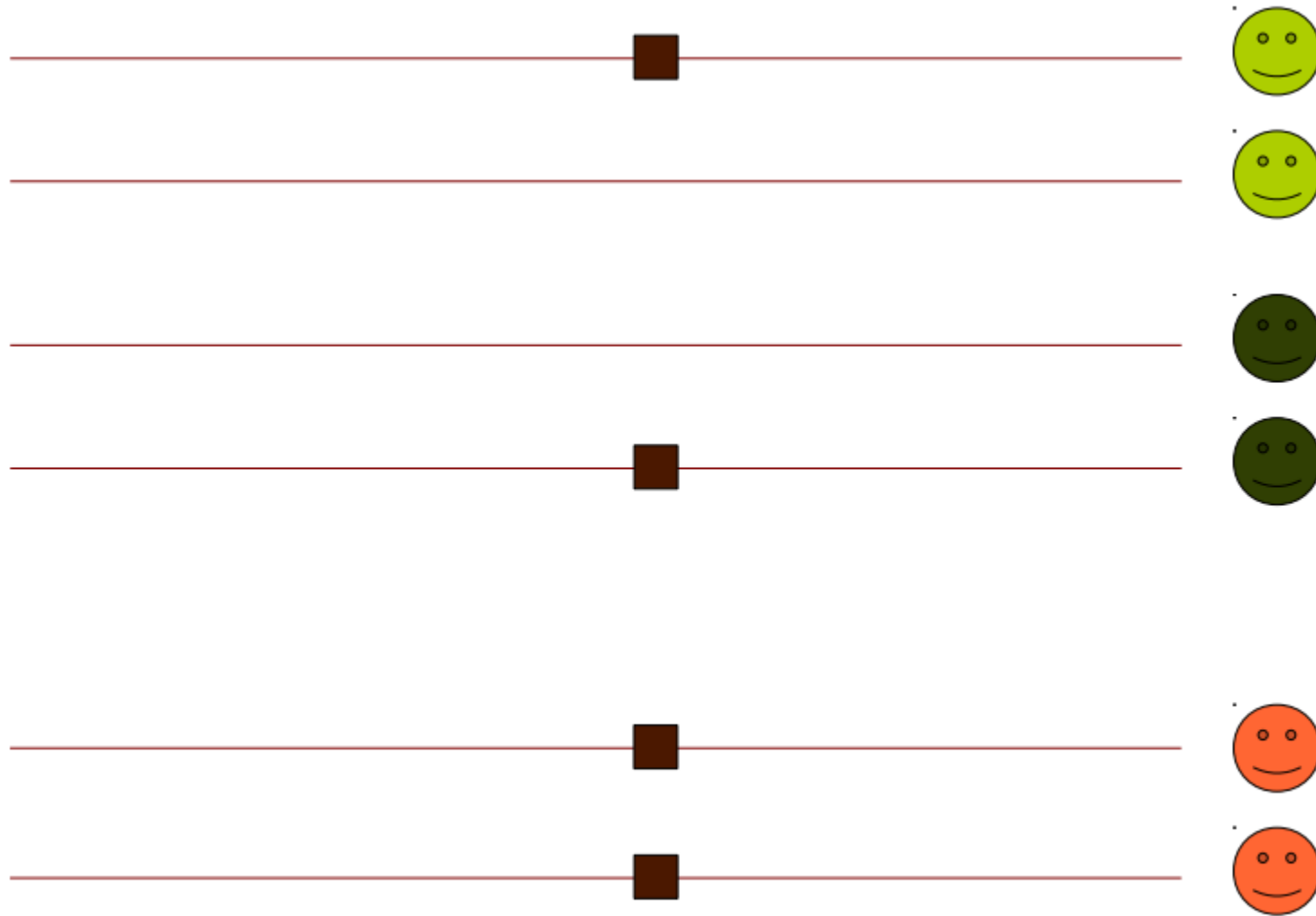
---



# Using family information

## Recessive homozygous

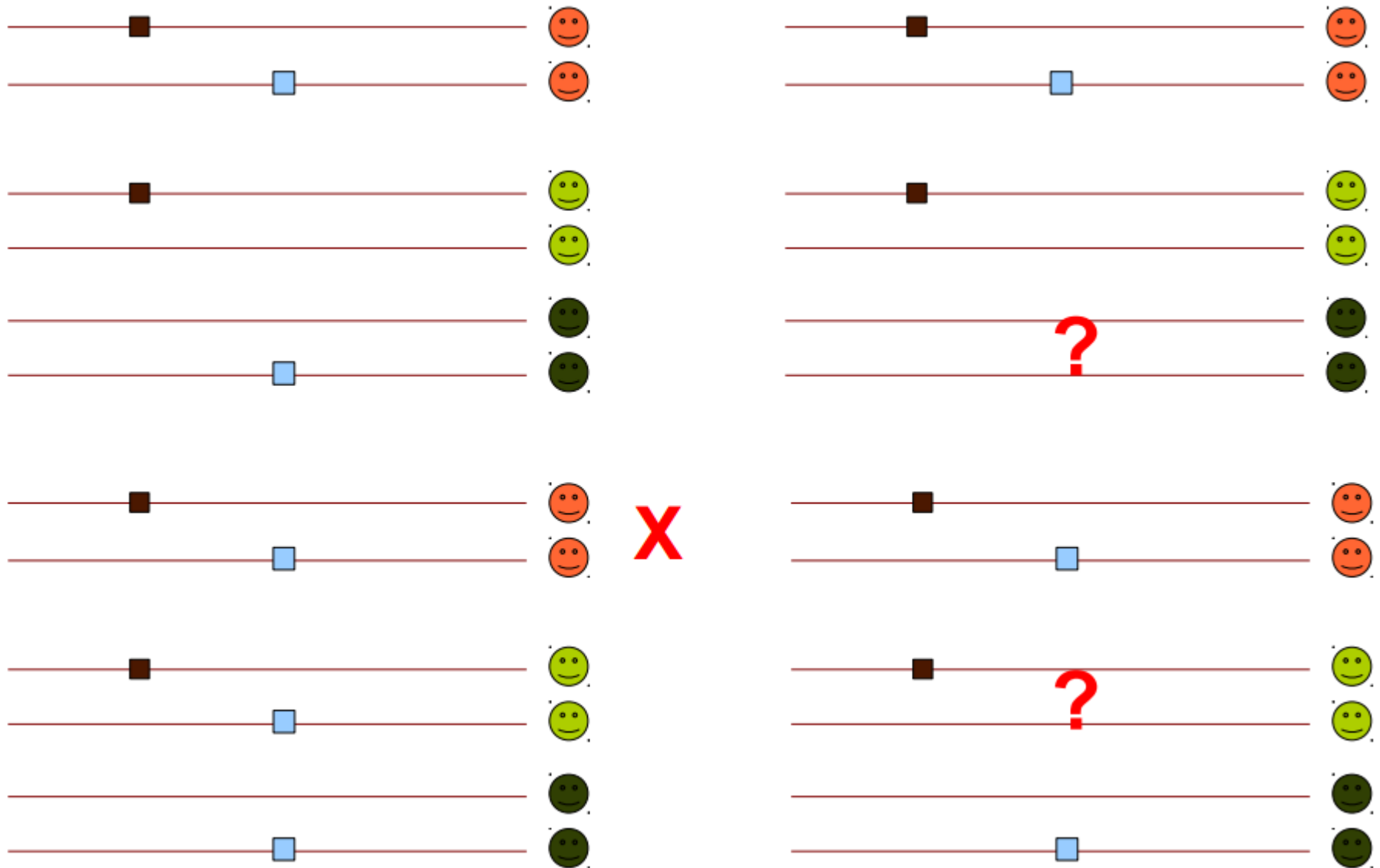
---





# Using family information

## Recessive - Compound heterozygosity



Variant analysis tool beta

logged in as ayuso Upload & Manage profile logout

home documentation tutorial about

Show jobs

Preprocess Analysis Visualization

Home RP-0859

Summary Variants and effect Genome Viewer

Filters

Reload Clear Search

Region +

Gene +

Stats +

Samples -

K529: 

☒

☒

☐

0/0 0/1 1/1

D056: 

☐

☐

☒

0/0 0/1 1/1

Controls +

Effect +

Variant Info

Variant	Alleles	Samples		SNP id	Controls (MAF)			Consq. Type	Polyphen	Sift	Conservation
		K529	D056		1000G	BIER	EVS				
gene_name: ACTR5 (1 Item)											
20:37396120	A>G	0/1	1/1	rs2245231	0.4231 (G)	0.4667 (G)	0.4416 (A)	exon_variant,non_synon...			
gene_name: ANKRD60 (1 Item)											
20:56807969	A>G	0/1	1/1	rs3818744	0.4785 (G)	0.4267 (A)	.	5KB_upstream_variant,e...			
gene_name: AURKA (1 Item)											
20:54961463	T>C	0/1	1/1	rs1047972	0.1557 (T)	0.2333 (T)	0.1622 (C)	exon_variant,non_synon...			
gene_name: BIRC7 (1 Item)											
20:61869826	C>T	0/1	1/1	rs2273487	0.4675 (T)	0.48 (C)	0.4343 (C)	exon_variant,DNAseI_h...			
26 variants											

Effect - 20:37396120 A>G

	Position chr:start:end	snp id	Samples	Consequence Type	Aminoacid Change	gene (Ensemblid)	transcript Id	feature Id	feature Name	feature Type	feature Biotype
featureId: ENSE00000844678 (3 Items)											
1	20:37396107-373...			exon_variant (SO:0001791)	.	ACTR5 (ENSG00...	ENST00000243903	ENSE00000844678	ACTR5	exon	protein_coding
2	20:37396107-373...			coding_sequence_variant (SO:000...	.	ACTR5 (ENSG00...	ENST00000243903	ENSE00000844678	ACTR5	exon	protein_coding
3	20:37396107-373...			non_synonymous_codon (SO:000...	IV - ATT/GTT (483)	ACTR5 (ENSG00...	ENST00000243903	ENSE00000844678	ACTR5	exon	protein_coding
featureId: H3K36me3 (18 Items)											
4	20:37377900-374...			regulatory_region_variant (SO:000...	.	.	.	H3K36me3	.	regulatory_region	.
5	20:37378100-374...			regulatory_region_variant (SO:000...	.	.	.	H3K36me3	.	regulatory_region	.
6	20:37378300-374...			regulatory_region_variant (SO:000...	.	.	.	H3K36me3	.	regulatory_region	.
7	20:37378450-374...			regulatory_region_variant (SO:000...	.	.	.	H3K36me3	.	regulatory_region	.
8	20:37382500-374...			regulatory_region_variant (SO:000...	.	.	.	H3K36me3	.	regulatory_region	.
9	20:37382650-374...			regulatory_region_variant (SO:000...	.	.	.	H3K36me3	.	regulatory_region	.
21 effects											



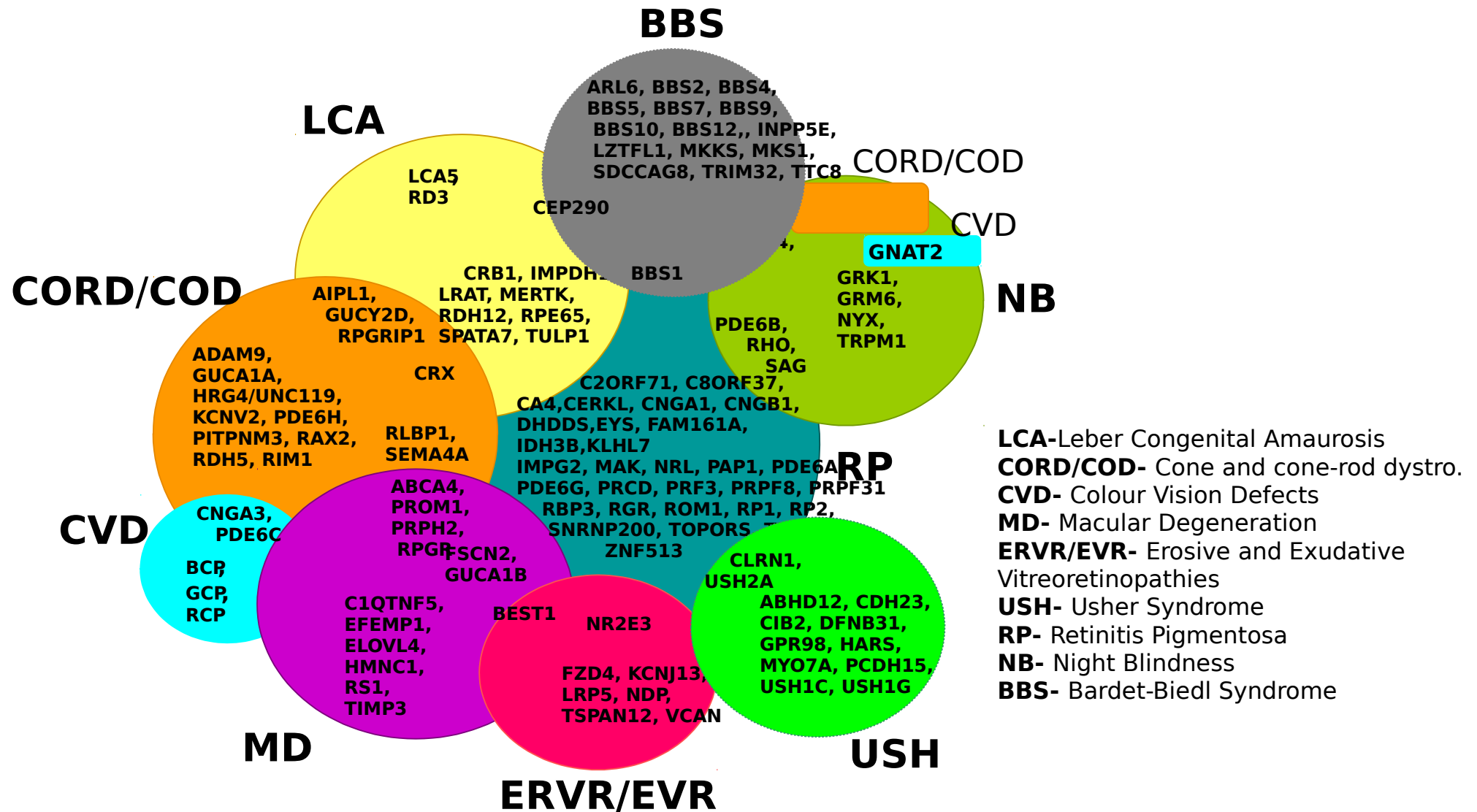
# Example with Inherited Retinal Dystrophies

---

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

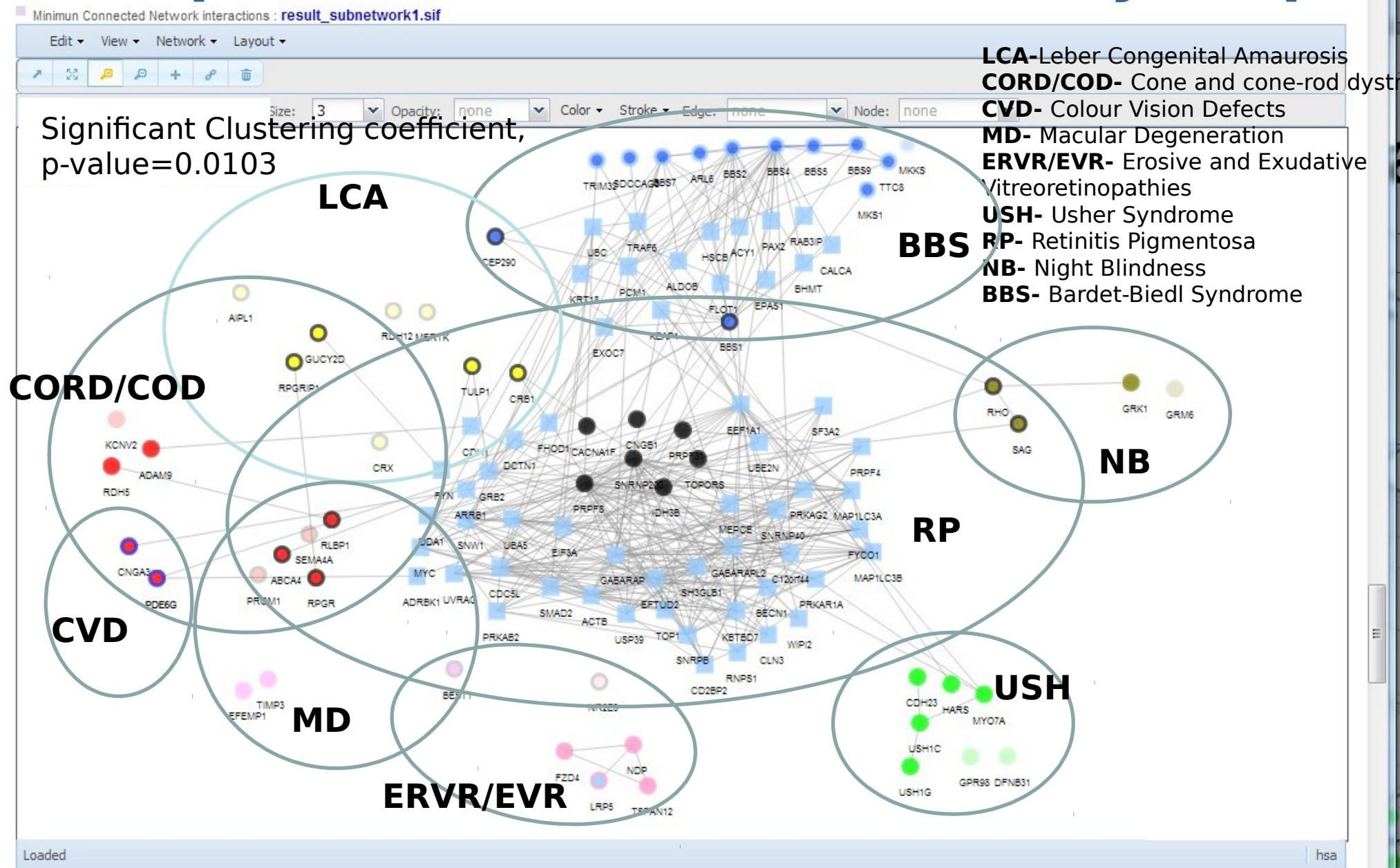
**Is genetic overlapping among IRDs related to protein int**

# Example with Inherited Retinal Dystrophies





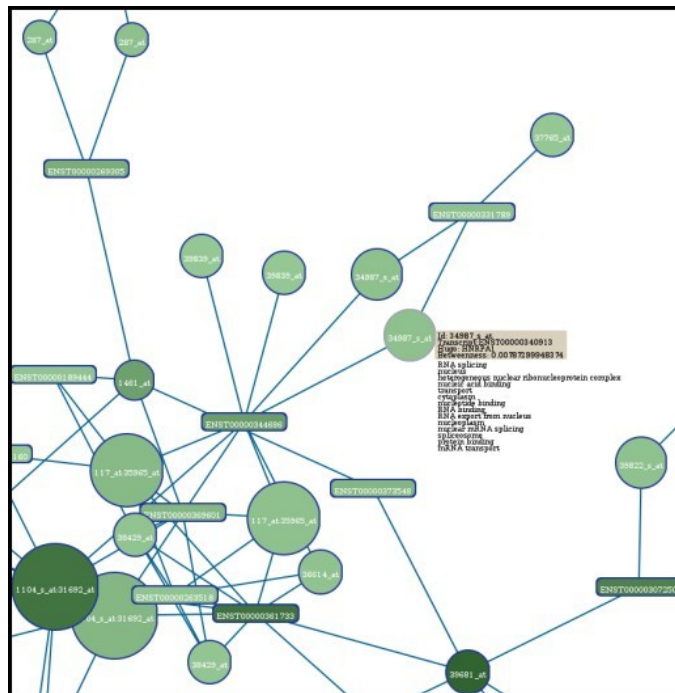
# Example with Inherited Retinal Dystrophies



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 proteins/genes (nodes) and evaluates the topological parameters of the within the interactome
  - Evaluates the list's cooperative behavior 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:

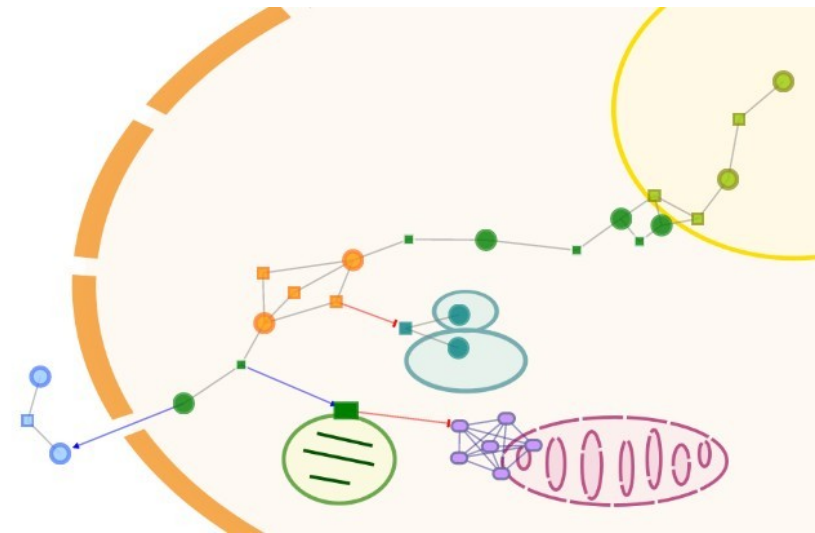
1. a list of **disease candidates** (ranked by their populational frequency)
2. a list of **genes** that are known to be **associated to the disease**

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

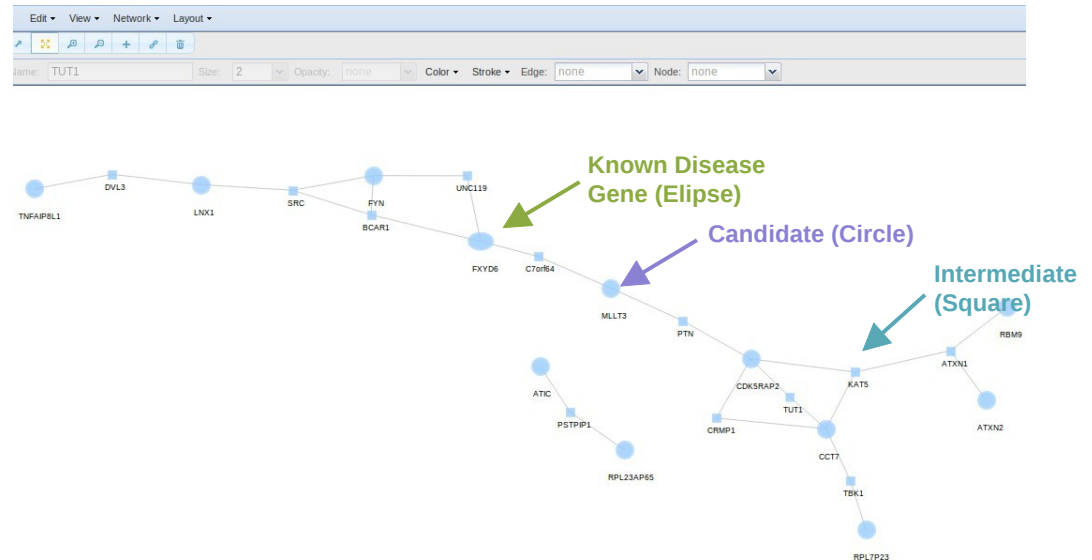
**Seed list:** Genes associated to Bipolar Disorder

**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

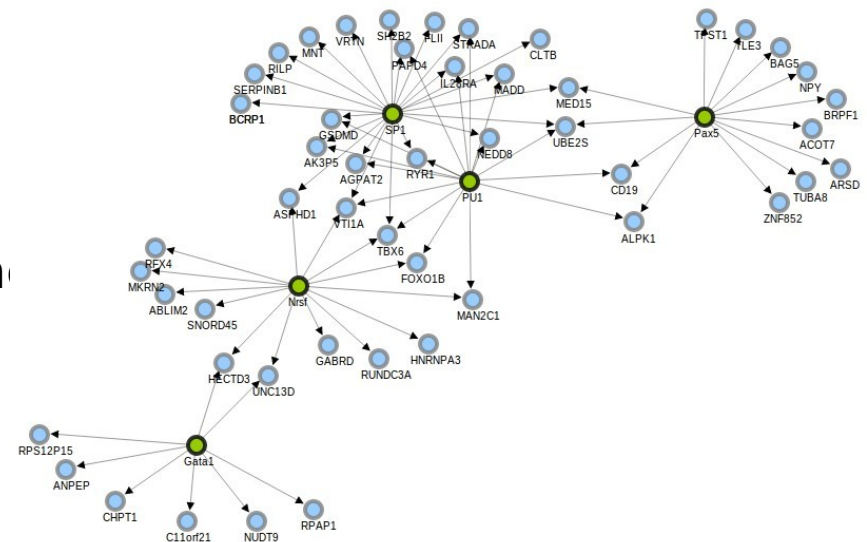
- Sometimes, the problem is not in the gene but in its regulators
- Tool for the **interpretation and visualization** of transcriptional (TFs) and post-transcriptional (miRNAs) **regulatory information**
- Designed to identify **common regulatory elements** in a list of genes
- RENATO maps these genes 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>

significant\_your\_annotation\_0.05.txt

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

Displaying 1 - 5 of 5



**THANK YOU.**