

# Patient stratification reveals the molecular basis of disease comorbidities

Beatriz Urda-García, Jon Sánchez-Valle, Rosalba Lepore and Alfonso Valencia

## Tutorial Web application

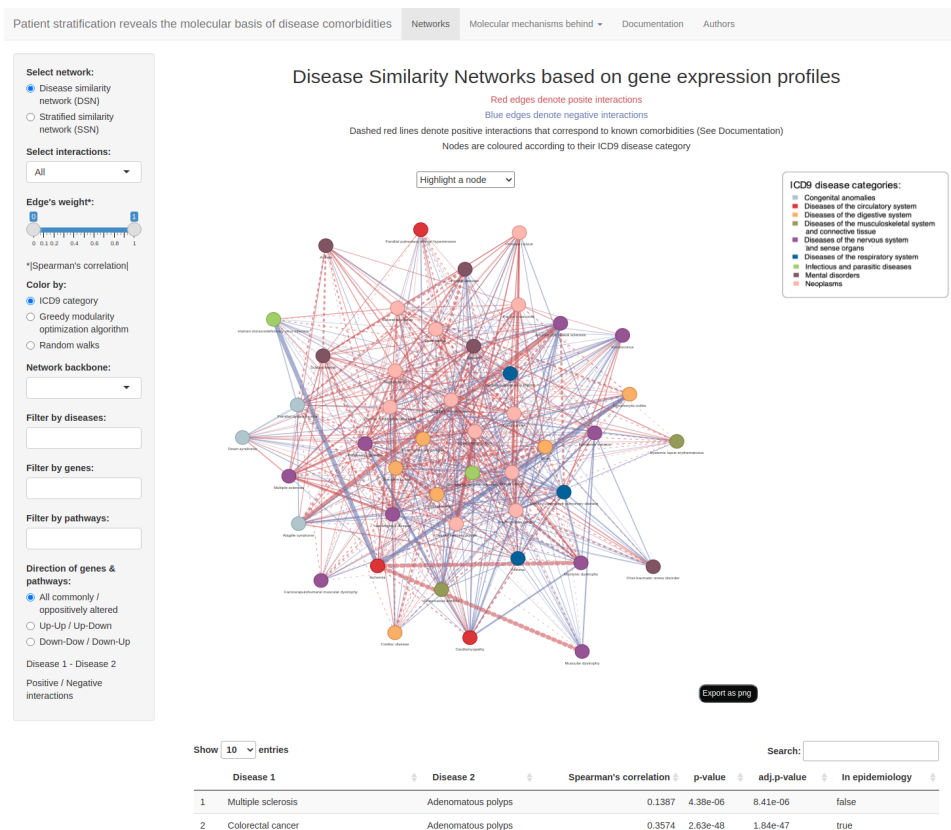
A functional and comprehensive resource to explore comorbidities at different resolution levels, accessible at <http://disease-perception.bsc.es/rgenexcom/>. This web application allows to explore the networks and molecular mechanisms presented in the publication *Patient stratification reveals the molecular basis of disease comorbidities*.

### Navigation bar:

- Networks
- Molecular Mechanisms behind
  - Diseases
  - Disease interactions
  - Get genes and pathways
- Documentation
- Authors

### Networks

This section allows the exploration of the Disease Similarity Network (DSN) and the Stratified Similarity Network (SSN).



## Functionalities:

**Select network:**

☒ Disease similarity network (DSN)

☐ Stratified similarity network (SSN)

**Select interactions:**

All ▼

**Edge's weight\*:**

0 1

0 0.1 0.2 0.4 0.6 0.8 1

\*|Spearman's correlation|

**Color by:**

☒ ICD9 category

☐ Greedy modularity optimization algorithm

☐ Random walks

**Network backbone:**

▼

**Filter by diseases:**

**Filter by genes:**

**Filter by pathways:**

**Direction of genes & pathways:**

☒ All commonly / oppositely altered

☐ Up-Up / Up-Down

☐ Down-Dow / Down-Up

Disease 1 - Disease 2

Positive / Negative interactions

Select the network of your choice (DSN or SSN)

Select the interactions (All, positive or negative interactions)

Filter the interactions by a given window of edge's weight, selected as the absolute value of the Spearman's correlation.

Color the nodes by their ICD9 category or community detection algorithms (greedy modularity optimization or random walks)

Show the network backbone of the current network Using the metric or ultrametric

Filter the network by nodes Shows the connections of the selected diseases and or meta-patients

Filter the network by shared genes or pathways Shows the connections in which the selected processes (gene or pathway) are involved.

Select a scenario (**Positive** / **Negative interactions**):

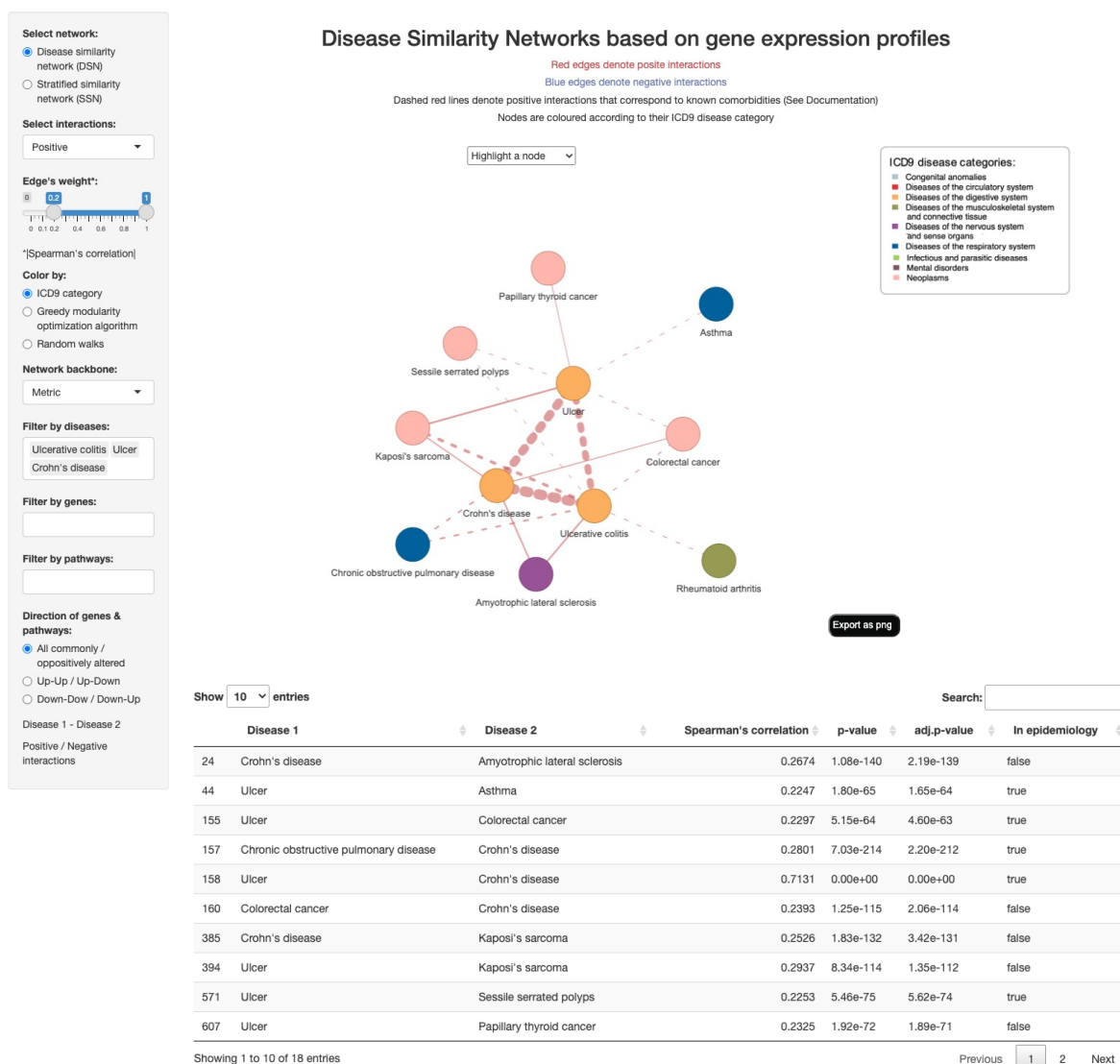
- ☐ All the interactions in which the processes are altered in the **same direction** / **opposite direction**
- ☐ The process is **up in both diseases** / **Up in the first disease and down in the second one**.
- ☐ The process is **down in both diseases** / **Down in the second one and up in the first one**.

\* The order of the diseases can be obtained from the table below (disease 1 and 2 are the first and second diseases respectively).

## Examples:

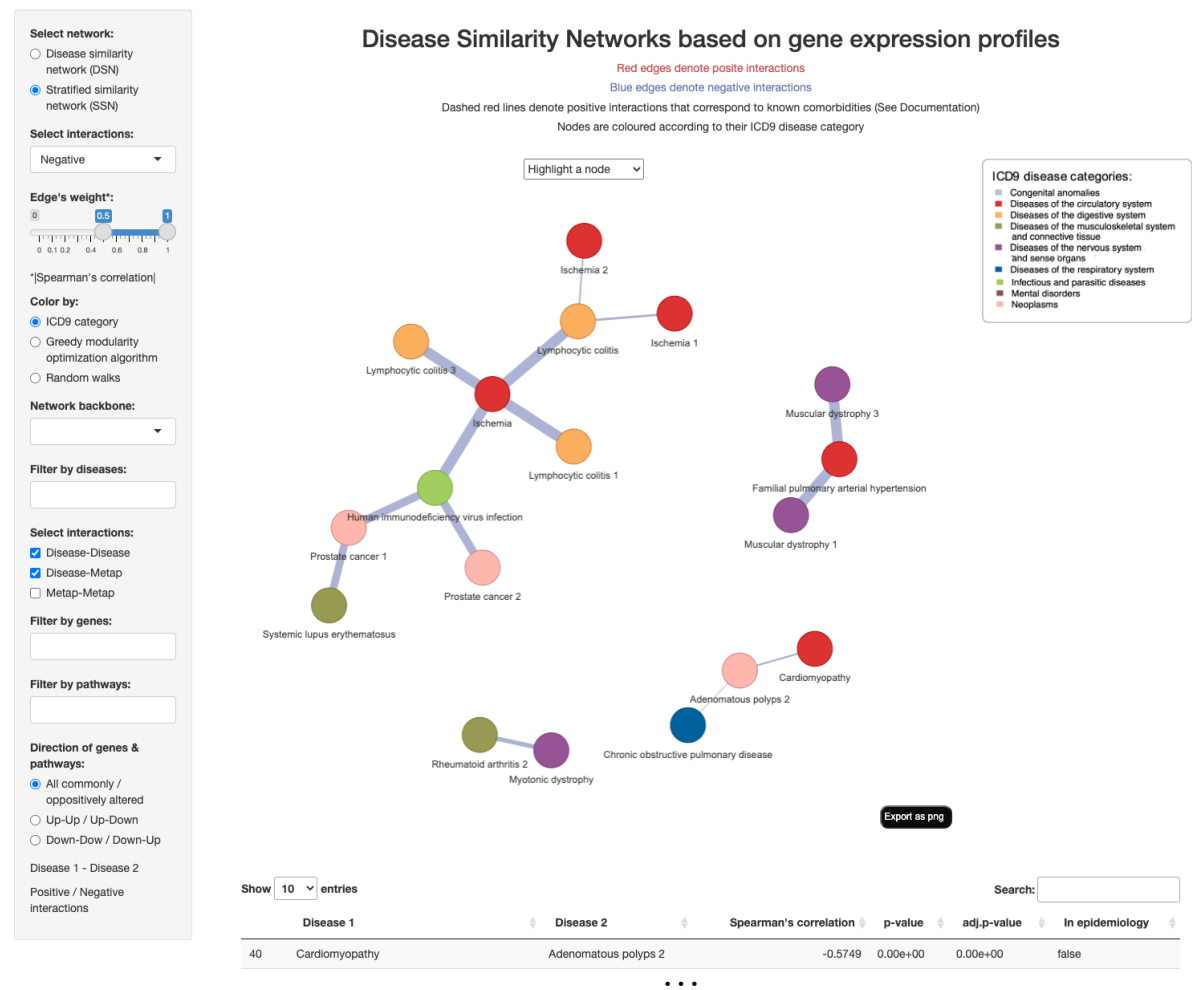
### 1. Exploring the disease-level positive links of ulcerative colitis, ulcer and Crohn's disease

We selected the positive interactions of the network at the disease level (DSN). We imposed a threshold for the edge's weight (Spearman's correlation  $\geq 0.2$ ) and we colored the nodes by their ICD9 category. Then, we were interested in the interactions contained in the network backbone using the metric closure. Finally, we selected the nodes of interest (ulcerative colitis, ulcer and Crohn's disease). We see the selected interactions, where dashed lines correspond to epidemiological interactions described in Hidalgo et al. All the information about these links can be explored in the dynamic table below.



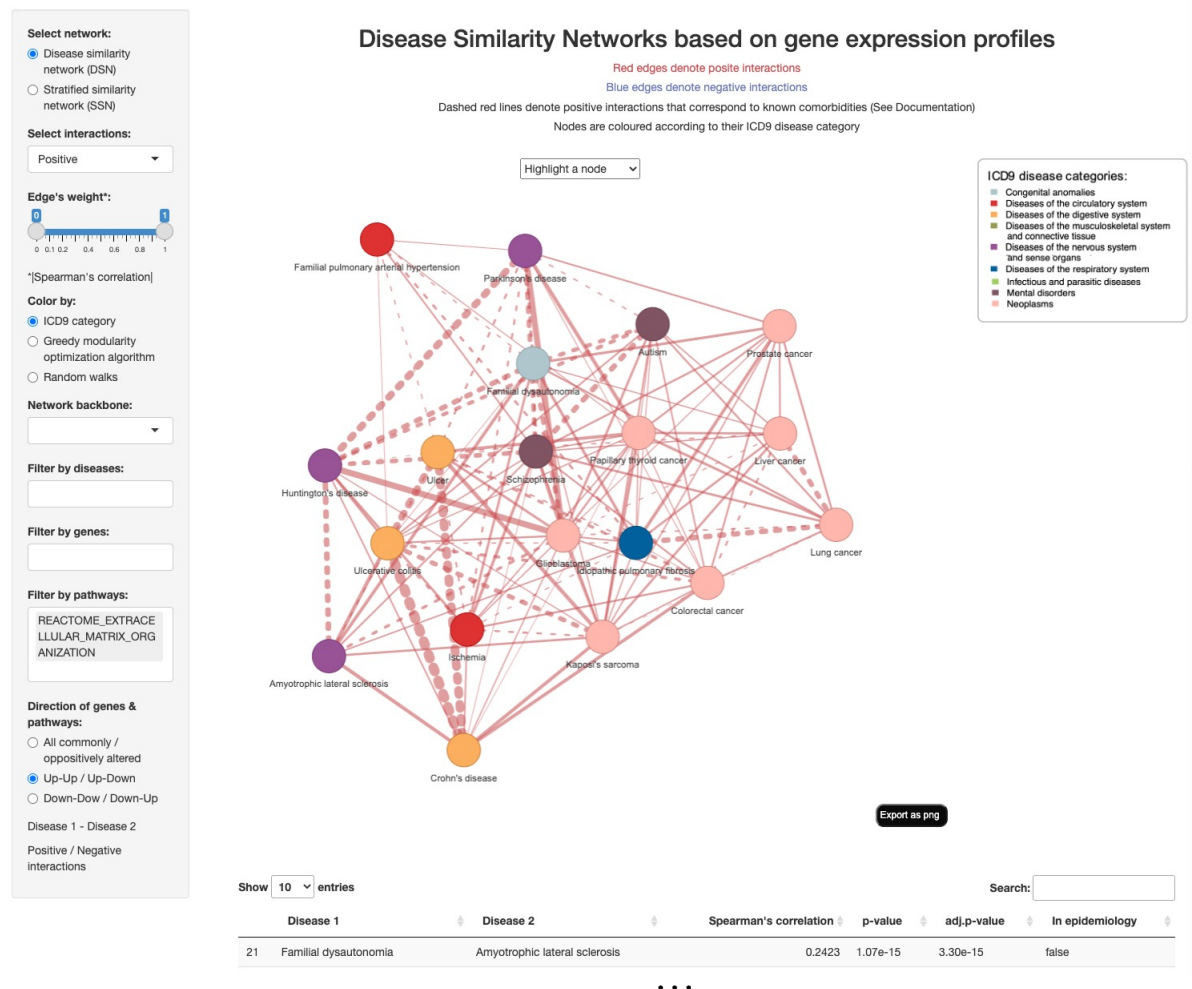
## 2. Exploring the strong negative links of the Stratified Similarity Network, excluding interactions between meta-patients

We selected the negative interactions of the SSN. We imposed a threshold for the edge's weight ( $|\text{Spearman's correlation}| \geq 0.5$ ). When the SSN is chosen, you can select the types of interactions that you are interested in (all, interactions between diseases, interactions between meta-patients and interactions between diseases and meta-patients). Here, we excluded the interactions between meta-patients.



### 3. Getting the positive interactions between diseases that share the overexpression of a given biological process

We are interested in the positive disease-disease links sharing the overexpression of the Extracellular Matrix Organization pathway from Reactome. To show this, we selected the positive interactions of the DSN. Then, we filtered by the interactions involving this pathway and we specified that we are interested in the interactions that share the overexpression of this pathway. As described in the manuscript, the diseases: ulcerative colitis, ulcer and Crohn's disease are comorbid, positively connected and share the overexpression of this process. Consequently, we observe them in the network connected with dashed red lines. You can filter by genes and or pathways from Reactome, KEGG and GO. If multiple processes are selected at the same time, it is shown the union of the interactions affected by these processes.



Extended documentation:

### **Disease Similarity Network (DSN)**

First, we implemented an RNA-seq pipeline to perform Differential Expression Analysis for each disease. Next, we computed the Spearman's correlation between the diseases' gene expression profiles. We kept significant interactions after multiple testing correction ( $FDR < 0.05$ ).

The obtained DSN contains positive and negative interactions, representing diseases with significantly similar and dissimilar gene expression profiles, respectively. Next, we evaluated the overlap of the positive interactions in the DSN with the epidemiological network from Hidalgo et al. Positive interactions in the DSN described in this epidemiological network are represented with red dashed lines.

### **Stratified Similarity Network (SSN)**

We stratified each disease into subgroups of patients with similar expression profiles (meta-patients) by applying the k-medoids clustering algorithm to its normalized and batch effect corrected gene expression matrix. Next, we performed Differential Expression Analyses for each meta-patient.

To analyze the disease subtype-associated comorbidities, we built the Stratified Similarity Network (SSN) connecting meta-patients and diseases based on the similarities of their gene expression profiles (following the same methodology described for the DSN). The resulting Stratified Similarity Network (SSN) contains three types of interactions: (1) the previously described disease-disease interactions, (2) interactions connecting different meta-patients and (3) interactions connecting meta-patients to diseases.

### **Network visualization**

The user can select the network (DSN or SSN) and interactions of interest (all, positive or negative interactions). A threshold for the edge's weight (the absolute value of the Spearman's correlation) can also be applied. By default, nodes are colored based on their International Code of Diseases 9 (ICD9) category. Moreover, community detection algorithms (greedy modularity optimization and random walks) can be computed. The DSN and SSN backbones can be extracted based on the metric or ultra-metric closure, following Simas *et al.* method.

Nodes can also be highlighted within the network and the interactions entailing specific nodes (diseases and or meta-patients), genes and pathways can be selected. Genes are represented with ensemble ids and REACTOME, KEGG and GO pathways are available. When selecting genes and pathways, positive interactions that share the alteration of such genes or pathways in the same / opposite direction will be shown for positive / negative interactions respectively. The user can also select only the interactions that are commonly up or downregulated for positive interactions or up-down and down-up for negative interactions. In the later cases, the table will clarify which diseases have the gene or pathways up vs down (Disease 1 - Disease 2).

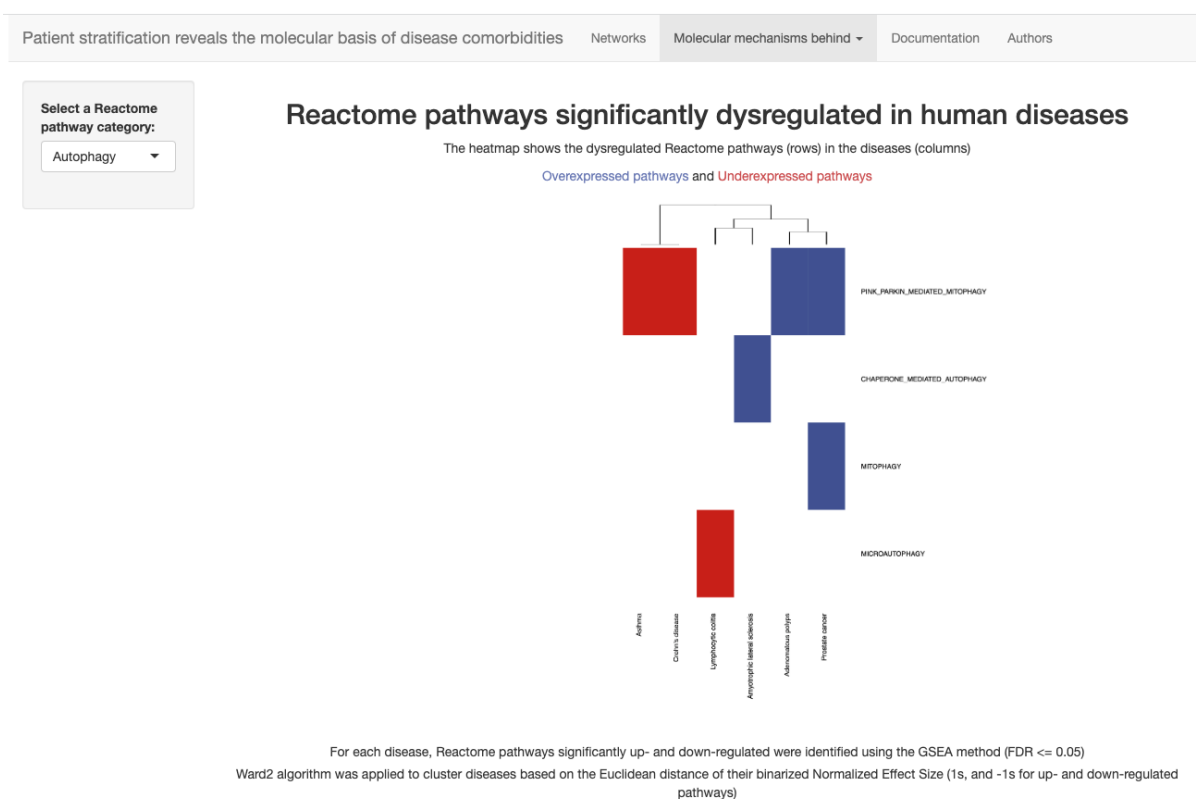
## Molecular Mechanisms behind

### Diseases

It shows the Reactome pathways significantly dysregulated in human diseases by pathway category. For each disease, Reactome pathways significantly over and underexpressed were identified using the GSEA method ( $FDR \leq 0.05$ ). Ward2 algorithm was applied to cluster diseases based on the Euclidean distance of their binarized Normalized Effect Size (1s, and -1s for over and underexpressed pathways). The heatmap shows the dysregulated Reactome pathways (rows) in the diseases (columns), where over and underexpressed pathways are blue and red colored respectively.

### Example:

Significantly dysregulated pathways for the Autophagy Reactome category.

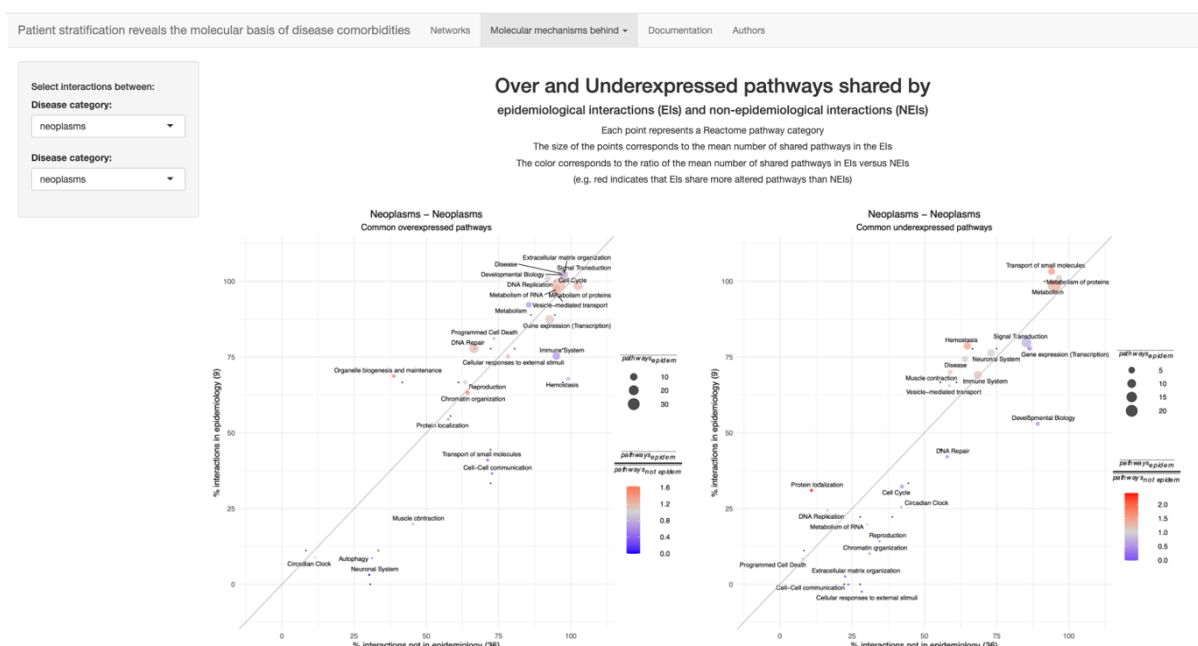


## Disease interactions

Over and underexpressed pathways behind epidemiological and not epidemiological interactions for each disease category pair. Percentage of epidemiological versus non epidemiological interactions that share overexpressed or underexpressed pathways. Each point represents a Reactome pathway category. The size of the points corresponds to the mean number of shared pathways in the epidemiological interactions. The color corresponds to the ratio of the mean number of shared pathways in epidemiological versus non epidemiological interactions.

### Example:

It shows the results for the interactions between neoplasms.



## Get genes and pathways

In this section you can access the differentially expressed genes and pathways in a given phenotype (disease or meta-patient) and commonly dysregulated in groups of phenotypes.

If you select one phenotype, you will get the table of dysregulated genes and pathways for that phenotype. You can filter the tables by selecting only the features that are significantly altered or by selecting only the over or underexpressed features.

If you select two or more phenotypes, you will get the genes or pathways that are significantly altered in all those phenotypes. Again, you can select only the over or the underexpressed features.



## Examples:

### 1. Get the genes significantly altered in a given phenotype: breast cancer meta-patient 1

We select the phenotype of interest (breast cancer meta-patient 1), we indicate that we want to get genes and the significantly altered ones. We get a table with the genes, logFC and adjusted p-value after multiple testing (FDR 5%).

[Patient stratification reveals the molecular basis of disease comorbidities](#) [Networks](#) [Molecular mechanisms behind ▾](#) [Documentation](#) [Authors](#)

Select disease(s):  
Breast cancer 1

Get:  
Genes ▾

Select features:  
☐ All  
☒ Significantly altered  
  
☒ All  
☐ Overexpressed  
☐ Underexpressed

### Genes and Pathways altered in diseases

or shared by disease groups and pairs

Show 10 ▾ entries Search:

	Gene.Symbol	logFC	adj.p.value
1	ENSG00000136158	-2.734	1.76e-53
2	ENSG00000109756	-1.712	8.25e-52
3	ENSG00000138685	-3.145	9.88e-48
4	ENSG00000104889	1.537	9.88e-48
5	ENSG00000119333	1.756	9.88e-48
6	ENSG00000166803	2.13	1.07e-47
7	ENSG00000160957	2.352	1.97e-47
8	ENSG00000162407	-2.429	2.09e-47
9	ENSG00000079462	2.193	6.56e-47
10	ENSG00000139946	-2.374	7.44e-47

Showing 1 to 10 of 9,891 entries Previous 1 2 3 4 5 ... 990 Next

### 2. Get the pathways significantly overexpressed in common in multiple phenotypes: ulcerative colitis, ulcer and Crohn's disease

We select the phenotypes of interest (ulcerative colitis, ulcer and Crohn's disease), we indicate that we want to get pathways, the significantly altered and overexpressed ones. We get the significantly overexpressed pathways in these disorders from KEGG, GO and REACTOME. We can search in the obtained table.

[Patient stratification reveals the molecular basis of disease comorbidities](#) [Networks](#) [Molecular mechanisms behind ▾](#) [Documentation](#) [Authors](#)

Select disease(s):  
Ulcerative colitis  
Ulcer  
Crohn's disease

Get:  
Pathways ▾

Select features:  
☐ All  
☒ Significantly altered  
  
☒ Overexpressed  
☐ Underexpressed

### Genes and Pathways altered in diseases

or shared by disease groups and pairs

Show 10 ▾ entries Search:

	Intersection
1	REACTOME_EXTRACELLULAR_MATRIX_ORGANIZATION
2	REACTOME_INTEGRIN_CELL_SURFACE_INTERACTIONS
3	KEGG_ECM_RECEPTOR_INTERACTION
4	REACTOME_ECM_PROTEOGLYCANS
5	GO_BLOOD_VESSEL_MORPHOGENESIS
6	REACTOME_COLLAGEN_FORMATION
7	REACTOME_INTERLEUKIN_4_AND_INTERLEUKIN_13_SIGNALING
8	GO_POSITIVE_REGULATION_OF_VASCULATURE_DEVELOPMENT
9	GO_EXTRACELLULAR_STRUCTURE_ORGANIZATION
10	REACTOME_ASSEMBLY_OF_COLLAGEN_FIBRILS_AND_OTHER_MULTIMERIC_STRUCTURES

Showing 1 to 10 of 570 entries Previous 1 2 3 4 5 ... 57 Next