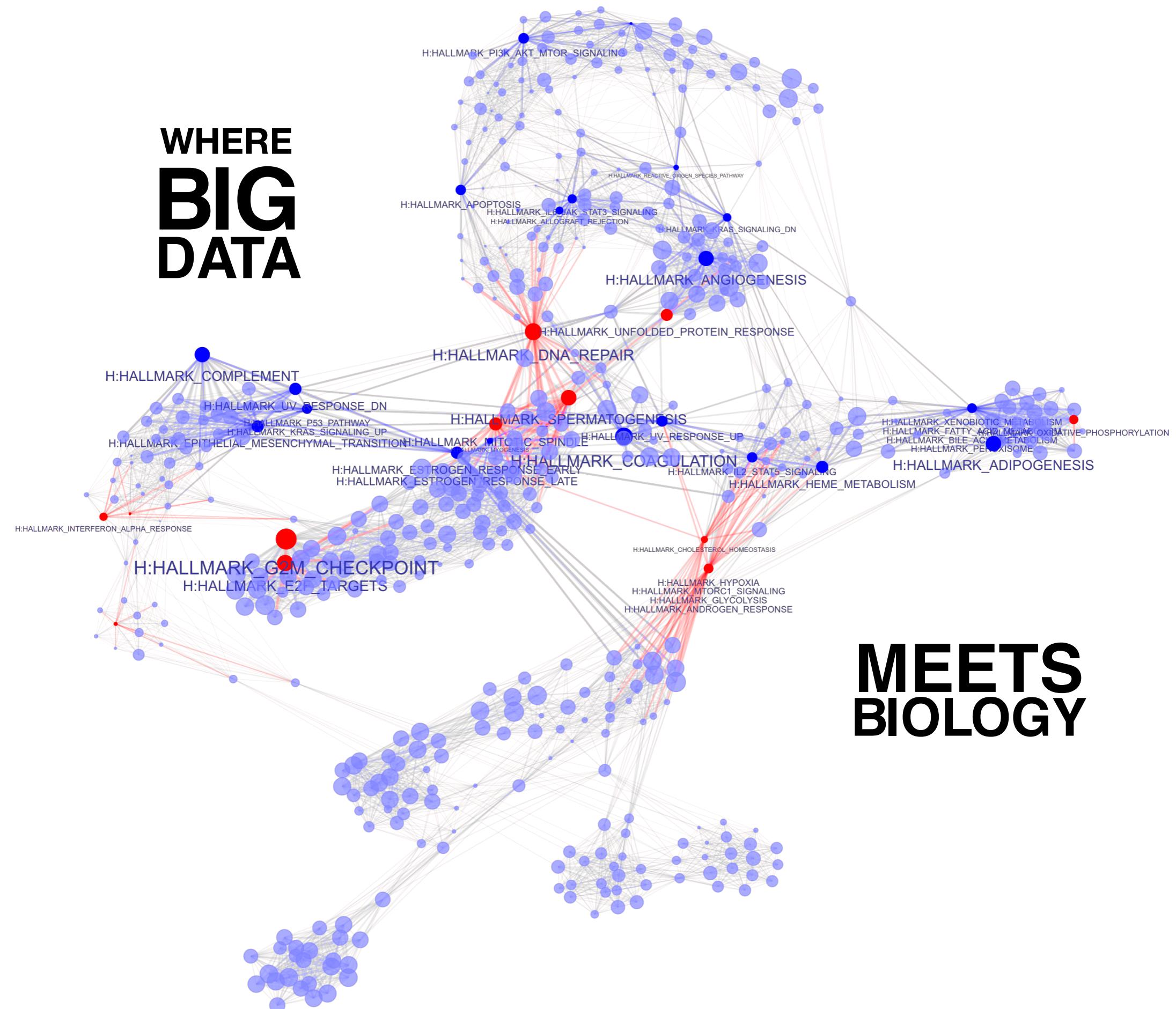


# Omics Playground (0.99)

WHERE  
**BIG**  
DATA



MEETS  
BIOLOGY



BigOmics Analytics Sagl, A spin-off of USI/IOR  
Via Vincenzo Vela 6 - CH6500 Bellinzona CHE-299.736.565  
Tel. +41 820 0368 - [www.bigomics.ch](http://www.bigomics.ch)



Institute of Oncology Research  
Institute of Oncology Research  
Via Vincenzo Vela 5 - CH6500 Bellinzona  
Tel. +41 91 820 0322 - [www.ior.iosi.ch](http://www.ior.iosi.ch)



Institute for Research in Biomedicine  
Institute for Research in Biomedicine  
Via Vincenzo Vela 6 - CH6500 Bellinzona  
Tel. +41 91 820 0300 - [www.irb.usi.ch](http://www.irb.usi.ch)

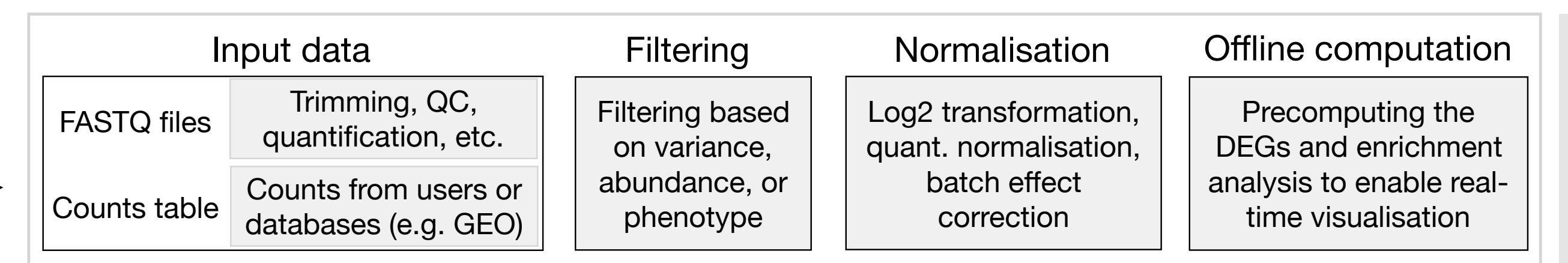
# The platform overview

A) The current version of our platform (0.99) is composed of two components.

B) The first component address the data cleaning and preprocessing, which includes preparing the input data, filtering, normalising and precomputing of statistics for some analyses.

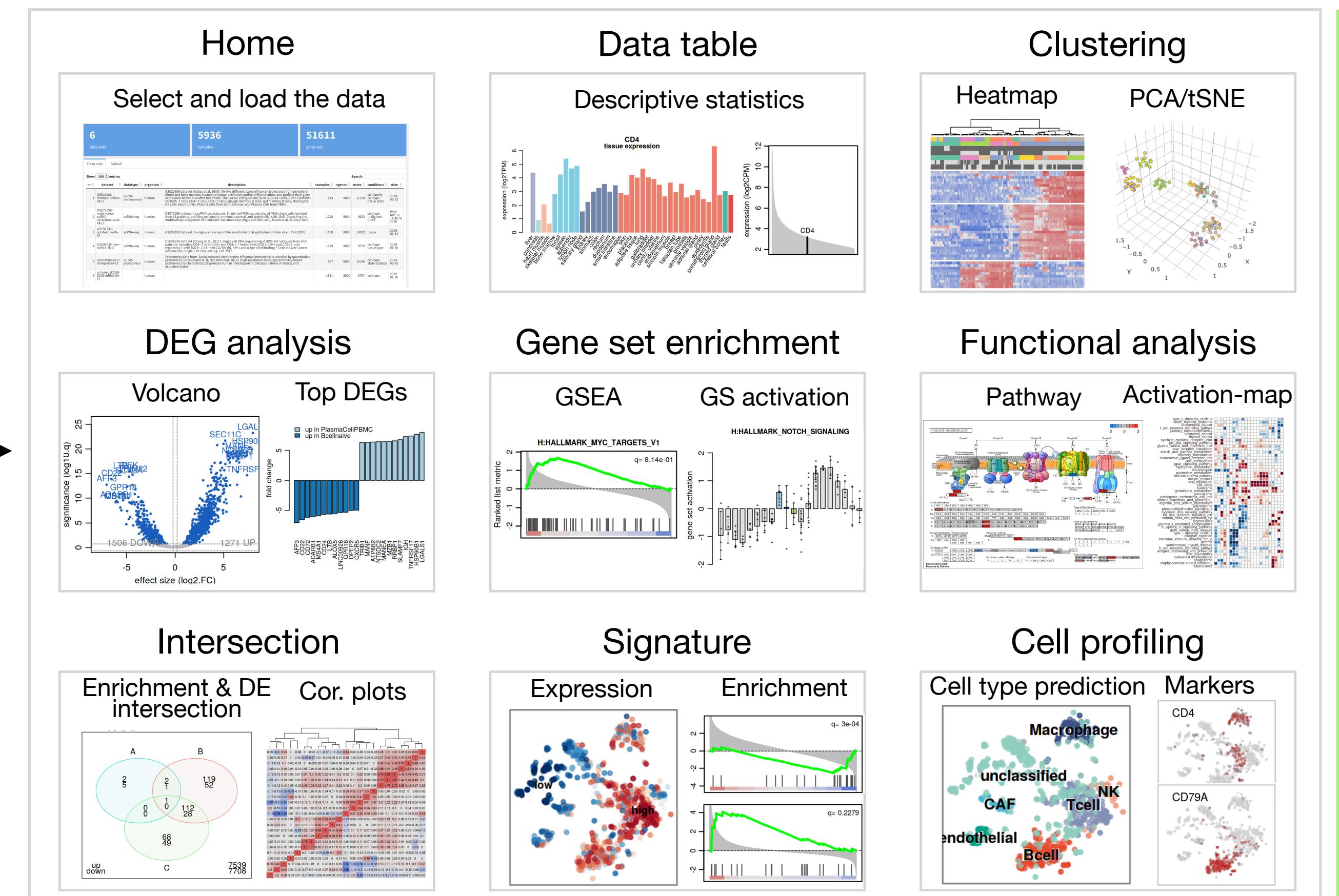


## Data cleaning & preprocessing



Offline

## Omics-playground interface



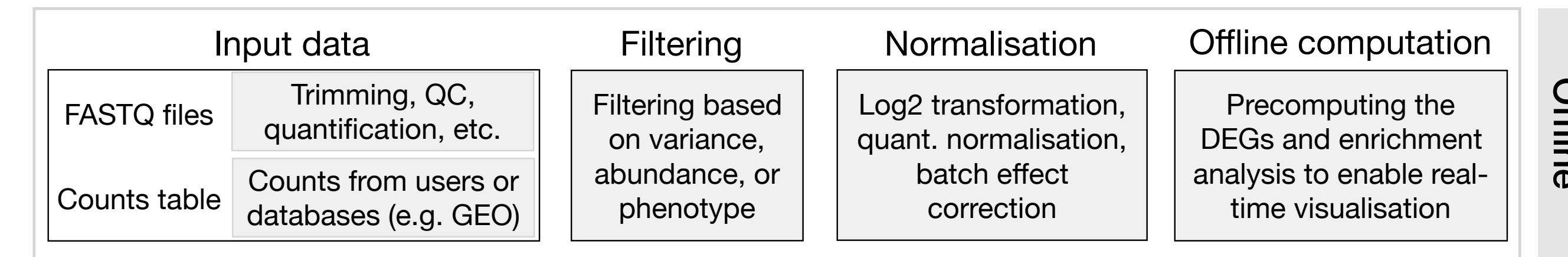
Online, real-time visualisation

C) The second part is composed of the online interface, which supports the real-time visualisation and interaction with users.



# Data cleaning & preprocessing

## Data cleaning & preprocessing



A) The data cleaning and preprocessing is performed offline using scripts to support real-time interaction by minimising user interface latency. The platform comes with the necessary scripts for data cleaning and preprocessing.

B) **Input data.** The platform requires the table of gene counts as input. Users can provide their own gene counts or download the relevant data from repositories such as GEO. If they have FASTQ files, we provide scripts to obtain gene counts through quality control, trimming, quantification of gene abundance, and so on. Although the script implements the Salmon, users can modify and use any other software, including Kallisto or Star.

C) **Filtering.** The data preprocessing includes some filtering criteria, such as filtering of genes based on variance, the expression across the samples, and the number of missing values. Similarly, samples can also be filtered based on the read quality, the total abundance, unrelated phenotype, or an outlier criterion.

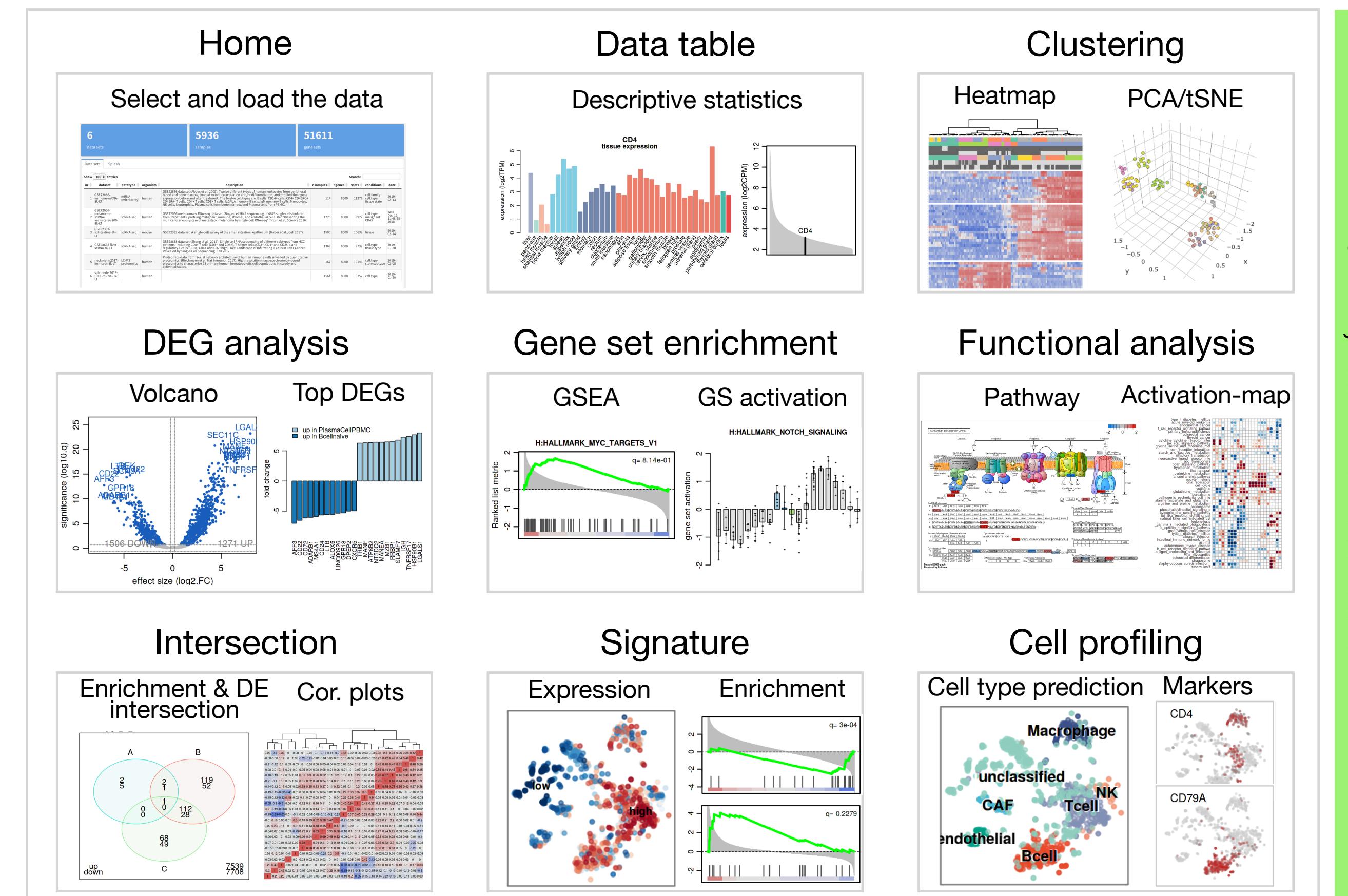
D) **Normalisation.** The raw counts are converted into the counts per million (CPM) and log2. Depending on the data set, a quantile normalization can be applied. Already known batches in the data can easily be corrected with the limma or ComBat. Unknown batches in the data, however, require a much deeper analysis to correct if there are any in the data. Although the sva package addresses the surrogate variable analysis to remove unknown batch effects, it is still an important open research question in itself.

E) **Offline computation.** We precompute the statistics for the differentially expressed genes (DEG) and gene set enrichment (GSE) analyses to accelerate the visualisation on the interface.

# Omics-playground interface

The interface contains nine modules, where the data analysis workflow is divided into subprocesses. After selecting and loading the data from the home page, users can proceed with any analysis on desire. There is no specific order between the analysis modules that users should follow, as the most of the statistics are precomputed offline in the previous step. A brief description and functionality of each module is provided in the next slides.

## Omics-playground interface



# Home page

A) The platform starts running from the home page. Basically, this page provides a general information about datasets injected into the platform.

The screenshot shows the Omics Playground interface. At the top, there is a navigation bar with the 'irb' logo, 'Omics Playground', and a 'Home' button (which is highlighted with a red box). Below the navigation bar, there are three large blue boxes displaying statistics: '6 data sets', '5936 samples', and '51611 gene sets'. A yellow box on the left contains the text 'C) Users can select and load the data of their interest and start the analysis from here.' with an arrow pointing to a dropdown menu labeled 'GSE72056-melanoma-scRNA-vsclusters-s200-8k-LT.pgx' and a 'Load dataset' button. The main area features a table with 6 entries, each representing a dataset. The table includes columns for nr, dataset, datatype, organism, description, nsamples, ngenes, nsets, conditions, and date. An upward arrow points from the text 'B)' in the yellow box below to the table header.

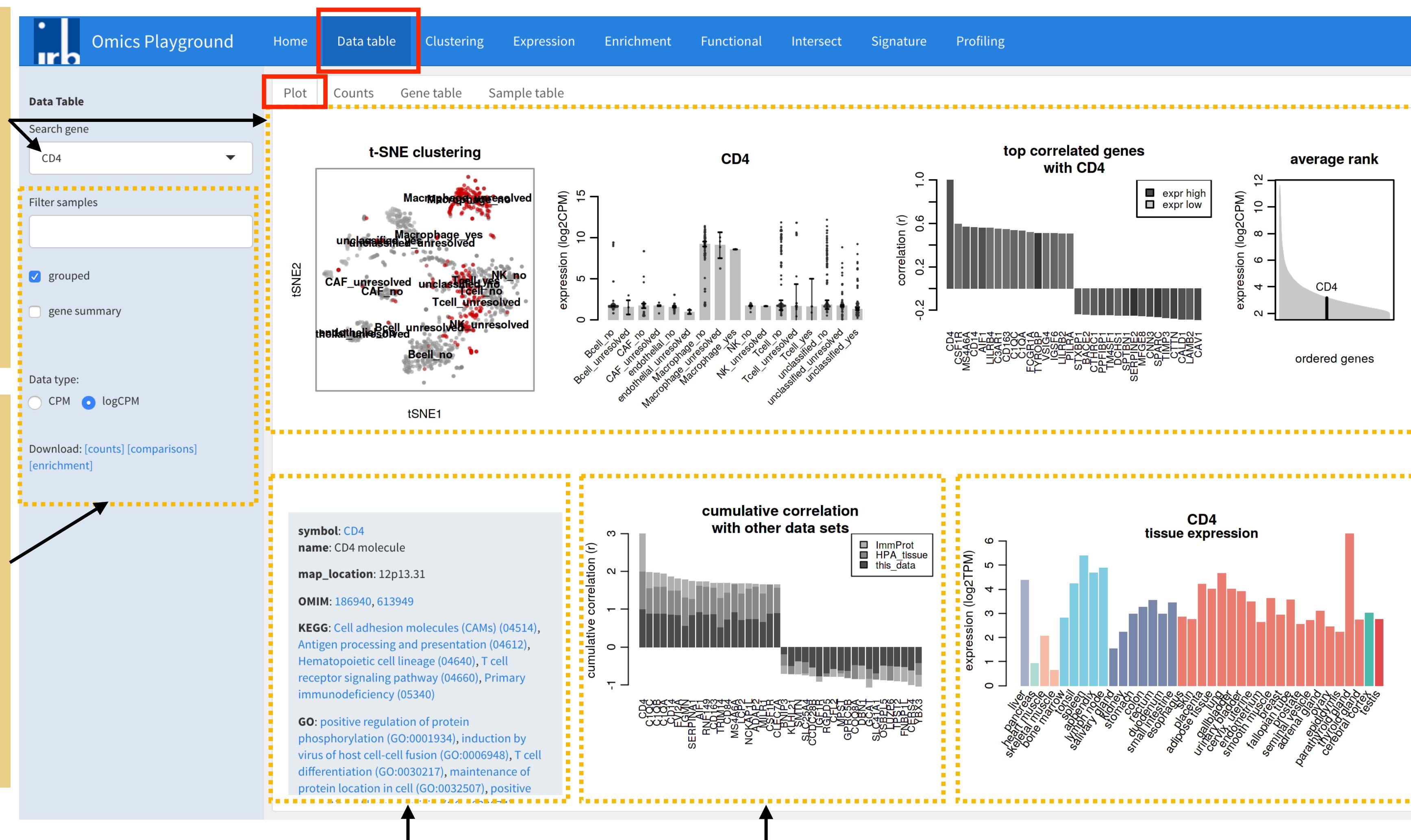
nr	dataset	datatype	organism	description	nsamples	ngenes	nsets	conditions	date
1	GSE22886-immune-mRNA-8k-LT	mRNA (microarray)	human	GSE22886 data set (Abbas et al, 2005). Twelve different types of human leukocytes from peripheral blood and bone marrow, treated to induce activation and/or differentiation, and profiled their gene expression before and after treatment. The twelve cell types are: B cells, CD14+ cells, CD4+ CD45RO+ CD45RA- T cells, CD4+ T cells, CD8+ T cells, IgG/IgA memory B cells, IgM memory B cells, Monocytes, NK cells, Neutrophils, Plasma cells from bone marrow, and Plasma cells from PBMC.	114	8000	11278	cell.family cell.type tissue state	2019-02-13
2	GSE72056-melanoma-scRNA-vsclusters-s200-8k-LT	scRNA-seq	human	GSE72056 melanoma scRNA-seq data set. Single-cell RNA sequencing of 4645 single cells isolated from 19 patients, profiling malignant, immune, stromal, and endothelial cells. Ref: 'Dissecting the multicellular ecosystem of metastatic melanoma by single-cell RNA-seq', Tirosh et al, Science 2016.	1225	8000	9922	cell.type malignant CD45	Wed Dec 12 11:48:58 2018
3	GSE92332-sclIntestine-8k-LT	scRNA-seq	mouse	GSE92332 data set. A single-cell survey of the small intestinal epithelium (Haber et al., Cell 2017).	1500	8000	10632	tissue	2019-02-14
4	GSE98638-liver-scRNA-8k-LT	scRNA-seq	human	GSE98638 data set (Zheng et al., 2017). Single cell RNA sequencing of different subtypes from HCC patients, including CD8+ T cells (CD3+ and CD8+), T helper cells (CD3+, CD4+ and CD25-), and regulatory T cells (CD3+, CD4+ and CD25high). Ref: Landscape of Infiltrating T Cells in Liver Cancer Revealed by Single-Cell Sequencing. Cell 2017.	1369	8000	9732	cell.type tissue.type	2019-01-30
5	rieckmann2017-immprot-8k-LT	LC-MS proteomics	human	Proteomics data from 'Social network architecture of human immune cells unveiled by quantitative proteomics' (Rieckmann et al, Nat Immunol. 2017). High-resolution mass-spectrometry-based proteomics to characterize 28 primary human hematopoietic cell populations in steady and activated states.	167	8000	10146	cell.type state subtype	2019-02-05
6	schmiedel2018-DICE-mRNA-8k-LT		human		1561	8000	9757	cell.type	2019-01-20

B) For each dataset, a brief description and the type of the data as well as the total number of samples, genes, genesets (or pathways), corresponding phenotypes and the collection date are reported in this tab.

# Data table module: plots tab

A) The data table page provides a descriptive statistical analysis on a gene level with visualisations.

B) For a selected gene by the user, the plots section displays figures related to the expression level of the gene, correlation to other genes, and average expression ranking within the dataset.



C) During the visual analysis, users can filter out some samples or collapse the samples by predetermined groups. It is also possible to visualize the information on a raw count level (CPM) rather than a logarithmic expression level (logCPM).

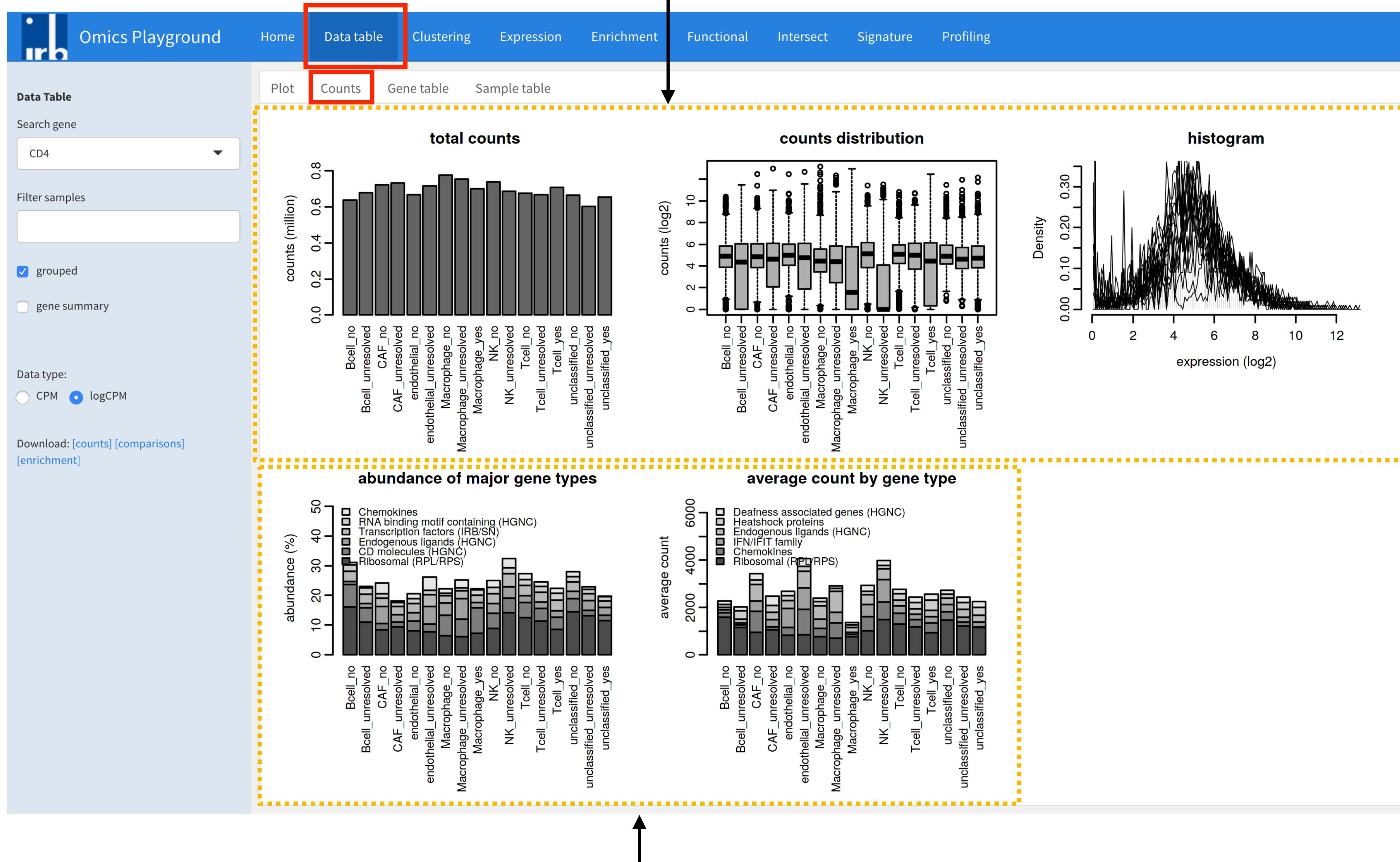
D) To find out more information from the literature, hyperlinks are provided to connect the selected gene to OMIM, KEGG and GO databases.

E) It also correlates the gene to the expressions of other genes across datasets such as ImmProt and HPA, and plots the cumulative correlation.

F) Furthermore, it displays the tissue expression for a selected gene using the genotype-tissue expression (GTEx) dataset.

# Data table module: counts tab

G) The total number of counts (abundance) per sample and their distribution among the samples are displayed in the **counts tab**.



H) For each sample, users can also see the percentage of counts in terms of major gene types such as CD molecules, kinases or RNA binding motifs.

# Data table module: gene table tab

The screenshot shows the Omics Playground interface with the 'Data table' tab selected (highlighted by a red box). The main content area displays a 'Gene table' (also highlighted by a red box) containing a list of genes and their correlation metrics across various samples. A yellow callout box with a black arrow points from the text below to the gene table.

**I) More elaborated correlation analysis across the samples can be performed under gene table section, where genes are ordered by the correlation with respect to the selected gene. Gene-wise average expression of a phenotype sample grouping is also presented in this table.**

gene	title	rho	SD	Bcell_no	Bcell_unresolved	CAF_no	CAF_unresolved	endothelial_no	endothelial_unresolved	M
CD4	CD4 molecule	1	3.696	1.804	1.848	1.918	1.589	1.501	0.984	
CSF1R	colony stimulating factor 1 receptor	0.576	2.792	1.683	1.383	1.45	1.589	1.464	0.984	
MS4A6A	membrane spanning 4-domains A6A	0.55	3.4	1.835	1.383	1.408	1.589	1.525	0.984	
LILRB4	leukocyte immunoglobulin like receptor B4	0.544	2.616	1.844	3.412	1.345	1.589	1.498	0.984	
CD14	CD14 molecule	0.541	3.633	1.663	1.383	1.925	1.968	2.558	0.984	
AIF1	allograft inflammatory factor 1	0.54	3.665	1.74	1.383	1.345	1.589	1.464	0.984	
CD163	CD163 molecule	0.529	2.614	1.725	1.383	1.345	1.589	1.464	0.984	
C3AR1	complement C3a receptor 1	0.528	2.685	1.676	1.383	1.431	1.589	1.464	0.984	
C1QC	complement C1q C chain	0.515	3.21	1.663	1.383	1.369	1.589	1.464	0.984	
C1QA	complement C1q A chain	0.511	3.286	1.663	1.383	1.345	1.589	1.586	0.984	
FCGR1A	Fc fragment of IgG receptor Ia	0.497	2.316	1.663	1.383	1.49	1.589	1.464	0.984	
IGSF6	immunoglobulin superfamily member 6	0.491	2.902	1.804	1.383	1.41	1.589	1.464	0.984	
VSIG4	V-set and immunoglobulin domain containing 4	0.49	2.547	1.689	1.383	1.345	1.589	1.464	0.984	
HCK	HCK proto-oncogene, Src family tyrosine kinase	0.489	2.717	1.905	1.383	1.345	1.589	1.515	0.984	
TYROBP	TYRO protein tyrosine kinase binding protein	0.488	4.691	1.788	1.383	1.345	1.589	1.632	0.984	
MSR1	macrophage scavenger receptor 1	0.488	2.486	1.663	1.383	1.345	1.589	1.522	0.984	
LILRB2	leukocyte immunoglobulin like receptor B2	0.488	2.41	2.222	1.383	1.403	1.589	1.616	0.984	
PILRA	paired immunoglobulin like type 2 receptor alpha	0.485	2.839	1.978	1.383	1.431	1.589	1.679	2.372	
FCER1G	Fc fragment of IgE receptor Ig	0.475	4.523	1.701	1.383	1.436	1.589	1.656	2.504	
C1QB	complement C1q B chain	0.463	3.663	1.704	1.383	1.588	1.589	1.505	0.984	
SLCO2B1	solute carrier organic anion transporter family me	0.455	2.295	1.68	1.383	1.895	2.338	2.266	4.979	
ADAP2	ArfGAP with dual PH domains 2	0.451	2.247	2.016	4.834	1.446	1.589	1.815	0.984	
TLR2	toll like receptor 2	0.448	2.439	1.726	1.383	1.468	1.589	1.537	0.984	
FGL2	fibrinogen like 2	0.446	3.102	2.702	1.383	2.506	2.689	2.236	1.637	
TNFSF13	TNF superfamily member 13	0.443	2.679	2.249	2.138	1.579	1.589	1.692	0.984	
MAFB	MAF bZIP transcription factor B	0.442	2.327	1.663	1.383	2.078	1.589	1.911	2.892	
MS4A4A	membrane spanning 4-domains A4A	0.439	3.18	3.022	1.383	1.601	1.589	2.098	0.984	
TBXAS1	thromboxane A synthase 1	0.436	2.734	2	1.383	1.345	1.589	1.464	0.984	
CCR1	C-C motif chemokine receptor 1	0.432	2.57	1.755	1.383	1.428	1.589	1.464	0.984	
FCGR3A	Fc fragment of IgG receptor IIIa	0.431	3.641	1.694	1.383	1.345	1.589	1.505	0.984	
TMEM176B	transmembrane protein 176B	0.43	3.381	1.693	1.383	5.201	3.004	3.428	5.446	
CSF3R	colony stimulating factor 3 receptor	0.428	2.22	1.699	1.383	1.345	1.589	1.464	0.984	
TREM2	triggering receptor expressed on myeloid cells 2	0.428	2.163	1.663	1.383	1.398	1.589	1.464	0.984	
ITGB2	integrin subunit beta 2	0.427	4.57	4.255	1.383	1.588	1.589	1.539	0.984	
TGFBI	transforming growth factor beta induced	0.426	3.501	1.873	1.383	6.32	8.2	1.639	0.984	
PLXDC2	plexin domain containing 2	0.423	2.747	1.697	1.383	4.251	2.513	3.173	3.991	
CD86	CD86 molecule	0.422	2.415	2.313	1.383	1.345	1.589	1.474	0.984	
FYB1	FYN binding protein 1	0.421	3.903	1.883	1.383	1.607	1.589	1.504	0.984	
CD162		0.419	8.35	2.651	1.383	1.457	1.609	4.61	0.934	

# Data table module: sample table tab

The screenshot shows the Omics Playground interface with the 'Data table' tab selected (highlighted by a red box). The main area displays a 'Sample table' (also highlighted by a red box) containing 46 rows of data. The columns are labeled: group, patient, cell.type, malignant, CD45, cluster, and .cell\_cycle. The data consists of various sample identifiers and their corresponding characteristics. A search bar and download options (Copy, CSV, PDF) are visible at the top of the table.

	group	patient	cell.type	malignant	CD45	cluster	.cell_cycle
cy60_1_cd_45_pos_3_A08_S296_comb	Bcell_no	CY60	Bcell	no	-	C3	S
cy60_1_cd_45_pos_3_B12_S312_comb	Bcell_no	CY60	Bcell	no	-	C3	S
cy60_1_cd_45_pos_3_C12_S324_comb	Bcell_no	CY60	Bcell	no	-	C3	G2M
cy60_1_cd_45_pos_3_D02_S326_comb	Bcell_no	CY60	Bcell	no	-	C3	G1
cy60_1_cd_45_pos_3_D04_S328_comb	Bcell_no	CY60	Bcell	no	-	C3	S
cy60_1_cd_45_pos_3_D05_S329_comb	Bcell_no	CY60	Bcell	no	-	C3	G1
cy60_1_cd_45_pos_3_E04_S340_comb	Bcell_no	CY60	Bcell	no	-	C3	G2M
cy60_1_cd_45_pos_3_E07_S343_comb	Bcell_no	CY60	Bcell	no	-	C3	G2M
cy60_1_cd_45_pos_3_F05_S353_comb	Bcell_no	CY60	Bcell	no	-	C3	G1
cy60_1_cd_45_pos_3_F07_S355_comb	Bcell_no	CY60	Bcell	no	-	C3	G1
cy60_1_cd_45_pos_3_G02_S362_comb	Bcell_no	CY60	Bcell	no	-	C3	G1
cy60_1_cd_45_pos_3_G06_S366_comb	Bcell_no	CY60	Bcell	no	-	C3	G1
cy60_1_cd_45_pos_3_G07_S367_comb	Bcell_no	CY60	Bcell	no	-	C6	G1
cy60_1_cd_45_pos_3_H06_S378_comb	Bcell_no	CY60	Bcell	no	-	C3	G2M
cy60_1_cd_45_pos_3_H10_S382_comb	Bcell_no	CY60	Bcell	no	-	C3	G1
cy60_1_cd_45_pos_4_A11_S11_comb	Bcell_no	CY60	Bcell	no	-	C3	G2M
cy60_1_cd_45_pos_4_B03_S15_comb	Bcell_no	CY60	Bcell	no	-	C3	G2M
cy60_1_cd_45_pos_4_D01_S37_comb	Bcell_no	CY60	Bcell	no	-	C6	G2M
cy60_1_cd_45_pos_4_D02_S38_comb	Bcell_no	CY60	Bcell	no	-	C3	G2M
cy60_1_cd_45_pos_4_D08_S44_comb	Bcell_no	CY60	Bcell	no	-	C3	G2M
cy60_1_cd_45_pos_4_E10_S58_comb	Bcell_no	CY60	Bcell	no	-	C3	G1
cy60_1_cd_45_pos_4_H01_S85_comb	Bcell_no	CY60	Bcell	no	-	C3	S
cy60_1_cd_45_pos_HC_pos_HLDAR_pos_A05_S965_comb	Bcell_no	CY60	Bcell	no	-	C3	G2M
cy60_1_cd_45_pos_HC_pos_HLDAR_pos_A09_S969_comb	Bcell_no	CY60	Bcell	no	-	C3	G1
cy60_1_cd_45_pos_HC_pos_HLDAR_pos_B06_S978_comb	Bcell_no	CY60	Bcell	no	-	C3	G1
cy60_1_cd_45_pos_HC_pos_HLDAR_pos_B10_S982_comb	Bcell_no	CY60	Bcell	no	-	C3	S
cy60_1_cd_45_pos_HC_pos_HLDAR_pos_C04_S988_comb	Bcell_no	CY60	Bcell	no	-	C3	G1
cy60_1_cd_45_pos_HC_pos_HLDAR_pos_C06_S990_comb	Bcell_no	CY60	Bcell	no	-	C3	S
cy60_1_cd_45_pos_HC_pos_HLDAR_pos_C07_S991_comb	Bcell_no	CY60	Bcell	no	-	C3	S
cy60_1_cd_45_pos_HC_pos_HLDAR_pos_D05_S1001_comb	Bcell_no	CY60	Bcell	no	-	C3	G1
cy60_1_cd_45_pos_HC_pos_HLDAR_pos_D06_S1002_comb	Bcell_no	CY60	Bcell	no	-	C3	G2M
cy60_1_cd_45_pos_HC_pos_HLDAR_pos_D10_S1006_comb	Bcell_no	CY60	Bcell	no	-	C3	G1
cy60_1_cd_45_pos_HC_pos_HLDAR_pos_E05_S1013_comb	Bcell_no	CY60	Bcell	no	-	C3	S
cy60_1_cd_45_pos_HC_pos_HLDAR_pos_F02_S1022_comb	Bcell_no	CY60	Bcell	no	-	C3	G1
cy60_1_cd_45_pos_HC_pos_HLDAR_pos_F03_S1023_comb	Bcell_no	CY60	Bcell	no	-	C3	G2M
cy60_1_cd_45_pos_HC_pos_HLDAR_pos_G05_S1037_comb	Bcell_no	CY60	Bcell	no	-	C3	S
cy60_1_cd_45_pos_HC_pos_HLDAR_pos_H02_S1046_comb	Bcell_no	CY60	Bcell	no	-	C3	S
cy60_1_cd_45_pos_HC_pos_HLDAR_pos_H03_S1047_comb	Bcell_no	CY60	Bcell	no	-	C3	G2M

J) More complete information about samples and their phenotype grouping can be found in the sample table.

# Clustering module: heatmap tab

A) The clustering module performs a holistic clustering analysis of the samples. The main output of this feature is twofold: i) It generates a sophisticated heatmap of samples and ii) It also provides a PCA/tSNE plot of samples obtained by PCA or tSNE algorithms.

B) The heatmap analysis can be performed on a gene level expression or gene set level expression in which, for each gene set (or pathway), an average expression is computed from the gene expression data using summary methods such as GSVA and ssGSEA.



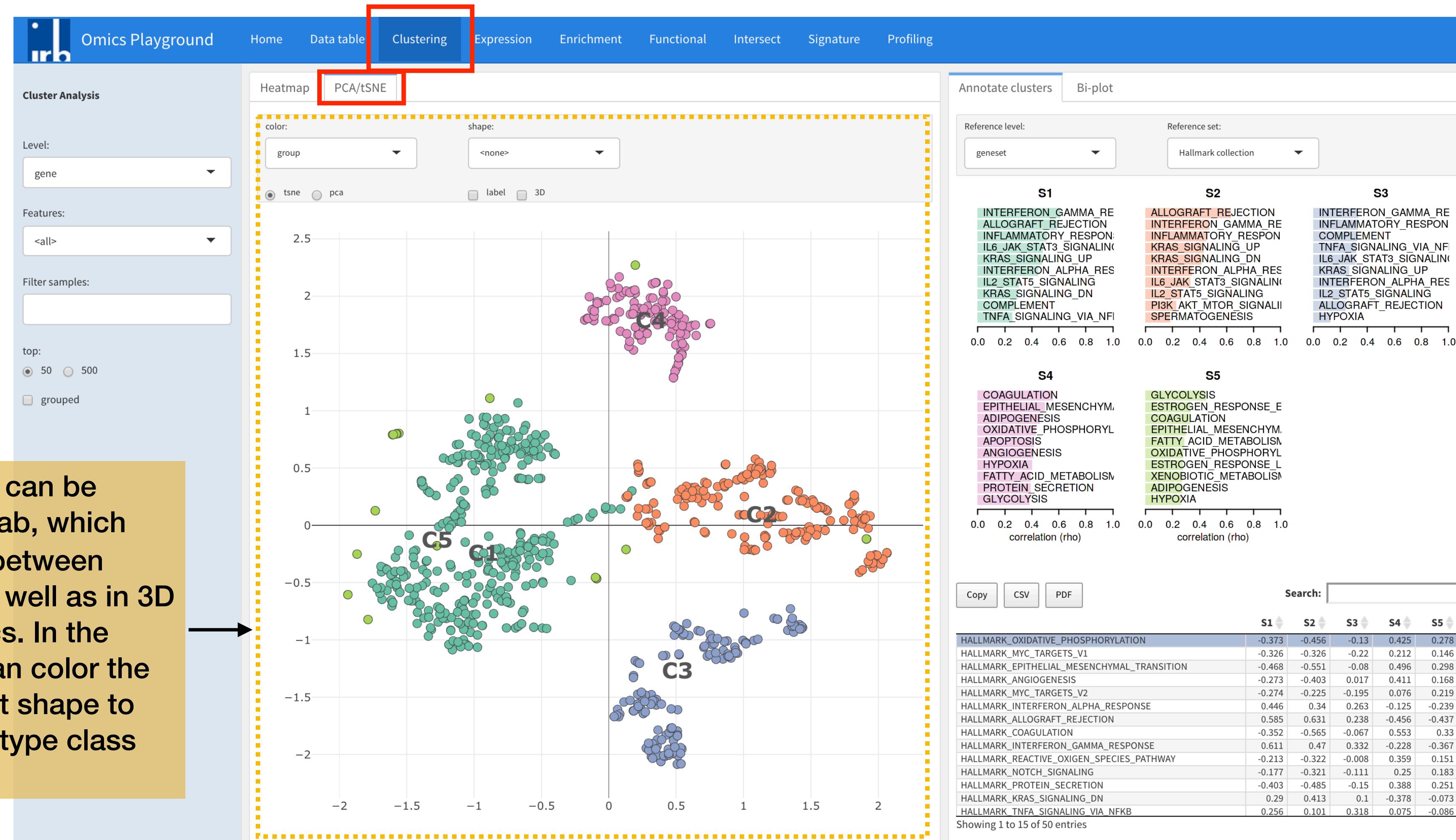
C) During the heatmap generation, users have various option that they can select, such as splitting the samples for a phenotype class provided in the data (eg, tissue, cell type, or gender). In addition, users have to specify the top  $N = \{50, 500\}$  features to be used in the heatmap for hierarchical clustering.

D) The criteria to select the top features are: 1) sd - features with the highest standard deviation across all the samples, 2) biomarker - features that are overexpressed in each phenotype class compared to the rest, 3) pca - principal components computed by `irlba` package.

E) Then the top features in the heatmap are divided into 5 clusters in terms of their expression. For each cluster, the platform provides a functional annotation under annotate cluster section using more than 42 reference databases from the literature, including but not limited to well-known databases such as MSigDB, KEGG and GO.

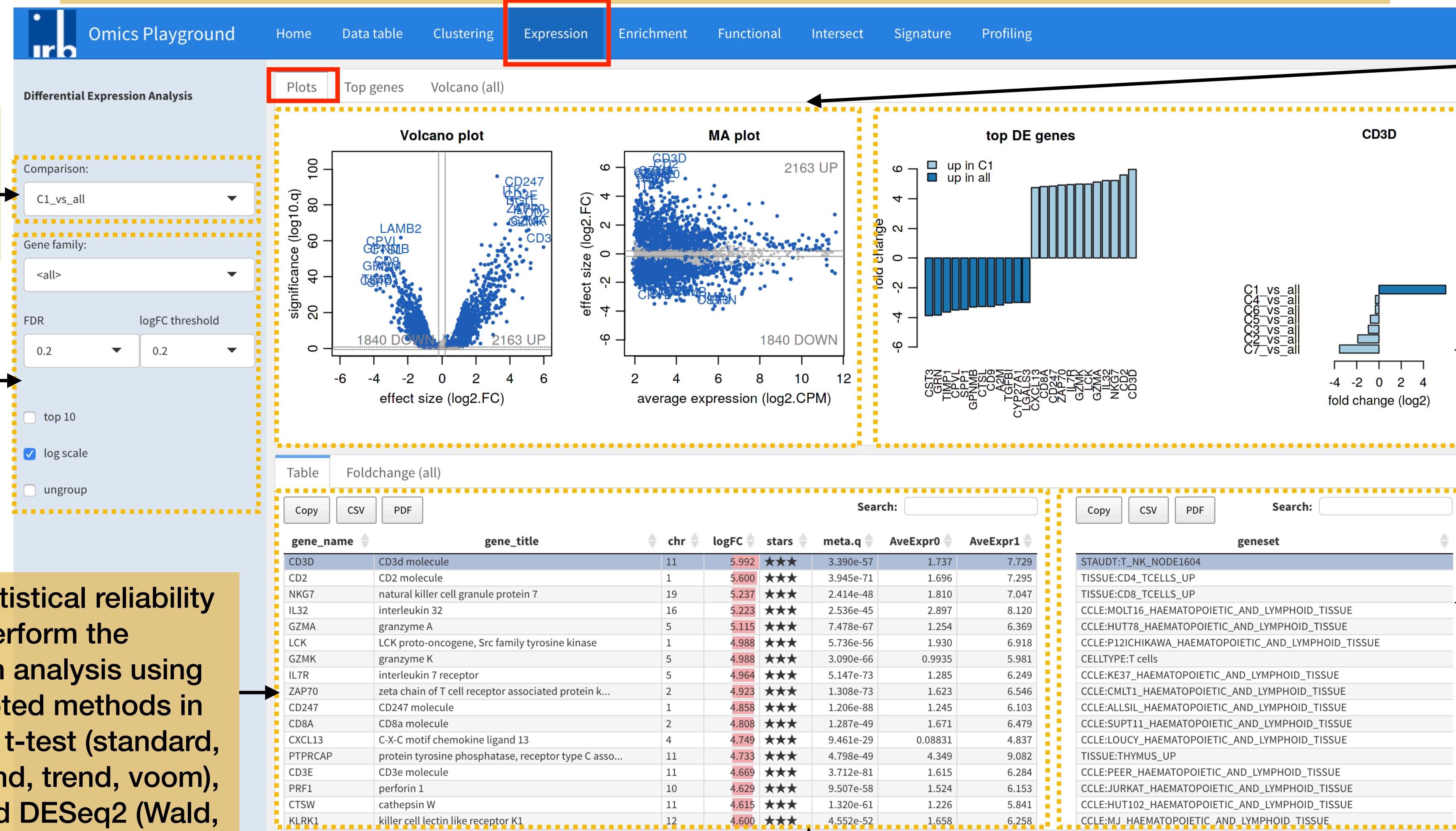
F) Furthermore, users can see the enrichment scores for the reference set used in annotating clusters

# Clustering module: PCA/tSNE tab



# Expression module: plots tab

A) This page hosts a differential gene expression analysis between conditions (i.e. tumor versus control), which is one of the fundamental analysis in the transcriptomics data analytics workflow.



G) The plots section also provides volcano plot and MA (an application of a Bland-Altman) plot.

F) By clicking on a gene, it is possible to see which genesets contain that gene from the geneset table located on the right , and check the differential expression status in other comparisons from the plots section.

E) For a selected comparison, the results of these methods are combined and reported under the table, where meta.q for a gene represents the worst q value among the three methods and the number of stars for a gene indicate how many methods identified significant q values ( $q < 0.05$ ). The table is interactive (scrollable, clickable); users can sort genes by logFC, meta.q, or average expression in either conditions.

# Expression module: top genes tab

**Differential Expression Analysis**

Comparison: C1\_vs\_all

Gene family: <all>

FDR logFC threshold: 0.2

top 10

log scale

ungroup

Plots Top genes Volcano (all)

Table Foldchange (all)

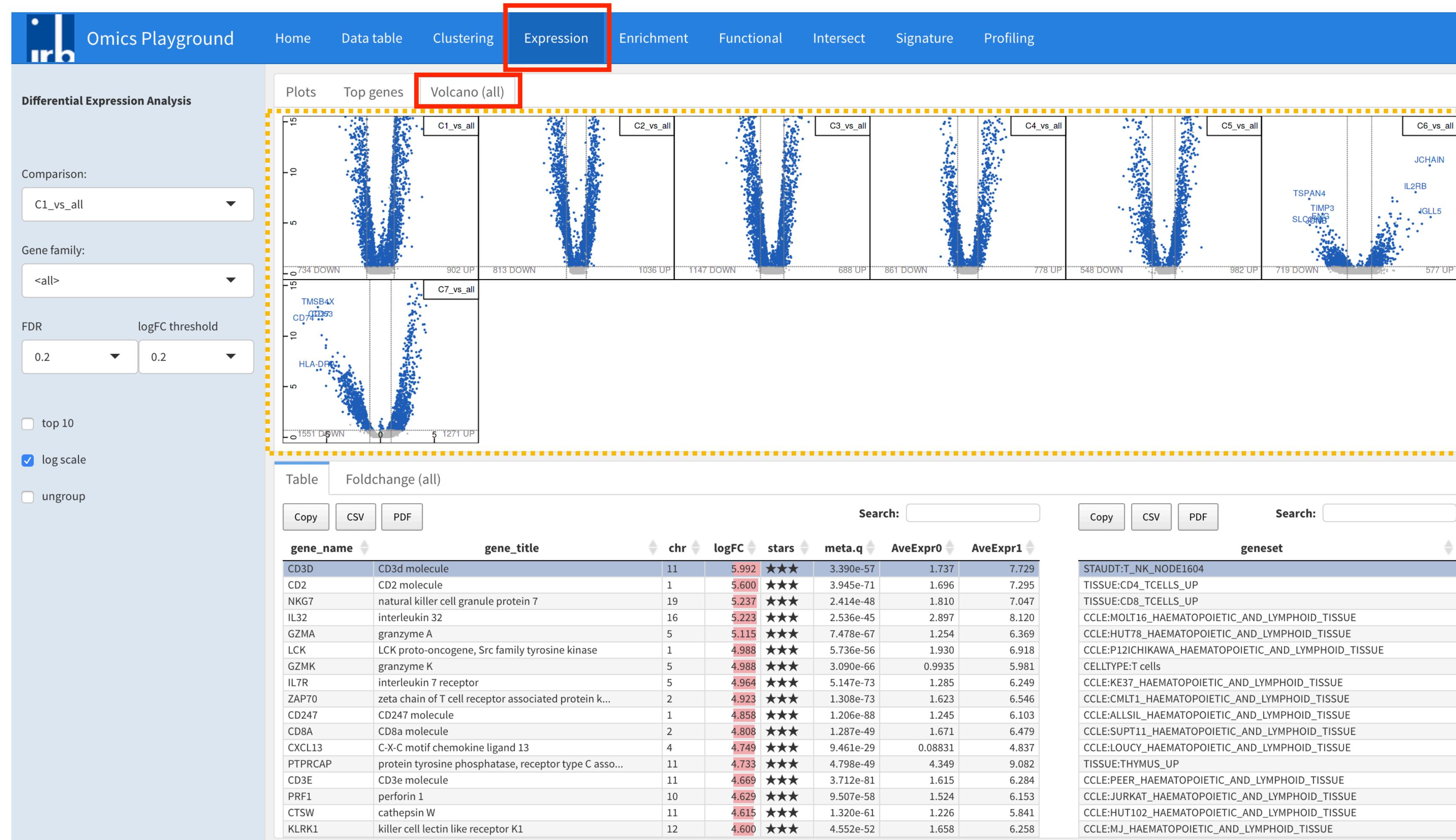
gene_name	gene_title	chr	logFC	stars	meta.q	AveExpr0	AveExpr1
CD3D	CD3d molecule	11	5.992	★★★	3.390e-57	1.737	7.729
CD2	CD2 molecule	1	5.600	★★★	3.945e-71	1.696	7.295
NKG7	natural killer cell granule protein 7	19	5.237	★★★	2.414e-48	1.810	7.047
IL32	interleukin 32	16	5.223	★★★	2.536e-45	2.897	8.120
GZMA	granzyme A	5	5.115	★★★	7.478e-67	1.254	6.369
LCK	LCK proto-oncogene, Src family tyrosine kinase	1	4.988	★★★	5.736e-56	1.930	6.918
GZMK	granzyme K	5	4.988	★★★	3.090e-66	0.9935	5.981
IL7R	interleukin 7 receptor	5	4.964	★★★	5.147e-73	1.285	6.249
ZAP70	zeta chain of T cell receptor associated protein k...	2	4.923	★★★	1.308e-73	1.623	6.546
CD247	CD247 molecule	1	4.858	★★★	1.206e-88	1.245	6.103
CD8A	CD8a molecule	2	4.808	★★★	1.287e-49	1.671	6.479
CXCL13	C-X-C motif chemokine ligand 13	4	4.749	★★★	9.461e-29	0.08831	4.837
PTPRCAP	protein tyrosine phosphatase, receptor type C asso...	11	4.733	★★★	4.798e-49	4.349	9.082
CD3E	CD3e molecule	11	4.669	★★★	3.712e-81	1.615	6.284
PRF1	perforin 1	10	4.629	★★★	9.507e-58	1.524	6.153
CTSW	cathepsin W	11	4.615	★★★	1.320e-61	1.226	5.841
KLRK1	killer cell lectin like receptor K1	12	4.600	★★★	4.552e-52	1.658	6.258

geneset

- STAUDT:T\_NK\_NODE1604
- TISSUE:CD4\_TCELLS\_UP
- TISSUE:CD8\_TCELLS\_UP
- CCLE:MOLT16\_HAEMATOPOIETIC\_AND LYMPHOID TISSUE
- CCLE:HUT78\_HAEMATOPOIETIC\_AND LYMPHOID TISSUE
- CCLE:P12ICHIKAWA\_HAEMATOPOIETIC\_AND LYMPHOID TISSUE
- CELLTYPE:T cells
- CCLE:KE37\_HAEMATOPOIETIC\_AND LYMPHOID TISSUE
- CCLE:CMLT1\_HAEMATOPOIETIC\_AND LYMPHOID TISSUE
- CCLE:ALLSIL\_HAEMATOPOIETIC\_AND LYMPHOID TISSUE
- CCLE:SUPT11\_HAEMATOPOIETIC\_AND LYMPHOID TISSUE
- CCLE:LOUCY\_HAEMATOPOIETIC\_AND LYMPHOID TISSUE
- TISSUE:THYMUS\_UP
- CCLE:PEER\_HAEMATOPOIETIC\_AND LYMPHOID TISSUE
- CCLE:JURKAT\_HAEMATOPOIETIC\_AND LYMPHOID TISSUE
- CCLE:HUT102\_HAEMATOPOIETIC\_AND LYMPHOID TISSUE
- CCLE:MJ\_HAEMATOPOIETIC\_AND LYMPHOID TISSUE

H) Furthermore, for top 10 differentially expressed genes within the selected comparison, average expression plots across the samples are displayed in top genes subsection.

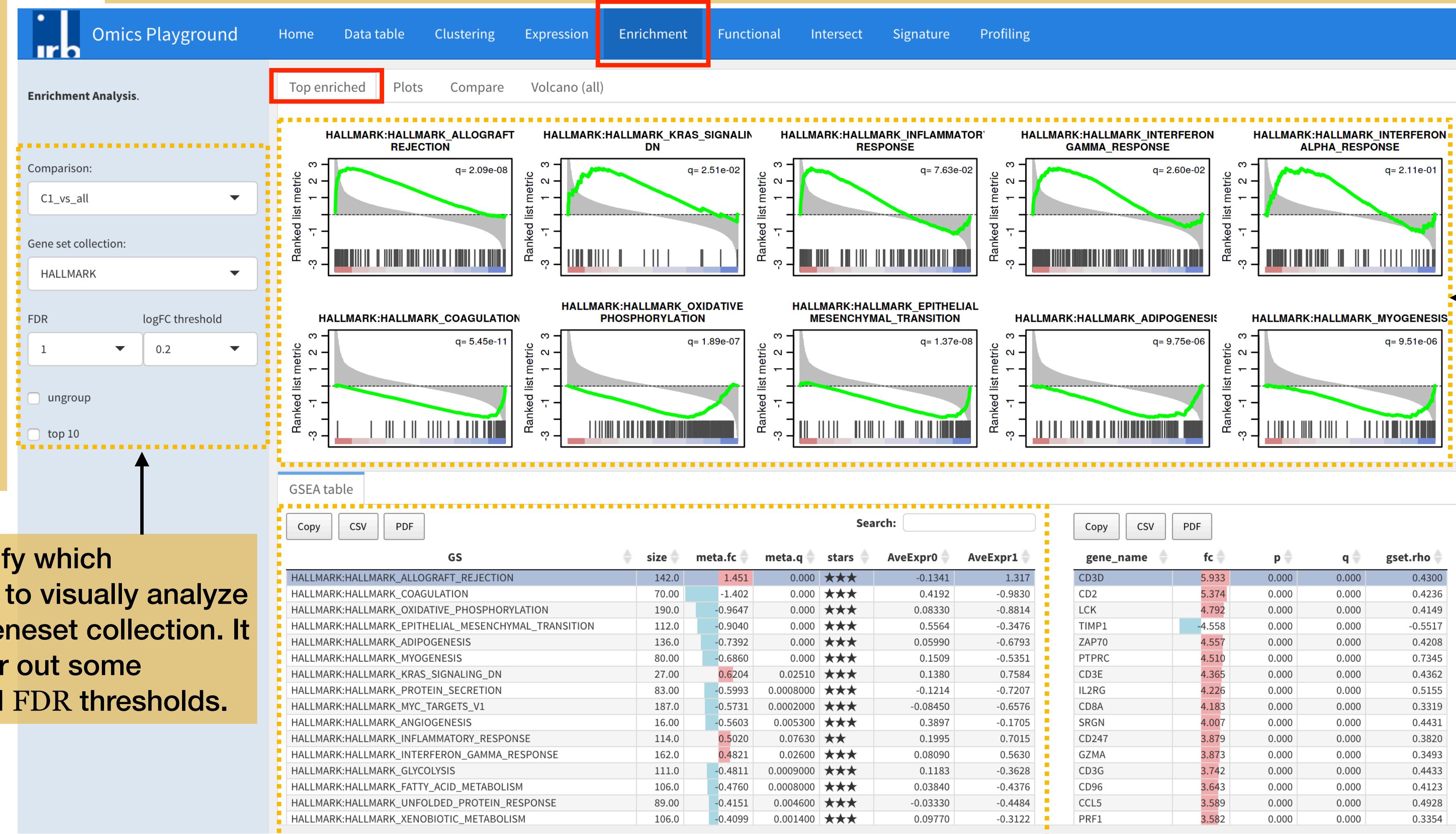
# Expression module: volcano (all) tab



I) Another important feature of the platform is that it can display volcano plots for all comparison simultaneously under **volcano (all)** section. This provides users an overview of a complete comparison space at once and facilitates to drive conclusions.

# Enrichment module: top enriched tab

A) Similar to the differential gene expression analysis, users can perform differential expression analysis on a geneset level in this page, which is referred to as gene set enrichment (GSE) analysis.



B) To perform the GSE, first, an expression data for each geneset (or pathway) is computed from the gene expression using summary methods such as GSVA and ssGSEA. The platform has more than 50.000 genesets (or pathways) in total that are divided into 30 geneset collections such as Hallmark, MSigDB, KEGG and GO.

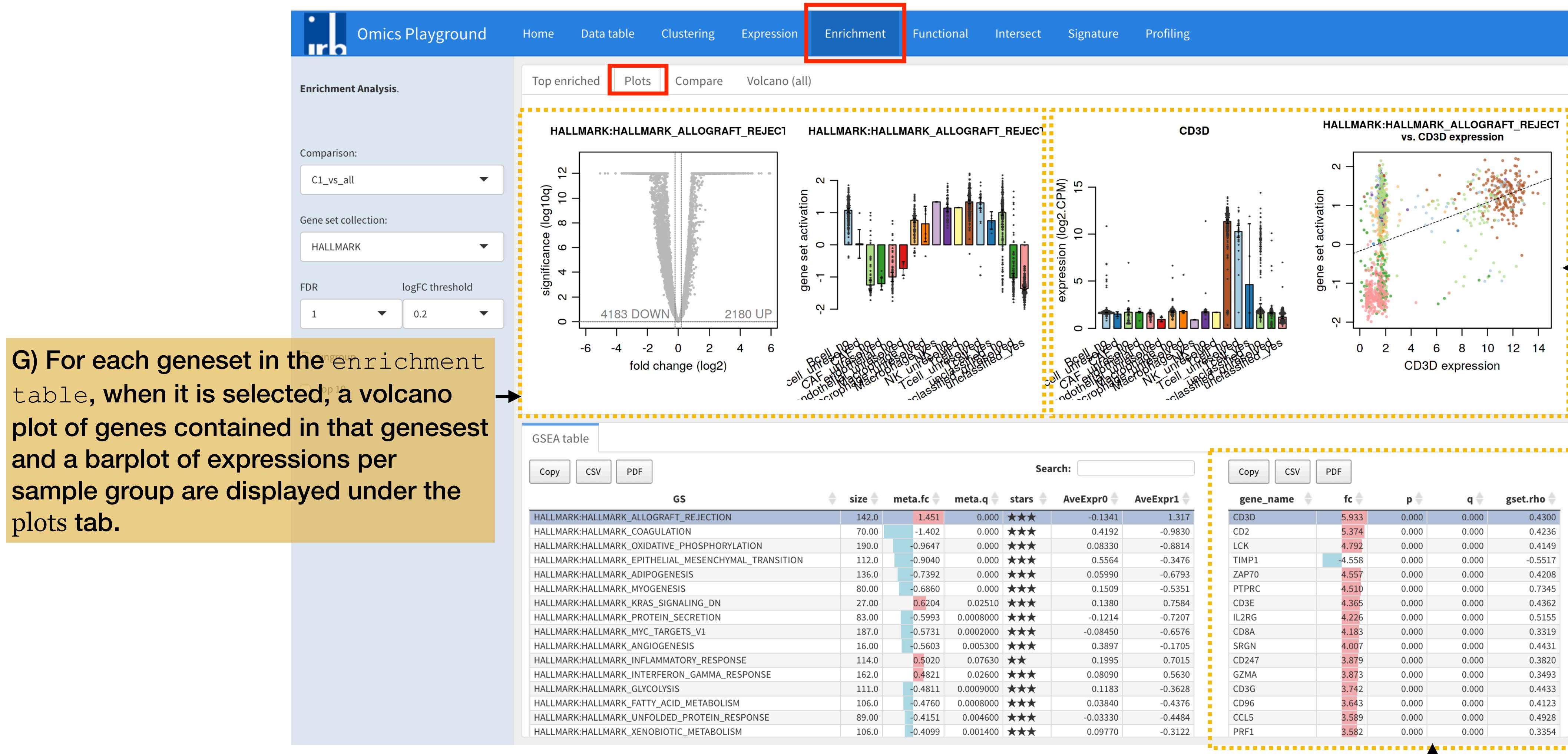
C) Users have to specify which comparison they want to visually analyze employing a certain geneset collection. It is also possible to filter out some genesets by logFC and FDR thresholds.

D) The platform performs enrichment analyses using seven different methods including Spearman rank correlation, GSVA, ssGSEA, Fisher's exact test, GSEA, camera and fry, and reports the combined result.

E) Then the combined result from the methods is displayed under enrichment table, where for each geneset the meta.q corresponds to the highest q value provided by the methods and the number of stars indicate how many methods identified the geneset as significant ( $q < 0.05$ ). The table is interactive; users can sort it by logFC, meta.q and starts.

F) For a selected comparison, top 10 differentially enriched geneses (pathways) are displayed under top enriched section.

# Enrichment module: plots tab



**H)** Additionally, the list of genes in that geneset are displayed in the second table on the right, and for every gene in that table,

**I)** It is possible to see the barplot of expressions per sample group and scatter plot of gene to geneset expressions under the plots tab.

# Enrichment module: compare tab

Omics Playground

Home Data table Clustering Expression **Enrichment** Functional Intersect Signature Profiling

**Enrichment Analysis.**

Comparison: C1\_vs\_all

Gene set collection: HALLMARK

FDR logFC threshold: 1 0.2

ungroup  
 top 10

Top enriched Plots Compare Volcano (all)

GSEA table

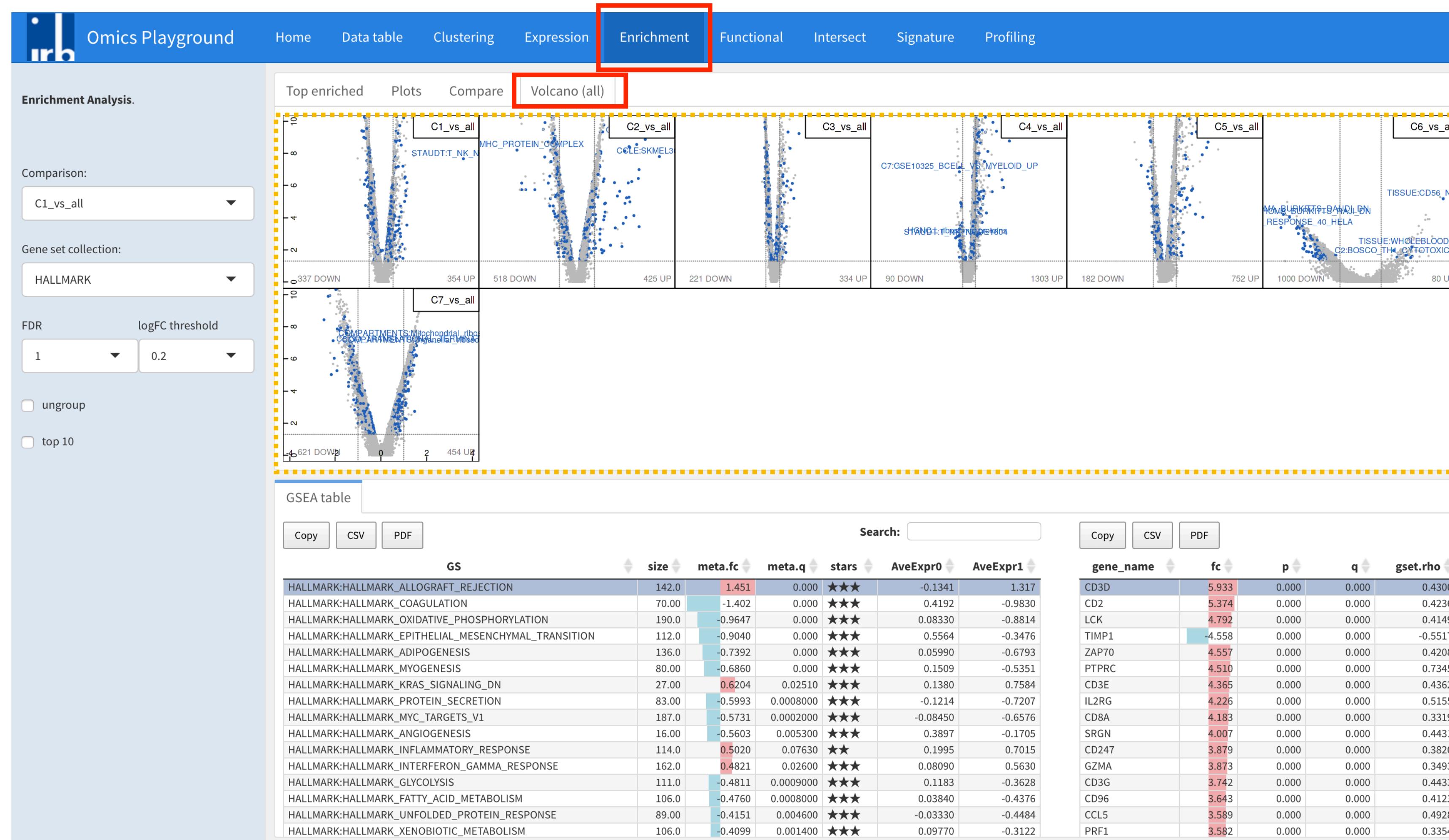
GS	size	meta.fc	meta.q	stars	AveExpr0	AveExpr1
HALLMARK:HALLMARK_ALLOGRAFT_REJECTION	142.0	1.451	0.000	★★★	-0.1341	1.317
HALLMARK:HALLMARK_COAGULATION	70.00	-1.402	0.000	★★★	0.4192	-0.9830
HALLMARK:HALLMARK_OXIDATIVE_PHOSPHORYLATION	190.0	-0.9647	0.000	★★★	0.08330	-0.8814
HALLMARK:HALLMARK_EPITHELIAL_MESENCHYMAL_TRANSITION	112.0	-0.9040	0.000	★★★	0.5564	-0.3476
HALLMARK:HALLMARK_ADIPGENESIS	136.0	-0.7392	0.000	★★★	0.05990	-0.6793
HALLMARK:HALLMARK_MYOGENESIS	80.00	-0.6860	0.000	★★★	0.1509	-0.5351
HALLMARK:HALLMARK_KRAS_SIGNALING_DN	27.00	0.6204	0.02510	★★★	0.1380	0.7584
HALLMARK:HALLMARK_PROTEIN_SECRETION	83.00	-0.5993	0.0008000	★★★	-0.1214	-0.7207
HALLMARK:HALLMARK_MYC_TARGETS_V1	187.0	-0.5731	0.0002000	★★★	-0.08450	-0.6576
HALLMARK:HALLMARK_ANGIOGENESIS	16.00	-0.5603	0.005300	★★★	0.3897	-0.1705
HALLMARK:HALLMARK_INFLAMMATORY_RESPONSE	114.0	0.5020	0.07630	★★	0.1995	0.7015
HALLMARK:HALLMARK_INTERFERON_GAMMA_RESPONSE	162.0	0.4821	0.02600	★★★	0.08090	0.5630
HALLMARK:HALLMARK_GLYCOLYSIS	111.0	-0.4811	0.0009000	★★★	0.1183	-0.3628
HALLMARK:HALLMARK_FATTY_ACID_METABOLISM	106.0	-0.4760	0.0008000	★★★	0.03840	-0.4376
HALLMARK:HALLMARK_UNFOLDED_PROTEIN_RESPONSE	89.00	-0.4151	0.004600	★★★	-0.03330	-0.4484
HALLMARK:HALLMARK_XENOBIOTIC_METABOLISM	106.0	-0.4099	0.001400	★★★	0.09770	-0.3122

Search:

gene_name	fc	p	q	gset.rho
CD3D	5.933	0.000	0.000	0.4300
CD2	5.374	0.000	0.000	0.4236
LCK	4.792	0.000	0.000	0.4149
TIMP1	-4.558	0.000	0.000	-0.5517
ZAP70	4.557	0.000	0.000	0.4208
PTPRC	4.510	0.000	0.000	0.7345
CD3E	4.365	0.000	0.000	0.4362
IL2RG	4.226	0.000	0.000	0.5155
CD8A	4.183	0.000	0.000	0.3319
SRGN	4.007	0.000	0.000	0.4431
CD247	3.879	0.000	0.000	0.3820
GZMA	3.873	0.000	0.000	0.3493
CD3G	3.742	0.000	0.000	0.4433
CD96	3.643	0.000	0.000	0.4123
CCL5	3.589	0.000	0.000	0.4928
PRF1	3.582	0.000	0.000	0.3354

J) For a selected geneset from the enrichment table, users can compare the differential expression status of that geneset for all other comparisons under the **Compare** section.

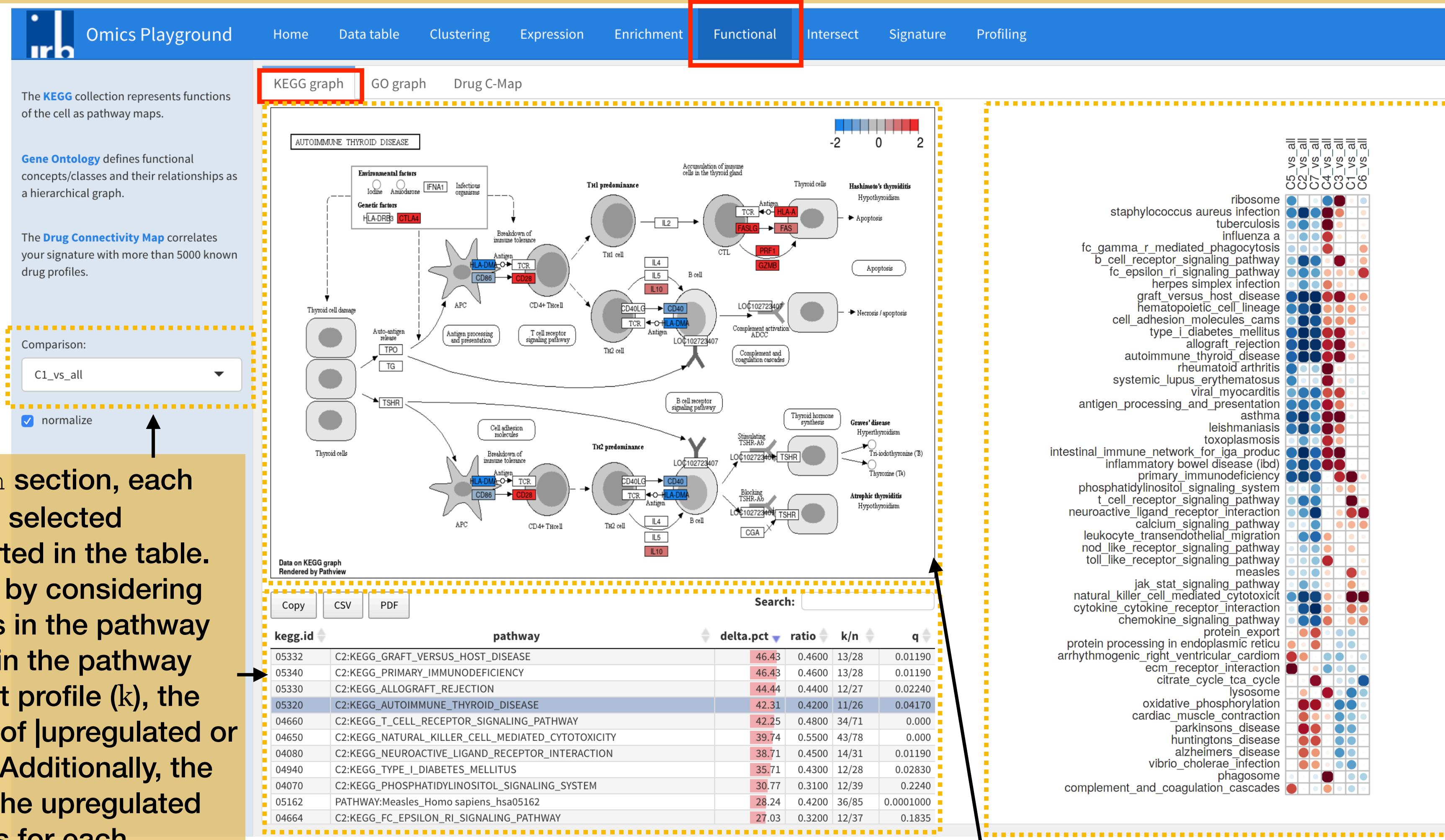
# Enrichment module: volcano (all) tab



K) Similarly, volcano plots of genesets for all comparisons are displayed under the volcano (all) tab. This allows users to have an overall picture across comparisons at the same time.

# Functional module: KEGG graph tab

A) This module provides higher level functional and visual analysis of the contrast space using the KEGG and GO graph structures. Given the profile of a particular contrast, it also searches for the closest drug profiles from the L1000 drug expression database.



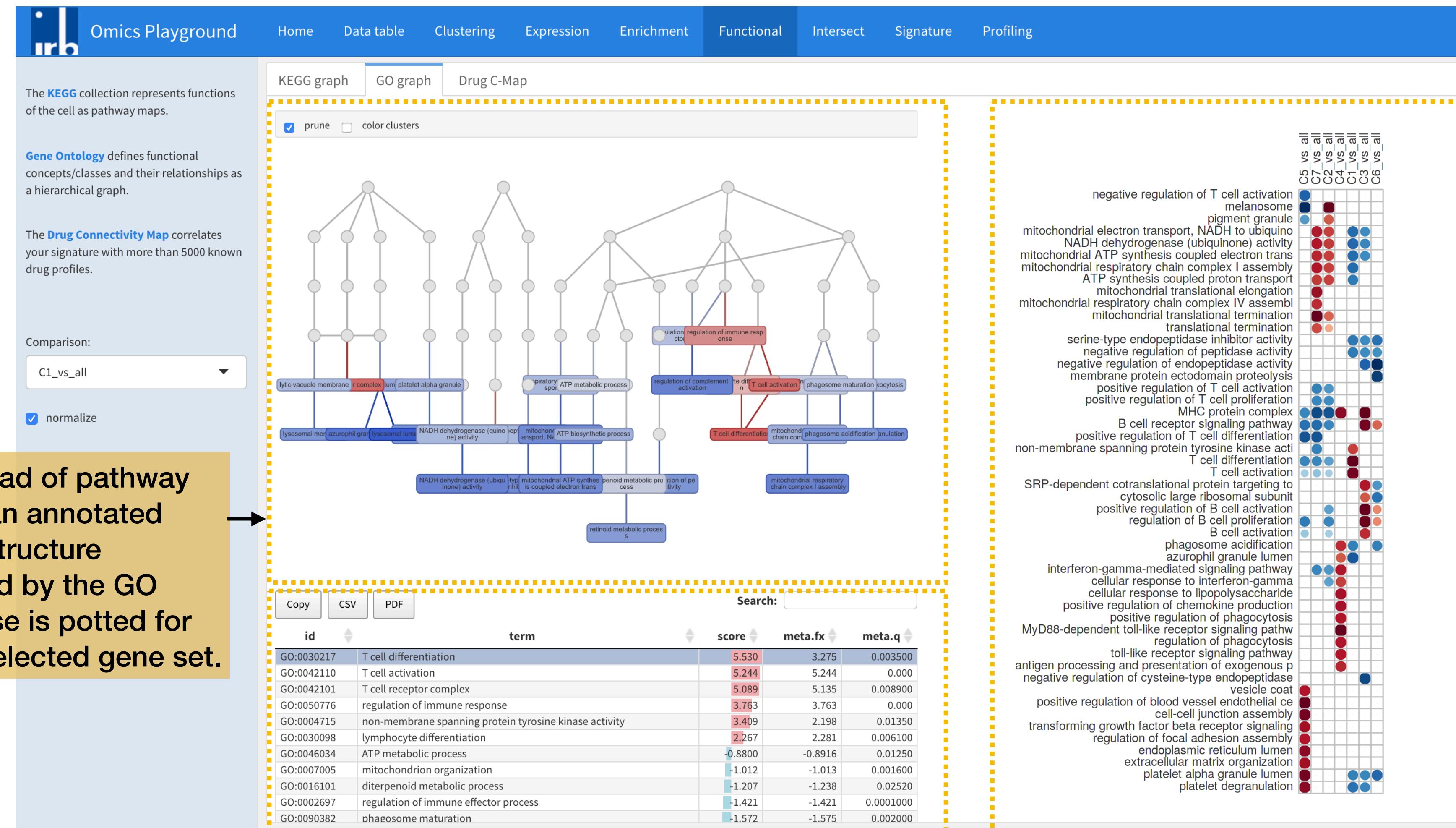
**B) KEGG** is a collection of manually curated pathways representing the current knowledge of molecular interactions, reactions and relation networks as pathway maps.

C) Within the KEGG graph section, each pathway is scored for the selected contrast profile and reported in the table. The scoring is performed by considering the total number of genes in the pathway ( $n$ ), the number of genes in the pathway supported by the contrast profile ( $k$ ), the ratio of  $k/n$ , and the ratio of |upregulated or downregulated genes|/ $k$ . Additionally, the table contains the list of the upregulated and downregulated genes for each pathway and a q value from the Fisher's test for the overlap. The table is interactive; enabling user to sort on different variables.

D) It is possible to visualize the pathway map by clicking on the pathway in the table, where genes are colored according to their upregulation (red) or downregulation (blue) in the contrast profile.

E) Another important feature of the platform is that it provides an activation-heatmap including the comparison of activation levels of pathways (or pathway keywords) across multiple contrast profiles. This facilitates to quickly see and detect the similarities between profiles in certain pathways.

# Functional module: GO graph tab



F) The GO database provides a computational representation of the current knowledge about roles of genes for many organisms in terms of molecular functions, cellular components and biological processes. All the features described under the KEGG graph tab, such as scoring the gene sets and drawing an activation-heatmap, can be performed for the GO database under the GO graph tab.

# Functional module: drug connectivity map (C-Map) tab

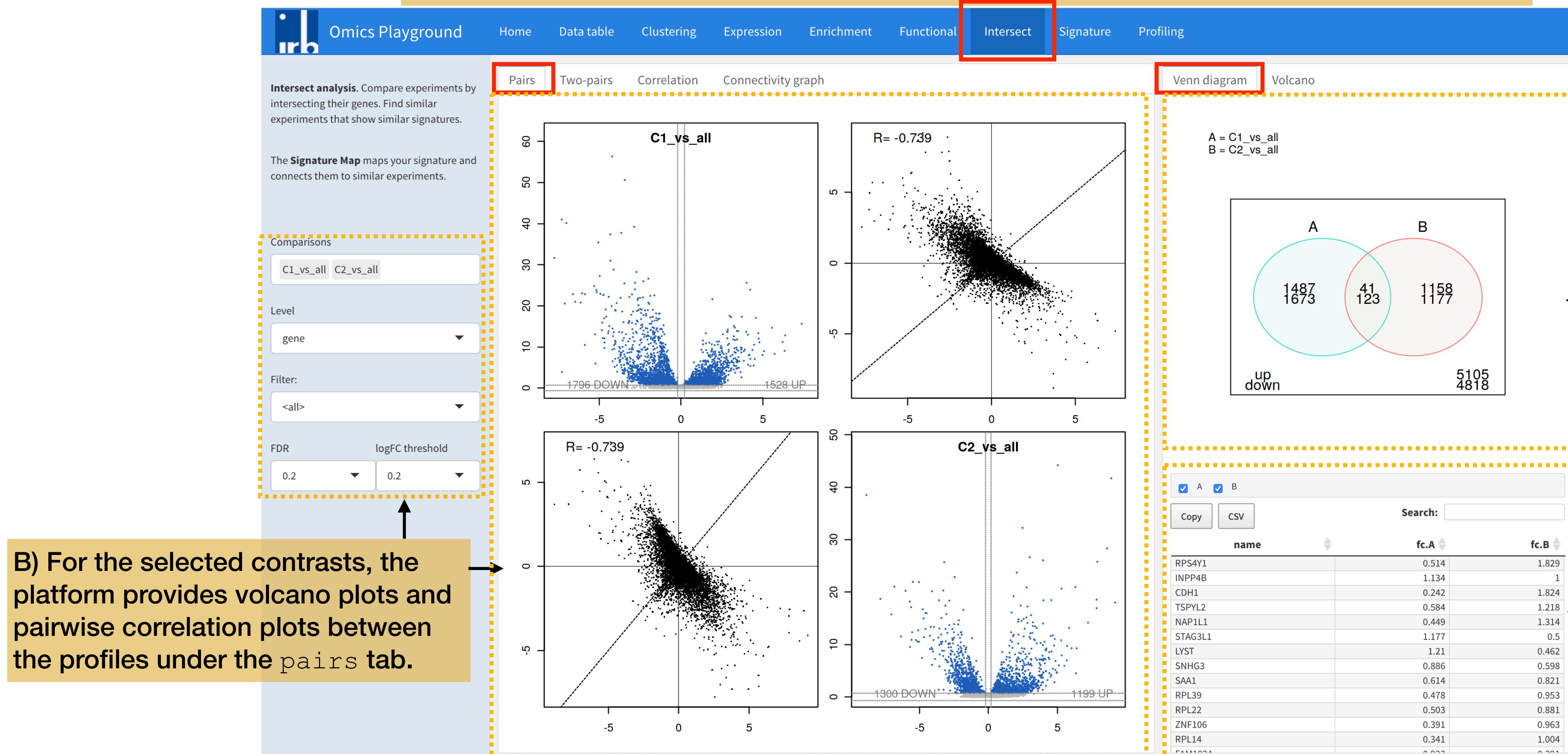
**H) The drug connectivity map (Drug C-Map) section correlates the selected contrast profile with more than 5000 known drug profiles from the L1000 database, and shows the top N=10 similar and opposite profiles by running GSEA algorithm on the contrast-drug profile correlation space.**

drug	NES	pval	padj	moa	target
CGP-60474	-3.5028	0.0031	0.0687	CDK inhibitor	CDK1 CDK2
emetine	3.4969	0.0013	0.0687	protein synthesis inhibitor	RPS2
narciclasine	3.4775	0.0014	0.0687		
AT-7519	-3.2489	0.003	0.0687	CDK inhibitor	CDK1 CDK2 CDK4 CDK5 CDK6 CDK7
triptolide	-3.1729	0.0035	0.0712	RNA polymerase inhibitor	RELA
BMS-387032	-3.1695	0.003	0.0687	CDK inhibitor cell cycle inhibitor MCL1 inhibitor	CDK2 CDK7 CDK9
PIK-75	-3.1236	0.0032	0.0688	DNA protein kinase inhibitor PI3K inhibitor	PIK3CA PIK3CB PIK3CD PIK3CG PIK3R1
homoharringtonine	2.9859	0.0016	0.0687	protein synthesis inhibitor	RPL3
alvocidib	-2.9799	0.0032	0.0688	CDK inhibitor	CDK1 CDK2 CDK4 CDK5 CDK6 CDK7
JNK-9L	-2.9201	0.0029	0.0687		
PAC-1	-2.721	0.0031	0.0687	caspase activator	CASP3
cephaeline	2.7201	0.0016	0.0687		

I) Another important feature of the platform is that it provides an activation-heatmap including the comparison of activation levels of pathways (or pathway keywords) across multiple contrast profiles. This facilitates to quickly see and detect the similarities between profiles in certain pathways.

# Intersection module: pairs + Venn diagram tabs

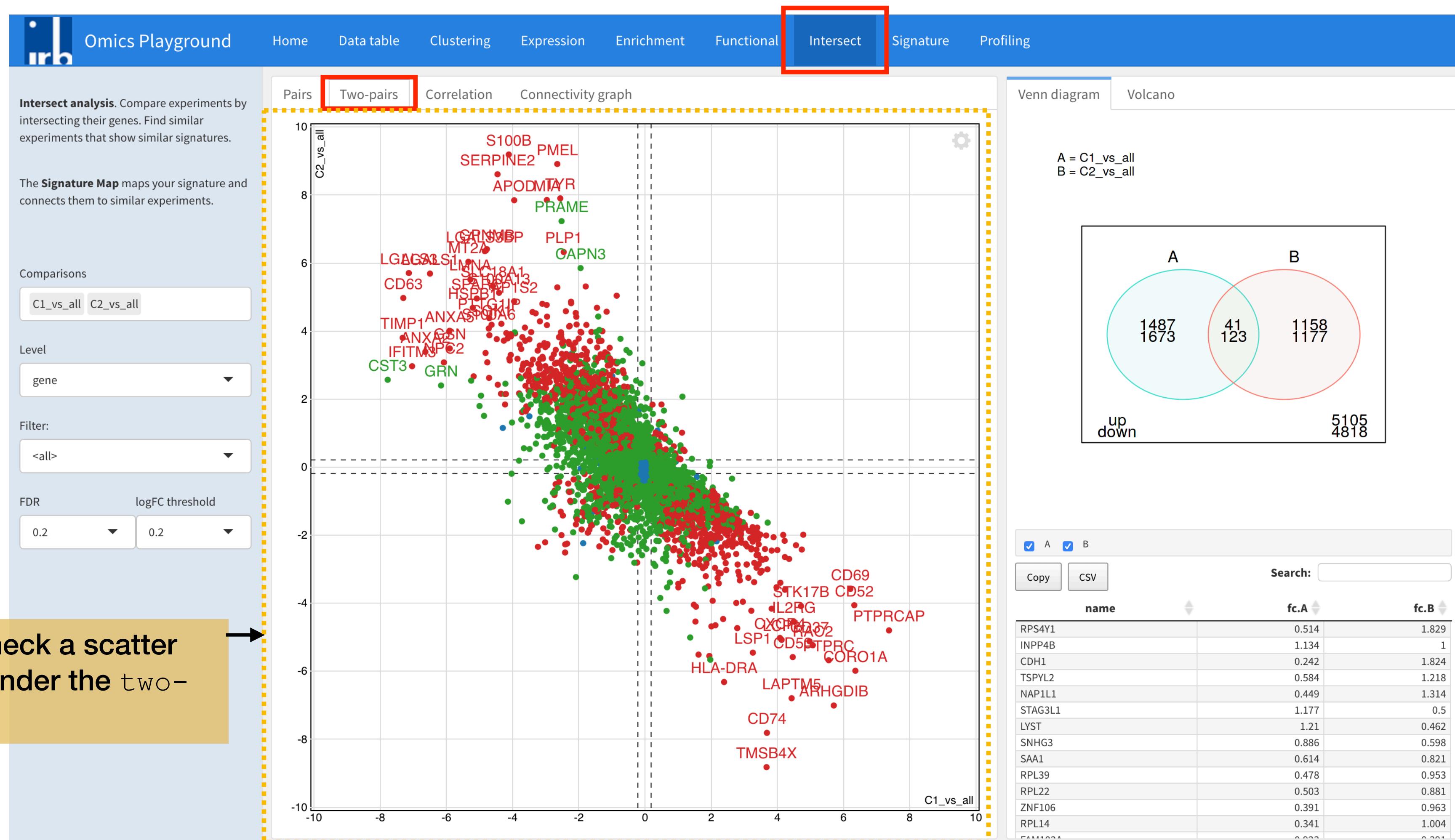
A) Intersection analysis module enables users to compare multiple contrasts by intersecting the genes of profiles. The main goal is to identify contrasts showing similar profiles.



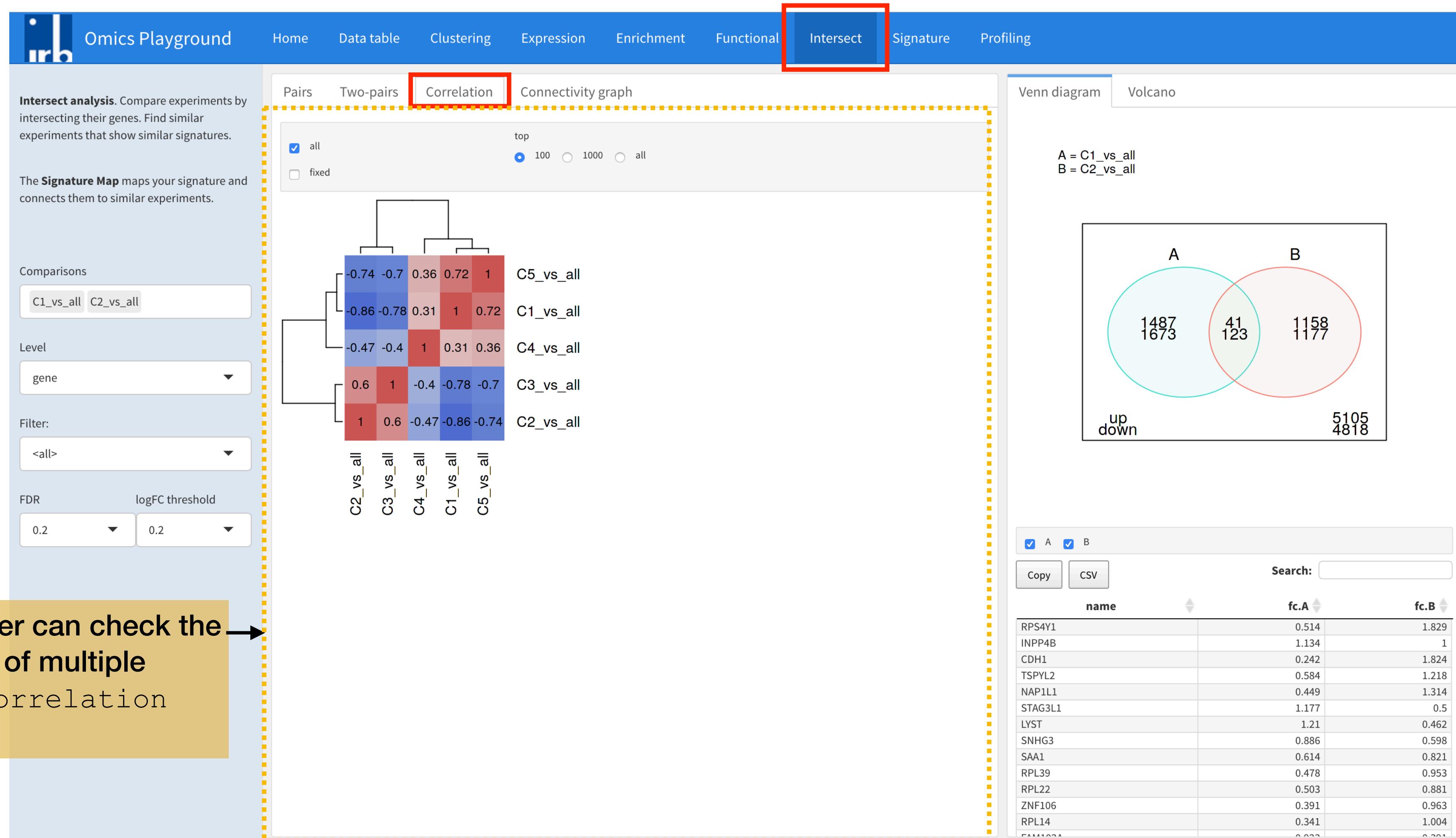
C) Simultaneously, it plots a Venn diagram with the number of intersecting genes between the profiles in venn diagram section.

D) The list of intersecting genes with further details is also reported in an interactive table below, where users can select and remove a particular contrast from the intersection analysis.

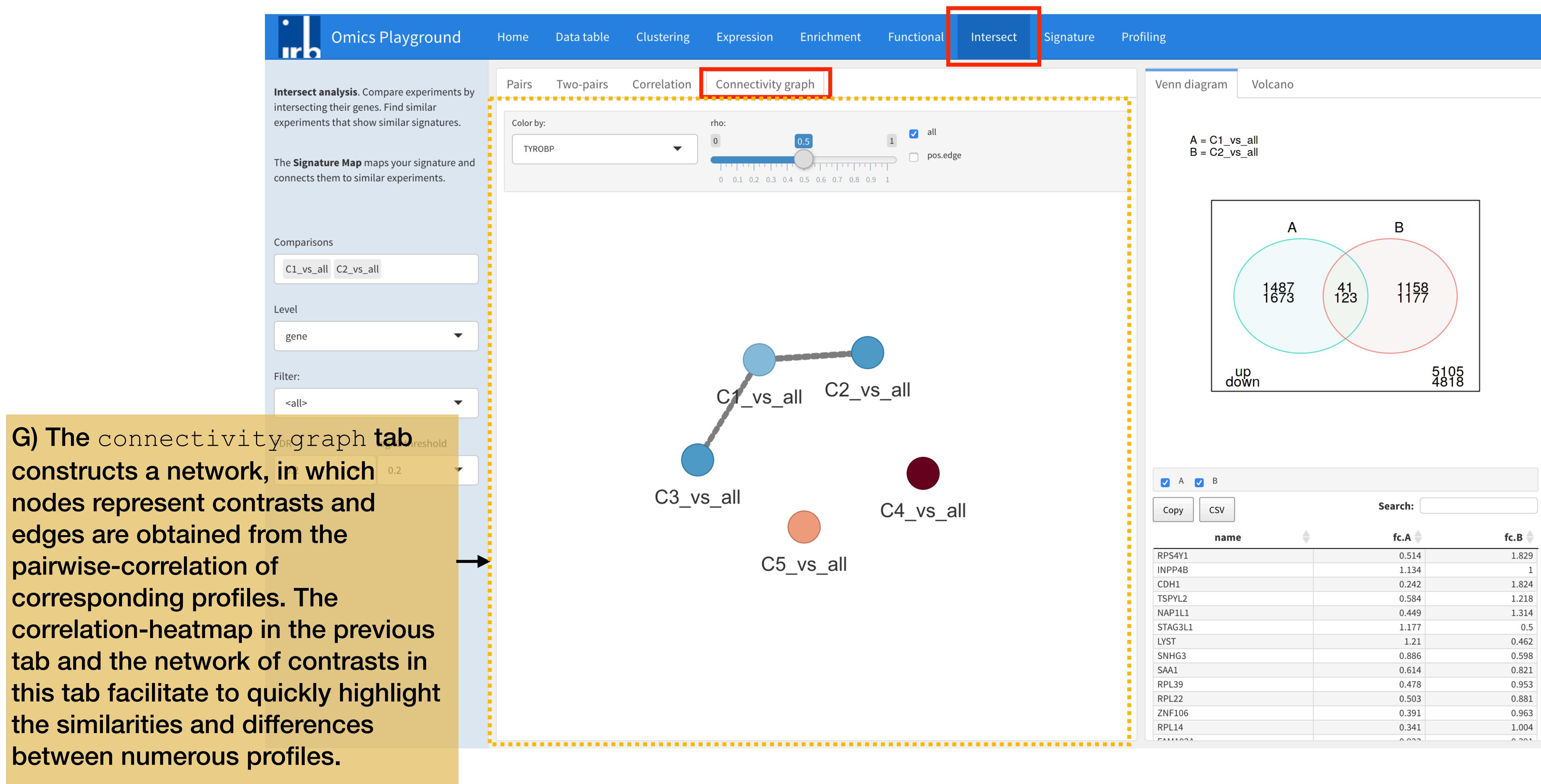
# Intersection module: two-pairs tab



# Intersection module: correlation tab

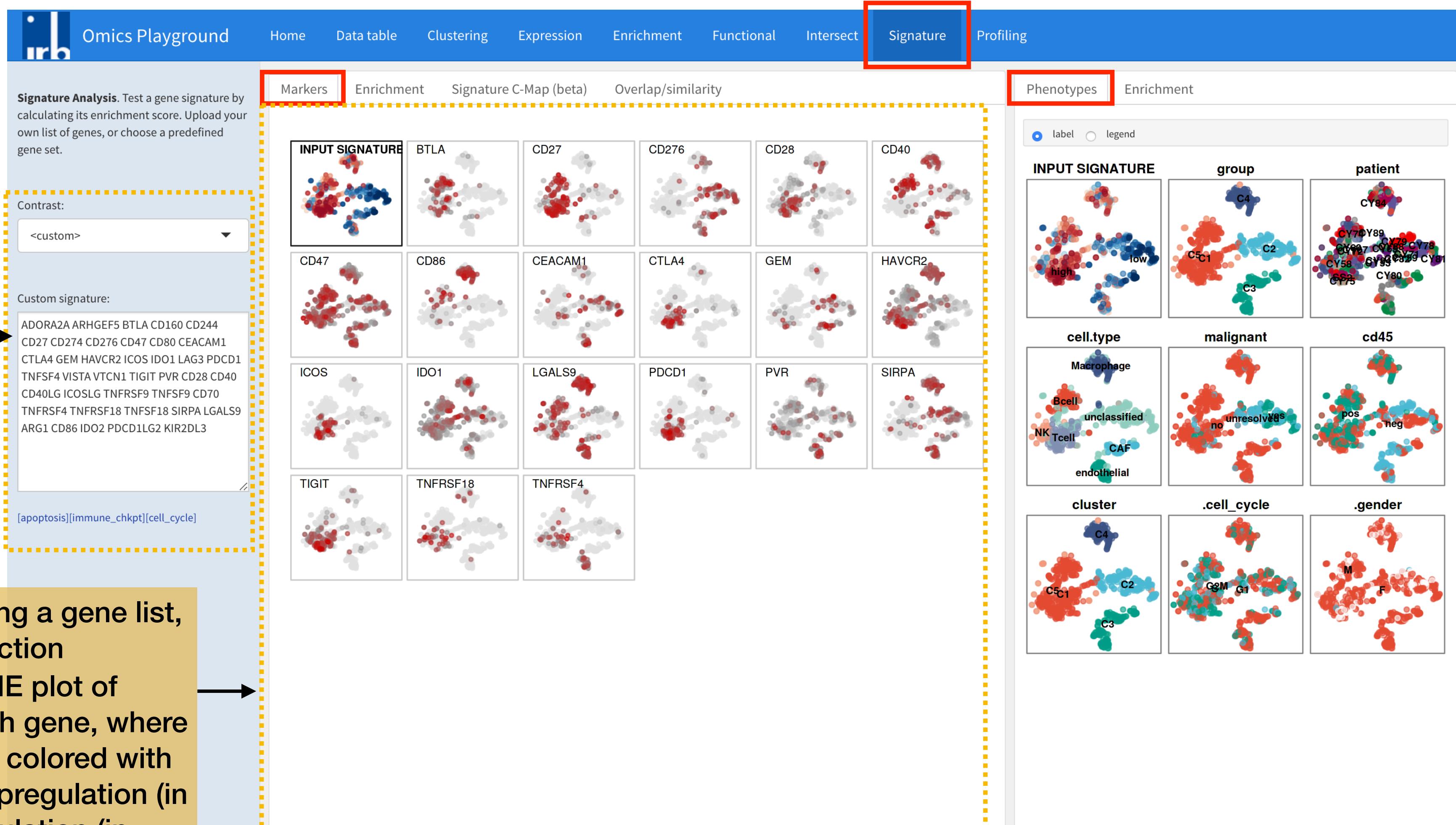


# Intersection module: connectivity graph tab



# Signature module: markers tab

A) In signature analysis module, users can test their gene signature by calculating an enrichment score.



B) They can use a sample list provided on the platform or upload their own gene list. Instead of a short list, a profile can also be selected, which is a complete gene list resulted from one of the contrasts in the analysis.

C) After uploading a gene list, the markers section produces a t-SNE plot of samples for each gene, where the samples are colored with respect to the upregulation (in red) or downregulation (in blue) of that particular gene.

# Signature module: enrichment tab

**D) The enrichment tab performs the enrichment analysis of the gene list against all contrasts by running the GSEA algorithm and plots enrichment outputs.**

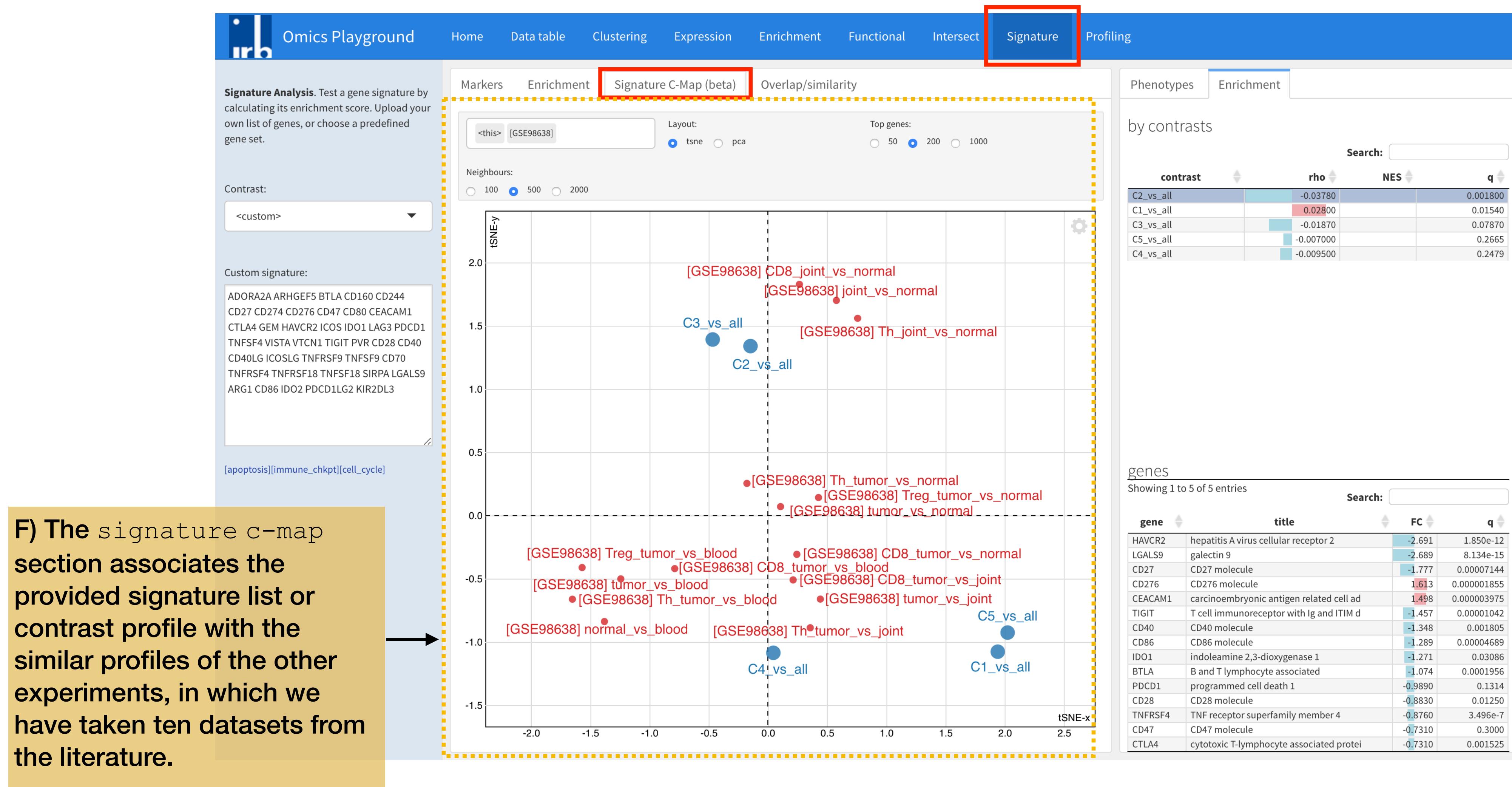
**E) The enrichment statistics can be found in this table**

contrast	rho	NES	q
C2_vs_all	-0.03780	0.001800	
C1_vs_all	0.02800	0.01540	
C3_vs_all	-0.01870	0.07870	
C5_vs_all	-0.007000	0.2665	
C4_vs_all	-0.009500	0.2479	

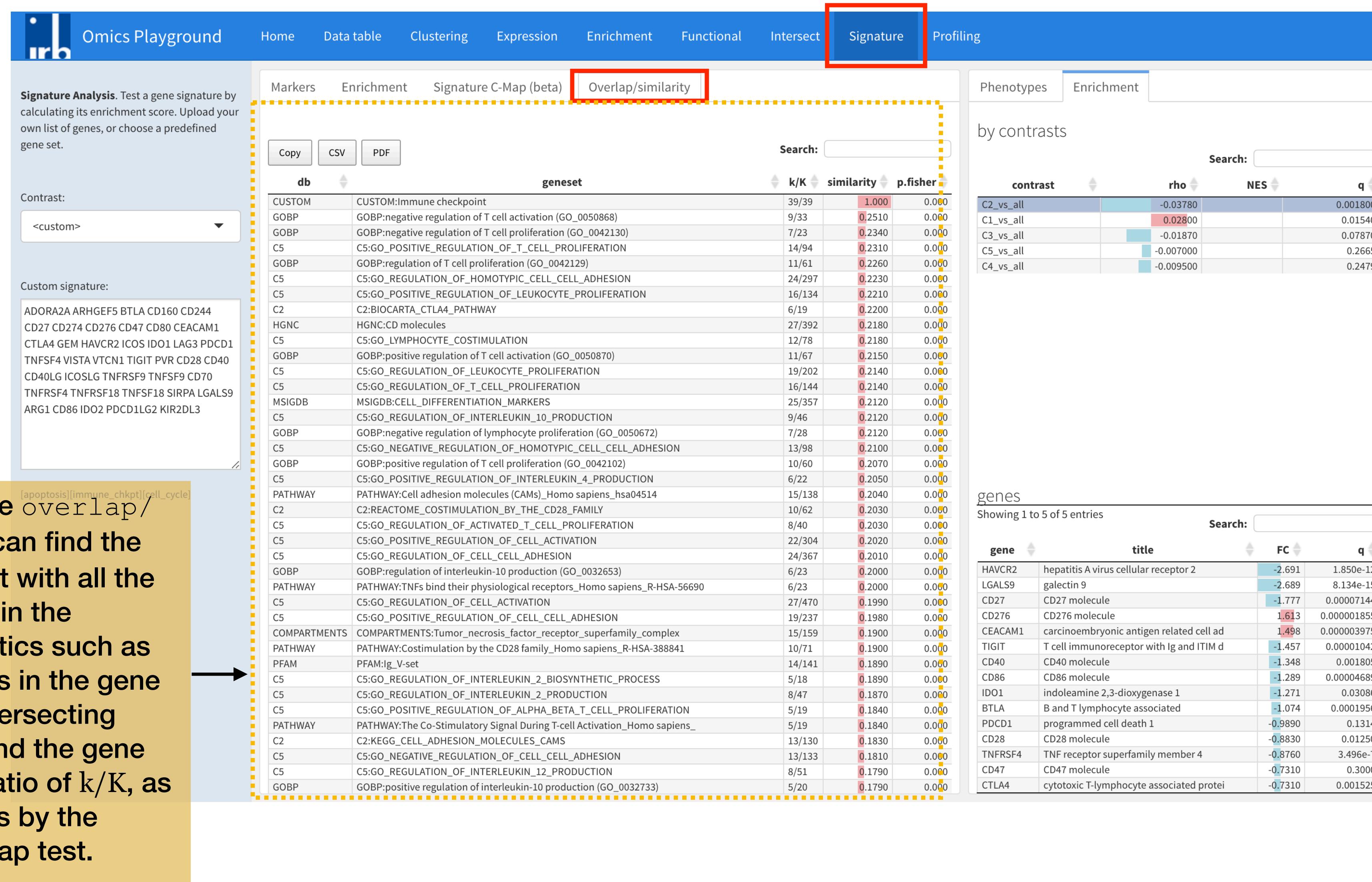
gene	title	FC	q
HAVCR2	hepatitis A virus cellular receptor 2	-2.691	1.850e-12
LGALS9	galectin 9	-2.689	8.134e-15
CD27	CD27 molecule	-1.777	0.00007144
CD276	CD276 molecule	1.613	0.000001855
CEACAM1	carcinoembryonic antigen related cell ad	1.498	0.000003975
TIGIT	T cell immunoreceptor with Ig and ITIM d	-1.457	0.00001042
CD40	CD40 molecule	-1.348	0.001805
CD86	CD86 molecule	-1.289	0.00004689
IDO1	indoleamine 2,3-dioxygenase 1	-1.271	0.03086
BTLA	B and T lymphocyte associated	-1.074	0.0001956
PDCD1	programmed cell death 1	-0.9890	0.1314
CD28	CD28 molecule	-0.8830	0.01250
TNFRSF4	TNF receptor superfamily member 4	-0.8760	3.496e-7
CD47	CD47 molecule	-0.7310	0.3000
CTLA4	cytotoxic T-lymphocyte associated protei	-0.7310	0.001525

# Signature module: signature connectivity map (c-map) tab



# Signature module: overlap/similarity tab

**G) Furthermore, under the overlap/similarity tab, users can find the similarity of their gene list with all the gene sets and pathways in the platform, including statistics such as the total number of genes in the gene set (K), the number of intersecting genes between the list and the gene set (k), the overlapping ratio of k/K, as well as the p and q values by the Fisher's test for the overlap test.**



db	geneset	k/K	similarity	p.fisher
CUSTOM	CUSTOM:Immune checkpoint	39/39	1.000	0.000
GOBP	GOBP:negative regulation of T cell activation (GO_0050868)	9/33	0.2510	0.000
GOBP	GOBP:negative regulation of T cell proliferation (GO_0042130)	7/23	0.2340	0.000
C5	C5:GO_POSITIVE_REGULATION_OF_T_CELL_PROLIFERATION	14/94	0.2310	0.000
GOBP	GOBP:regulation of T cell proliferation (GO_0042129)	11/61	0.2260	0.000
C5	C5:GO_REGULATION_OF_HOMOTYPIC_CELL_CELL_ADHESION	24/297	0.2230	0.000
C5	C5:GO_POSITIVE_REGULATION_OF_LEUKOCYTE_PROLIFERATION	16/134	0.2210	0.000
C2	C2:BIOCARTA_CTL4_PATHWAY	6/19	0.2200	0.000
HGNC	HGNC:CD molecules	27/392	0.2180	0.000
C5	C5:GO_LYMPHOCYTE_COSTIMULATION	12/78	0.2180	0.000
GOBP	GOBP:positive regulation of T cell activation (GO_0050870)	11/67	0.2150	0.000
C5	C5:GO_REGULATION_OF_LEUKOCYTE_PROLIFERATION	19/202	0.2140	0.000
C5	C5:GO_REGULATION_OF_T_CELL_PROLIFERATION	16/144	0.2140	0.000
MSIGDB	MSIGDB:CELL_DIFFERENTIATION_MARKERS	25/357	0.2120	0.000
C5	C5:GO_REGULATION_OF_INTERLEUKIN_10_PRODUCTION	9/46	0.2120	0.000
GOBP	GOBP:negative regulation of lymphocyte proliferation (GO_0050672)	7/28	0.2120	0.000
C5	C5:GO_NEGATIVE_REGULATION_OF_HOMOTYPIC_CELL_CELL_ADHESION	13/98	0.2100	0.000
GOBP	GOBP:positive regulation of T cell proliferation (GO_0042102)	10/60	0.2070	0.000
C5	C5:GO_POSITIVE_REGULATION_OF_INTERLEUKIN_4_PRODUCTION	6/22	0.2050	0.000
PATHWAY	PATHWAY:Cell adhesion molecules (CAMs)_Homo sapiens_hsa04514	15/138	0.2040	0.000
C2	C2:REACTOME_COSTIMULATION_BY_THE_CD28_FAMILY	10/62	0.2030	0.000
C5	C5:GO_REGULATION_OF_ACTIVATED_T_CELL_PROLIFERATION	8/40	0.2030	0.000
C5	C5:GO_POSITIVE_REGULATION_OF_CELL_ACTIVATION	22/304	0.2020	0.000
C5	C5:GO_REGULATION_OF_CELL_CELL_ADHESION	24/367	0.2010	0.000
GOBP	GOBP:regulation of interleukin-10 production (GO_0032653)	6/23	0.2000	0.000
PATHWAY	PATHWAY:TNFs bind their physiological receptors_Homo sapiens_R-HSA-56690	6/23	0.2000	0.000
C5	C5:GO_REGULATION_OF_CELL_ACTIVATION	27/470	0.1990	0.000
C5	C5:GO_POSITIVE_REGULATION_OF_CELL_CELL_ADHESION	19/237	0.1980	0.000
COMPARTMENTS	COMPARTMENTS:Tumor_necrosis_factor_receptor_superfamily_complex	15/159	0.1900	0.000
PATHWAY	PATHWAY:Costimulation by the CD28 family_Homo sapiens_R-HSA-388841	10/71	0.1900	0.000
PFAM	PFAM:Ig_V-set	14/141	0.1890	0.000
C5	C5:GO_REGULATION_OF_INTERLEUKIN_2 BIOSYNTHETIC_PROCESS	5/18	0.1890	0.000
C5	C5:GO_REGULATION_OF_INTERLEUKIN_2 PRODUCTION	8/47	0.1870	0.000
C5	C5:GO_POSITIVE_REGULATION_OF_ALPHA_BETA_T_CELL_PROLIFERATION	5/19	0.1840	0.000
PATHWAY	PATHWAY:The Co-Stimulatory Signal During T-cell Activation_Homo sapiens	5/19	0.1840	0.000
C2	C2:KEGG_CELL_ADHESION_MOLECULES_CAMS	13/130	0.1830	0.000
C5	C5:GO_NEGATIVE_REGULATION_OF_CELL_CELL_ADHESION	13/133	0.1810	0.000
C5	C5:GO_REGULATION_OF_INTERLEUKIN_12 PRODUCTION	8/51	0.1790	0.000
GOBP	GOBP:positive regulation of interleukin-10 production (GO_0032733)	5/20	0.1790	0.000

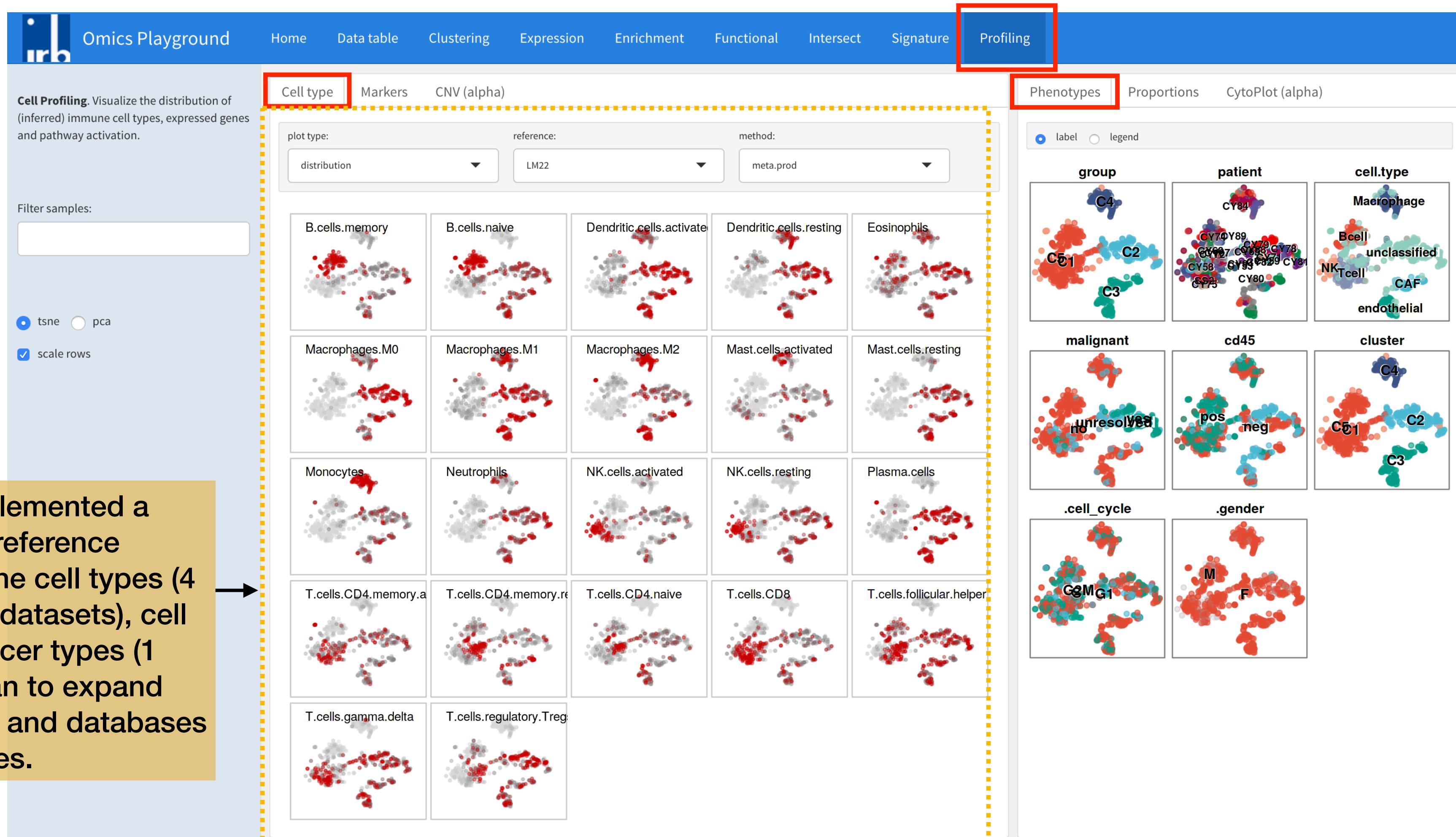
contrast	rho	NES	q
C2_vs_all	-0.03780	0.001800	
C1_vs_all	0.02800	0.01540	
C3_vs_all	-0.01870	0.07870	
C5_vs_all	-0.007000	0.2665	
C4_vs_all	-0.009500	0.2479	

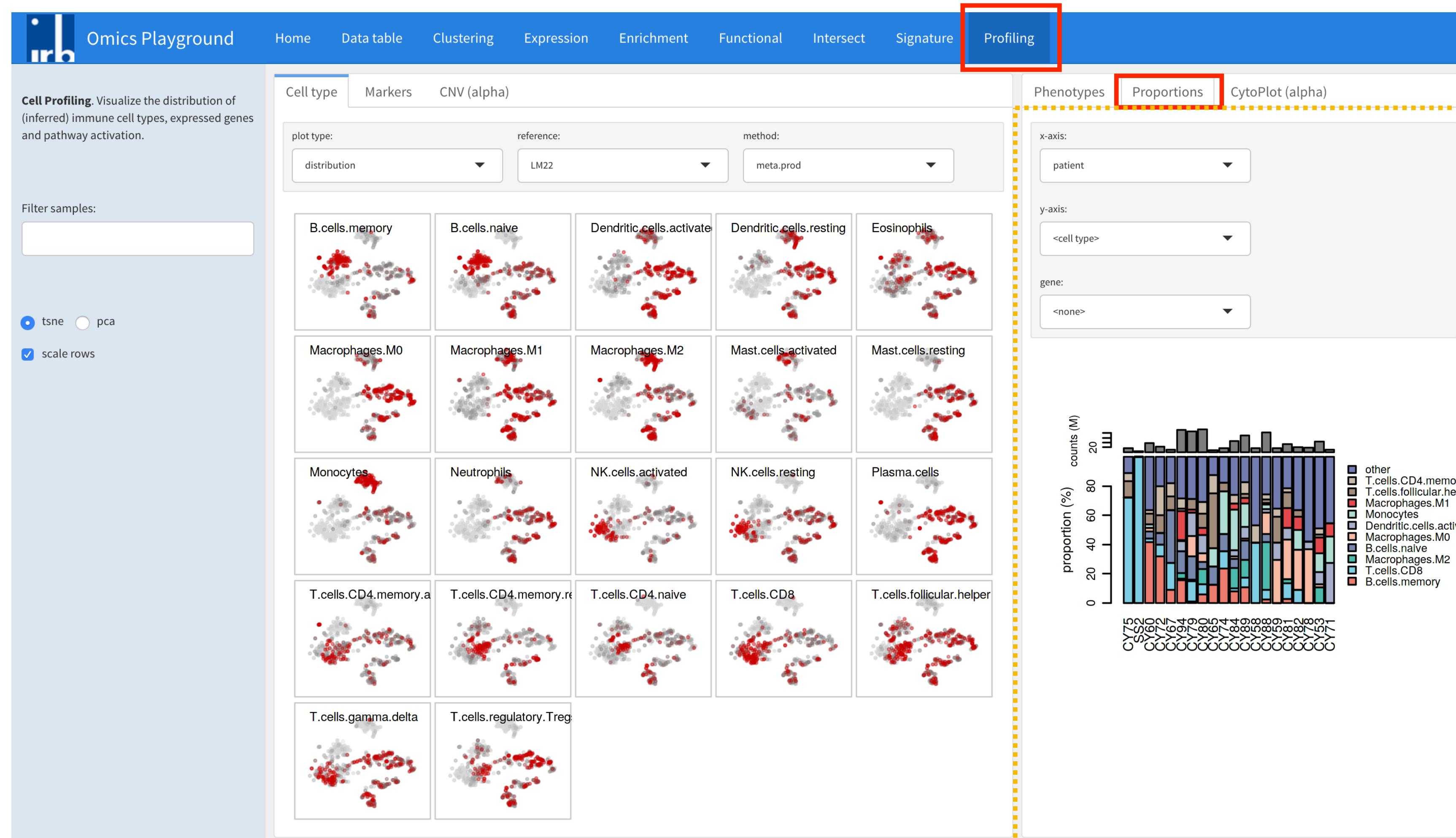
genes	title	FC	q
HAVCR2	hepatitis A virus cellular receptor 2	-2.691	1.850e-12
LGALS9	galectin 9	-2.689	8.134e-15
CD27	CD27 molecule	-1.777	0.00007144
CD276	CD276 molecule	1.613	0.000001855
CEACAM1	carcinoembryonic antigen related cell ad	1.498	0.000003975
TIGIT	T cell immunoreceptor with Ig and ITIM d	-1.457	0.00001042
CD40	CD40 molecule	-1.348	0.001805
CD86	CD86 molecule	-1.289	0.00004689
IDO1	indoleamine 2,3-dioxygenase 1	-1.271	0.03086
BTLA	B and T lymphocyte associated	-1.074	0.0001956
PDCD1	programmed cell death 1	-0.9890	0.1314
CD28	CD28 molecule	-0.8830	0.01250
TNFRSF4	TNF receptor superfamily member 4	-0.8760	3.496e-7
CD47	CD47 molecule	-0.7310	0.3000
CTLA4	cytotoxic T-lymphocyte associated protei	-0.7310	0.001525

# Cell profiling module: cell type tab

A) The cell profiling module infers the types of cells using the prediction methods and reference datasets from the literature.

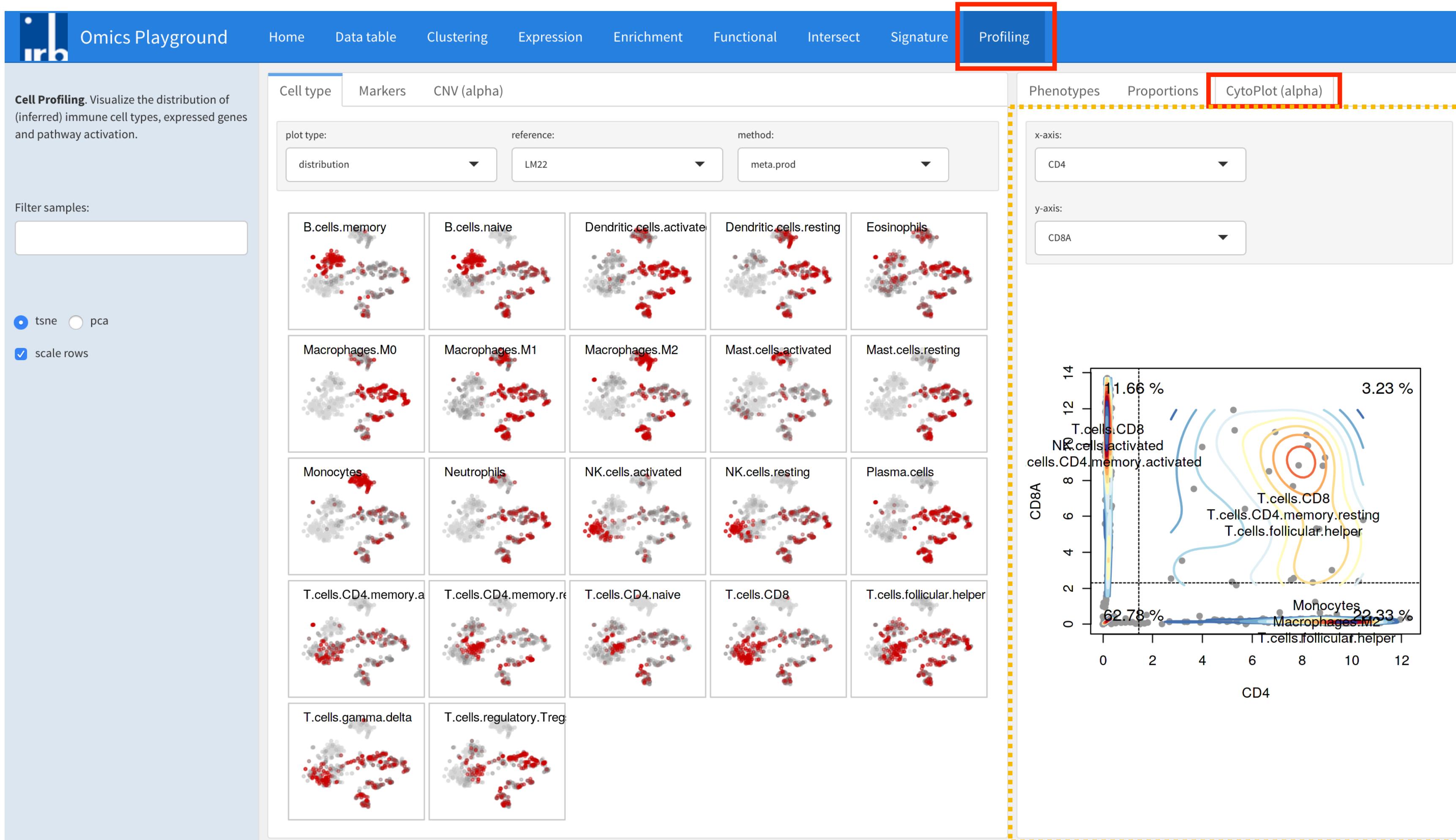


# Cell profiling module: proportions tab



C) Although this feature is very suitable for a single-cell sequencing data, it provides useful information about the proportion of different cell types in samples obtained by the bulk sequencing method.

# Cell profiling module: cytometry-like (CytoPlot) tab



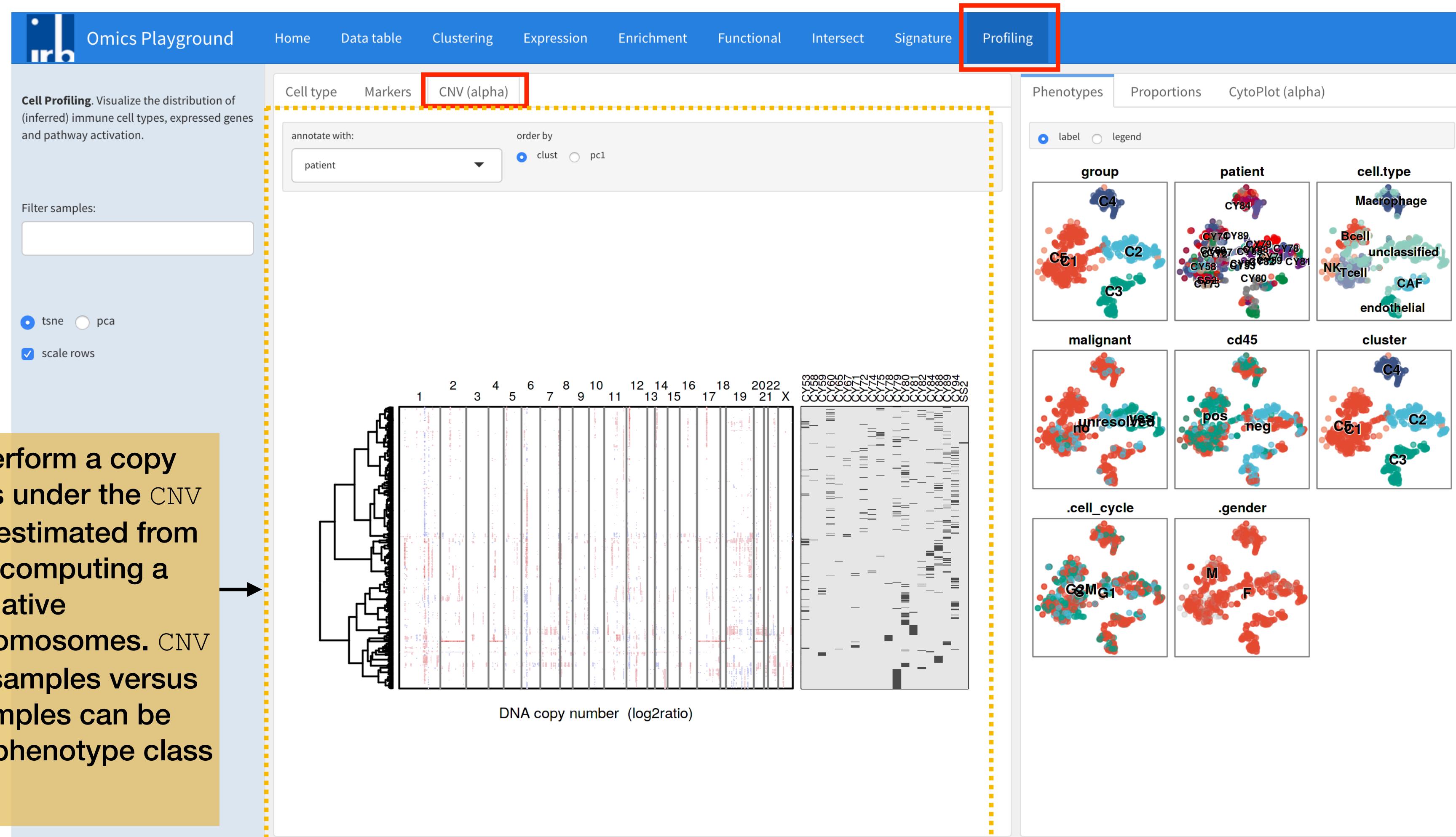
D) For each combination of gene pairs, the platform can generate a cytometry-like plot of samples under the cytoplot tab. The aim of this feature is to observe the distribution of samples in relation to the selected gene pairs. For instance, when applied to single-cell sequencing data from immunological cells, it can mimic flow cytometry analysis and distinguish T helper cells from the other T cells by selecting the CD4 and CD8 gene combination.

# Cell profiling module: markers tab

**E) The markers section provides potential marker genes, which are the top N=36 genes with the highest standard deviation within the expression data across the samples. For every gene, it produces a t-SNE plot of samples, with samples colored in red when the gene is overexpressed in corresponding samples. Users can also restrict the marker analysis by selecting a particular functional group in which genes are divided into 89 groups, such as chemokines, transcription factors, genes involved in immune checkpoint inhibition, and so on.**

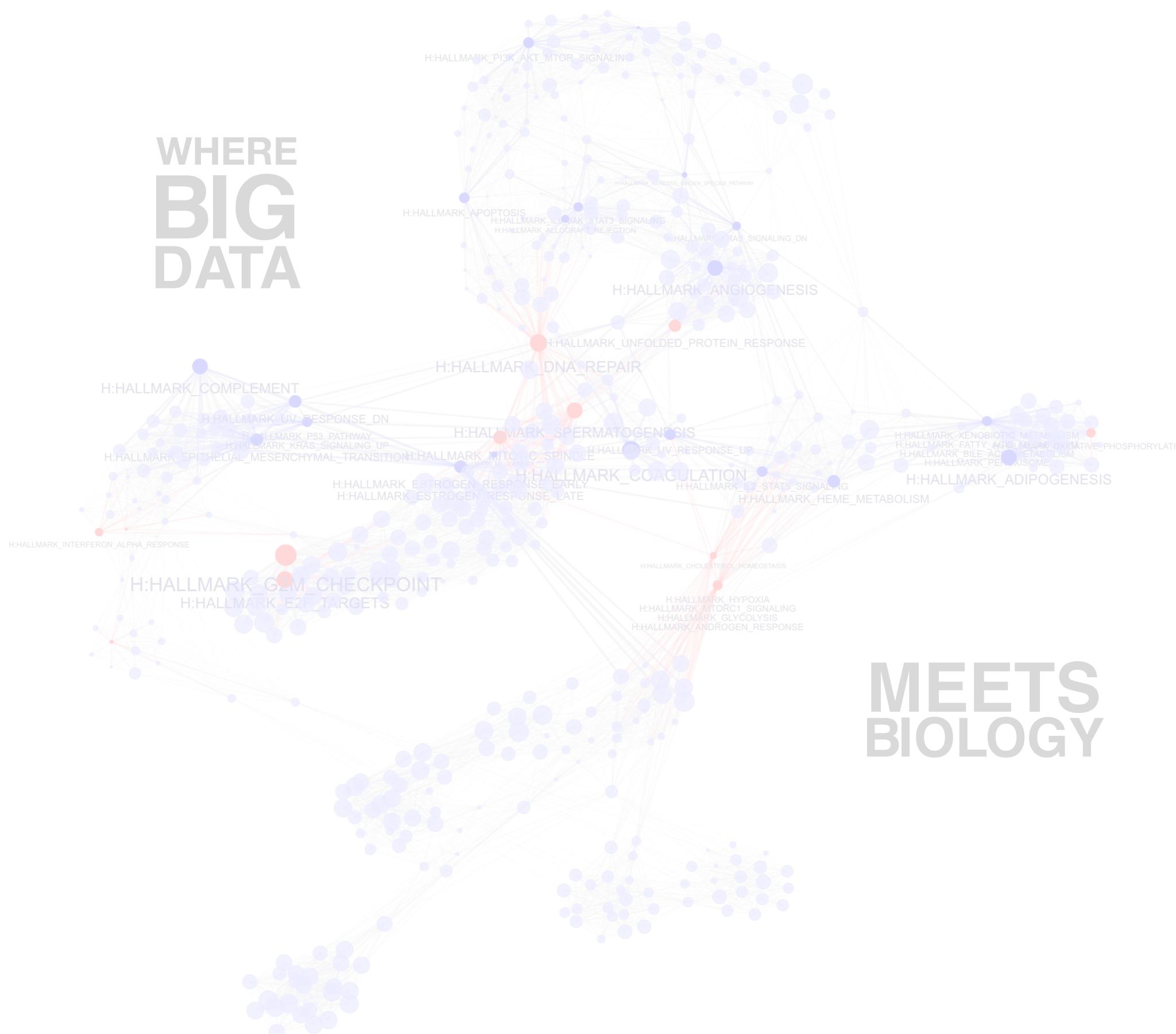
The screenshot shows the Omics Playground interface with the 'Profiling' tab selected. The 'Markers' tab is highlighted with a red box. The main area displays a grid of t-SNE plots for various genes, with a yellow dashed border around the grid. Each plot shows sample distribution with red dots indicating overexpression. To the right, there are other phenotypic and clustering plots. A legend at the top right indicates 'label' (blue dot) and 'legend' (white circle). The phenotypic plots include 'group' (C4, C5, C2, C3), 'patient' (CY84, CY74, CY89, CY97, CY98, CY78, CY58, CY93, CY99, CY101, CY75, CY80), 'cell.type' (Macrophage, Bcell, unclassified, NK\_Tcell, CAF, endothelial), 'malignant' (Unresolved), 'cd45' (pos, neg), 'cluster' (C4, C5, C2, C3), '.cell\_cycle' (G0, G1, S, M), and '.gender' (M, F).

# Cell profiling module: copy number variations (CNV) tab



# The end. Thank you!

Omics Playground (0.99)



BigOmics Analytics Sagl, A spin-off of USI/IOR  
Via Vincenzo Vela 6 - CH6500 Bellinzona CHE-299.736.565  
Tel. +41 820 0368 - [www.bigomics.ch](http://www.bigomics.ch)



Institute of Oncology Research  
Institute of Oncology Research  
Via Vincenzo Vela 5 - CH6500 Bellinzona  
Tel. +41 91 820 0322 - [www.ior.iosi.ch](http://www.ior.iosi.ch)



Institute for Research in Biomedicine  
Institute for Research in Biomedicine  
Via Vincenzo Vela 6 - CH6500 Bellinzona  
Tel. +41 91 820 0300 - [www.irb.usi.ch](http://www.irb.usi.ch)