Cross-talk quantification in molecular networks with application to pathway-pathway and cell-cell interactions.

Authors list:
Chiodi A, institute of biomedical technologies, national research council (CNR), 20054 Segrate (Milan), Italy
Pelucchi P, institute of biomedical technologies, national research council (CNR), 20054 Segrate (Milan), Italy
Mosca E, institute of biomedical technologies, national research council (CNR), 20054 Segrate (Milan), Italy

Motivations: Disease phenotypes are interpreted as the consequence of interactions among molecular processes affected by a series of molecular alterations. Current omics approaches underline the importance of such interactions, focusing on molecular mechanisms, by using manually curated pathways or gene sets. As there are no clear boundaries between gene sets, cross-talks between them are relevant to investigate disease phenotypes.
Here we present Ulisse, a method to (1) quantify cross-talks between gene sets, with application to pathways and intercellular cross-talks; (2) investigate the role of the genes involved in cross-talks, via functional relevance analysis, in terms of regulated processes/cell types. As a proof-of-concept, we studied pathway and cell-cell cross-talks in normal and tumoral breast samples, and compared the obtained results with other available tools: PathNet for pathway cross-talk, ICELLNET and SingleCellSignalR for intercellular cross-talk.

Results: Cross-talk analysis allowed to identify dysregulated interactions between tumor progression key mechanisms, while intercellular cross-talks describe well-known interactions of the tumor with the microenvironment. Through functional relevance analysis, Ulisse was able to pinpoint well known key players of the tumor, as well as potential therapeutic targets. The comparison of Ulisse to other packages highlighted the good performance of our package. In conclusion, Ulisse proved to be a valuable tool to study cross-talks, and provide a newly implemented approach to analyse involved genes.

Availability and implementation: Ulisse is an R package and available at https://github.com/emosca-cnr/Ulisse

Contact: ettore.mosca@itb.cnr.it

Supplementary information: Link to supplementary tables + reference to on-line supplementary material

1. INTRODUCTION

The understanding of how gene-level alterations translate into disease phenotypes is a major challenge in life sciences. The experience of reductionist approaches like genome-wide association studies – where millions of single nucleotide variations are independently tested for association with a phenotype – underlined the key role of molecular *interactions* (Boyle, Li and Pritchard 2017). Nowadays, network medicine approaches interpret disease phenotypes as the consequence of various interacting molecular processes that are affected by a series of molecular alterations (Barabási, Gulbahce and Loscalzo 2011). However, the experimental detection of interactions is hard to perform and, despite relevant progresses (Porras *et al.* 2020), there is not a universally accepted "reference" model of the interactome (Huang *et al.* 2018; Mosca *et al.* 2021).

The largest amount of experiments are based on technologies that provide data about "building blocks" (e.g., genes, proteins) rather than interactions between them. Typically, omics assays results in long gene lists whose interpretation is a major challenge, and pathway enrichment analysis is used to gain insights from those lists. However, pathway definitions result from the analytical approach of dividing a complex network of biochemical processes into "compartments" that are associated with particular functional meanings. Indeed, there are no boundaries between pathways and their interactions – often referred to as cross-talks – have been shown to be of interest to dissect relevant regulatory mechanisms (Mosca *et al.* 2012a, 2012b) identify effective drug combinations in cancer (Jaeger *et al.* 2017), investigate complex diseases phenotype (Hu *et al.* 2017).

Molecular interactions go beyond cell boundaries. The study of cell-cell cross-talk (or communication) has been recently raised thanks to the advent of single cell RNA sequencing technologies (scRNA-seq), which allows to obtain expression data of different cell-types in a tissue, even with spatial data. This information allows to infer the interactions between different cells, like the tumoral cells with their microenvironment. These studies are far from being resolutive, as these represent inference that need to be experimentally validated, but provide testable hypothesis of cellular behaviour in different conditions (Armingol *et al.* 2021).

Here, we describe Ulisse, a method to screen (weighted) gene lists for the presence of cross-talks between a collection of gene sets **(Fig. 1)**. The resulting cross-talks are analysed to identify key genes in terms of the variety of (i) different gene sets they are in cross-talk with and (ii) interacting partners involved in the cross-talks. As a proof-of-concept, we applied

Ulisse to identify the cross-talks between pathways affected by molecular alterations in breast cancer (BC) subtypes, and cell-cell communication in breast cancer. We also compared the applicability of our tool to other already available in literature: PathNet for pathway cross-talks (Dutta, Wallqvist and Reifman 2012); SingleCellSignalR and ICELLNET for cell-cell communication (Noël *et al.* 2020; Cabello-Aguilar *et al.* 2021).

2. Methods

2.1. A general score for cross-talk

We introduce a general score to quantify the cross-talk between gene sets. This score can be used to study both pathway cross-talk and intercellular cross-talk, given that both the state of a pathway and that of a group of cells can be represented as a gene set associated with weights, and their cross-talk can be quantified by means of gene-gene interactions.

Given a set of genes of interest $L = \{g\}$, a set $\Gamma = \{G_1\}$ of gene sets G to be screened for cross-talk, a gene network N with adjacency matrix $\mathbf{A} = \{a\}$ and a $|L| \times |\Gamma|$ matrix of gene weights $\mathbf{W} = \{w\}$ (**Fig. 1A**), the cross-talk C between any gene set pair $\{G_x, G_y\} \subseteq \Gamma$ is defined as the sum of the weighted products between the interacting genes $\{g_i, g_j\} \in L$ (**Fig. 1B**):

$$C(G_x, G_y) = \sum_{g_i \in G'_x} \sum_{g_j \in G'_y} a_{ij} w_{ix} w_{jy}$$

where we distinguish two definitions for the sets G'_x and G'_y:

- $G'_x = G_x \backslash G_y$ and $G'_y = G_y \backslash G_x$, to exclude genes belonging to both gene sets in the cross-talk, as in the case of pathway cross-talks, to avoid counting intra-pathway interactions between g_i and g_j;
- $G'_x = G_x$ and $G'_y = G_y$, to include genes belonging to both gene sets, as in the case of cell-cell cross-talks, to count the intercellular interaction between g_i and g_j, given that they belong to two cell clusters.

Note that **A** is typically sparse and therefore only some genes contribute to the cross-talk. Gene weights can be different for every gene set, which is important to model the different statistics that a gene can have in different gene sets, as in cell cluster markers; when gene weights are not available, the elements of **W** are set to 1.

The gene sets that establish cross-talks can be organized in a cross-talk network $T = (V, E)$, where vertices are gene sets and edges connect gene sets with the necessary and sufficient condition $C(G_x, G_y) > 0$ (**Fig. 1C**).

To provide a statistical assessment of the observed value of $C(G_x, G_y)$, we consider an empirical null model that focuses on the interactions between G_x and G_y (**Fig. 1D**). Precisely, we generate a number k of random gene set pairs $\{G_x^k, G_y^k\}$ matched in size with $\{G_x, G_y\}$ and calculate the corresponding $C^k(G_x^k, G_y^k)$ with **W** consisting of all 1s. These values are used to obtain the empirical *p*-value that estimates the probability of observing the same or a higher number of interactions between G_x and G_y:

$$p(G_x, G_y) = \frac{\#\{C^k(G_x^k, G_y^k) \geq C(G_x, G_y)\} + 1}{n_k + 1}$$

The empirical False Discovery Rate (FDR) associated with an observed value of $C_{z=1}$, in the context of $z = \{1, 2, \dots, n_z\}$ observed values, is estimated as the ratio between the fraction of all permuted scores C_z^k, *k* scores for each *z,* that are equal or greater than $C_{z=1}$, and the fraction of the observed scores C_z that are equal or greater than $C_{z=1}$ (Subramanian *et al.* 2005):

$$q_{z=1} = \frac{\#\{C_z^k \geq C_{z=1}\}}{n_k n_z} \cdot \left(\frac{\#\{C_z \geq C_{z=1}\}}{n_z}\right)^{-1}.$$

2.2. Gene classification based on cross-talks

Once a number of cross-talks has been detected, it is interesting to "go back" to genes and identify the key players of the observed interactions. To this aim, we introduce *functional diversity* and *interactor diversity*.

The *functional diversity* of a gene $g_i \in G_x$ captures the number $n_f(i)$ of different gene sets G_y that (i) are in cross-talk with any gene set G_x and (ii) contain at least a gene g_j that interacts with g_i:

$$n_f(i) = \#\{G_y \mid \exists\, g_j \in G_y' \wedge a_{ij} = 1 \wedge C(G_x, G_y) > 0\}$$

where $G_y' = G_y \backslash G_x$. In other words, $n_f(i)$ counts the number of gene sets that are *affected* by g_i. If Γ is part of a broader set of gene sets, $\Gamma \subset U$, e.g. {pathways that contain at least a gene in L} $\subset$ {all pathways}, then it is possible to define the *relative functional diversity* as a means to normalize $n_f(i)$ by the number of all gene sets $G_u \in U$ that can establish a cross-talk with any gene set G_x by means of interactions between g_i and $g_l \in \Lambda \supseteq L$ (**Fig. 1E**):

$$n'_f(i) = \frac{n_f(i)}{\#\{G_u \mid \exists g_l \in G'_u \wedge a_{il} = 1 \wedge C(G_x, G_u) > 0\}}$$

where $G'_u = G_u \backslash G_x$. Intuitively, $n'_f(i)$ counts the fraction of gene sets that are affected by g_i in relation to all gene sets that can be affected by g_i.

The *interactor diversity* of g_i quantifies the number $n_d(i)$ of its interactors that contribute to at least a cross-talk interacting with g_i:

$$n_d(i) = \#\{g_j \mid \exists g_j \in G'_y \wedge a_{ij} = 1 \wedge C(G_x, G_y) > 0\}.$$

As with *functional diversity*, we consider the *relative interactor diversity* **(Fig 1E)** to normalize $n_d(i)$ by the number of interactors of g_i in all possible cross-talks in U **(Fig. 1E)**:

$$n'_d(i) = \frac{n_d(i)}{\#\{g_l \mid \exists\, g_l \in G'_u \wedge a_{il}=1 \wedge C(G_x,G_u)>0\}}.$$

We remark that the calculation of functional diversity and interactor diversity does not require weights.

Lastly, we define the *functional relevance* f_i of the interactors of g_i as the log-ratio between functional diversity and interactor diversity:

$$f_i = \log_2\left(\frac{n_f(i)}{n_d(i)}\right)$$

and the relative *functional relevance* $f'_i = \log_2(n'_f(i)/n'_d(i)$.

Taken together, the above quantities suggest the following gene classification (**Fig. 1F, G**):

- *kinless* genes: genes that affect many gene sets (high $n_f(i)$) by means of many interactors, high $n_d(i)$;
- *provincial* genes: genes that affect a few gene sets, low $n_f(i)$, by means of a few interactors, low $n_d(i)$;
- *genes with kinless interactors*: genes that affect many gene sets, high $n_f(i)$, by means of a few interactors, low $n_d(i)$; therefore high f_i;
- *genes with provincial interactors*: genes that affect a few gene sets, low $n_f(i)$, by means of many interactors, high $n_d(i)$; therefore low f_i.

The same classification can be obtained using relative quantities.

2.3. Implementation

Ulisse is implemented as an R package and is available at the URL https://github.com/emosca-cnr/Ulisse. Cross-talks are quantified through the function:

```
ct <- gs_cross_talk(gs_list = G, gene_network_adj = A)
```

which requires in input a list of gene sets (*G*, with optional weights) and an adjacency matrix (*A*) that represents a gene-gene interaction network. The documentation of Ulisse describes how these input data can be obtained using public repositories of gene-pathway associations, gene-gene interactions and ligand-receptors interactions, and the various parameters that can be tuned **(Supplementary material)**. The functional relevance is assessed using the homonymous function:

```
fr <- functional_relevance(ct = ct, adj = A, method =
      "count", ct_null = NULL)
```

which requires the output of gs_cross_talk(), the adjacency matrix and the type of functional relevance to be quantified ("count" for absolute values, "relative" for relative values). See the documentation for further details **(Supplementary material)**.

2.4. Datasets and tools used in case studies

2.4.1. Somatic mutation and gene expression

Breast cancer mutation and expression data were obtained by means of TCGAbiolinks (v2.25.2) (Colaprico *et al.* 2016) from TCGA, and pre-processed as described in the package tutorials (**Supplementary Material**). A gene-level mutation score was defined for All BCs and for each considered subtype (Basal, Her2, Luminal A and B), as the fraction of patients in which the gene carried at least a mutation. Analogously, a gene-level differential expression score was defined as $-\log10(p) * |\log2(f)|$, where f is the fold change between tumors and matched controls, and p is the differential expression p-value ("glmLRT" model through limma package (Ritchie *et al.* 2015)). To mitigate the presence of outliers, the gene mutation score was normalized over the mean of the top 10 highest scores and the resulting values higher than 1 were set to 1 (**Supplementary Figure 1**). The gene differential expression score was normalized over its maximum (**Supplementary Figure 1**). The top 300 genes of each list with the highest scores were considered as "altered" genes for cross-talk analysis and their scores were used as weights. After normalization, the two gene lists (mutated and differentially

expressed genes) were combined into a gene list ranked by gene score. The mean of the two scores (mutation and expression) was used if a gene appeared two times between the top 300 (**Supplementary Material, Supplementary Table 1-5**). Gene-gene interactions were collected from STRING v11.5; only high-confidence (STRING score > 700) interactions were considered, excluding "text mining" from the list of evidences. mSigDB "Hallmarks" gene sets (Liberzon *et al.* 2015), were collected through the package "msigdbr" v7.4.1 (Dolgalev 2022). Empirical nulls used to estimate $p(G_x, G_y)$ were defined using 99 permutations. Functional relevance and relative functional relevance analysis were performed by using only significant cross-talks (p-value ≤ 0.01 and FDR ≤ 0.2); relative functional relevance was calculated by using a general model build on whole Hallmark database (**Supplementary Material**).

2.4.2. Single cell RNA sequencing

Two breast single cell datasets were analysed: the first is composed by four normal breast organoid samples (N4, N5, N6, N7) from GSE113197 (Nguyen *et al.* 2018); the second is composed by three tumoral samples (CID4290A, CID44991 and CID4515) from GSE176078 (Wu *et al.* 2021), selected by considering number of cells and cell-types. All the samples were downloaded from GEO and analysed with standard Seurat v4 workflow to obtain cell clusters (**Supplementary Material**) (Hao *et al.* 2021). To analyse intercellular cross-talk, expression or differential expression data were used, together with a Ligand-Receptor (LR) network. For each sample, normalized data were used to define an expression threshold to identify expressed genes. Expression scores for the genes was thus calculated as the frequency of expression in the cells of a cluster (**Supplementary Material**). Differentially expressed (DE) genes were identified using Seurat default function ("FindAllMarkers()") and defined as those genes with log2(*f*) ≥ 0.5 and FDR ≤ 0.05 (Benjamini-Hochberg method (Benjamini and Hochberg 1995)). DE scores were calculated as the -log10(*FDR*) * log2(*f*) (**Supplementary Material**).

LR communication network was obtained from omnipathR package (v3.0.4) as described in Supplementary Material (Türei, Korcsmáros and Saez-Rodriguez 2016; Valdeolivas, Turei and Gabor 2019). Ulisse intercellular cross-talk analyses were performed using both expression and DE lists in all samples with 99 permutations.

The intercellular cross-talks of sample CID4515 were studied using the cell-types identified by the authors (Wu *et al.* 2021) on expression data (**Supplementary Material**).

Significant cross-talk (FDR ≤ 0.1) were further evaluated to study genes involved through Functional relevance analysis (**Supplementary Material**).

2.4.3. Cross-talk tools

All BC top genes obtained from mutational and differential expression analysis were analysed with PahtNet (v1.32), by using Hallmarks database together with STRING network.

SingleCellSignalR (v1.4) calculates a score for each gene-gene interaction considering only paracrine or autocrine signals. It also integrates DE data to identify specific gene interactions. For each sample, SingleCellSignalR was used to obtain scores for both paracrine and autocrine signals considering DE data. SingleCellSignalR ligand-receptor scores for a given cluster-cluster pair were summed to obtain a unique representative score for each pair of clusters.

ICELLNET (v1.0.1) returns a score for each directed link between cluster pairs. The package standard workflow was applied, which involves cleaning and scaling of the data by using ICELLNET functions. As the package returns directed score, it calculates two scores for each cluster pair. Were summed these to obtain a unique undirected score that can be compared to Ulisse's communication score.

Further details about PathNet, SingleCellSinalR and ICELLNET data analyses are listed in Supplementary Material.

2.4.4. Robustness against dataset subsampling

The sensitivity of a tool to the input data was assessed generating subsamples composed of the 70% of the cells of the corresponding samples, and then quantifying the performance of each tool in recovering – analysing subsampled datasets – the cross-talks obtained using original samples. The performance was estimated by means of the partial area under the ROC curve (pAUC):

$$\mathrm{pAUC}_N = \frac{1}{N \cdot P} \sum_{i=1}^{N} P_i$$

where P is the total number of positives (cluster-cluster pairs found in the sample), P_i is the number of positives (cluster-cluster pairs found in both sample and subsample) that score higher than the i-th highest scoring negative (cluster-cluster pair found only in the subsample) and N is the number of top high scoring negatives considered for the calculation of pAUC_N. Here, we tested robustness for P = 20 and i = {1, 3, 5, 10}.

3. Results

3.1. Cross-talks between MSigDB Hallmark gene sets

We used Ulisse to screen the existing cross-talks between biological pathways. To this aim, we considered the "Hallmark" gene sets from MSigDB, which "summarize and represent specific well-defined biological states or processes" (Liberzon *et al.* 2015), and high confidence gene-gene interactions from STRING (**Supplementary Material**).

We obtained an almost fully connected cross-talk network, with a total of 1233 cross-talks between the 50 Hallmarks **(Figure 2A)**. Cross-talk strength varies between 3 orders of magnitude (from 10^1 to 10^3), with the strongest cross-talks occurring between "allograft rejection" (genes up-regulated in the immune response to nonself antigens) and "interferon alpha (gamma) response" (genes involved in immune and autoimmune responses), while the weakest are the interactions of "MYC targets" with "angiogenesis", "cholesterol homeostasis" and "coagulation" pathways (**Figure 2A**). Functional relevance analysis resulted in 3472 genes involved in 1-50 cross-talks with 1-226 interactors, with functional relevance scores ranging from -3.28 to 5.61 **(Figure 2B, Supplementary Table 7)**. Most of these are genes with provincial genes (both low n_f and n_d, i.e.: *RBM14*, *DMAC2L*), genes with kinless interactors (low n_d and high n_f, i.e.: *C5AR1*, *ALCAM*, *ADD2, C4BPB* and *SHROOM2*) and kinless genes (both high n_f and n_d, i.e.: *SRC*, *UBC*, *MAPK1*, *CDK1*), with a complete lack of genes with provincial interactors (high n_d and low n_f).

3.2. Cross talks between pathways hit by molecular alterations in breast cancer (BC)

We applied Ulisse to quantify the altered cross-talks between biological pathways in All BCs and by BC subtype (**Tab. 1**), based on the genes associated with the most significant molecular alterations (mutations and matched tumor-normal differential expression), "Hallmark" MSigDB pathway definitions and high confidence gene-gene interactions from STRING.

3.2.1. Cross-talks in BC

We found the highest number of significantly altered cross-talks in Basal BCs and, on the opposite, only a few in HER2 subtype (**Fig. 3A, Tab. 1**). Interestingly, this 11-fold difference (75 / 7) between the two subtypes arises from much lower differences in terms of altered genes, affected pathways and average number of altered genes per pathway (**Tab. 1**). This result suggests a particular role of cross-talks in Basal BCs, where about 30% of altered genes dysregulate cross-talks (**Tab. 1**).

In every subtype we found a relevant number (from 19% to 65%) of specific altered cross-talks and 1 cross-talk in common (**Fig. 3A**), between "adipogenesis" and "epithelial-mesenchymal transition" (**Fig. 3B-F**). Basal and Luminal subtypes share 4 altered cross-talks, involving "estrogen response (late)", "glycolysis" and "PI3K-AKT-mTOR", which are well known key players of BC tumorigenesis (Cavazzoni *et al.* 2012; Nahta 2012; O'Regan and Paplomata 2013; Paplomata and O'regan 2014) (**Fig. 3B-F**). As expected by their clinical characteristics, the two Luminal subtypes share a higher number of altered cross-talks with each other (39.5% of Luminal A, 48.6% of Luminal B) than other subtypes (**Fig. 3B**).

We found that each subtype shares a relevant number (from 24% to 29%) of altered cross-talks with those emerged from the analysis of All BCs, but, unexpectedly, none of these is in common between all subtypes and All BCs, even excluding HER2 subtype that has only a few altered cross-talks (**Fig. 3A, Tab. 1**).

3.2.2. Functional relevance in BC

We quantified the functional relevance and relative functional relevance of genes that mediate the significant cross-talks (mediator genes hereafter) found in All BCs and in each subtype **(Fig. 4, Supplementary Tables 8-12)**. As expected, the number of mediator genes in each subtype **(Fig. 4A)** is proportional to the number of cross-talks **(Fig. 3A)**, with Luminal A and B sharing the highest number of mediators (53.9% and 47% of the genes, respectively, **Fig. 4B**). The 4 subtypes share 13 mediators, which are the 10.5% of all mediators observed in Basal, 27.6% in Her2, 20.6% in Luminal A and 18% in Luminal B **(Fig 4B)**, and all of them mediate significantly altered cross-talks in All BCs.

Most of the genes lay around the diagonals of the functional relevance plots, with high prevalence of provincial genes and some kinless genes **(Fig. 4C-G)**. Luminal A and B show some genes with kinless interactors, and only Basal shows genes with provincial interactors. Among kinless genes of All and subtypes BCs, we found a number of well-known key players

of BC: *TP53* is central in P53 pathway, whose alteration is involved in tumor progression and invasion, as under stress conditions (Vogelstein, Lane and Levine 2000) it induces cell-cycle inhibition (Chan *et al.* 1999), apoptosis (Oda *et al.* 2000a, 2000b) and prevents blood vessels formation (Hendrix 2000); *PIK3CA* and *AKT1* genes are part of the PI3K-AKT-mTOR pathway which is involved in cell proliferation (Paplomata and O'regan 2014); *CREBBP* is involved in cell growth, differentiation, apoptosis and transformation (Goodman and Smolik 2000), exerts a critical role in the biology of cancer stem cells (DiMeo *et al.* 2009) and represents a potential therapeutic target in BC subtypes (Ring *et al.* 2018; Waddell *et al.* 2021). Among the genes with provincial interactors we found *KIF20A* (Basal subtype, **Fig. 4D**), which is involved in intracellular and chromosome transport during mitosis. Overexpression of this gene is correlated to poor overall survival, while suppression results in inhibition of cancer cells (Nakamura *et al.* 2020; Jin *et al.* 2023).

In relative functional relevance analysis (**Fig 4H-L**), the distributions of the mediator genes in the plots are skewed toward low values of both n_d' and n_f', except for the Basal subtype. Among All BCs kinless genes we identify genes involved in tumor progression and poor prognosis: *STAG1* dysfunction could trigger tumorigenesis (Yun *et al.* 2016), as is involved in the organization of chromosomal DNA topology and modulation of expression of oestrogen-responsive genes associated with breast cancer progression (Antony *et al.* 2015); *FLNC* is part of a family that mediates cytoskeleton remodelling and whose members are associated with poor prognosis in various cancer (Kamil *et al.* 2019); *G0S2* regulates several biological processes and is silenced in many tumors (Heckmann *et al.* 2013), therefore identified as possible tumor suppressor in BC by blocking *MYC* activity (Yim *et al.* 2016), antagonizing PI3K-mTOR signaling and increasing therapy effectiveness (Yim *et al.* 2017); *MMP11* is a gene involved in degradation of the external matrix and cancer progression macrophages, whose overexpression enhance migration of *HER2* positive cancer cells (Kang *et al.* 2022). Basal subtype shows a high number of kinless genes and one gene with provincial interactor, *CASQ2*, which has been recently found to be involved in modulation of cancer-microenvironment interactions. Overexpression of this gene also results in more aggressive phenotype (Kim *et al.* 2022). Luminal B genes mostly behave as provincial, except for a kinless gene (*KRT17*) and a gene halfway between kinless and genes with kinless interactors (*PTX3*). *KRT17* is associated with various tumors, and it was recently demonstrated that reduced expression of this gene in BC significantly correlates with poor prognosis (Tang *et al.* 2022). *PTX3* is involved in the regulation of innate immunity, inflammation, angiogenesis and remodelling of the extracellular

matrix. Overexpression of this gene is observed in many tumors, included BC, but its involvements is still under analysis. PTX3 was identified as a potential target in BC where angiogenesis is involved in the pathogenesis, especially in hormone-dependent tumors (Bogdan *et al.* 2022). Considering relative functional relevance score, among the genes with high score in all the cases we identified most of the abovementioned key players and other relevant genes, like *NF1*, whose mutation leads to a cancer-predisposing syndrome that is associated to higher predisposition to BC and mortality (Suarez-Kelly *et al.* 2019).

3.3. Comparison among pathway cross-talks of Ulisse and PathNet

PathNet is a package to perform pathway analysis using topological information (Dutta, Wallqvist and Reifman 2012). PathNet "contextual analysis" quantifies pathway cross-talks similarly to Ulisse, with a focus on gene expression data. There are relevant differences between the two approaches. To statistically evaluate the cross-talks, Ulisse builds a null model of the expected interactions between two pathways, while PathNet permutes gene weights. Besides, Ulisse provides a detailed output for each cross-talk (e.g., score, p-value, FDR, genes involved) occurring between gene sets that show at least an interaction, while PathNet returns a matrix of p-values between all-pairs of pathways. Lastly, Ulisse uses cross-talk results to suggest relevant genes via functional relevance analysis.

To compare the two approaches, we tested them on the same input and used the same filtering threshold to identify significant cross-talks (p-value $\leq$ 0.05, **Supplementary Material**). PathNet and Ulisse returned 23 and 80 significant cross-talks, respectively, showing low percentage of shared cross-talks (12.6%), despite the percentage of shared pathways is high (53.6%, **Supplementary Figure 3**).

3.4. Cell-cell communication in healthy and tumor breast mammary gland

As stated above, the "mathematical machinery" of Ulisse can also be used to quantify the cross-talk (communication) between cell-types/clusters, using LR interactions and cluster

markers as input gene sets $\{G_x, G_y\}$. Firstly, we assessed whether Ulisse produces results that are comparable with those of SingleCellSignalR and ICELLNET. Then, we illustrate the application of Ulisse to study the intercellular cross-talks between the various cell-types of a BC sample.

3.4.1. Comparison between Ulisse, SingleCellSignalR and ICELLNET

We run the three approaches on four normal breast tissue and three BCs samples. We considered two possible ways of defining cluster markers in Ulisse and SingleCellSignalR ("-expression" and "-DEG") and two possible ways of cluster-cluster cross-talk ranking in Ulisse ("-score" and "-FDR") (**Supplementary Material**).

We calculated Spearman correlation coefficient between the cluster-cluster communication scores of the various tools and implementation, and found a relevant variability, as emerged in other studies (Jin *et al.* 2021; Dimitrov *et al.* 2022). Indeed, in the majority of cases, we observed both positive and negative correlations (i) on the same pair of tools depending on the sample (**Fig. 5A**) and (ii) on the same sample depending on the tool (**Fig. 5B**). Ulisse-Score and ICELLNET showed the highest consistency across samples, obtaining a median correlation coefficient of 0.77 and the lowest variance (0.004, **Fig. 5A**). We assessed the performance (pAUC) against dataset subsampling and found that the various tools performed very closely, with few exceptions (**Fig. 5C**).

3.4.2. Cell-cell communication in breast cancer microenvironment

We used Ulisse to infer the intercellular cross-talks between the 9 different cell-types of the BC sample CID4515 (Wu *et al.* 2021) and found a total of 8 significant links between 6 cell-types (FDR ≤ 0.1) (**Fig. 6, Supplementary Table 13**). Cancer epithelial cells communicate – at various strengths – with most of the cell-types, except B-cells, Normal epithelial and Plasmablasts. Cancer cells have higher cross-talk scores with myeloid, CAFs (Cancer Associated Fibroblasts) and endothelial cells, reflecting tumor-microenvironment interactions: the communication network involves CAFs which promote tumorigenic features supporting cancer proliferation and metastasis, myeloid cells recruitment (with inflammation induction) and angiogenesis mediated by Endothelial cells. The intercellular network emerged from our analysis is coherent with the described main cell-cell communications in the tumor

microenvironment, exerted by secretion of several molecules and vesicles targeting various cell-types, but also mechanical interactions (Elwakeel and Weigert 2021; Mao *et al.* 2021; Deng and Fleming 2022).

The intercellular cross-talks are mediated by 1070 genes, that can be specific of a cell-type (cell-type specific markers) or general (genes that emerged as markers of all cell-types), and show n_f and n_d values ranging from 1 to 6 cell-types and 1 to 110 interactor genes respectively **(Supplementary material, Supplementary Figure 2 and Supplementary Table 14)**. We focused on genes present in Cancer epithelial gene sets (780 genes mediating interactions with 1-6 cell-types and 1-110 genes, **Figure 7A**), identifying Cancer-specific genes, that behave as mediator genes only in cancer cells (**Figure 7B**), and Cancer-shared genes, that behave as mediator genes in all 6 cell-types involved in the significant intercellular cross-talks (**Figure 7C**). Functional diversity (n_f) here represents the cell-types with whom a gene mediates a communication. By suing this value is possible to identify we can identify particular (low n_f values) and diffused (high n_f values) gene-communications.

Among Cancer-specific genes (**Figure 7B**), examples among particular gene-communications we identify genes involved in the communication with specific cell types (n_f = 1), like PTPRG (with CAF) and SCN2A (with Endothelial cells); while diffused gene-communications involve all microenvironment cell-types (n_f = 5; i.e.: *TCN1*, *MST1R*, *ERRB2*, *WNT7B*, *FGRF2*).

In Cancer-shared genes (**Figure 7C**), particular gene-communications allow to identify genes involved in communications with specific cell-types (n_f = 2, e.g., *GNAQ*, *HSP09AA1* and *ECE1* used to communicate with Cancer epithelial and Endothelial cells, *CLEC2B* and *BAG6* with Cancer epithelial and T-cells).

4. DISCUSSION

We presented a general framework to assess the presence of cross-talks between gene sets, with a focus on intra-cellular cross talks (between pathways) and inter-cellular cross-talks (between cell-types), and a statistical assessment of the number of interactions observed. Besides, cross-talk mediating genes are classified based on their "functional relevance", that is, proportional to the (relative) number of interacting genes/gene sets.

The comparison of cross-talk analysis with existing tools revealed relevant differences in methodology as well as implementation. The comparison of cross-talks (inter-cellular and intra-cellular) are intrinsically complex due to the absence of a ground truth, and to this is added the types of output returned by the tools. Ulisse and PathNet use two different empirical nulls, which is probably the main factor that determined different results on the same input. Ulisse focuses on interactions between pairs of gene sets and, in line with this, the empirical null models the expected interactions between the two sets. Therefore, it supports analyses without gene weights. As regards inter-cellular cross-talks Ulisse, SingleCellSignalR, and ICELLNET provide slightly different pieces of information. In this context, Ulisse is focused on assessing cross-talks between cell clusters and, differently from the other two tools, it provides a statistical assessment. SingleCellSignalR is focused on gene-gene communications, while ICELLNET studies directed communications also by considering sub-units. Taken together, our results suggest that Ulisse is more similar to ICELLNET than SingleCellSignalR.

The proof-of-concept on BC provides a picture of dysregulated pathway interactions determined by gene mutations as well as gene expression changes. We considered this multi-omic approach to reach a more complete description of the molecular alterations that might affect the cross-talks. Interestingly, we observed a relevant variation in the number of dysregulated cross-talks between subtypes. Despite this finding deserves further study, by using for example other pathway definitions, it shows how the analysis of pathway cross-talk can shed light on another "layer" of complexity. Further assessment of dysregulated pathway communications may deepen the understanding of subtype-specific mechanisms of tumor progressions. More generally, the analysis of pathway cross-talks complements current pathway analysis approaches in the aim of relating a pathological phenotype to its gene-level determinants.

Functional relevance analysis showed its usefulness discriminating the various modalities in which well-known BC key players behave as regulators of one or more cellular functions. These results can be relevant for the choice of actionable targets. Indeed, genes that affect several cross-talks interacting with multiple cellular functions (kinless genes) are interesting targets for therapy, but – at the same time – could be associated with a wide spectrum of negative side effects. In this context, relative functional relevance provides a means to collect more selective targets for therapy, as it prioritizes genes that mediate cross-talks with less but

more disease-specific cellular functions, as exemplified by G0S2 and STAG1 genes (Yim *et al.* 2017; Antony, Chin and Horsfield 2021).

The inter-cellular cross-talk analysis of the BC dataset confirmed the well-known core network of communications between tumor and its microenvironment, which involves cancer epithelial cells with Cancer Associated Fibroblasts (CAFs), immune cells (in particular myeloid infiltrates) and endothelial cells. This network of communication underlies many tumor promoting functions supporting tumor growth, chronic inflammation and angiogenesis, by secretion of growth factors and other soluble molecules, vesicles and also mechanic interactions among cells and extracellular matrix (Elwakeel and Weigert 2021; Mao *et al.* 2021; Hu *et al.* 2022; Guo *et al.* 2023).

An interesting opportunity provided by Ulisse, which we leave as matter of future studies, is the joint analysis of intra-cellular and inter-cellular cross talks on the same scRNA-seq dataset. Indeed, such analysis could reconstruct a map that integrates the communications between different cell states with the pathway cross-talks activated within each cell state.

In conclusion, our study supports the relevance of cross-talk analysis as a complement to current pathway analysis and cell-cell communication approaches. Besides, the prioritization of cross-talk mediator genes based on functional relevance suggests key players that can be selected as potential actionable targets.

5. Acknowledgements

The results shown here are in part based upon data generated by the TCGA Research Network: https://www.cancer.gov/tcga.

6.Funding

This work was supported by Fondazione Regionale per la Ricerca Biomedica (Regione Lombardia) (ERAPERMED2018-233 FindingMS GA 779282)

7. Figures

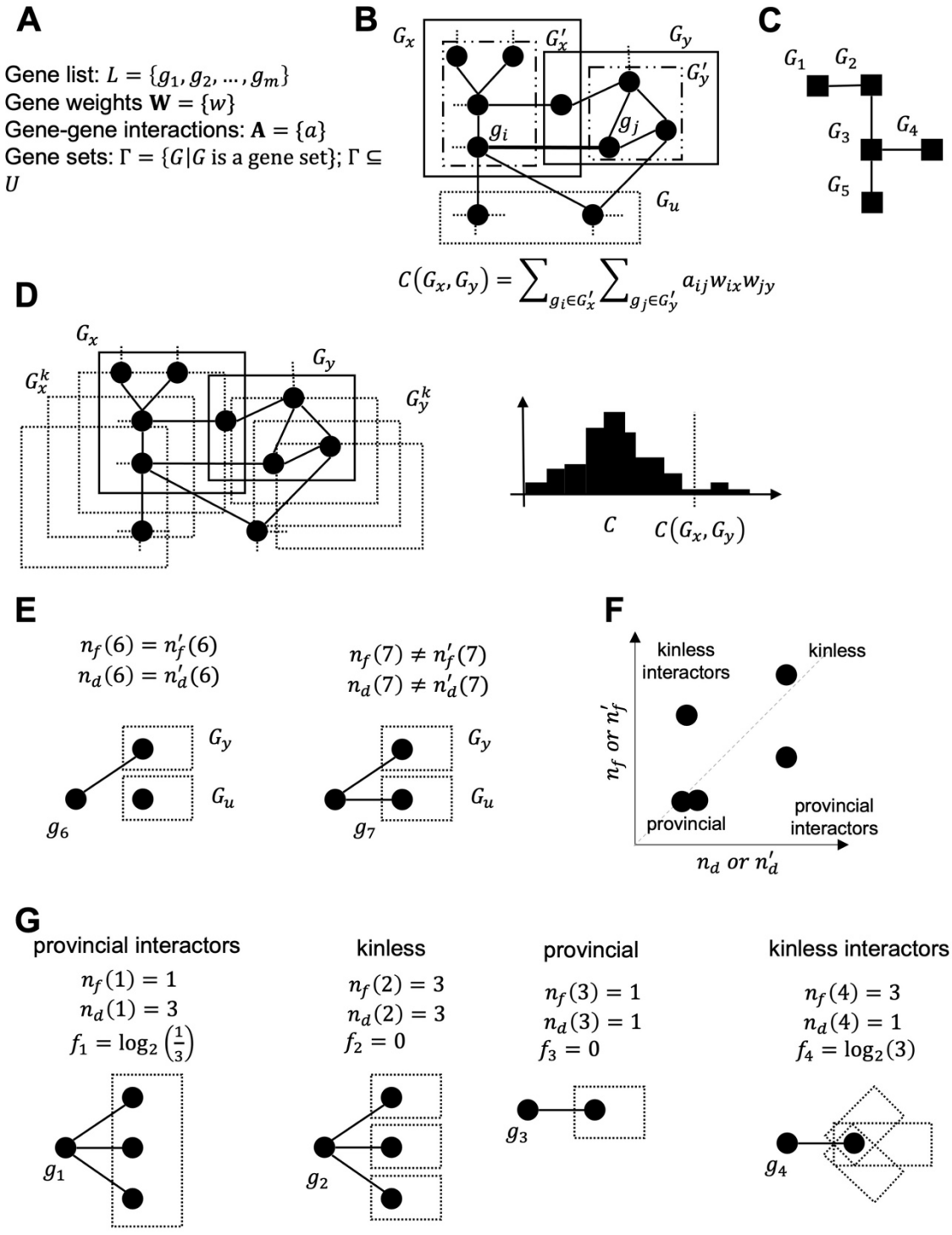
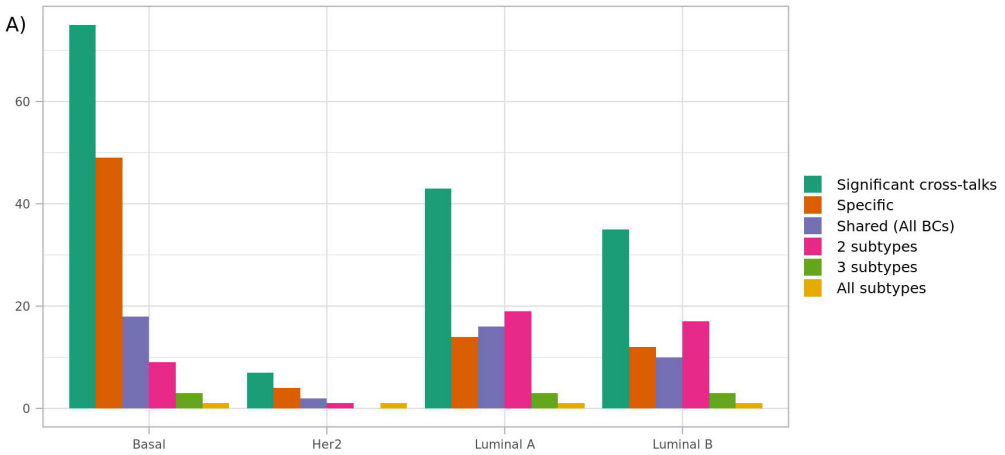
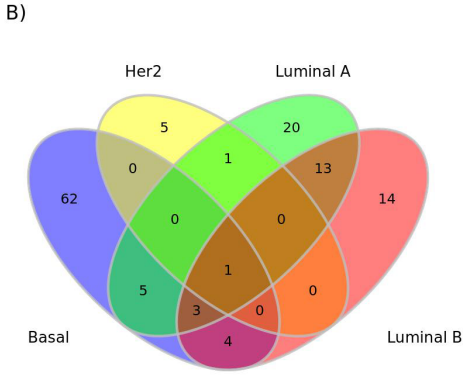

A
Gene list: $L = \{g_1, g_2, \dots, g_m\}$
Gene weights $\mathbf{W} = \{w\}$
Gene-gene interactions: $\mathbf{A} = \{a\}$
Gene sets: $\Gamma = \{G | G \text{ is a gene set}\}$; $\Gamma \subseteq U$
B
G_x
G'_x
G_y
G'_y
g_i
g_j
G_u
$C(G_x, G_y) = \sum_{g_i \in G'_x} \sum_{g_j \in G'_y} a_{ij} w_{ix} w_{jy}$
C
G_1
G_2
G_3
G_4
G_5
D
G_x
G_x^k
G_y
G_y^k
C
$C(G_x, G_y)$
E
$n_f(6) = n'_f(6)$
$n_d(6) = n'_d(6)$
$n_f(7) \neq n'_f(7)$
$n_d(7) \neq n'_d(7)$
G_y
G_u
g_6
g_7
F
kinless interactors
kinless
provincial
provincial interactors
n_f or n'_f
n_d or n'_d
G
provincial interactors
$n_f(1) = 1$
$n_d(1) = 3$
$f_1 = \log_2\left(\frac{1}{3}\right)$
kinless
$n_f(2) = 3$
$n_d(2) = 3$
$f_2 = 0$
provincial
$n_f(3) = 1$
$n_d(3) = 1$
$f_3 = 0$
kinless interactors
$n_f(4) = 3$
$n_d(4) = 1$
$f_4 = \log_2(3)$
g_1
g_2
g_3
g_4

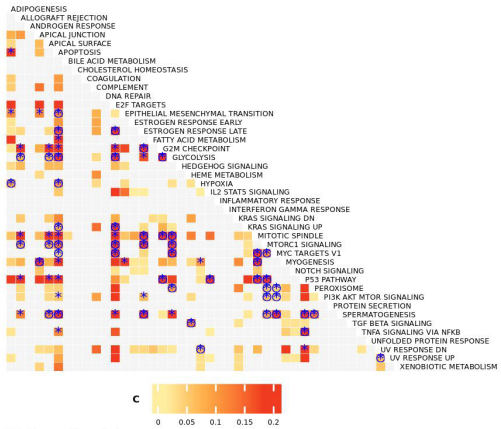
Figure 1 – Cross-talk quantification and gene classification. A) Involved quantities; B) Cross-talk definition; C) cross-talk network; D) empirical null model; E) absolute vs relative functional diversity and interactor diversity; F) gene classification; G) examples of genes classification.

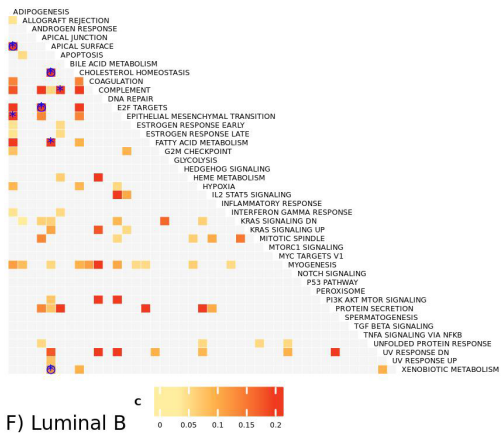
Figure 2 – Cross-talks between MSigDB "Hallmarks" gene sets. A) Heatmap of cross-talk values

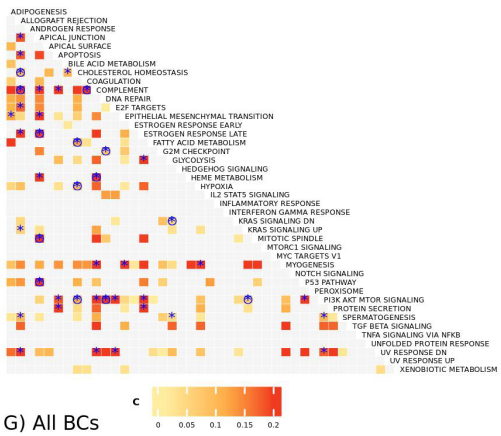
between Hallmarks gene sets (gray indicates no cross-talk); (B) functional (n_f) and interactor (n_d) diversities of the genes involved in the cross talks; points are colored by functional relevance (f).

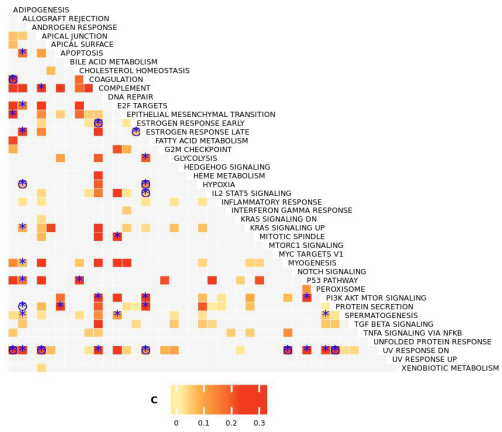

A)
60
40
20
0
Basal
Her2
Luminal A
Luminal B
Significant cross-talks
Specific
Shared (All BCs)
2 subtypes
3 subtypes
All subtypes


B)
Her2
Luminal A
Basal
Luminal B
5
20
0
1
13
62
14
0
0
1
5
0
3
0
4


C) Basal
ADIPOGENESIS
ALLOGRAFT REJECTION
ANDROGEN RESPONSE
APICAL JUNCTION
APICAL SURFACE
APOPTOSIS
BILE ACID METABOLISM
CHOLESTEROL HOMEOSTASIS
COAGULATION
COMPLEMENT
DNA REPAIR
E2F TARGETS
EPITHELIAL MESENCHYMAL TRANSITION
ESTROGEN RESPONSE EARLY
ESTROGEN RESPONSE LATE
FATTY ACID METABOLISM
G2M CHECKPOINT
GLYCOLYSIS
HEDGEHOG SIGNALING
HEME METABOLISM
HYPOXIA
IL2 STAT5 SIGNALING
INFLAMMATORY RESPONSE
INTERFERON GAMMA RESPONSE
KRAS SIGNALING DN
KRAS SIGNALING UP
MITOTIC SPINDLE
MTORC1 SIGNALING
MYC TARGETS V1
MYOGENESIS
NOTCH SIGNALING
P53 PATHWAY
PEROXISOME
PI3K AKT MTOR SIGNALING
PROTEIN SECRETION
SPERMATOGENESIS
TGF BETA SIGNALING
TNFA SIGNALING VIA NFKB
UNFOLDED PROTEIN RESPONSE
UV RESPONSE DN
UV RESPONSE UP
XENOBIOTIC METABOLISM
c
0
0.05
0.1
0.15
0.2


D) Her2
ADIPOGENESIS
ALLOGRAFT REJECTION
ANDROGEN RESPONSE
APICAL JUNCTION
APICAL SURFACE
APOPTOSIS
BILE ACID METABOLISM
CHOLESTEROL HOMEOSTASIS
COAGULATION
COMPLEMENT
DNA REPAIR
E2F TARGETS
EPITHELIAL MESENCHYMAL TRANSITION
ESTROGEN RESPONSE EARLY
ESTROGEN RESPONSE LATE
FATTY ACID METABOLISM
G2M CHECKPOINT
GLYCOLYSIS
HEDGEHOG SIGNALING
HEME METABOLISM
HYPOXIA
IL2 STAT5 SIGNALING
INFLAMMATORY RESPONSE
INTERFERON GAMMA RESPONSE
KRAS SIGNALING DN
KRAS SIGNALING UP
MITOTIC SPINDLE
MTORC1 SIGNALING
MYC TARGETS V1
MYOGENESIS
NOTCH SIGNALING
P53 PATHWAY
PEROXISOME
PI3K AKT MTOR SIGNALING
PROTEIN SECRETION
SPERMATOGENESIS
TGF BETA SIGNALING
TNFA SIGNALING VIA NFKB
UNFOLDED PROTEIN RESPONSE
UV RESPONSE DN
UV RESPONSE UP
XENOBIOTIC METABOLISM
c
0
0.05
0.1
0.15
0.2


E) Luminal A
ADIPOGENESIS
ALLOGRAFT REJECTION
ANDROGEN RESPONSE
APICAL JUNCTION
APICAL SURFACE
APOPTOSIS
BILE ACID METABOLISM
CHOLESTEROL HOMEOSTASIS
COAGULATION
COMPLEMENT
DNA REPAIR
E2F TARGETS
EPITHELIAL MESENCHYMAL TRANSITION
ESTROGEN RESPONSE EARLY
ESTROGEN RESPONSE LATE
FATTY ACID METABOLISM
G2M CHECKPOINT
GLYCOLYSIS
HEDGEHOG SIGNALING
HEME METABOLISM
HYPOXIA
IL2 STAT5 SIGNALING
INFLAMMATORY RESPONSE
INTERFERON GAMMA RESPONSE
KRAS SIGNALING DN
KRAS SIGNALING UP
MITOTIC SPINDLE
MTORC1 SIGNALING
MYC TARGETS V1
MYOGENESIS
NOTCH SIGNALING
P53 PATHWAY
PEROXISOME
PI3K AKT MTOR SIGNALING
PROTEIN SECRETION
SPERMATOGENESIS
TGF BETA SIGNALING
TNFA SIGNALING VIA NFKB
UNFOLDED PROTEIN RESPONSE
UV RESPONSE DN
UV RESPONSE UP
XENOBIOTIC METABOLISM
c
0
0.05
0.1
0.15
0.2


F) Luminal B
ADIPOGENESIS
ALLOGRAFT REJECTION
ANDROGEN RESPONSE
APICAL JUNCTION
APICAL SURFACE
APOPTOSIS
BILE ACID METABOLISM
CHOLESTEROL HOMEOSTASIS
COAGULATION
COMPLEMENT
DNA REPAIR
E2F TARGETS
EPITHELIAL MESENCHYMAL TRANSITION
ESTROGEN RESPONSE EARLY
ESTROGEN RESPONSE LATE
FATTY ACID METABOLISM
G2M CHECKPOINT
GLYCOLYSIS
HEDGEHOG SIGNALING
HEME METABOLISM
HYPOXIA
IL2 STAT5 SIGNALING
INFLAMMATORY RESPONSE
INTERFERON GAMMA RESPONSE
KRAS SIGNALING DN
KRAS SIGNALING UP
MITOTIC SPINDLE
MTORC1 SIGNALING
MYC TARGETS V1
MYOGENESIS
NOTCH SIGNALING
P53 PATHWAY
PEROXISOME
PI3K AKT MTOR SIGNALING
PROTEIN SECRETION
SPERMATOGENESIS
TGF BETA SIGNALING
TNFA SIGNALING VIA NFKB
UNFOLDED PROTEIN RESPONSE
UV RESPONSE DN
UV RESPONSE UP
XENOBIOTIC METABOLISM
c
0
0.1
0.2
0.3

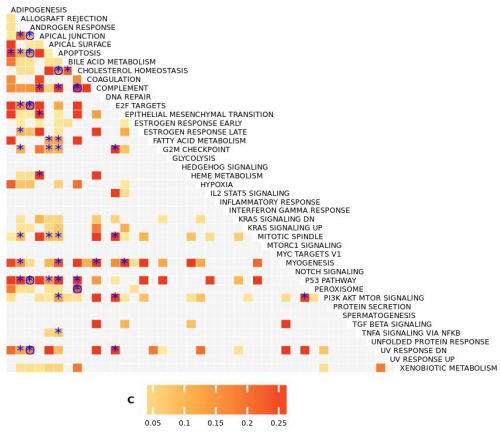

G) All BCs
ADIPOGENESIS
ALLOGRAFT REJECTION
ANDROGEN RESPONSE
APICAL JUNCTION
APICAL SURFACE
APOPTOSIS
BILE ACID METABOLISM
CHOLESTEROL HOMEOSTASIS
COAGULATION
COMPLEMENT
DNA REPAIR
E2F TARGETS
EPITHELIAL MESENCHYMAL TRANSITION
ESTROGEN RESPONSE EARLY
ESTROGEN RESPONSE LATE
FATTY ACID METABOLISM
G2M CHECKPOINT
GLYCOLYSIS
HEDGEHOG SIGNALING
HEME METABOLISM
HYPOXIA
IL2 STAT5 SIGNALING
INFLAMMATORY RESPONSE
INTERFERON GAMMA RESPONSE
KRAS SIGNALING DN
KRAS SIGNALING UP
MITOTIC SPINDLE
MTORC1 SIGNALING
MYC TARGETS V1
MYOGENESIS
NOTCH SIGNALING
P53 PATHWAY
PEROXISOME
PI3K AKT MTOR SIGNALING
PROTEIN SECRETION
SPERMATOGENESIS
TGF BETA SIGNALING
TNFA SIGNALING VIA NFKB
UNFOLDED PROTEIN RESPONSE
UV RESPONSE DN
UV RESPONSE UP
XENOBIOTIC METABOLISM
c
0.05
0.1
0.15
0.2
0.25

Figure 3 – Cross-talks between MSigDB "Hallmarks" gene sets in BCs. A) Number of significant cross-talks in BC subtypes. B) Overlap between cross-talks in BC subtypes. C-G) Heatmaps of cross-talks values, where grey colors indicate that there is no cross-talks; stars indicate significant cross-talks; circles indicate significant and specific cross-talks.

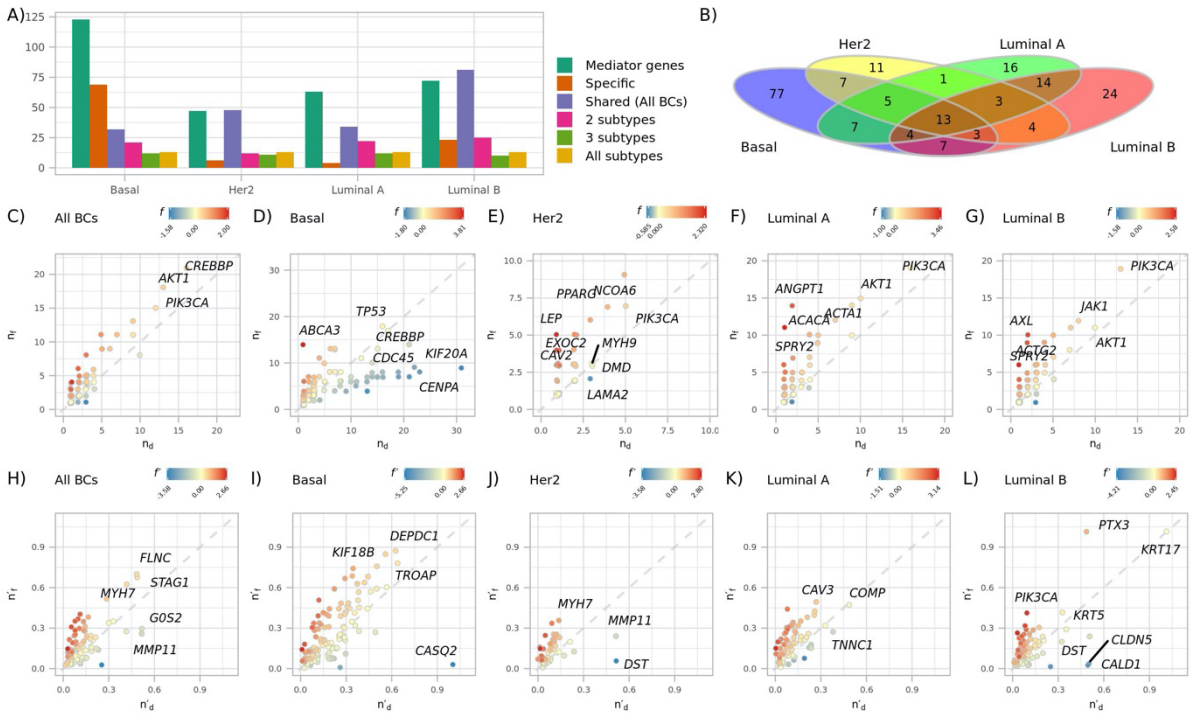


Figure 4 – Functional relevance of BC genes. A) Number of genes involved in significant cross-talks (”Mediator genes”). B) Mediator genes shared between subtypes. C-G) Functional (nf) and interactor (nd) diversities of the genes involved in the cross talks; points are colored by functional relevance (f). (H-L) Relative functional (nf) and interactor (nd) diversities of the genes involved in the cross talks; points are colored by relative functional relevance (f’).

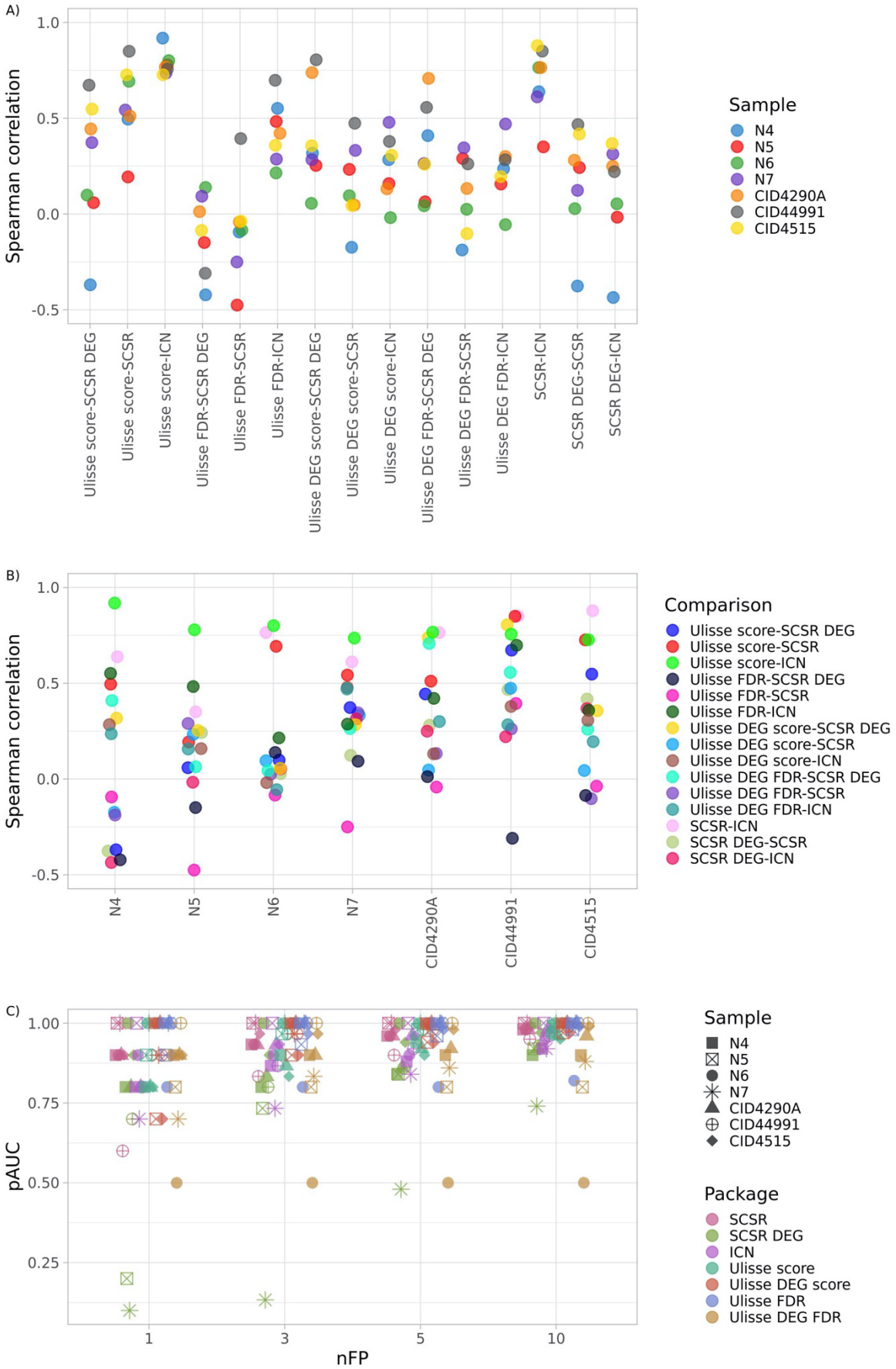


Figure 5 – Comparison of approaches for intercellular cross-talk quantification. Spearman

correlation (A, B) and pAUC score C) between approaches in various samples. Both statistics are calculated by using Ulisse CCC score ("Ulisse score" or "Ulisse DEG score") or the -log10 of FDR ("Ulisse FDR" or "Ulisse DEG FDR"). ICN = ICELLNET; SCSR = SingleCellSignalR; SCSR DEG = SingleCellSignalR with DEG specificity.

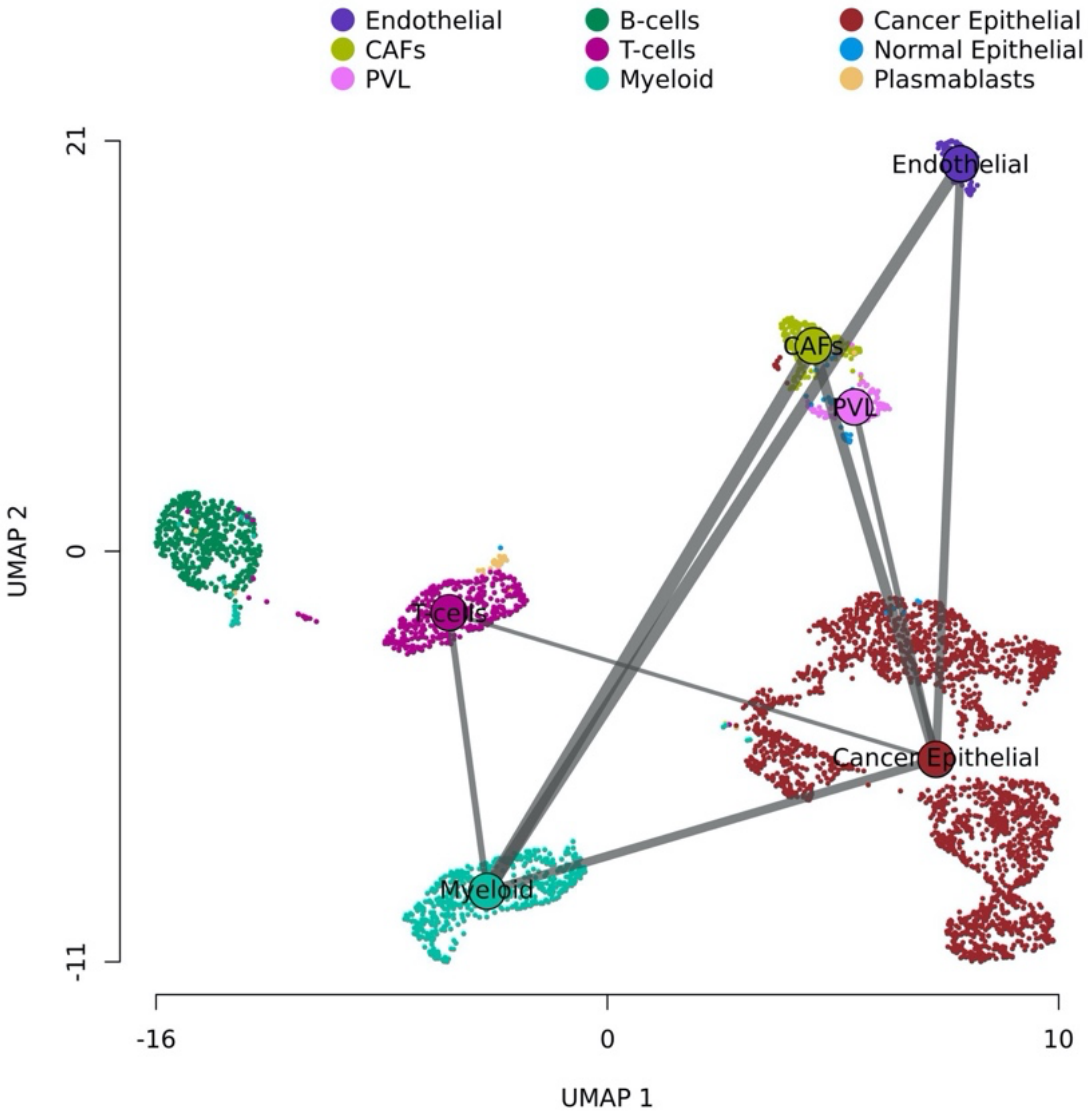


Figure 6 – Intercellular cross-talk in BC sample CID4515. UMAP visualization of cells colored by cell type and the corresponding intercellular cross-talk network identified through Ulisse. Edge width is proportional to communication score.

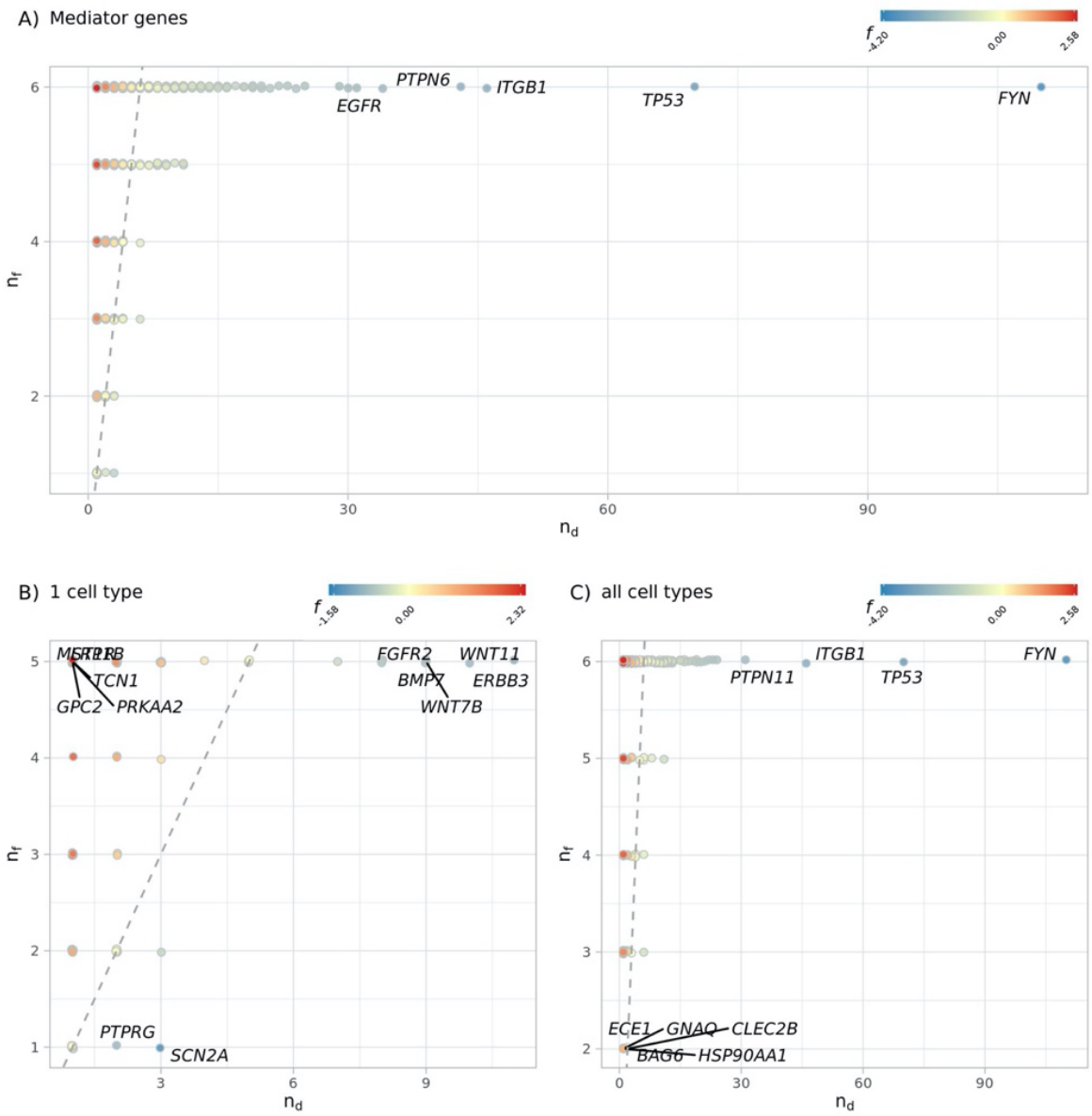


Figure 7 – Functional relevance of genes expressed in "cancer epithelial" cells and involved in intercellular cross-talks of CID4515 BC sample. Functional relevance of only the genes present in Cancer Epithelial gene set. Functional (nf) and interactor (nd) diversities of (A) cancer-specific genes, (B) mediators expressed in Cancer Epithelial cells and (C) cancer genes expressed in all cell-types. Points are colored by functional relevance (f).

8.Tables

Table 1 – Pathway cross-talk analysis in BC. Number of samples, altered genes, pathways and cross-talk found by subtype.

	Basal	**Her2**	**Luminal A**	**Luminal B**	**All BCs**
Mutation data # samples (% All BCs)	175 (17.7%)	76 (7.7%)	497 (50.4%)	194 (19.7%)	987
Expression data # samples (% All BCs)	187 (15.3%)	85 (6.9%)	552 (45.0%)	215 (17.5%)	1226
# Top list altered genes (mutated and DE)	428	367	394	393	402
# Pathways	32	35	29	34	36
# Altered genes per pathway	13.38	10.49	13.59	11.56	11.17
# Significant cross-talks (FDR < 25%)	**75**	**7**	**43**	**35**	**42**
# Specific cross-talks (% Significant)	49 **(65%)**	4 **(57%)**	14 **(33%)**	12 **(34%)**	8 **(19%)**
# Shared with All BCs (%)	18 **(24%)**	2 **(29%)**	16 **(37%)**	10 **(29%)**	42
# Mediator genes (% Top list)	123 **(28.7%)**	47 **(12.8%)**	63 **(16%)**	72 **(18.3%)**	81 **(20.1%)**

9.References

Antony J, Chin CV, Horsfield JA. Cohesin Mutations in Cancer: Emerging Therapeutic Targets. *Int J Mol Sci* 2021;**22**:6788.

Antony J, Dasgupta T, Rhodes JM *et al.* Cohesin modulates transcription of estrogen-responsive genes. *Biochim Biophys Acta* 2015;**1849**:257–69.

Armingol E, Officer A, Harismendy O *et al.* Deciphering cell–cell interactions and communication from gene expression. *Nat Rev Genet* 2021;**22**, DOI: 10.1038/s41576-020-00292-x.

Barabási A-L, Gulbahce N, Loscalzo J. Network medicine: a network-based approach to human disease. *Nat Rev Genet* 2011;**12**:56–68.

Benjamini Y, Hochberg Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society: Series B (Methodological)* 1995;**57**:289–300.

Bogdan M, Meca A-D, Turcu-Stiolica A *et al.* Insights into the Relationship between Pentraxin-3 and Cancer. *Int J Mol Sci* 2022;**23**:15302.

Boyle EA, Li YI, Pritchard JK. An Expanded View of Complex Traits: From Polygenic to Omnigenic. *Cell* 2017;**169**:1177–86.

Cabello-Aguilar S, Alame M, Kon-Sun-Tack F *et al.* SingleCellSignalR: Inference of intercellular networks from single-cell transcriptomics. *Nucleic Acids Res* 2021;**48**, DOI: 10.1093/NAR/GKAA183.

Cavazzoni A, Bonelli MA, Fumarola C *et al.* Overcoming acquired resistance to letrozole by targeting the PI3K/AKT/mTOR pathway in breast cancer cell clones. *Cancer Lett* 2012;**323**:77–87.

Chan TA, Hermeking H, Lengauer C *et al.* 14-3-3Sigma is required to prevent mitotic catastrophe after DNA damage. *Nature* 1999;**401**:616–20.

Colaprico A, Silva TC, Olsen C *et al.* TCGAbiolinks: an R/Bioconductor package for integrative analysis of TCGA data. *Nucleic Acids Res* 2016;**44**:e71–e71.

Deng J, Fleming JB. Inflammation and Myeloid Cells in Cancer Progression and Metastasis. *Front Cell Dev Biol* 2022;**9**, DOI: 10.3389/fcell.2021.759691.

DiMeo TA, Anderson K, Phadke P *et al.* A novel lung metastasis signature links Wnt signaling with cancer cell self-renewal and epithelial-mesenchymal transition in basal-like breast cancer. *Cancer Res* 2009;**69**:5364–73.

Dimitrov D, Türei D, Garrido-Rodriguez M *et al.* Comparison of methods and resources for cell-cell communication inference from single-cell RNA-Seq data. *Nat Commun* 2022;**13**:3224.

Dolgalev I. msigdbr: MSigDB Gene Sets for Multiple Organisms in a Tidy Data Format. 2022.

Dutta B, Wallqvist A, Reifman J. PathNet: A tool for pathway analysis using topological information. *Source Code Biol Med* 2012;**7**, DOI: 10.1186/1751-0473-7-10.

Elwakeel E, Weigert A. Breast Cancer CAFs: Spectrum of Phenotypes and Promising Targeting Avenues. *Int J Mol Sci* 2021;**22**:11636.

Goodman RH, Smolik S. CBP/p300 in cell growth, transformation, and development. *Genes Dev* 2000;**14**:1553–77.

Guo Z, Zhang H, Fu Y *et al.* Cancer-associated fibroblasts induce growth and radioresistance of breast cancer cells through paracrine IL-6. *Cell Death Discov* 2023;**9**:6.

Hao Y, Hao S, Andersen-Nissen E *et al.* Integrated analysis of multimodal single-cell data. *Cell* 2021;**184**:3573-3587.e29.

Heckmann BL, Zhang X, Xie X *et al.* The G0/G1 switch gene 2 (G0S2): regulating metabolism and beyond. *Biochim Biophys Acta* 2013;**1831**:276–81.

Hendrix MJC. De-mystifying the mechanism(s) of maspin. *Nat Med* 2000;**6**:374–6.

Hu D, Li Z, Zheng B *et al.* Cancer-associated fibroblasts in breast cancer: Challenges and opportunities. *Cancer Commun* 2022;**42**:401–34.

Hu Y, Yang Y, Fang Z *et al.* Detecting pathway relationship in the context of human protein-protein interaction network and its application to Parkinson's disease. *Methods* 2017;**131**:93–103.

Huang JK, Carlin DE, Yu MK *et al.* Systematic Evaluation of Molecular Networks for Discovery of Disease Genes. *Cell Syst* 2018;**6**:484-495.e5.

Jaeger S, Igea A, Arroyo R *et al.* Quantification of pathway cross-talk reveals novel synergistic drug combinations for breast cancer. *Cancer Res* 2017;**77**, DOI: 10.1158/0008-5472.CAN-16-0097.

Jin S, Guerrero-Juarez CF, Zhang L *et al.* Inference and analysis of cell-cell communication using CellChat. *Nat Commun* 2021;**12**:1088.

Jin Z, Peng F, Zhang C *et al.* Expression, regulating mechanism and therapeutic target of KIF20A in multiple cancer. *Heliyon* 2023;**9**:e13195.

Kamil M, Shinsato Y, Higa N *et al.* High filamin-C expression predicts enhanced invasiveness and poor outcome in glioblastoma multiforme. *Br J Cancer* 2019;**120**:819–26.

Kang SU, Cho SY, Jeong H *et al.* Matrix metalloproteinase 11 (MMP11) in macrophages promotes the migration of HER2-positive breast cancer cells and monocyte recruitment through CCL2–CCR2 signaling. *Laboratory Investigation* 2022;**102**:376–90.

Kim JH, Lee E, Yun J *et al.* Calsequestrin 2 overexpression in breast cancer increases tumorigenesis and metastasis by modulating the tumor microenvironment. *Mol Oncol* 2022;**16**:466–84.

Liberzon A, Birger C, Thorvaldsdóttir H *et al.* The Molecular Signatures Database (MSigDB) hallmark gene set collection. *Cell Syst* 2015;**1**:417.

Mao X, Xu J, Wang W *et al.* Crosstalk between cancer-associated fibroblasts and immune cells in the tumor microenvironment: new findings and future perspectives. *Mol Cancer* 2021;**20**:131.

Mosca E, Alfieri R, Maj C *et al.* Computational modeling of the metabolic states regulated by the kinase Akt. *Front Physiol* 2012a;**3 NOV**:418.

Mosca E, Barcella M, Alfieri R *et al.* Systems biology of the metabolic network regulated by the Akt pathway. *Biotechnol Adv* 2012b;**30**:131–41.

Mosca E, Bersanelli M, Matteuzzi T *et al.* Characterization and comparison of gene-centered human interactomes. *Brief Bioinform* 2021;**22**, DOI: 10.1093/bib/bbab153.

Nahta R. Pharmacological strategies to overcome HER2 cross-talk and Trastuzumab resistance. *Curr Med Chem* 2012;**19**:1065–75.

Nakamura M, Takano A, Thang P *et al.* Characterization of KIF20A as a prognostic biomarker and therapeutic target for different subtypes of breast cancer. *Int J Oncol* 2020;**57**:277–88.

Nguyen QH, Pervolarakis N, Blake K *et al.* Profiling human breast epithelial cells using single cell RNA sequencing identifies cell diversity. *Nat Commun* 2018;**9**, DOI: 10.1038/s41467-018-04334-1.

Noël F, Massenet-Regad L, Carmi-Levy I *et al.* ICELLNET: a transcriptome-based framework to dissect intercellular communication. *bioRxiv* 2020:2020.03.05.976878.

Oda E, Ohki R, Murasawa H *et al.* Noxa, a BH3-only member of the Bcl-2 family and candidate mediator of p53-induced apoptosis. *Science* 2000a;**288**:1053–8.

Oda K, Arakawa H, Tanaka T *et al.* p53AIP1, a potential mediator of p53-dependent apoptosis, and its regulation by Ser-46-phosphorylated p53. *Cell* 2000b;**102**:849–62.

O'Regan R, Paplomata. New and emerging treatments for estrogen receptor-positive breast cancer: focus on everolimus. *Ther Clin Risk Manag* 2013;**9**:27.

Paplomata E, O'regan R. The PI3K/AKT/mTOR pathway in breast cancer: targets, trials and biomarkers. *Ther Adv Med Oncol* 2014;**6**:154–66.

Porras P, Barrera E, Bridge A *et al.* Towards a unified open access dataset of molecular interactions. *Nat Commun* 2020;**11**:6144.

Ring A, Nguyen C, Smbatyan G *et al.* CBP/β-Catenin/FOXM1 Is a Novel Therapeutic Target in Triple Negative Breast Cancer. *Cancers (Basel)* 2018;**10**, DOI: 10.3390/CANCERS10120525.

Ritchie ME, Phipson B, Wu D *et al.* limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res* 2015;**43**:e47–e47.

Suarez-Kelly LP, Yu L, Kline D *et al.* Increased breast cancer risk in women with neurofibromatosis type 1: a meta-analysis and systematic review of the literature. *Hered Cancer Clin Pract* 2019;**17**:12.

Subramanian A, Tamayo P, Mootha VK *et al.* Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A* 2005;**102**:15545–50.

Tang S, Liu W, Yong L *et al.* Reduced Expression of KRT17 Predicts Poor Prognosis in HER2high Breast Cancer. *Biomolecules* 2022;**12**:1183.

Türei D, Korcsmáros T, Saez-Rodriguez J. OmniPath: guidelines and gateway for literature-curated signaling pathway resources. *Nature Methods 2016 13:12* 2016;**13**:966–7.

Valdeolivas A, Turei D, Gabor A. OmnipathR: client for the OmniPath web service. 2019.

Vogelstein B, Lane D, Levine AJ. Surfing the p53 network. *Nature 2000 408:6810* 2000;**408**:307–10.

Waddell A, Mahmud I, Ding H *et al.* Pharmacological Inhibition of CBP/p300 Blocks Estrogen Receptor Alpha (ERα) Function through Suppressing Enhancer H3K27 Acetylation in Luminal Breast Cancer. *Cancers (Basel)* 2021;**13**, DOI: 10.3390/CANCERS13112799.

Wu SZ, Al-Eryani G, Roden DL *et al.* A single-cell and spatially resolved atlas of human breast cancers. *Nat Genet* 2021;**53**, DOI: 10.1038/s41588-021-00911-1.

Yim CY, Bikorimana E, Khan E *et al.* G0S2 represses PI3K/mTOR signaling and increases sensitivity to PI3K/mTOR pathway inhibitors in breast cancer. *Cell Cycle* 2017;**16**:2146–55.

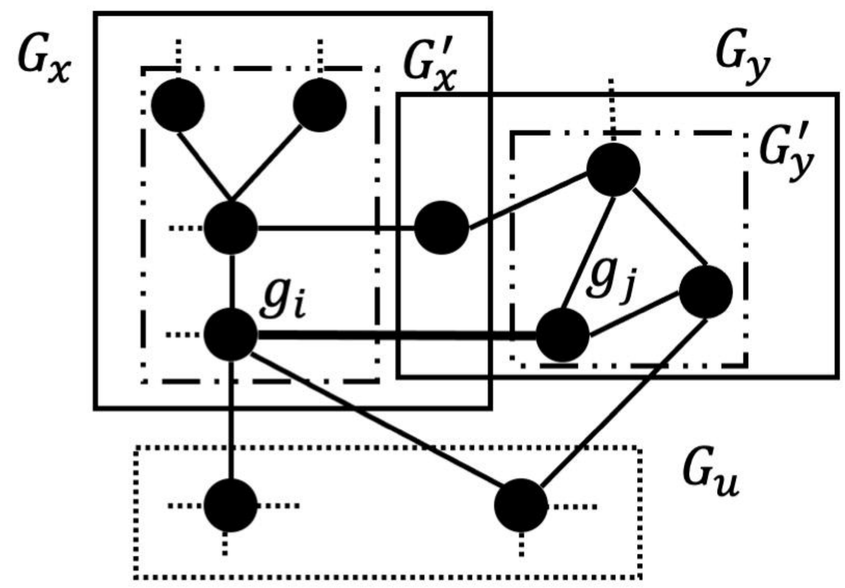
Yim CY, Sekula DJ, Hever-Jardine MP *et al.* G0S2 Suppresses Oncogenic Transformation by Repressing a MYC-Regulated Transcriptional Program. *Cancer Res* 2016;**76**:1204–13.

Yun J, Song SH, Kang JY *et al.* Reduced cohesin destabilizes high-level gene amplification by disrupting pre-replication complex bindings in human cancers with chromosomal instability. *Nucleic Acids Res* 2016;**44**:558–72.

A

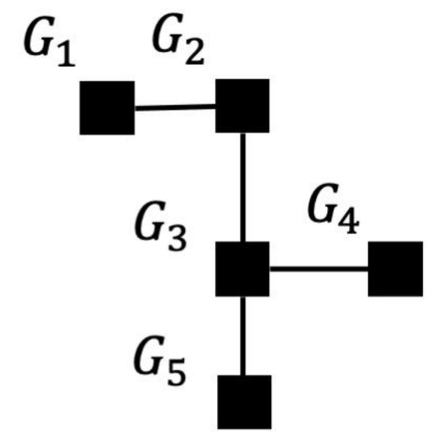
Gene list: $L = \{g_1, g_2, \dots, g_m\}$
Gene weights $\mathbf{W} = \{w\}$
Gene-gene interactions: $\mathbf{A} = \{a\}$
Gene sets: $\Gamma = \{G | G \text{ is a gene set}\}$; $\Gamma \subseteq U$

B

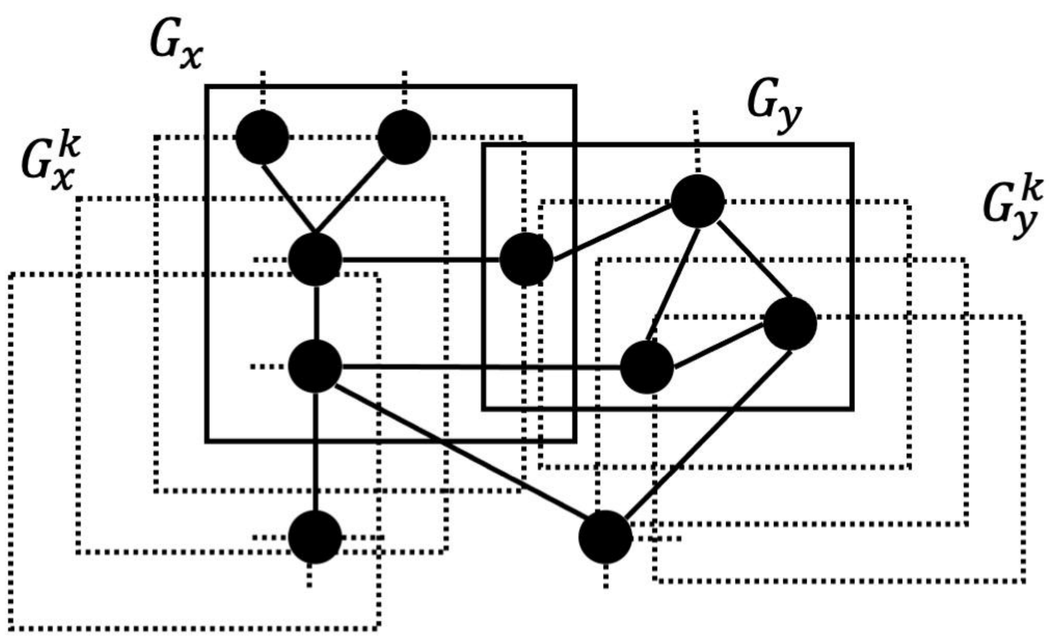


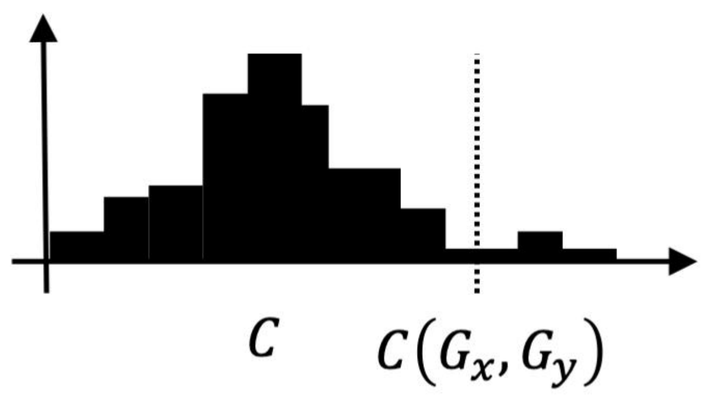
$$C(G_x, G_y) = \sum_{g_i \in G'_x} \sum_{g_j \in G'_y} a_{ij} w_{ix} w_{jy}$$

C



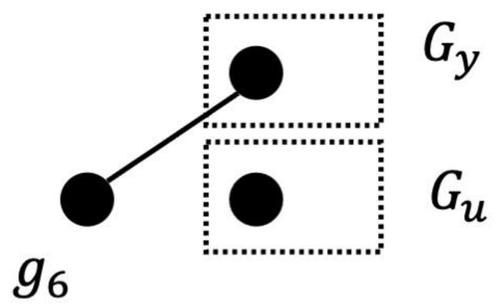
D



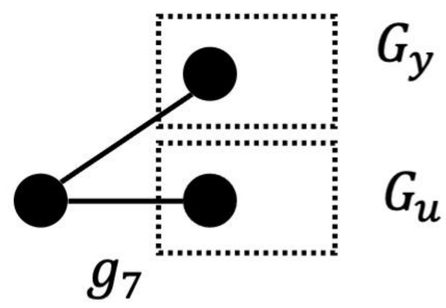


E

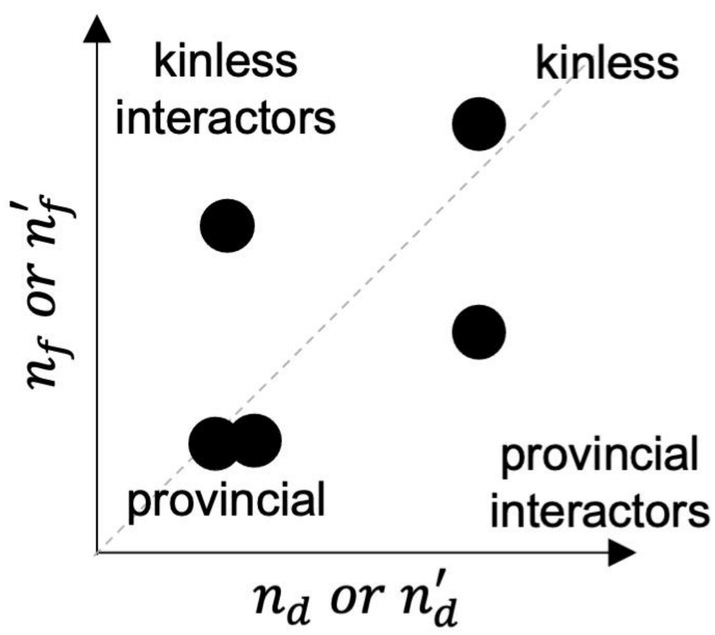
$n_f(6) = n'_f(6)$
$n_d(6) = n'_d(6)$



$n_f(7) \neq n'_f(7)$
$n_d(7) \neq n'_d(7)$



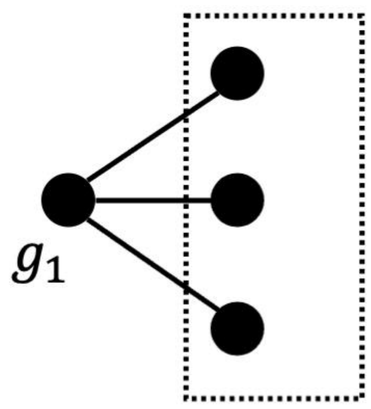
F



G

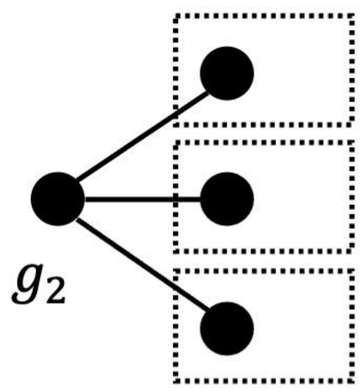
provincial interactors

$n_f(1) = 1$
$n_d(1) = 3$
$f_1 = \log_2\left(\frac{1}{3}\right)$



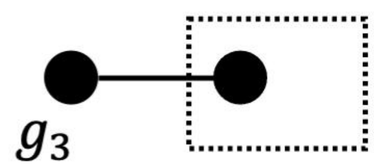
kinless

$n_f(2) = 3$
$n_d(2) = 3$
$f_2 = 0$



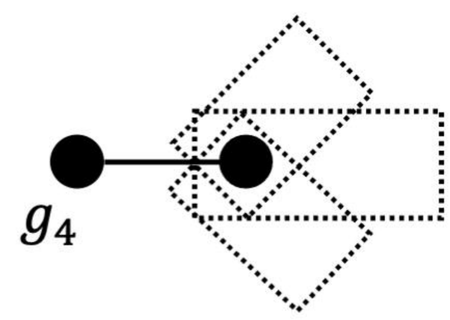
provincial

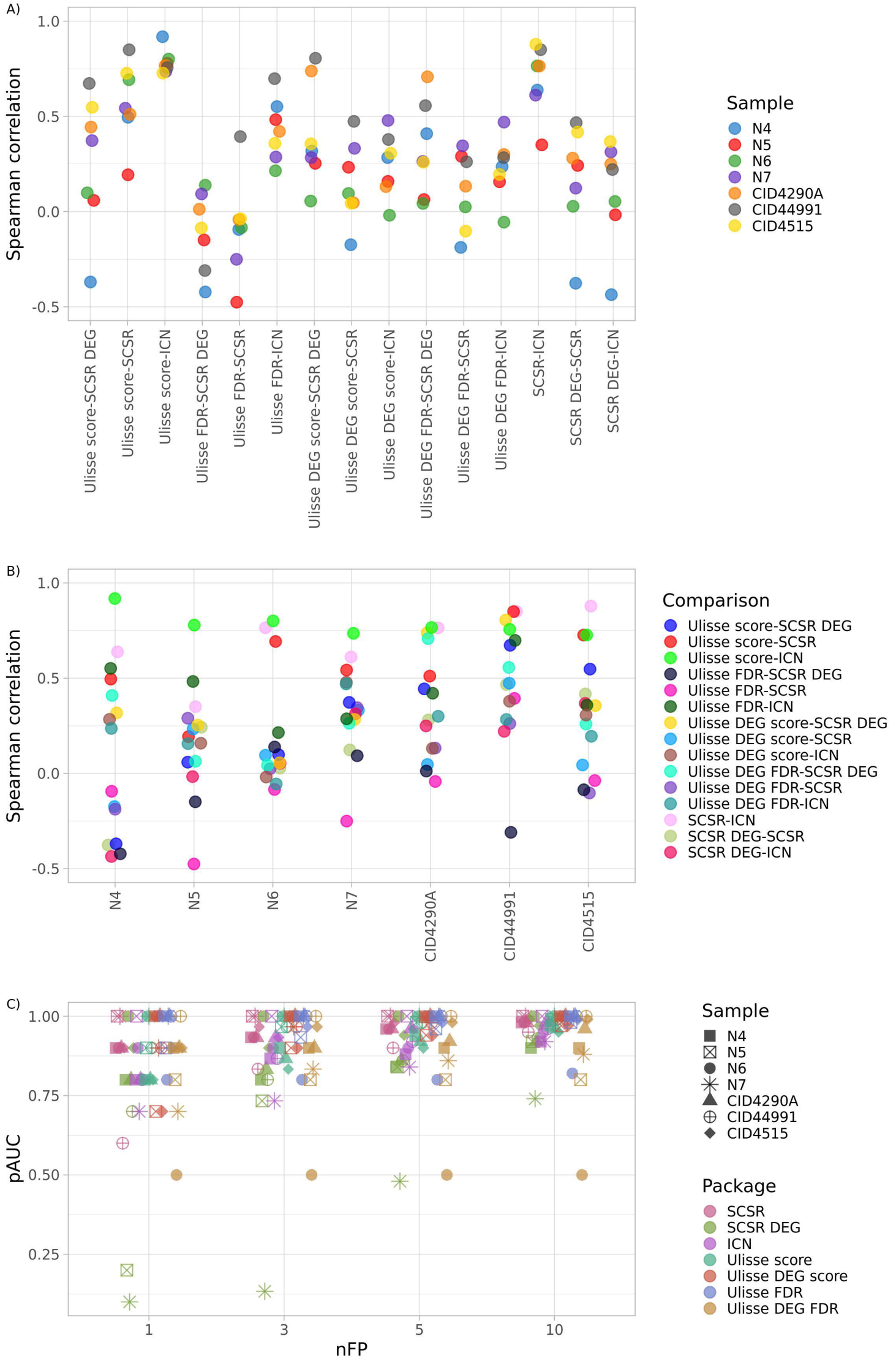
$n_f(3) = 1$
$n_d(3) = 1$
$f_3 = 0$

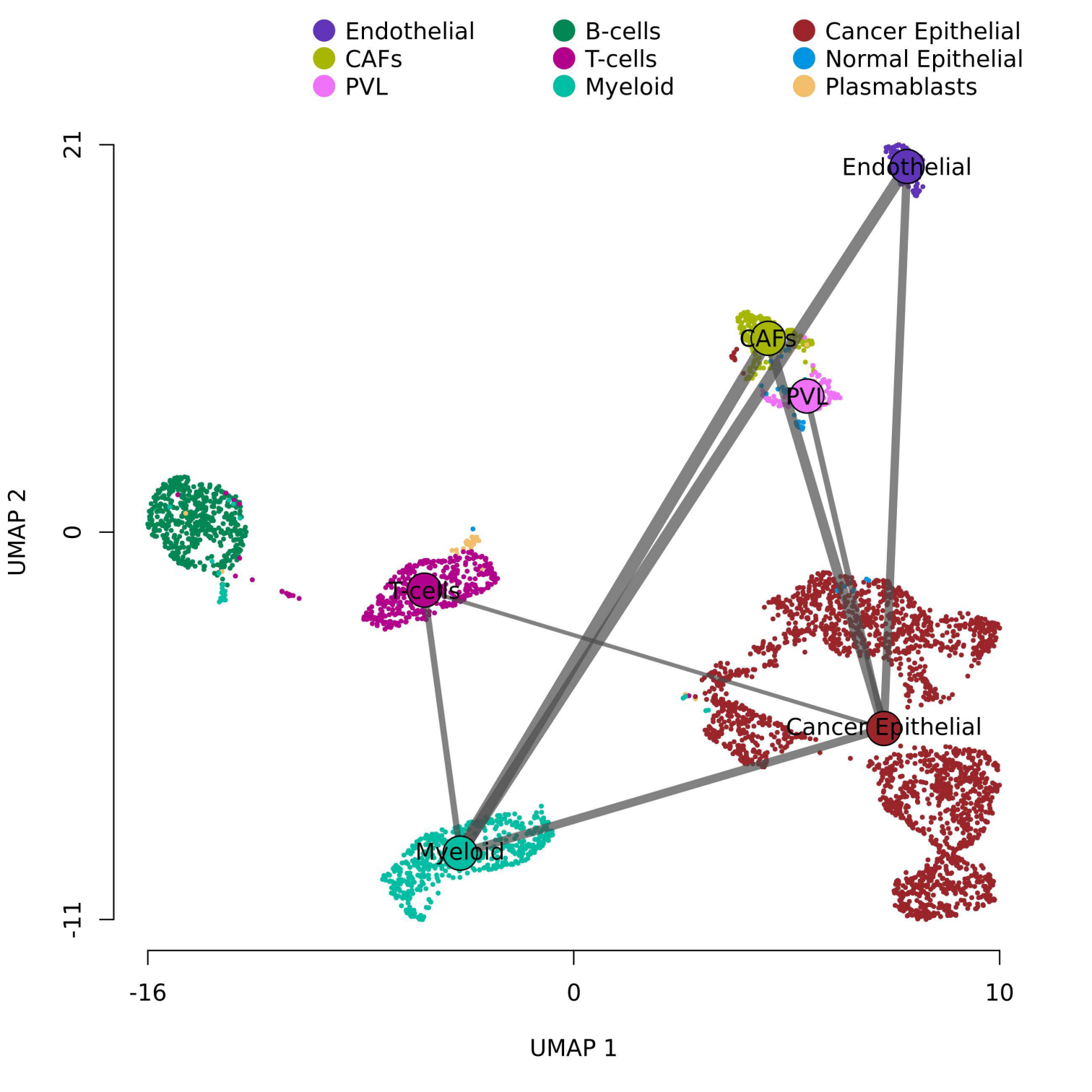


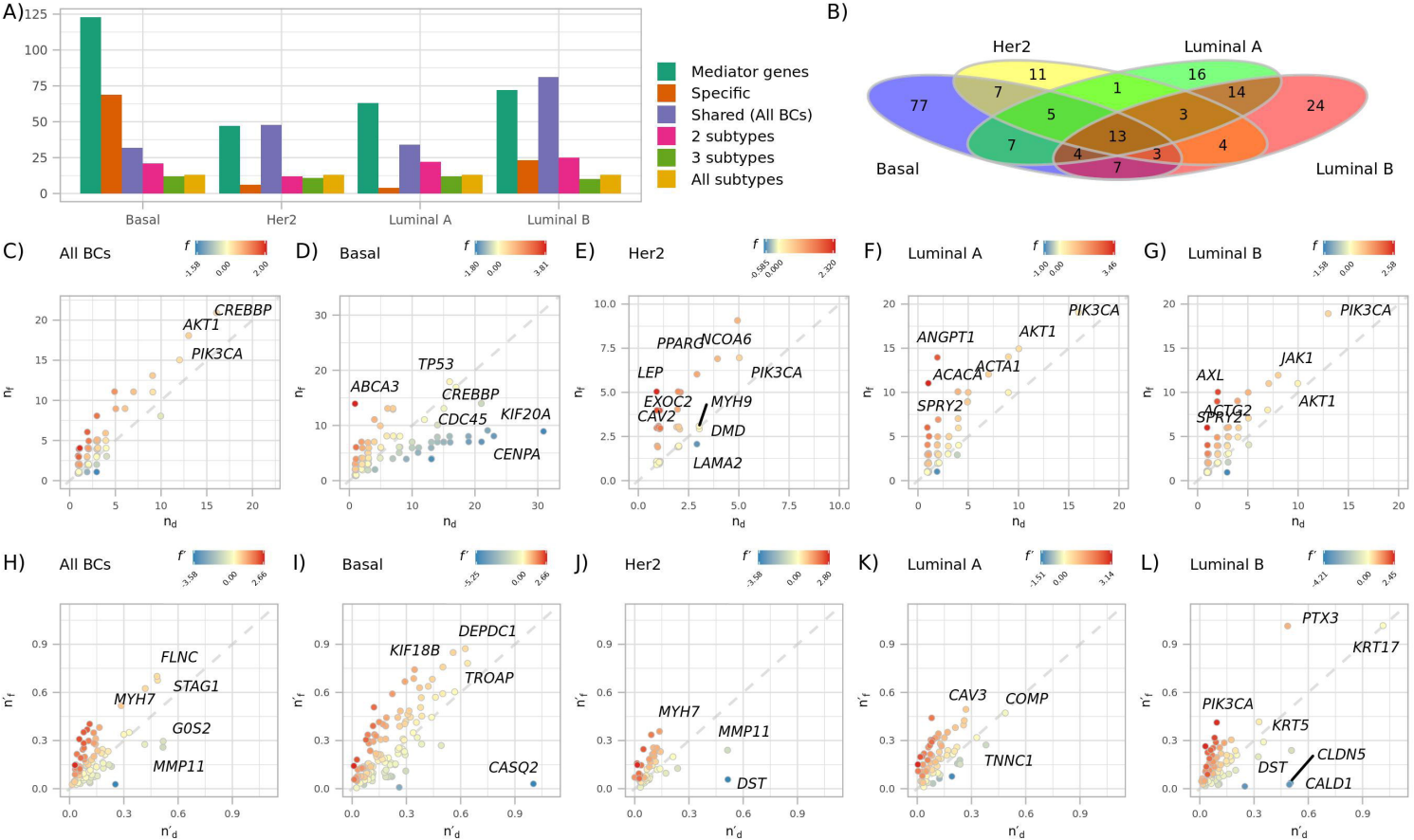
kinless interactors

$n_f(4) = 3$
$n_d(4) = 1$
$f_4 = \log_2(3)$

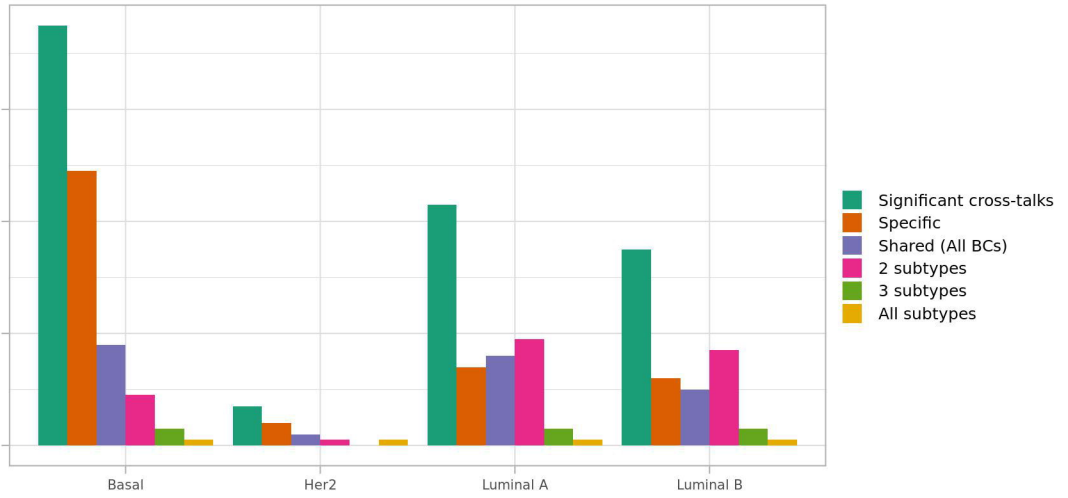



A)
Spearman correlation
1.0
0.5
0.0
-0.5
Ulisse score-SCSR DEG
Ulisse score-SCSR
Ulisse score-ICN
Ulisse FDR-SCSR DEG
Ulisse FDR-SCSR
Ulisse FDR-ICN
Ulisse DEG score-SCSR DEG
Ulisse DEG score-SCSR
Ulisse DEG score-ICN
Ulisse DEG FDR-SCSR DEG
Ulisse DEG FDR-SCSR
Ulisse DEG FDR-ICN
SCSR-ICN
SCSR DEG-SCSR
SCSR DEG-ICN
Sample
N4
N5
N6
N7
CID4290A
CID44991
CID4515
B)
Spearman correlation
1.0
0.5
0.0
-0.5
N4
N5
N6
N7
CID4290A
CID44991
CID4515
Comparison
Ulisse score-SCSR DEG
Ulisse score-SCSR
Ulisse score-ICN
Ulisse FDR-SCSR DEG
Ulisse FDR-SCSR
Ulisse FDR-ICN
Ulisse DEG score-SCSR DEG
Ulisse DEG score-SCSR
Ulisse DEG score-ICN
Ulisse DEG FDR-SCSR DEG
Ulisse DEG FDR-SCSR
Ulisse DEG FDR-ICN
SCSR-ICN
SCSR DEG-SCSR
SCSR DEG-ICN
C)
pAUC
1.00
0.75
0.50
0.25
1
3
5
10
nFP
Sample
N4
N5
N6
N7
CID4290A
CID44991
CID4515
Package
SCSR
SCSR DEG
ICN
Ulisse score
Ulisse DEG score
Ulisse FDR
Ulisse DEG FDR

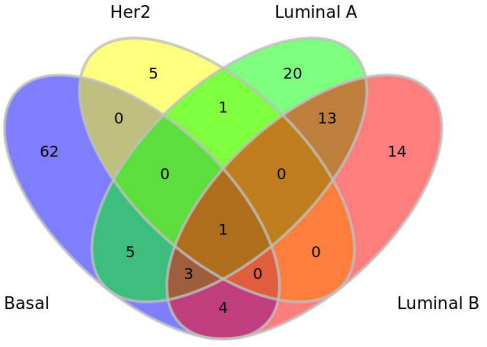

Endothelial
CAFs
PVL
B-cells
T-cells
Myeloid
Cancer Epithelial
Normal Epithelial
Plasmablasts
Endothelial
CAFs
PVL
T-cells
Cancer Epithelial
Myeloid
21
0
-11
UMAP 2
-16
0
10
UMAP 1


A)
Mediator genes
Specific
Shared (All BCs)
2 subtypes
3 subtypes
All subtypes
Basal
Her2
Luminal A
Luminal B
B)
Her2
Luminal A
Basal
Luminal B
C) All BCs
D) Basal
E) Her2
F) Luminal A
G) Luminal B
CREBBP
AKT1
PIK3CA
TP53
ABCA3
CREBBP
CDC45
KIF20A
CENPA
PPARD
NCOA6
LEP
PIK3CA
EXOC2
MYH9
CAV2
DMD
LAMA2
ANGPT1
AKT1
ACACA
ACTA1
SPRY2
PIK3CA
AXL
JAK1
AKT1
NCTG2
SPRY2
PIK3CA
H) All BCs
I) Basal
J) Her2
K) Luminal A
L) Luminal B
FLNC
STAG1
MYH7
G0S2
MMP11
DEPDC1
KIF18B
TROAP
CASQ2
MYH7
MMP11
DST
CAV3
COMP
TNNC1
PTX3
KRT17
PIK3CA
KRT5
CLDN5
DST
CALD1

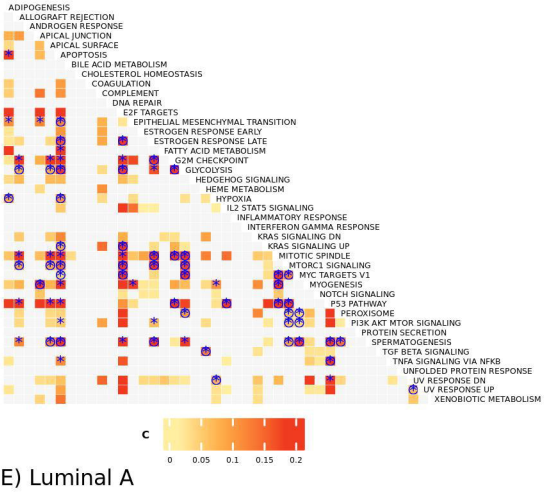
A)



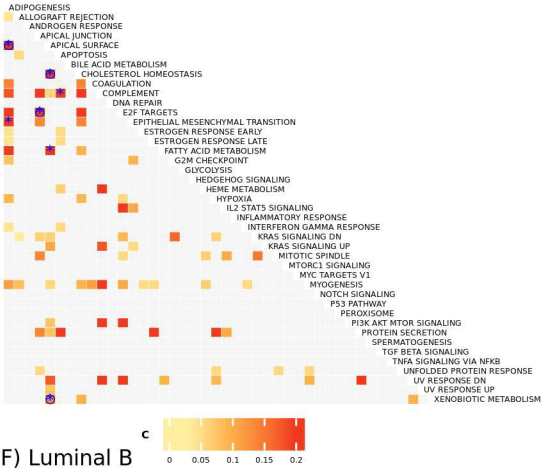
B)



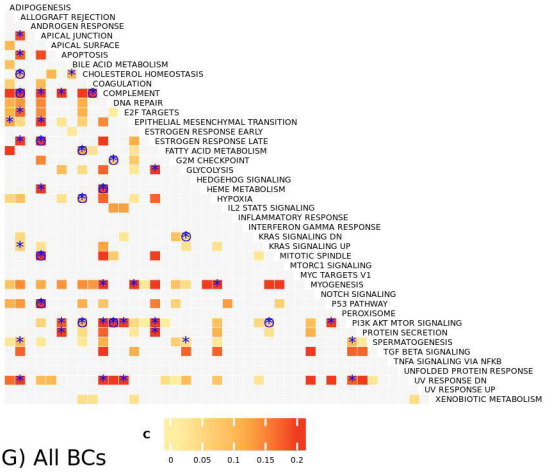
C) Basal



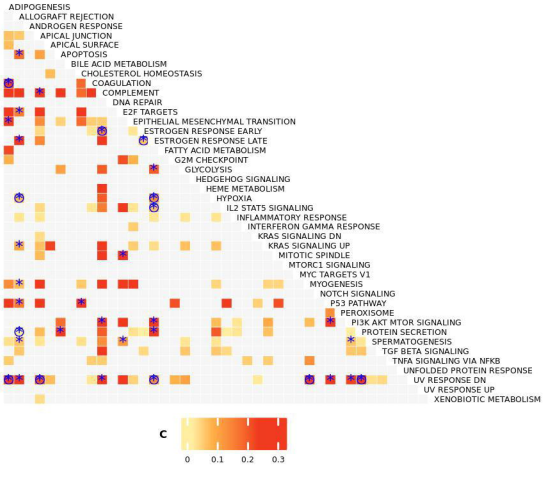
D) Her2



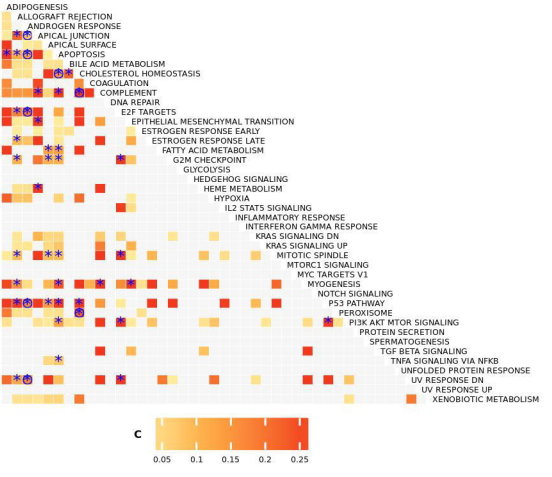
E) Luminal A

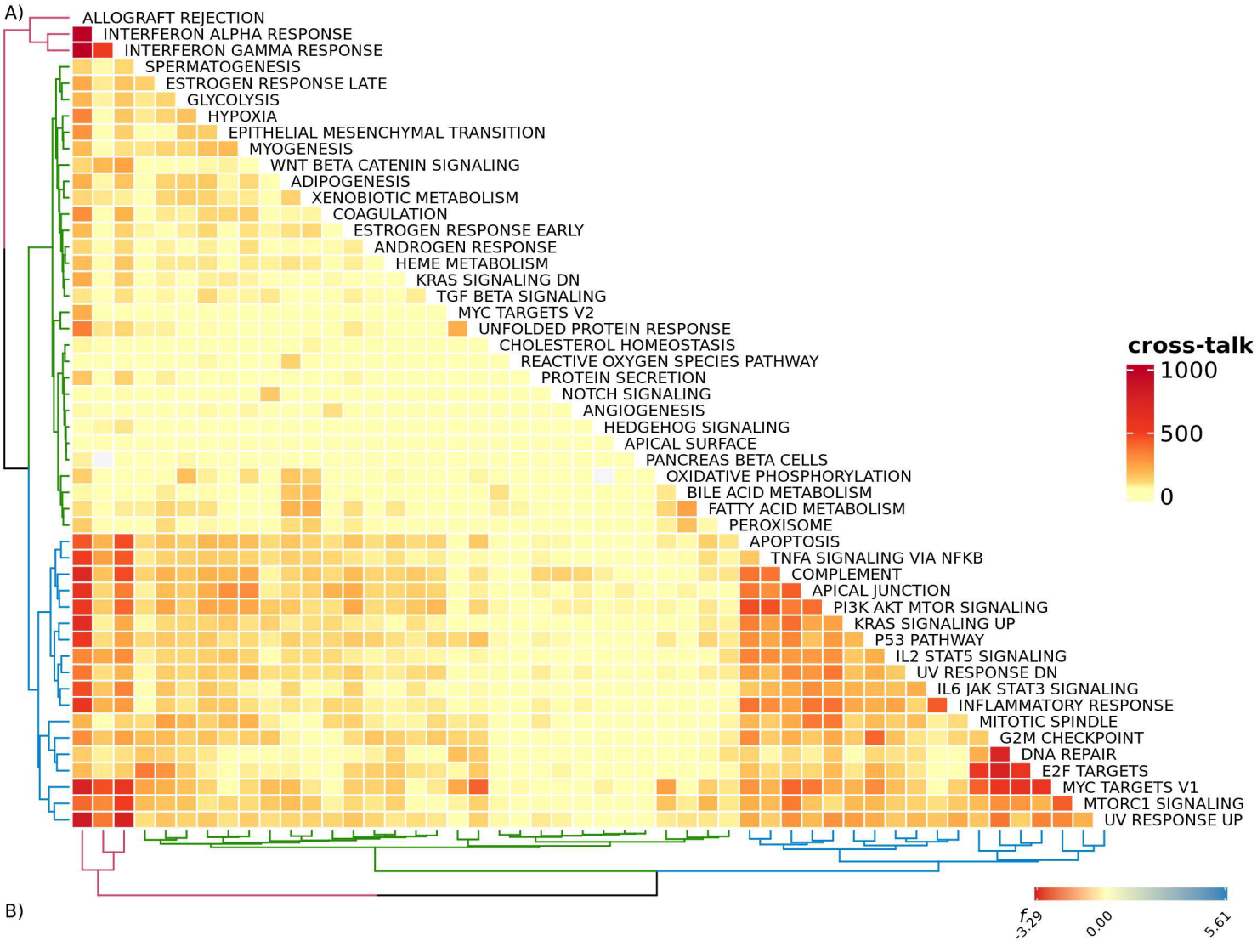


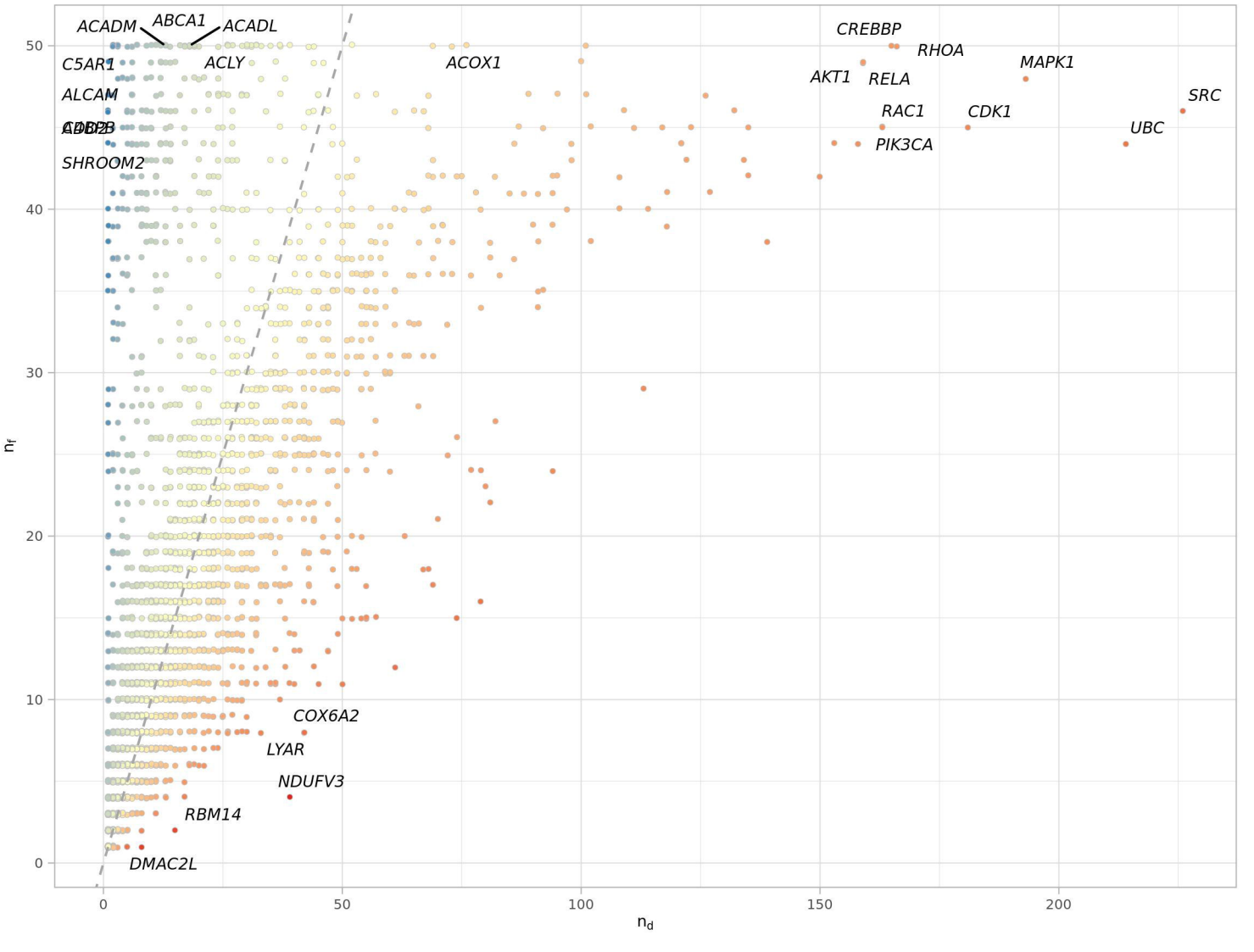
F) Luminal B

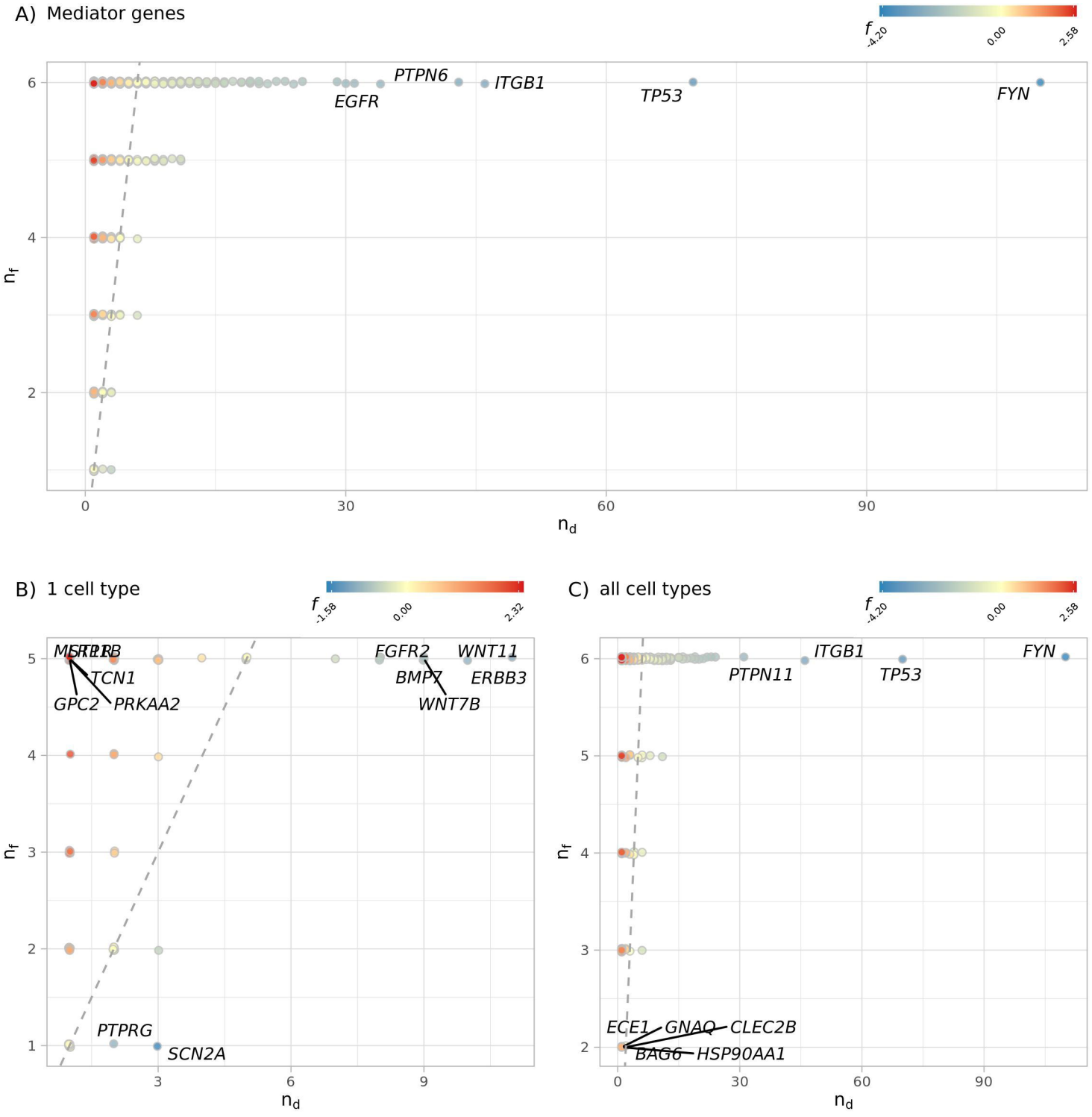


G) All BCs




A)
ALLOGRAFT REJECTION
INTERFERON ALPHA RESPONSE
INTERFERON GAMMA RESPONSE
SPERMATOGENESIS
ESTROGEN RESPONSE LATE
GLYCOLYSIS
HYPOXIA
EPITHELIAL MESENCHYMAL TRANSITION
MYOGENESIS
WNT BETA CATENIN SIGNALING
ADIPOGENESIS
XENOBIOTIC METABOLISM
COAGULATION
ESTROGEN RESPONSE EARLY
ANDROGEN RESPONSE
HEME METABOLISM
KRAS SIGNALING DN
TGF BETA SIGNALING
MYC TARGETS V2
UNFOLDED PROTEIN RESPONSE
CHOLESTEROL HOMEOSTASIS
REACTIVE OXYGEN SPECIES PATHWAY
PROTEIN SECRETION
NOTCH SIGNALING
ANGIOGENESIS
HEDGEHOG SIGNALING
APICAL SURFACE
PANCREAS BETA CELLS
OXIDATIVE PHOSPHORYLATION
BILE ACID METABOLISM
FATTY ACID METABOLISM
PEROXISOME
APOPTOSIS
TNFA SIGNALING VIA NFKB
COMPLEMENT
APICAL JUNCTION
PI3K AKT MTOR SIGNALING
KRAS SIGNALING UP
P53 PATHWAY
IL2 STAT5 SIGNALING
UV RESPONSE DN
IL6 JAK STAT3 SIGNALING
INFLAMMATORY RESPONSE
MITOTIC SPINDLE
G2M CHECKPOINT
DNA REPAIR
E2F TARGETS
MYC TARGETS V1
MTORC1 SIGNALING
UV RESPONSE UP
cross-talk
1000
500
0
f_c
-3.29
0.00
5.61


B)
ACADM
ABCA1
ACADL
C5AR1
ACLY
ACOX1
ALCAM
ADBB2
SHROOM2
CREBBP
RHOA
MAPK1
AKT1
RELA
SRC
RAC1
CDK1
UBC
PIK3CA
COX6A2
LYAR
NDUFV3
RBM14
DMAC2L
n_f
n_d
0
10
20
30
40
50
0
50
100
150
200


A) Mediator genes
f
-4.20
0.00
2.58
PTPN6
ITGB1
EGFR
TP53
FYN
n_f
n_d
B) 1 cell type
f
-1.58
0.00
2.32
MSRB2
TCN1
GPC2
PRKAA2
FGFR2
WNT11
BMP7
ERBB3
WNT7B
PTPRG
SCN2A
n_f
n_d
C) all cell types
f
-4.20
0.00
2.58
ITGB1
FYN
PTPN11
TP53
ECE1
GNAQ
CLEC2B
BAG6
HSP90AA1
n_f
n_d