

A Framework for Analyzing Omics Data using Routes of Biological Pathways

Doctoral Dissertation Defense

Ph.D. Candidate: Pujan Joshi

Major Advisor: Dong-Guk Shin

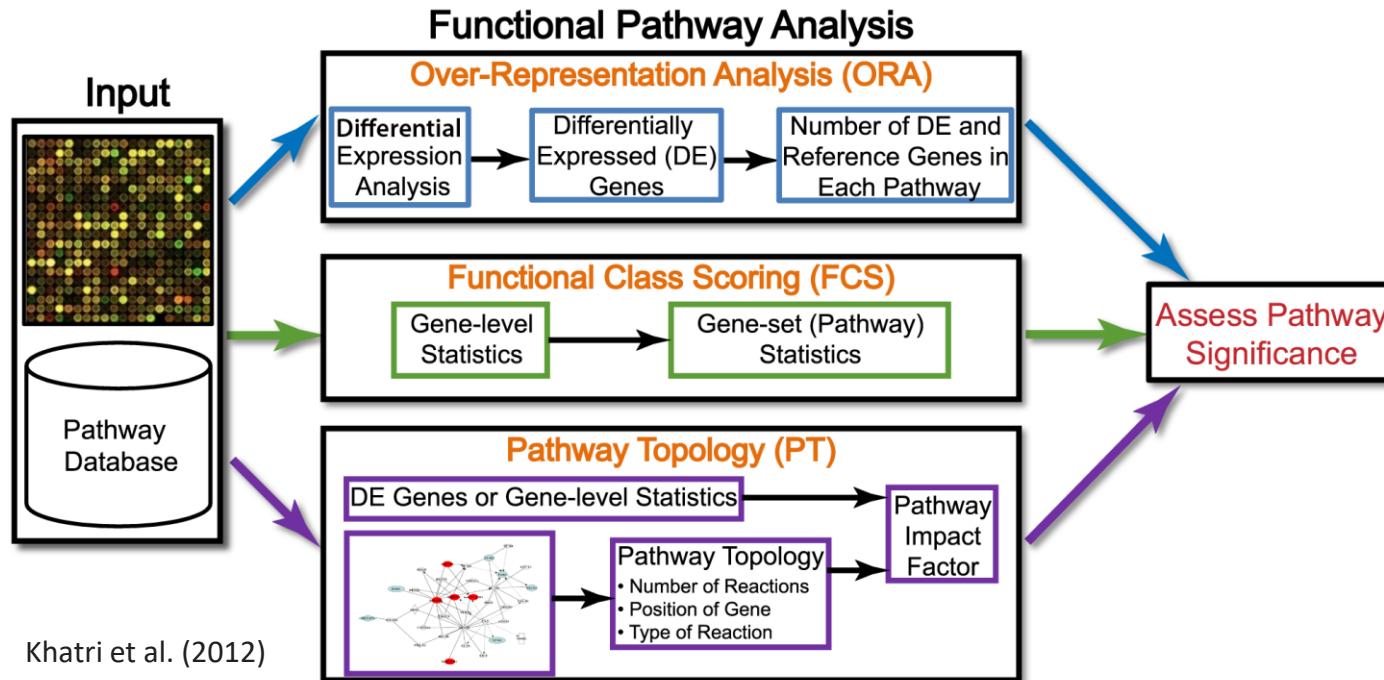
Associate Advisors: Charles Giardina, Sheida Nabavi

Review Committee Members: Ion Mandoiu, Derek Aguiar

Date/Time: Thursday, April 21, 2022, 11:00 AM

Department of Computer Science and Engineering, University of Connecticut, Storrs, CT 06269

Background



ORA – counts number of DE genes and computes enrichment statistics. E.g., GOSTAT

FCS – uses gene level statistics or log₂ fold change to find enrichment. E.g., GSEA

PT – uses relations and positions of genes that compose the pathway. E.g., SPIA, PARADIGM.

Motivation

- Scoring whole pathway limits our understanding on functions triggered by sub-pathways.
- Sub-pathways may trigger different or even opposite cell behavior. (e.g., Apoptosis and Cell Survival are both included in KEGG Apoptosis pathway)
- **Precise pathway routes need to be assessed and scored to understand cellular signaling mechanism more accurately.**

Organization of the talk

- Part 1: Route based Pathway Analysis of Cohorts (rPAC)
- Part 2: Route based Pathway Analysis of Crosstalk (rPAX)
- Part 3: Route based analysis of Higher Order Pathways (rHOP)
- Part 4: Route based pathway analysis of Single Cell Data

Part 1: Route based Pathway Analysis of Cohort (rPAC)

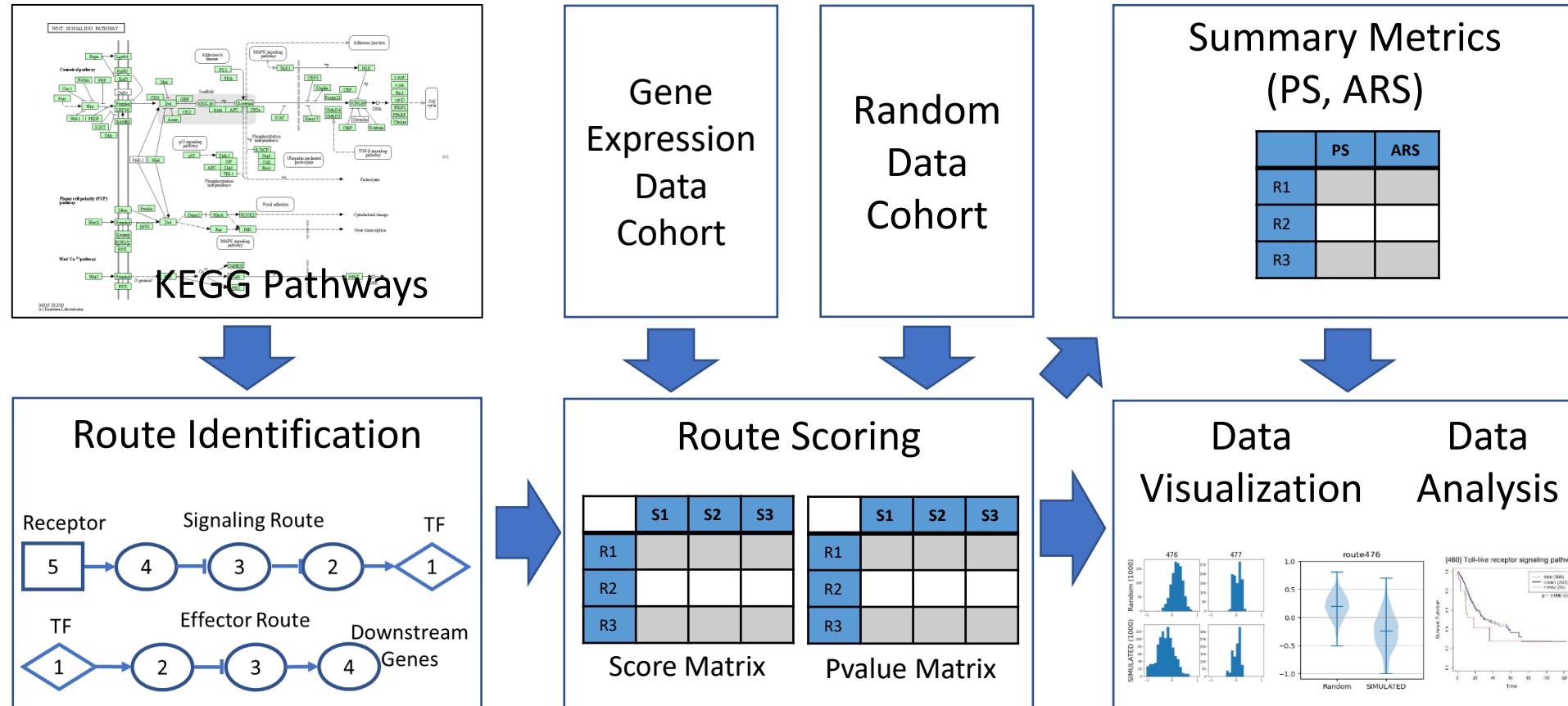
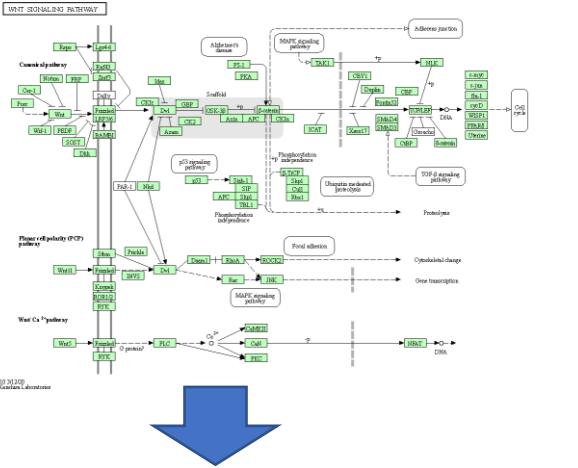


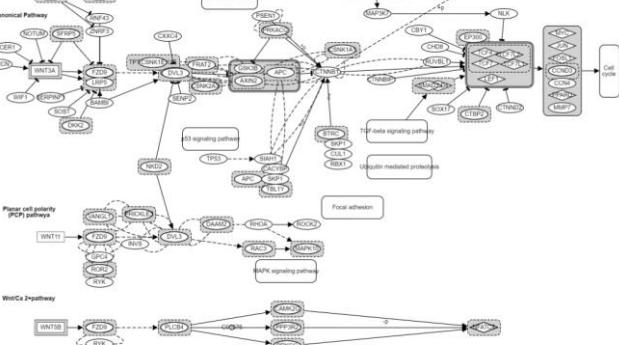
Figure: Overview of rPAC Framework

KEGG Pathways and Cohort Data

KEGG
KGML Format



JSON
Format



KEGG signaling pathways are downloaded in KGML format and converted into JSON format for subsequent processing.

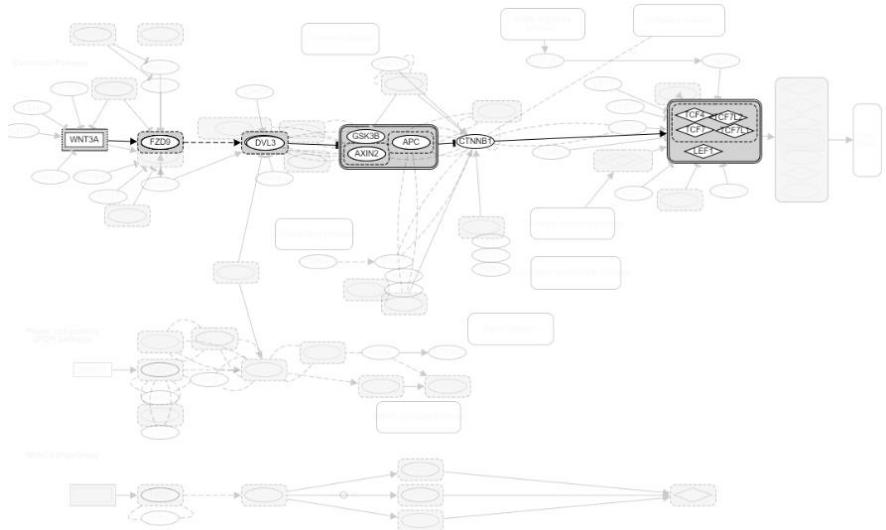
Gene Expression Data Matrix

	Sample 1	Sample 2	Sample 3	...	Sample N
Gene1	R ₁₁	R ₁₂	R ₁₃		R _{1N}
Gene2	R ₂₁	R ₂₂	R ₂₃		R _{2N}
...					
GeneM	R _{M1}	R _{M2}	R _{M3}		R _{MN}

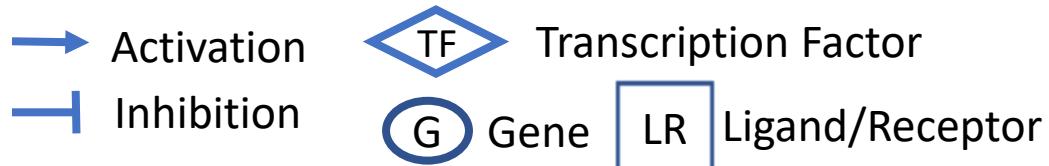
$$R_{gj} = \log_2 \left(\frac{E_{gt} + 1}{E_{gc} + 1} \right)$$

where E_{gt} is expression value of gene g in test sample and E_{gc} is the average expression value of the gene in control population.

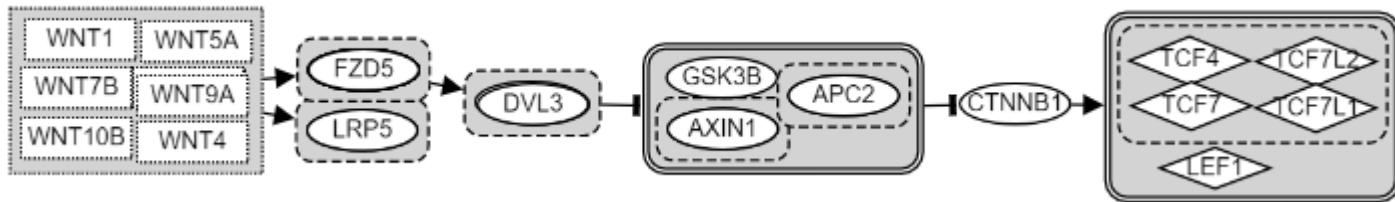
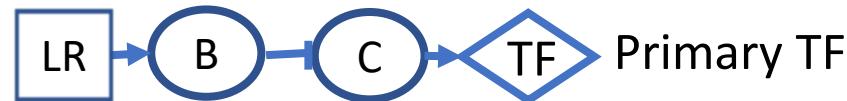
Route Identification



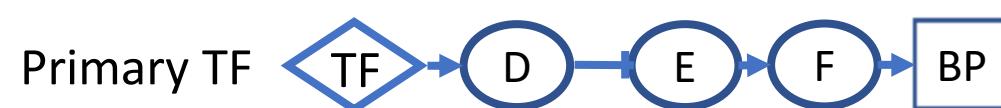
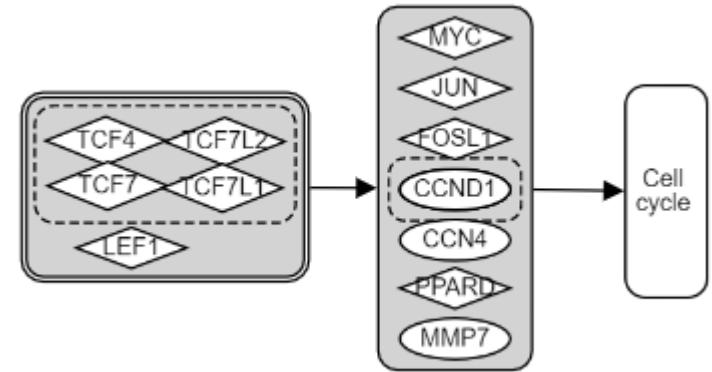
Depth First Search (DFS) to identify **TF-centric routes** that are **mediated** by Transcription Factor (TF). The mediator TF (or TF complex) is called **primary TF**.



Signaling route captures signal transduction from cell surface into the nucleus of the cell. **Ligand/Receptor (LR)** to **Transcription Factor (TF)**.

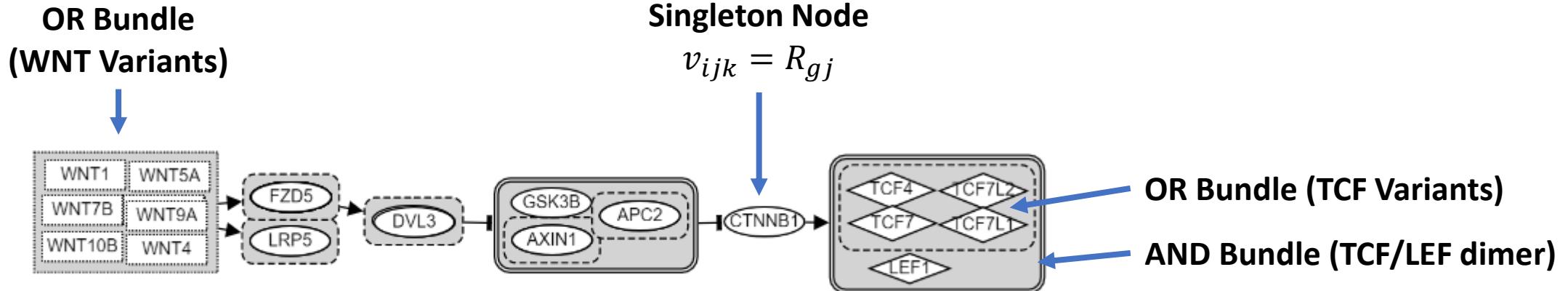


Effector route captures second part of the cellular process which is **transcription factor binding** to regulate target gene expression. **Transcription Factor (TF)** to **Biological Process (BP)**.



Node Evaluation

Node Value (v_{ijk}) represents value for node k in route i for sample j .



Bundle Node Evaluation

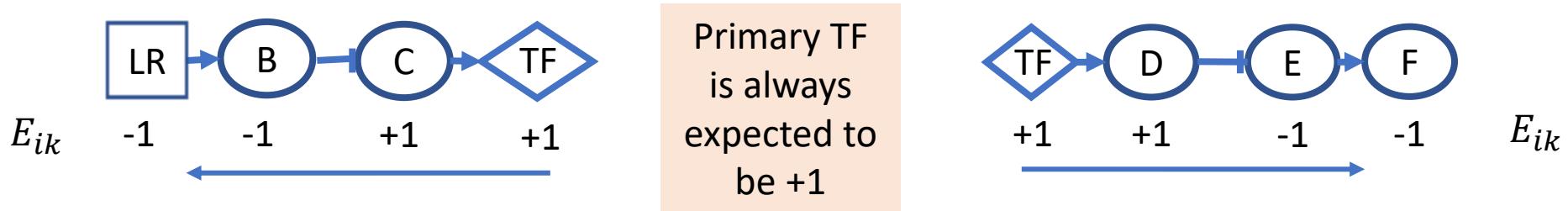
$$v_{ijk} = \begin{cases} \frac{\sum_{m=1}^M v_{mjk}}{M} * U_{ijk} & U_{ijk} \geq U_{min} \\ \frac{\sum_{n=1}^N v_{njk}}{N} * (1 - U_{ijk}) & U_{ijk} < U_{min} \end{cases}$$

- (i) M and N are the numbers of up and down regulated nodes respectively in the bundle,
- (ii) v_{mjk} and v_{njk} are values of up-regulated and down-regulated nodes in the bundle,
- (iii) $U_{ijk} = \frac{M}{n_k}$, proportion of up-regulation
- (iv) n_k is the total number of nodes in bundle node k .

U_{min} is a hyper-parameter and represents minimum proportion of up-regulation required to assign activation/inhibition to a bundle. Empirically identified to be 0.5 for AND bundle and 0.2 for OR bundle.

Route Activity Score

Expected value (E_{ik}) is assigned to every node k in route i based on the relation of the node with primary TF (i.e., +1 if relationship between node k and TF is positive, and -1 if relationship is negative)



Contribution (C_{ijk}) of node k in route i for sample j is computed as:

$$C_{ijk} = \begin{cases} C_{max} & v_{ijk} > C_{max} \\ v_{ijk} & -C_{max} \leq v_{ijk} \leq C_{max} \\ -C_{max} & v_{ijk} < -C_{max} \end{cases}$$

where C_{max} (default value 1) is the maximum allowed contribution of a node.

Activity score (S_{ij}) for each sample j in given cohort is computed as:

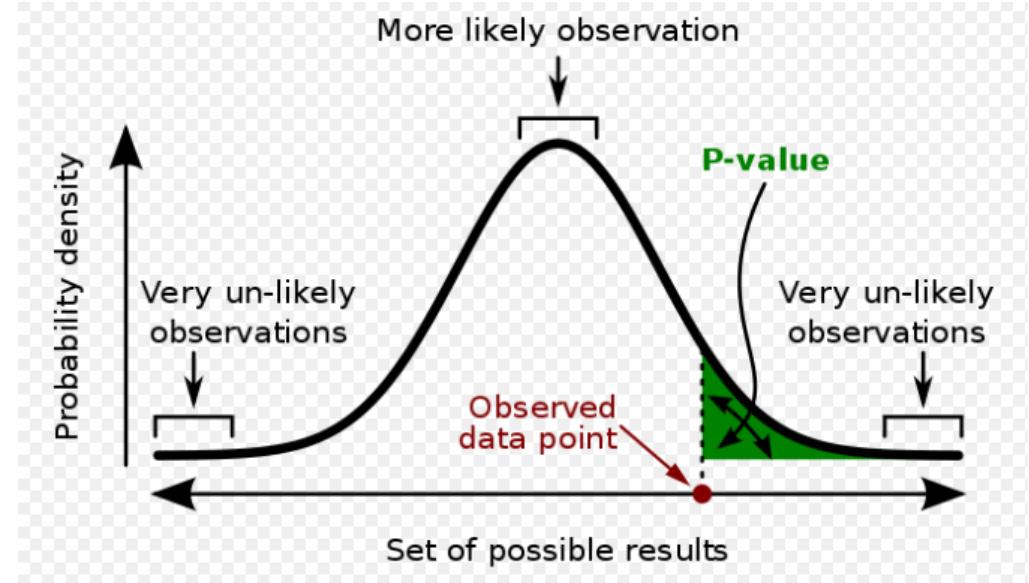
$$S_{ij} = \frac{1}{n_i} * \sum_{k=1}^{n_i} C_{ijk} * E_{ik} \quad \text{where } n_i \text{ is number of nodes in route } i$$

S_{ij} ranges from -1 (highly suppressed) to +1 (highly activated)

P-value and Summary Metrics (PS & ARS)

- 1000 random samples (gene values from a standard normal distribution). Route activity score, S_{ir}^* was computed for each route i and random sample r . (NULL Distribution)
- A p-value (shaded green area) is the probability of an observed (or more extreme) result in the NULL distribution.

$$p_{ij} = 2 * \min[\Pr(S_{ir}^* \geq S_{ij} | H_0), \Pr(S_{ir}^* \leq S_{ij} | H_0)]$$



Summary metrics for each route i :

Proportion of Significance $PS_i = \frac{1}{J} * \sum_{j=1}^J Z(i,j)$ where $Z(i,j) = \begin{cases} 1 & P_{ij} < P_t \\ 0 & otherwise \end{cases}$

Average Route Score $ARS_i = \frac{1}{J} * \sum_{j=1}^J S_{ij}$ where J is the number of samples in given cohort and P_t is p-value threshold (default 0.05).

Results and Discussion

- Case study on large amount of simulated data.
- Case study on three epithelial cancer cohorts from The Cancer Genome Atlas (TCGA)
 - Breast Invasive Carcinoma (BRCA)
 - Colon Adenocarcinoma (COAD)
 - Stomach Adenocarcinoma (STAD).
- 1695 total routes were identified from 70 KEGG signaling pathways. Route scores, p-values, ARS & PS were computed using rPAC framework.

Case study with simulated data

Log2 Ratio value R_{ik} for each gene k in route i was simulated using a multivariate normal distribution

$$R_{ik} = E_{ik} * g + t$$

where,

- E_{ik} is the expectation for a gene, +1, 0 or -1.
- g represents magnitude of differential expression of a gene and comes from a normal distribution with mean μ_g and standard deviation σ_g .
- t represents random variation and is sampled from standard normal distribution.

Large number of simulated data cohorts (size 500 each) were generated with varying μ_g and σ_g for 35 routes from 13 unconnected KEGG pathways

Power ($1 - \beta$) is the ratio of perturbed pathways successfully identified as significant (FDR qVal < 0.05 for GSEA, $p < 0.05$ for rPAC and $pG < 0.05$ for SPIA) over the total perturbed pathways

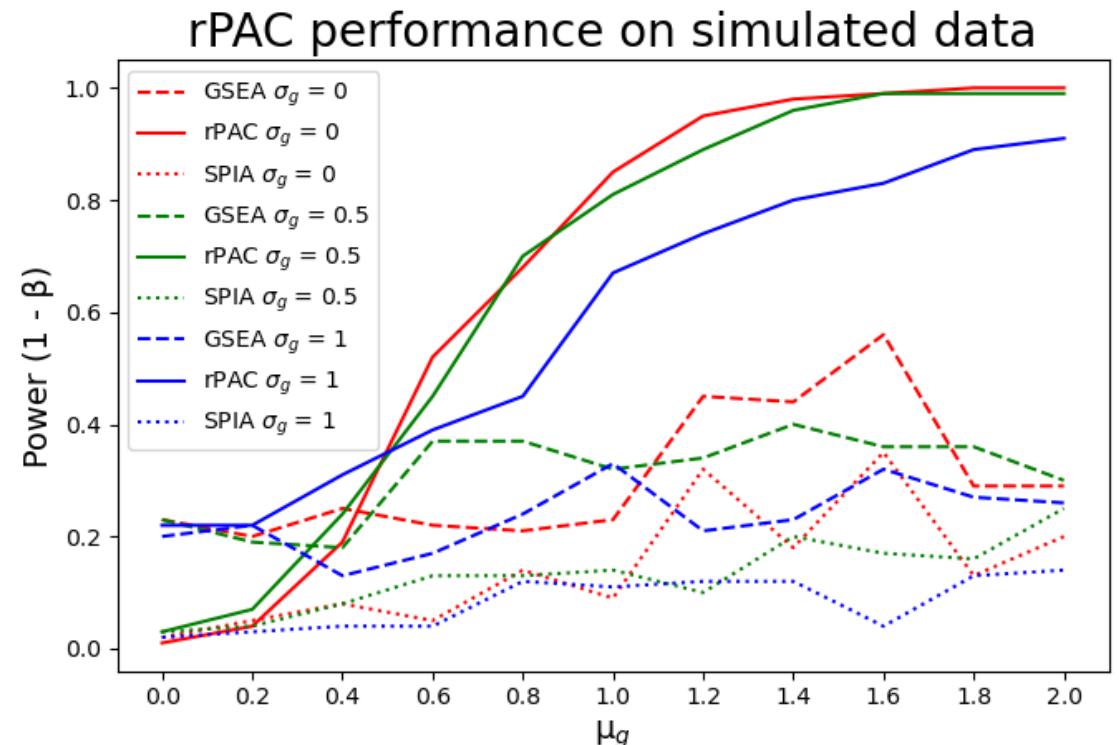


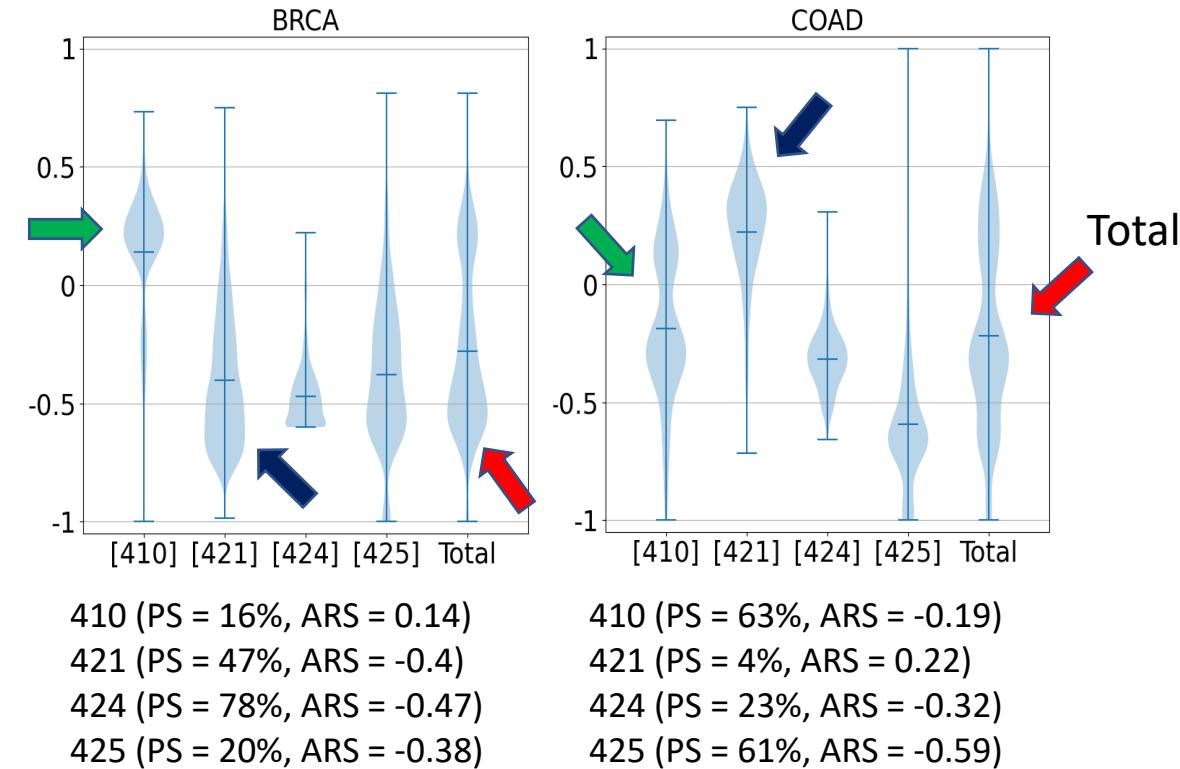
Figure: Power curve on simulated data at various differential gene expression level

Route activity variation within pathways

cellular activity detection at a resolution beyond the pathway level

ID	Route (p1 - Signaling route, p2 - Effector Route)	BRCA		COAD	
		PS	ARS	PS	ARS
410	Thyroid hormone signaling pathway(p2) PIK3CA, PIK3CB, PIK3CD, PIK3R1, PIK3R2, PIK3R3, ATP1B4, ATP1A1, ATP1A2, ATP1A3, ATP1A4, ATP1B1, ATP1B2, ATP1B3, FXYD2	0.16	0.14	0.63	-0.19
421	Thyroid hormone signaling pathway(p2) THRA, NOTCH4, NOTCH3, NOTCH2, NOTCH1, BMP4	0.47	-0.40	0.04	0.22
424	Thyroid hormone signaling pathway(p2) THRA, MYH6, MYH7, ATP2A2, ATP2A1, ATP2A3, PLN	0.78	-0.47	0.23	-0.32
425	Thyroid hormone signaling pathway(p2) PIK3CA, PIK3CB, PIK3CD, PIK3R1, PIK3R2, PIK3R3, RCAN2	0.20	-0.38	0.61	-0.59

Routes in Thyroid hormone signaling pathway



It is difficult to make meaningful assessments of activity without considering route-level variability.

Comparison between cancers

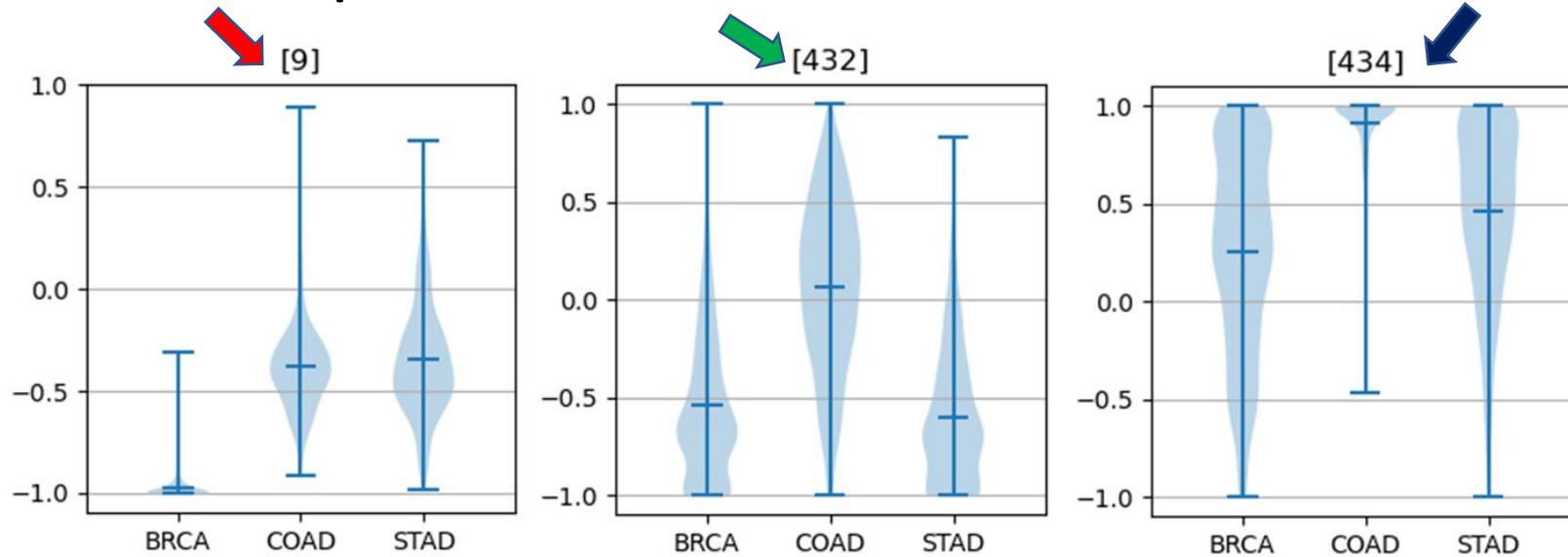


Figure: Route 9 (Adipocytokine signaling) and Routes 432 & 434 (Tight junction)

Route 9 leads from the activation of the leptin receptor to the activation of the PPARA/RXR dimer. PPARA has demonstrable anti-proliferative effects in human and rodent models [Tyagi et al. (2011)].

Route 432 leads from c-Jun to the transcription of the CD1 family of antigen presenting proteins and has been associated with breast cancer [Coventry et al. (2002)].

Route 434 leads to the transcription of CCND1, a cell cycle regulator involved in colon cancer progression [Qie et al. (2016)].

Subtype analysis on BRCA

Study of five subtypes of BRCA: luminal A, luminal B, HER2+, basal, and normal-like.

Route 35 leads to the expression of perilipin 1 (PLIN1), a lipolysis inhibitor. PLIN1's downregulation allows for greater-than-normal lipolytic activity. Lipolysis and lipogenesis are common in cancer cells, as they both support the fatty acid demands of rapid membrane synthesis [Zaidi et al. (2013)]

Th1 and Th2 cell differentiation pathway **route 377** connects STAT5 to GATA3, leading to Th2 differentiation which can support tumor growth by inhibiting a patient's cytotoxic T cell response to cancer cells [Zhao et al. (2019)], and basal breast cancers tend to have worse patient outcomes than luminal A or normal-like cancers [Dai et al. (2015)].

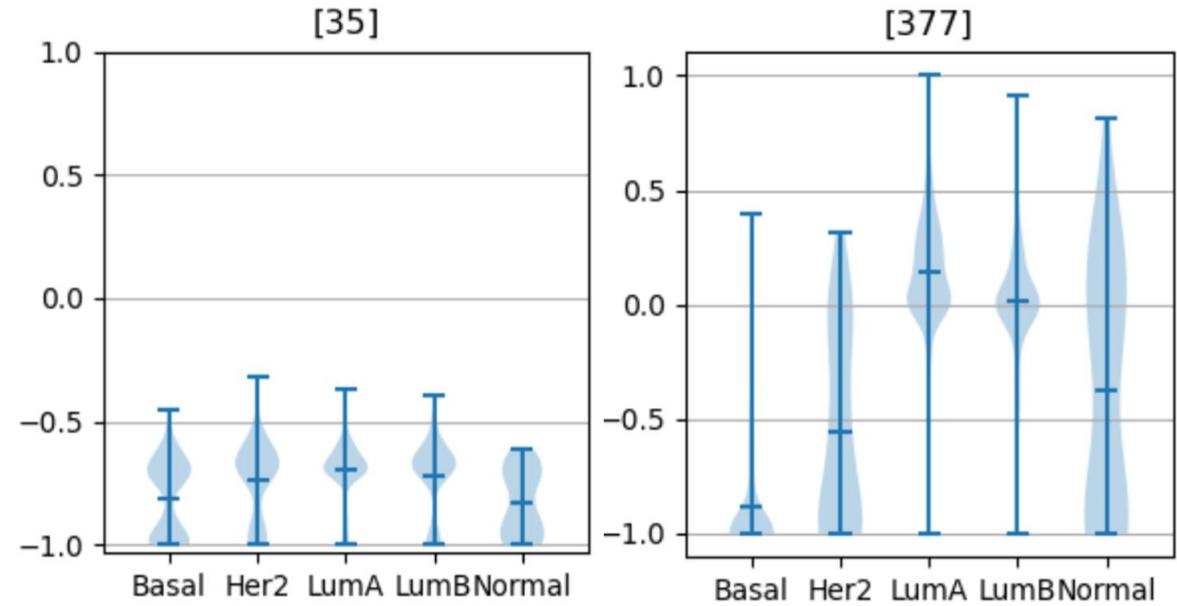


Figure: Violin plots for route 35 (Apelin signaling pathway) and route 377 (Th1 and Th2 cell differentiation) for BRCA subtypes. Note that route 35 is universally suppressed in BRCA and route 377 is a prominent differentiator of subtypes.

Kaplan-Meier (K-M) Survival Analysis

Route 258 (Fig. b) is part of a feedback route which downregulates p53, well-documented prognostic indicator in cancers [Robles and Harris (2009)].

Apelin route 46 (Fig. c), facilitates EMT by regulation of E-cadherin which is associated with both progression and metastasis in breast cancers [Padmanaban et al. (2019)].

JAK/STAT route 190 (Fig. d) begins with cytokine activation of the JAK/STAT and leads to the dimerization of STAT family transcription factors. Suppression of a cytokine dependent feedback loop in this pathway is associated with breast cancer [Ghafouri-Fard et al. (2018)].

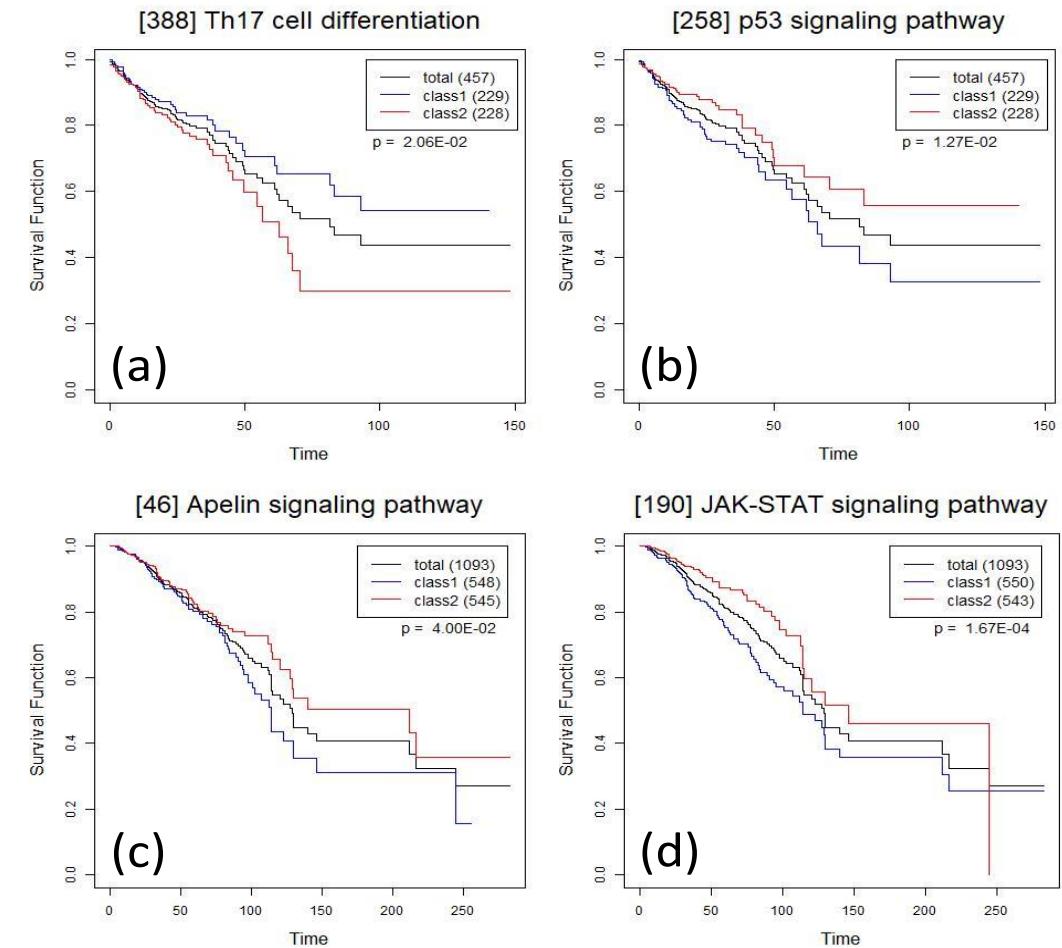


Figure: (a, b) route 388 & 258 (COAD) (c, d) route 46 & 190 (BRCA). **Class1** – samples with score lower or at median. **Class2** – samples with score higher than the median.

Related Work

GSEA and SPIA were run with all three cancer cohorts on KEGG. PS scores were calculated using FDR q-val < 0.05 for GSEA and $P_G < 0.05$ for SPIA.

Table: PS scores from GSEA, rPAC, and SPIA (selected routes). The rPAC framework generates scores at route-level resolution distinguishing roles of various routes within a pathway. GSEA and SPIA assign scores for entire pathway with no route-level distinction.

Route 9 is BRCA Differentiator

Route 10 has pan-cancer role
[Vansau et al., 2013]

APC Mutation,
WNT targets
high in COAD

Route 432 is associated in BRCA

[Coventry et al., 2012]

Route 434 is associated with COAD

[Qie et al., 2016]

ID	Route (p1 - Signaling, p2 - Effector)	GSEA (FDRq < 0.05)			rPAC (p < 0.05)			SPIA ($P_G < 0.05$)		
		BRCA	COAD	STAD	BRCA	COAD	STAD	BRCA	COAD	STAD
9	Adipocytokine signaling pathway(p1) LEP, LEPR, PPARA, RXRA, RXRB, RXRG	0.62	0.41	0.41	0.99	0.10	0.10	0.14	0.13	0.01
10	Adipocytokine signaling pathway(p1) ADIPOQ, ..., RXRB, RXRG	0.62	0.41	0.41	0.35	0.35	0.23	0.14	0.13	0.01
281	PPAR signaling pathway(p2) PPARA, RXRA, RXRB, ..., PLIN5, PLIN2	0.83	0.90	0.93	0.86	0.82	0.64	0.01	0.04	0.01
299	PPAR signaling pathway(p2) PPARG, RXRA, ..., ME1, ME3	0.83	0.90	0.93	0.69	0.00	0.11	0.01	0.04	0.01
432	Tight junction(p2) JUN, CD1A, CD1B, CD1C, CD1D, CD1E	0.49	0.76	0.18	0.69	0.10	0.63	0.01	0.01	0.02
434	Tight junction(p2) YBX3, CDK4, SYMPK, CCND1	0.49	0.76	0.18	0.08	0.75	0.17	0.01	0.01	0.02
471	Wnt signaling pathway(p1)	0.72	0.12	0.32	0.14	0.04	0.07	0.02	0.01	0.06
472	Wnt signaling pathway(p2)	0.72	0.12	0.32	0.12	0.61	0.13	0.02	0.01	0.06

Part 1 (rPAC) - Conclusion

- Scoring whole pathway limits our understanding of functions triggered by sub-pathways.
- Here we describe a computational scoring system designed to assess signaling through individual routes within a pathway.
- Using this approach on cancer data sets, we identified discrete routes that can distinguish various cancer cohorts and subgroups.

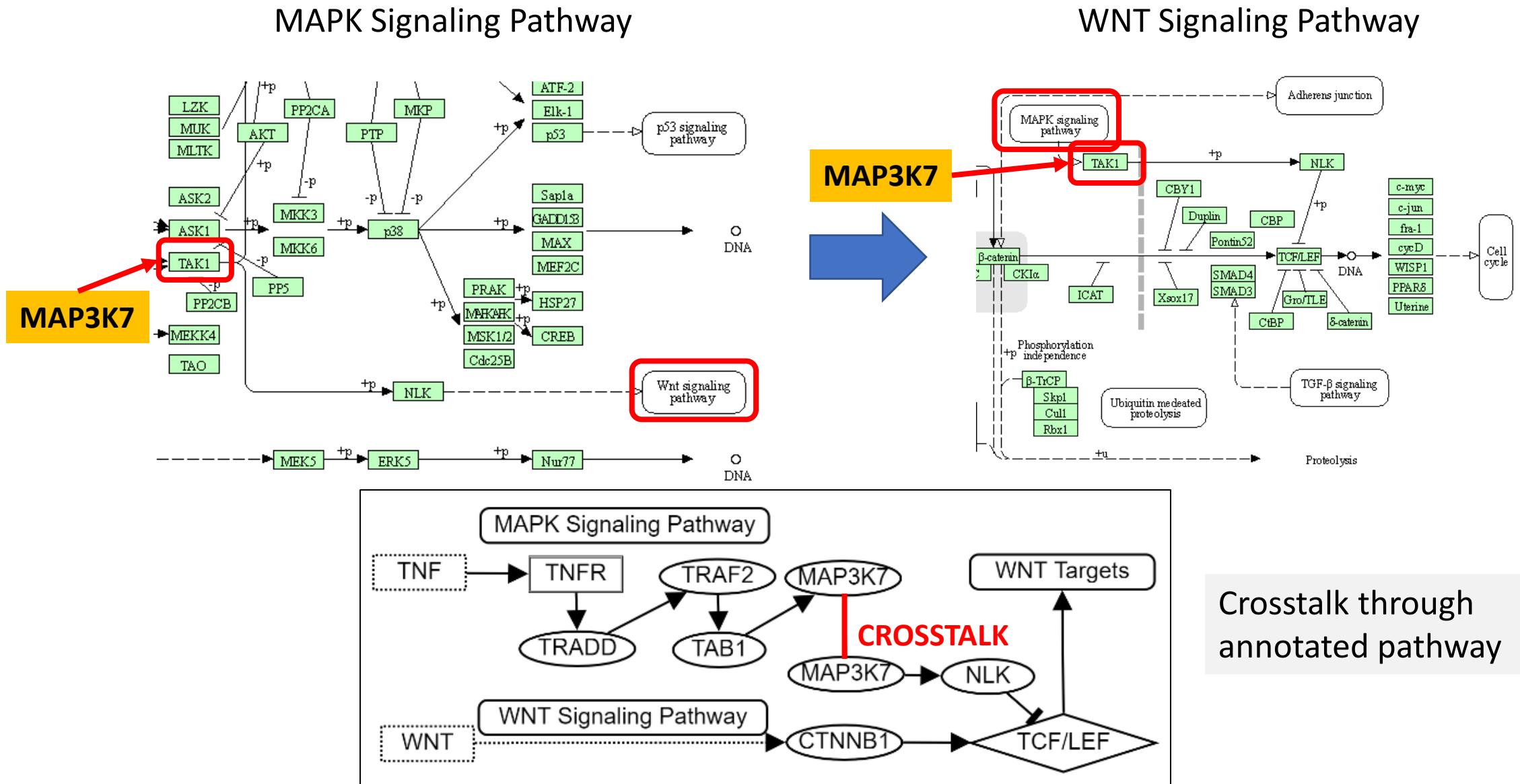
Organization of the talk

- Part 1: Route based Pathway Analysis of Cohorts (rPAC)
- Part 2: Route based Pathway Analysis of Crosstalk (rPAX)
- Part 3: Route based analysis of Higher Order Pathways (rHOP)
- Part 4: Route based pathway analysis of Single Cell Data

Part 2: Route based Pathway Analysis of Crosstalk (rPAX) - INTRODUCTION

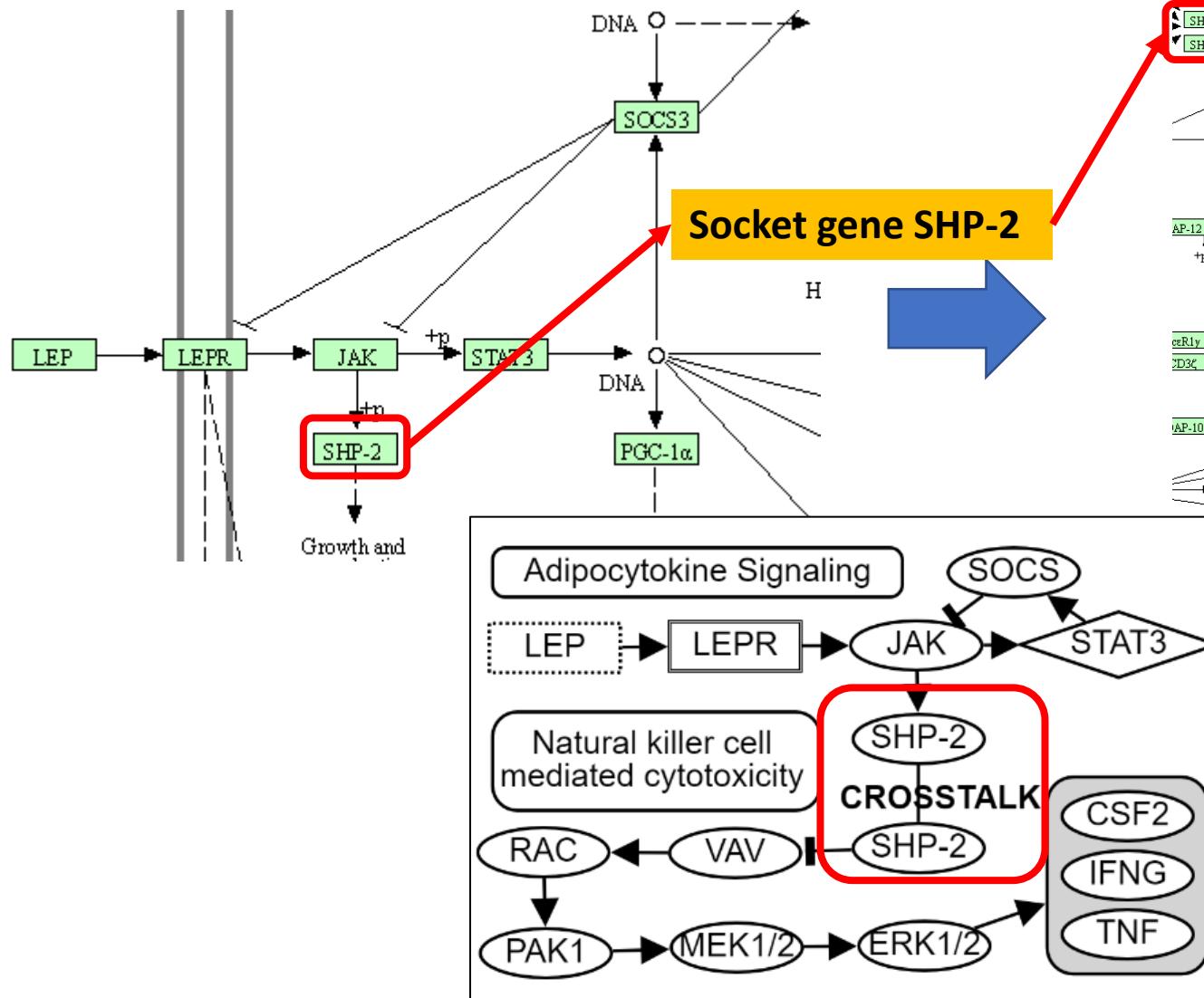
- Although analyzing **intra-pathway routes** in **rPAC** was informative, pathways in the cell do not work independently.
- Signal transduction pathways can affect each other through **crosstalk** between one or more shared components.
- Upon activation of a receptor in one pathway, signaling cascade can result in the activation of transcription factors (TFs) that activate genes in other pathways.
- Crosstalk in signaling pathways is well-studied and has been implicated in several cancers.

Crosstalk Route Identification

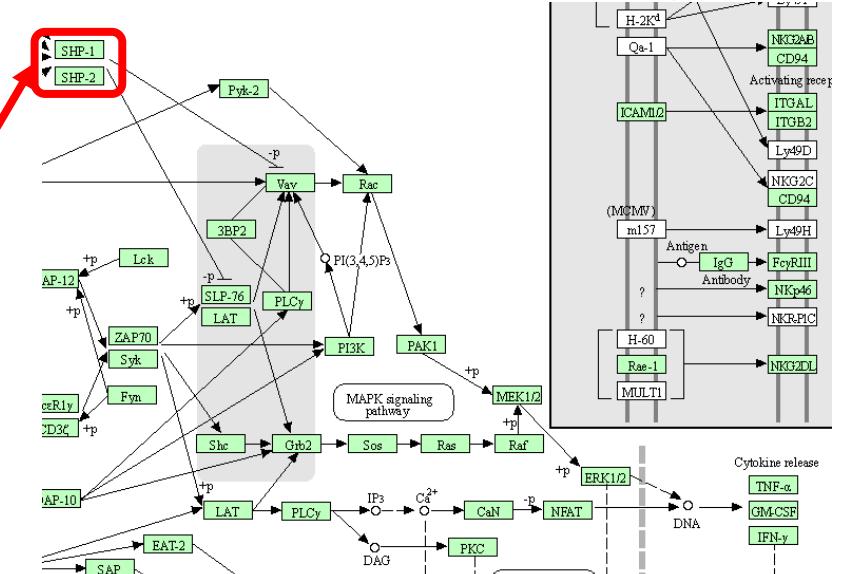


Crosstalk Route Identification

Adipocytokine Signaling Pathway



Natural Killer Cell Mediated Cytotoxicity



Crosstalk through common nodes (genes)

Route based Pathway Analysis of Crosstalk (rPAX)

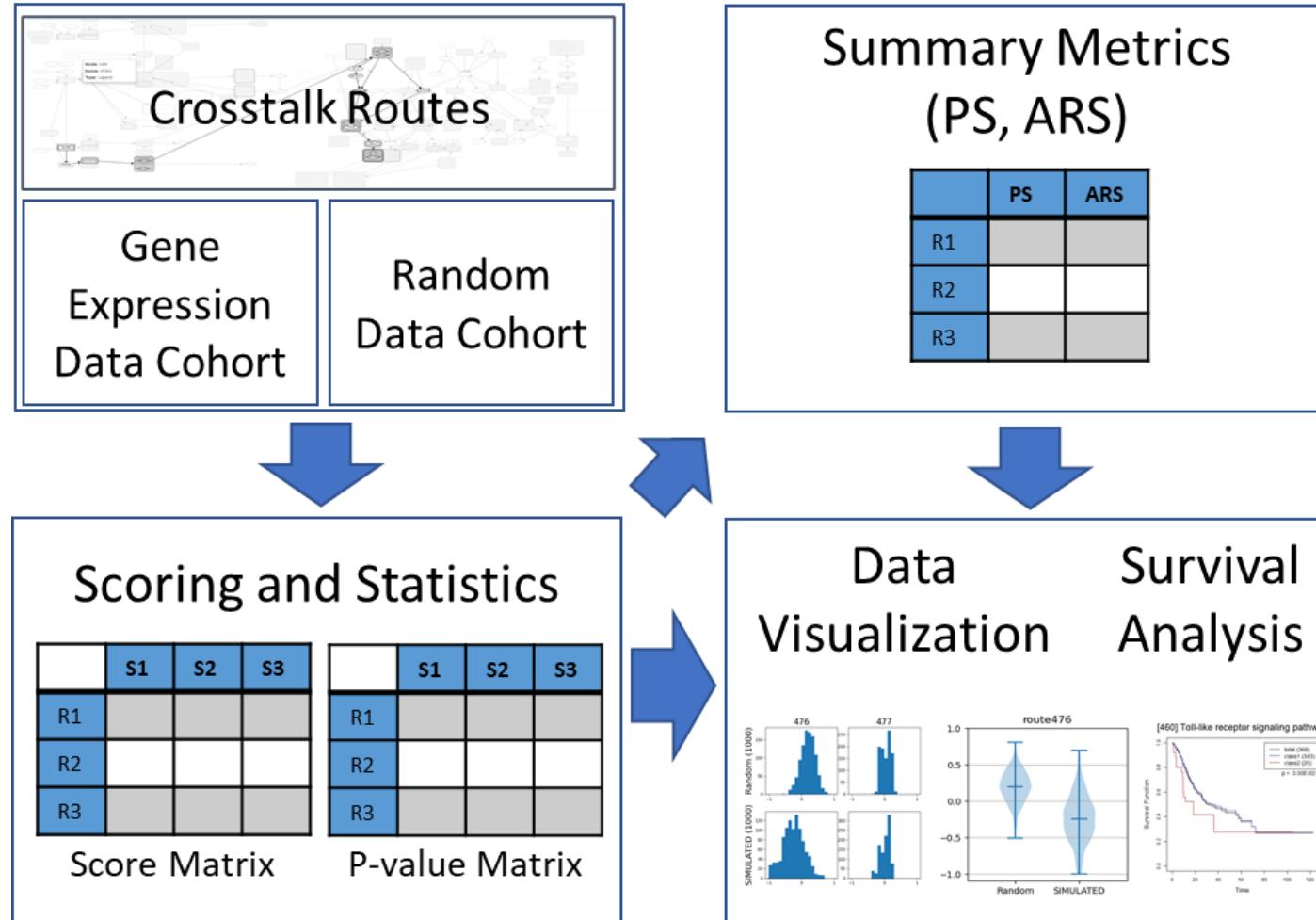


Figure: Overview of rPAX framework.

Highlights

- Crosstalk (**X**) Signaling Route (XSR) originates from ligand/receptor of one pathway and ends at TF of another pathway.
- Crosstalk (**X**) Effector Route (XER) originate from TF of one pathway and jumps to ligand of another pathway.
- Now that we have introduced large number of routes, Benjamini-Hochberg Procedure (BH procedure) is used to control FDR from multiple hypothesis testing.

$$FDR = \frac{pval * \#tests}{pval_rank}$$

Part 2 (rPAX) - Results and Discussion

- Case study on **five** cancer cohorts from TCGA.
 - Breast Invasive Carcinoma (BRCA - 1095 tumor and 127 normal tissue samples)
 - Colon Adenocarcinoma (COAD – 458 tumor and 63 normal tissue samples)
 - Stomach Adenocarcinoma (STAD – 372 tumor and 35 normal tissue samples)
 - Glioblastoma (GBM – 168 tumor and 6 normal tissue sample)
 - Skin Cutaneous Melanoma (SKCM – 443 tumor and 29 normal tissue samples)
- **15782 (15163 XSR, and 619 XER)** total cross routes were identified from 70 KEGG signaling pathways. Route scores, p-values, ARS & PS were computed.
- Survival Analysis was performed using K-M plots and p-value FDR were computed using Benjamini-Hochberg procedure to keep results with FDR < 10%.

Global signaling activity comparison of cancers

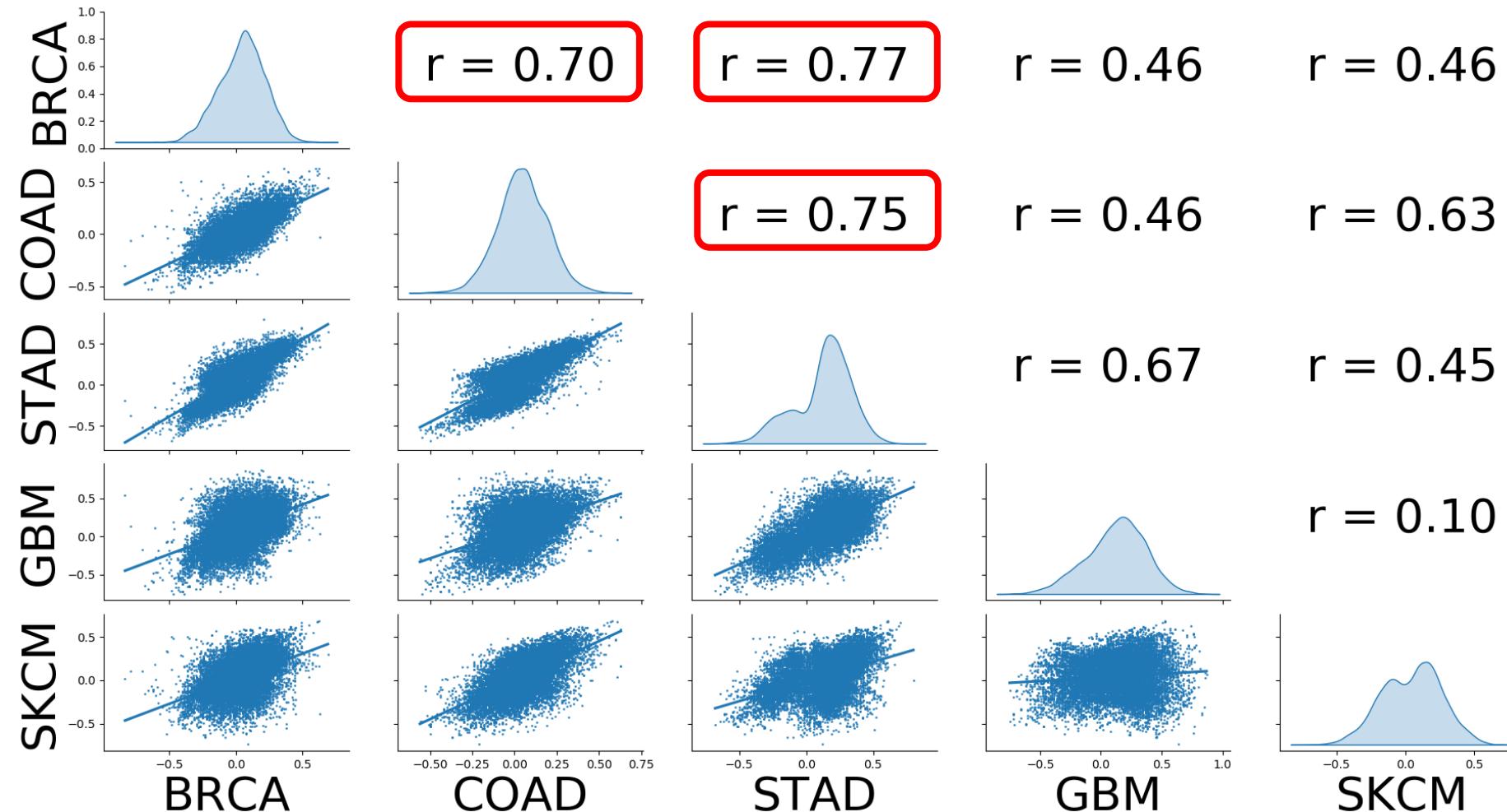


Figure: Scatter plot of ARS in cancer cohort pairs. Three epithelial cancers (BRCA, COAD, and STAD) are more correlated with each other than non-epithelial cancers (GBM and SKCM).

Comparison between cancers

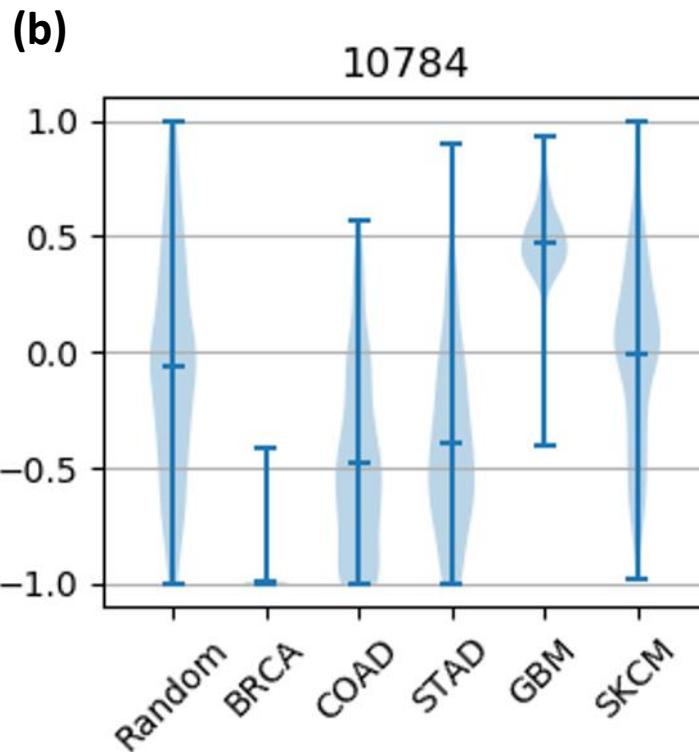
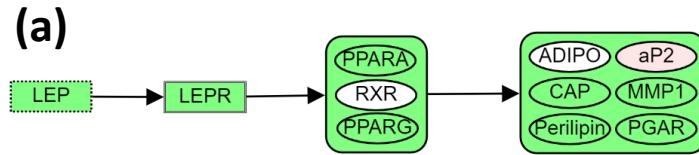


Figure: (a) Detail organization of route 10784. (b) Violin plot distribution of scores for route 10784 in different cancer cohorts.

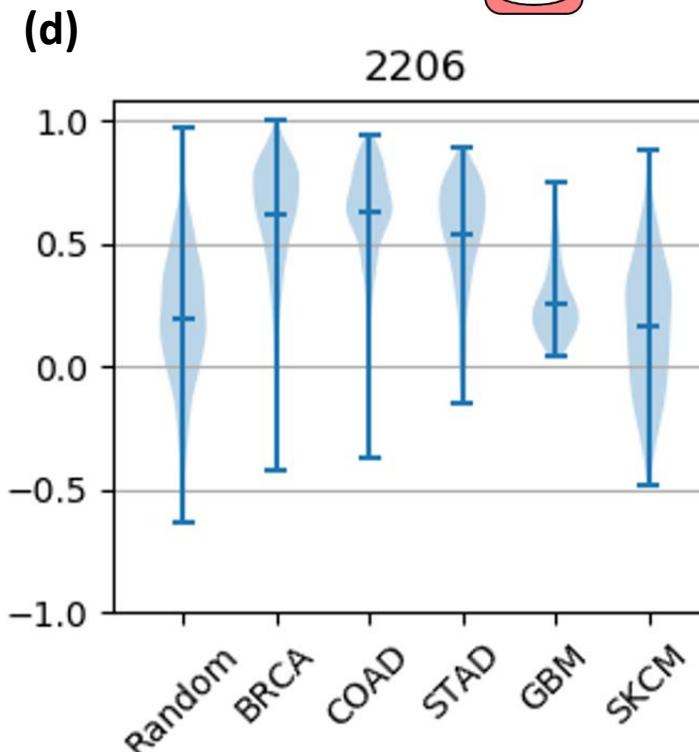
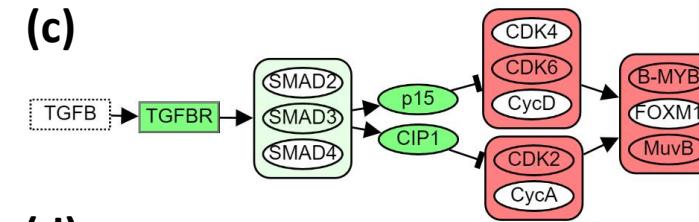
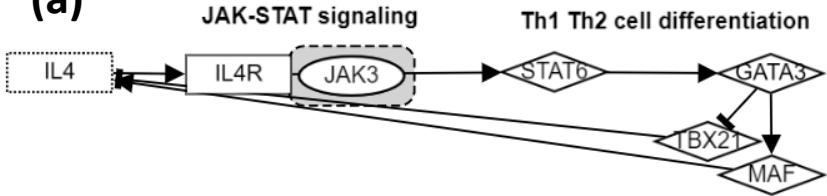


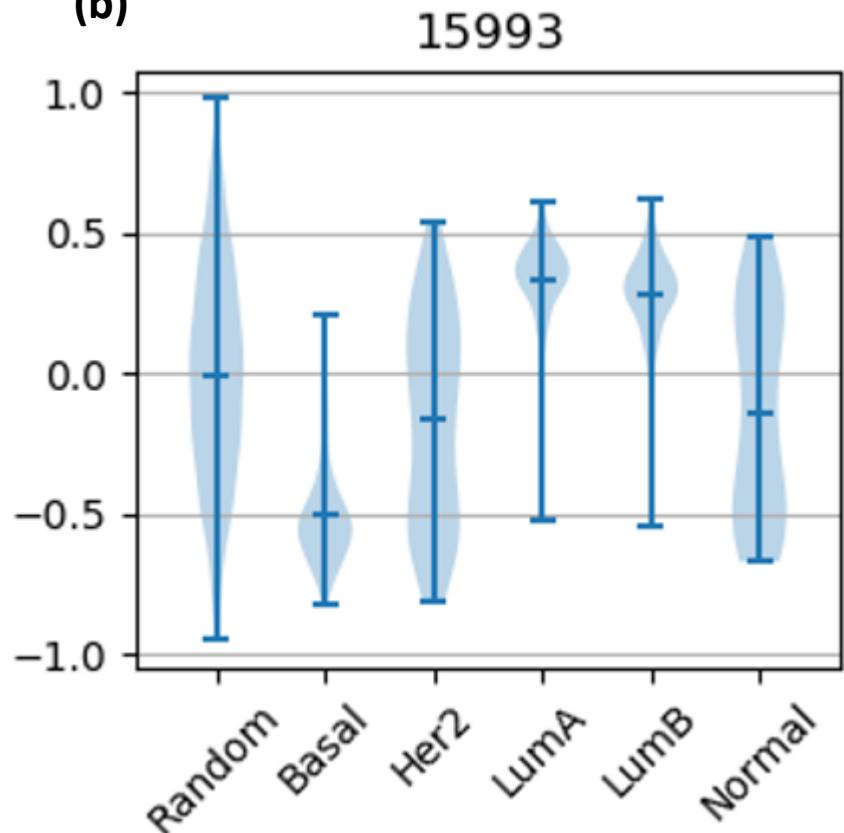
Figure: (a) Detail organization of route 2206. (b) Violin plot distribution of scores for route 2206 in different cancer cohorts.

Subtype analysis on BRCA

(a)



(b)

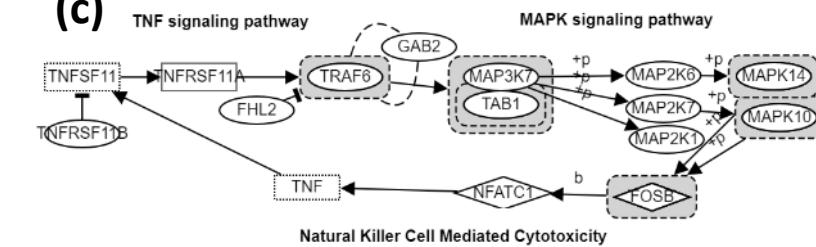


Route 15933 is the best predictor of breast cancer subtypes.

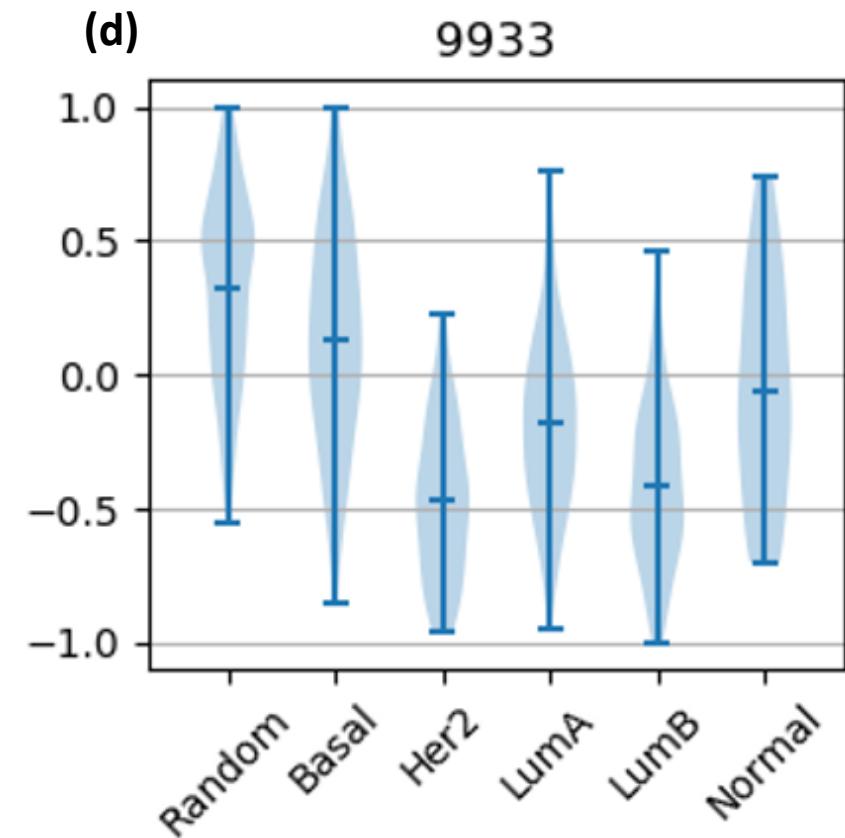
IL4 is a B cell promoting cytokine, suggests B cell and humoral immune response suppression in Basal like breast cancers. The humoral immune response is a strong prognostic indicator for breast cancer [Schmidt et al., 2008]

In Route 9933, since NFATc1 is a pivotal T cell TF, the suppression of this positive feedback loop may be important for suppressing the immune response in HER2-positive breast cancers.

(c)



(d)



Survival Analysis (BRCA Cohort)

Route 229

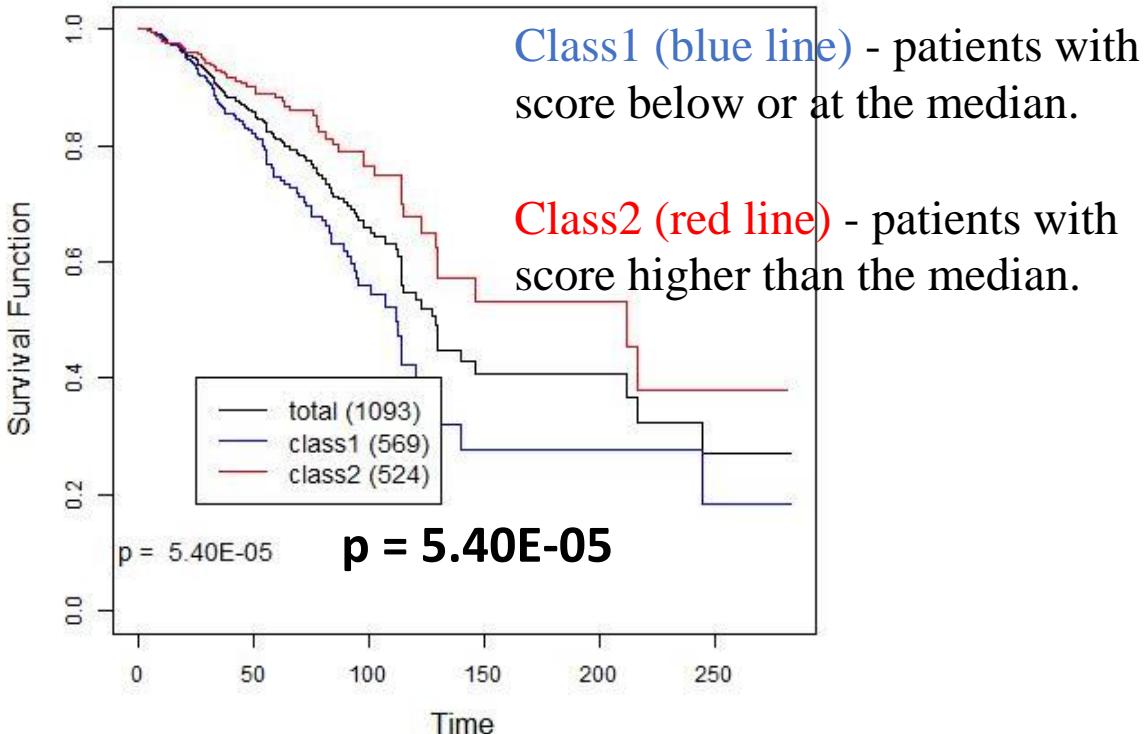


Figure: K-M plot for patients in BRCA cohort separated at the median scores of crosstalk route 229. Survival analysis was performed in R using survival package

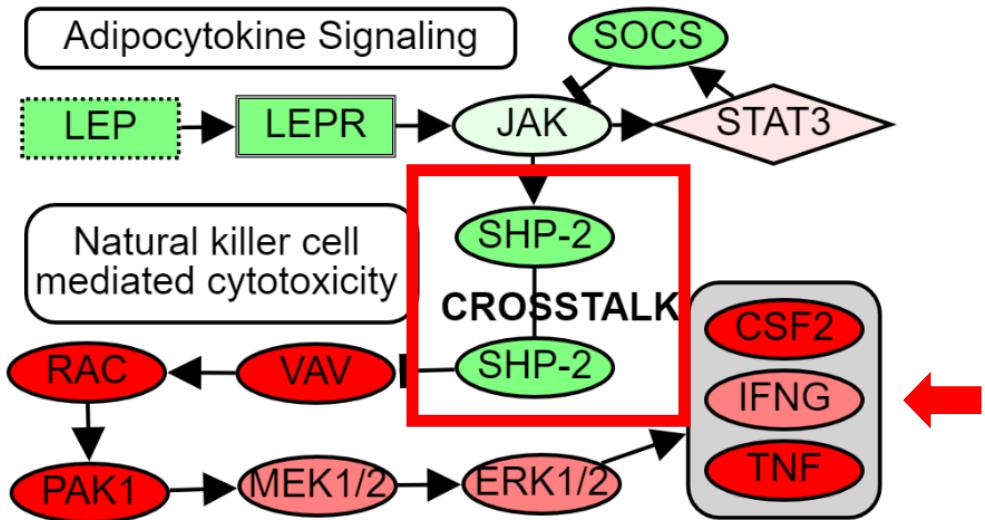


Figure: Cross route 229. Green background represents downregulation and red represents upregulation

IFNG stimulates the activity of natural killer cells and cytotoxic T lymphocytes, both of which can kill cancer cells (consistent with K-M plot observation).

SHP-2, socket gene in this crosstalk, is a current target for drug development efforts [Chen et al., 2020]

Part 2 (rPAX) Conclusion

- Pathways interact with each other via common components with a phenomenon called **crosstalk**.
- A computational framework is developed that can identify and score all possible crosstalk between pathway routes.
- Using this approach on cancer data sets, we identified discrete routes that can distinguish various cancer types and subtypes.
- K-M analysis revealed crosstalk routes critical for patient survival.

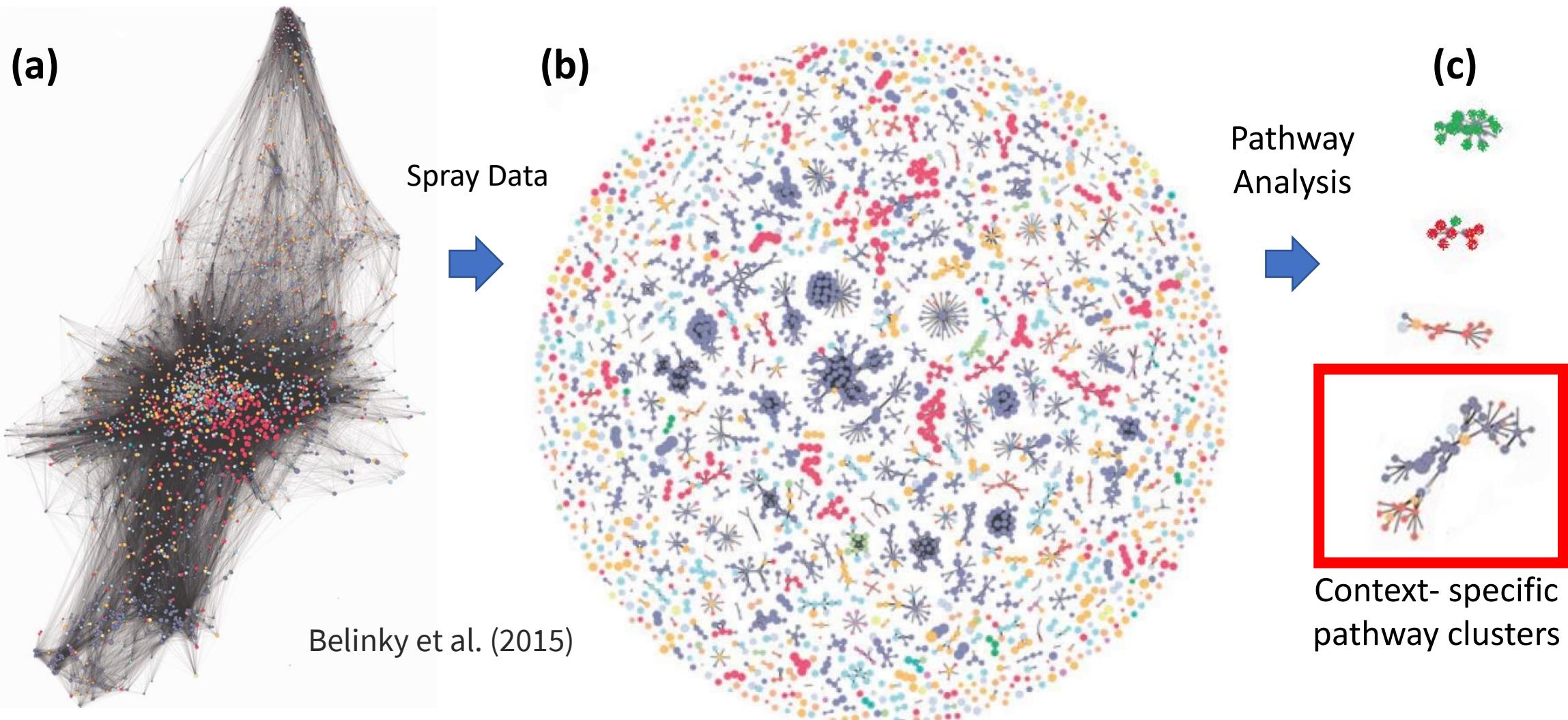
Organization of the talk

- Part 1: Route based Pathway Analysis of Cohorts (rPAC)
- Part 2: Route based Pathway Analysis of Crosstalk (rPAX)
- Part 3: Route based analysis of Higher Order Pathways (rHOP)
- Part 4: Route based pathway analysis of Single Cell Data

Part 3: Route based analysis of Higher Order Pathways (rHOP) - INTRODUCTION

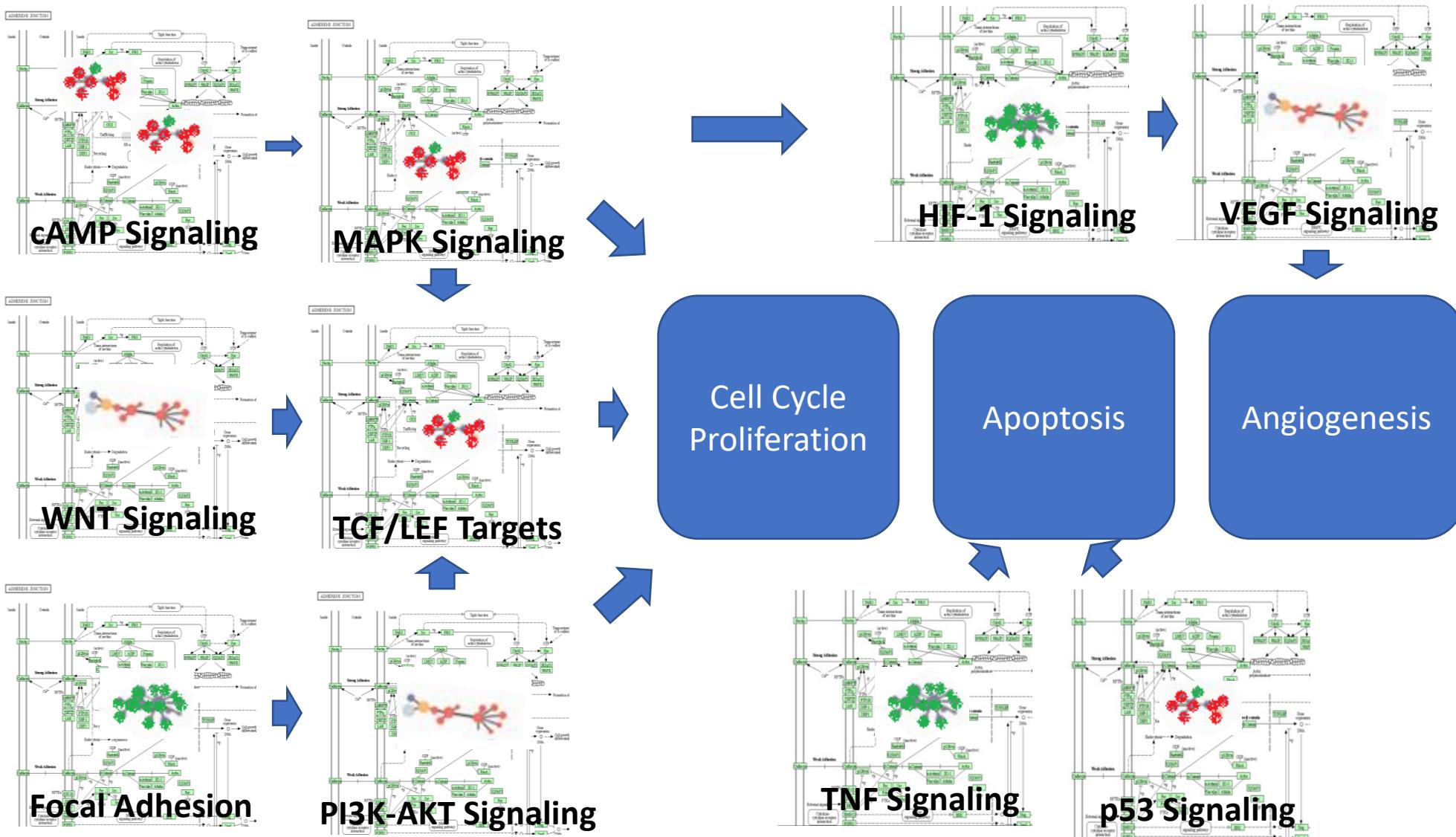
- Biological processes interact with each other to form higher level cellular process, often in a context specific manner.
- Identification of interactions between pathway routes can provide insights to detailed mechanism of whole cell function.
- Lack of consideration of such inter-pathway dependencies is one of the major limitations of standard pathway analysis tools which investigate individual pathways independently.

Higher order pathway analysis - Overview



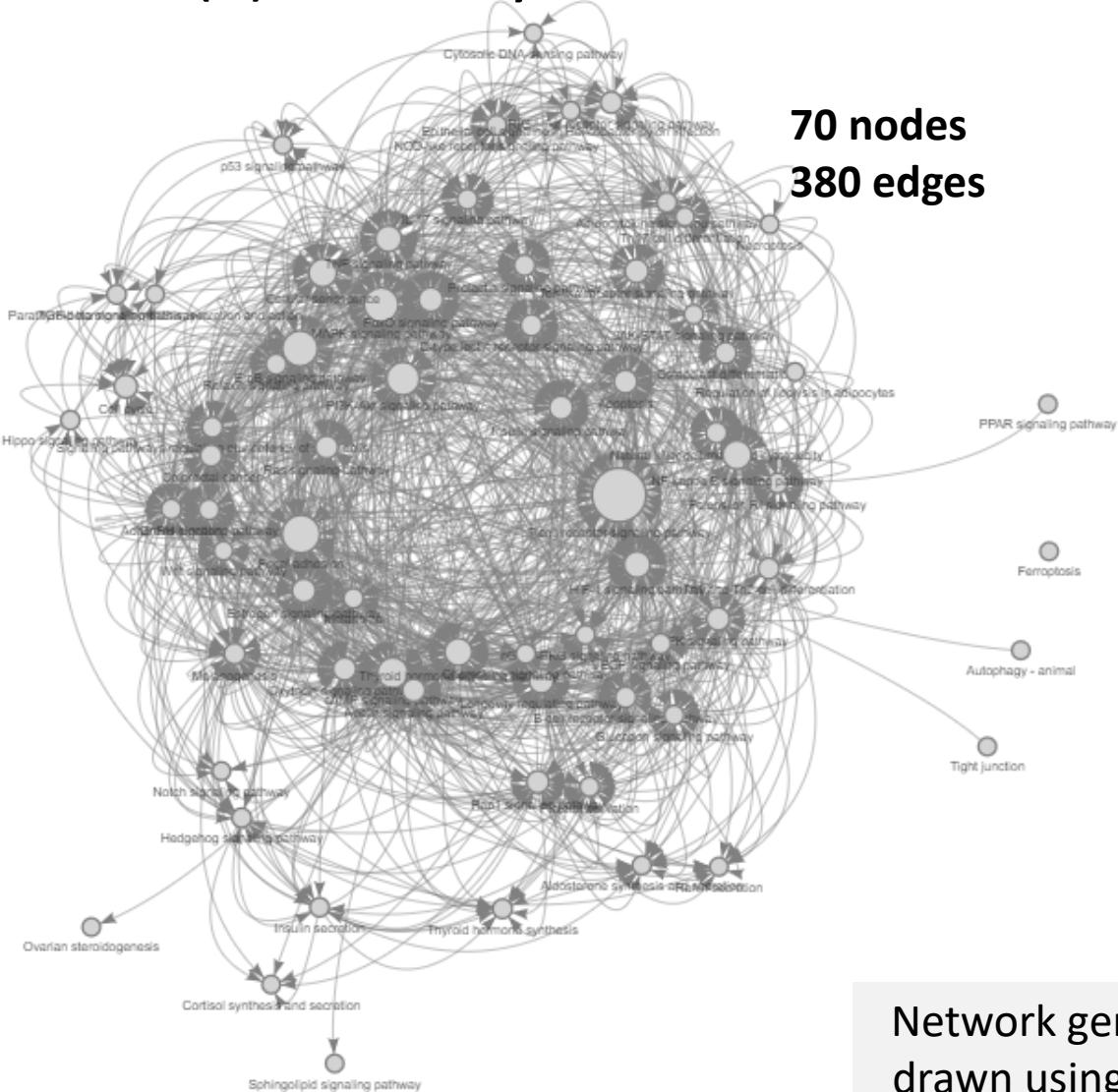
(a) Network representation of pathways (b) Sprayed with transcriptomics data (c) Isolated Routes

Context Specific Pathway Cluster (Example)

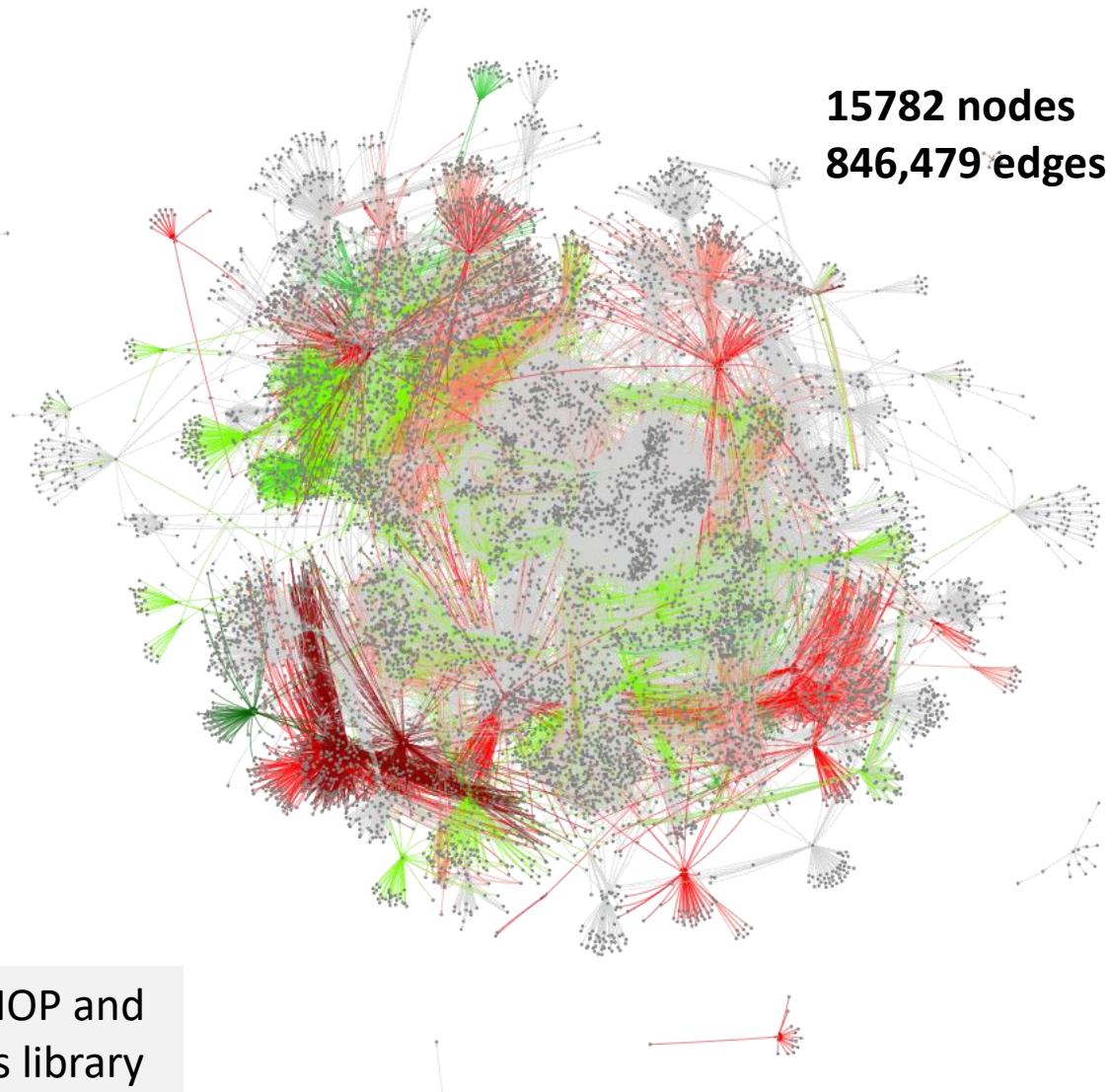


Global View of System Biology

(a) Pathway interactions



(b) Route interactions

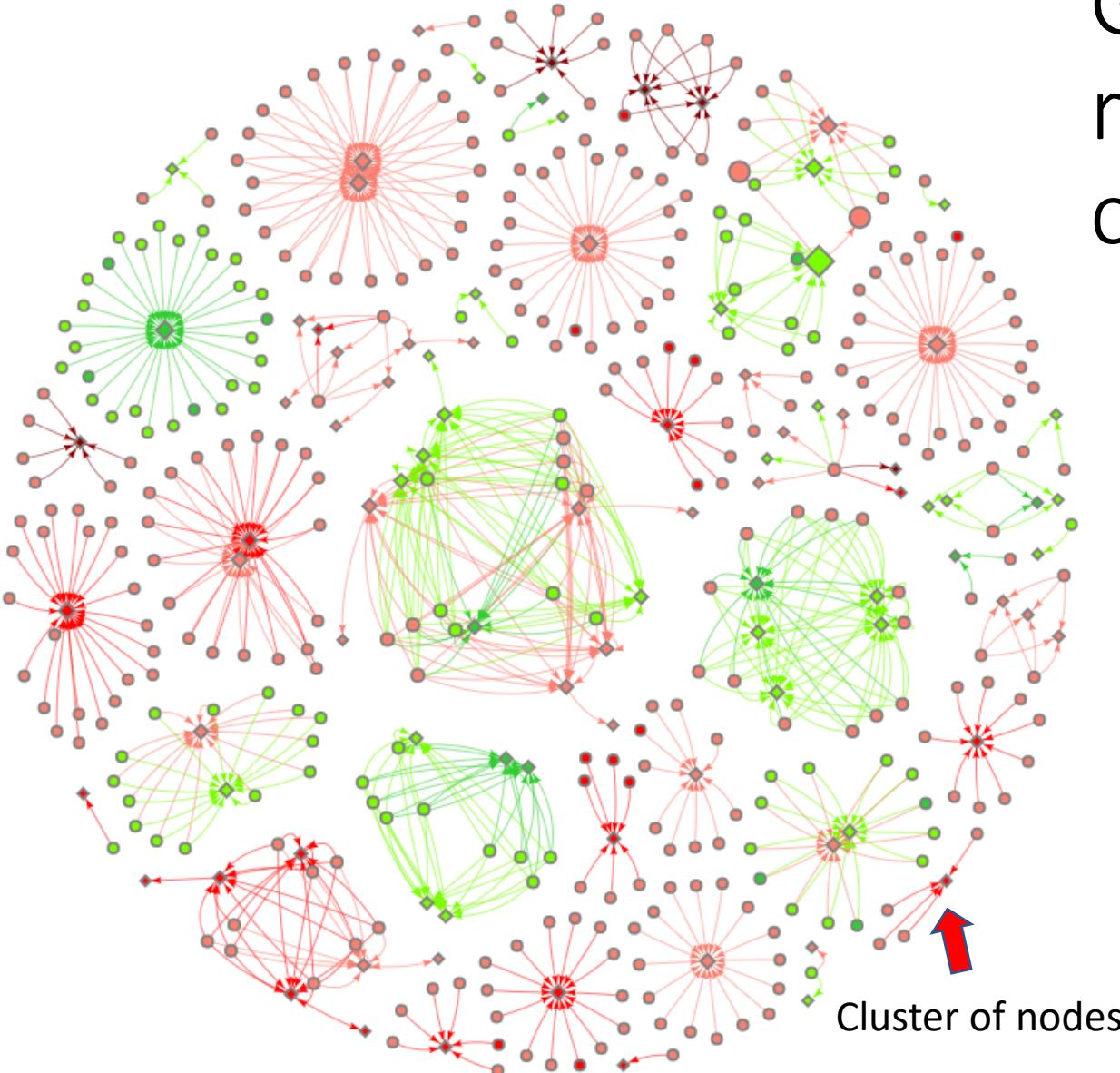


rHOP
Analysis



Network generated by rHOP and
drawn using python Pyvis library

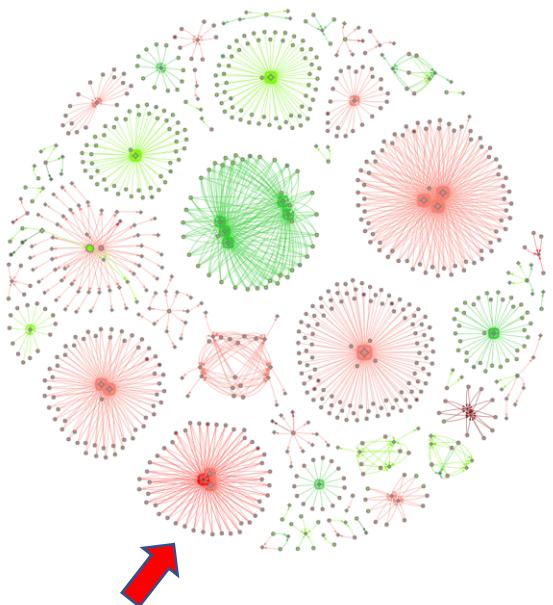
Global view of significantly regulated super routes in colon cancer ($|ARS| \geq 0.3$)



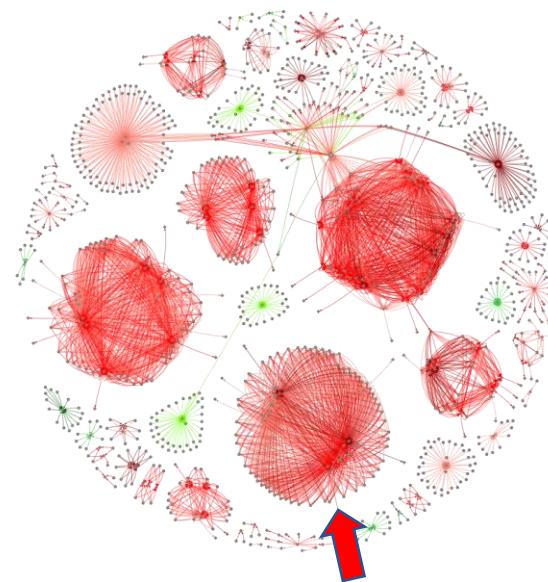
- Each node is a crosstalk route.
- Size of a node represents the degree of the node.
- Color of a node represents the regulation information of the route. Green represents down-regulation and red represents up-regulation.
- Shape of nodes are associated with their type. Circular nodes are signaling routes (or P1 routes), and diamond shaped nodes are effector routes (or P2 routes).

Global views of significantly regulated pathway routes in various cancer types

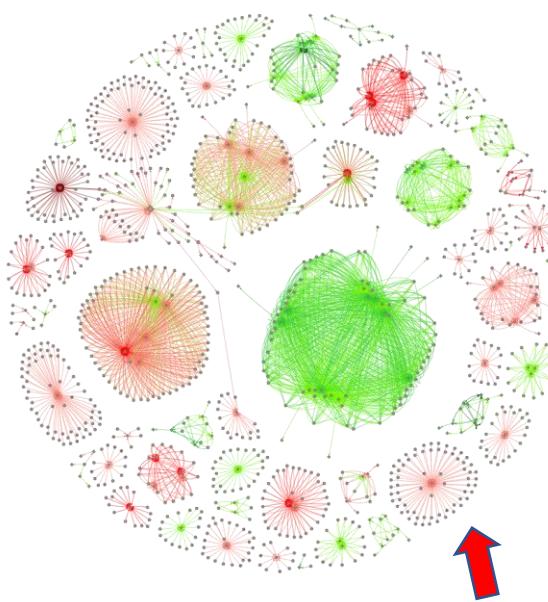
Breast Cancer
(BRCA)



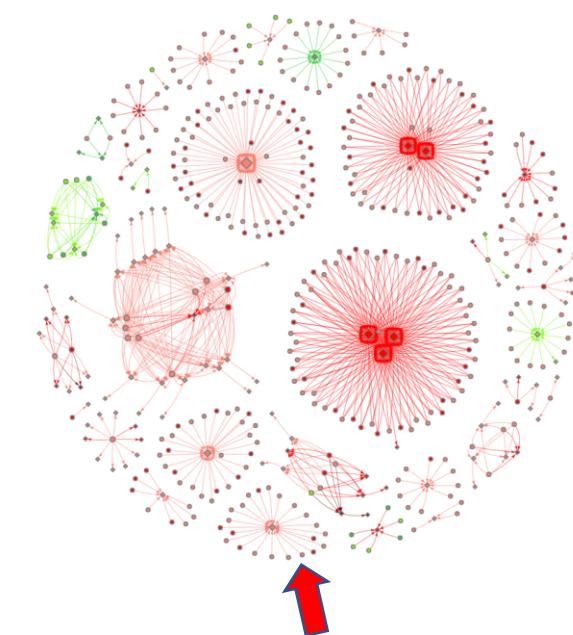
Brain Cancer
(GBM)



Skin Cancer
(SKCM)



Stomach Cancer
(STAD)



A higher order pathway (HOP) is a cluster of routes that are well-connected within the cluster but is disconnected (or minimally connected) with rest of the network.

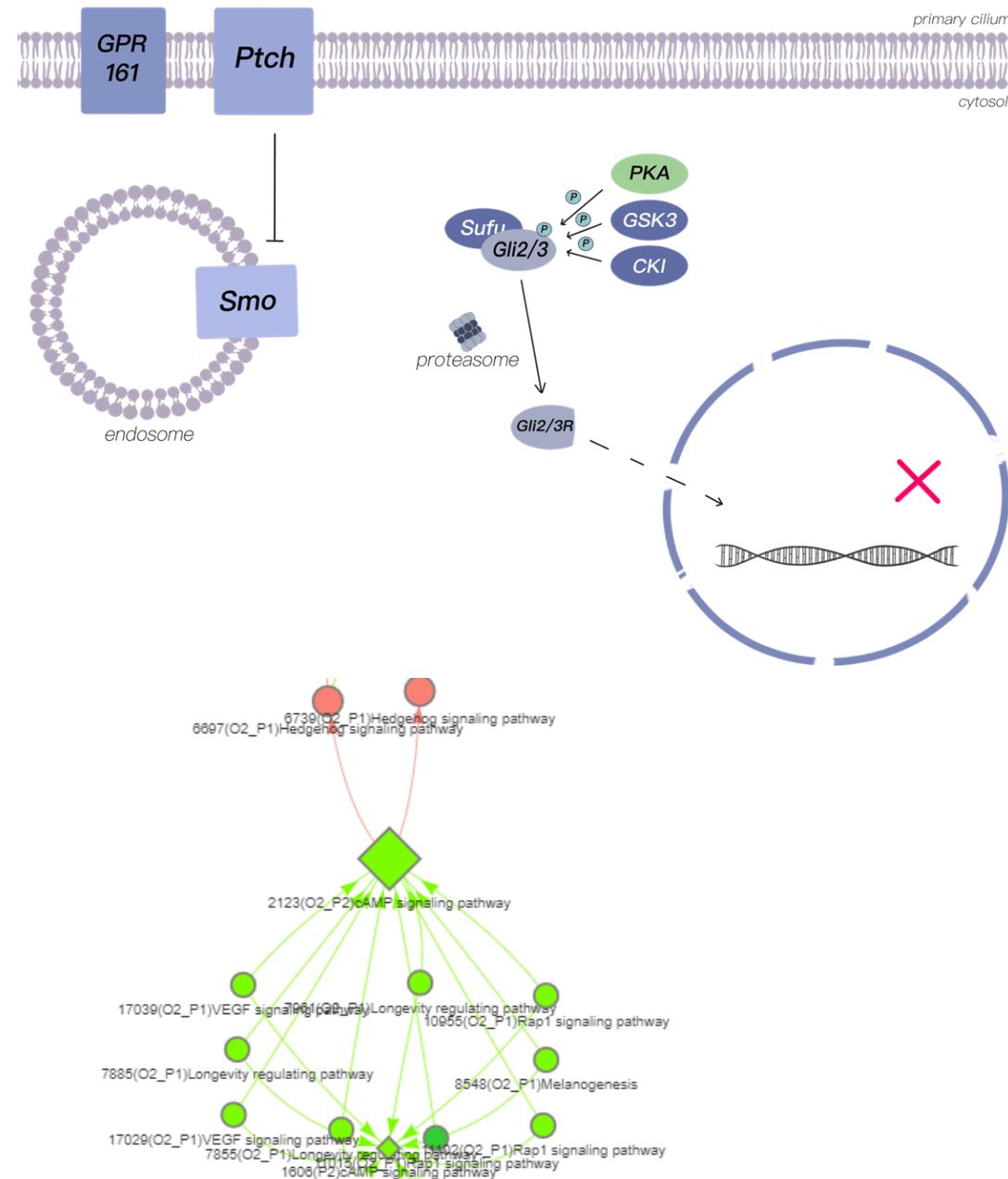
One Higher Order Pathway (HOP) in colon cancer



An example HOP identified by rHOP framework in colon cancer. Circular nodes are signaling routes (or P1 routes), and diamond shaped nodes are effector routes (or P2 routes). The figure illustrates the interaction between cAMP signaling pathway and FoxO, with Hedgehog as the intermediary, and with Rap1 signaling pathway towards the beginning of signaling.

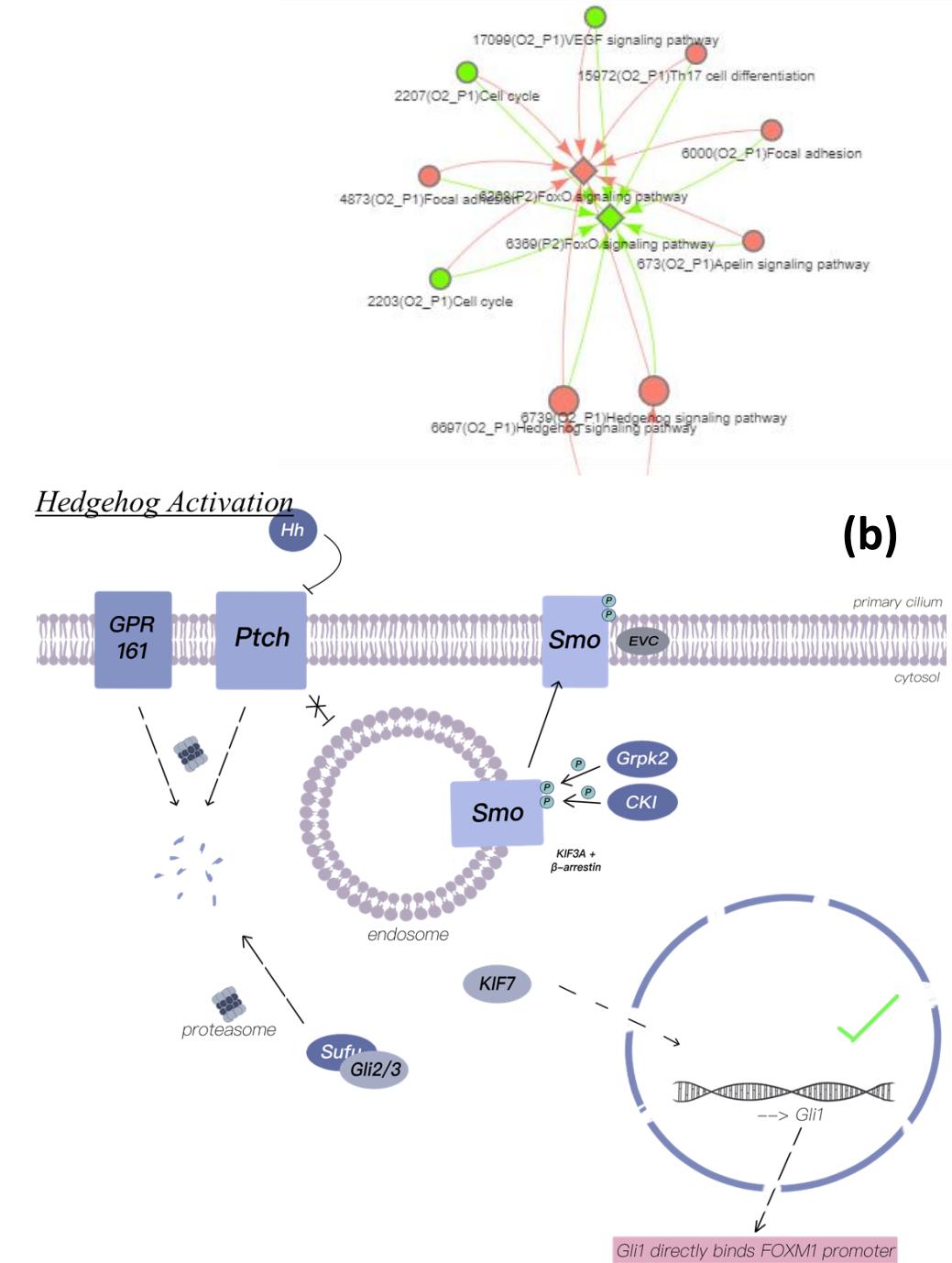
Hedgehog Inactivation

(a)

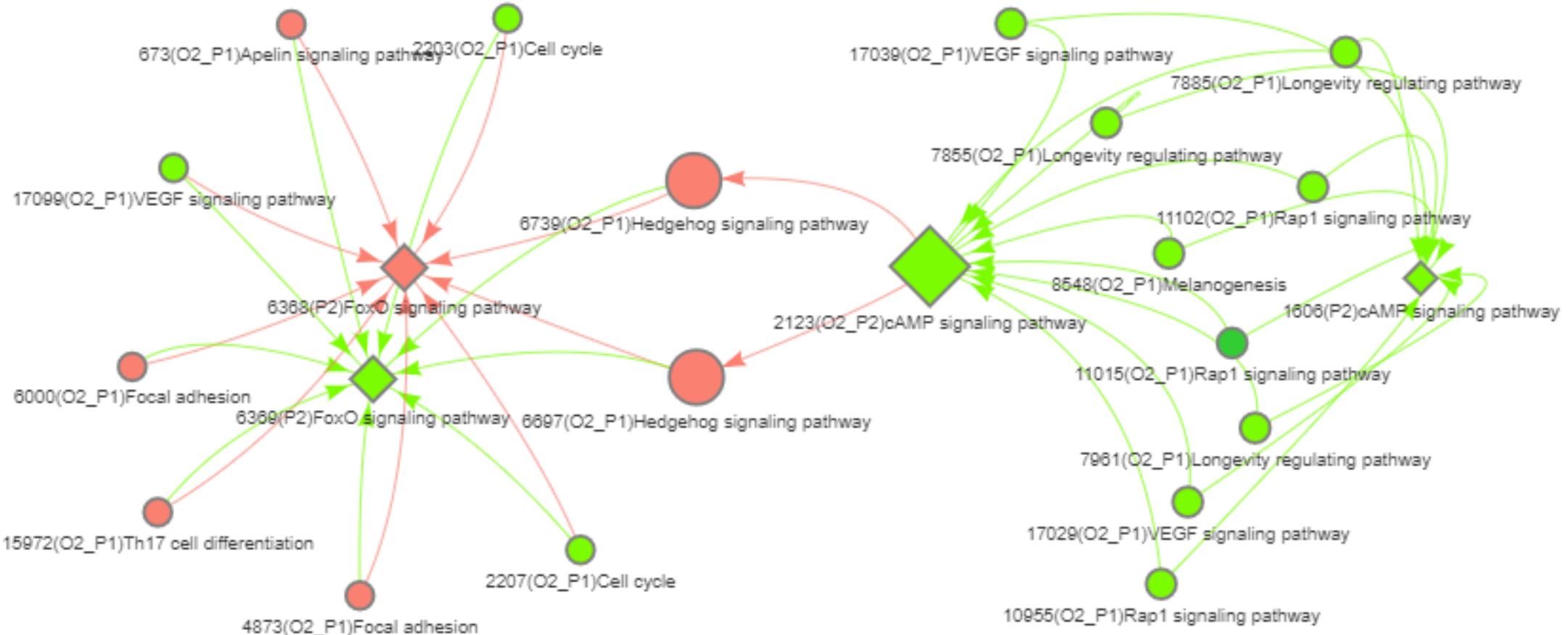


Hedgehog Activation

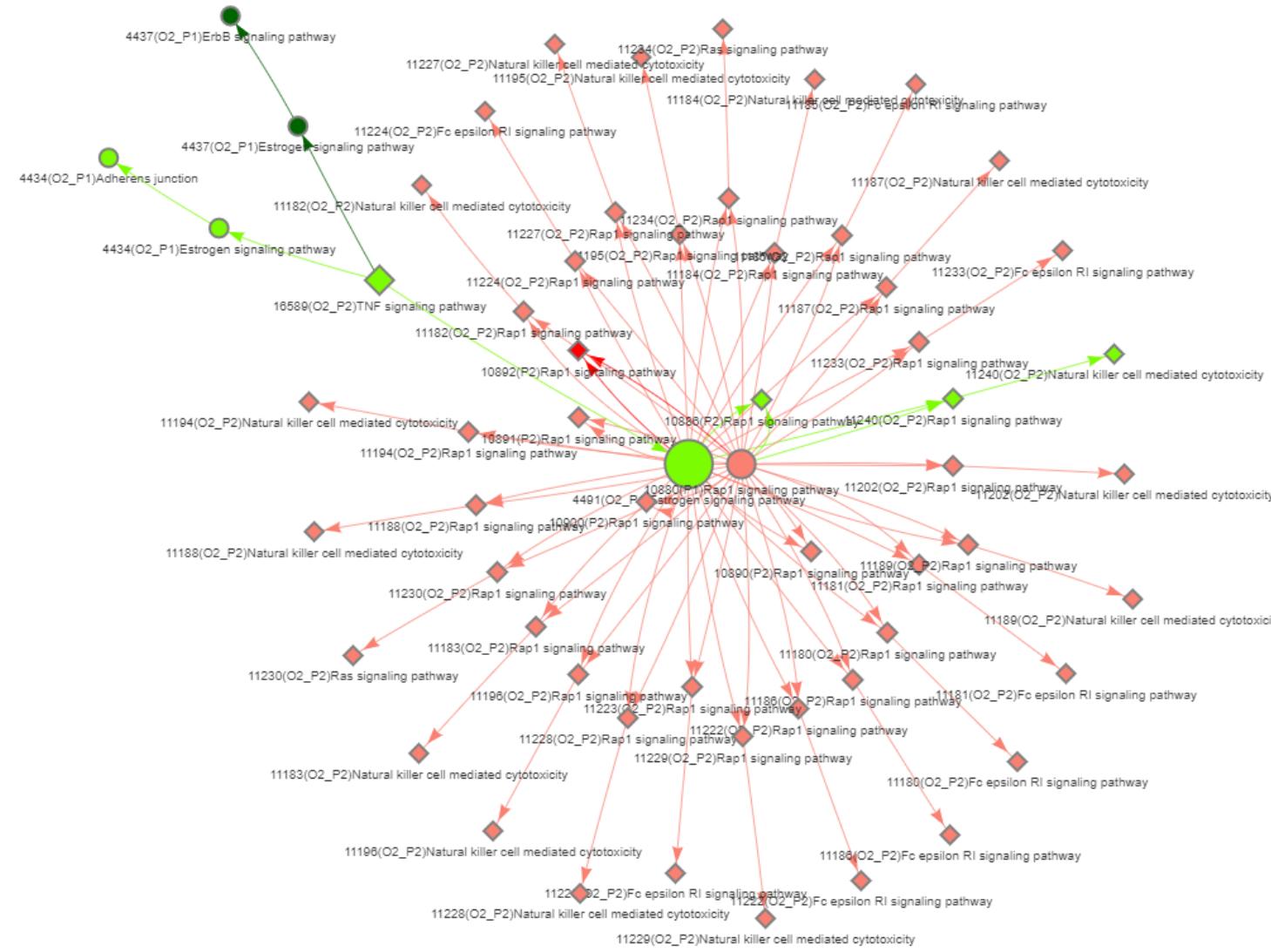
(b)



Higher Order Pathway (HOP)



An example HOP in breast cancer



Estrogen Signaling

Rap1 signaling

Rap1 signaling

Natural Killer Cell Mediated Cytotoxicity

Fc Epsilon
RI

Ras signaling

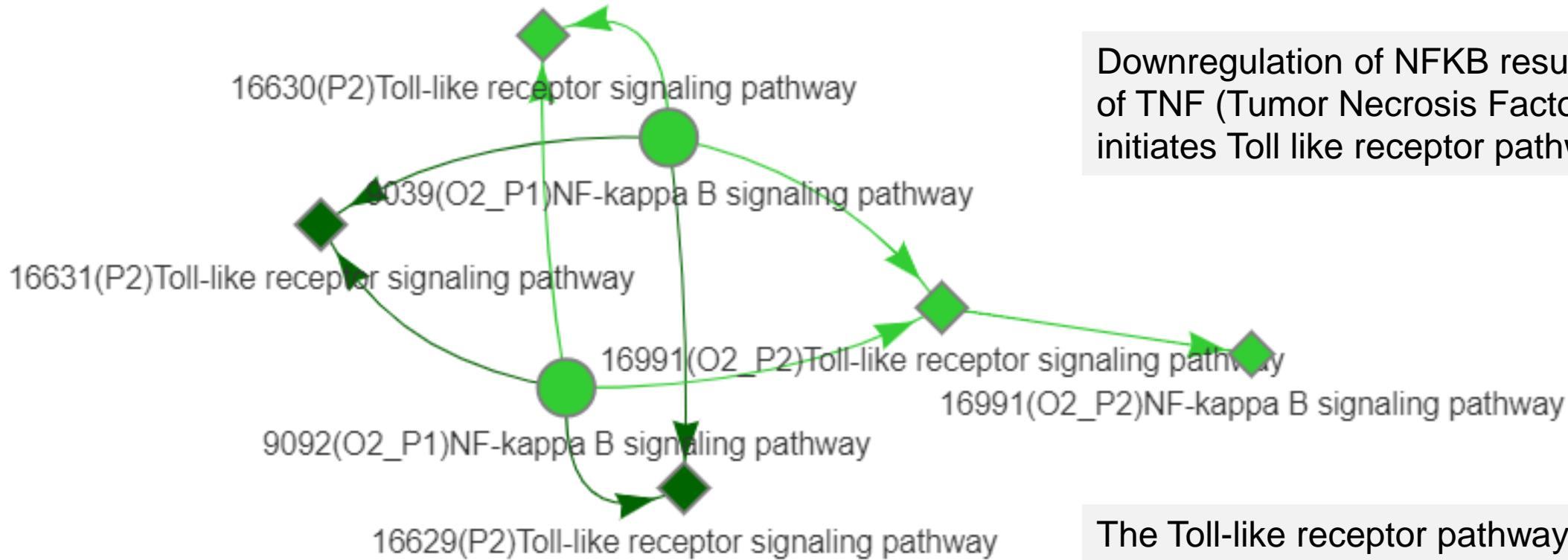
TNF signaling

Others

Estrogen

ErbB

An example HOP in skin cancer



Downregulation of NFKB results in the loss of TNF (Tumor Necrosis Factor) ligand that initiates Toll like receptor pathway.

The Toll-like receptor pathway is responsible for producing more of the TNF and IL1B ligands which regulate the NFKB pathway, thereby forming a **positive feedback loop**.

Related Work on combining pathways

- Dynamic Enrichment Analysis [Hansen et al., 2017]
- Curated Sub-Cellular Processes (SCP) from literature
- Enrichment analysis on SCPs, such as Fisher's exact test
- Enrichment analysis on combined SCPs that are highly ranked (based on common genes)
- Pathways and SCPs are considered as gene sets.
- Various components (e.g., ligand, receptors, TFs) are considered as elements of set.
- Our route-based methods overcome both issues.

Part 3 (rHOP) Conclusion

- Different biological processes may engage sets of similar and/or different pathways in different.
- The rHOP framework is designed to identify interactions between crosstalk routes that together give rise to higher level cell function.
- Using these interactions, higher order pathways (HOPs) are generated and presented in a noble view.
- Analysis of five cancer data sets revealed numerous HOPs, some of which are validated by extensive literature survey.

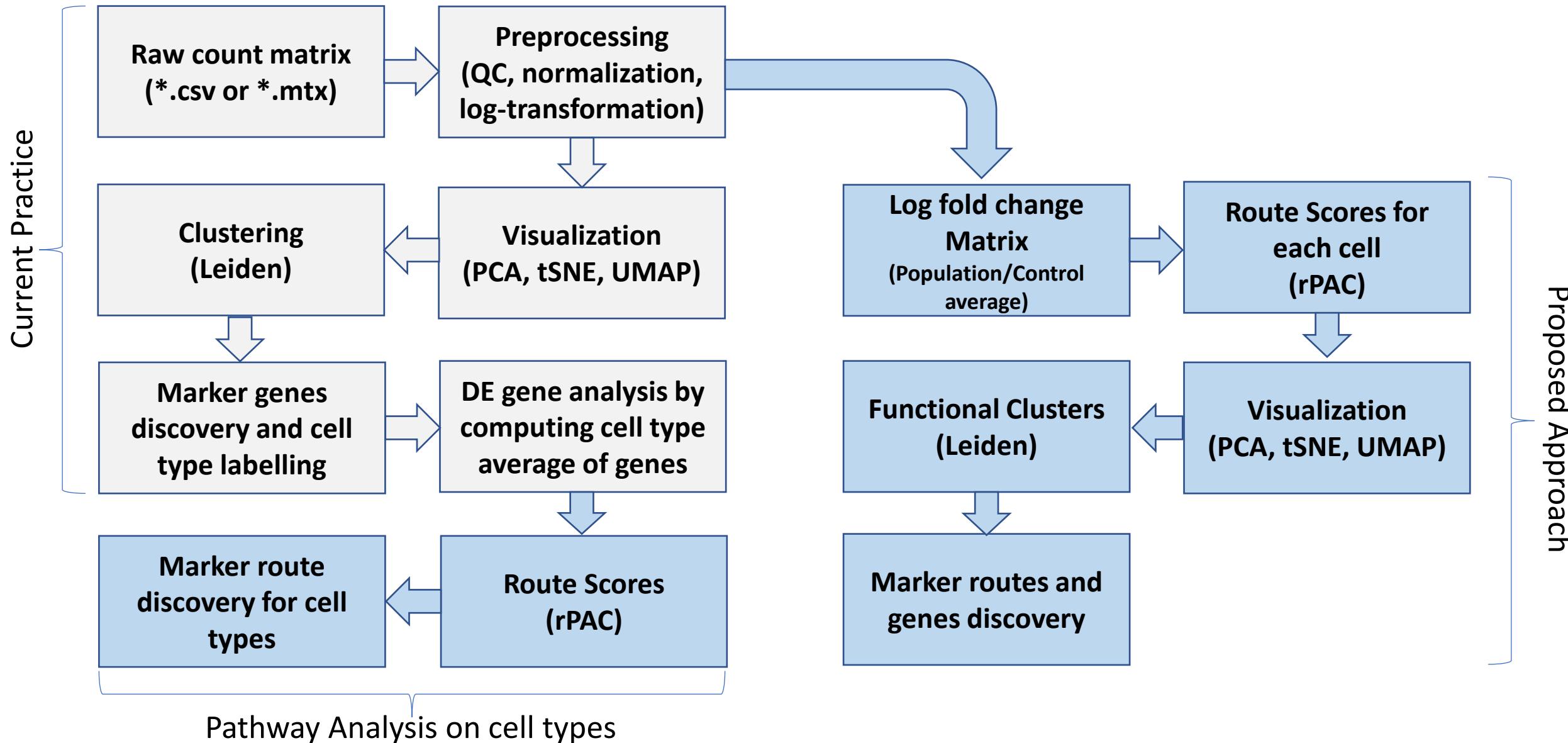
Organization of the talk

- Part 1: Route based Pathway Analysis of Cohorts (rPAC)
- Part 2: Route based Pathway Analysis of Crosstalk (rPAX)
- Part 3: Route based analysis of Higher Order Pathways (rHOP)
- Part 4: Route based pathway analysis of Single Cell Data

Part 4: Route Based Pathway Analysis on Single Cell Transcriptomics Data

- Bulk RNA sequencing provides an average readout of RNA content in a tissue and is influenced by the prevalent cell type.
- To characterize rare population and understand the mechanism behind complex diseases (e.g., cancer, Alzheimer's disease, etc.), higher resolution data is required which is provided by single cell RNA sequencing (scRNA-seq) data.
- The single-cell data, however, come with unique computational challenges emerged from technical limitations and transient nature of RNA.

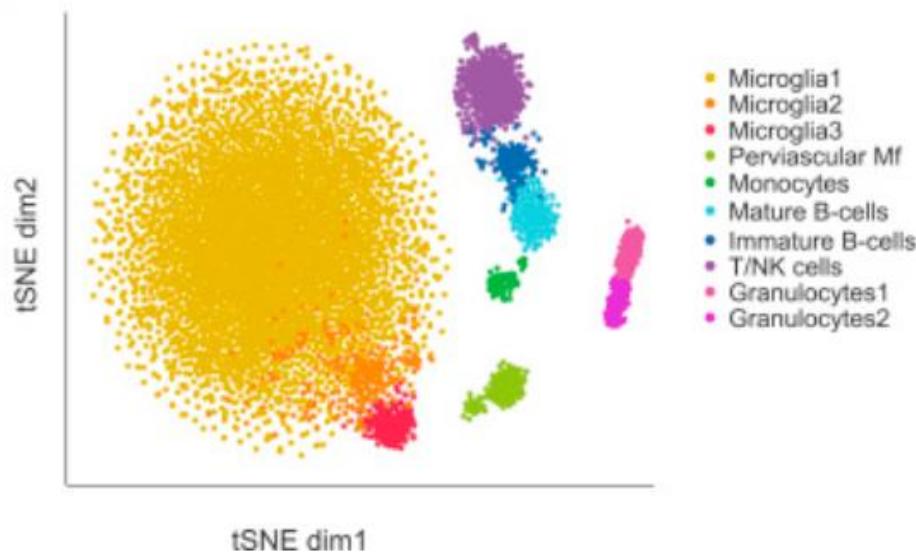
Overall Pathway Analysis Workflow



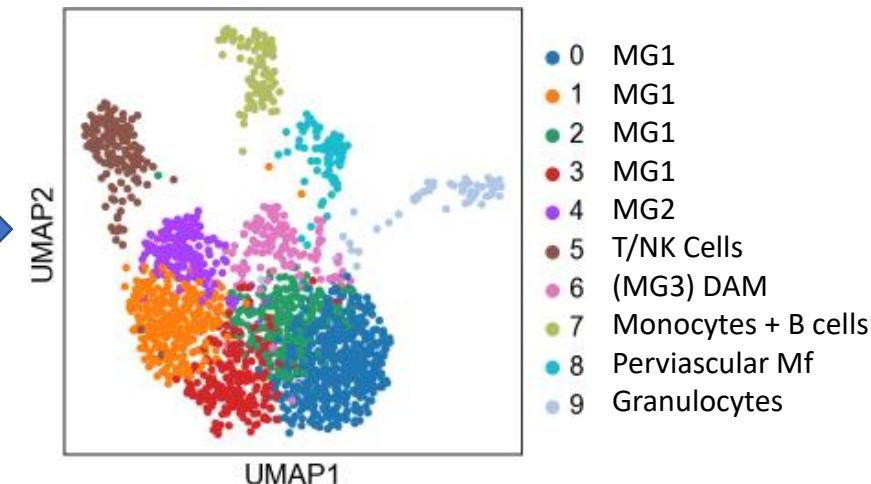
Single Cell RNA-seq Datasets

GSE	Context	No. of Cells	No. of Cell Types	Publication
GSE98969	Alzheimer's Disease (Mouse)	8K	10	Keren-Shaul et al., 2017
GSE174367	Alzheimer's Disease (Human)	60K	7	Morabito et al., 2021
GSE104782	Osteoarthritis (Human)	1600	7	Ji et al., 2019
NA	Articular Cartilage (Human)	10K	NA	Lotz et al., 2021

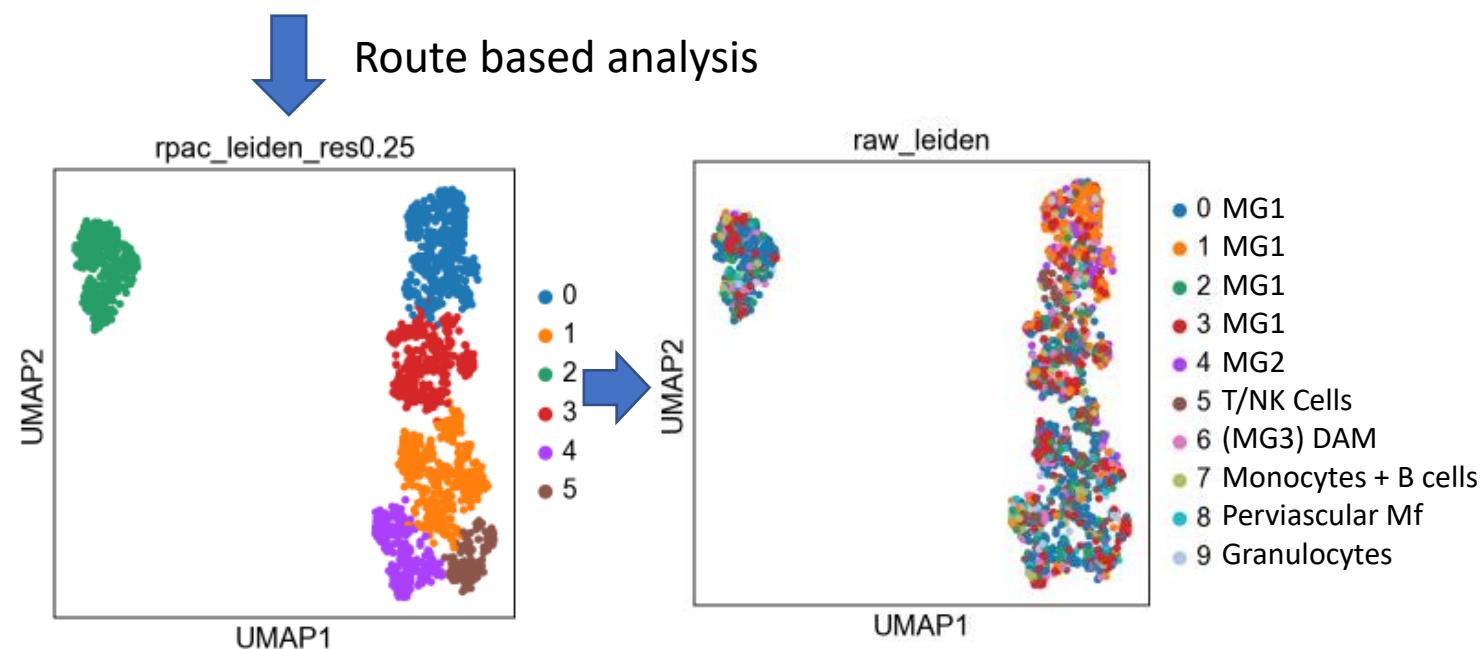
From Publication (8K cells)

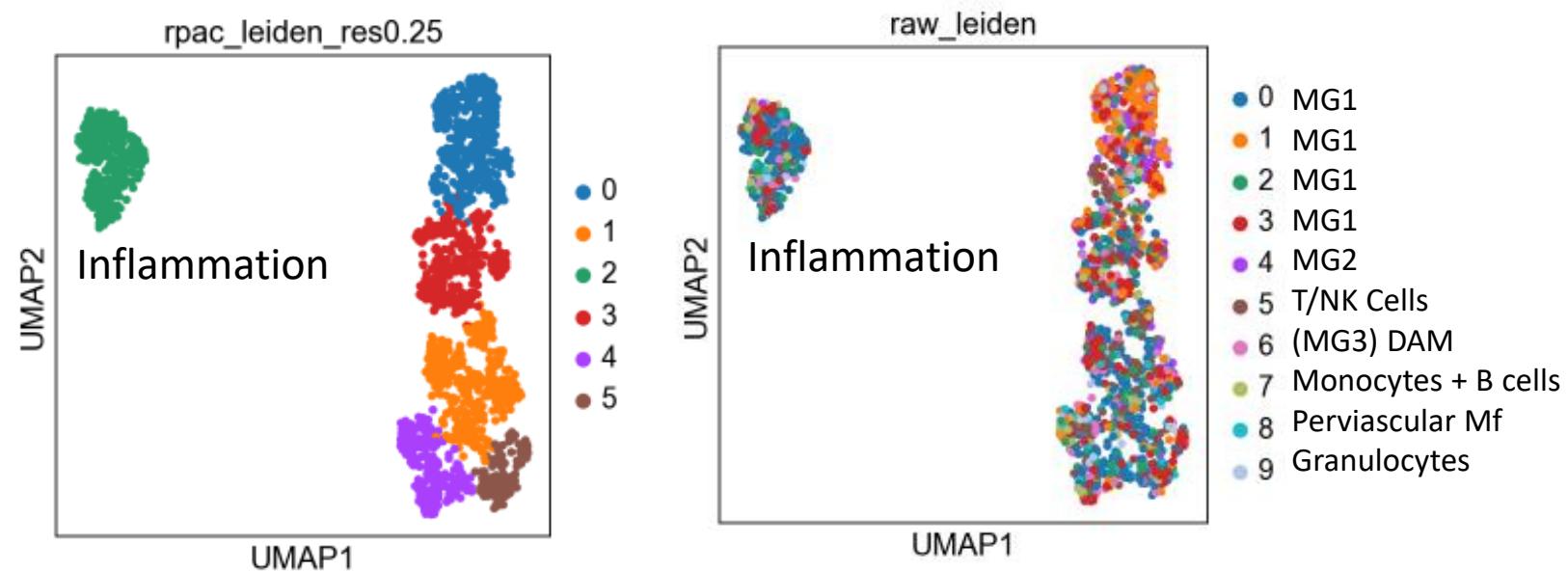


Gene based clustering (30K features)



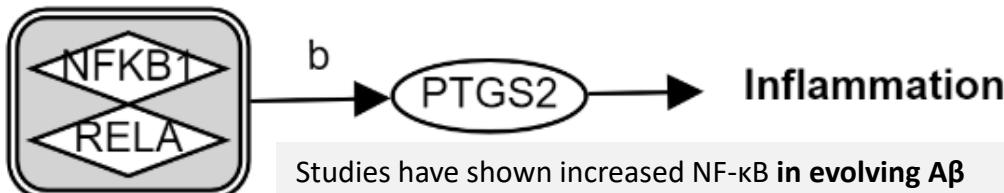
Route based clustering with 1695 features (route scores)



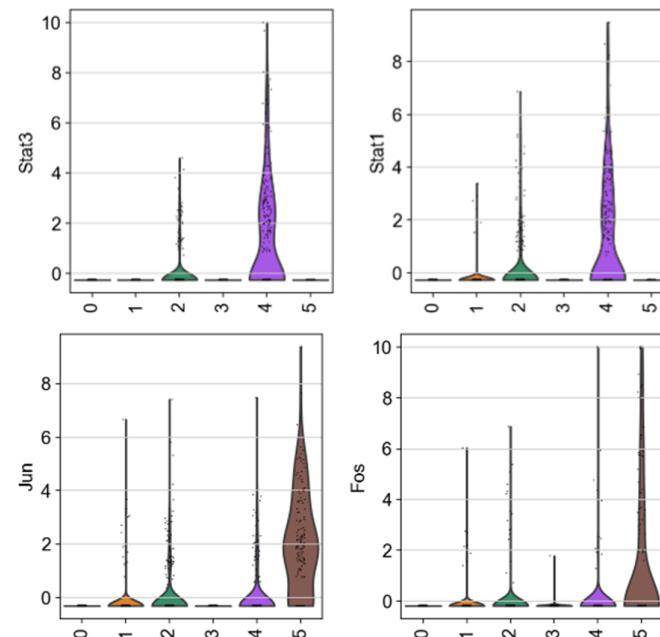
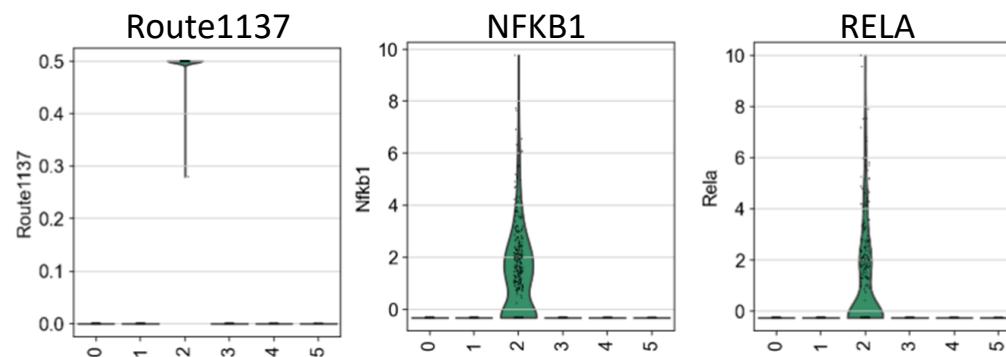


Alzheimer's disease is believed to occur when **abnormal amounts of amyloid beta (A β)**, accumulating extracellularly as **amyloid plaques** and **tau proteins**, or intracellularly as **neurofibrillary tangles**, form in the brain, affecting neuronal functioning and connectivity, resulting in a progressive loss of brain function.

Route1137 NF-kappa B signaling pathway



Studies have shown increased NF- κ B in evolving A β deposits than the areas surrounding more mature plaques. [Jones et al., 2017]

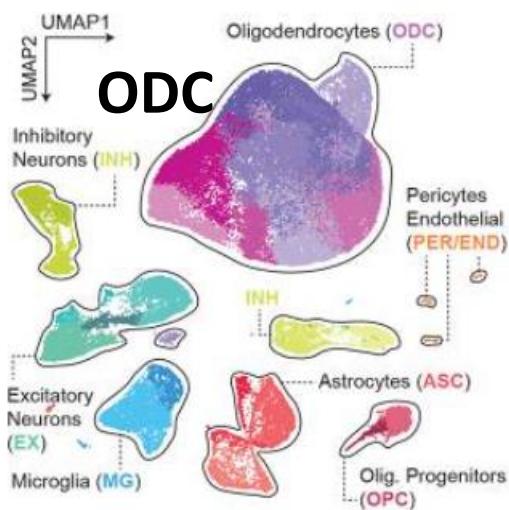


Stat1/Stat3 negatively regulates spatial memory formation and **mediates the memory-impairing effect of amyloid-beta (A β)**. [Hsu et al., 2014]

A β is the main component of plaques found in the brains with Alzheimer's disease.

Jun/Fos is activated in neurons of the Alzheimer's disease brain and suggests its involvement in abnormal processes, ranging from **tau phosphorylation** to neuronal death. [Okazawa et al., 2002]

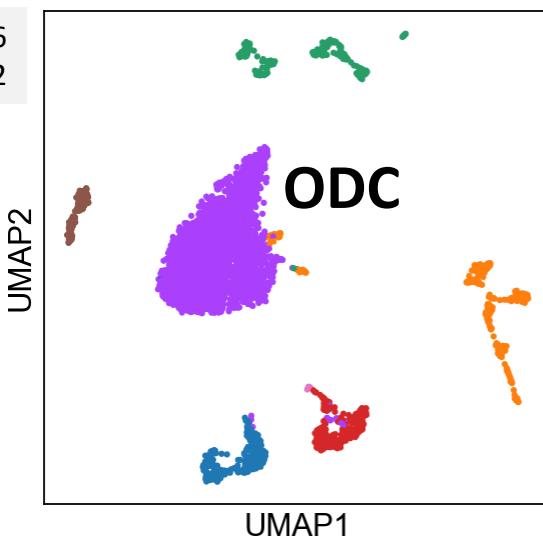
From Publication



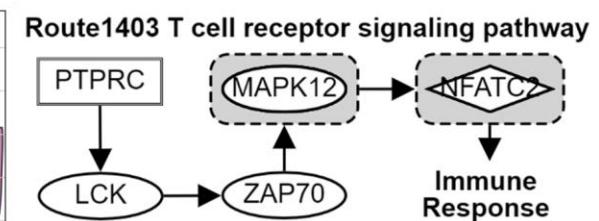
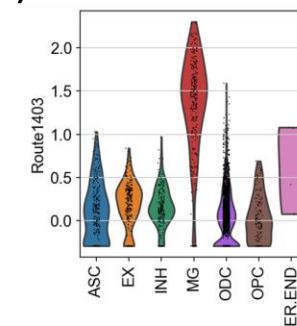
ARI = 0.96
AMI = 0.92



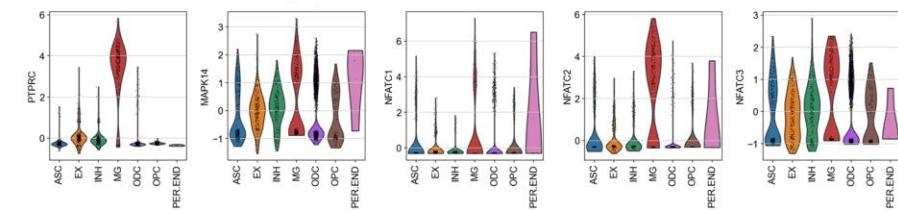
Gene based clustering (30K features)



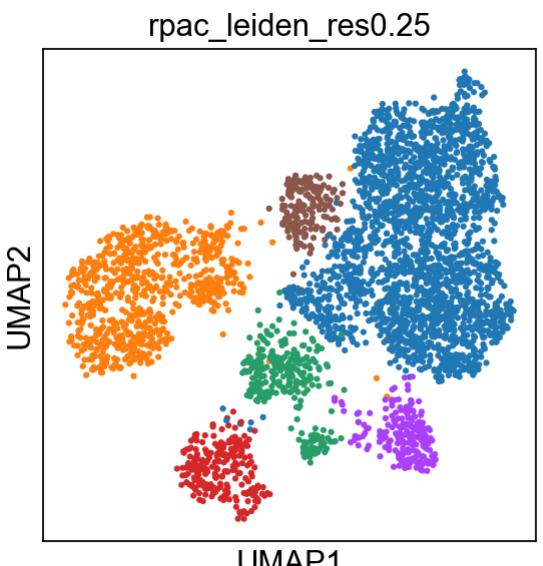
- ASC
- EX
- INH
- MG
- ODC
- OPC
- PER.END



Microglia is regulated by NFAT in Alzheimer's disease. [PMID: 20631193]



Route based analysis

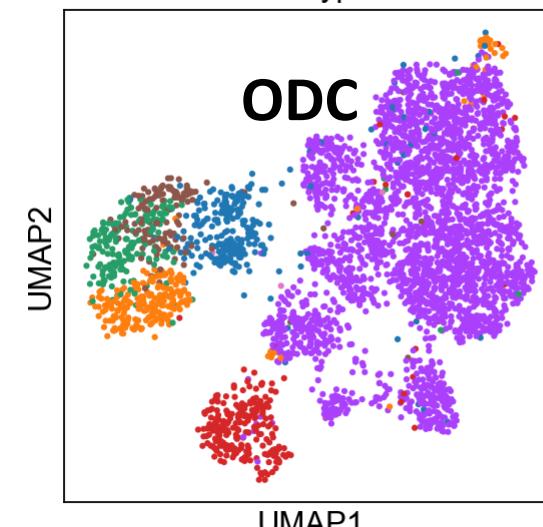


- 0
- 1
- 2
- 3
- 4
- 5



Route based clustering with 1695 features (route scores)

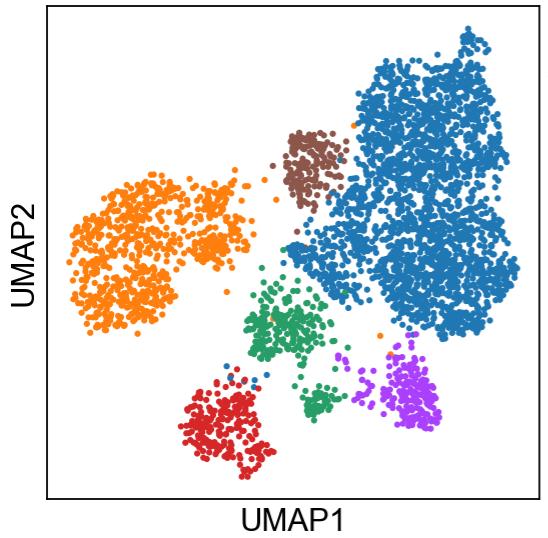
CellType



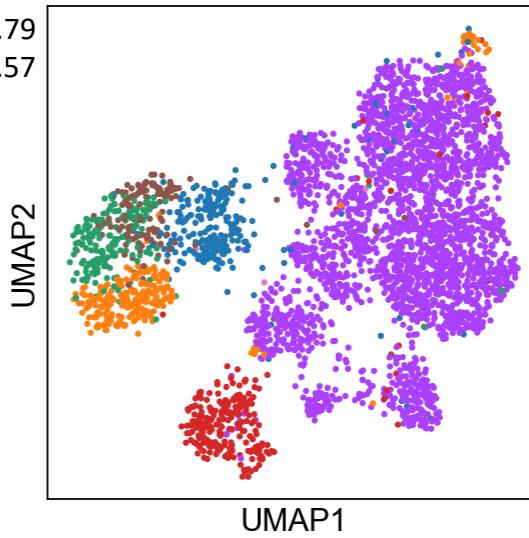
- ASC
- EX
- INH
- MG
- ODC
- OPC
- PER.END

GSE174367 - Alzheimer's Disease (AD) [3797 cells from 90 years old female (sample 17) with Stage C Plaque]

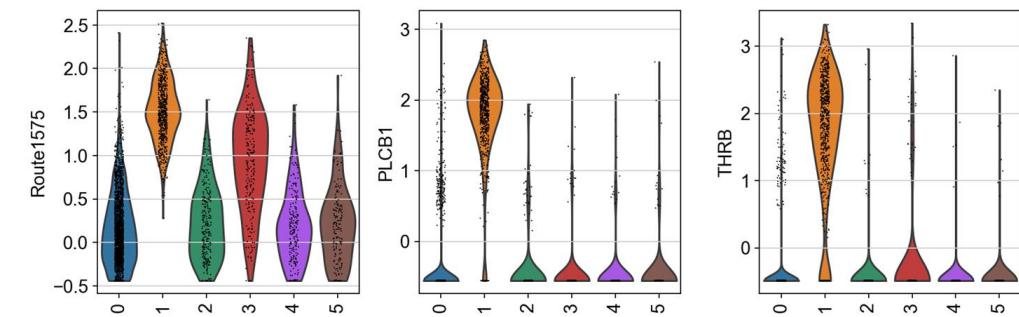
rpac_leiden_res0.25



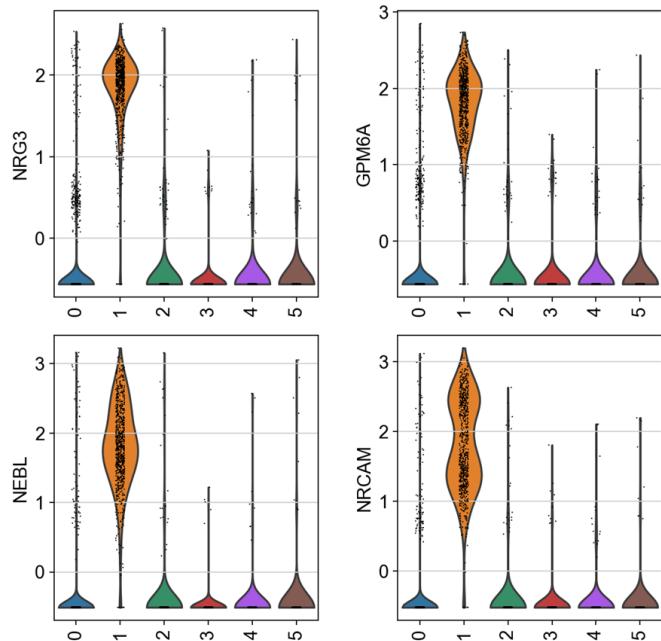
CellType



- ASC
- EX
- INH
- MG
- ODC
- OPC
- PER.END



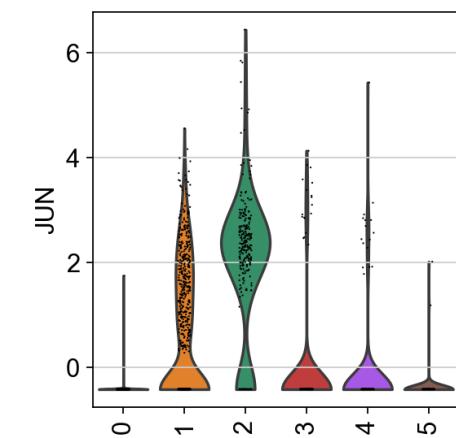
THRB is reported to be correlated with neuroserpin and HuD levels in brain to induce AD. [Subhadra et al., 2013]



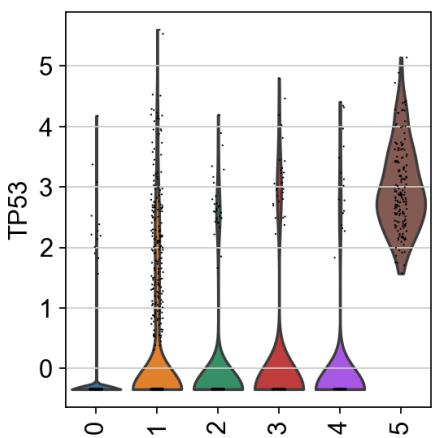
NRG3 and NRCAM are associated with risk in Alzheimer's disease.
[PMID: 24061483, 30833305]

GPM6A gene, was recently proposed as a gene-target in various neuropsychiatric disorders where it could also be used as a biomarker.
PMID: 34017241

NEBL was identified as highly expressed in eight brain regions and recognized as a risk factor for cognitive decline and dementia.
PMID: 28460139

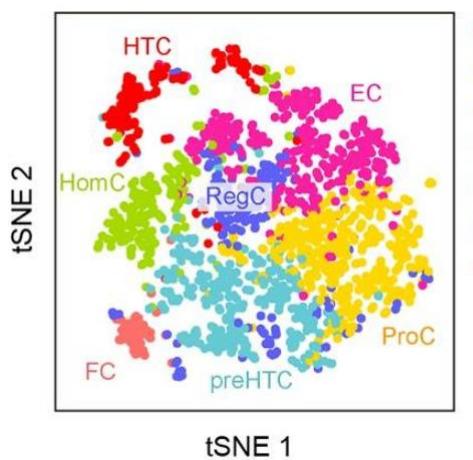


Jun is activated in neurons of the Alzheimer's disease brain and suggests its involvement in abnormal processes, ranging from tau phosphorylation to neuronal death. PMID: 11954673

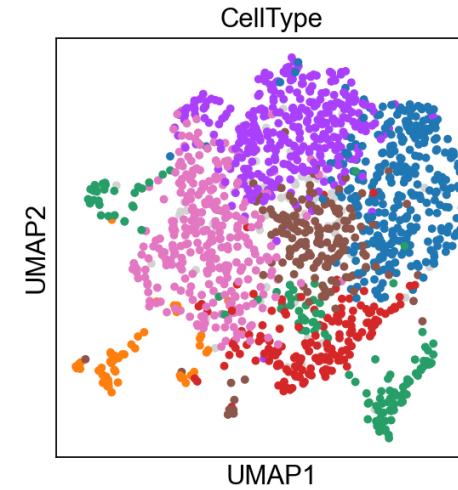


TP53 is known for cellular senescence. The conformational variant of p53 represents a potential peripheral biomarker that could detect AD at its earliest stages. [Abate et al., 2020]

From publication

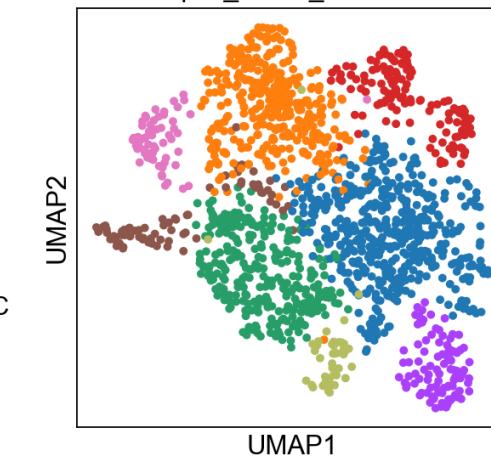


Gene based



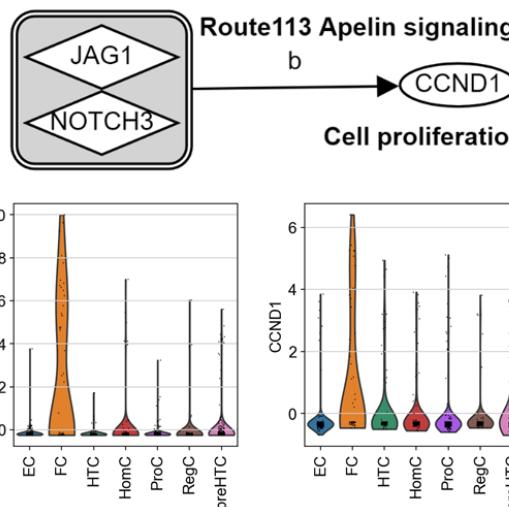
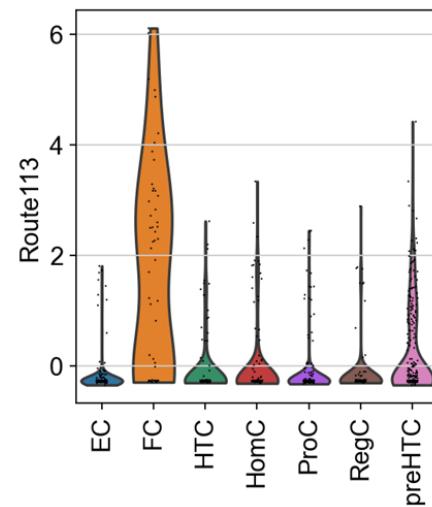
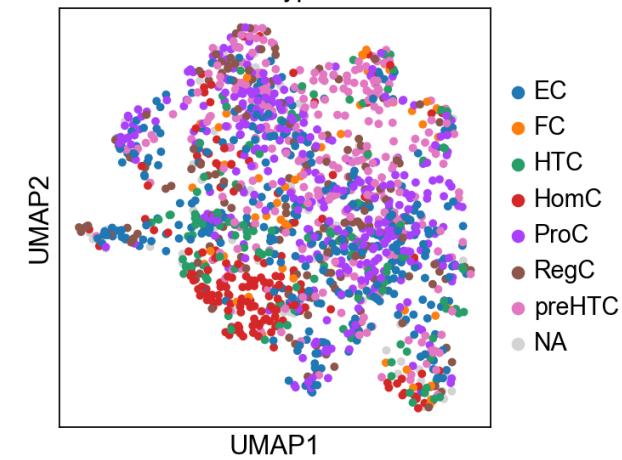
Route based

rpac_leiden_res0.35



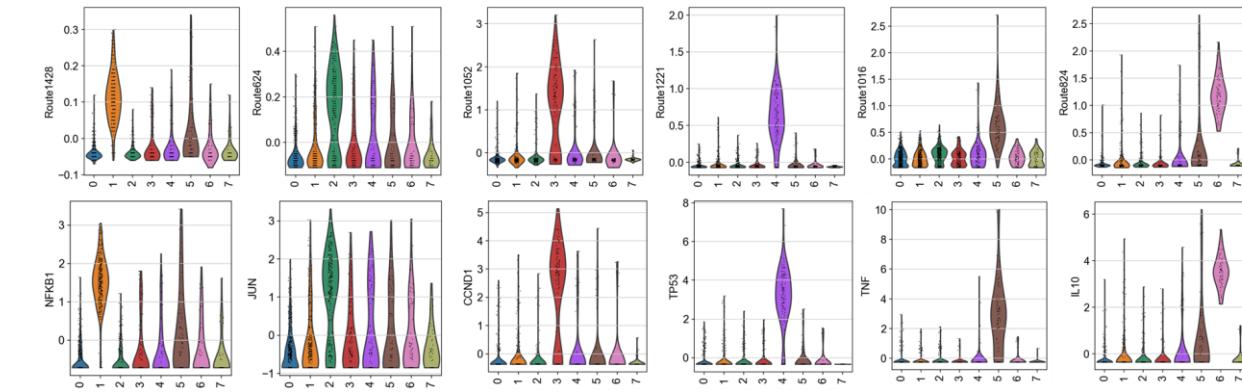
Route based

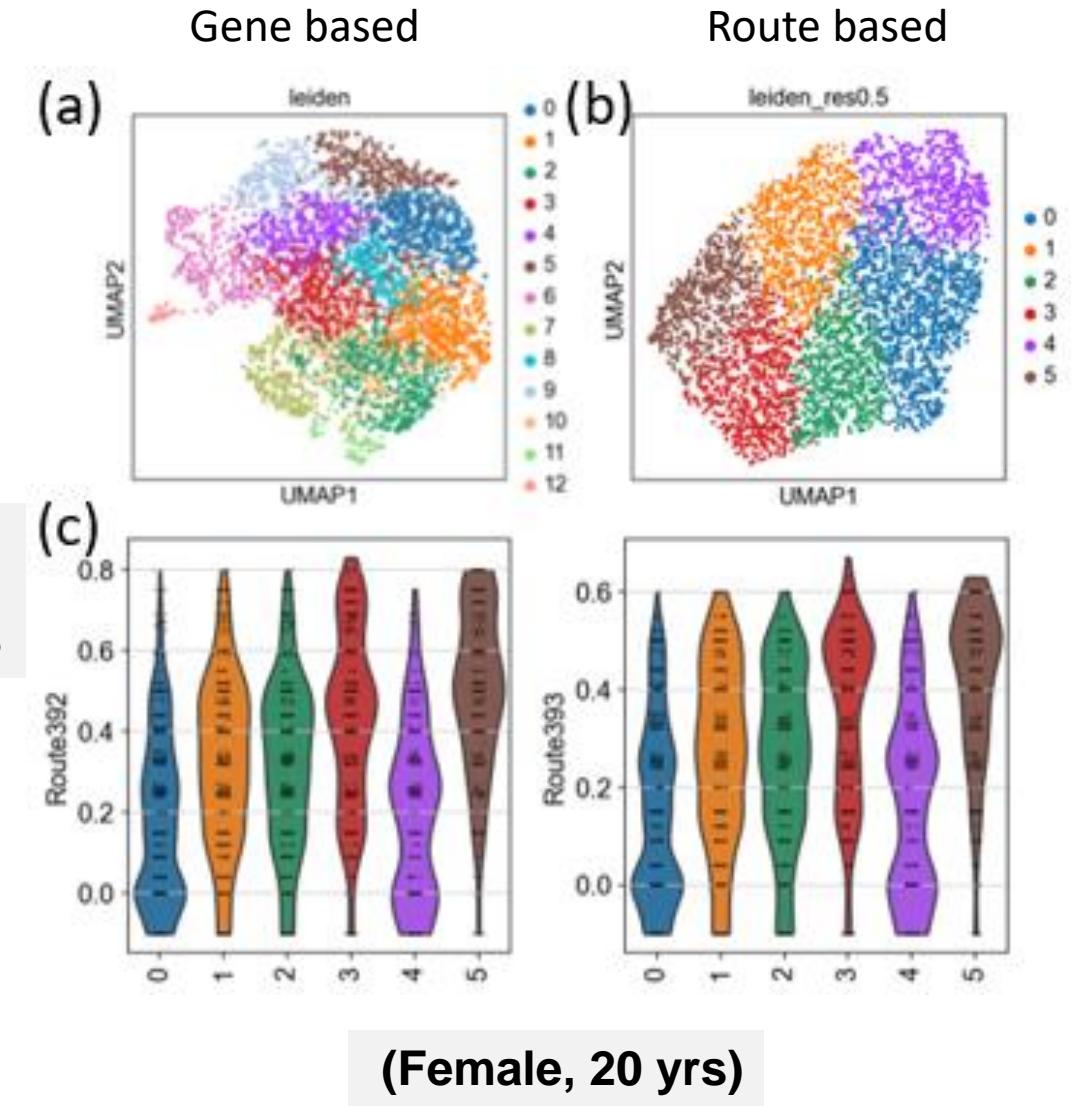
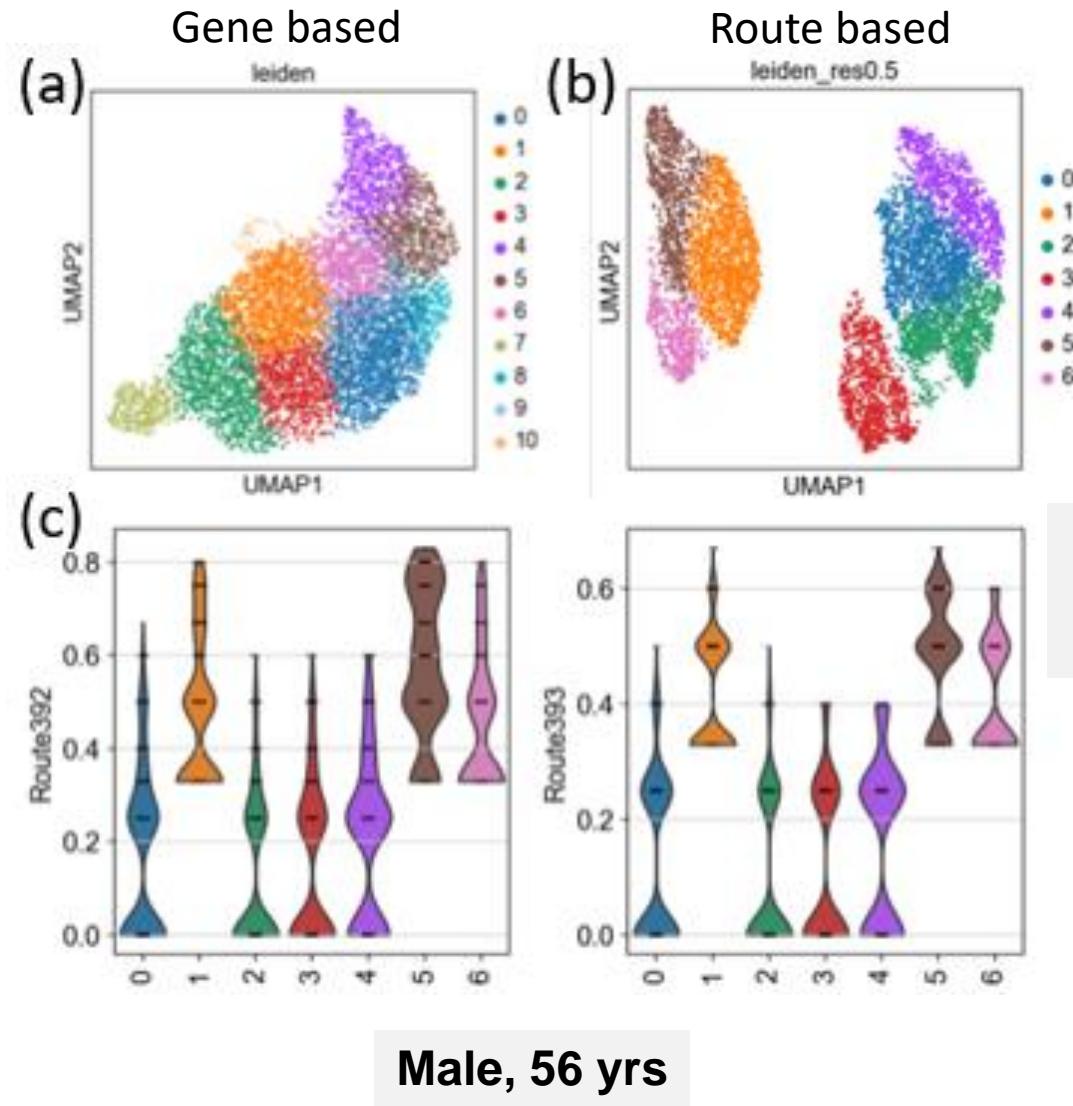
CellType



Route and gene markers based on cell types

Route and gene markers based on functional groups of cells





Conclusion

- Signaling mechanism can occur through perturbation of discrete routes within pathways. So, route-based analysis is preferred instead of pathway-based approach.
- A comprehensive analysis on five human cancer types was conducted to demonstrate importance and effectiveness of route-based approach. Survival analysis based on routes (intra/inter-pathway) revealed important predictors of survival.
- Inter pathway (crosstalk) routes were abstracted to identify and visualize higher order pathways (HOPs) in cancer.
- Route-based pathway analysis was also carried out in scRNA-seq datasets to identify cell-type marker routes and to group cells into “functional” clusters.

Future Work

- Inference of cell-cell interactions (CCIs) using crosstalk routes.
- Currently, ligand–receptor pairs are used to infer intercellular communication from the coordinated expression of genes
- This can be improved by considering the effector (P2) routes (begins with TF and ends with ligand) in sender cells and signaling (P1) route (which begin from receptor) in the receiver cells.
- Route regulation information can provide better confidence on the CCI inference. This can be validated with spatial genomics data to see association of inferred interacting cell pairs with their physical distance in the tissue.

Recent Selected Publications

1. P. Joshi, H. Wang, S. Jaramillo, S. Hong, C. Giardina, and D. Shin, "**Identification of Crosstalk between Biological Pathway Routes in Cancer Cohorts,**" IEEE International Conference on Bioinformatics and Biomedicine (BIBM), December 2021.
2. P. Joshi, B. Basso, H. Wang, S.H. Hong, C. Giardina, and D.G. Shin, "**rPAC: Route based pathway analysis for cohorts of gene expression data sets,**" Methods, vol. 198, pp. 76-87, 2022. PMID: 34628030. Epub Oct 7, 2021.
3. P. Joshi, H. Wang, B. Basso, S. Hong, C. Giardina, and D. Shin, "**A Framework for Route Based Pathway Analysis of Gene Expression Data,**" International Conference on Computational Biology and Bioinformatics (ICCBB '20), ACM, NY, Dec. 27–29, 2020.
4. P. Joshi, B. Basso, H. Wang, S. Hong, C. Giardina, and D. Shin, "**Identification of Key Biological Pathway Routes in Cancer Cohorts,**" IEEE International Conference on Bioinformatics and Biomedicine (BIBM), Virtual Event, South Korea, December 16-19, 2020.
5. P. Joshi, A. Vijaykumar, B. Enkhmandakh, DG. Shin, M. Mina, D. Bayarsaihan, "**The chromatin accessibility landscape in the dental pulp of mouse molars and incisors,**" Acta Biochimica Polonica, vol. 69(1):131-138, 2022. PMID: 35226446.
6. P. Joshi, A. Vijaykumar, B. Enkhmandakh, M. Mina, DG. Shin, D. Bayarsaihan, "**Genome-wide distribution of 5hmC in the dental pulp of mouse molars and incisors,**" The Journal of Biochemistry, Volume 171, Issue 1, Pages 123–129, 2022. PMID: 34676418.
7. H. Wang, P. Joshi, SH. Hong, P. Maye, D. Rowe, DG. Shin, "**Predicting the targets of IRF8 and NFATc1 during osteoclast differentiation using the machine learning method framework cTAP,**" BMC Genomics, vol. 23, 14, 2022.
8. H. Wang, P. Joshi, SH. Hong, DJ. Shin, and DG. Shin, "**ctBuilder: A framework for building pathway crosstalk by combining single cell data with bulk cell data,**" 2021 IEEE International Conference on Bioinformatics and Biomedicine (BIBM), December 2021.
9. Y. Zhao, P. Joshi and D. Shin, "**Recurrent Neural Network for Gene Regulation Network Construction on Time Series Expression Data,**" IEEE International Conference on Bioinformatics and Biomedicine (BIBM), 2019.

References

1. Khatri P, Sirota M, Butte AJ, "Ten Years of Pathway Analysis: Current Approaches and Outstanding Challenges. PLOS Computational Biology 8(2): e1002375, 2012.
2. S. Tyagi, S. Sharma, P. Gupta, A. Saini and C. Kaushal, "The peroxisome proliferator-activated receptor: A family of nuclear receptors role in various diseases," Journal of Advanced Pharmaceutical Technology and Research, vol. 2, pp. 236-240, 2011.
3. B.J. Coventry, P. Lee, D. Gibbs and D.N.J. Hart, "Dendritic cell density and activation status in human breast cancer – CD1a, CMRF-44, CMRF-56 and CD-83 expression," British Journal of Cancer, vol. 86, pp. 546-551, 2002.
4. S. Qie and J.A. Diehl, "Cyclin D1, cancer progression, and opportunities in cancer treatment," J Mol Med, vol. 94, pp. 1313-1326, 2016.
5. C. Zheng, H. Hu, P. Hong, Q. Zhang, W.W. Xu, Q. He and B. Li, "'Significance of integrin-linked kinase (ILK) in tumorigenesis and its potential implication as a biomarker and therapeutic target for human cancer," American journal of cancer research, vol. 9, no. 1, pp. 186-197, 2019.
6. A.I. Robles and C.C. Harris, "Clinical Outcomes and Correlates of TP53 Mutations and Cancer," Cold Spring Harbor Perspectives in Biology, vol. 2, pp. a001016, 2009.
7. V. Padmanaban, I. Krol, Y. Suhail, B.M. Szczerba, N. Aceto, J.S. Bader and A.J. Ewald, "E-cadherin is required for metastasis in multiple models of breast cancer," Nature (London), vol. 573, pp. 439-444, 2019.
8. S. Ghafouri-Fard, V.K. Oskooei, I. Azari and M. Taheri, "Suppressor of cytokine signaling (SOCS) genes are downregulated in breast cancer," World Journal of Surgical Oncology, vol. 16, pp. 226, 2018.

References (..continued)

9. H. Chen, S. Libring, K.V. Ruddaraju, J. Miao, L. Solorio, Z.Y. Zhang, and M.K. Wendt, "SHP2 is a multifunctional therapeutic target in drug resistant metastatic breast cancer," *Oncogene*, vol. 39, pp. 7166-7180, Dec. 2020.
10. M.N. Vansaun, "Molecular Pathways: Adiponectin and Leptin Signaling in Cancer," *Clinical Cancer Research*, vol. 19, pp. 1926-1932, Apr 15, 2013.
11. M. Schmidt, D. Böhm, C. von Törne, E. Steiner, A. Puhl, H. Pilch, H.A. Lehr, J.G. Hengstler, H. Kölbl and M. Gehrmann, "The humoral immune system has a key prognostic impact in node-negative breast cancer," *Cancer Res.*, vol. 68, pp. 5405-5413, Jul 1. 2008.
12. Frida Belinky, Noam Nativ, Gil Stelzer, Shahar Zimmerman, Tsippi Iny Stein, Marilyn Safran, Doron Lancet, "PathCards: multi-source consolidation of human biological pathways", Database, 2015.
13. L. Wadi, M. Meyer, J. Weiser, L.D. Stein and J.ü Reimand, "Impact of outdated gene annotations on pathway enrichment analysis," *Nature Methods*, vol. 13, pp. 705-706, 2016.
14. W. Gao, G. Ye, L. Liu and L. Wei, "The downregulation of Rap1 GTPase-activating protein is associated with a poor prognosis in colorectal cancer and may impact on tumor progression," *Oncology Letters*, vol. 15, pp. 7661-7668, 2018.
15. Morabito S, Miyoshi E, Michael N, Shahin S et al. Single-nucleus chromatin accessibility and transcriptomic characterization of Alzheimer's disease. *Nat Genet* 2021 Aug;53(8):1143-1155. PMID: 34239132

Thank You

Questions?