Executable network models of integrated multi-omics data

Mukta G. Palshikar[1], Xiaojun Min[2], Alexander Crystal[2], Jiayue Meng, Shannon Hilchey[3], Martin Zand[3,4], Juilee Thakar[1,4,5,6,7*]

[1] Biophysics, Structural, and Computational Biology Program, University of Rochester School of Medicine and Dentistry, Rochester, USA. [2]University of Rochester College of Arts, Sciences and Engineering, Rochester, USA. [3]Division of Nephrology, University of Rochester School of Medicine and Dentistry, Rochester, USA. [4]University of Rochester Clinical & Translational Science Institute, Rochester, USA. [5]Department of Biostatistics and Computational Biology, University of Rochester School of Medicine and Dentistry, Rochester, USA. [6]Department of Microbiology and Immunology, University of Rochester School of Medicine and Dentistry, Rochester, USA. [7]Department of Biomedical Genetics, University of Rochester School of Medicine and Dentistry, Rochester, USA.

Outline of manuscript

1) Multi-omics network modeling and pathway enrichment analysis with mBONITA
    1) Outline methods for data integration, rule inference, node modulation scores, and pathway analysis (Figure 1)
2) mBONITA identifies mechanisms of hypoxia-mediated chemotaxis in RAMOS B cells (pathway analysis with mBONITA)
    1) Supplement – transcriptomics data analysis (Supplementary Figure 1)
    2) Correlation between omics datasets (Figure 2A and B)
    3) Pathway analysis with mBONITA on multiomics data (Figure 2C)
3) Pathway-based prioritization of genes in a signaling network with mBONITA
    1) Node importance score: show a case study of a LSP1/HIF1A-centric signaling network
4) Benchmarking of mBONITA
    1) Rule inference: Supplement - show that mBONITA identifies a smaller rule set from combined omics data than from individual datasets
    2) Pathway analysis:
        1) mBONITA identifies more significant pathways than:
            1) PaintOmics:
            2) LeapR:
            3) ActivePathways:
            4) CAMERA:
            5) BONITA (TBD)

            Show pathways in supplement figures & tables
        2) mBONITA identifies different node importance scores from BONITA:
            1) Supplement – low correlations between node importance score from single omics and multi-omics data (ie, a comparison to mBONITA)

Main figures

1. Figure 1: moBONITA integrates information from multiple omics datasets to learn a consensus set of logic rules for simulation and perturbation of prior knowledge networks
2. Figure 2: mBONITA identifies mechanisms of hypoxia-mediated chemotaxis from a multi-omics dataset from RAMOS B cells grown under three conditions (A) 1505 genes were profiled in all three omics datasets (median log2-abundance > 0) (B) The multi-omics datasets showed low inter-dataset correlations. Distinct experimental conditions are indicated by colors and shapes as shown in the legend. (C) KEGG pathways identified by mBONITA as significantly dyresgulated in the three contrasts. Pathways are defined as differentially regulated if the Benjamini-Hochberg corrected p-value < 0.05.
3. Figure 3: Pathway-based prioritization of genes in a signaling network with mBONITA. Node importance score: show a case study of a LSP1/HIF1A-centric signaling network (TO BE DONE). (A) Network figure (B) Heatmap of node modulation scores. This is a placeholder/draft figure showing node modulation scores for each dataset/contrast combination for the B cell receptor signaling network. NB – this is just IS * RA, need to multiply by std.dev as well.
4. Figure 4: Benchmarking of mBONITA. Numbers of differentially regulated KEGG pathways identified from combination multi-omics data by tested methods in three contrasts (A) 19%O2,CyA- vs 1%O2,CyA- (B) 1%O2,CyA+ vs 1%O2,CyA- (C) 19%O2,CyA- vs 1%O2,CyA+ (D). Pathways known to be involved in the hypoxia-mediated response to CyA, Only pathways identified as significant from a combined dataset by at least one method are shown. Pathways are defined as differentially regulated if the Benjamini-Hochberg corrected p-value < 0.05.

Supplementary Materials

- Supplementary Table 1 Experimental conditions in the three datasets from RAMOS B cells. Conditions that are in all datasets are highlighted in red.
- Suppplementary Figure 1: Transcriptomics analysis of RAMOS B cells grown under three conditions. (A) Numbers of differentially expressed genes identified by DESeq2 in all three contrasts (absolute log2-fold change > 0.5 and Bonferroni-adjusted p-value < 0.05) (B) z-scored RPM values of DE genes identified in all/any contrast. Experimental conditions are indicated by colors as shown in the legend. (C) Over-representation analysis of DE genes in all three contrasts (unadjusted p-value < 0.05). Complete tables of DE genes and over-represented pathways may be found in the Supplementary Data.
- Supplementary Table 2: KEGG Pathways involved in the HIF1A-mediated response of B cells to hypoxia and CyA
- Supplementary Figure 2: Rule inference from all three datasets – (A) Rule set sizes, (B) Importance scores (Spearman correlations between 0.5 and 0.8, p << 0.01). See Supplementary Table 3 for all correlation coefficients.
- Supplementary Table 3: Spearman correlation between importance scores
- Supplementary Figure 3: Comparison of mBONITA-PA to BONITA-PA Numbers of differentially regulated pathways identified from combination multi-omics data by mBONITA in three contrasts (A) 19%O2,CyA- vs 1%O2,CyA- (B) 1%O2,CyA+ vs 1%O2,CyA- (C) 19%O2,CyA- vs 1%O2,CyA+. Pathways are defined as differentially regulated if the Benjamini-Hochberg corrected p-value < 0.05.
- Supplementary Figure 4: Pathway analysis with Bonita. All p-values are Bonferroni-corrected and are < 0.01. The top 10 pathways with the lowest p-values are shown. A complete table of significantly dysregulated pathways may be found in the Supplementary Data. (a) Proteomics (top 10 pathways with the lowest p-values are shown) (b) Phosphoproteomics (top 10 pathways with the lowest p-values are shown) (c) Transcriptomics (top 4 pathways with the lowest p-values are shown) (d) TO BE ADDED. Multiomics network. Contrasts are color-coded as shown in the legend.
- Supplementary File 1: mBONITA-PA results Excel workbook - pvalues_concatenated_20220816
- Supplementary File 2: PaintOmics results (paintomics_allResults.csv)
- Supplementary File 3: leapR results (leapR_allResults.csv)
- Supplementary File 4: CAMERA results (camera_allResults.csv)
- Supplementary File 5: ActivePathways results (activePathways_allResults.csv)
- Supplementary File 6: BONITA results
- Supplementary Figure 5: paintOMICS results
- Supplementary Figure 6: leapR results
- Supplementary Figure 7: ActivePathways
- Supplementary Figure 8: CAMERA + Fisher results

Main figures and tables

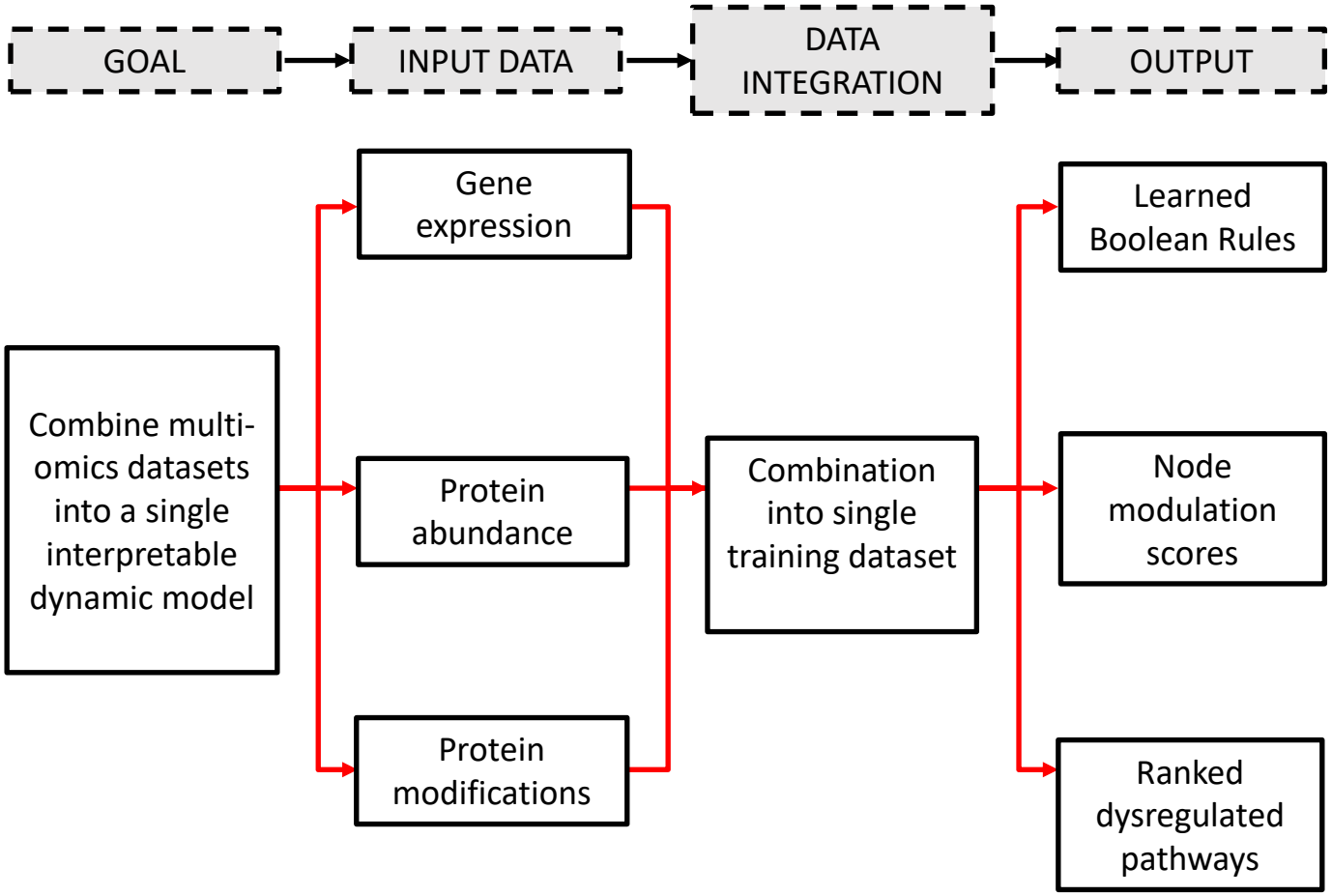


Figure 1 moBONITA integrates information from multiple omics datasets to (A) learn a consensus set of logic rules for a prior knowledge networks (PKNs) (B) simulate and perturb PKNs in silico (C) calculate condition-specific node modulation scores, and (D) perform pathway analysis.

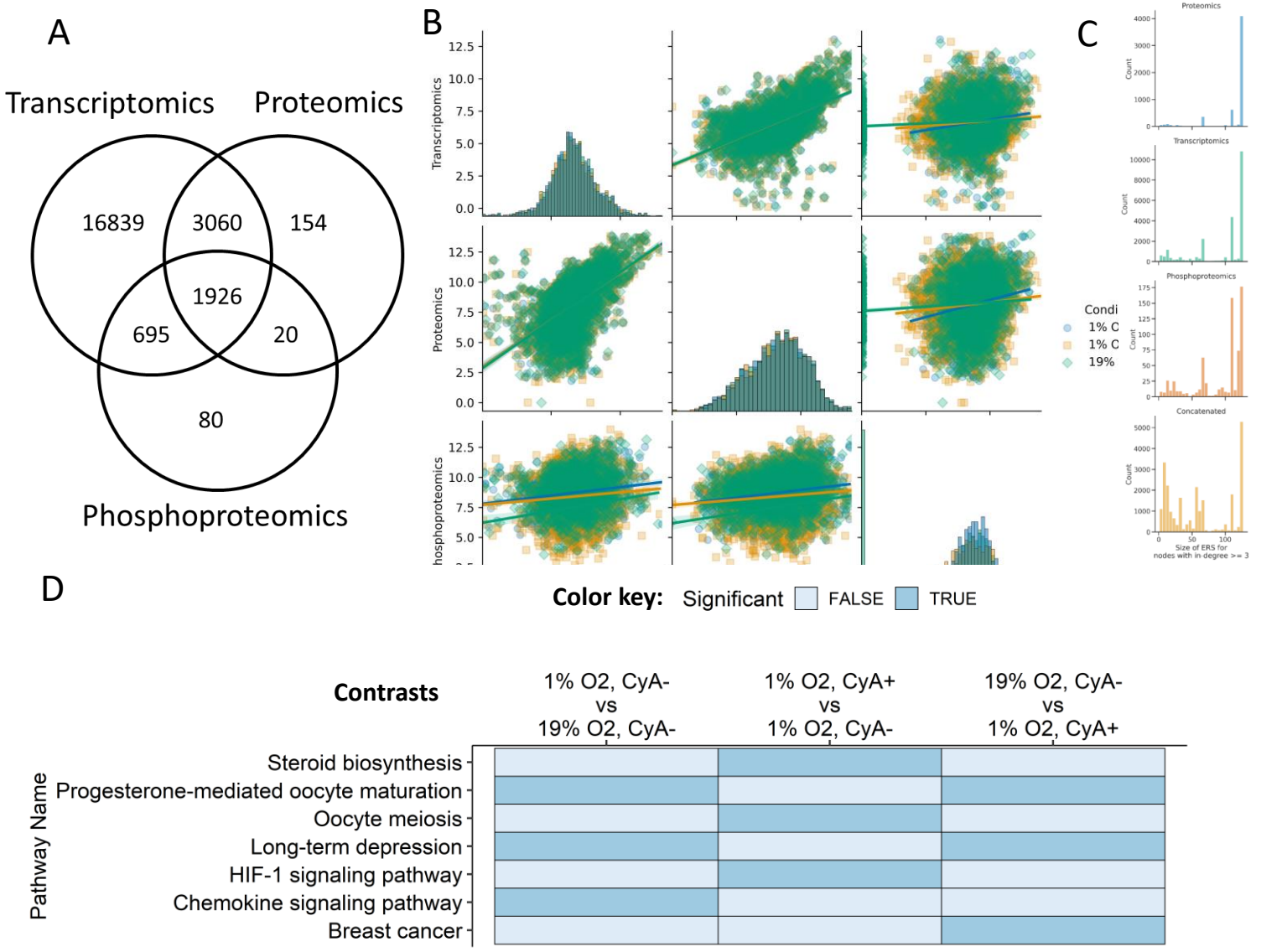


Figure 2: mBONITA identifies mechanisms of hypoxia-mediated chemotaxis from a multi-omics dataset from RAMOS B cells grown under three conditions (A) 1505 genes were profiled in all three omics datasets (median log2-abundance > 0) (B) The multi-omics datasets showed low inter-dataset correlations. Distinct experimental conditions are indicated by colors and shapes as shown in the legend. (C) Rule set sizes (ERS) for each omics dataset and for the concatenated dataset. (D) KEGG pathways identified by mBONITA as significantly dyresgulated in the three contrasts. Pathways are defined as differentially regulated if the Benjamini-Hochberg corrected p-value < 0.05.

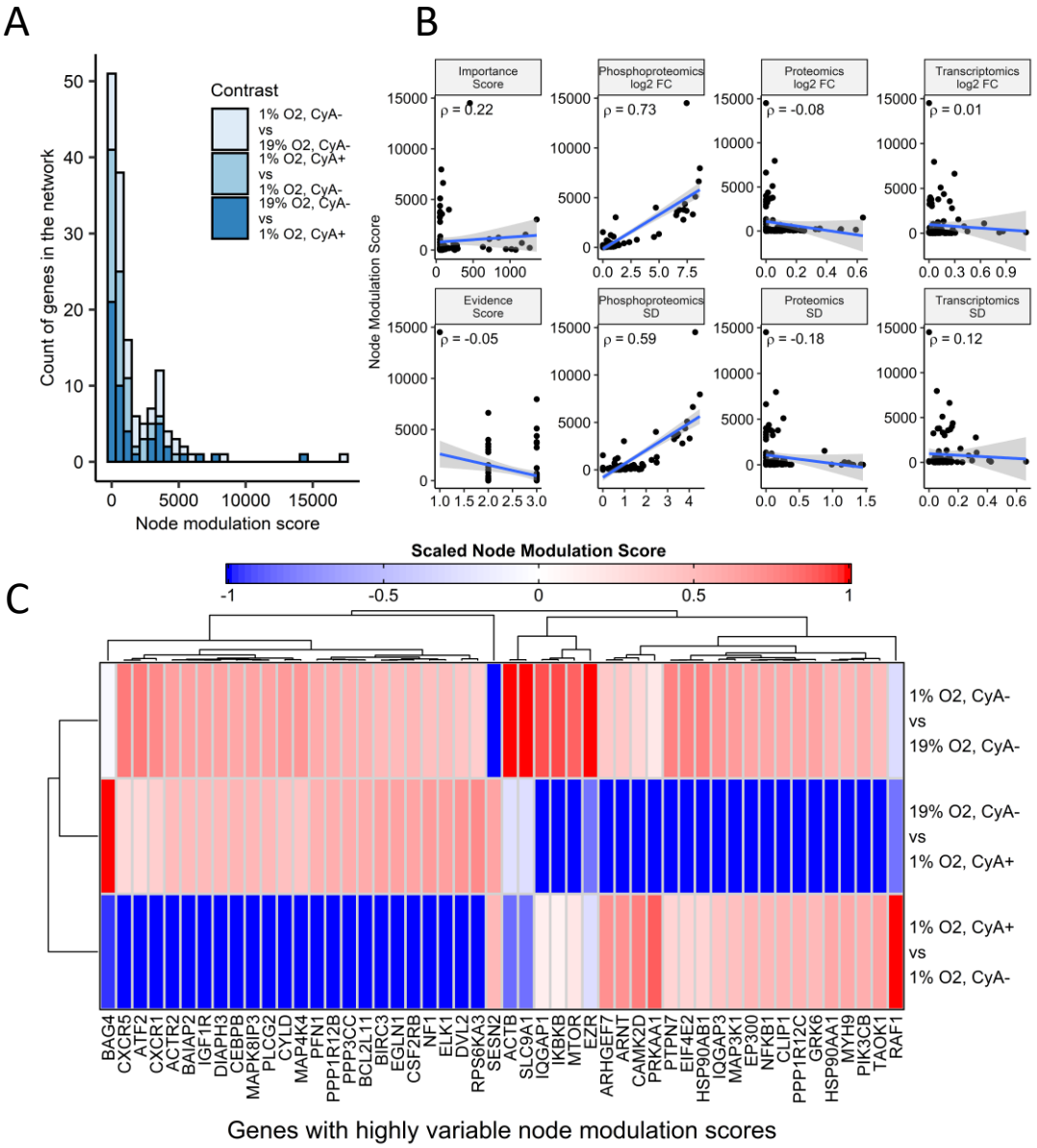


Figure 3: Pathway-based prioritization of genes in a LSP1/HIF1A-centric signaling network with mBONITA. (A) Distribution of node modulation scores (Nm) for this network, for the three tested contrasts. (B) Correlation between Nm and its individual components. (C) Nm for the top 50 most variable genes in the network (in terms of Nm), for the three tested contrasts.

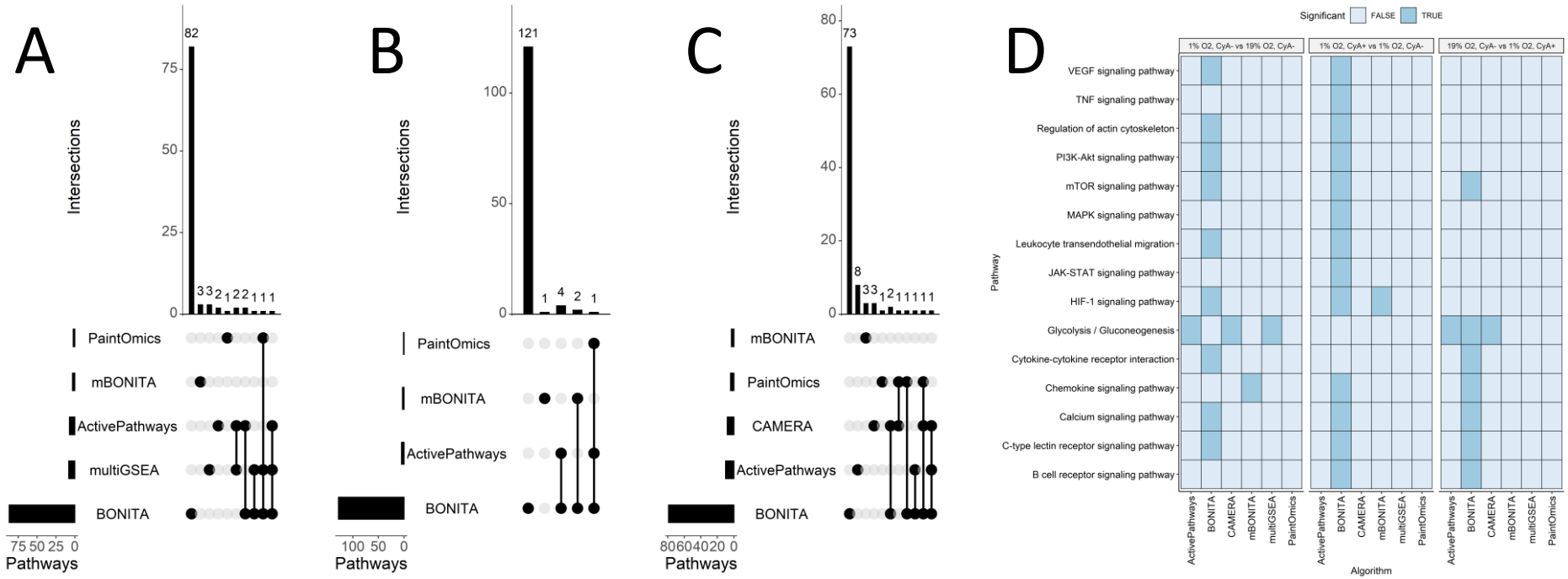
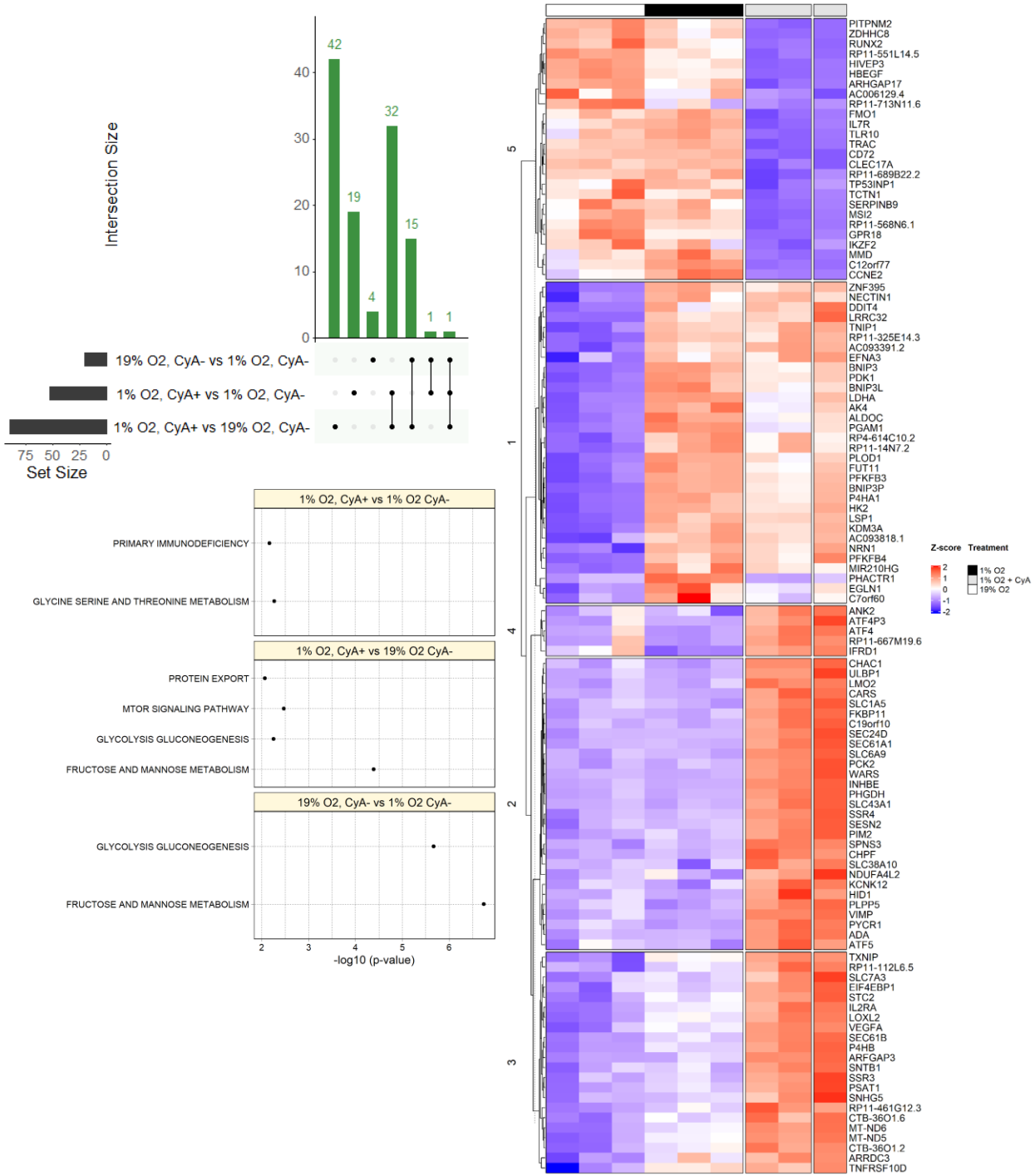


Figure 4: Benchmarking of mBONITA. Numbers of differentially regulated KEGG pathways identified from combination multi-omics data by tested methods in three contrasts (A) 19%O2,CyA- vs 1%O2,CyA- (B) 1%O2,CyA+ vs 1%O2,CyA- (C) 19%O2,CyA- vs 1%O2,CyA+ (D). Pathways known to be involved in the hypoxia-mediated response to CyA, Only pathways identified as significant from a combined dataset by at least one method are shown. Pathways are defined as differentially regulated if the Benjamini-Hochberg corrected p-value < 0.05.

Supplementary figures and tables

Data type	Oxygen	CyA	CXCL12
Proteomics	19%	"-"	"-"
	1%	"-"	"-"
	1%	"+"	"-"
Transcriptomics	19%	"-"	"-"
	19%	"+"	"-"
	1%	"-"	"-"
	1%	"+"	"-"
Phosphoproteomics	19%	"+"	"+"
	19%	"-"	"+"
	19%	"+"	"-"
	19%	"-"	"-"
	1%	"+"	"+"
	1%	"-"	"+"
	1%	"+"	"-"
	1%	"-"	"-"

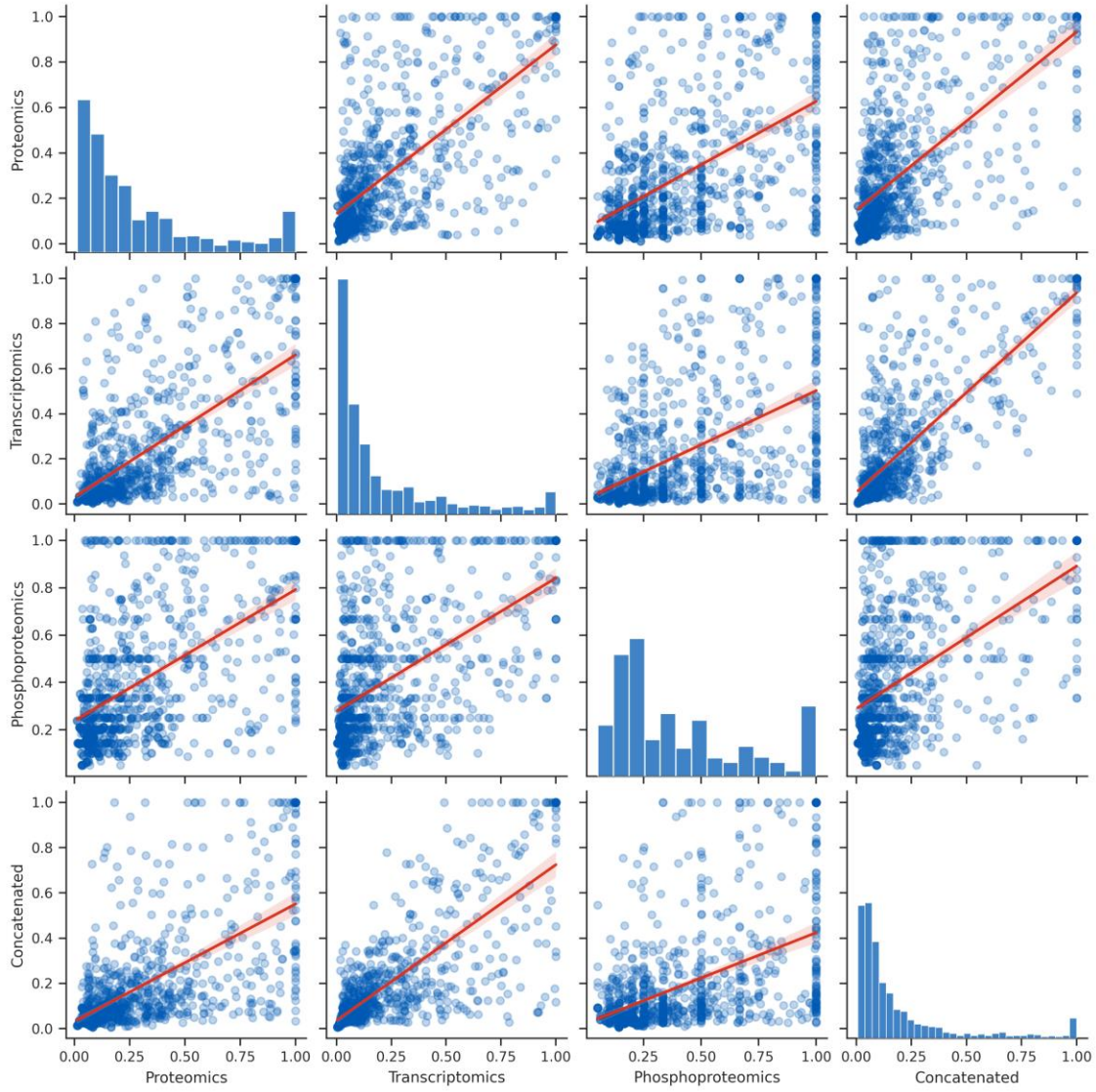
Supplementary Table 1 Experimental conditions in the three datasets from RAMOS B cells. Conditions that are in all datasets are highlighted in red.



Suppplementary Figure 1: Transcriptomics analysis of RAMOS B cells grown under three conditions. (A) Numbers of differentially expressed genes identified by DESeq2 in all three contrasts (absolute log2-fold change > 0.5 and Bonferroni-adjusted p-value < 0.05) (B) z-scored RPM values of DE genes identified in all/any contrast. Experimental conditions are indicated by colors as shown in the legend. (C) Over-representation analysis of DE genes in all three contrasts (unadjusted p-value < 0.01). Complete tables of DE genes and over-represented pathways may be found in Supplementary Tables 2 and 3 respectively.

Supplementary Table 2: KEGG Pathways involved in the HIF1A-mediated response of B cells to hypoxia and CyA

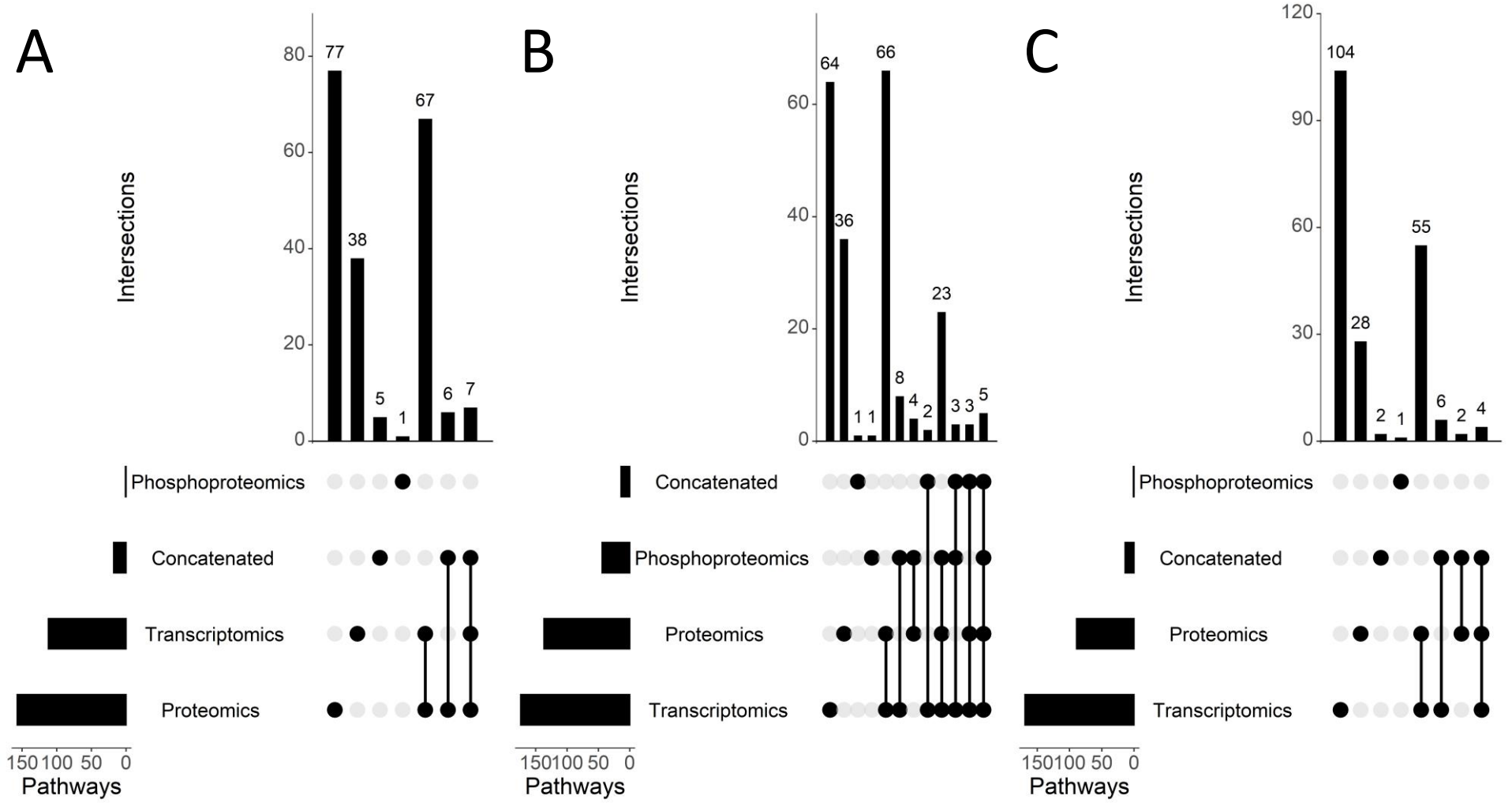
	KEGG Pathway	KEGG Code
1	MAPK signaling	hsa04010
2	Chemokine signaling	hsa04062
3	NF-kappa B signaling	hsa04064
4	HIF-1 signaling	hsa04066
5	mTOR signaling	hsa04150
6	PI3K-Akt signaling	hsa04151
7	VEGF signaling	hsa04370
8	Cell adhesion	hsa04514
9	C-type lectin receptor signaling	hsa04625
10	JAK-STAT signaling	hsa04630
11	TNF signaling	hsa04668
12	Leukocyte transendothelial migration	hsa04670
13	Regulation of actin cytoskeleton	hsa04810



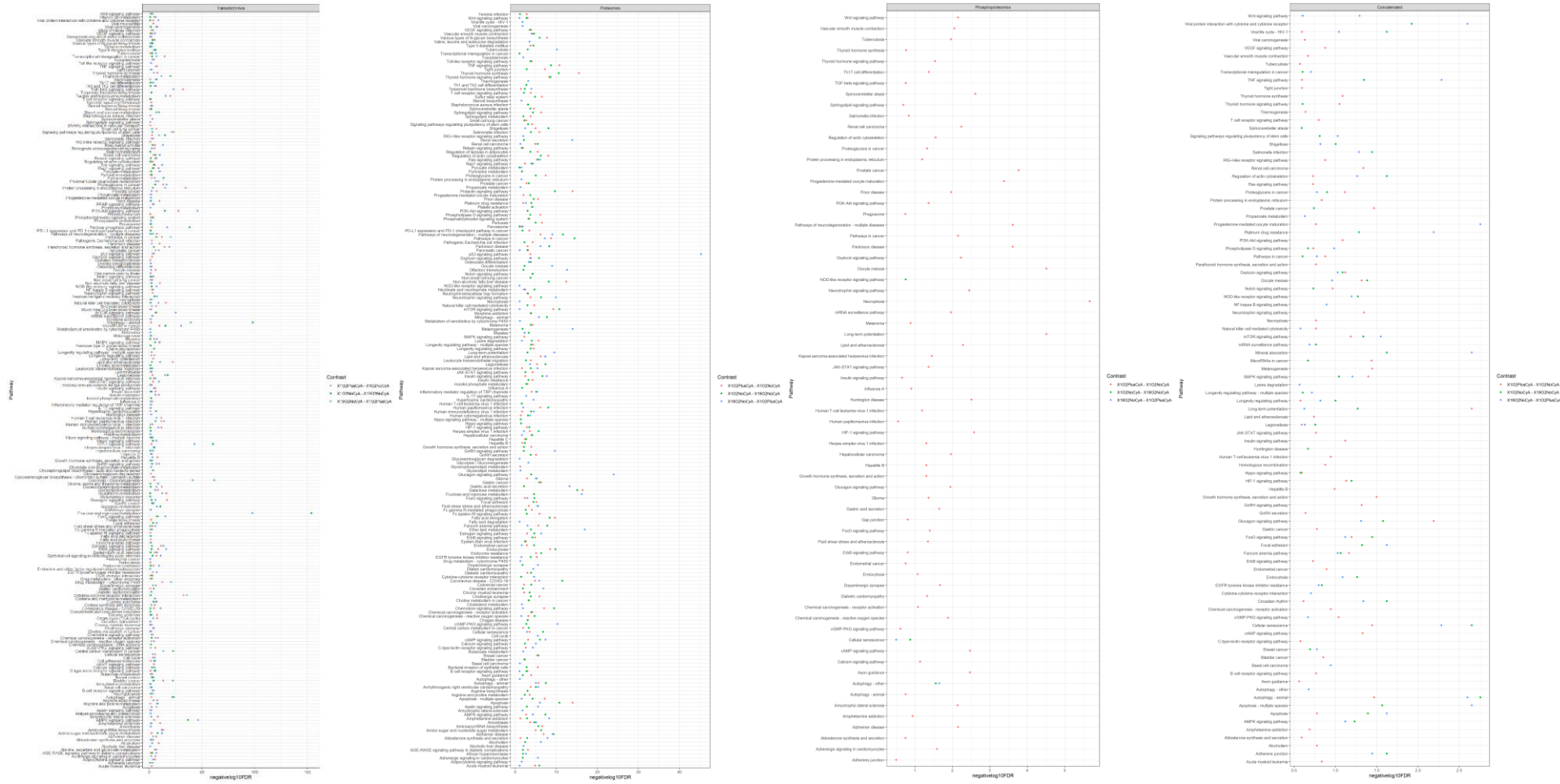
Supplementary Figure 2: Rule inference from all three datasets Importance scores (Spearman correlations between 0.5 and 0.8, $p << 0.01$). See Supplementary Table 3 for all correlation coefficients.

Supplementary Table 3: Spearman correlation between importance scores

Dataset	Concatenated	Phosphoproteomics	Proteomics	Transcriptomics
Concatenated	1	0.494465	0.661879	0.796354
Phosphoproteomics	0.494465	1	0.539542	0.532783
Proteomics	0.661879	0.539542	1	0.707195
Transcriptomics	0.796354	0.532783	0.707195	1



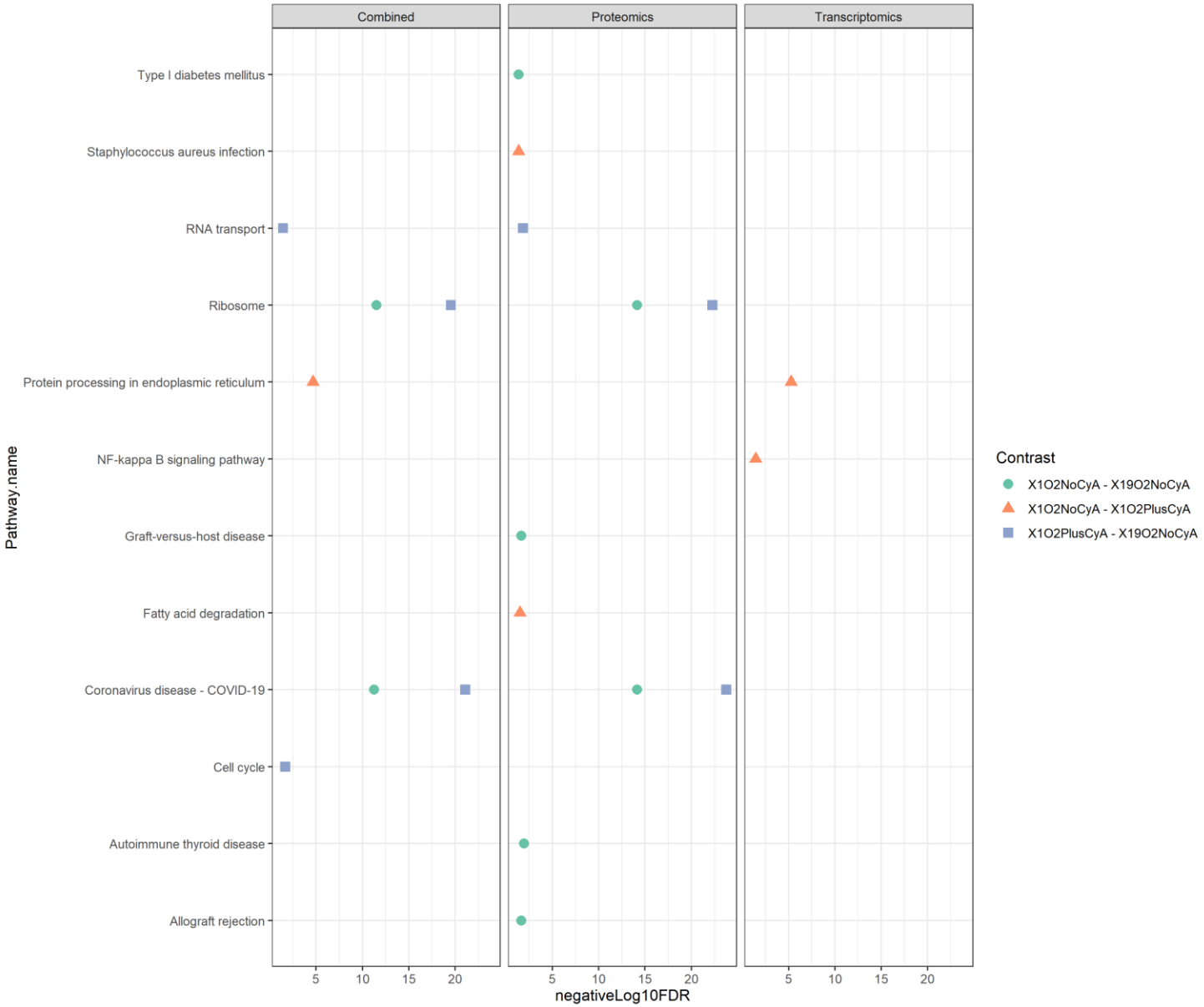
Supplementary Figure 3: Comparison of mBONITA-PA to BONITA-PA Numbers of differentially regulated pathways identified from combination multi-omics data by mBONITA in three contrasts (A) 19%O2,CyA- vs 1%O2,CyA- (B) 1%O2,CyA+ vs 1%O2,CyA- (C) 19%O2,CyA- vs 1%O2,CyA+. Pathways are defined as differentially regulated if the Benjamini-Hochberg corrected p-value < 0.05.



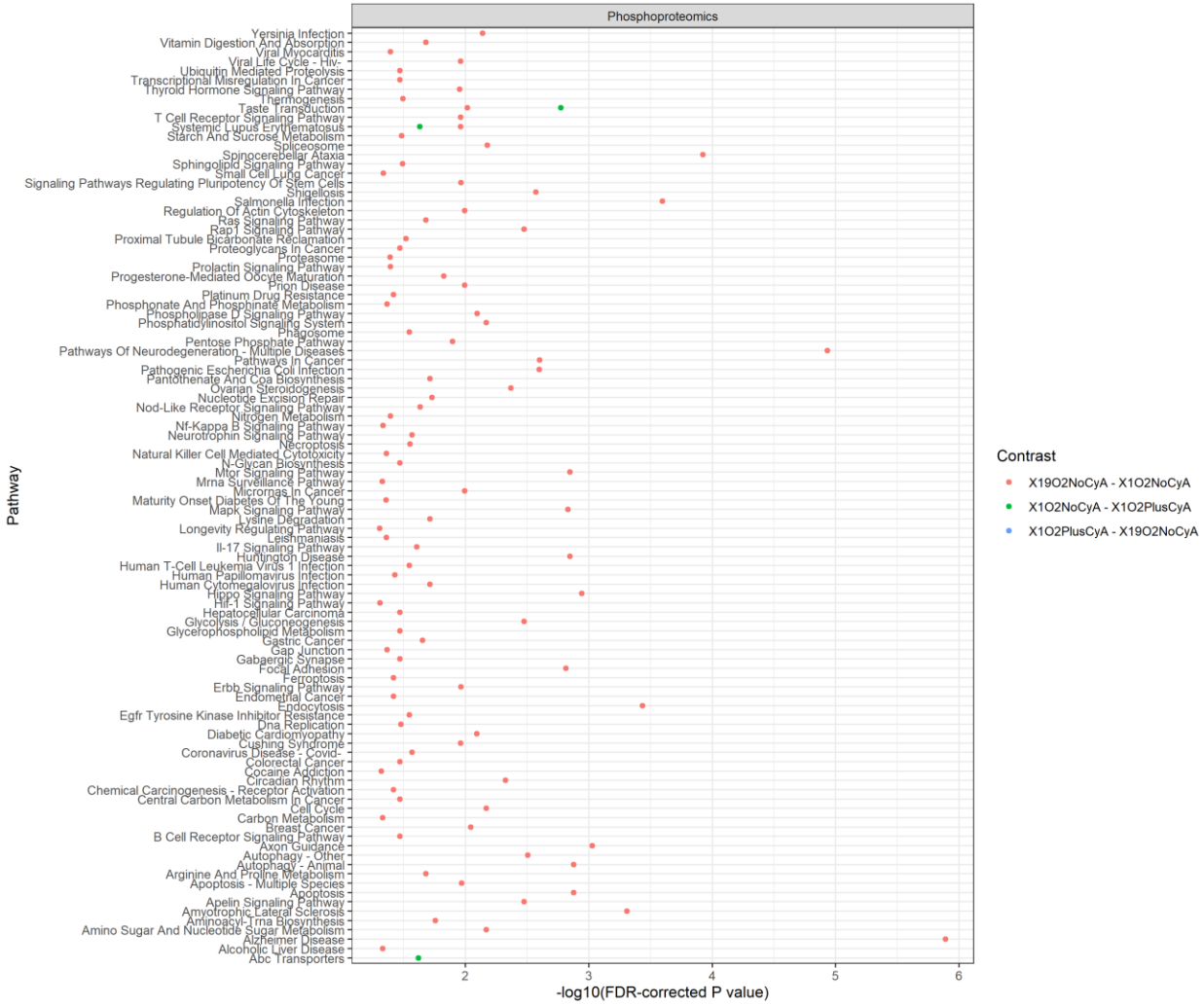
Supplementary Figure 4: Pathway analysis with Bonita and mBonita. All p-values are Bonferroni-corrected and are < 0.01. The top 10 pathways with the lowest p-values are shown. A complete table of significantly dysregulated pathways may be found in the Supplementary Data (A) Transcriptomics (B) Proteomics (C) Phosphoproteomics (D) mBonita

- Supplementary File 1: mBONITA-PA results Excel workbook - pvalues_concatenated_20220816
- Supplementary File 2: PaintOmics results (paintomics_allResults.csv)
- Supplementary File 3: leapR results (leapR_allResults.csv)
- Supplementary File 4: CAMERA results (camera_allResults.csv)
- Supplementary File 5: ActivePathways results (activePathways_allResults.csv)
- Supplementary File 6: BONITA results
- Supplementary File 7: multiGSEA results (multiGSEA_allResults.csv)

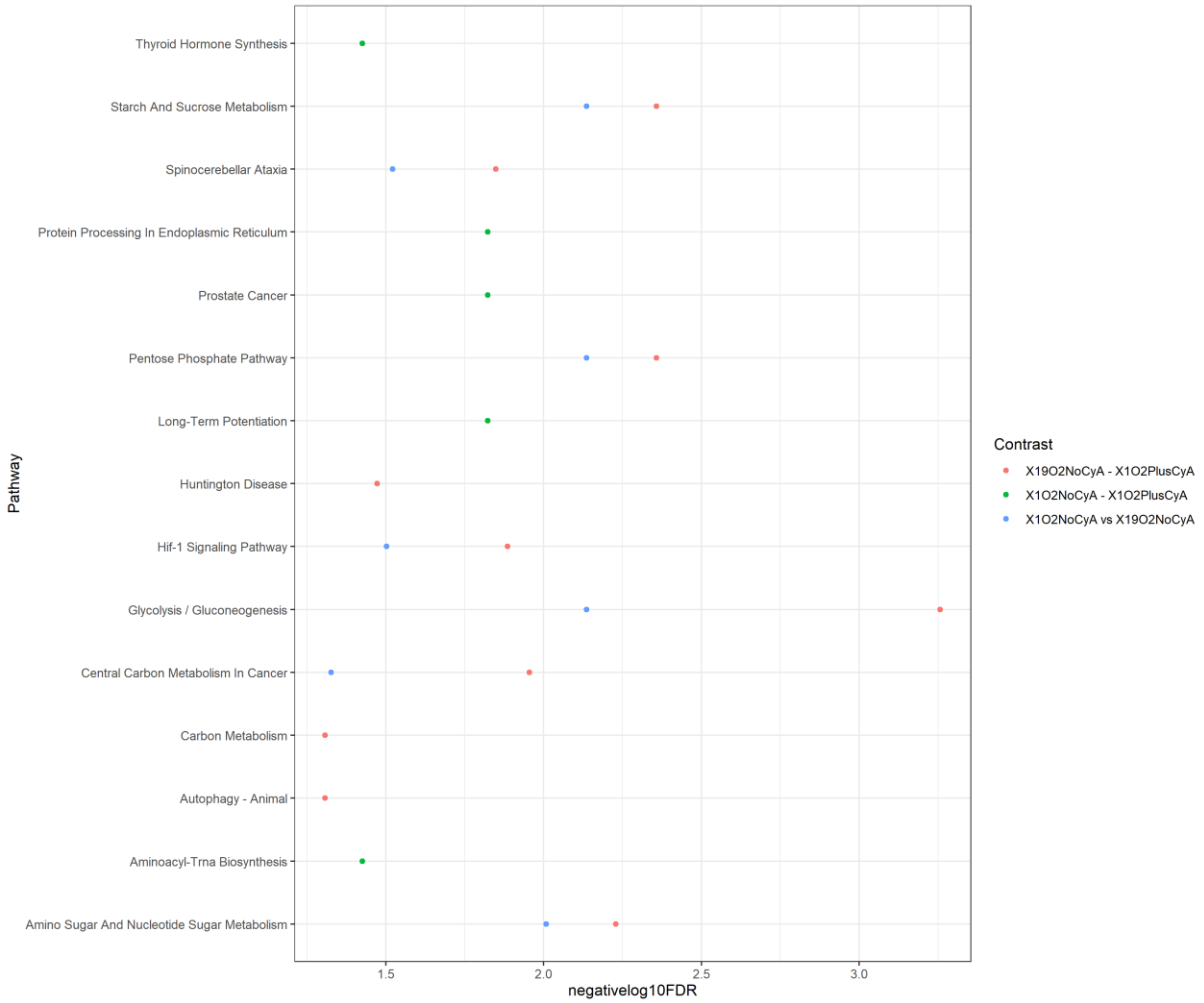
Supplementary Figure 5: paintOMICS results



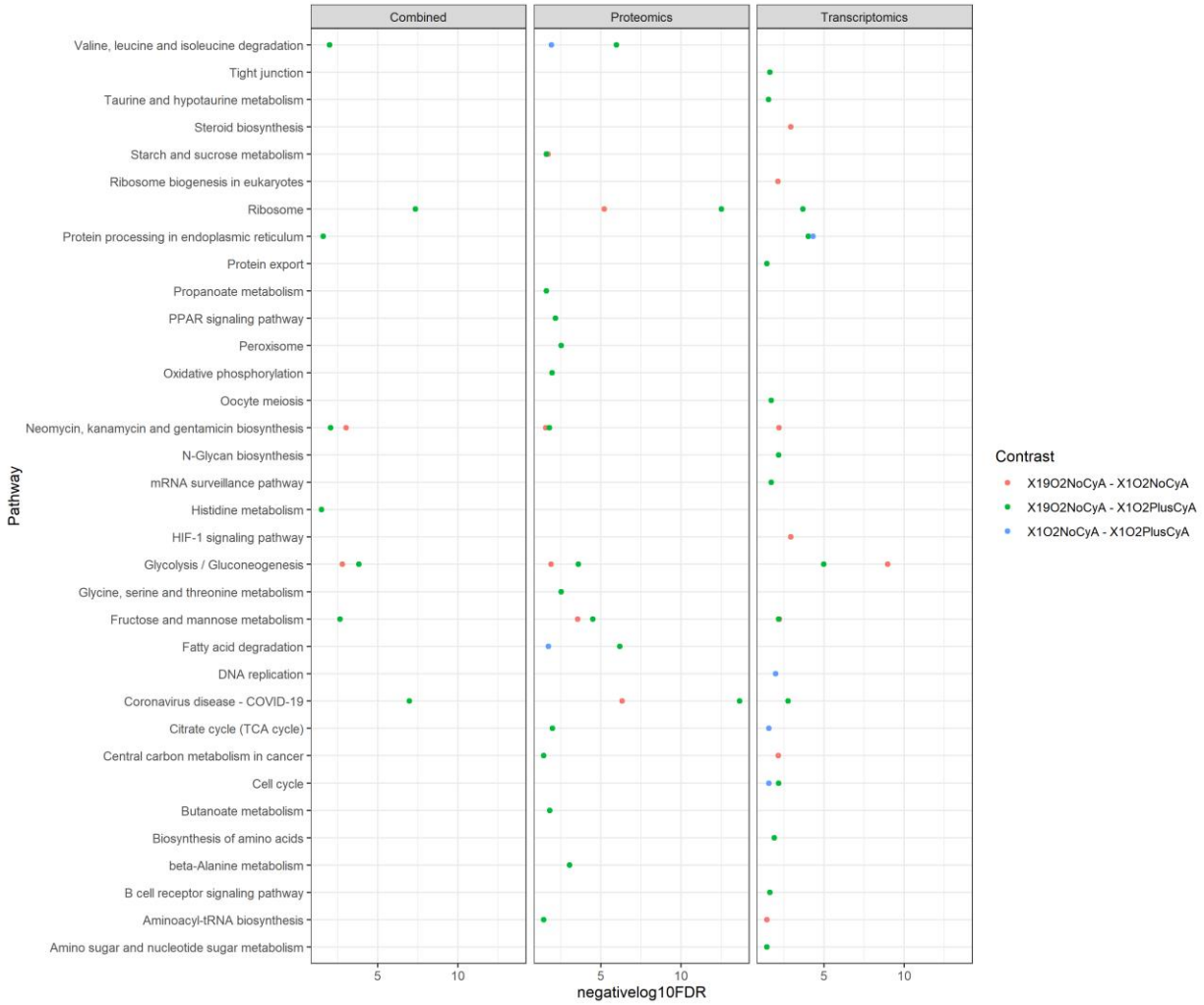
Supplementary Figure 6: leapR results



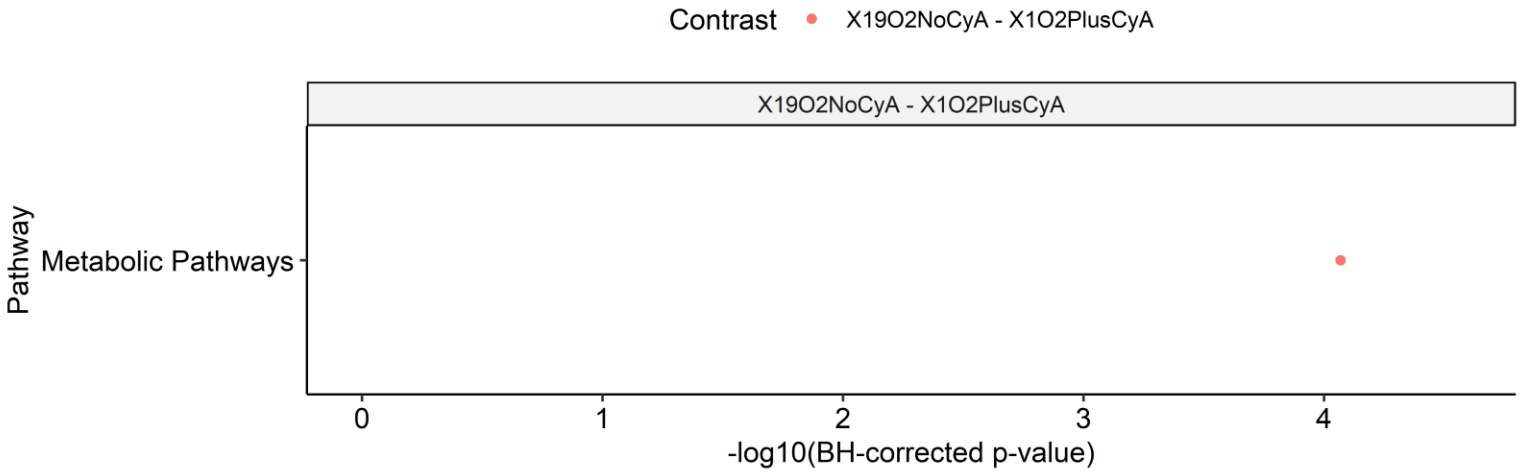
Supplementary Figure 7: ActivePathways



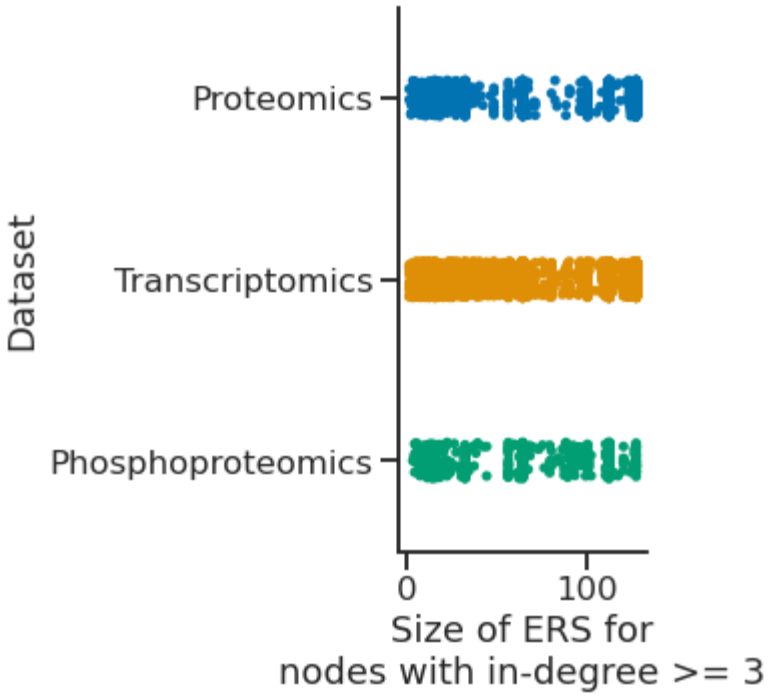
Supplementary Figure 8: CAMERA + Fisher results

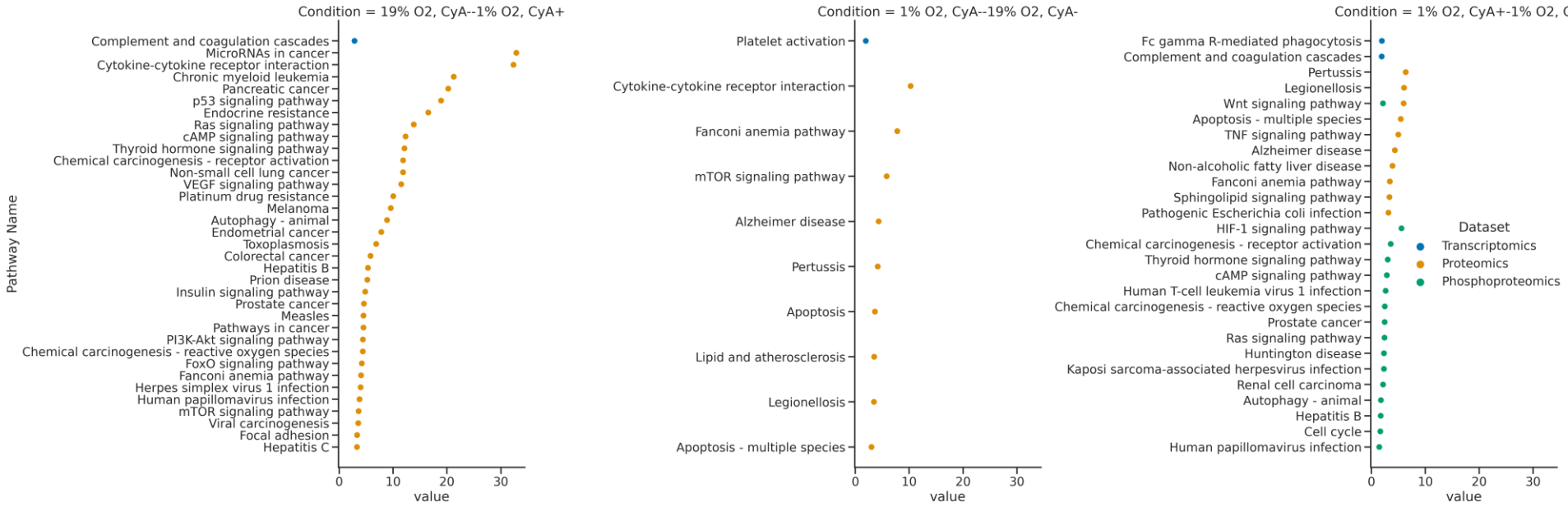


Supplementary Figure 9: multiGSEA results

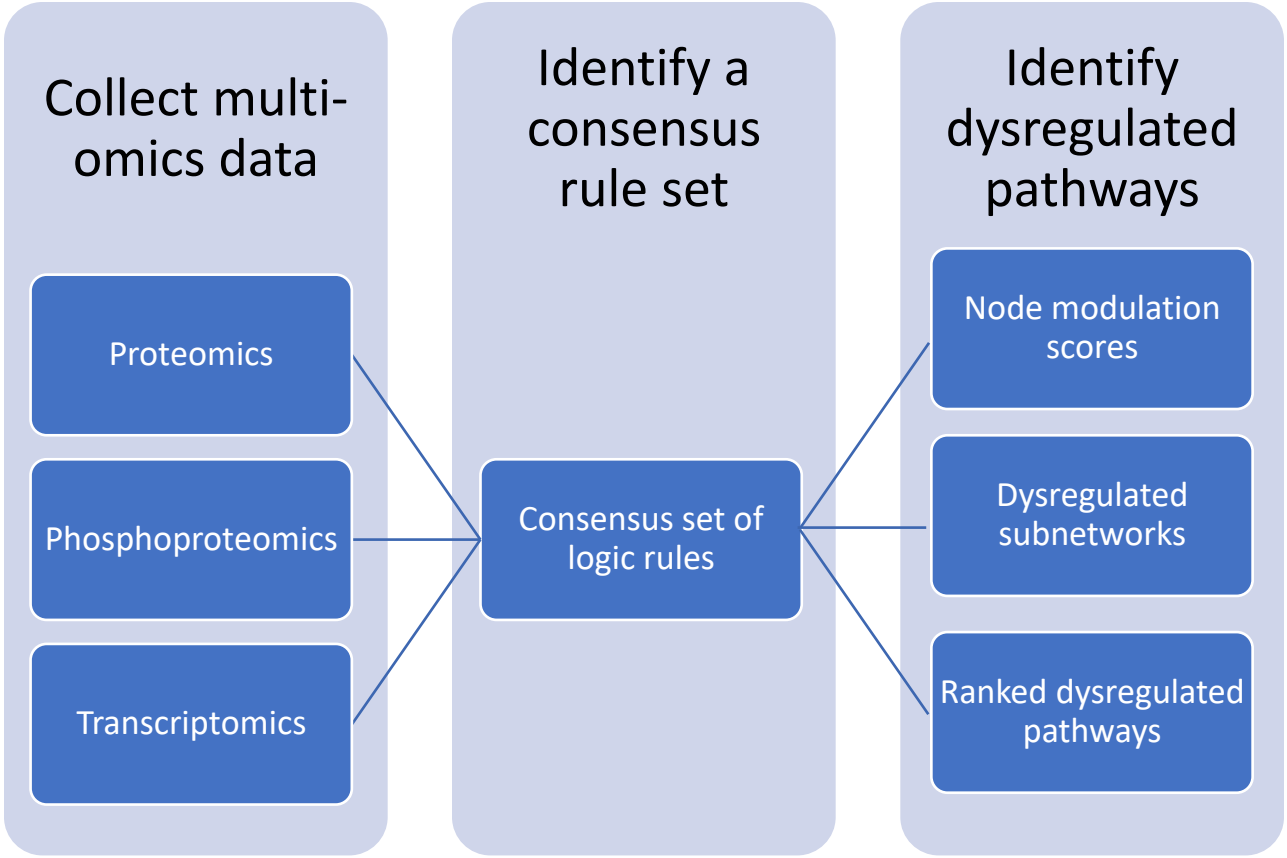


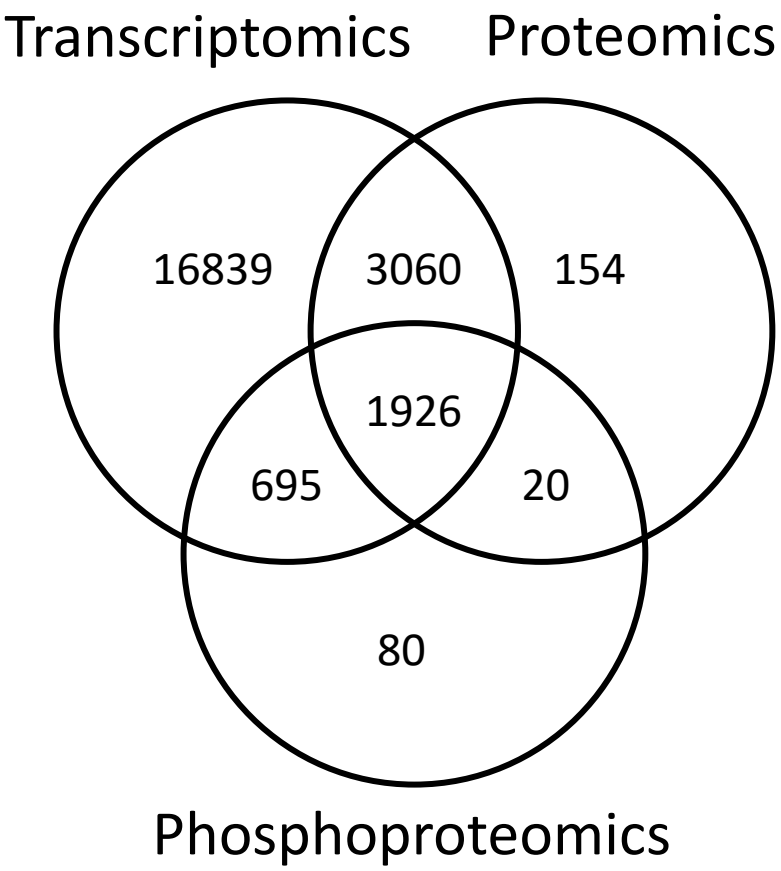
Drafts of figures

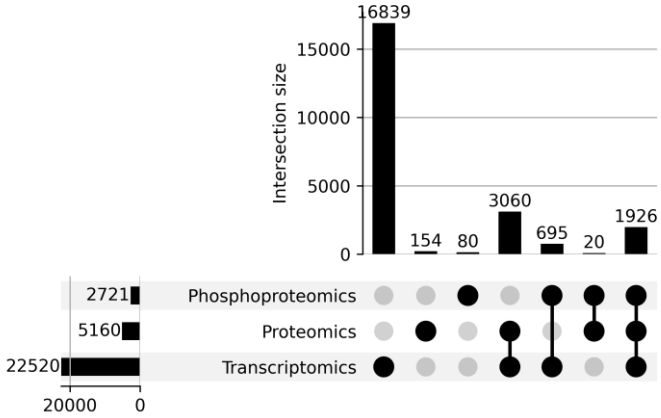

Proteomics
Transcriptomics
Phosphoproteomics
Dataset
0
100
Size of ERS for
nodes with in-degree >= 3

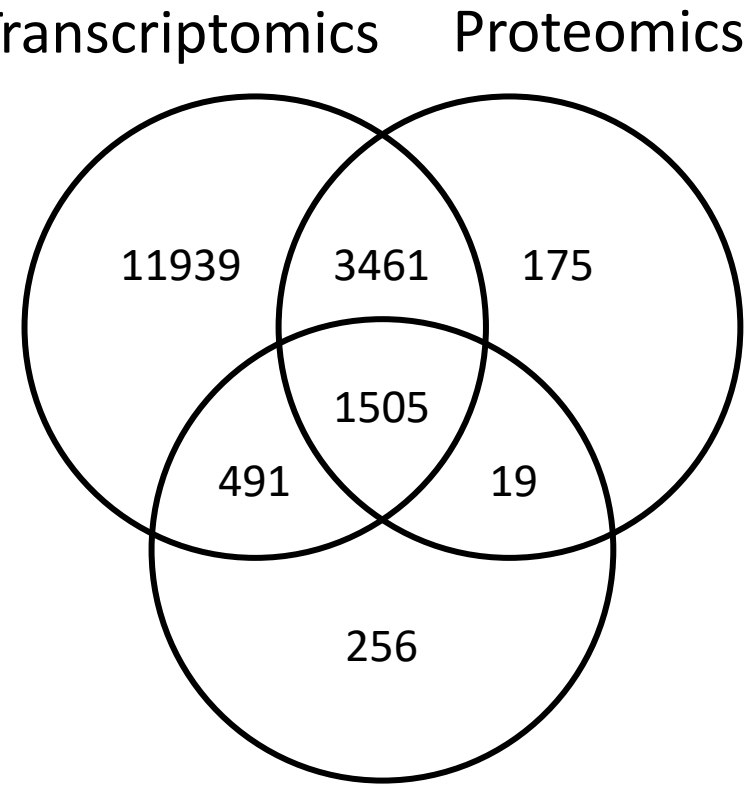


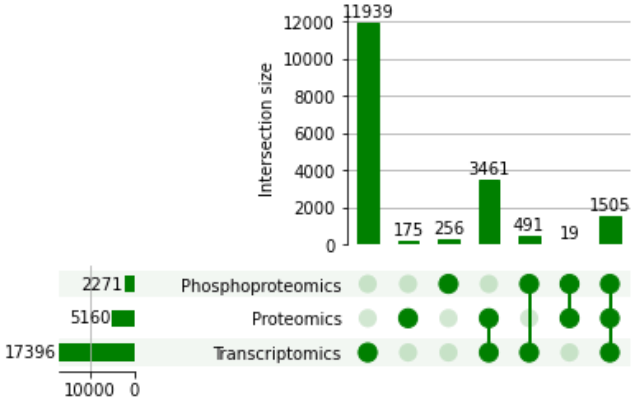
Alternate for Figure 5: Pathway analysis with Bonita (a) Proteomics (Bonferonni-adjusted p value < 0.001) (b) Transcriptomics (unadjusted p value < 0.05) (c) Phosphoproteomics (unadjusted p value < 0.05). (d) TO BE ADDED. Multiomics network. Contrasts are color-coded as shown in the legend.


Collect multi-omics data
Proteomics
Phosphoproteomics
Transcriptomics
Identify a consensus rule set
Consensus set of logic rules
Identify dysregulated pathways
Node modulation scores
Dysregulated subnetworks
Ranked dysregulated pathways


Transcriptomics
Proteomics
16839
3060
154
1926
695
20
80
Phosphoproteomics


Intersection size
16839
15000
10000
5000
0
154
80
3060
695
20
1926
2721
5160
22520
Phosphoproteomics
Proteomics
Transcriptomics
20000
0


Transcriptomics
Proteomics
11939
3461
175
1505
491
19
256
Phosphoproteomics


Intersection size
12000
10000
8000
6000
4000
2000
0
11939
175
256
3461
491
19
1505
2271
5160
17396
Phosphoproteomics
Proteomics
Transcriptomics
10000
0

OLD

C

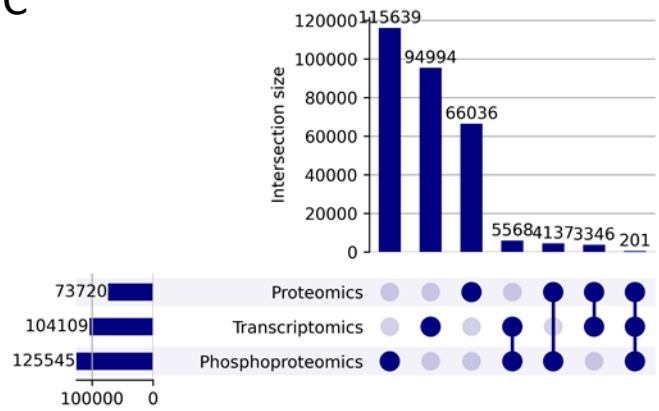
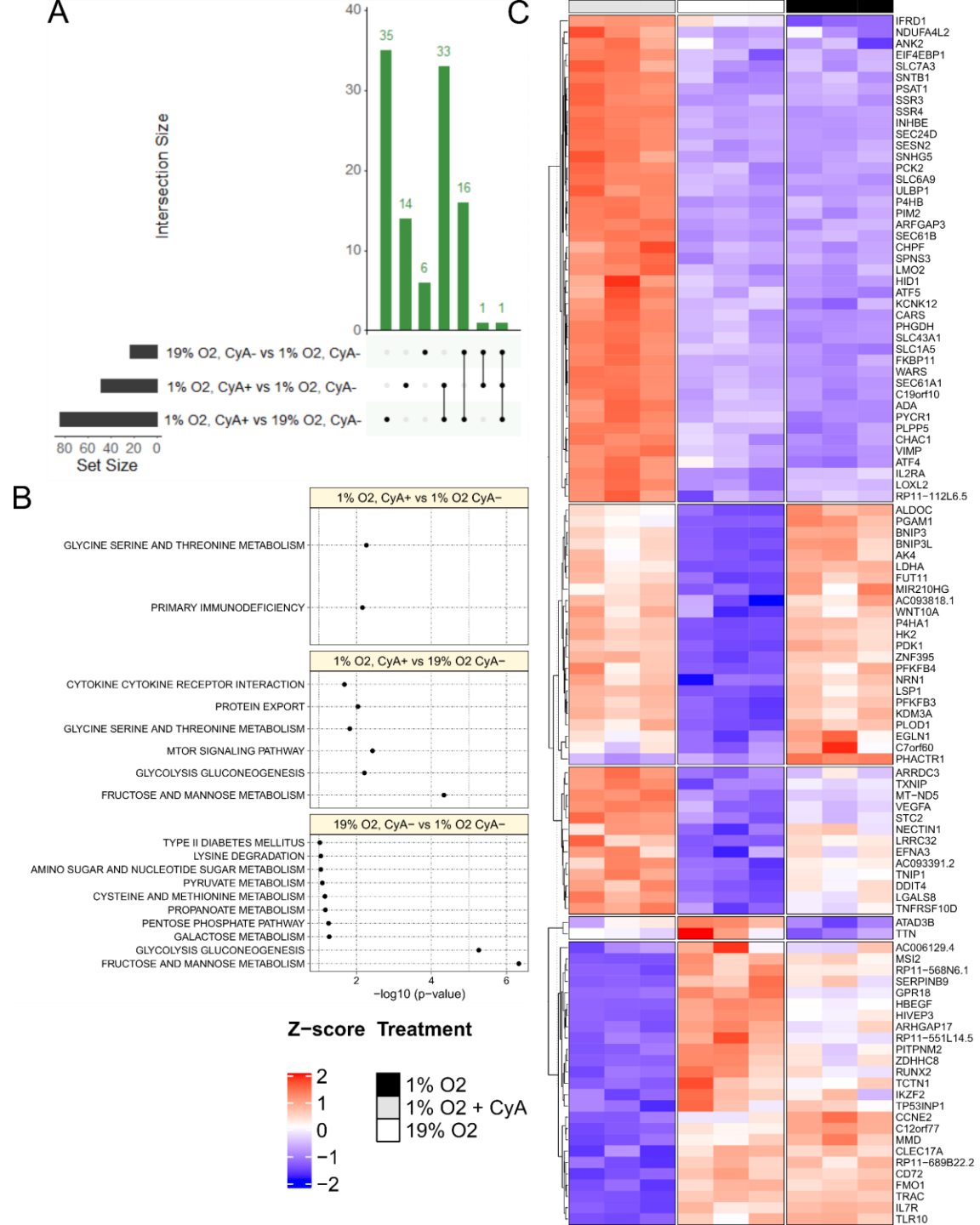


Figure 2, part 2: (C) Correlation networks constructed using three multi-omics datasets showed little overlap between edges. (D) The genes involved in the overlapping edges between datasets were enriched for gene sets related to x,y,z. Enrichment analysis for other intersections are shown in the supplement.



Suppplementary Figure 1 (old): Transcriptomics analysis of RAMOS B cells grown under three conditions. (A) Numbers of differentially expressed genes identified by DESeq2 in all three contrasts (absolute log2-fold change > 0.5 and Bonferroni-adjusted p-value < 0.05) (B) z-scored RPM values of DE genes identified in all/any contrast. Experimental conditions are indicated by colors as shown in the legend. (C) Over-representation analysis of DE genes in all three contrasts (unadjusted p-value < 0.05). Complete tables of DE genes and over-represented pathways may be found in the Supplementary Data.

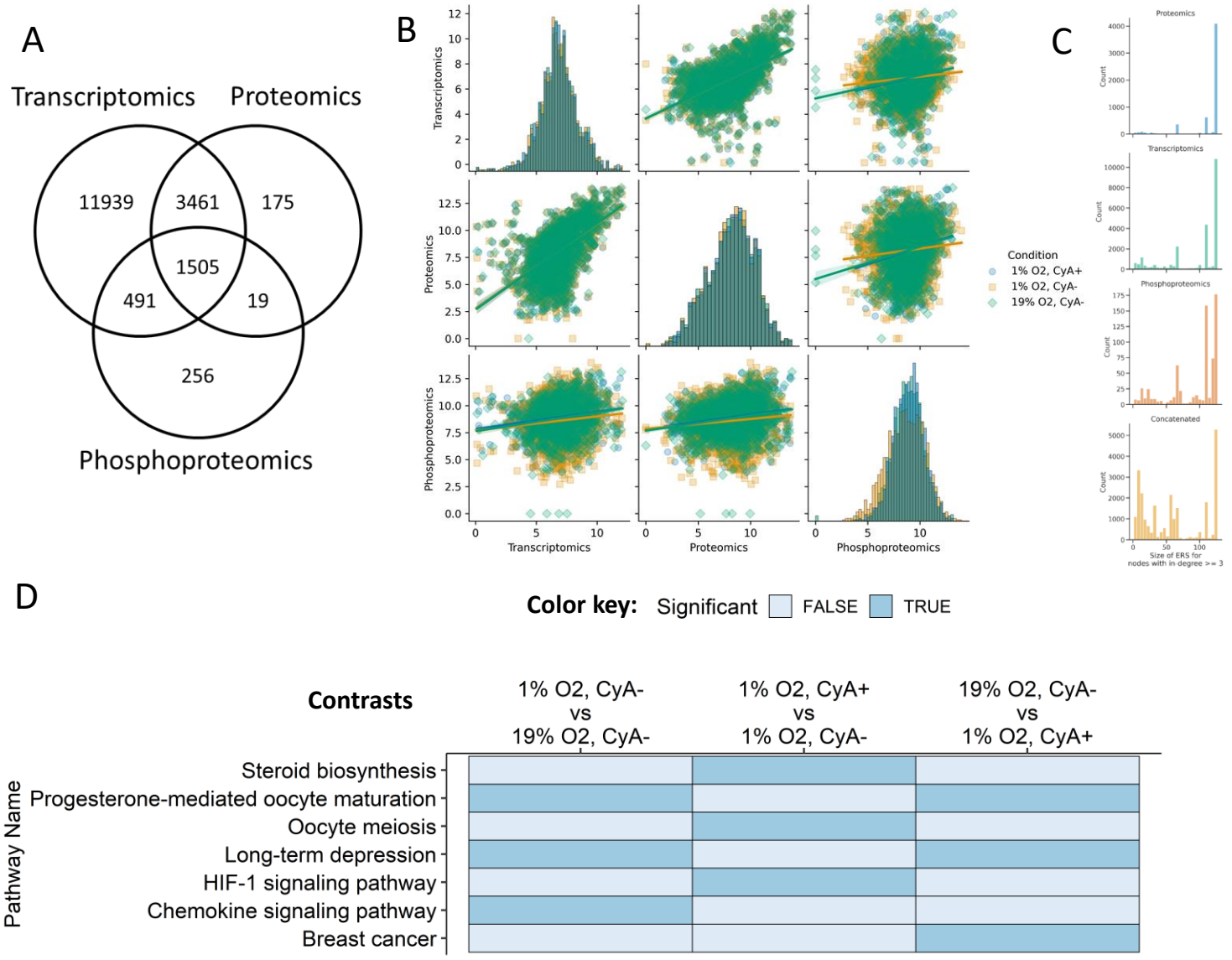
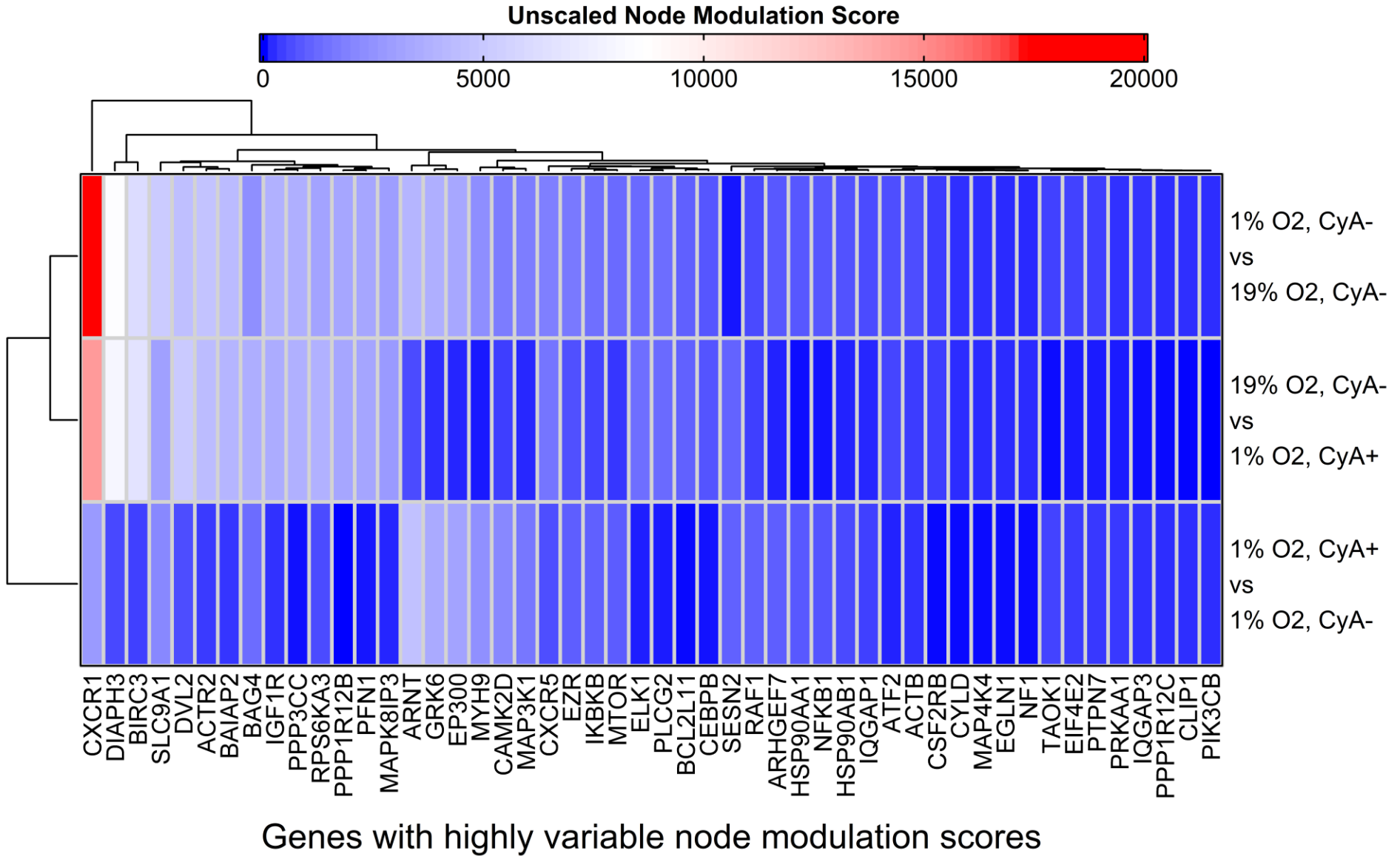
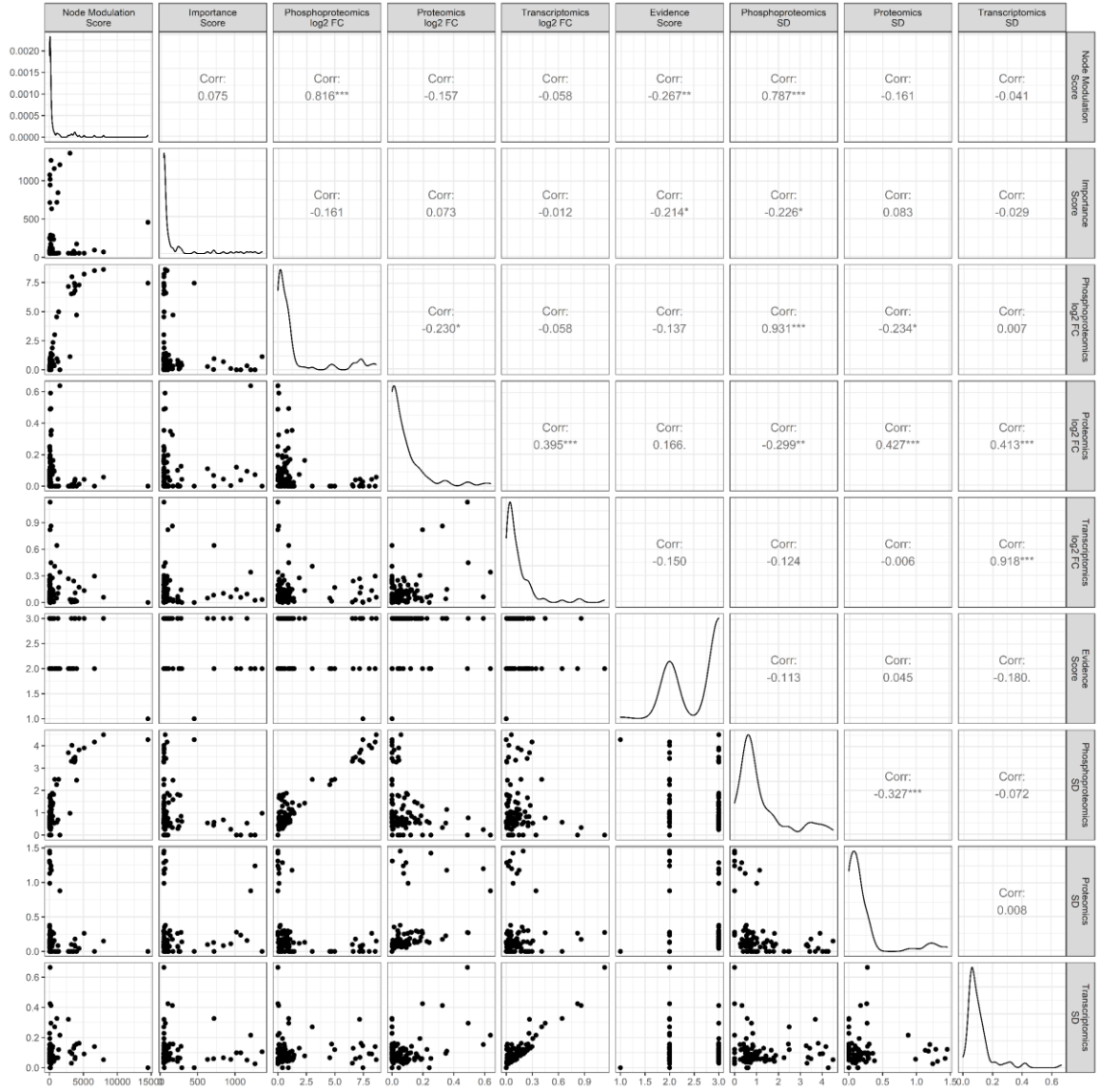


Figure 2: mBONITA identifies mechanisms of hypoxia-mediated chemotaxis from a multi-omics dataset from RAMOS B cells grown under three conditions (A) 1505 genes were profiled in all three omics datasets (median log2-abundance > 0) (B) The multi-omics datasets showed low inter-dataset correlations. Distinct experimental conditions are indicated by colors and shapes as shown in the legend. (C) Rule set sizes (ERS) for each omics dataset and for the concatenated dataset. (D) KEGG pathways identified by mBONITA as significantly dyresgulated in the three contrasts. Pathways are defined as differentially regulated if the Benjamini-Hochberg corrected p-value < 0.05.



Genes with highly variable node modulation scores



Correlation between and distributions of components of the node modulation score