Executable network models of integrated multi-omics data

**Mukta G. Palshikar1, Xiaojun Min2, Alexander Crystal2, Jiayue Meng, Shannon Hilchey3, Martin Zand3,4, Juilee Thakar1,4,5,6,7\***

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

# Abstract

* Integrate 3 multi-omics datasets from the same system to create multi-layered networks for which Boolean rules can be inferred.
* **Network-based integration of multiomics data suggests a mechanism for HIF-1α-mediated chemotaxis in B cells.**

# Introduction

Modern sequencing technology allows biologists to study conditions at different molecular levels by generating ‘omics’ datasets such as transcriptomics (mRNA levels), proteomics (protein levels) and phosphoproteomics (phosphoprotein levels). These datasets are extraordinarily rich and allow sophisticated inferences about molecular signaling individually. In particular, pathway enrichment analysis allows the identification of dysregulated biological processes by two classes of methods (1, 2) – over-representation analysis and functional class scoring, or topology-based pathway enrichment analysis. However, technical and biological variability between these layered datasets present challenges for integrative computational analyses.

Discrete-state modeling characterizes network topologies with Boolean rules or gates that define signal flow through the network. These networks can be simulated either synchronously or asynchronously to identify limit cycles or attractors that correspond to network-specific phenotypes. We have recently published two algorithms that infer regulatory rules for PKNs from omics data (3, 4). These inferred regulatory rules are used to perturb and simulate networks *in silico* to calculate the influence of nodes over signaling through the network. This perturbation-based scores are combined with expression data to perform pathway analysis. However, both these methods rely on information from a single omics training dataset to perform both rule inference and pathway analysis.

Here, we present a method multi-omics Boolean Omics Network Invariant Time Analysis (mBONITA) to (a) use multiple layers of omics data to improve inference of regulatory rules, (b) use expression from all these layers to calculate node importance scores and identify condition-related subnetworks and (c) perform pathway analysis that incorporates information from multiple molecular layers. We demonstrate the utility of this algorithm on a multi-omics dataset from RAMOS B cells grown under hypoxic conditions and treated with cyclosporine A (CyA) and CXCL12. This dataset consists of three layers – transcriptomics, proteomics, and phosphoproteomics. We have previously published the analysis and validation of the proteomics (5) and phosphoproteomics dataset (*in preparation*) and present the conventional analysis of the transcriptomics dataset in this manuscript. These datasets have limited correlation in accordance with previous studies. We show that our method can effectively use this multi-omics dataset in combination with PKNs from KEGG (6) and WikiPathways (7, 8) to (a) identify an accurate and reliable regulatory rule set for PKNs, (b) use this improved rule set to calculate node modulation scores that incorporate all available expression information and the network topology and (c) use these modulation scores to identify dysregulated pathways that suggest a mechanism for hypoxia-mediated chemotaxis in B cells. We also demonstrate this on a multi-omics dataset from persons with renal cancer (9). We used mBONITA to identify pathways that are significantly dysregulated in the three contrasts in our multi-omics dataset, including pathways that are not significantly dysregulated in individual datasets. We compare these pathways to those identified by our competitor algorithms PaintOmics4 (10, 11), CAMERA (12) in combination with Fisher’s method of p-value combination as suggested in ReactomeGSA (13), LeapR (14) , and ActivePathways (15), and show that mBONITA identifies the most relevant pathways to these conditions. We use mBONITA to calculate node modulation scores for a large signaling network describing the HIF1A-mediated signaling in B cells and identify that are highly modulated in this condition. We show that the genes identified by mBONITA show improved condition specificity and contain strong candidates for experimental validation.

In this manner, we show here that mBONITA uses a Boolean rule-based network propagation method that integrates multiple sources of omics/perturbation data to prioritize nodes in an interaction network, can identify subnetworks of highly modulated genes, and can perform pathway analysis using these multiple sources of omics data to present a complete picture of dysregulated signaling in the condition under study.

# Results

## Multi-omics network modeling and pathway enrichment analysis with mBONITA

***Corresponds to Figure 1***

1. Learn Boolean rules for PKNs from combined omics datasets
2. Simulate networks with combined omics datasets and calculate importance scores
3. Use the new importance score with RA from all three datasets to calculate pathway modulation scores/pvalues

## Pathway-based prioritization of genes in a signaling network with mBONITA

***Corresponds to Figure 2***

Outline:

Rule inference – mBONITA identifies a small set of rules from combined omics data

Node importance scores – mBONITA prioritizes genes according to their influence over signal flow

Case study – LSP1/HIF1A centric network, node importance scores,

We used BONITA to identify Boolean rules for all three datasets as described in the Methods. Boolean rule inference was more accurate for some molecular layers than for others. Figure 2A shows that the average size of the equivalent rule set (ERS) for nodes with in-degree >= 3, which is used as a proxy for BONITA’s ability to narrow down the state space of Boolean rules, was significantly different for the same networks between different datasets (t-test, adjusted p-value < 0.05). We also compared the node importance scores between datasets and found that there was a similarly low correlation between them (Figure 2B). Both these observations underscore the difference in signaling information that can be obtained from these molecular layers and the importance of a meaningful combination of abundance measures to learn node importance scores.

## mBONITA identifies mechanisms of hypoxia-mediated chemotaxis in RAMOS B cells

## mBONITA outperforms competitor algorithms on two test datasets

* Identifies more significant pathways than
  + PaintOmics:
  + LeapR:
  + ActivePathways:
  + CAMERA:
  + BONITA on individual datasets
    - Compare only selected pathways:
      * mBONITA doesn’t pick up any pathways not in other datasets, in fact the set of significant pathways is smaller
    - Across all KEGG pathways:
      * mBONITA picks up more pathways than Bonita
* Identifies different node importance scores from:
  + BONITA:

# Materials and Methods

## Transcriptomics data collection and analysis

*Insert information from Zand lab & GRC.*

Differentially expressed (DE) genes were identified using DESeq2 (16). The R package ‘ashr’ was used for log fold change shrinkage (17). Genes with a Benjamini-Hochberg adjusted p-value < 0.05 and an absolute log-fold change > 0.5 were identified as being DE. Heatmaps were prepared using ComplexHeatmap (18). Over-representation analysis of DE genes was performed with the R package clusterprofiler, using gene sets of canonical KEGG pathways from the MSigDB database (6, 19, 20). Gene sets were identified as being over-represented if the un-adjusted p-value was less than 0.05.

## Data processing

Proteomics and phosphoproteomics data were collected and processed as described in (21) and (22) respectively. We retained only samples from the experimental conditions represented in all three datasets (Table 1). In the case of the proteomics and phosphoproteomics datasets, we mapped protein names to gene names using Entrez and retained these gene names for downstream analysis, for consistency between datasets. We discarded observations for genes whose median value was 0.

## BONITA experiments, networks used, Bonita3 package

We reimplemented our previously published algorithm BONITA (23) in Python3, resulting in a significant improvement in speed. We used this implementation of BONITA-RD to infer Boolean rules individually for the three multi-omics datasets. For each of these experiments, we used all KEGG networks with an overlap of 5 or more genes with the training dataset. We used BONITA-PA to perform pathway analysis to compare the three conditions that were profiled in all three omics datasets (Table 1).

## Integration strategy

## Comparison to other methods

Parameters for other packages here:

PaintOmics:

LeapR:

ActivePathways:

CAMERA:

BONITA:

**Other multi-omics pathway analysis methods:**

1. **DONE - reactomeGSA** – applies CAMERA + Reactome to each dataset separately, returns individual p-values, correlations between datatypes. Adapted this method – used CAMERA + KEGG on each dataset separately, used a p-value combination method (Stouffer’s) to get an overall p-value per pathway.
2. **DONE - paintOmics** (<https://doi.org/10.1093/nar/gkac352>, <https://doi.org/10.1093/nar/gky466>): Fisher’s exact test for each dataset, joint pathway enrichment p-value for all omics data is computed by applying either Fisher combined probability test or Stouffer’s method. Other interesting pathway visualization features.
3. **DONE - ActivePathways** ([doi:10.1038/s41467-019-13983-9](https://doi.org/10.1038/s41467-019-13983-9)): “From a matrix of p-values, ActivePathways creates a ranked gene list where genes are prioritised based on their combined significance of in the series of omics datasets provided in the input matrix. The ranked gene list includes the most significant genes first. ActivePathways then performs a ranked hypergeometric test to determine if a pathway (i.e., a gene set with a common functional annotation) is enriched in the ranked gene list, by performing a series of hypergeometric tests (also known as Fisher’s exact tests). In each such test, a larger set of genes from the top of the ranked gene list is considered. At the end of the series, the ranked hypergeometric test returns the top most significant p-value from the series, corresponding to the point in the ranked gene list where the pathway enrichment reached the greatest significance of enrichment. This approach is useful when the genes in our ranked gene list have varying signals of biological importance in the input omics datasets, as the test identifies the top subset of genes that are the most relevant to the enrichment of the pathway.”
4. Tried a pca-based method PathwayPCA (<https://doi.org/10.1002/pmic.201900409>) and couldn’t get the R package to work. Update – this was my error, it’s not really a pathway analysis method and only tests association of pathways with specific categorical phenotypes, no way to test a contrast.
5. **DONE - leapR** (<https://doi.org/10.1021/acs.jproteome.0c00963>): “Many applications of enrichment compare one group of samples (case) against another group (control) with the goal of identifying pathways that have significantly different abundance in this comparison. The leapR package accomplishes this in the enrichment\_comparison (see [Figure 1](https://pubs.acs.org/doi/10.1021/acs.jproteome.0c00963) and [Table 1](https://pubs.acs.org/doi/10.1021/acs.jproteome.0c00963)) using a *t* test in which the overall abundance of the pathway members is summarized in distributions for the case and control groups and then compared. Output from this analysis will yield *p*-values for each input pathway that indicate the significance of enrichment. Examining the mean abundance from each condition will provide an idea of the effect size and the direction of enrichment—that is, is the pathway more abundant in the case or control condition? A small effect size can still yield very significant *p*-values, but these kinds of results must be treated with caution.”
6. **NOT SURE WHETHER TO INCLUDE - MGSEA** (<https://doi.org/10.1186/s12859-019-2716-6>): adaptation of GSEA to multiomics data. Haven’t looked at this code yet, it is just a supplementary R file.
   1. Not to be confused with multiGSEA (<https://doi.org/10.1186/s12859-020-03910-x>) which just applies GSEA to each sample individually and uses a pvalue combination method or with MOGSA (<https://doi.org/10.1074/mcp.TIR118.001251>) which is a single-sample method

# Discussion

**Manuscript outline:**

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

* + 1. 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)

# Figures and Tables

## 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 datasets from RAMOS B cells grown under three conditions (pathway analysis with mBONITA) (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 )Pathways known to be involved in the hypoxia-mediated response to CyA, Only pathways identified as significant from a combined dataset by mBONITA are shown. 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:

1. Supplementary Table 1 Experimental conditions in the three datasets from RAMOS B cells. Conditions that are in all datasets are highlighted in red.
2. Supplementary 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.
3. Supplementary Table 2: KEGG Pathways involved in the HIF1A-mediated response of B cells to hypoxia and CyA
4. 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.
5. Supplementary Table 3: Spearman correlation between importance scores
6. 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 is < 0.05.
7. 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.
8. Supplementary File 1: mBONITA-PA results Excel workbook - pvalues\_concatenated\_20220816
9. Supplementary File 2: PaintOmics results (paintomics\_allResults.csv)
10. Supplementary File 3: leapR results (leapR\_allResults.csv)
11. Supplementary File 4: CAMERA results (camera\_allResults.csv)
12. Supplementary File 5: ActivePathways results (activePathways\_allResults.csv)
13. Supplementary File 6: BONITA results
14. Supplementary Figure 5: paintOMICS results
15. Supplementary Figure 6: leapR results
16. Supplementary Figure 7: ActivePathways
17. Supplementary Figure 8: CAMERA + Fisher results

# Article and Author Information

## Authors' contributions

*Please see* [*https://casrai.org/credit/*](https://casrai.org/credit/) *for a list of possible author roles in the CRediT system.*

**MGP:** Conceptualization, methodology, software, validation, formal analysis, investigation, data curation, writing (original draft), writing (review & editing), visualization; **XM:** Formal analysis; **AC:** Formal analysis; **SH:** Resources, writing (review & editing); **MZ:** Resources, writing (review & editing); **JT:** Conceptualization, methodology, software, validation, formal analysis, investigation, resources, data curation, writing (original draft), writing (review & editing), visualization, supervision, project administration, funding acquisition

## Competing interests

The authors declare that they have no competing interests.

## Funding

The study was supported by U.S. National Institutes of Health. MGP is supported by R01 AI134058. JT is supported by UM1 AI069511, P30 AI078498, R01 AI134058 and R21 AI136668. (*Insert funding information for other authors*)

## Acknowledgements

We would like to thank all past and present members of the Thakar Lab and the Zand Lab at the University of Rochester, and the Jun Qu Lab at SUNY Buffalo for helpful discussions. The Center for Integrated Research Computing at the University of Rochester provided high-performance computing resources.

# References

1. Ma J, Shojaie A, Michailidis G. A comparative study of topology-based pathway enrichment analysis methods. BMC Bioinformatics. 2019;20(1):546.

2. Ihnatova I, Popovici V, Budinska E. A critical comparison of topology-based pathway analysis methods. Plos One. 2018;13(1):e0191154.

3. Palshikar MG, Palli R, Tyrell A, Maggirwar S, Schifitto G, Singh MV, et al. Executable models of pathways built using single-cell RNA seq data reveal immune signaling dysregulations in people living with HIV and atherosclerosis. medRxiv. 2022:2022.03.07.22271522.

4. Palli R, Palshikar MG, Thakar J. Executable pathway analysis using ensemble discrete-state modeling for large-scale data. Plos Computational Biology. 2019;15(9).

5. Hilchey SP, Palshikar MG, Emo JA, Li D, Garigen J, Wang J, et al. Cyclosporine a directly affects human and mouse b cell migration in vitro by disrupting a hIF-1 αdependent, o(2) sensing, molecular switch. BMC Immunol. 2020;21(1):13.

6. Kanehisa M, Furumichi M, Sato Y, Ishiguro-Watanabe M, Tanabe M. KEGG: integrating viruses and cellular organisms. Nucleic acids research. 2021;49(D1):D545-D51.

7. Martens M, Ammar A, Riutta A, Waagmeester A, Slenter DN, Hanspers K, et al. WikiPathways: connecting communities. Nucleic acids research. 2021;49(D1):D613-D21.

8. Palshikar MG, Hilchey SP, Zand MS, Thakar J. WikiNetworks: translating manually created biological pathways for topological analysis. Bioinformatics. 2021:btab699.

9. Dugourd A, Kuppe C, Sciacovelli M, Gjerga E, Gabor A, Emdal KB, et al. Causal integration of multi-omics data with prior knowledge to generate mechanistic hypotheses. Mol Syst Biol. 2021;17(1):e9730.

10. Hernández-de-Diego R, Tarazona S, Martínez-Mira C, Balzano-Nogueira L, Furió-Tarí P, Pappas GJ, Jr., et al. PaintOmics 3: a web resource for the pathway analysis and visualization of multi-omics data. Nucleic acids research. 2018;46(W1):W503-W9.

11. Liu T, Salguero P, Petek M, Martinez-Mira C, Balzano-Nogueira L, Ramšak Ž, et al. PaintOmics 4: new tools for the integrative analysis of multi-omics datasets supported by multiple pathway databases. Nucleic acids research. 2022;50(W1):W551-W9.

12. Wu D, Smyth GK. Camera: a competitive gene set test accounting for inter-gene correlation. Nucleic acids research. 2012;40(17):e133.

13. Griss J, Viteri G, Sidiropoulos K, Nguyen V, Fabregat A, Hermjakob H. ReactomeGSA - Efficient Multi-Omics Comparative Pathway Analysis. Mol Cell Proteomics. 2020;19(12):2115-25.

14. Danna V, Mitchell H, Anderson L, Godinez I, Gosline SJC, Teeguarden J, et al. leapR: An R Package for Multiomic Pathway Analysis. Journal of Proteome Research. 2021;20(4):2116-21.

15. Paczkowska M, Barenboim J, Sintupisut N, Fox NS, Zhu H, Abd-Rabbo D, et al. Integrative pathway enrichment analysis of multivariate omics data. Nature Communications. 2020;11(1):735.

16. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 2014;15(12):550.

17. Stephens M. False discovery rates: a new deal. Biostatistics. 2017;18(2):275-94.

18. Gu Z, Eils R, Schlesner M. Complex heatmaps reveal patterns and correlations in multidimensional genomic data. Bioinformatics. 2016;32(18):2847-9.

19. Wu T, Hu E, Xu S, Chen M, Guo P, Dai Z, et al. clusterProfiler 4.0: A universal enrichment tool for interpreting omics data. Innovation (N Y). 2021;2(3):100141.

20. Liberzon A, Subramanian A, Pinchback R, Thorvaldsdottir H, Tamayo P, Mesirov JP. Molecular signatures database (MSigDB) 3.0. Bioinformatics. 2011;27(12):1739-40.

21. Hilchey SP, Palshikar MG, Li D, Garigen J, Cipolla V, Thakar J, et al. Cyclosporine A Directly Affects Human and Mouse B cell Migration <em>in vitro</em> by Disrupting a HIF-1<em>α</em> Dependent, O<sub>2</sub> Sensing, Molecular Switch. bioRxiv. 2019:622514.

22. Hilchey SP, Palshikar MG, Shen S, Rasam S, Mendelson ES, Emo JA, et al. LSP1 Attenuates Human B Cell Migration at Physiological Oxygen Levels, as Revealed by Phosphoproteomics Analysis. In preparation. 2022.

23. Palli R, Palshikar MG, Thakar J. Executable pathway analysis using ensemble discrete-state modeling for large-scale data. Plos Computational Biology. 2019;15(9):e1007317.

# Supplementary Data

## Transcriptomics data suggests mechanisms of HIF1A-mediated chemotaxis disrupted by CyA

Differentially expressed (DE) genes and over-represented KEGG gene sets were identified as described in the Methods. The 24 genes that were DE between the `19% O2, CyA-` and `1% O2, CyA-` conditions (i.e., the DE genes driven solely by the response to hypoxia) were enriched for gene sets related to glucose metabolism, as expected (Figure 1 A-C). These genes included LSP1, EGLN1, BNIP3/BNIP3L, and ALDOC, which were upregulated in cells grown at 1% oxygen. All these genes are well-known regulators of cellular responses to cells to hypoxia (*many citations here*). BNIP3/BNIP3L and EGLN1 are direct downstream targets of HIF1A. While the over-representation analysis of the 49 genes that were DE between the `1% O2, CyA+` and `1% O2, CyA-` conditions (i.e., the DE genes driven solely by the response to CyA under physiological hypoxic conditions) did not identify gene sets directly linked to either hypoxia response or immunosuppression (Figure 1C).

VEGFA , PHGDH (<https://aacrjournals.org/cancerres/article/77/22/6321/622947/PHGDH-as-a-Key-Enzyme-for-Serine-Biosynthesis-in>, https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5286373/), PSAT1 (regulated by ATF4/5)

ADA - <https://pubmed.ncbi.nlm.nih.gov/16670267/>

PIM2 - <https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0088301>, <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7783746/>

However, these DE genes included ATF4/5, SESN2 (HIF1A-induced, <http://www.aginganddisease.org/EN/10.14336/AD.2019.0320>),

## Multi-omics datasets showed low inter-dataset consistency

We profiled expression of mRNA, proteins, and phosphorylated proteins from RAMOS B cells grown under different oxygenation conditions and treatment with CyA and CXCL12, and retained only those conditions that were profiled in all three omics datasets for downstream analysis (Table 1). After processing and filtering the datasets as described in the Methods, we found that only 1505 genes were profiled in all three datasets (Figure 1A). The transcriptomic and proteomic data for these 1505 genes were moderately well correlated (Spearman correlation = 0.58, p-value < 0.01). However, the phosphoproteomics data showed low correlation to both the transcriptomics and proteomics data (Spearman correlation = 0.15 and 0.17 respectively, p-value < 0.01) (Figure 1B). We next attempted to find whether the intra-dataset correlations were consistent between datasets, i.e., whether the same signaling processes could be identified in all three datasets. We constructed a correlation network for each dataset as described in the Methods.

The overlap between genes that were highly correlated between pairs of datasets was similarly low (Figure 2). We constructed a network comprised of all KEGG networks, performed network propagation using TIEDIE, and identified modules of genes that were significantly differentially expressed between pairs of conditions. A representative comparison between B cells grown at 1% oxygen with CyA and without CyA is shown in Figure 1.