OMICS Data Integration: Why and How?

Ali Shojaie
Department of Biostatistics
University of Washington
faculty.washington.edu/ashojaie
Part I: WHY?
Why Integrate Omics Data?

Biology is complex, heterogenous and structured!
Why Integrate Omics Data?

Biology is complex, heterogenous and structured!
Why Integrate Omics Data?

Comprehensive view of biology requires looking at multiple types of *omics* data (TCGA, ENCODE, etc)

⇒ integrative analysis of multiple **structured** omics data
Why Integrate Omics Data?

Possible reasons:
Why Integrate Omics Data?

Possible reasons:

1. To confirm or narrow down omics signals
Why Integrate Omics Data?

Possible reasons:

1. To *confirm* or narrow down omics signals
2. To *complement* or boost omics signals
Why Integrate Omics Data?

Possible reasons:

1. To confirm or narrow down omics signals
2. To complement or boost omics signals
3. To glean systems perspective
Reconstructing targetable pathways in lung cancer by integrating diverse omics data

O. Alejandro Balbin, John R. Prensner, Anirban Sahu, Anastasia Yocum, Sunita Shankar, Rohit Malik, Damian Fermin, Saravana M. Dhanasekaran, Benjamin Chandler, Dafydd Thomas, David G. Beer, Xuhong Cao, Alexey I. Nesvizhskii & Arul M. Chinnaiyan

Global ‘multi-omics’ profiling of cancer cells harbours the potential for characterizing the signalling networks associated with specific oncogenes. Here we profile the transcriptome, proteome and phosphoproteome in a panel of non-small cell lung cancer (NSCLC) cell lines in order to reconstruct targetable networks associated with KRAS dependency. We develop a two-step bioinformatics strategy addressing the challenge of integrating these disparate datasets. We first define an ‘abundance-score’ combining transcript, protein and phospho-protein abundances to nominate differentially abundant proteins and then use the Prize Collecting Steiner Tree algorithm to identify functional sub-networks. We identify three modules centred on KRAS and MET, LCK and PAK1 and β-Catenin. We validate activation of these proteins in KRAS-dependent (KRAS-Dep) cells and perform functional studies defining LCK as a critical gene for cell proliferation in KRAS-Dep but not KRAS-independent NSCLCs. These results suggest that LCK is a potential druggable target protein in KRAS-Dep lung cancers.
Omics data integration can lead to new discoveries...

Cell Systems

Integrated Transcriptome and Proteome Analyses Reveal Organ-Specific Proteome Deterioration in Old Rats

Graphical Abstract

Authors
Alessandro Ori, Brandon H. Toyama, Michael S. Harris, ..., Nicholas T. Ingolia, Martin W. Hetzer, Martin Beck

Correspondence
ingolia@berkeley.edu (N.T.I.), hetzer@salk.edu (M.W.H.), mbeck@embl.de (M.B.)

In Brief
Ori et al. quantified the molecular alterations that occur between young and old rats in two organs: brain and liver. By integrating genomic and proteomic measurements, the authors were able to reveal that changes in translation are the primary cause of protein level alterations during aging. However, they also identified other levels of regulation such as protein localization and phosphorylation that co-participate in modifying the proteome in old animals.
Integrated analysis of global proteome, phosphoproteome, and glycoproteome enables complementary interpretation of disease-related protein networks

Jong-Moon Park1,*, Ji-Hwan Park2,*, Dong-Gi Mun3,*, Jingi Bae3,*, Jae Hun Jung4, Seunghoon Back3, Hangyeore Lee3, Hokeun Kim3, Hee-Jung Jung6, Hark Kyun Kim5, Hookeun Lee3, Kwang Pyo Kim4, Dahee Hwang2,6 & Sang-Won Lee3

Multi-dimensional proteomic analyses provide different layers of protein information, including protein abundance and post-translational modifications. Here, we report an integrated analysis of protein expression, phosphorylation, and N-glycosylation by serial enrichments of phosphorylation and N-glycosylation (SEPG) from the same tissue samples. On average, the SEPG identified 142,106 unmodified peptides of 8,625 protein groups, 18,846 phosphopeptides (15,647 phosphosites), and 4,019 N-glycopeptides (2,634 N-glycosites) in tumor and adjacent normal tissues from three gastric cancer patients. The combined analysis of these data showed that the integrated analysis additively improved the coverages of gastric cancer-related protein networks; phosphoproteome and N-glycoproteome captured predominantly low abundant signal proteins, and membranous or secreted proteins, respectively, while global proteome provided abundances for general population of the proteome. Therefore, our results demonstrate that the SEPG can serve as an effective approach for multi-dimensional proteome analyses, and the holistic profiles of protein expression and PTMs enabled improved interpretation of disease-related networks by providing complementary information.
Integrated Analyses Identify a Master MicroRNA Regulatory Network for the Mesenchymal Subtype in Serous Ovarian Cancer

Da Yang,1,7,11 Yan Sun,1,7,11 Limei Hu,1,11 Hong Zheng,8,11 Ping Ji,1 Chad V. Pecot,6 Yanrui Zhao,8 Sheila Reynolds,9 Hanyin Cheng,1,12 Rajesha Rupaimoole,2 David Cogdell,1 Matti Nykter,10 Russell Broaddus,1 Cristian Rodriguez-Aguayo,4 Gabriel Lopez-Berestein,4,5 Jinsong Liu,1 Ilya Shmulevich,9 Anil K. Sood,2,3,5,7 Kexin Chen,8,7 and Wei Zhang1,5,7

1Department of Pathology
2Department of Gynecologic Oncology and Reproductive Medicine
3Department of Cancer Biology
4Department of Experimental Therapeutics
5Center for RNA and Non-Coding RNA
6Division of Cancer Medicine
The University of Texas MD Anderson Cancer Center, Houston, TX 77030, USA
7Department of Pathology
8Department of Epidemiology and Biostatistics
9Tianjin Medical University Cancer Institute and Hospital, Tianjin 300060, China
10Institute for Systems Biology, Seattle, WA 98103, USA
11Tampere University of Technology, Tampere 33101, Finland
12These authors contributed equally to this work

Summary

Integrated genomic analyses revealed a miRNA-regulatory network that further defined a robust integrated mesenchymal subtype associated with poor overall survival in 459 cases of serous ovarian cancer (OvCa) from The Cancer Genome Atlas and 560 cases from independent cohorts. Eight key miRNAs, including miR-506, miR-141, and miR-200a, were predicted to regulate 89% of the targets in this network. Follow-up functional experiments illustrate that miR-506 augmented E-cadherin expression, inhibited cell migration and invasion, and prevented TGFβ-induced epithelial-mesenchymal transition by targeting SNAI2, a transcriptional repressor of E-cadherin. In human OvCa, miR-506 expression was correlated with decreased SNAI2 and VIM, elevated E-cadherin, and beneficial prognosis. Nanoparticle delivery of miR-506 in orthotopic OvCa mouse models led to E-cadherin induction and reduced tumor growth.
Part II: HOW?
More on Omics Data Integration

Broadly, two existing integration approaches:
More on Omics Data Integration

Broadly, two existing integration approaches:

- **Horizontal integration** *(same variables, different studies/subjects)*

\[
\begin{align*}
X_3 \ (n_3 \times p) \\
X_2 \ (n_2 \times p) \\
X_1 \ (n_1 \times p)
\end{align*}
\]

Variables/features 1, ..., p
More on Omics Data Integration

Broadly, two existing integration approaches:

- **Vertical integration** (different platforms/variables, same subjects)
Existing Approaches: Meta Analysis
## Existing Approaches: Meta Analysis

### Horizontal integration

<table>
<thead>
<tr>
<th>Variables/features 1, ..., p</th>
</tr>
</thead>
<tbody>
<tr>
<td>$X_3 \ (n_3 \times p)$</td>
</tr>
<tr>
<td>$X_2 \ (n_2 \times p)$</td>
</tr>
<tr>
<td>$X_1 \ (n_1 \times p)$</td>
</tr>
</tbody>
</table>
Existing Approaches: Meta Analysis

**Horizontal integration**

- Used extensively in GWAS (especially in consortiums)
Existing Approaches: Meta Analysis

**Horizontal integration**

- Used extensively in GWAS (especially in consortiums)
- Used to boost the signal (larger sample size)
Existing Approaches: **Meta Analysis**

**Horizontal integration**

- Used extensively in GWAS (especially in consortiums)
- Used to *boost the signal* (larger sample size)
- Used to *confirm previous findings* (reproducibility)
# Existing Approaches: Meta Analysis

<table>
<thead>
<tr>
<th>Method</th>
<th>Description</th>
<th>Advantages</th>
<th>Disadvantages</th>
<th>Main software used</th>
</tr>
</thead>
<tbody>
<tr>
<td>P value meta-analysis</td>
<td>Simplest meta-analytical approach</td>
<td>Allows meta-analysis when effects are not available</td>
<td>Direction of effect is not always available; inability to provide effect sizes; difficulties in interpretation</td>
<td>METAL, GWAMA, R packages</td>
</tr>
<tr>
<td>Fixed effects</td>
<td>Synthesis of effect sizes. Between-study variance is assumed to be zero</td>
<td>Effects readily available through specialized software</td>
<td>Results may be biased if a large amount of heterogeneity exists</td>
<td>METAL, GWAMA, R packages</td>
</tr>
<tr>
<td>Random effects</td>
<td>Synthesis of effect sizes. Assumes that the individual studies estimate different effects</td>
<td>Generalizability of results</td>
<td>Power deserts in discovery efforts; may yield spuriously large summary effect estimates when there are selection biases</td>
<td>GWAMA, R packages</td>
</tr>
<tr>
<td>Bayesian approach</td>
<td>Incorporates prior assessment of the genetic effects</td>
<td>Most direct method for interpretation of results as posterior probabilities given the observed data</td>
<td>Methodologically challenging; GWAS-tailored routine software not available; subjective prior information used</td>
<td>R packages</td>
</tr>
<tr>
<td>Multivariate approaches</td>
<td>Incorporates the possible correlation between outcomes or genetic variants</td>
<td>Increased power can identify variants that conventional meta-analysis do not reveal using the same data sets</td>
<td>Computationally intensive; software not available for all analyses; some may require individual-level data</td>
<td>GCTA for multi-locus approaches</td>
</tr>
<tr>
<td>Other extensions</td>
<td>A set of different approaches that allows for the identification of multiple variants across different diseases</td>
<td>Summary results of previous meta-analyses can be used</td>
<td>May need additional exploratory analyses for the identification of variants; prone to systematic biases</td>
<td>Software developed by the authors of the proposed methodologies</td>
</tr>
</tbody>
</table>

Existing Approaches: **Direct Integration**
Existing Approaches: Direct Integration

**Vertical integration**

\[ X_3 (n \times p_3) \]  \[ X_2 (n \times p_2) \]  \[ X_1 (n \times p_1) \]

Can result in way-too-many variables
Existing Approaches: Direct Integration

Vertical integration

\[
X_3 \ (n \times p_3) \quad X_2 \ (n \times p_2) \quad X_1 \ (n \times p_1)
\]

The simplest approach: concatenate the variables!!
Existing Approaches: **Direct Integration**

**Vertical integration**

![Diagram](image)

The simplest approach: **concatenate** the variables!!

- Can result in way-too-many variables
- Can discern **conditional associations** with phenotype $y$
Existing Approaches: **Kernel-Based Methods**

Vertical integration using kernel regression

Penalized regression in terms of a kernel

\[ K^d = \arg\min_d \sum_{n=1}^{n} \sum_{\Omega} y K_{d(\Omega)}^2 + \lambda \sum_{\Omega} K_{d(\Omega)}^2 \]

Ideal for predicting \( y \) in the dual space

Also used to test for association between \( y \) and \( X \) (SKAT)
Existing Approaches: Kernel-Based Methods

Vertical integration using kernel regression
Existing Approaches: **Kernel-Based Methods**

**Vertical** integration using **kernel regression**

**Kernel Regression**

\[
y = X_{nxp} \beta + e
\]

Ideal for predicting **y** in the dual space

Also used to test for association between **y** and **X** (SKAT)
Existing Approaches: **Kernel-Based Methods**

**Vertical integration using kernel regression**

**Kernel Regression**

- Penalized regression in terms of a kernel $K$

$$
\hat{\delta} = \operatorname{argmin}_{\delta \in \mathbb{R}^n} || y - K\delta ||_2^2 + \lambda \| \delta \|_K^2
$$
Existing Approaches: **Kernel-Based Methods**

**Vertical integration using kernel regression**

**Kernel Regression**

- Penalized regression in terms of a kernel $K$

$$\hat{\delta} = \arg\min_{\delta \in \mathbb{R}^n} \| y - K\delta \|_2^2 + \lambda \|\delta\|_K^2$$

- Ideal for predicting $y$ in the dual space
- Also used to test for association between $y$ and $X$ (SKAT)
Existing Approaches: **Kernel-Based Methods**

Can define different kernels (or feature maps) for different omics data types

Existing Approaches: Kernel-Based Methods
Existing Approaches: **Kernel-Based Methods**

- **Early** integration

\[
\begin{align*}
X_1 (nxp_1) &+ X_2 (nxp_2) = X (nx(p_1+p_2)) \\
K_{nxn} &+ \partial + e
\end{align*}
\]
Existing Approaches: **Kernel-Based Methods**

- **Intermediate** integration

\[
X_1 (nxp_1) \quad X_2 (nxp_2) \quad K_1 (nxn) \quad K_2 (nxn) \quad K (nxn)
\]

\[
y = K_{nxn} \partial + e
\]
Existing Approaches: **Kernel-Based Methods**

- **Late** integration

\[ X_1(n \times p_1) + X_2(n \times p_2) = K_1(n \times n) + K_2(n \times n) \]

\[ y = K_1(n \times n) \partial_1 + K_2(n \times n) \partial_2 + e \]
Kernel-Penalized Regression$^1$

$^1$Randolph et al (2018)
Kernel-Penalized Regression\textsuperscript{1}

\[ y = X_{n \times p} \beta + e \]

\[ y = K_{n \times n} \delta + e \]

\textsuperscript{1}Randolph et al (2018)
Kernel-Penalized Regression\(^1\)

\[ y = X_{nxp} \beta + e \]

\[ y = K_{nxn} \delta + e \]

- How can we incorporate network information?
- How can we evaluate association of individual omics measures and the response (biomarker discovery)?

\(^1\)Randolph et al (2018)
Kernel-Penalized Regression

Use the duality between the feature space \( \mathbb{R}^p \) and the observation space \( \mathbb{R}^n \) – formally, the duality diagram (Escoufier (1977), ...)

Can incorporate additional structure, e.g., network information

Can also incorporate multiple data omics data
Kernel-Penalized Regression

- Use the *duality* between the feature space ($\mathbb{R}^p$) and the observation space ($\mathbb{R}^n$) – formally, the duality diagram (Escoufier (1977), ...)

\[
\begin{align*}
\mathbb{R}^p & \leftrightarrow X^\top \quad \mathbb{R}^n \\
X & \downarrow \\
\end{align*}
\]
Kernel-Penalized Regression

- Use the *duality* between the feature space \( \mathbb{R}^p \) and the observation space \( \mathbb{R}^n \) – formally, the duality diagram (Escoufier (1977), ...)
- Can incorporate additional structure, e.g., network information
Kernel-Penalized Regression

- Use the *duality* between the feature space ($\mathbb{R}^p$) and the observation space ($\mathbb{R}^n$) – formally, the duality diagram (Escoufier (1977), ...)
- Can incorporate additional structure, e.g., network information
- Can also incorporate multiple data omics data
Example: Integrating Metabolomics Data

Integrating targeted ($X_1$) and unbiased ($X_2$) metabolomics profiling data for the same subjects
Example: Integrating Metabolomics Data

Integrating targeted \((X_1)\) and unbiased \((X_2)\) metabolomics profiling data for the same subjects

LOO CV-prediction error for the original data

<table>
<thead>
<tr>
<th></th>
<th>Lasso</th>
<th>Ridge</th>
<th>KPR</th>
</tr>
</thead>
<tbody>
<tr>
<td>MSE</td>
<td>29.00</td>
<td>30.68</td>
<td>27.25</td>
</tr>
</tbody>
</table>
Example: Integrating Metabolomics Data

Integrating targeted \((X_1)\) and unbiased \((X_2)\) metabolomics profiling data for the same subjects

LOO CV-prediction error for the original data

<table>
<thead>
<tr>
<th></th>
<th>Lasso</th>
<th>Ridge</th>
<th>KPR</th>
</tr>
</thead>
<tbody>
<tr>
<td>MSE</td>
<td>29.00</td>
<td>30.68</td>
<td>27.25</td>
</tr>
</tbody>
</table>

MSE for estimation of regression coefficients, based on our \(\beta^*\)
Another Example: Analysis of Microbiome Data

Simulation setup

- **Simulate** the outcome based on real microbiome data:
  - we use the data from Yatsunenko et al (2012) consisting of \( p = 495 \) taxa for \( n = 100 \) subjects with \( y = \log(\text{age}) \)
  - the original study showed that a 2D-MDS based on the phylogenetic tree captures the pattern in response (left)
  - we generate \( y^* \) similarly in a phylogenetically-informed PCR (right)
Another Example: Analysis of Microbiome Data

Simulation results

![Graph 1: Squared Error in Estimation vs. R-Squared](image1)

![Graph 2: H-Squared Error in Prediction vs. R-Squared](image2)
Existing Approaches: **Unsupervised Learning Methods**
Existing Approaches: **Unsupervised Learning Methods**

*Vertical integration*
Existing Approaches: Unsupervised Learning Methods

**Vertical integration**
- Integrative clustering

Existing Approaches: Unsupervised Learning Methods

**Vertical integration**

- Integrative clustering
Existing Approaches: **Unsupervised Learning Methods**

**Vertical integration**

- Integrative clustering

Existing Approaches: **Unsupervised Learning Methods**

**Vertical integration**
Existing Approaches: **Unsupervised Learning Methods**

**Vertical integration**
- Integrative dimension reduction
Existing Approaches: Unsupervised Learning Methods

**Vertical integration**

- Integrative dimension reduction
  - **Canonical Correlation Analysis (CCA)**, which looks for correlated omics measures — see, e.g. Witten et al (2009)
Existing Approaches: Unsupervised Learning Methods

**Vertical integration**

- **Integrative dimension reduction**
  - **Canonical Correlation Analysis** (CCA), which looks for correlated omics measures — see, e.g. Witten et al (2009)
  - **Integrative Matrix Factorization** (PCA, etc) — see Lock et al (2013); Argelaguet et al (2018)
Part III: Extensions
Metabolomic profiles delineate potential role for sarcosine in prostate cancer progression

Annu Sreekrishnan,1,2,3,4, Laila M. Poisson,1,2,3,4, Thekkelmycky M. Rajendiran,1,2,3,4, Anjali P. Khani,1,2,3,4, Qi Cao,1,3,4, LinDan Yu,1,2, Jiyanthi Damyanwana,1,2,3,4, Rohit Mehra,1,2,3,4, Robert J. Longrado,1,2,4, Yong Li,1,2,3,4, Mukesh K. Nyali,1,2,3,4, Aarif Ahsan,1,2,3,4, Shanker Kalyana Sundaram,1,2,3,4, Bo Han,1,2,3,4, Xuehong Cao,1,2,3,4, Jionan Yuan,1,2,3,4, Gilbert S. Omang,1,2,3,4, Debashis Ghosh,1,2,3,4, Subramaniyan Pennathur,1,2,3,4, Danny C. Alexander,1,2,3,4, Alvin Berger,1,2,3,4, Jeffrey R. Shuster,1,2,3,4, John T. Wei,1,2,3,4, Sooryanarayana Varambally,1,2,3,4, Christopher Beecher,1,2,3,4, & Arul M. Chinnaiyan1,2,3,4,5,6

Multiple, complex molecular events characterize cancer development and progression.1-5. Deciphering the molecular networks that distinguish aggressive prostate cancer from metastatic disease may lead to the identification of critical biomarkers for cancer invasion and disease aggressiveness. Although gene and protein expression have been extensively profiled in human tumors, little is known about the global metabolomic alterations that characterize metastatic progression. Using a combination of high-throughput liquid- and gas-chromatography-based mass spectrometry, we profiled more than 1,125 metabolites across 282 clinical samples related to prostate cancer (45 cancer and 137 cancer and were differential (Wilcoxon P < 0.05), with a false discovery rate (FDR) of 7.9%. Furthermore, for urine, 36 out of 585 (6%) metabolites were differential (Wilcoxon P < 0.05), with an FDR of 0.7%. Thus, our initial focus was directed towards understanding the urine metabolomic profiles as they showed more robust alterations.

These samples were derived from benign adjacent prostate (n = 10), clinically localized prostate cancer (n = 12, PCA) and metastatic prostate cancer (n = 14) patients. Selection of metastatic tissue samples from different sites (see Supplementary Table 2) minimized characterization of analytes specific to cells of non-prostatic origin. In total, high throughput profiling of the urine metabolomic dataset...
Prostate Cancer...

- Prostate cancer (PCa) is the most common cancer in men
- About 221K new cases per year in the US
- About 28K deaths per year in the US – second leading cause of deaths in cancers (behind lung cancer)
- 5-year survival rate for localized PCa is nearly 100%
Prostate Cancer...(ctd.)

- PCa is driven by multiple factors & many genes implicated (androgen receptor, the TMPRSS2-ETS gene family fusion, BRCA1 and BRCA2)
- The prostate glands require androgen to work properly
- Androgen hormonal therapy is widely used in older patients (over 75 years) rather than radical prostatectomy or radiation therapy
- Castrate Resistant PCa (CRPC) does not respond to hormone (androgen) treatments or gets worse with hormone therapy
- poor survival prognostics for CRPC patients: mean survival time ≤ 2 yrs
- Precise molecular alterations driving CRPC not well-understood
Data from Sreekumar et al (2009)

- Transcriptomic and metabolomic data for 12 PCa and 16 benign adjacent tissue samples
- Mostly matched samples, but few unmatched!
- Given the small sample size, need to
  - preserve all samples
  - reduce dimension
Omics Data Integration: Beyond Vertical and Horizontal

What if we have data on different platforms, but the samples don’t match?
What if we have data on different platforms, but the samples don’t match?
More on Omics Data Integration

Solution: Use pathways as the common dimension!
More on Omics Data Integration

**Solution**: Use *pathways* as the common dimension!
Rank-Based Integration

Metabolomics and Transcriptomics data from non-matching samples
Rank-Based Integration
Metabolomics and Transcriptomics data from non-matching samples
Rank-Based Integration
Metabolomics and Transcriptomics data from non-matching samples
Step 1: Rank-Based Integration

Rankings vs Integrative Score

Gene rank

Log(integrative score)

Metab rank

Riboflavin metabolism

Aminosugars metabolism

- Riboflavin metabolism
- Biotin metabolism
- Aminosugars metabolism
- Valine, leucine and isoleucine biosynthesis
Step 2: Network Enrichment Analysis

Network permutation test to identify key pathways with active neighbors
Putting Things Together
Rank-based integrative pathway scores vs. network enrichment p-values

Integrative Score

-\log_{10}(P(\text{net perm}))

-\log(0.01)
-\log(0.03)
-\log(0.1)
-\log(0.3)
-\log(1)

\begin{itemize}
  \item Riboflavin metabolism
  \item Biotin metabolism
  \item Amino sugar metabolism
  \item Valine, leucine and isoleucine biosynthesis
  \item Cysteine metabolism
\end{itemize}

\[\Rightarrow \text{Aminosugar Metabolism, or Hexosamine Biosynthesis Pathway (HBP)}\]
Clinical Relevance of HBP
Expressions of HBP Genes in PCa
GNPNAT1 Expression in PCa
Therapeutic Potential
Therapeutic Potential

- HBP components elevated in localized PCa, but down-regulated in castrate resistant PCa (CRPC)
- Genetic loss of function experiments for GNPNAT1 in CRPC-like cells led to increased proliferation and aggressiveness, in vitro and in vivo
Addition of HBP metabolite UDP-N-acetylglucosamine to CRPC-like cells reduced the expression of cell cycle genes and attenuated tumor cell proliferation, both in vitro and in vivo; also demonstrated additive efficacy when combined with enzalutamide in vitro.
Inhibition of the hexosamine biosynthetic pathway promotes castration-resistant prostate cancer

Akash K. Kaushik\textsuperscript{1,2,*}, Ali Shojai\textsuperscript{3,*}, Katrin Panzitt\textsuperscript{1,†,*}, Rajni Sonavane\textsuperscript{1}, Harene Venghatakrishnan\textsuperscript{4,5}, Mohan Manikkam\textsuperscript{1}, Alexander Zaslavsky\textsuperscript{4,5}, Vasanta Putluri\textsuperscript{1}, Vihas T. Vasu\textsuperscript{6}, Yiqing Zhang\textsuperscript{1}, Ayesha S. Khan\textsuperscript{7}, Stacy Lloyd\textsuperscript{1}, Adam T. Szafran\textsuperscript{1}, Subhamoy Dasgupta\textsuperscript{1}, David A. Bader\textsuperscript{1}, Fabio Stossi\textsuperscript{1}, Hangwen Li\textsuperscript{4,5}, Susmita Samanta\textsuperscript{1}, Xuhong Cao\textsuperscript{4,5,8}, Efrosini Tsouko\textsuperscript{7}, Shixia Huang\textsuperscript{1,9}, Daniel E. Frigo\textsuperscript{7,10}, Lawrence Chan\textsuperscript{1,2}, Dean P. Edwards\textsuperscript{1,9}, Benny A. Kaipparettu\textsuperscript{11}, Nicholas Mitsiades\textsuperscript{1}, Nancy L. Weigel\textsuperscript{1}, Michael Mancini\textsuperscript{1}, Sean E. McGuire\textsuperscript{1}, Rohit Mehra\textsuperscript{4,8}, Michael M. Ittmann\textsuperscript{12}, Arul M. Chinnaiyan\textsuperscript{4,5,8,13}, Nagireddy Putluri\textsuperscript{1}, Ganesh S. Palapattu\textsuperscript{4,5}, George Michailidis\textsuperscript{14,*} & Arun Sreekumar\textsuperscript{1,2,9}
Other Related Projects
Other Related Projects

- **FDR control** for omics data integration (multivariate test statistics)²

1 Alishahi, Ehyaei & S., A generalized Benjamini-Hochberg procedure for multivariate hypothesis testing
Other Related Projects

- **FDR control** for omics data integration (multivariate test statistics)\(^2\)
- **Integrative multi-layer network analysis**\(^3\)

---

1. Alishahi, Ehyaei & S., A generalized Benjamini-Hochberg procedure for multivariate hypothesis testing
Other Related Projects

- **FDR control** for omics data integration (multivariate test statistics)$^2$
- **Integrative multi-layer network analysis**$^3$

---

1. Alishahi, Ehyaei & S., A generalized Benjamini-Hochberg procedure for multivariate hypothesis testing
Other Related Projects

- **FDR control** for omics data integration (multivariate test statistics)\(^2\)
- **Integrative multi-layer network analysis**\(^3\)

\(^1\) Alishahi, Ehyaei & S., A generalized Benjamini-Hochberg procedure for multivariate hypothesis testing
\(^2\) Zhang et al (2018)
What’s Next?
What’s Next?

Network-based integration of omics data over multiple subpopulations (horizontal and vertical!)