Nature Methods 13.9 (2016): 770-776

Revealing disease-associated pathways by network integration of untargeted metabolomics

Leila Pirhaji, Pamela Milani, Mathias Leidl, Timothy Curran, Julian Avila-Pacheco, Clary B Clish, Forest M White, Alan Saghatelian & Ernest Fraenkel

Presented by Yang Yang

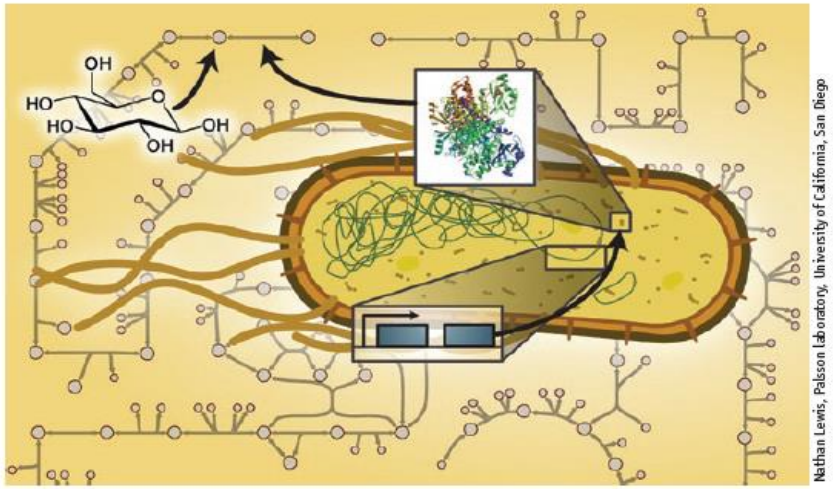
September 26, 2016

Outline

- Background
- Methods
- Results
- Conclusion

Background

- Integrative analysis of metabolomics with other omic data is a crucial step to identifying disease etiology
- **Global measurements of metabolites (untargeted metabolomics)**
  - ➢ Liquid chromatography-mass spectrometry (LC-MS)
- **Ambiguity in global metabolite identification**
- Relatively few of metabolite features are characterized
  - ➢ Tandem mass spectrometry (MS/MS) experiments



Genomic, proteomic and metabolomic data can all be integrated using genome-scale metabolic network reconstructions.

Baker, M. Metabolomics: from small molecules to big ideas. *Nat. Methods* **8**, 117–121 (2011).

Background

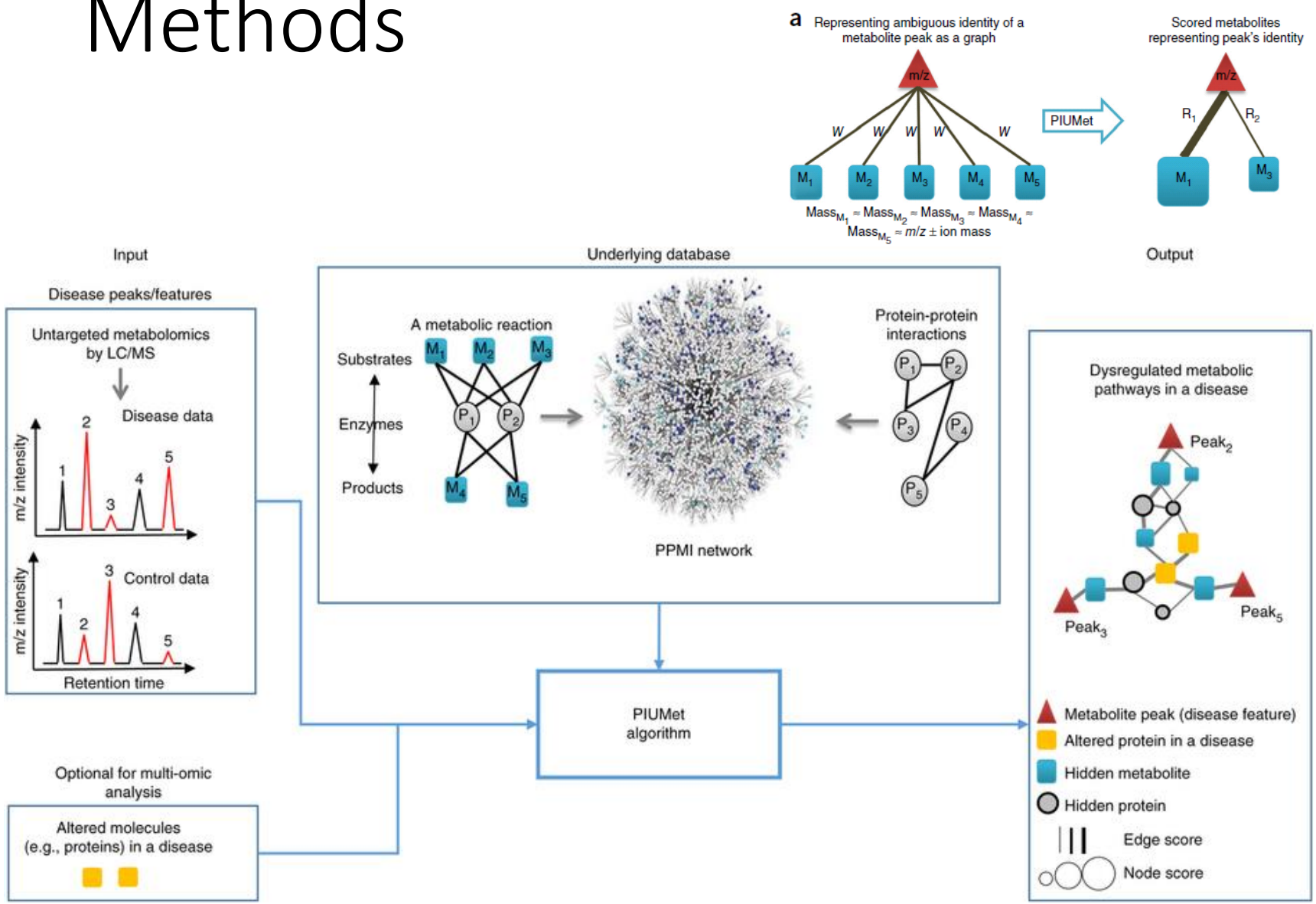
PIUMet: a network-based approach, prize-collecting Steiner forest algorithm for integrative analysis of untargeted metabolomics

Goal of the proposed method:

Infer molecular pathways and components via integrative analysis of metabolite features, without requiring their identification

- Infer putative metabolites corresponding to features and molecular mechanisms underlying their dysregulation
- Provide statistical methods that account for uncertainty in experimental data and the network
- Integrative analysis of untargeted metabolomics with other omics data

Methods



Compile a PPMI network from biochemical reactions and protein-protein interactions

↓

Link disease features to metabolites with the same mass

↓

Eliminate bias for highly connected nodes by assigning negative prizes

↓

Infer a subnetwork connecting disease features using PCSF optimization

↓

Genereate a family of solution by adding noise to the PPMI edge weights

↓

Merge these networks to captures metabolic network complexity

↓

Calculate robustness of the nodes and edges

↓

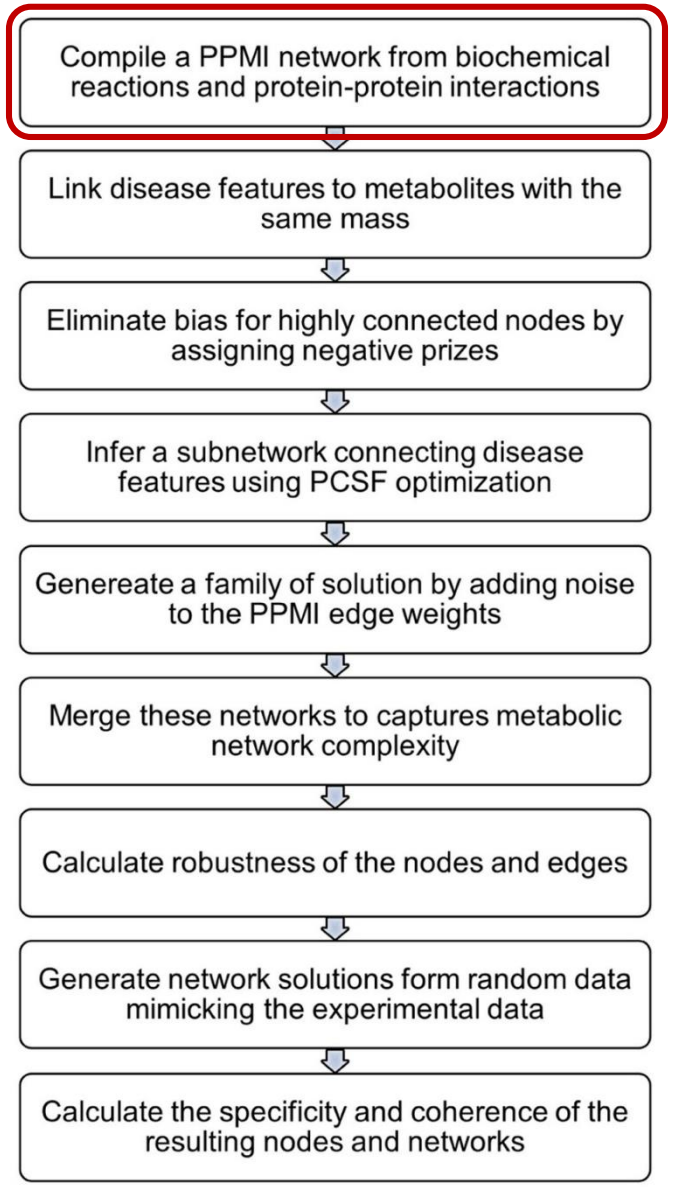
Generate network solutions form random data mimicking the experimental data

↓

Calculate the specificity and coherence of the resulting nodes and networks

Methods

- Construct the PPMI (protein-protein and protein-metabolite interactions) network
- Integrate three databases
  - **iRefIndex**: database of protein-protein interactions
  - **HMDB**: database of human metabolites
    - Include association of transporters and enzymes with metabolites
  - **Recon2**: database of metabolic reactions
- Confidence scores with PPMI edges are obtained from source databases



Methods

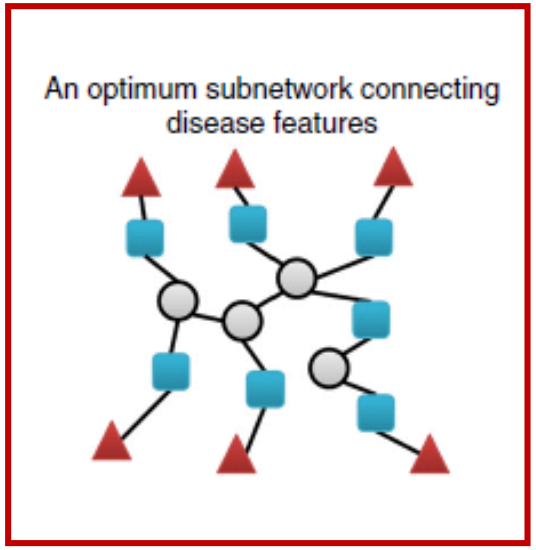
- PIUMet Searches PPMI interactome for a subnetwork connecting disease features using **Prize-Collecting Steiner Forest (PCSF)** optimization

$$f'(F) = \beta \sum_{n \notin N_F} p(n) + \sum_{e \in E_F} c(e) + \omega \times k$$

Prize of node $p(n)$: $-\log(p \text{ value})$

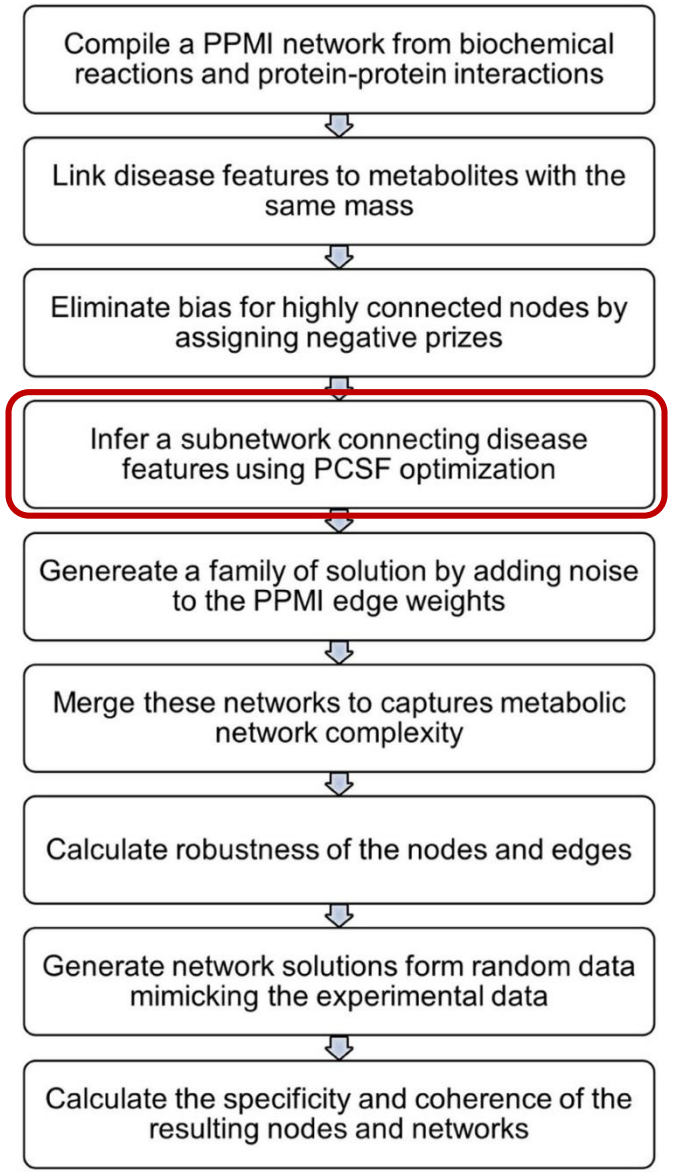
Cost of edge $c(e)$: 1- PPMI edge weight

k: number of trees in the forest



➢ Message passing approach

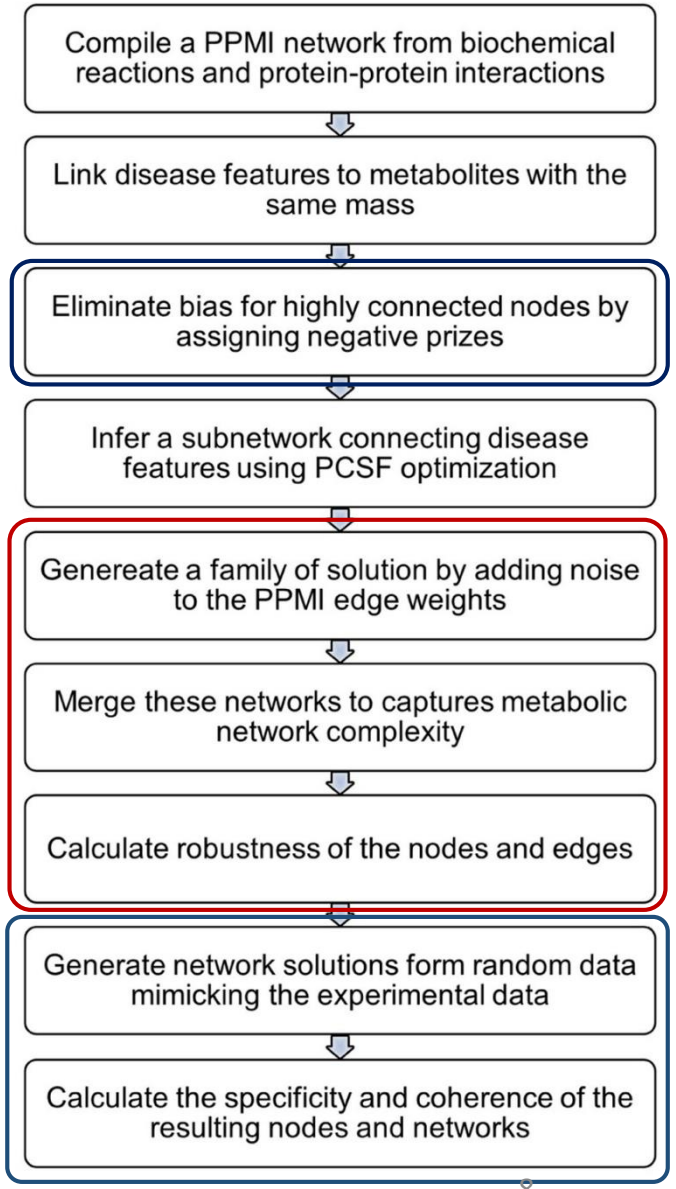
Bailly-Bechet, M., Braunstein, A., Pagnani, A., Weigt, M. & Zecchina, R. Inference of sparse combinatorial-control networks from gene-expression data: a message passing approach. BMC Bioinformatics 11, 355 (2010).



Methods

Features to improve the accuracy

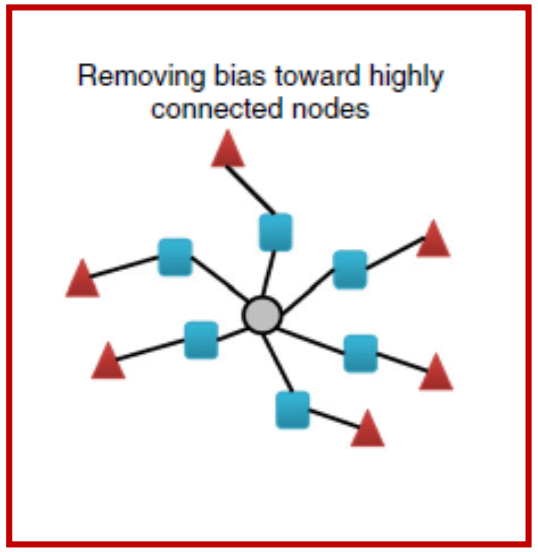
- Eliminate bias toward highly connected nodes in PPMI
- Generate a family of networks to infer the complex interactions
- Calculate disease-specific score for each resulting node and each network

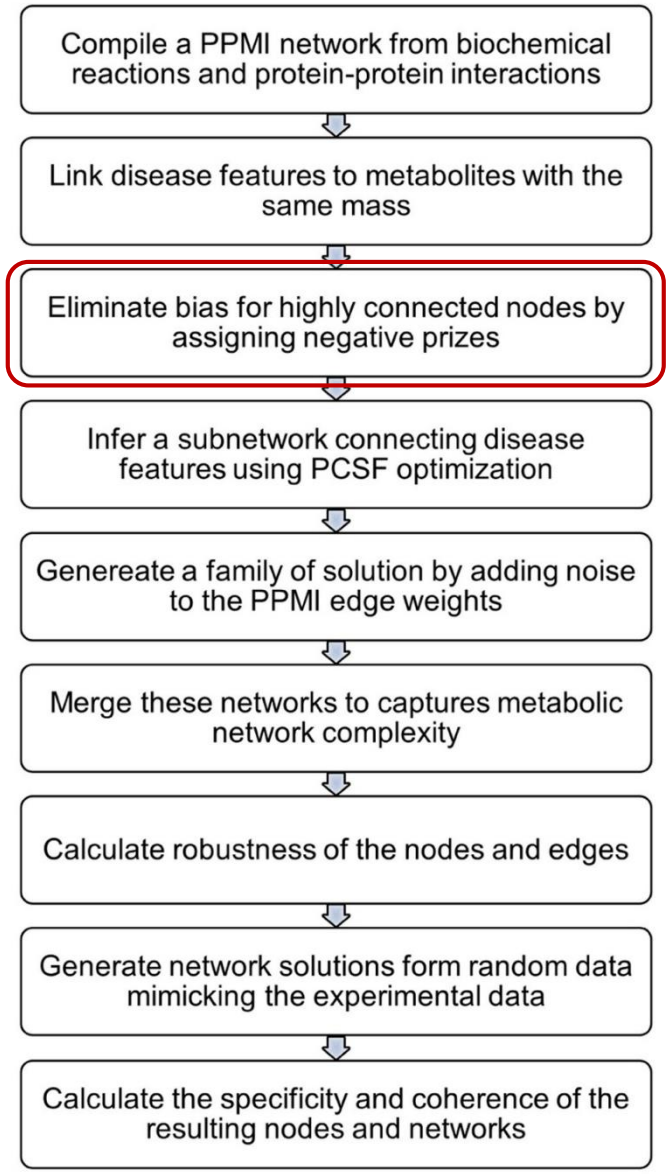


Methods

- Eliminate bias toward highly connected nodes in PPMI by penalizing the inclusion of high-degree nodes
  - Negative prize assigned to nonterminal nodes according to their degree
  - Nodes' degree distribution differ between metabolite and protein sets

$$p(n) = \begin{cases} -\mu \times \text{degree(n) if } \text{n} \in \text{PPMI}(P) \\ -\mu \times \text{degree(n)}^2 \text{ if } \text{n} \in \text{PPMI}(M) \end{cases}$$



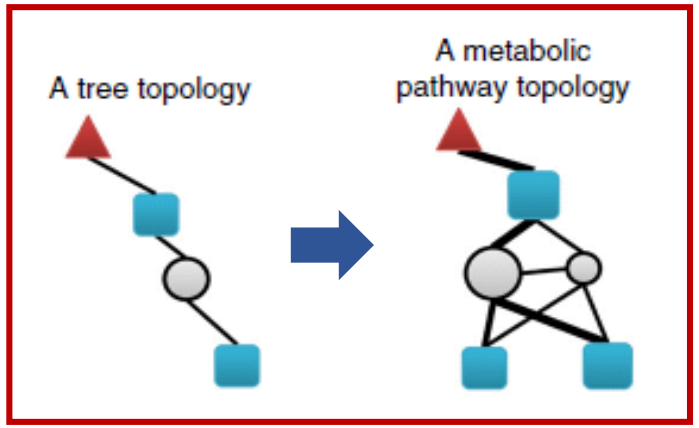


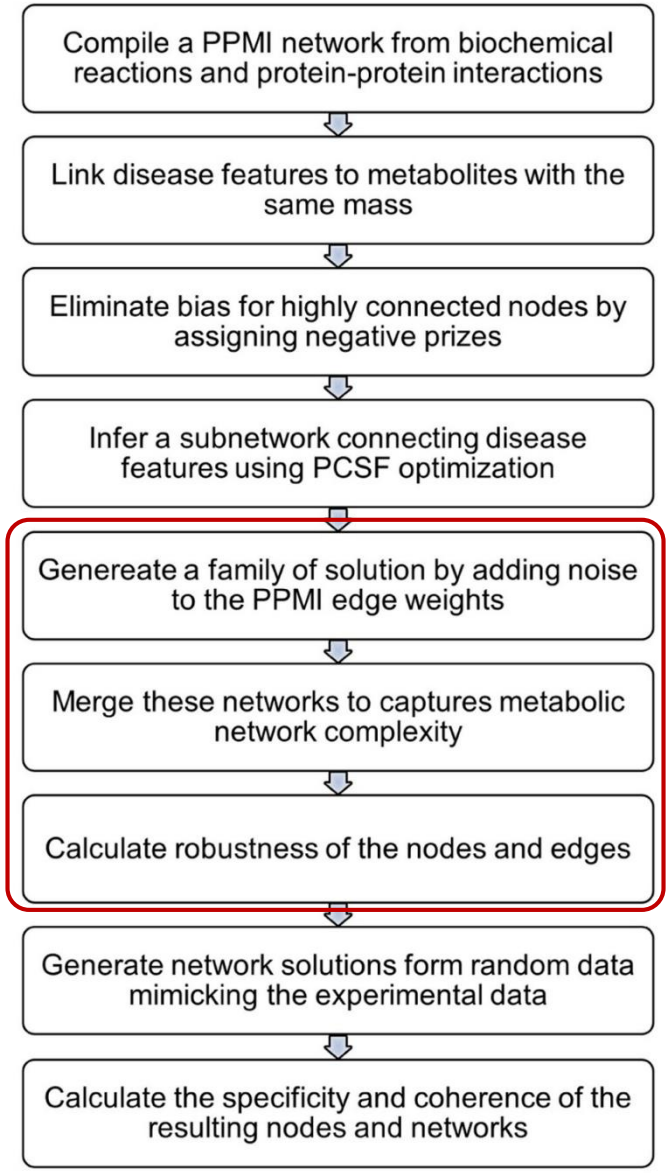
Methods

- Generate a family of networks to infer the complex interactions
  - Add small random noise to the PPMI edge weights
  - Union of multiple solutions form the network
  - Calculate robustness score

$$R_{n_i} = \frac{\sum_{j=1}^{R} f_{n_i,j}}{\sum_{i=1}^{N}\sum_{j=1}^{R} f_{n_i,j}}$$

$$f_{n_i,j} = \begin{cases} 1 \text{ if } n_i \in F_j(n) \\ 0 \text{ otherwise} \end{cases}$$



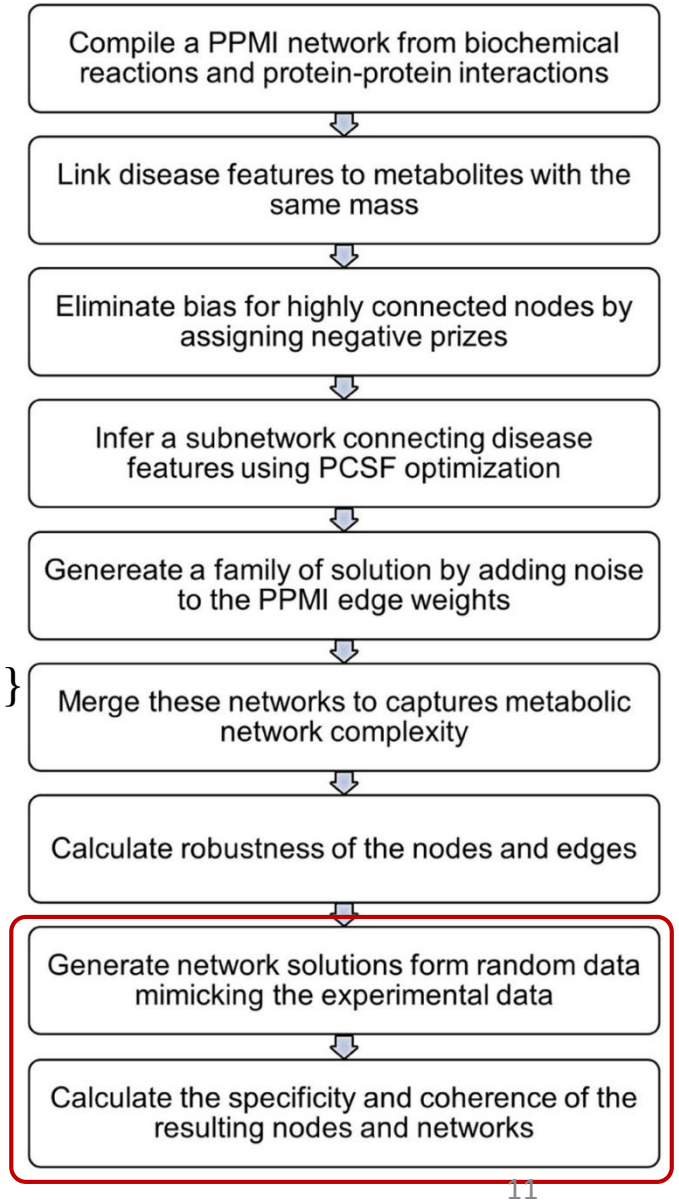


Methods

- Calculate disease-specific score for each resulting node and each network
  - Define Detectable Metabolite Feature (DMF) to mimick experimental data
  - Random sampling from DMF to generate solutions
  - Calculate node specificity and network specificity

$$DMF = \{\text{dmf} \mid \exists M_{\text{dmf}} : \text{PPMI}(M) \bigcap \text{m/ z}_{\text{min}} \le m / z_{\text{dmf}} \le m / z_{\text{max}} \bigcap \exists M_{\text{dmf}} : \text{detectable superclass}\}$$

$$\text{Node specificity}(\text{n}_\text{i}) = \frac{\sum_{i=1}^{R} f_i}{R} \qquad f_{n_i,j} = \begin{cases} 1 \text{ if } n \in RF_j(n) \\ 0 \text{ otherwise} \end{cases}$$

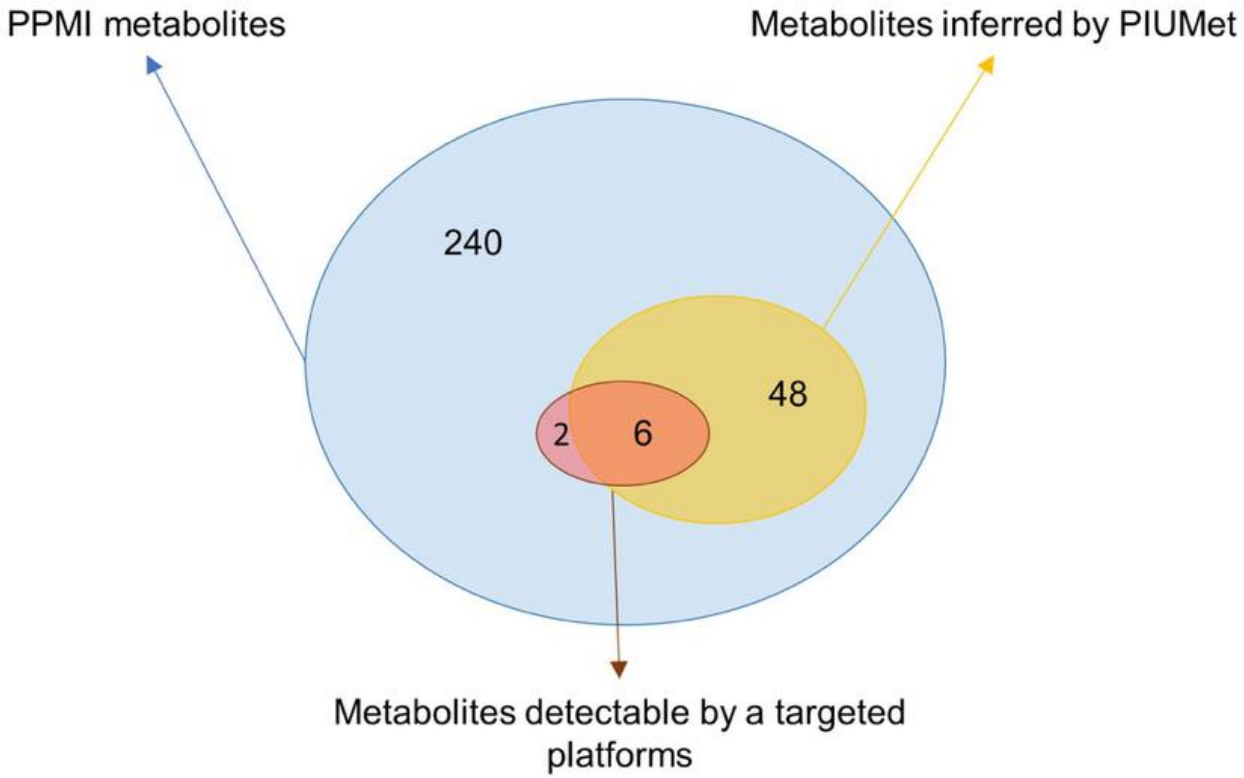


Results

- Analyze untargeted lipidomics and phosphoproteomics from a cell-line model of Huntingyton's Disease (HD)
- Use conditionally immortalized striatal cell lines (STHdh)
  - Control cell: wild type embryos (STHdh Q7)
  - Disease cell: knock-in embryos (STHdh Q111)
- Verification of experimental results
  - Targeted metabolomic platform
  - Two reverse-phase LC methods and MS data acquired in positive and negative ionization modes using two LC-MS system
  - Multiparameter high-content imaging for the analysis of cell apoptosis
  - Protein extraction and western blot analysis

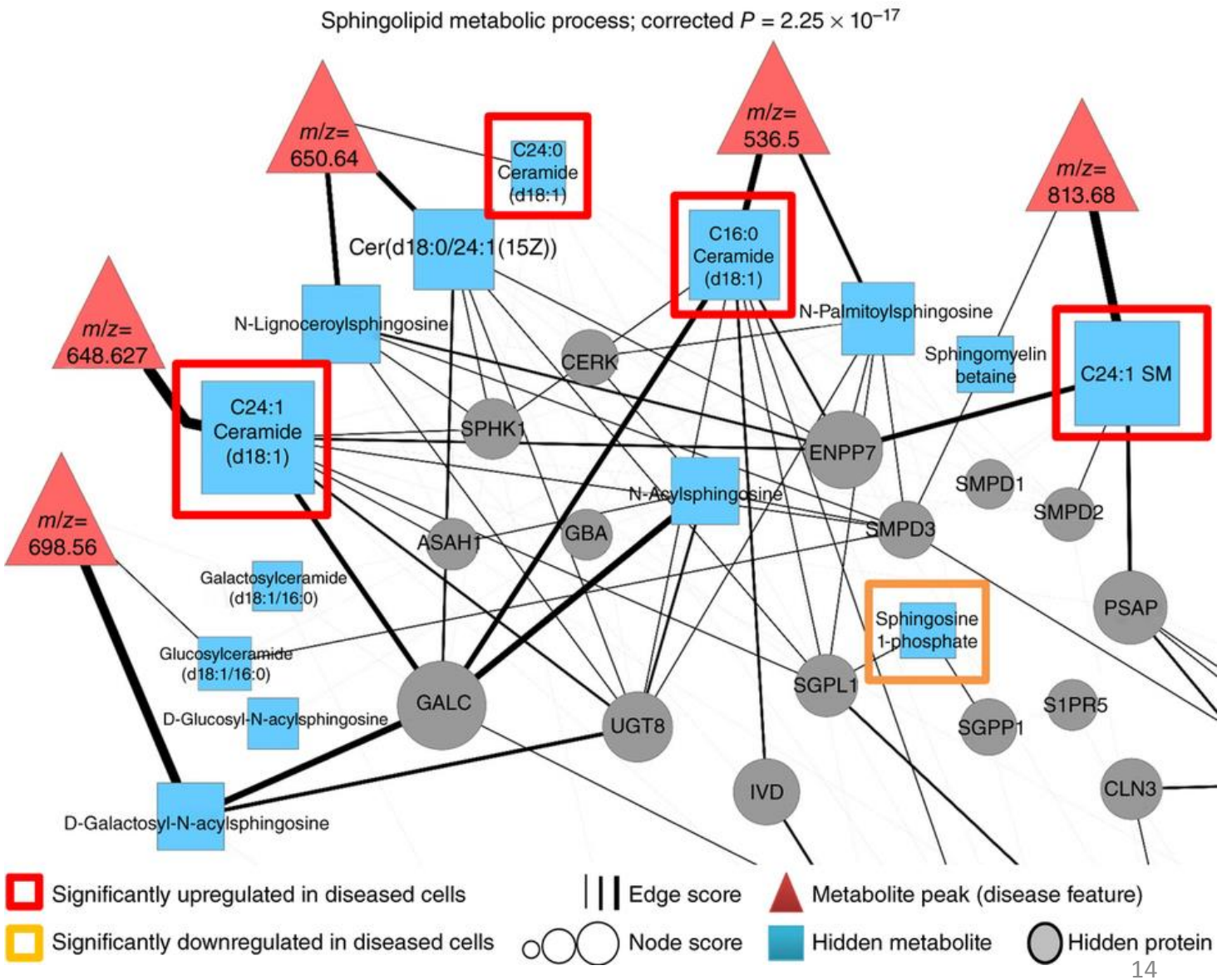
Results

- Analyze untargeted lipidomics and phosphoproteomics from a cell-line model of Huntingyton's Disease (HD)
- Use conditionally immortalized striatal cell lines (STHdh Q7 and STHdh Q111)



Results

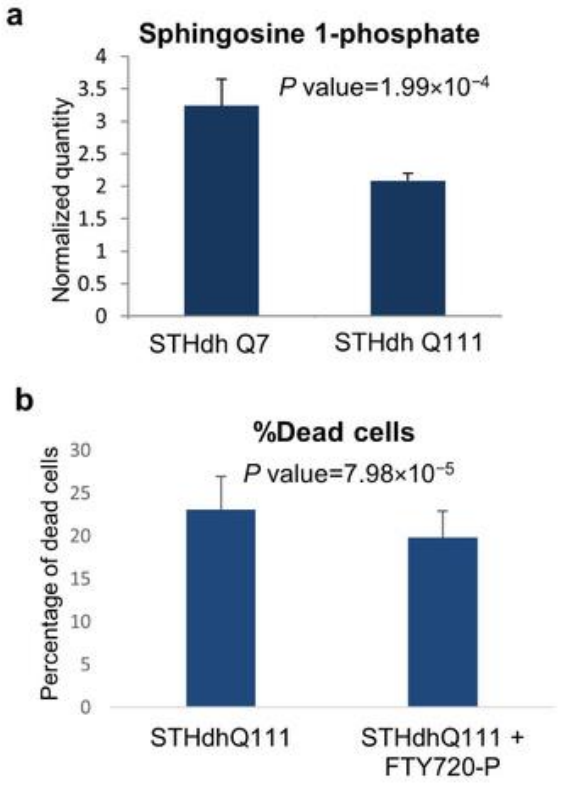
- PIUMet subnetwork shows altered sphingolipid metabolism in the HD model
- PIUMet identified S1P as hidden metabolite

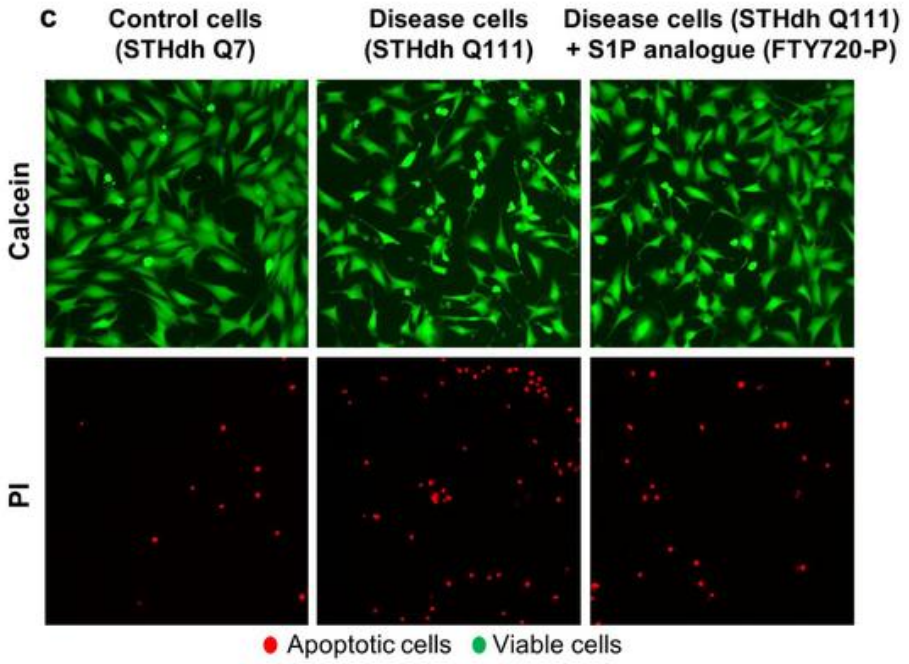


Results

- PIUMet subnetwork shows altered sphingolipid metabolism in the HD model
- PIUMet identified S1P as hidden metabolite

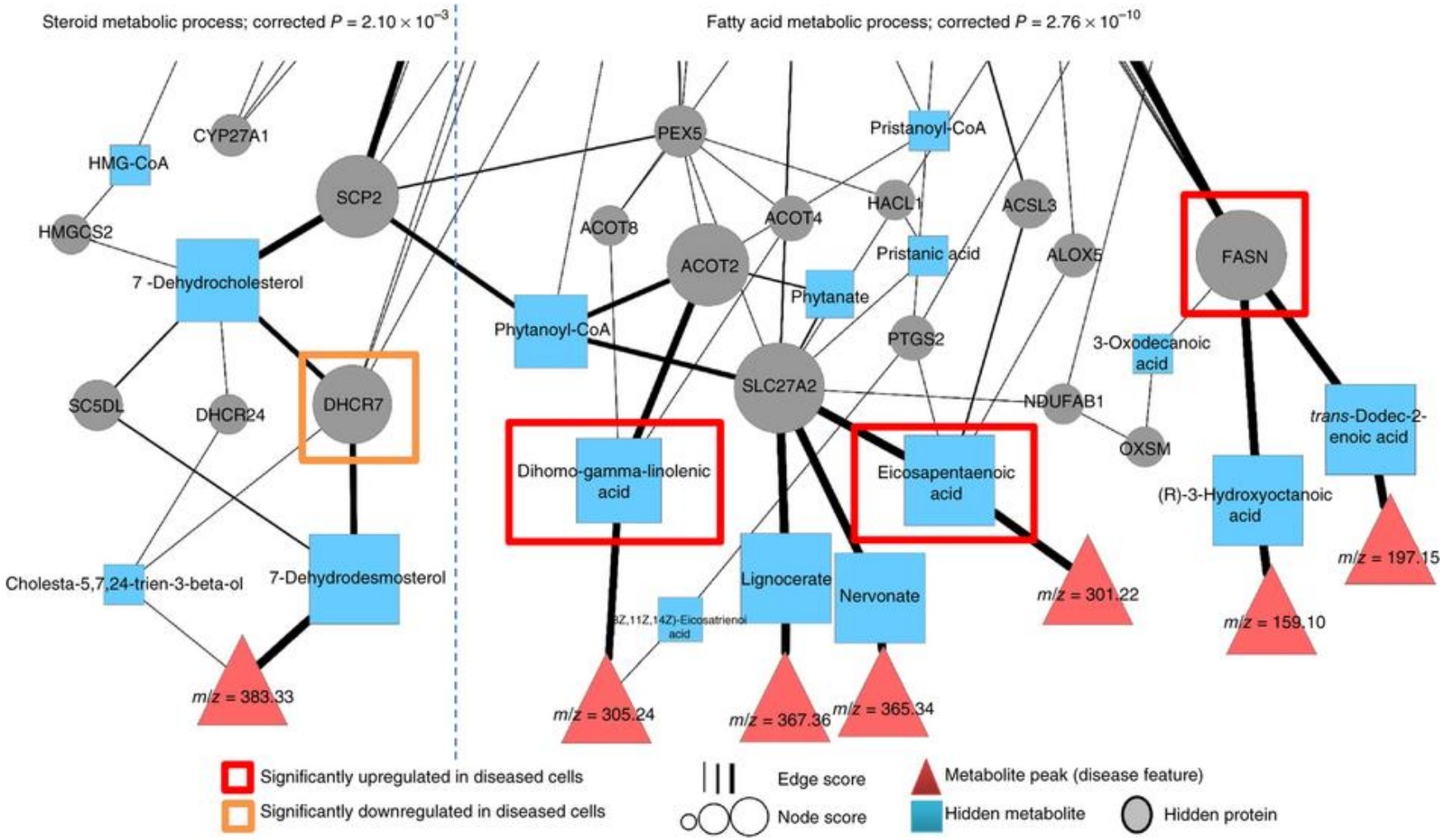
➢ S1P is a key signaling molecule that activates antiapoptotic pathways





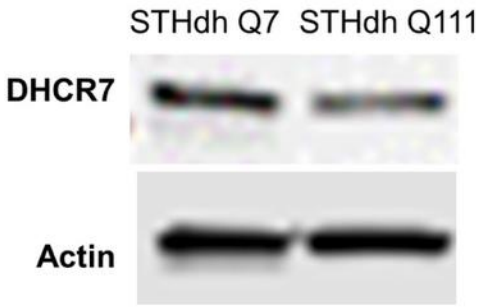
Results

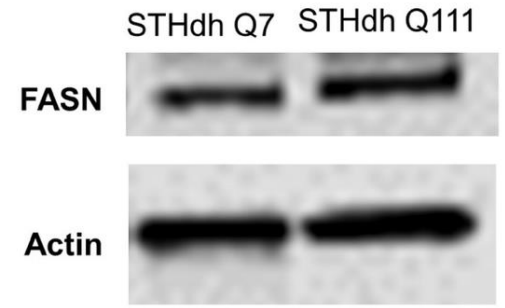
- PIUMet discovers an altered steroid metabolism network in the HD model
- PIUMet subnetwork is also associated with fatty acid metabolism

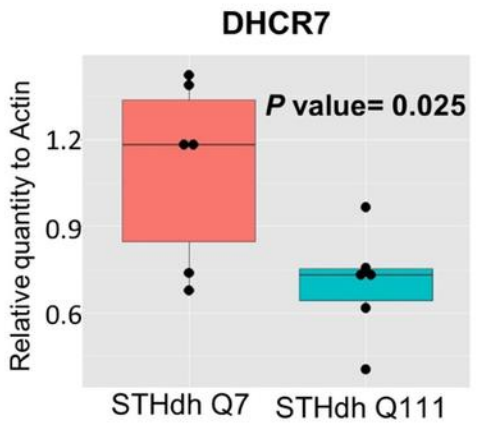


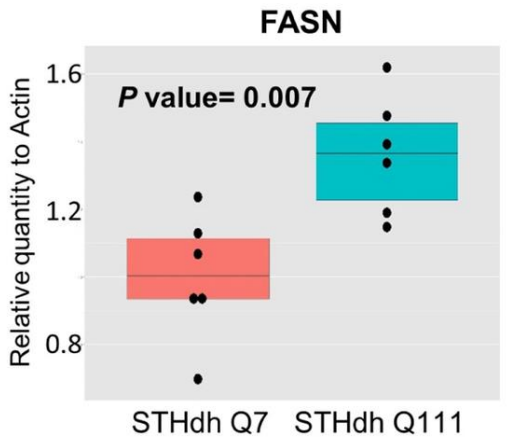
Results

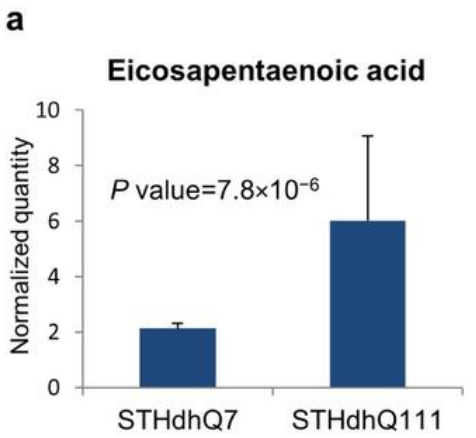
- PIUMet discovers an altered steroid metabolism network in the HD model
- PIUMet subnetwork is associated with fatty acid metabolism

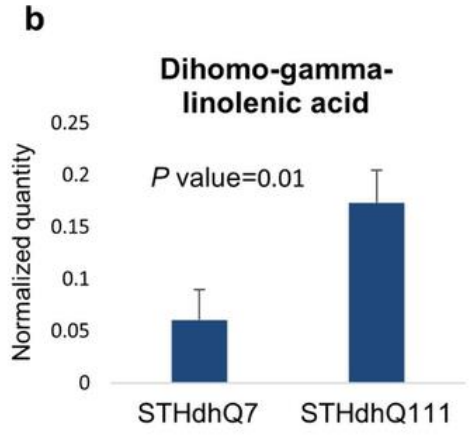






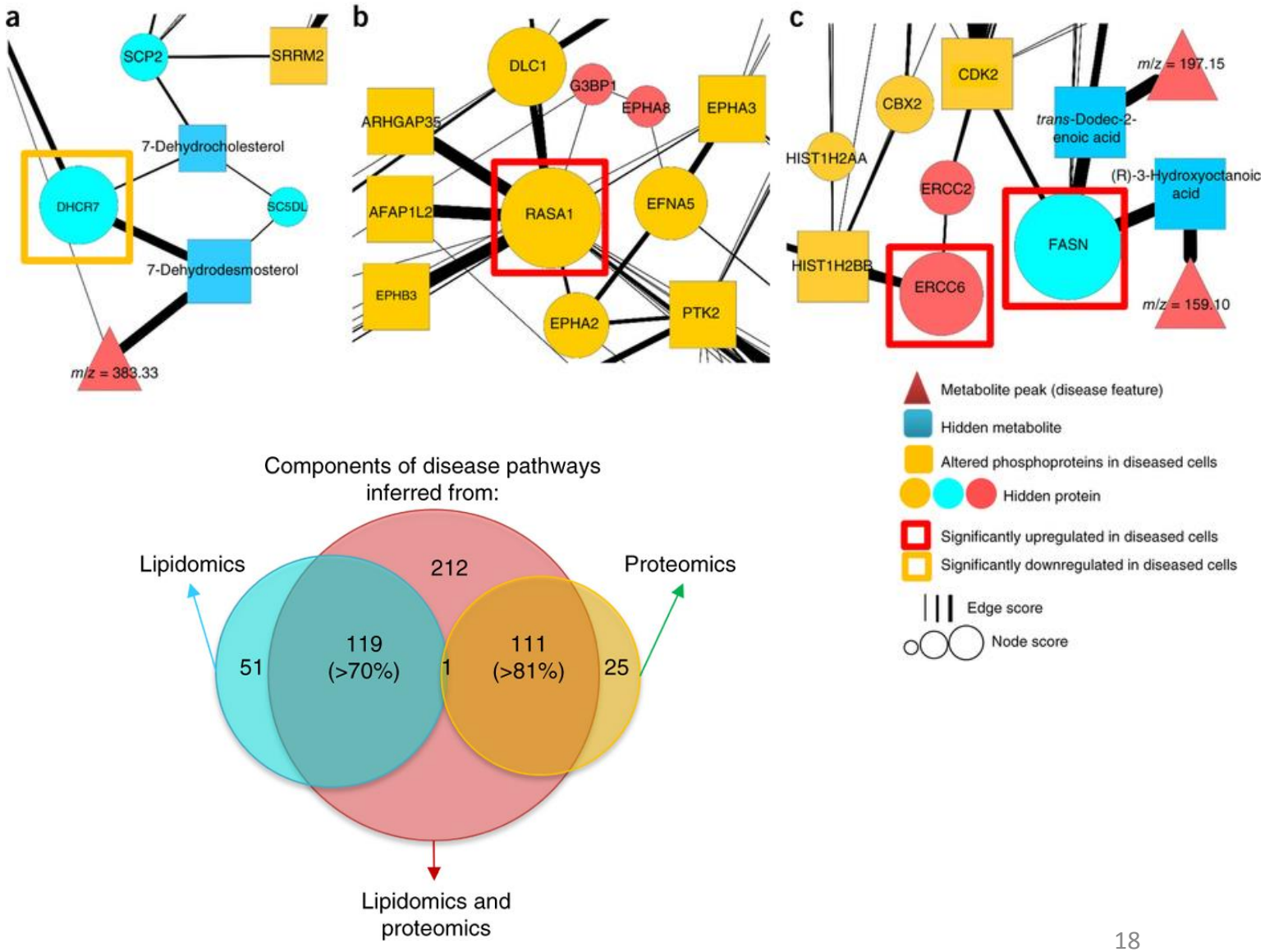






Results

- Integrate metabolomics with phosphoproteomics
- Multi-omic analysis inferred the majority of hidden components obtained from individual analysis
- It also revealed new disease-associated molecules



Conclusion

- PIUMet leverages known metabolic reactions and protein-protein interactions to analyze the ambiguous assignment of metabolomics features
- PIUMet can be used to identify dysregulated metabolic networks
- PIUMet is a general approach that be applied to other diseases
- It has the potential to systematically discover new molecular mechanisms
- It prioritizes metabolite features for experimental validation
- It contextualizes metabolomics with other system-level molecular data

Thank You!