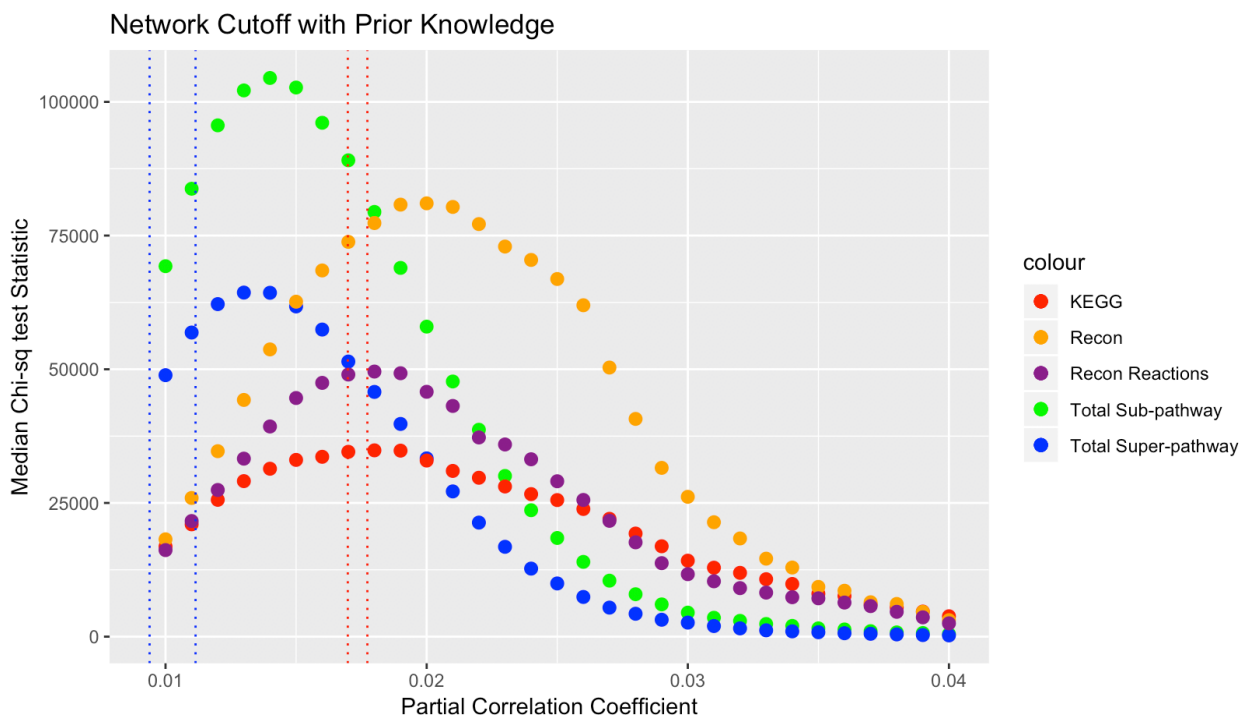


Supplementary Materials

SUPPLEMENTARY FIGURE 1 - OPTIMUM NETWORK CUTOFF SEARCH FOR QMDIAB DATA	2
SUPPLEMENTARY FIGURE 2 - QMDIAB ENRICHED MODULE COMPOSITION	3
SUPPLEMENTARY FIGURE 3 – ROS/MAP ENRICHED MODULE COMPOSITION	4
SUPPLEMENTARY FIGURE 4 – EDGE REPRESENTATION IN MODENTIFY AND MEGENA NETWORKS	10
SUPPLEMENTARY FIGURE 5 - QMDIAB ANALYSIS WITH WGCNA CLUSTERING METHOD	11
SUPPLEMENTARY FIGURE 6 – ENRICHMENT PEAK THRESHOLD METRICS AND ANALYSIS	13
SUPPLEMENTARY FIGURE 7 - ENRICHMENT PEAK IDENTIFICATION AND “PIGGY-BACKERS”	15
REFERENCES	17

Supplementary Figure 1 - Optimum network cutoff search for QMDiab data



Supplementary Figure 1. Prior knowledge-informed optimum network cutoff search for QMDiab data.

Cutoff optimization graph for QMDiab dataset using annotations from the KEGG database (red) and Recon 3D database (molecular annotations in orange, reaction annotations in purple), as well as the sub- (green) and super- (blue) pathway annotations of the Metabolon platform. Bonferroni-adjusted p-value cutoffs are demarcated in blue vertical dashed lines at $p = 0.01$ (left) and $p = 0.05$ (right), while FDR-adjusted cutoffs are similarly demarcated in red.

Using the method outlined in Benedetti et al¹, biomolecules of the QMDiab dataset were annotated using annotations from the KEGG² and Recon 3D³ databases, as well as sub- and super- pathway annotations that describe biochemical processes and broader metabolite groups, respectively, for molecules measured on Metabolon platforms. Briefly, we created a Gaussian graphical model (GGM) of the QMDiab biomolecules and set the network cutoff at a partial correlation coefficient ranging from 0.01 to 0.04 (a partial correlation value greater than the cutoff would indicate an edge present in the network). For each annotation set, we performed a Chi-square test on an overlap between the GGM and the known pathways. Specifically, we calculated a contingency table which classified pairs of biomolecules based on whether an edge between them appears in the GGM and they share the same annotation (true positive), they only share a GGM edge (false positive), they only share an annotation (false negative), or they share neither an edge nor an annotation (true negative). The optimal cutoff is defined as that with the highest test statistic. For the QMDiab dataset, the optimal cutoff varied per annotation set, with the omic specific annotations (sub- and super- pathways) being optimal at a more stringent cutoff, while the cross-omic annotation sets (KEGG and Recon) being optimal at a higher coefficient. This lack of agreement illustrates the correlation-based platform bias within the QMDiab dataset, as there does not exist a single optimal correlation cutoff.

Supplementary Figure 2 - QMDiab Enriched Module Composition

Cluster ID	Size	Enrichment	Height	Dataset Distribution	Pathway Distribution	Cluster ID	Size	Enrichment	Height	Dataset Distribution	Pathway Distribution
5051	192	0.54	0.88			3714	3	0.67	0.55		
4081	7	0.86	0.63			3771	3	1	0.56		
1869	6	0.67	0.22			71	2	1	0		
3275	6	0.67	0.46			72	2	1	0		
2663	5	0.8	0.35			75	2	1	0		
3057	4	0.75	0.42			91	2	1	0		
1816	3	1	0.21			872	2	1	0.09		
2264	3	0.67	0.28			885	2	1	0.1		
2452	3	0.67	0.31			956	2	1	0.1		
2529	3	0.67	0.32			1584	2	1	0.18		
						2268	2	1	0.28		

Dataset Annotations

- Metabolon Urine
- Metabolon Plasma (HD4)
- Metabolon Plasma (HD2)
- Somalogic Proteomics
- Chenomix NMR Metabolomics
- Biocrates Metabolomics
- Nightingale Metabolomics
- Metabolon Saliva
- Metabolon Lipidomics

Pathway Annotations

- Amino acid
- No pathway information
- Carbohydrate
- Lipid
- Peptide
- Xenobiotics
- Energy
- Nucleotide
- Cofactors and Vitamins

Supplementary Figure 2. Table of QMDiab module information. Listed are the modules' ID (for cross referencing with **Supplementary Data 1**), size (number of molecules), enrichment (proportion of significant phenotype-associated molecules), height on the hierarchy, and distributions of the members' datasets and pathways (Metabolon platforms only).

Supplementary Figure 3 – ROS/MAP Enriched Module Composition

Cluster ID	Size	Enrichment		Height	Platform Distribution	Pathway Distribution	Cluster ID	Size	Enrichment		Height	Platform Distribution	Pathway Distribution
		NFT	CD						NFT	CD			
7124	165	0.52	0.42	0.8			6191	22	0.73	0.27	0.74		
4366	105	0.53	0.3	0.63			3523	19	0.52	0.05	0.58		
5258	88	0.57	0.52	0.68			4232	18	0.56	0.28	0.63		
4031	64	0.22	0.56	0.61			4908	16	0.94	0.56	0.66		
6566	62	0.52	0.26	0.76			1590	16	0.13	0.56	0.43		
5396	52	0.6	0.25	0.69			767	16	0.81	0.56	0.32		
6658	44	0.3	0.66	0.77			5090	15	0.93	0.53	0.67		
4157	38	0.82	0.53	0.63			3575	14	0.64	0.71	0.59		
795	37	0	0.51	0.32			4216	13	0.54	1	0.63		
4978	31	0.65	0.52	0.67			5311	12	0.83	0.83	0.69		
3924	29	0.69	0.55	0.61			1841	12	0.83	0.67	0.45		
2183	28	0.61	0.43	0.48			751	12	0.75	0.58	0.32		
4570	25	0.64	0.12	0.65			3447	11	0.1	0.55	0.58		
7643	22	0.73	0.41	0.85			3089	11	0.55	0.1	0.55		
							4559	10	0.2	0.8	0.64		
							3614	10	0.8	0.6	0.59		

Platform Annotations

- Metabolomics
- Proteomics

Pathway Annotations

- Amino acid
- No pathway information
- Carbohydrate
- Lipid
- Peptide
- Xenobiotics
- Energy
- Nucleotide
- Cofactors and Vitamins

Cluster ID	Size	Enrichment		Height	Platform Distribution	Pathway Distribution
		NFT	CD			
7344	9	0.56	0	0.82		
6353	9	0.67	0.33	0.75		
6109	9	0.33	0.56	0.73		
5718	9	0.56	0.33	0.71		
4109	9	0.56	0.11	0.62		
2931	9	0.89	0.56	0.54		
2626	9	0.56	0.22	0.52		
6223	8	0.38	0.63	0.74		
7442	7	0.57	0	0.83		
5432	7	0.57	0.43	0.69		
5189	7	0.14	0.57	0.68		
5029	7	0.86	0.29	0.67		
4637	7	0.43	0.71	0.65		
2749	7	0.29	0.71	0.53		

Cluster ID	Size	Enrichment		Height	Platform Distribution	Pathway Distribution
		NFT	CD			
1552	7	0.86	0.57	0.42		
3770	6	0.17	1	0.6		
2261	6	0.83	0.5	0.49		
1435	6	0.67	0.83	0.41		
137	6	0.67	0.17	0.13		
5765	5	0	0.6	0.71		
5320	5	0.6	0.6	0.69		
4502	5	0.8	0	0.64		
4411	5	0.2	0.6	0.64		
2499	5	1	0.6	0.51		
2251	5	0.8	0.2	0.49		
2244	5	0.6	0.6	0.49		
1847	5	0.4	0.6	0.45		
1551	5	0.6	0.2	0.42		
996	5	0.6	0.2	0.36		
535	5	0	0.6	0.27		

Platform Annotations

- Metabolomics
- Proteomics

Pathway Annotations

- Amino acid
- No pathway information
- Carbohydrate
- Lipid

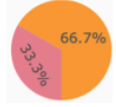
- Peptide
- Xenobiotics
- Energy
- Nucleotide
- Cofactors and Vitamins

Cluster ID	Size	Enrichment		Height	Platform Distribution	Pathway Distribution
		NFT	CD			
5694	4	0.75	0.75	0.71		
4736	4	0.75	0	0.66		
4267	4	0.5	0.75	0.63		
3530	4	0.75	1	0.58		
3429	4	0	0.75	0.58		
3333	4	1	0.75	0.57		
3060	4	0.75	0.25	0.55		
3002	4	0.75	0.75	0.55		
2091	4	1	0.25	0.47		
1684	4	0.25	1	0.44		
610	4	0.75	0.75	0.29		
5983	3	1	1	0.72		
5922	3	0.67	0	0.72		
5753	3	0.33	0.67	0.71		

Cluster ID	Size	Enrichment		Height	Platform Distribution	Pathway Distribution
		NFT	CD			
5642	3	1	0	0.7		
5287	3	0	0.67	0.69		
5121	3	0.33	0.67	0.68		
4849	3	0	0.67	0.66		
4781	3	0.67	0.33	0.66		
4757	3	0.67	0.33	0.66		
4617	3	0.67	0.33	0.65		
4525	3	0.67	1	0.64		
4470	3	0.33	1	0.64		
4427	3	0.67	0	0.64		
4325	3	0.33	0.67	0.63		
4290	3	0.33	0.67	0.63		
4022	3	0	0.67	0.61		
3986	3	0	0.67	0.61		
3840	3	0.33	0.67	0.6		
3718	3	0	0.67	0.59		

Platform Annotations	Pathway Annotations	Peptide
Metabolomics	Amino acid	Xenobiotics
Proteomics	No pathway information	Energy
	Carbohydrate	Nucleotide
	Lipid	Cofactors and Vitamins

Cluster ID	Size	Enrichment		Height	Platform Distribution	Pathway Distribution
		NFT	CD			
3639	3	0.33	0.67	0.59	100%	100%
3678	3	0.67	1	0.59	100%	100%
3480	3	0.67	0.33	0.58	100%	100%
3396	3	0.67	1	0.57	100%	100%
3241	3	0	0.67	0.56	100%	100%
3166	3	1	0.33	0.56	100%	100%
2636	3	0.67	0	0.52	100%	100%
2332	3	1	1	0.5	100%	100%
2309	3	0.67	0	0.49	100%	100%
2276	3	0	0.67	0.49	100%	100%
2214	3	0.33	0.67	0.49	100%	100%
2005	3	0.67	1	0.47	100%	100%
2001	3	0.67	0.33	0.47	100%	100%
1737	3	1	0.67	0.44	100%	100%

Cluster ID	Size	Enrichment		Height	Platform Distribution	Pathway Distribution
		NFT	CD			
1540	3	1	0.67	0.42	100%	100%
1493	3	0	0.67	0.42	100%	100%
1189	3	0.67	1	0.38	100%	
762	3	0.67	0.67	0.32	100%	100%
657	3	0.67	0	0.3	100%	100%
500	3	1	0.67	0.26	100%	100%
36	3	0.67	0	0.05	100%	100%
5	3	1	0	0.02	100%	100%
6214	2	0	1	0.74	100%	100%
5407	2	1	0	0.69	100%	100%
4491	2	0.5	1	0.64	100%	100%
4462	2	1	0.5	0.64	100%	100%
4342	2	0	1	0.63	100%	100%
4175	2	0	1	0.62	100%	100%
4085	2	1	0.5	0.62	100%	100%
3882	2	1	0.5	0.6	100%	100%

Platform Annotations

- Metabolomics
- Proteomics

Pathway Annotations

- Amino acid
- No pathway information
- Carbohydrate
- Lipid

- Peptide
- Xenobiotics
- Energy
- Nucleotide
- Cofactors and Vitamins

Cluster ID	Size	Enrichment		Height	Platform Distribution	Pathway Distribution
		NFT	CD			
3823	2	1	0	0.6	100%	100%
3742	2	1	0	0.6	100%	100%
3578	2	0	1	0.59	100%	100%
3502	2	0	1	0.58	100%	100%
3470	2	1	0	0.58	100%	100%
3344	2	1	0.5	0.57	100%	100%
3292	2	0	1	0.57	100%	100%
3225	2	0.5	1	0.56	100%	50% Amino acid, 50% Energy
3112	2	1	0.5	0.55	100%	100%
3074	2	0.5	1	0.55	100%	100%
3044	2	0	1	0.55	100%	100%
3023	2	0	1	0.55	100%	100%
2960	2	0.5	1	0.54	100%	100%
2932	2	1	0.5	0.54	100%	100%

Cluster ID	Size	Enrichment		Height	Platform Distribution	Pathway Distribution
		NFT	CD			
2715	2	0	1	0.53	100%	100%
2707	2	1	0	0.52	100%	100%
2683	2	1	0	0.52	100%	100%
2619	2	1	0.5	0.52	100%	100%
2604	2	1	1	0.52	100%	100%
2526	2	0	1	0.51	100%	100%
2496	2	0	1	0.51	100%	100%
2462	2	1	0	0.51	100%	100%
2396	2	1	1	0.5	100%	50% Peptide, 50% Xenobiotics
2125	2	0	1	0.48	100%	100%
2038	2	1	1	0.47	100%	50% Energy, 50% Cofactors and Vitamins
2020	2	1	0	0.47	100%	100%
2010	2	0	1	0.47	100%	100%
1928	2	1	1	0.46	100%	100%
1896	2	1	0.5	0.46	100%	100%
1814	2	0	1	0.45	100%	100%

Platform Annotations

- Metabolomics
- Proteomics

Pathway Annotations

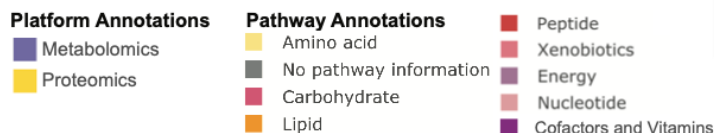
- Amino acid
- No pathway information
- Carbohydrate
- Lipid

- Peptide
- Xenobiotics
- Energy
- Nucleotide
- Cofactors and Vitamins

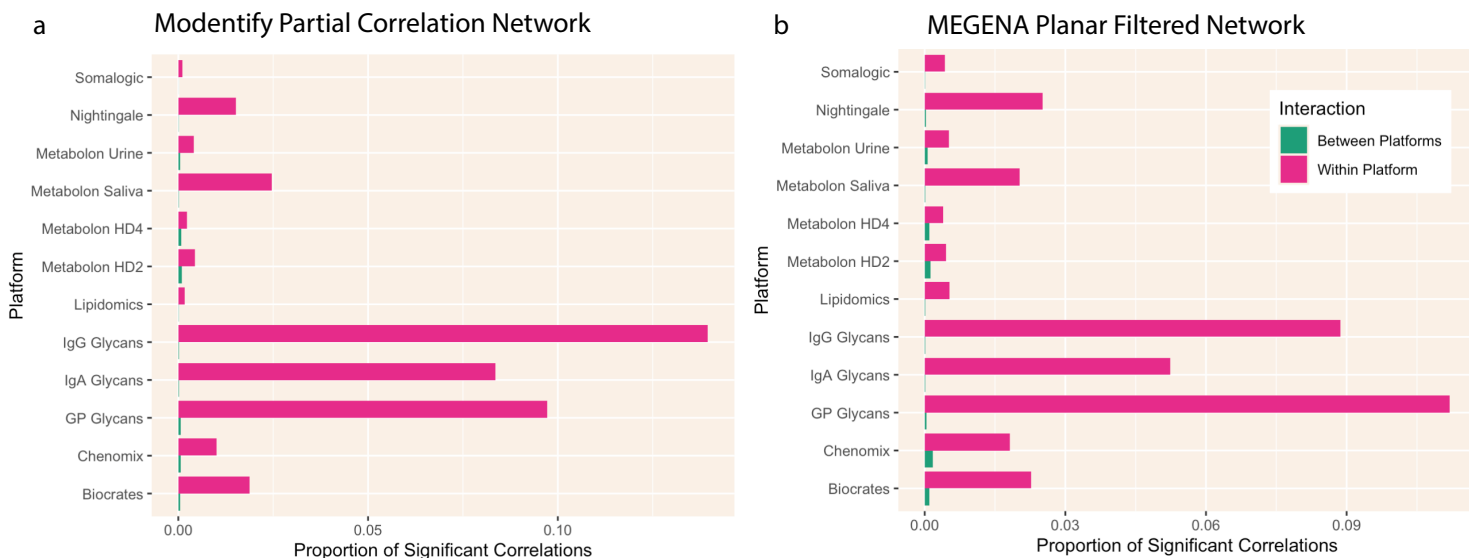
Cluster ID	Size	Enrichment		Height	Platform Distribution	Pathway Distribution
		NFT	CD			
1717	2	0	1	0.44	100%	100%
1529	2	1	0.5	0.42	100%	100%
1476	2	1	1	0.41	100%	100%
1474	2	1	0	0.41	100%	100%
1462	2	0	1	0.41	100%	100%
1315	2	1	1	0.39	100%	100%
1281	2	0	1	0.39	100%	100%
1181	2	1	0	0.38	100%	100%
1125	2	1	1	0.37	100%	100%
1116	2	1	1	0.37	100%	100%
1029	2	0	1	0.36	100%	100%
969	2	1	0	0.35	100%	100%
882	2	0.5	1	0.34	100%	100%
845	2	1	1	0.33	100%	100%

Cluster ID	Size	Enrichment		Height	Platform Distribution	Pathway Distribution
		NFT	CD			
821	2	1	0	0.33	100%	100%
769	2	1	1	0.32	100%	50% 50%
682	2	1	1	0.3	100%	100%
562	2	1	0	0.27	100%	100%
211	2	1	1	0.16	100%	100%
123	2	0	1	0.12	100%	100%

Supplementary Figure 3. Table of ROS/MAP module information. Listed are the modules' ID (for cross referencing with **Supplementary Data 2**), size (number of molecules), Neurofibrillary Tangles (NFT) and Cognitive Decline (CD) phenotype enrichment (proportion of significant phenotype-associated molecules), height on the hierarchy, and distributions of the members' datasets and pathways (metabolomics platform only).

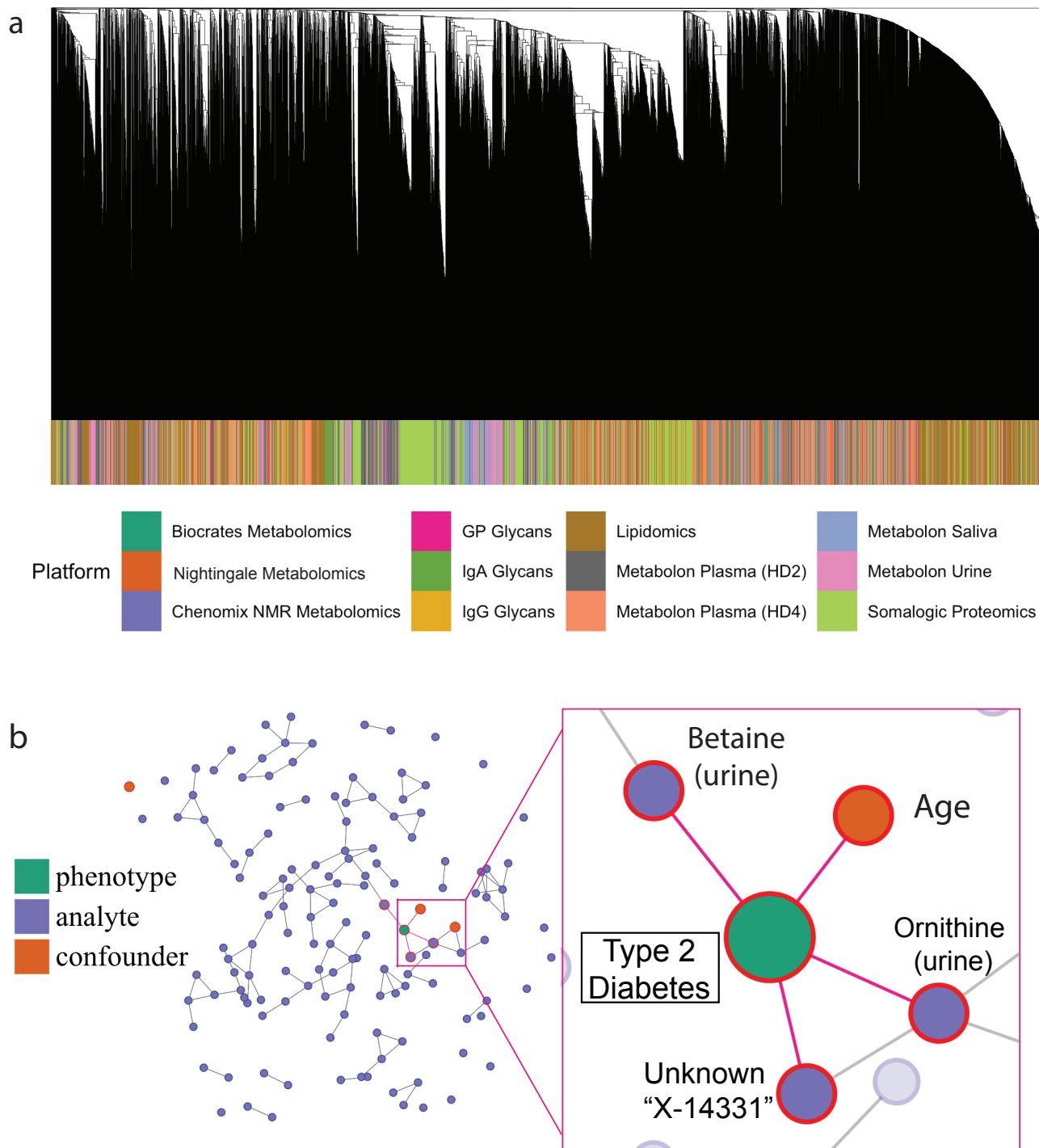


Supplementary Figure 4 – Edge representation in MoDentify and MEGENA networks



Supplementary Figure 4. Proportion of edges within platform and between platforms, normalized by all within-platform and between-platform edges, respectively of the **a)** MoDentify Partial Correlation Network with p-value cutoff of 0.01 (Bonferroni adjusted) and **b)** the MEGENA Planar Filtered Network. The platform correlation bias affects both methods similarly, overrepresenting within-platform edges as compared to between-platform edges.

Supplementary Figure 5 - QMDiab analysis with WGCNA clustering method



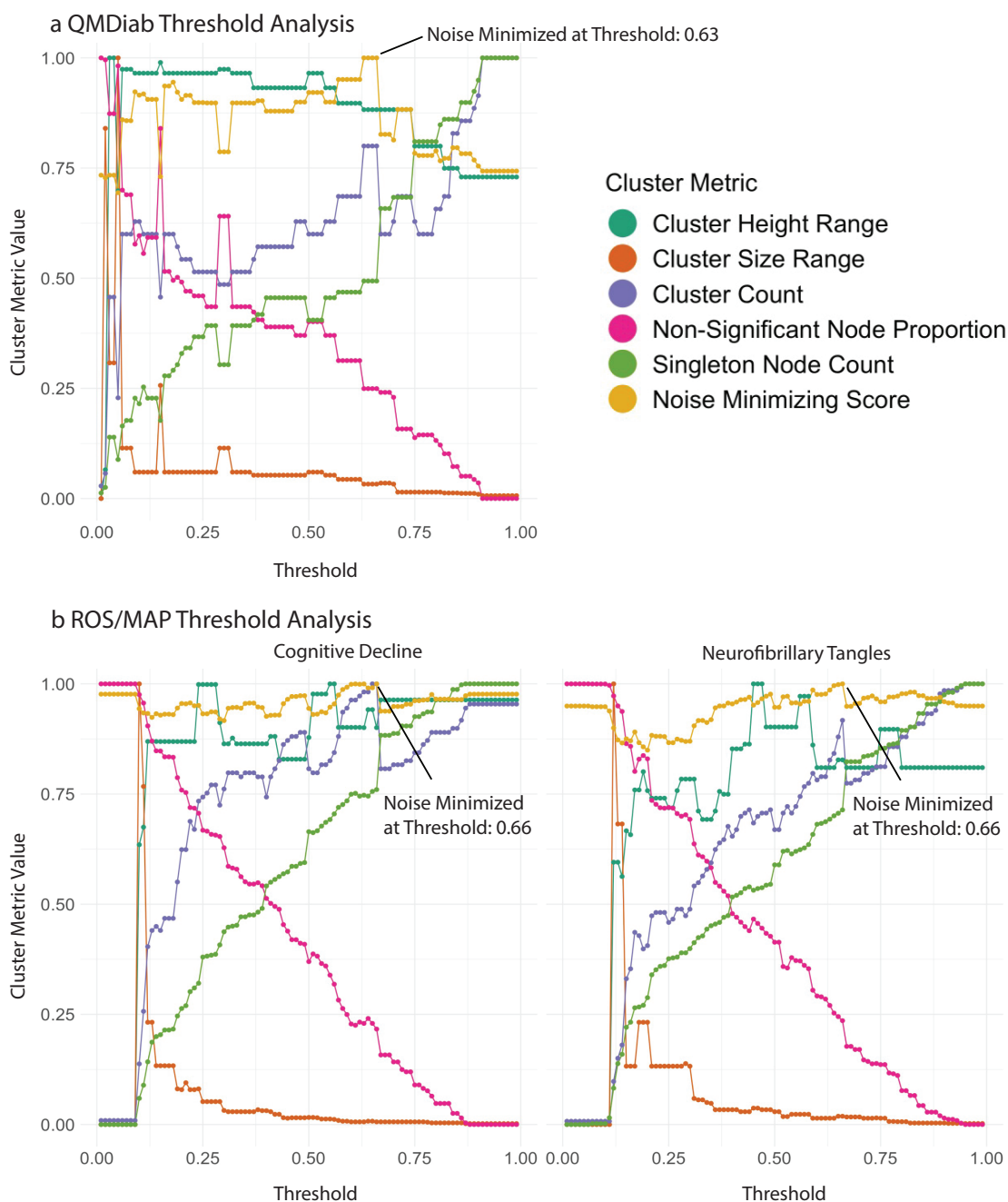
Supplementary Figure 5. a) Platform distribution in the WGCNA hierarchical structure on the QMDiab dataset. Strong intra-platform correlations can be seen for proteomics (green) and to a lesser extent for urine metabolomics (pink). **b)** Mixed graphical model of the 153 molecules in the largest cluster of the WGCNA analysis with phenotype and confounders. To the right, a zoomed in view of nodes with edges to the Type 2 Diabetes phenotype which include ornithine and betaine in urine along with the confounder age and one unknown molecule. Both ornithine in urine and that specific unknown were found with AutoFocus' original hierarchical clustering method.

To investigate how a clustering method can affect AutoFocus cluster analysis, the QMDiab dataset was rerun using WGCNA's TOM-based hierarchical structure⁴. This structure was similarly susceptible to the problem of platform bias within the resulting hierarchy; while the TOM-based hierarchical clustering more substantially distributed the lipidomics data with the other platforms as compared to correlation-based clustering, the proteomics data was largely segregated, as large regions of the tree contain only proteins (**Supplementary Figure 5a**).

The T2D enrichment step identified 13 modules within the TOM hierarchy, including a large, multi-omic cluster of 153 molecules that contained many of the same molecules as the energy metabolism module found with the correlation-based hierarchy. In contrast to the results in the main manuscript, no bone degradation proteins were identified in this cluster, potentially due to the proteomics clustering bias mentioned above. This difference in module composition also changed the MGM driver analysis, as all the drivers identified in the 153-molecule module were from urine platforms (**Supplementary Figure 5b**). Despite this, both ornithine and one unknown metabolite (labeled "X-14331") were maintained across these two analyses.

It is worth noting that when the hierarchy is created using a TOM-based distance matrix and average linkage as is used in the original WGCNA analysis⁴, the resulting tree can have a very high maximum depth (the maximum number of internal nodes between the tree's root and a leaf). In the instance of large datasets such as QMDiab, a high maximum depth hinders tree visualization due to memory limitations. As such, results presented here were explored without the aid of AutoFocus' visual interface.

Supplementary Figure 6 – Enrichment Peak Threshold Metrics and Analysis



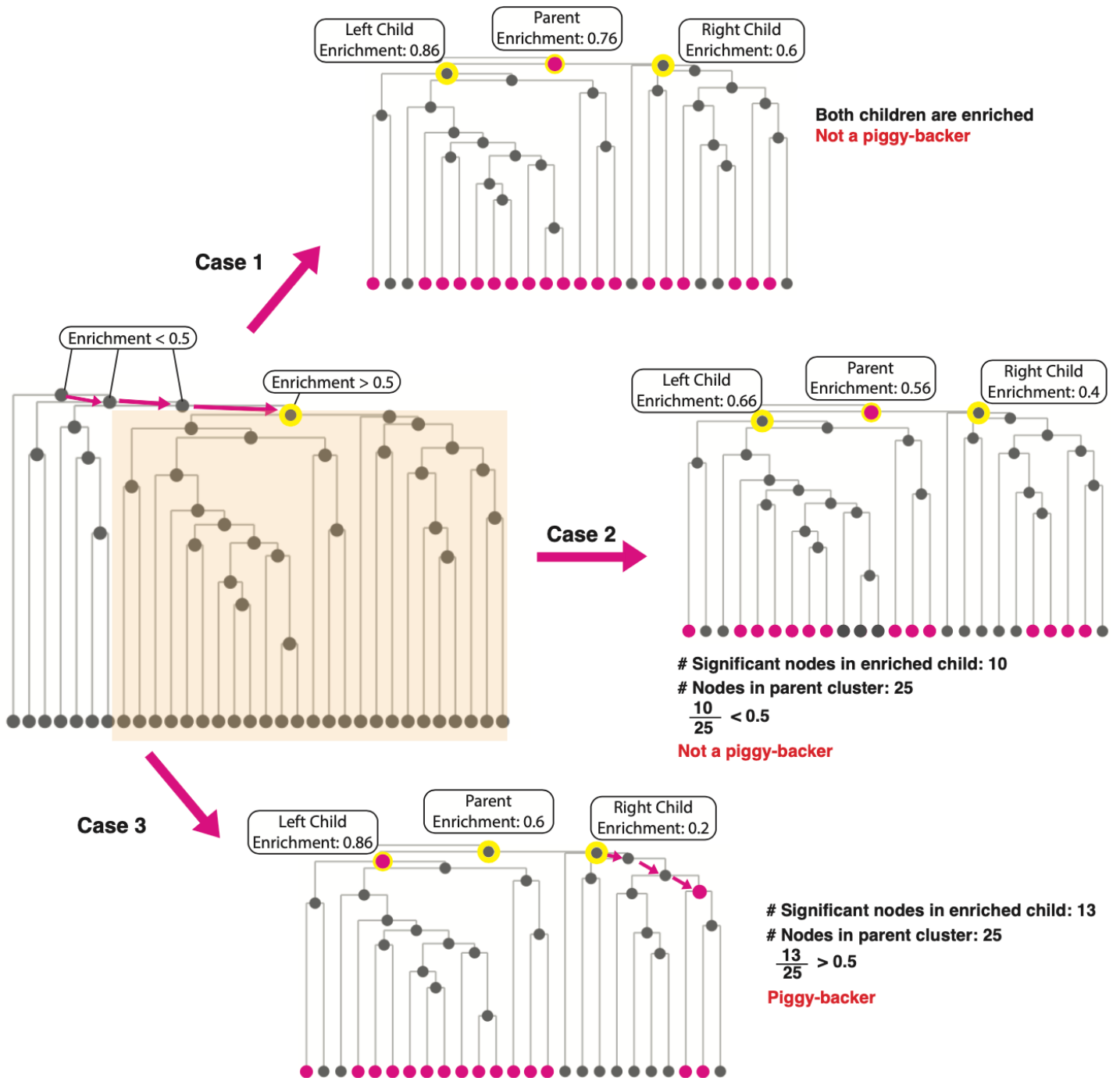
Supplementary Figure 6. Enrichment Threshold impact on cluster features analysis for the a) QMDiab dataset, and b) ROS/MAP dataset on both Cognitive Decline and Neurofibrillary tangles phenotypes.

The AutoFocus R package contains functionality to perform an analysis assessing the impact of enrichment threshold on the returned clusters. The metrics assessed in this analysis are 1. The range of cluster heights (height of the highest cluster minus height of the lowest cluster), 2. The range of cluster sizes (size of the largest cluster minus size of the smallest cluster), 3. The number of clusters, 4. The proportion of nodes in clusters that are not significantly associated to the desired phenotype ($\frac{\# \text{ non-significant nodes in clusters}}{\# \text{ of all nodes in clusters}}$), and 5. The number of significant nodes not included in any cluster with size greater than 2 (also called singletons).

In **Supplementary Figure 6**, these 5 metrics are calculated across thresholds between 0.01 and 0.99 at increasing increments of 0.01 for the QMDiab dataset, and the ROS/MAP dataset for both cognitive decline and neurofibrillary tangles. Each metric plotted is normalized to that metric's maximum across all thresholds, putting it on a scale of 0 to 1. For all three analyses, the cluster height and size ranges stabilize after a threshold of ~0.3. As expected, the number of singletons increases and the non-significant node proportion decreases as the threshold becomes more stringent, only returning clusters with very dense signal.

In order to provide a rough guideline for a potential threshold choice, we derived a 6th metric, the “Noise Minimizing Score” that combines the non-significant node proportion with the singletons, which is maximized when both are low, indicating the returned clusters are including significant molecules while excluding non-significant nodes. This is calculated by summing the normalized non-significant node proportion and the normalized number of singletons, inverting that summation such that higher indicates more optimal, and then normalizing to the maximum score across all thresholds to restrict the scale from 0 to 1. The maximum Noise Minimizing score is indicated in **Supplementary Figure 6**. Interestingly, this score peaks at a threshold of 0.66 for both phenotypes in the ROS/MAP dataset.

Supplementary Figure 7 - Enrichment Peak Identification and “Piggy-backers”



Supplementary Figure 7. Example visualization of “piggy-backer” filtration process through example scenarios. The enrichment threshold in this example is set to 0.5. Note that the piggy-backer threshold is the same as the enrichment threshold set when scanning the hierarchy for clusters; it is not an independent threshold.

When traversing the hierarchical structure from root to leaves to identify enrichment peaks, we run into the issue of “piggy-backers”, which are defined as peaks that reach the enrichment threshold only due to one child reaching the enrichment threshold, and the joining of the two children dilutes the signal (reduces the fraction of significant molecules in the cluster). To filter out these piggy-backers, we implement the following workflow:

Once an internal node in the hierarchical structure is found that surpasses the user defined enrichment threshold (in this example, the threshold will be set as 0.5), we initially label it as a peak (**Supplementary Figure 7, left**). This node will be referred to as the “Parent”. It consists of an internal node descendant to the left (“Left Child”) and to the right (“Right Child”). This will result in three possible situations:

1. Both the Left Child and the Right Child also surpass the enrichment threshold. In this case, both children are enriched for signal, meaning neither is diluting the signal and therefore the Parent is a true peak and not a piggy-backer (**Supplementary Figure 7, Case 1**).
2. Only one Child meets the enrichment threshold. Without loss of generality, let the Left Child have a higher enrichment than the Right Child. In this case:
 - a. If the number of significant children in the Left Child is not enough for the Parent cluster to surpass the threshold (i.e. $\frac{\# \text{ significant children in Left Child cluster}}{\# \text{ nodes in Parent Cluster}} < 0.5$), this means that the significant nodes in the Right Child were contributing to rather than diluting the Parent Cluster’s enrichment, and therefore the Parent is a true peak and not a piggy-backer (**Supplementary Figure 7, Case 2**).
 - b. If the number of significant children in the Left Child *is* enough for the Parent cluster to surpass the threshold (i.e., $\frac{\# \text{ significant children in Left Child cluster}}{\# \text{ nodes in Parent Cluster}} > 0.5$), this means that the Parent cluster could be enriched only due to the Left Child without the contribution of the Right Child. The Parent is then deemed a piggy-backer and the Left Child is labeled as a peak. The Left Child will then undergo the same piggy-backer identification process as the Parent, and the tree will continue to be traversed down the Right Child (**Supplementary Figure 7, Case 3**).

References

1. Benedetti, E. *et al.* A strategy to incorporate prior knowledge into correlation network cutoff selection. *Nat. Commun.* **11**, 5153 (2020).
2. Kanehisa, M. & Goto, S. KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Res.* **28**, 27–30 (2000).
3. Brunk, E. *et al.* Recon3D enables a three-dimensional view of gene variation in human metabolism. *Nat. Biotechnol.* **36**, 272–281 (2018).
4. Langfelder, P. & Horvath, S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics* **9**, 559 (2008).