Bioinformatics, YYYY, 0–0
doi: 10.1093/bioinformatics/xxxxx
Advance Access Publication Date: DD Month YYYY
Manuscript Category

Subject Section

MoNET: an R package for multi-omic network analysis

Jin Li[1,#], Feng Chen[1,#], Hong Liang[1,*], Jingwen Yan[2,*]

[1]College of Intelligent Systems Science and Engineering, Harbin Engineering University, Harbin, China; [2]Department of BioHealth Informatics, School of Informatics and Computing, Indiana University Purdue University Indianapolis, Indianapolis, USA

#These authors contributed to the work equally and should be regarded as co-first authors.

*To whom correspondence should be addressed.

Associate Editor: XXXXXXX

Received on XXXXX; revised on XXXXX; accepted on XXXXX

Abstract

Motivation: The increasing availability of multi-omic data has enabled the discovery of disease biomarkers in different scales. Understanding the functional interaction between multi-omic biomarkers is becoming increasingly important due to its great potential for providing insights of the underlying molecular mechanism.
Results: Leveraging multiple biological network databases, we integrated the relationship between SNPs, genes/proteins and metabolites, and developed an R package MoNET for multi-omic network analysis. This new tool enables users to not only track down the interaction of SNPs/genes with metabolome level, but also trace back for the potential risk variants/regulators given altered genes/metabolites. MoNET is expected to advance our understanding of the multi-omic findings by unveiling their trans-omic interactions and is likely to generate new hypotheses for further validation.
Availability: The MoNET package is freely available on https://github.com/JW-Yan/MONET
Contact: jingyan@iupui.edu, lh@hrbeu.edu.cn
Supplementary information: Supplementary data are available at *Bioinformatics* online.

1 Introduction

With recent volume escalation in -omics data (Saykin, et al., 2015; Shen, et al., 2012), multi-omics is becoming increasingly essential in disease research with a potential to reveal functional decoupling among multiple types of -omics molecules (Acharjee, et al., 2016; Hasin, et al., 2017; Huang, et al., 2017; Yan, et al., 2018). While substantial effort has been dedicated to identify disease biomarkers from different -omics layers (i.e., genetic variations, genes, proteins and metabolites related to disease) (Chase Huizar, et al., 2020; Hurtado, et al., 2018; Kiebish and Narain, 2019), the next key step is to functionally validate these multi-omic findings to further understand how they interact with each other or with other -omics layers (Edwards, et al., 2013; Visscher, et al., 2012). While existing network analysis tools are mostly on protein interaction network, a new tool with multi-omic network relationships will provide extra insights into the functional interactions of multi-omic findings.

Recon2 (Noronha, et al., 2019) is one of the most comprehensive reconstruction of the human metabolic network. It provides directed connections between proteins and metabolites, indicating the catalytic or regulatory roles of proteins in metabolism. While protein expression is certainly the downstream event of gene expression, expression level of genes is partly the regulatory output of the interactions between transcription factors and target binding sites. Genetic variation (i.e., single nucleotide polymorphism (SNP)) within the target binding sites could potentially disrupt or create a new transcription factor binding site (TFBS) (Kumar, et al., 2017) and therefore affect the expression level of downstream genes. Taken together, integrating these relationships could open new opportunities for exploring the potential cascade effect of disease risk variations on the downstream genes, proteins and metabolites.

In this work, we built a multi-omic network by integrating metabolic-protein interactions, functional protein interactions and SNP-Gene relationships and developed a multi-omic network analysis tool, Multi-omic

This is the author's manuscript of the article published in final edited form as:
Li, J., Chen, F., Liang, H., & Yan, J. (2022). MoNET: An R package for multi-omic network analysis. Bioinformatics, 38(4), 1165–1167. https://doi.org/10.1093/bioinformatics/btab722

Network Explorer Tool (MoNET), for discovery of trans-omic paths and subnetwork related to SNPs, genes and/or metabolites of interest. Results generated from this tool is expected to further our understanding of the disease markers by revealing their functional interaction with other -omics layers, which can serve as new hypotheses to be tested.

2 Implementation

The integrated network in MoNET is built on top of three unique data resources, RECON2 (Noronha, et al., 2019), STRING (Szklarczyk, et al., 2019) and SNP2TFBS (Kumar, et al., 2017). While RECON2 is the most comprehensive metabolic network reconstruction with protein-metabolite interactions (i.e., gene-metabolite interaction), STRING is a database with known and predicted protein associations and SNP2TFBS is a database of regulatory SNPs affecting predicted transcription factor binding site (i.e., SNP-gene interactions). We integrated these different types of interactions from three databases using the gene symbols and entrez IDs, which serves as the foundation of the downstream network analysis in MoNET. Details of the interaction data preparation can be found in Text S1. The integrated network provides connections among SNPs, genes/proteins and metabolites, characterizing both genetic regulation and metabolic reactions (Fig. S1, Table S1). MoNET is designed with two main functional modules enabling trans-omic interaction query of disease markers, and analysis of trans-omic paths.

2.1 Trans-omic interaction query of disease markers

With the goal to search for the trans-omic interactions of disease biomarkers, MoNET is specifically designed to find the functional paths/subnetworks from query input (e.g., SNPs/genes of interest) to other -omics layers or between multiple types of query input. This new tool can not only help track down the potential interaction of SNPs, genes/proteins with metabolome level, but also enables user to trace back for the potential risk variants/regulators given altered metabolites. Users can provide one or more types of query input and/or specify one target output. Identification of paths containing different types of -omic features is performed via an unweighted breadth-first searching as implemented in the R package igraph (Csardi and Nepusz, 2005). MoNET supports metabolite query by HMDB ID, KEGG ID, CHEBI ID or PubChem ID, gene query by Entrez ID, Ensemble Gene ID, Gene Name, Enzyme IDs, RefSeq IDs, and SNP query by rsID. The identified paths are stored in a comma-delimited text file for further analysis, in which each line represents a path and contains a list of -omics features.

2.2 Analysis of trans-omic paths

Subsequent analysis of trans-omic paths includes subnetwork extraction, node annotation, pathway enrichment, and visualization. A subnetwork including all output paths can be visualized and exported as a tab-delimited text file, a compatible input format for other network analysis tools like Cytoscape (Shannon, et al., 2003). The current version allows users to download the image of the subnetwork in PDF format with the force-directed network layout. For all genes involved in the subnetwork, further pathway enrichment analysis can be performed using Enrichr (Kuleshov, et al., 2016). In addition, MoNET also provides the annotation of all the SNPs, genes and metabolites in the subnetwork, e.g., mapping between 5 types of gene IDs, and between 7 types of metabolite related IDs. Separate annotation files for each type of -omics feature in the subnetwork will be generated. A detailed description of annotation is given in Text S2.

3 Application

Alzheimer's disease (AD) is a neurodegenerative disease characterized by progressive deterioration of memory and other cognitive functions followed by loss of ability to perform daily activities. It has been affecting an increasing number of aging populations and has now become a public health crisis due to no validated disease modifying treatment. Multi-omic data has been increasingly collected and analyzed in AD for biomarker discovery. Here, we used MoNET to explore the functional interactions of 2,774 AD risk SNPs with downstream genes and metabolites (Lambert, et al., 2013). With path length restricted to [3,4], MoNET returned 9,416 paths cutting across -omics layers, including 173 SNPs, 159 genes and 702 metabolites.

More specifically, there are 15 trans-omic paths along which AD risk SNPs are functionally associated with downstream metabolites through intermediate genes and proteins. With the subnetwork containing all the returned paths, top 10% hubs are located within 75 trans-omic paths (Supplementary material; Figure S4). Taken together, these paths include seven genes (*APOE*, *APOC2*, *APOC1*, *SLC24A4 CD2AP*, *INPP5D*, *BIN1*), the corresponding proteins and metabolites like nicotinamide adenine dinucleotide (NAD+) (Table S4). Of note, these seven genes are all among the top 20 AD risk genes and *APOE* is known as the top one risk gene for AD. The metabolite with the highest degree in the subnetwork is NAD+. NAD+ levels decrease with aging and experimentally boosting NAD+ could reduce the production of beta amyloid and improve cognitive function in Alzheimer's disease. In several recent studies, dietary supplements of NAD+ was suggested as a potential treatment for AD (Fang, et al., 2019; Hou, et al., 2018; Liu, et al., 2013).

In summary, the SNP-gene-metabolite paths formed a subnetwork with top AD risk genes and AD-related metabolites. This suggests that network analysis results in MoNET can not only help functionally annotate those SNPs but also to some extent enable the prioritization of risk genes. These SNP-gene-metabolite paths provide new hypotheses for the molecular mechanism underlying AD and suggest potential molecular targets that warrant further validation. Details of this application can be found in supplementary material Text S3. Detailed network analysis of trans-omic subnetwork identified from AD risk SNPs is available in Text S4.

4 Conclusion

MoNET is a network analysis tool built on top of integrated multi-omic network. It is designed to identify the interactions between SNPs, genes and/or metabolites of interest and other -omics layers. MoNET is expected to benefit the emerging multi-omics analysis by advancing our understanding of existing -omic findings. The trans-omic paths and subnetwork returned from MoNET will provide functional annotation for disease markers from trans-omic perspective and are likely to generate new hypothesis for further validation.

Funding

This work was partly supported by NIH/NIA R21 AG066135, and by NSF 1942394 and 1755836. The work was also supported by the National Natural Science Foundation of China (61773134, 61803117 and 61901063); the Humanities and Social Science Fund of Ministry of Education of China (19YJCZH120);

the Natural Science Foundation of Heilongjiang Province of China (YQ2019F003) and the Science and Technology Plan Project of Changzhou (CE20205042).

Conflict of Interest: none declared.

References

Acharjee, A.*, et al.* (2016) Integration of multi-omics data for prediction of phenotypic traits using random forest. *BMC Bioinformatics*, **17**(Suppl. 5), 180.

Chase Huizar, C., Raphael, I. and Forsthuber, T.G. (2020) Genomic, proteomic, and systems biology approaches in biomarker discovery for multiple sclerosis. *Cell Immunol*, **358**, 104219.

Csardi, G. and Nepusz, T. (2005) The Igraph Software Package for Complex Network Research. *InterJournal Complex Systems*, **1695**.

Edwards, S.L.*, et al.* (2013) Beyond GWASs: illuminating the dark road from association to function. *Am J Hum Genet*, **93**(5), 779-97.

Fang, E.F.*, et al.* (2019) Mitophagy inhibits amyloid-beta and tau pathology and reverses cognitive deficits in models of Alzheimer's disease. *Nat Neurosci*, **22**(3), 401-412.

Hasin, Y., Seldin, M. and Lusis, A. (2017) Multi-omics approaches to disease. *Genome Biol*, **18**(1), 83.

Hou, Y.*, et al.* (2018) NAD(+) supplementation normalizes key Alzheimer's features and DNA damage responses in a new AD mouse model with introduced DNA repair deficiency. *Proc Natl Acad Sci U S A*, **115**(8), E1876-E1885.

Huang, S., Chaudhary, K. and Garmire, L.X. (2017) More Is Better: Recent Progress in Multi-Omics Data Integration Methods. *Front Genet*, **8**, 84.

Hurtado, M.O., Kohler, I. and de Lange, E.C. (2018) Next-generation biomarker discovery in Alzheimer's disease using metabolomics - from animal to human studies. *Bioanalysis*, **10**(18), 1525-1546.

Kiebish, M.A. and Narain, N.R. (2019) Enabling biomarker discovery in Parkinson's disease using multiomics: challenges, promise and the future. *Per Med*, **16**(1), 5-7.

Kuleshov, M.V.*, et al.* (2016) Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. *Nucleic Acids Res*, **44**(W1), W90-7.

Kumar, S., Ambrosini, G. and Bucher, P. (2017) SNP2TFBS - a database of regulatory SNPs affecting predicted transcription factor binding site affinity. *Nucleic Acids Res*, **45**(D1), D139-D144.

Lambert, J.C.*, et al.* (2013) Meta-analysis of 74,046 individuals identifies 11 new susceptibility loci for Alzheimer's disease. *Nat Genet*, **45**(12), 1452-1458.

Liu, D.*, et al.* (2013) Nicotinamide forestalls pathology and cognitive decline in Alzheimer mice: evidence for improved neuronal bioenergetics and autophagy procession. *Neurobiol Aging*, **34**(6), 1564-80.

Noronha, A.*, et al.* (2019) The Virtual Metabolic Human database: integrating human and gut microbiome metabolism with nutrition and disease. *Nucleic Acids Res*, **47**(D1), D614-D624.

Saykin, A.J.*, et al.* (2015) Genetic studies of quantitative MCI and AD phenotypes in ADNI: Progress, opportunities, and plans. *Alzheimers Dement*, **11**(7), 792-814.

Shannon, P.*, et al.* (2003) Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res*, **13**(11), 2498-504.

Shen, E.H., Overly, C.C. and Jones, A.R. (2012) The Allen Human Brain Atlas: comprehensive gene expression mapping of the human brain. *Trends Neurosci*, **35**(12), 711-4.

Szklarczyk, D.*, et al.* (2019) STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Res*, **47**(D1), D607-D613.

Visscher, P.M.*, et al.* (2012) Five years of GWAS discovery. *Am J Hum Genet*, **90**(1), 7-24.

Yan, J.*, et al.* (2018) Network approaches to systems biology analysis of complex disease: integrative methods for multi-omics data. *Brief Bioinform*, **19**(6), 1370-1381.

Supplementary Material

Text S1: Integration of multi-omic network

We built a multi-omic network characterizing the relationships across multiple types of omics layers by combining metabolic network (Noronha, et al., 2019), protein interaction network (Szklarczyk, et al., 2019) and relationships between transcription factor binding sites (TFBS) and SNPs (Kumar, et al., 2017).

The human metabolic reactions were obtained from RECON2 metabolism resource (Noronha, et al., 2019) (https://www.vmh.life/#human/all), which details all known metabolic reactions occurring in at least one cell type in the human body. The original metabolic interactions in RECON2 including different relationships: between metabolite and reaction, between enzyme and reaction, between complex and reaction respectively. We extracted the enzyme-reaction and metabolite-reaction relationships from raw data and combined them to generate a metabolic network involving enzyme-coding genes, metabolites and reactions. For each reaction, there are one input metabolite, one output metabolite and one or more enzyme-coding genes involved. Following previous studies, we removed 78 current metabolites from the final metabolic network. These metabolites are involved in more than 100 reactions, like water, proton, oxygen and ATP, and are of less interest to our analysis (Vitkup, et al., 2006).

The human protein interaction network were downloaded from a quality-controlled database STRING (Szklarczyk, et al., 2019) (https://string-db.org/). The protein-protein associations consist of direct (physical) as well as indirect (functional) interactions, deriving from high throughput experimental data, from the mining of databases and literature, and from predictions based on genomic context analysis. As the interactions in STRING are redundant, we kept only one record of each interaction with a maximum score. The original interaction scores, expressing an approximate confidence of the association being true given all the available evidence, were normalized to the interval (0,1).

The relationships between genes and SNPs were created based on the data in SNP2TFBS (Kumar, et al., 2017) (http://ccg.vital-it.ch/snp2tfbs/). SNP2TFBS is a database of regulatory SNPs affecting predicted transcription factor binding site. We followed the original paper and kept only SNPs that significantly change the TF binding affinity ($p<3e-6$). Since transcription factors bind to the upstream regulatory elements of genes and stimulate or inhibit gene expression, we made a connection between a SNP and a gene if that SNP significantly affects its upstream transcription factor binding site.

All the functional interactions from the SNP2TFBS (Kumar, et al., 2017), STRING (Szklarczyk, et al., 2019) and RECON2 databases (Noronha, et al., 2019) were integrated into a multi-omic network based on the gene symbol and entrez ID (Fig. S1). In total, the multi-omic network currently consists of 2,521,055 interactions and covers 1,413,101 components including 1,373,322 SNPs, 27,589 genes, 7,289 reactions and 4,901 metabolites (Table S1).

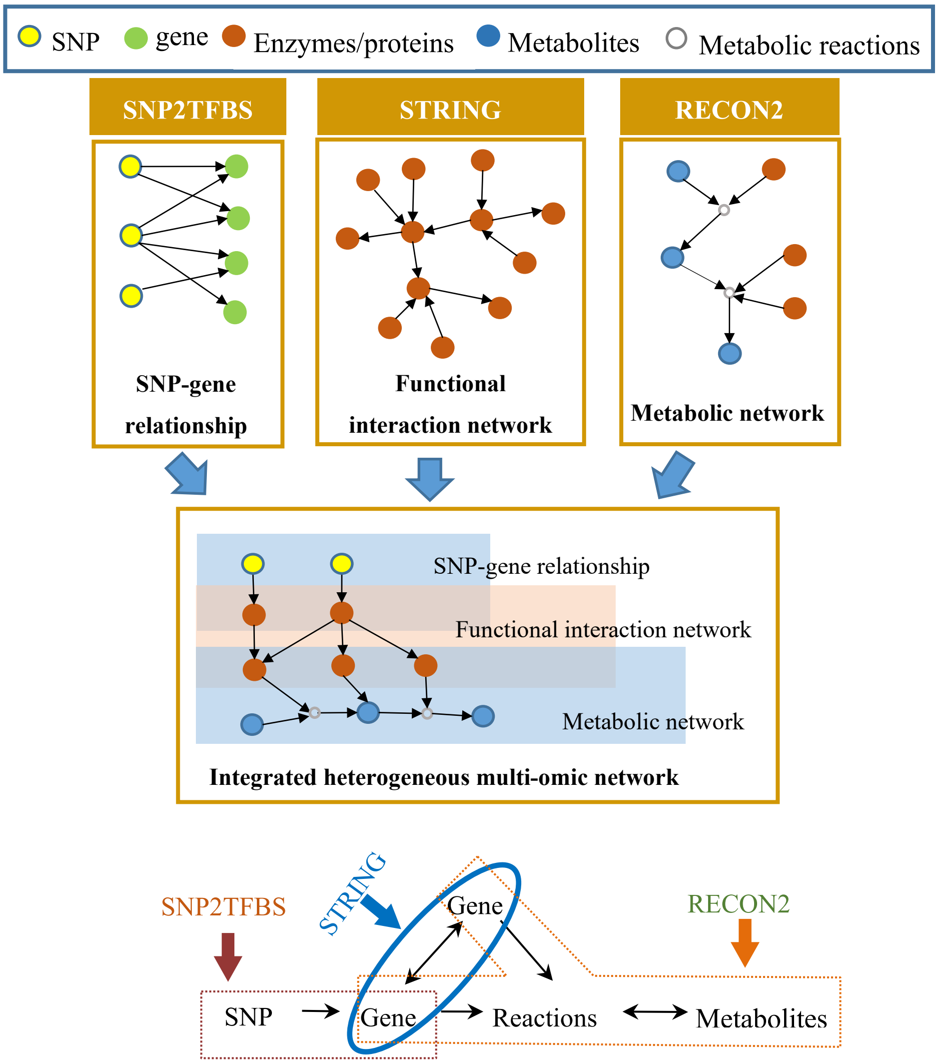

Figure S1. Integration of heterogeneous multi-omic network

Table S1. Summary information of the integrated multi-omic network

	SNPTFBS	STRING	RECON2			Total
Interaction Type	S-G*	G-G*	G-R*	R-M*	M-R*	—
Number of Interactions	2,534,603	1,075,246	10,294	10,611	9,292	2,521,055
TF	170	—	—			—
SNP/TFBS	1,373,322					1,373,322
Gene/Protein	—	15,868	1,730			27,589
Reaction		—	7,289			7,289
Metabolite			4,901			4,901

*S:SNP, G: Gene, R: metabolic reaction, M: metabolite.

Text S2: Annotation of nodes in the integrated multi-omic network

The annotation information of a SNP includes the position of the SNP, which chromosome it belongs to, which gene it is located in or near, its start position and end position, name, stable identifier in NCBI (https://www.ncbi.nlm.nih.gov/), and the relative position of the SNP and the gene (inside, upstream or downstream). Each SNP is assigned to a gene according to the coordinate if it is located within a gene according to the hg19 genome build. If a SNP is not located within a gene, then its nearest gene is listed.

Genes were annotated by HGNC Biomart (https://biomart.genenames.org) (Braschi, et al., 2019). Annotation was conducted in the hg19 assembly. The annotation of a gene includes its stable identifier in NCBI, the official gene symbol, the chromosome the gene located on, the ensemble gene identifier, gene-regulated enzyme entries that indicate the hierarchical functional classes to which they belong and the reference sequence (RefSeq) identifiers for the gene.

Annotation information of both reactions and metabolites were directly obtained from RECON2 (Noronha, et al., 2019). The annotation information of a reaction includes the reaction's name, a confidence score indicating the evidence supporting reactions' inclusion into the human metabolic reconstruction, the reference of the reaction. The annotation information of a metabolite includes its descriptive name, full name, formula, and identifiers from human metabolome database, KEGG database, PubChem database and Chemical Entities of Biological Interest (ChEBI) database.

Text S3: Application of MoNET

We used MoNET to annotate the function of AD risk SNPs on other -omics layers from a large-scale meta-analysis of Alzheimer's disease (AD) (Igap and Schellenberg, 2012). More specifically, 2,774 SNPs with meta-analysis P-value less than 10^{-5} were extracted from the International Genomics of Alzheimer's Project (IGAP) (Lambert, et al., 2013) as candidates for multi-omic network analysis. Among these, 257 SNPs were located within the TF binding sites, and thus were used as the query list in MoNET.

MoNET traversed the whole multi-omic network to search for the functional interactions of those SNPs with the downstream genes and metabolites. To identify a reasonable default path length for search, we plotted the distribution of all the paths between all molecules (Fig. S2). As shown in Figure S2, the number of paths with length between the interval [5,9] is larger than 10^8. Searching those paths would result in a great burden to the computation. To save the computation time, we limited our search for the paths with length less than 5. In other words, we only focused on the SNP-gene-reaction-metabolite paths with less than 3 intermediate genes. Using function SNPquery(), MoNET identified 9,416 trans-omic paths. Example paths are shown in Table S2. A subnetwork composed of all identified paths is generated by function SubNetRFW() (Fig. S3). Overall, the downstream influence on metabolome level consists of 173 SNPs, 159 genes, 732 reactions and 702 metabolites. Annotation files of these molecules were also generated by function SubNetRFW(). For genes involved in the identified paths, SubNetRFW() also performed pathway enrichment analysis in the 2019 KEGG human pathway database. Top enriched pathways are listed in Table S3.

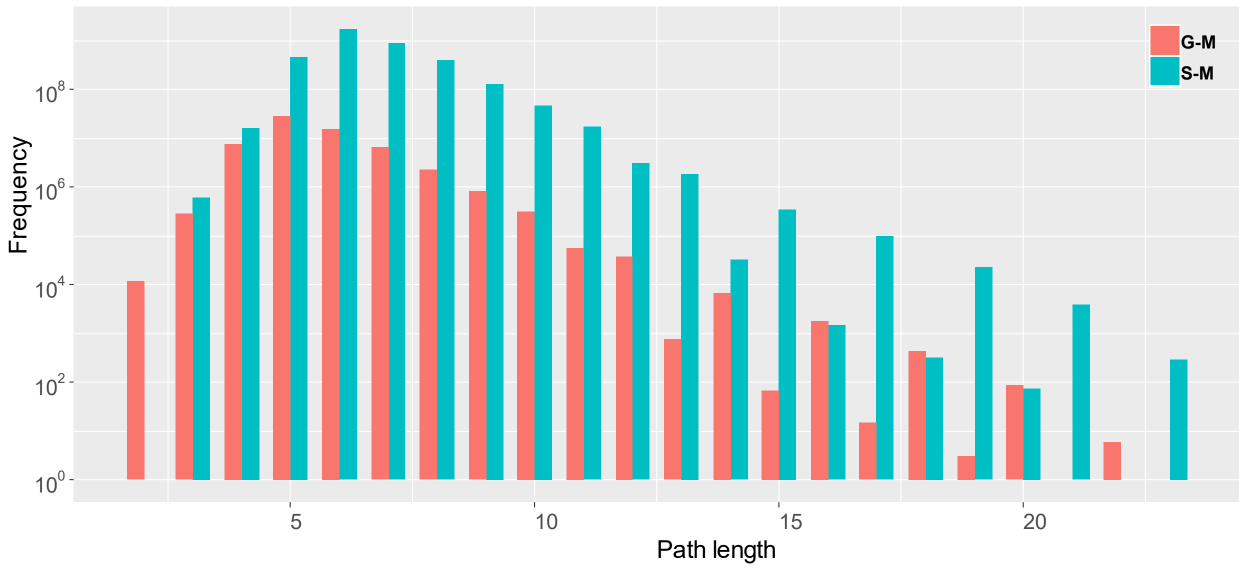

Figure S2. The length distribution of pairwise shortest paths from SNPs to metabolites (Green) and from genes to metabolites (Red).

Table S2. Example paths including *APOE* or *APOC1*

rs769446, *APOE*, *LRP2*, R32, 25hvitd3[c]
rs769446, *APOE*, *PLA2G7*, R2906, ac[e]
rs769450, *APOE*, *PI4K2A*, R3022, pail4p_hs[c]
rs484195, *APOC1*, *PLA2G7*, R2906, ac[e]
rs5117, *APOC1*, *PLA2G7*, R2906, ak2lgchol_hs[e]
rs78959900, *APOC1*, *LPL*, R2439, Rtotal[c]

Figure S3. The subnetwork with all identified paths from query SNPs to metabolites**.**

Table S3. Top enrichment results of 159 genes involved in identified paths

Term	Adjusted.P.value	Odds.Ratio	Genes
Carbohydrate digestion and absorption	1.16E-26	107.57	*MGAM; ATP1A4; PIK3CD; ATP1A3; ATP1B4; ATP1A2; PIK3R2; ATP1B3; PIK3CB; PIK3R1; ATP1A1; ATP1B2; SLC2A5; ATP1B1; HK2; HK1; HK3; PIK3CA; HKDC1*
Cholesterol metabolism	5.32E-20	65.18	*ABCA1; ABCG8; ABCG5; LRP1; APOA1; LCAT; LPL; APOC3; LRP2; CYP7A1; LIPC; APOC2; APOC1; CD36; APOE; APOB*
Phosphatidylinositol signaling system	2.04E-19	33.52	*INPPL1; PTEN; PIK3CD; OCRL; PIK3R2; PIK3R1; SYNJ2; PIK3CB; PIK3C2A; INPP5B; SYNJ1; PIK3CA; INPP5D; INPP5E; INPP5J; PI4KA; PLCG1; PI4KB; PI4K2A*
Inositol phosphate metabolism	7.60E-19	41.55	*INPPL1; PTEN; PIK3CD; OCRL; PIK3CB; SYNJ2; PIK3C2A; INPP5B; PIK3CA; SYNJ1; INPP5D; INPP5E; INPP5J; PI4KA; PLCG1; PI4KB; PI4K2A*
Aldosterone-regulated sodium reabsorption	4.31E-17	73.52	*ATP1A4; PIK3CD; ATP1A3; ATP1B4; ATP1A2; PIK3R2; ATP1B3; PIK3CB; PIK3R1; ATP1A1; ATP1B2; ATP1B1; PIK3CA*
Type II diabetes mellitus	4.98E-14	47.56	*HK3; PKM; PKLR; PIK3CA; HKDC1; PIK3CD; PIK3R2; PIK3CB; PIK3R1; HK2; GCK;HK1*
Thyroid hormone signaling pathway	7.49E-13	20.36	*ATP1A4; ATP1B4; PIK3CD; ATP1A3; ATP1B3; PIK3R2; ATP1A2; ATP1B2; PIK3R1; PIK3CB; ATP1A1; ATP1B1; SLC9A1; PIK3CA; PLCG1*
Proximal tubule bicarbonate reclamation	2.17E-12	84.97	*ATP1A4; ATP1A3; ATP1B4; ATP1A2; ATP1B3; SLC25A10; ATP1A1; ATP1B2; ATP1B1*
cAMP signaling pathway	2.32E-12	12.93	*ATP1A4; ATP1B4; PIK3CD; ATP1A3; ATP1B3; PIK3R2; ATP1A2; ATP1B2; PIK3R1; PIK3CB; ATP1A1; ATP1B1; PLD1; PLD2; SLC9A1; PIK3CA; ACOX1; PDE4B*

Text S4: Trans-omic interaction of AD risk SNPs with genes and metbolites

As mentioned in section Text S3, MoNET identified 9,416 trans-omic paths involving 159 genes, 732 reactions and 702 metabolites. Considering that the influence of AD risk SNPs on downstream molecules is most efficient along the shortest paths, and trans-omic paths covering these hub molecules are more vulnerable, we further narrowed down to the shortest paths and those that cover hub molecules.

In total, there are 15 SNP-Gene-Reaction-Metabolite shortest paths. Figure S4a is the subnetwork consisting of these paths. As shown in Figure S4a, the subnetwork includes potassium (K+), calcium (Ca2+), four genes (*APOE*, *APOC2*, *APOC1*, *SLC24A4*) and their assembly protein (Table S4). Genes *APOE*, *APOC2*, *APOC1*, *SLC24A4* are top risk genes with significant association with AD (Karch and Goate, 2015; Lambert, et al., 2013; Stage, et al., 2016; Van Cauwenberghe, et al., 2016). The K+ level was found significantly decreased across all intracellular compartments in the AD brain (Roberts, et al., 2016). The impaired neuronal Ca2+ handling rendered aging brain vulnerable to AD (Mattson and Arumugam, 2018).

The molecules with high degree (hub molecules) usually have important impact on upstream and/or downstream pathways. Changes in hub molecules is likely to cause the interruption of the pathway, so that paths with hub molecules are more vulnerable. In the subnetwork with all returned paths, molecules with top 10% degree were extracted, including 133 genes and 26 metabolites and 39 reactions. Among all the returned paths, 75 of them have these hub molecules in the entire path. Figure S4b shows a subnetwork composed of nodes from these 75 trans-omic paths, including three AD risk genes (*CD2AP*, *INPP5D*, *BIN1*) and hydrogen phosphate (Table S4). The expression of upstream regulator of hydrogen phosphate, *INPP5D*, has been reported to increase with the progression of AD in the 5xFAD mouse model, predominantly in plaque-associated microglia (Tsai, et al., 2021). Another noteworthy molecule is nicotinamide adenine dinucleotide (NAD+), which has the highest degree of metabolites in all the identified trans-omic paths. It has been reported that NAD+ augmentation inhibited cognitive decline in AD animal models and was recently suggested as potential treatment for AD (Fang, et al., 2019; Hou, et al., 2018; Liu, et al., 2013).

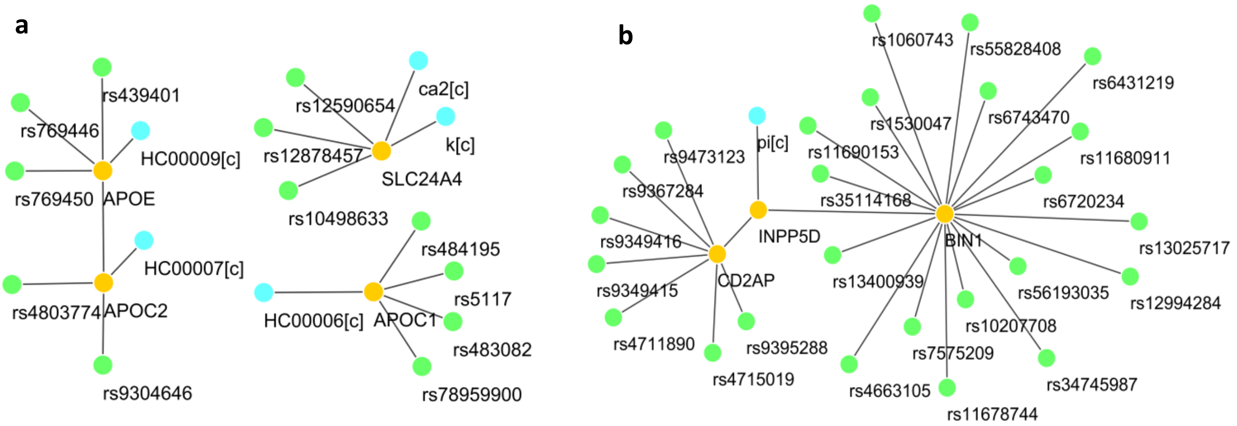

Figure S4. Subnetwork consists of identified trans-omic paths with metabolic reactions omitted. **a**, subnetwork consisting of 15 shortest trans-omic paths. **b**, subnetwork consisting of trans-omic paths that covering hub molecules.

Table S4. A summary of genes and metabolites in the subnetwork consisting of shortest trans-omic paths and paths that cover hub molecules.

	Attribute	ID	Names	Degree	Rank in top 10%
S_1	Gene	348	*APOE*	60	Yes
		344	*APOC2*	31	Yes
		341	*APOC1*	29	Yes
		123041	*SLC24A4*	10	Yes
	Metabolite	HC00009[c]	ApoE assembly	1	No
		HC00007[c]	ApoC2 assembly	1	No
		HC00006[c]	ApoC1 assembly	1	No
		ca2[c]	calcium (Ca2+)	2	No
		k[c]	potassium (K+)	3	No
S_2	Gene	274	*BIN1*	78	Yes

		23607	*CD2AP*	25	Yes
		3635	*INPP5D*	33	Yes
	Metabolite	pi[c]	hydrogen phosphate	39	Yes
MHD	Metabolite	nadh[x]	nicotinamide adenine dinucleotide (NAD+)	76	Yes

S_1, subnetwork consists of nodes from shortest trans-omic paths; S_2, subnetwork consists of nodes from trans-omic paths that covering hub molecules; MHD, metabolite with highest degree.

References

Braschi, B., *et al.* Genenames.org: the HGNC and VGNC resources in 2019. *Nucleic acids research* 2019;47(D1):D786-D792.

Fang, E.F., *et al.* Mitophagy inhibits amyloid-beta and tau pathology and reverses cognitive deficits in models of Alzheimer's disease. *Nat Neurosci* 2019;22(3):401-412.

Hou, Y., *et al.* NAD(+) supplementation normalizes key Alzheimer's features and DNA damage responses in a new AD mouse model with introduced DNA repair deficiency. *Proc Natl Acad Sci U S A* 2018;115(8):E1876-E1885.

Igap and Schellenberg, G.D. International Genomics of Alzheimer's Disease Project (IGAP) genome-wide association study. *Alzheimer's & Dementia* 2012;8(4, Supplement):P101.

Karch, C.M. and Goate, A.M. Alzheimer's disease risk genes and mechanisms of disease pathogenesis. *Biol Psychiatry* 2015;77(1):43-51.

Kumar, S., Ambrosini, G. and Bucher, P. SNP2TFBS - a database of regulatory SNPs affecting predicted transcription factor binding site affinity. *Nucleic acids research* 2017;45(D1):D139-D144.

Lambert, J.C., *et al.* Meta-analysis of 74,046 individuals identifies 11 new susceptibility loci for Alzheimer's disease. *Nat Genet* 2013;45(12):1452-1458.

Liu, D., *et al.* Nicotinamide forestalls pathology and cognitive decline in Alzheimer mice: evidence for improved neuronal bioenergetics and autophagy procession. *Neurobiol Aging* 2013;34(6):1564-1580.

Mattson, M.P. and Arumugam, T.V. Hallmarks of Brain Aging: Adaptive and Pathological Modification by Metabolic States. *Cell Metab* 2018;27(6):1176-1199.

Noronha, A., *et al.* The Virtual Metabolic Human database: integrating human and gut microbiome metabolism with nutrition and disease. *Nucleic acids research* 2019;47(D1):D614-D624.

Roberts, B.R., *et al.* Rubidium and potassium levels are altered in Alzheimer's disease brain and blood but not in cerebrospinal fluid. *Acta Neuropathol Commun* 2016;4(1):119.

Stage, E., *et al.* The effect of the top 20 Alzheimer disease risk genes on gray-matter density and FDG PET brain metabolism. *Alzheimers Dement (Amst)* 2016;5:53-66.

Szklarczyk, D., *et al.* STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic acids research* 2019;47(D1):D607-D613.

Tsai, A.P., *et al.* INPP5D expression is associated with risk for Alzheimer's disease and induced by plaque-associated microglia. *Neurobiology of disease* 2021;153:105303.

Van Cauwenberghe, C., Van Broeckhoven, C. and Sleegers, K. The genetic landscape of Alzheimer disease: clinical implications and perspectives. *Genet Med* 2016;18(5):421-430.

Vitkup, D., Kharchenko, P. and Wagner, A. Influence of metabolic network structure and function on enzyme evolution. *Genome Biol* 2006;7(5):R39.