MULTI-OMIC INTEGRATION VIA SIMILARITY NETWORK FUSION TO DETECT MOLECULAR SUBTYPES OF AGEING

Paper Summary

Zhengwei Song

COLUMBIA
Columbia University
Department of Neurology

Outline

Five-modal Data:

- RNA Sequencing
- DNA methylation
- Histone acetylation
- Proteomics
- Metabolomics

Similarity network fusion (SNF) →

Internal Validity: APN, ADM - less, better

External Validity: Association with neuropathology

Top contributor: NMI

Three-modal Data:

- Histone acetylation
- DNA methylation
- RNA Sequencing

Subtype membership →

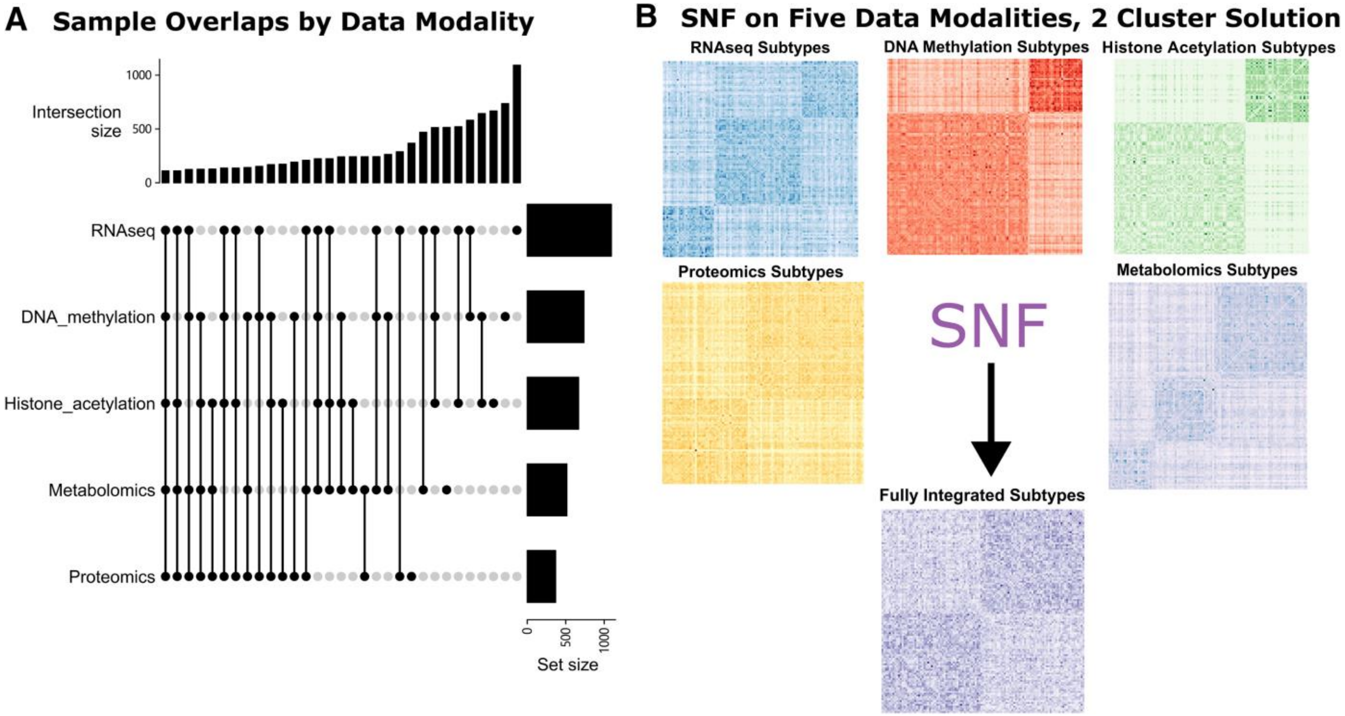
Association with

- Longitudinal cognitive decline
- Molecular features

SNF by R *SNFtool* package

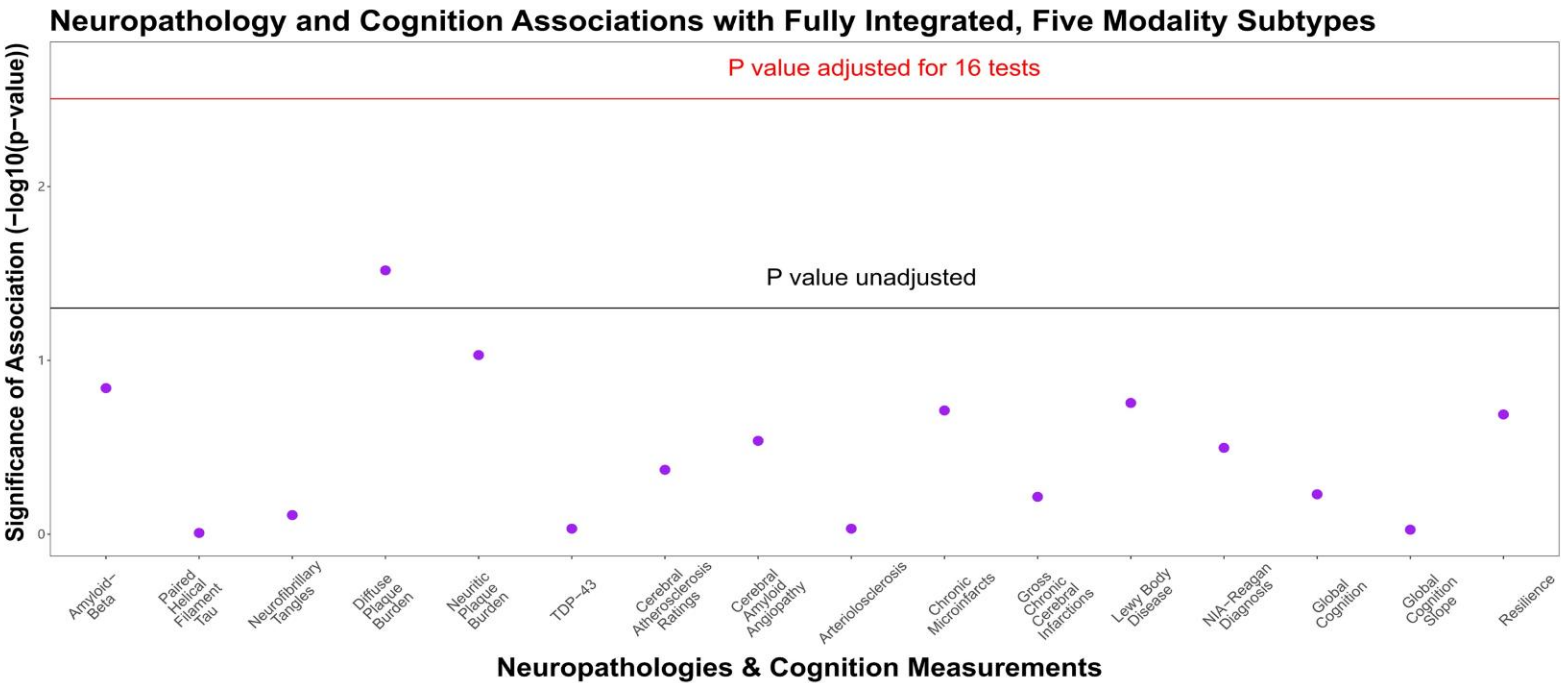
- Data : 111 (all five) overlapping samples across five data modalities
- Parameters:
  - K = 40, α = 0.5, T = 50
  - K: # of neighbors used to construct the similarity matrices
  - α: hyper-parameter used in the scaling of edge weights
  - T: # of iterations
- Method:
  - Spectral clustering by rotation cost and eigen-gap method
- Result

SNF by R *SNFtool* package

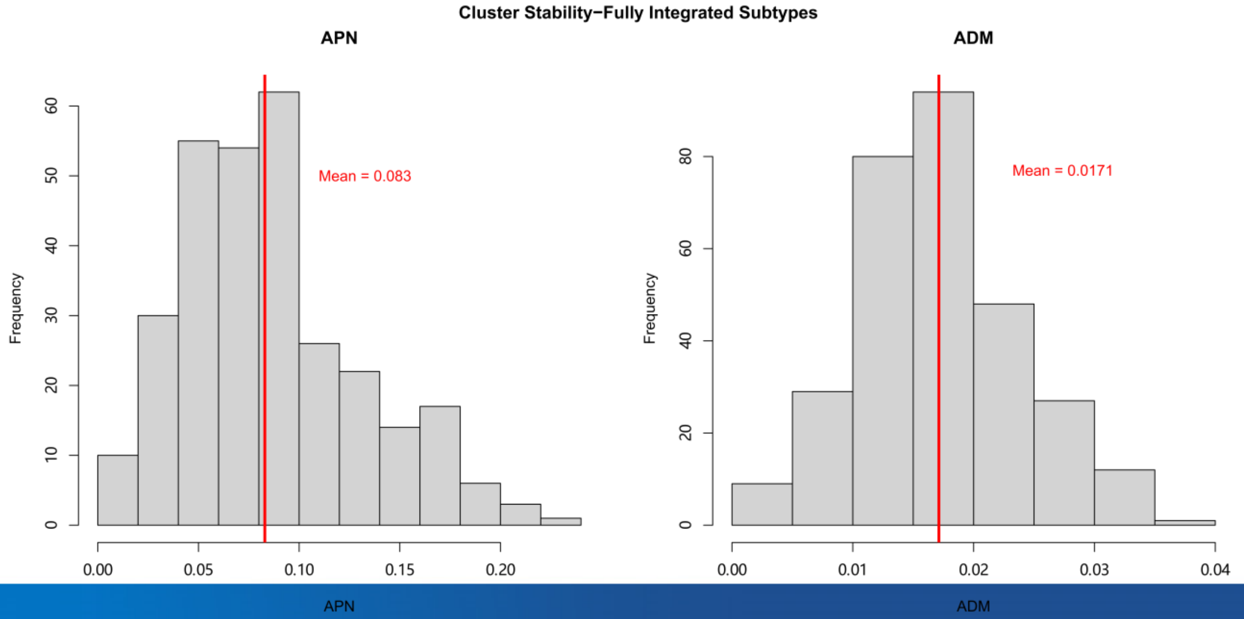


- Result
    - Optimal 2 clusters by both cost and eigen-gap method

Validity by R clValid package



- Result
  - External Validity: $p_{raw} = 0.03$, but $p_{bf} > 0.05$
  - However, internal stable (APN = 8.7%, ADM = 0.02)



Top Contributors

- Top modality contributors to the fused network

NMI	Fused Network	RNASeq	DNA-Methylation	Histone Acetylation	Proteomic	Metabolomic
Fused Network	1	**0.15**	**0.18**	**0.38**	0.04	0.05

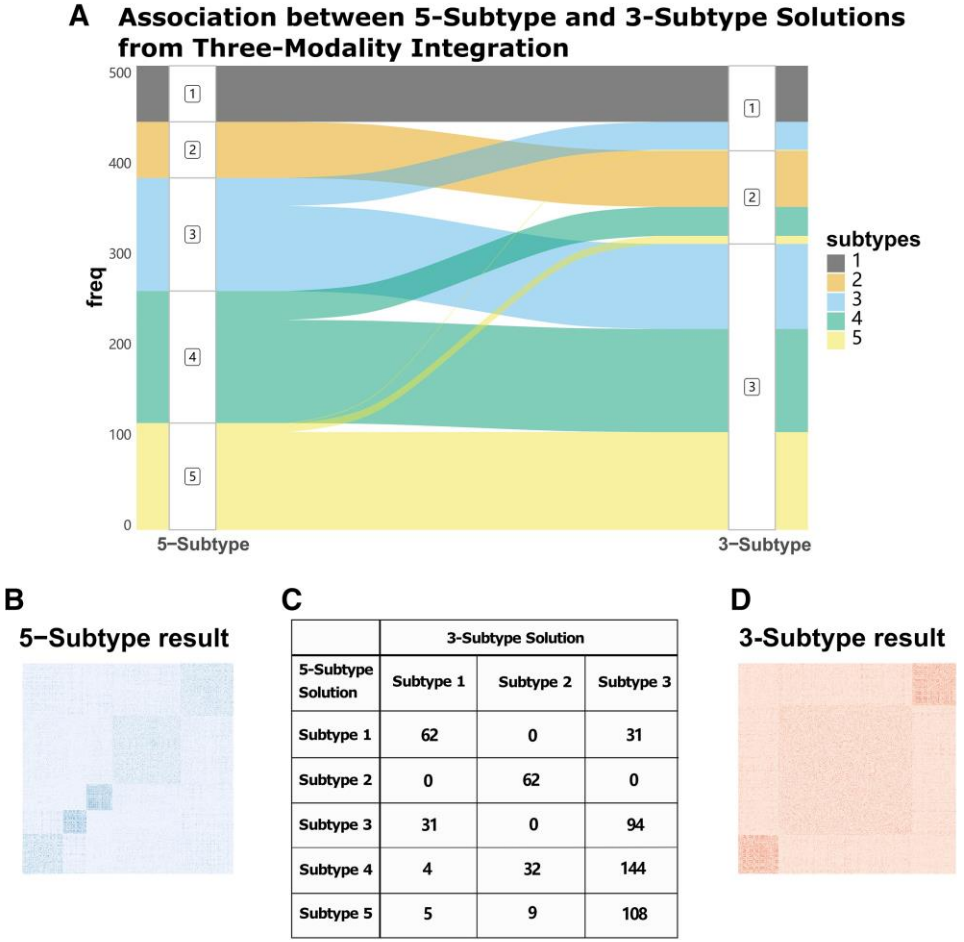
- 513 non-missing overlapping samples across three modalities

SNF: Secondary by R *SNFtool* package

- Data : 513 overlapping samples
- Result
    - ○ 3 clusters by Eigen-gap (B)
    - ○ 5 clusters by rotation cost (D)
    - ○ Strong overlap b/w two solutions (A, C)

      ($\chi 2$: $p = 2.2*10_{-16}$)

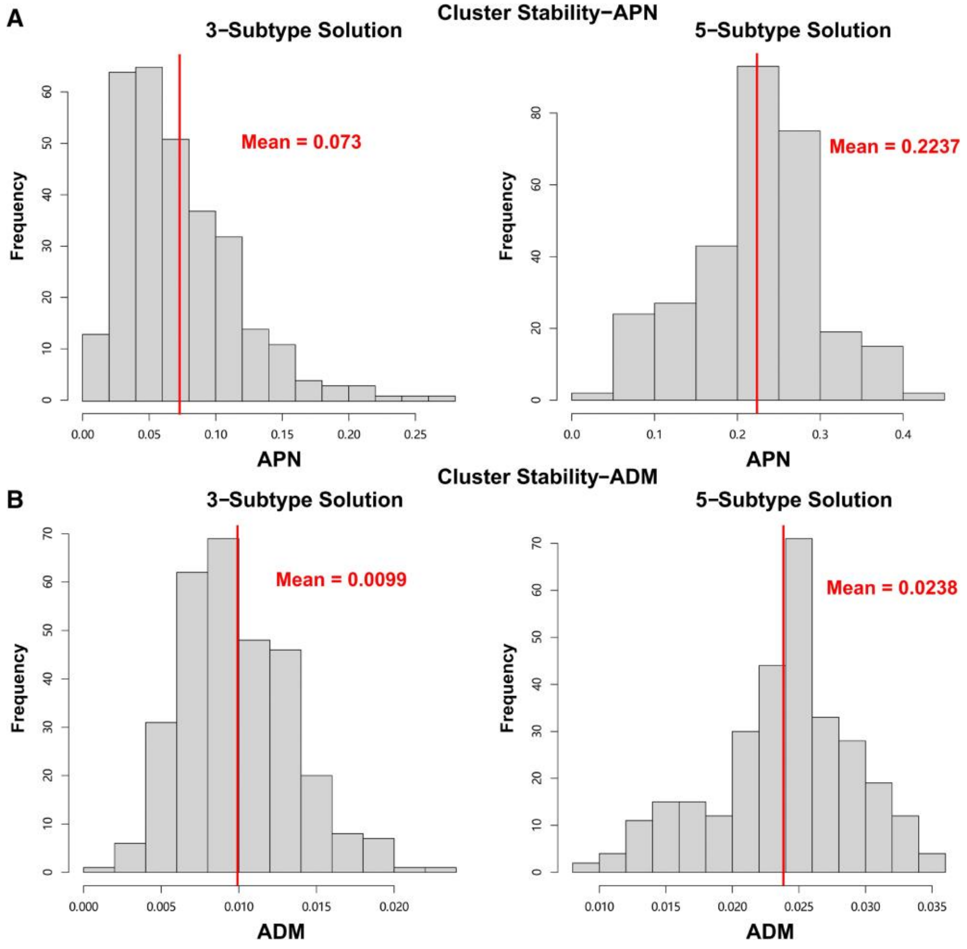
  i.e. subtype 4, 5 are largely represented by subtype 3



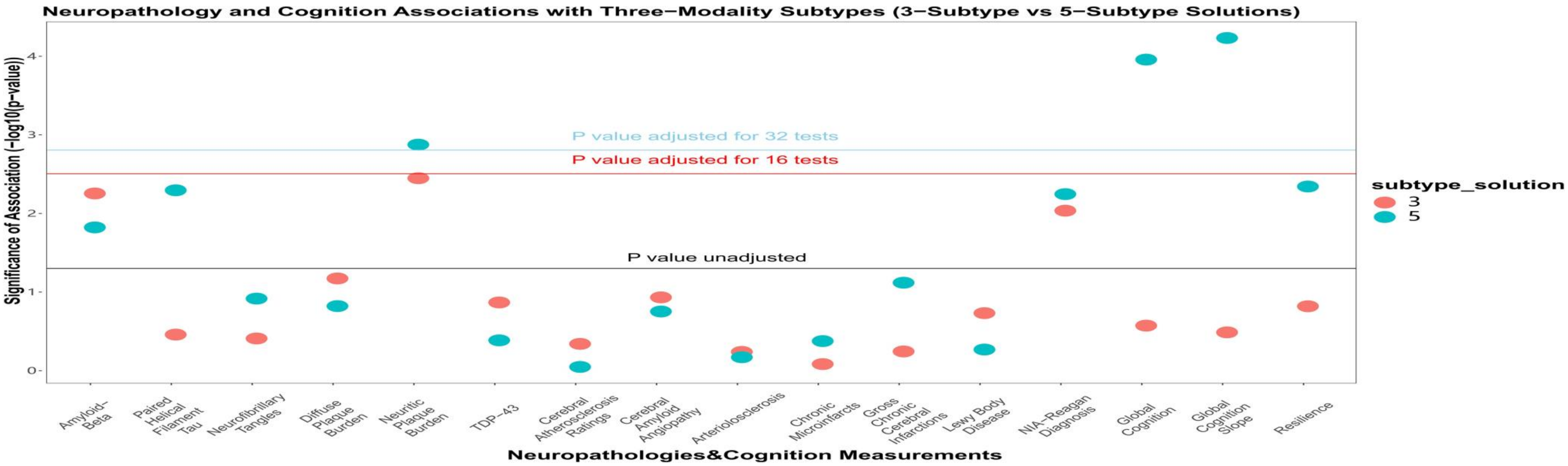
	3-Subtype Solution		
5-Subtype Solution	Subtype 1	Subtype 2	Subtype 3
Subtype 1	62	0	31
Subtype 2	0	62	0
Subtype 3	31	0	94
Subtype 4	4	32	144
Subtype 5	5	9	108

SNF: Secondary by R *SNFtool* package

- Compare two solutions by Internal Stability (APN, ADM)
  - Both measures were better for 3-subtype solution
  - Though those for 5-subtype were also well



External Validity

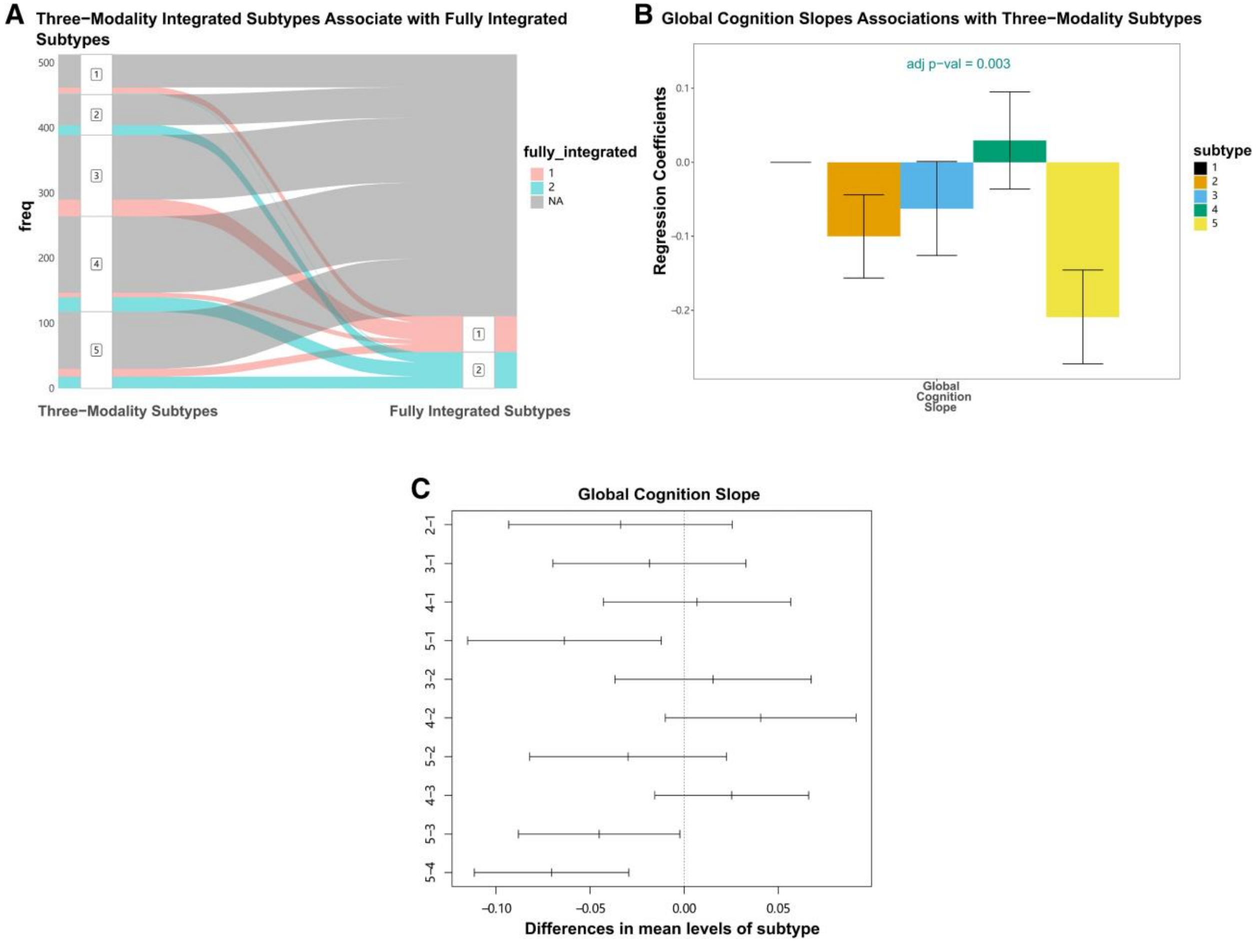


- 5-subtype was preferred by external validity, while 3-subtype was preferred by internal v
    - Subtype membership significantly associated with global cognition, rate of cognitive decline under 5-subtype while not observed under 3-subtype
- Therefore, 5-subtype solution would be further probed

i.e. three-modal data divided by 5 subtypes

Association with longitudinal cognitive decline

- SNF Consistence across sample size
  - Substantial overlap (chi-sq $p=8.1*10^{-9}$) (A)
- Result:
  - Subtype 5 had the worst global cognitive performance at last visit and the fastest rate of cognitive decline, while subtype 4 had the best and slowest ones (B)
  - The diff is significant by pairwise tests against all other subtypes(C)

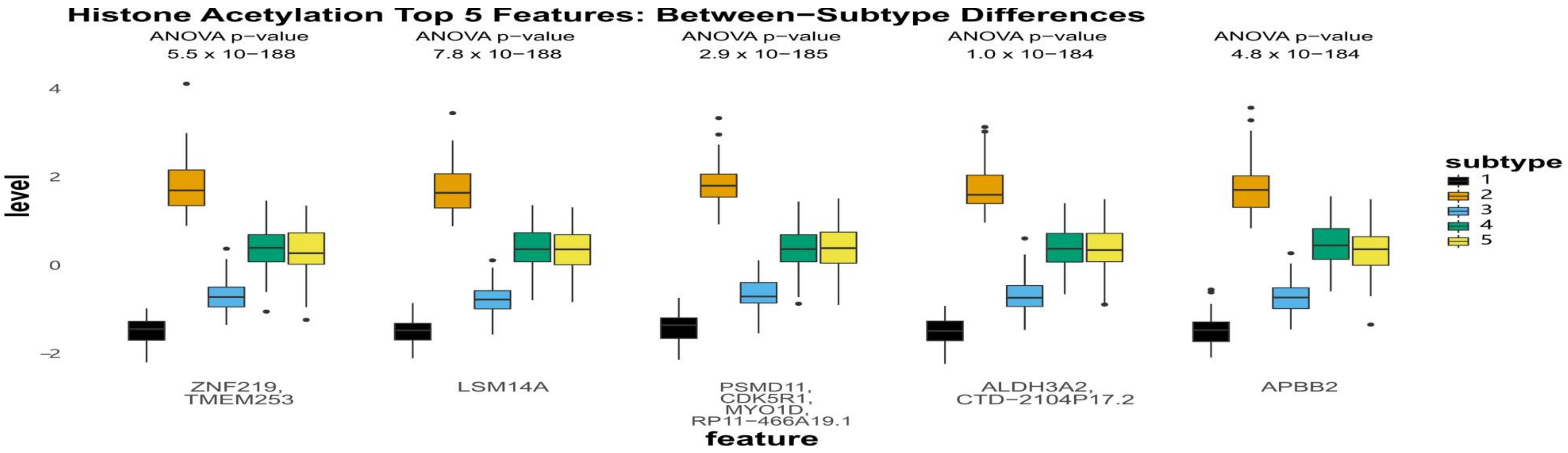


Association with Molecular features

- Method:
  - Top features from each data by One-way ANOVA

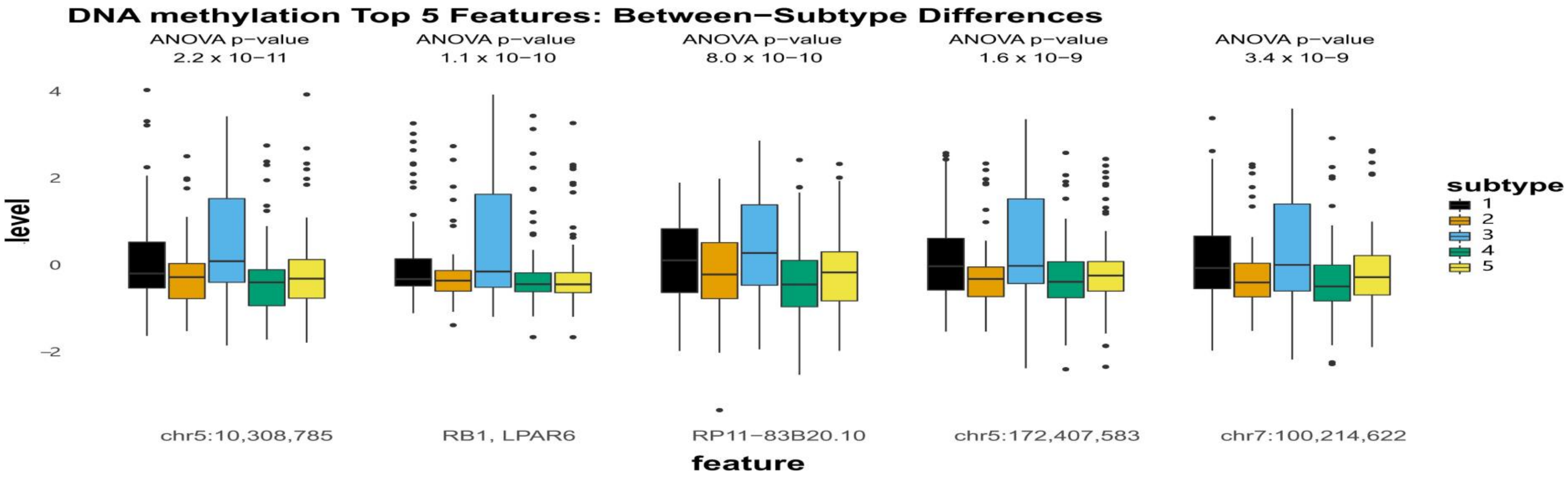
Data Modalities	Feature Name	Genomic Region/Gene
RNASeq	ENSG00000145882	PCYOX1L
	ENSG00000254561 …	NECTIN1 …
DNA-methylation	cg26878318	chr5:10,308,785
	cg21512370 …	RB1, LPAR6 …
H3K9 Histone Acetylation	peak7234	ZNF219, TMEM253
	peak12821 …	LSM14A …

Association with Molecular features



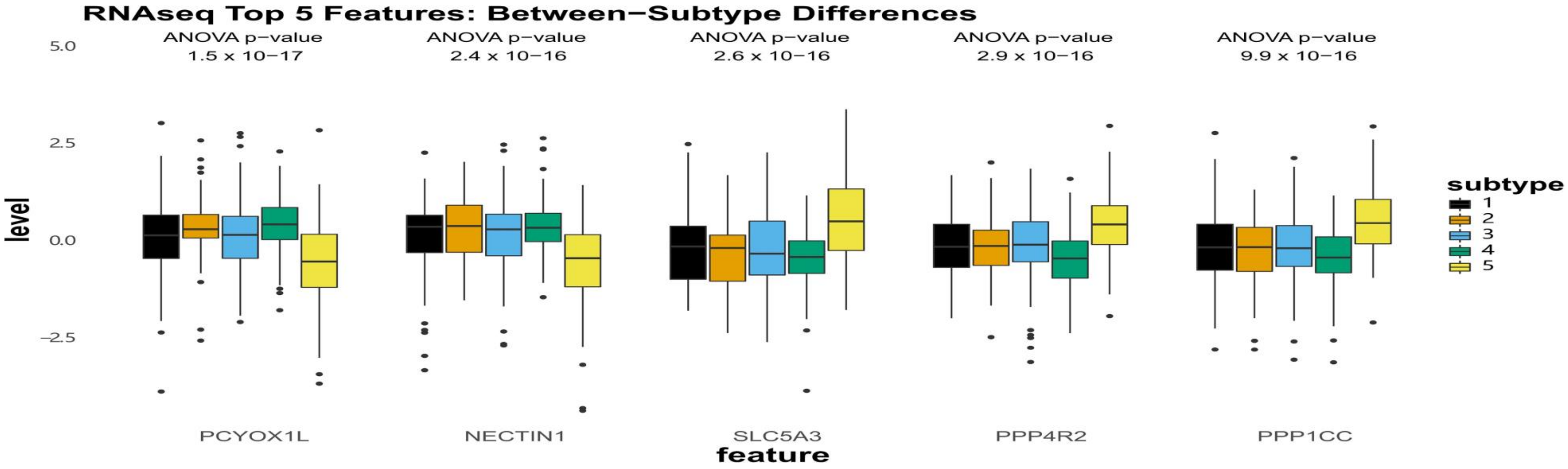
- Acetylation
  - Subtype 1 had the lowest level while subtype 2 had the highest level
  - Both subtype 4 & 5 had intermediate acetylation

Association with Molecular features



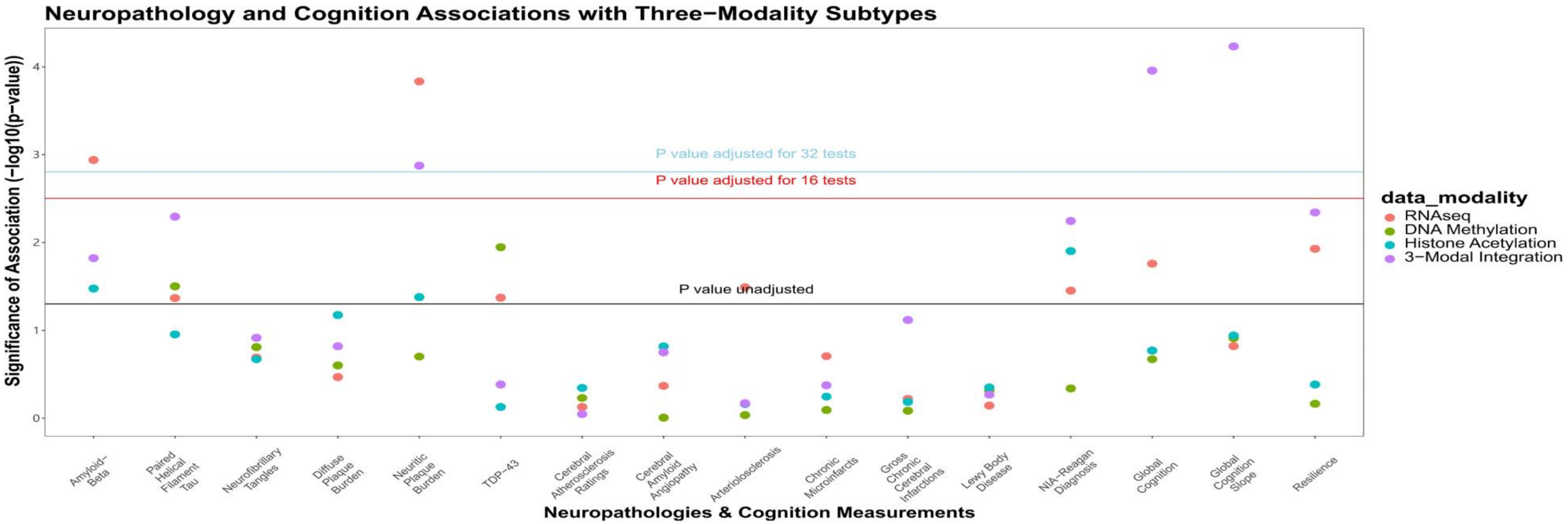
- DNA methylation
  - CpG sites showed differential methylation at sites annotated to RB1, LPAR6, and RP11-83B20.10, as well as intergenic regions on chromosome 5 and 7
  - Though no consistent pattern related to the cognition-associated subtype 5 was observed

Association with Molecular features



- RNA sequencing
  - The top subtype-associated RNA sequencing features revealed lower levels of PCYOX1L and NECTIN1, as well as higher levels of SLC5A3, PPP4R2, and PPP1CC in subtype 5 specifically compared to all other subtypes

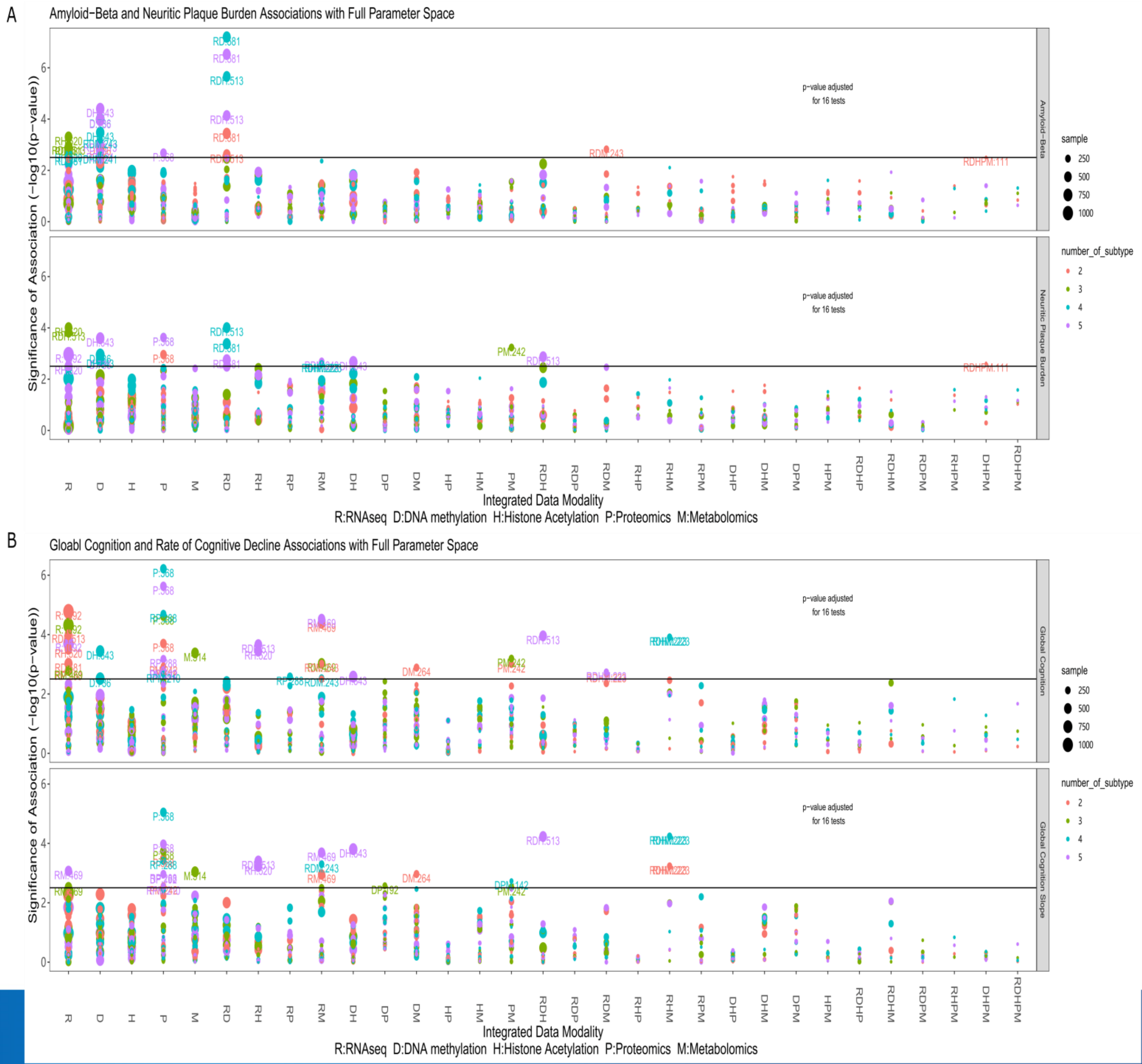
Comparison with single modality subtypes


Neuropathology and Cognition Associations with Three−Modality Subtypes

- Integrated subtypes had unique associations with cognitive performance and decline
  - None of the unimodal subtypes showed more significant associations than three-modal, five-cluster subtypes on global cognitive performance

Sensitivity analysis

- Substantial variability in external validity was observed across different selections of sample size, data modalities and cluster number
  - e.g. Amyloid, Cognitive outcomes



Conclusion & Discussion

- Association of multi-omic subtypes with cognition
  - subtypes were significantly associated with individuals' rates of cognitive decline and levels of beta-amyloid neuropathology
- Sensitivity analysis of integrated subtypes
  - combining all five modalities revealed two subtypes that did not show significant external validity in terms of neuropathology and cognition
- Association of RNAseq subtypes with neuropathology
  - significant associations were found between RNAseq subtypes and neurofibrillary tangles and amyloid-beta
- Importance of histone acetylation
  - histone acetylation (H3K9ac) provided the most information to the fused subtypes
- Identification of molecular features
  - the genes that differentiated the cognition-associated subtype 5 from other subtypes, including PCYOX1L and NECTIN1

Columbia | Neurology

Unused slides / Appendix

Data processing

- Uniform multi-omic feature post-processing
  - For each omic modality (top 20 components from PCA), test associations in pair between

    Age of death, Sex, PMI, Study cohort
- Residual cognition
  - Last available global cognitive measurement ~ 13 pathological indicators
- Missing data for SNF
  - Random Forest

Internal Validity by R *clValid* package

- Methods :
    - Resampling: 80% participants, 300 random draws
    - CV
- Indicators
    - APN, ADM
    - χ^2 statistic: Compare independence b/w subtyping solutions

Top molecular features

- Methods :
  - One-way ANOVA b/w subtypes for each omic modality
- Indicators
  - P-value from F-test

Association of subtypes with neuropathology, cognition and residual cognition

- Data:
    - Subtype: dummy variables
    - Latency: time diff (in years) b/w the last study visit and age of death
- Models:
    - A: **Neuropathologies** ~ subtype + age of death + sex + education + PMI + study + APOEε4
    - B: **Cognitive measurements** ~ subtype + **latency** + age of death + sex + education + PMI + study + APOEε4
- Indicators
    - Bonferroni adjusted p-value from omnibus F-tests

External Validity

- Sensitivity analyses across data modalities, sample sizes and cluster numbers
- Parameters:
  - d: 31 data modalities (combining any of five datasets)
  - n (sample sizes): 111 (all five) to 1092 (RNA seq) participants from 31 data modalities
  - c: 2 to 5 clusters
  - j: cognitive outcome
  - m: # of data modalities being fused
- Models:
  - A: $-\log(p_j) \sim m$
  - B: $-\log(p_j) \sim n$
  - C: $-\log(p_j) \sim c$

SNF

- Unimodal: 2 cluster solution from SNF from 111 overlapping samples
- Subtypes:
  - RNA seq: 3
  - DNA methylation: 2
  - histone acetylation: 2
  - proteomics: 2
  - metabolomics: 3
- Models: