

Integrating Multivariate Biomarkers And High-dimensional Multi-omics Data To Identify Integrative Aging Subtypes



COLUMBIA UNIVERSITY
MEDICAL CENTER

Boyi Hu¹, Yuanjia Wang², Annie J. Lee¹

¹Department of Neurology, Columbia University, USA

²Department of Biostatistics, Columbia University, USA

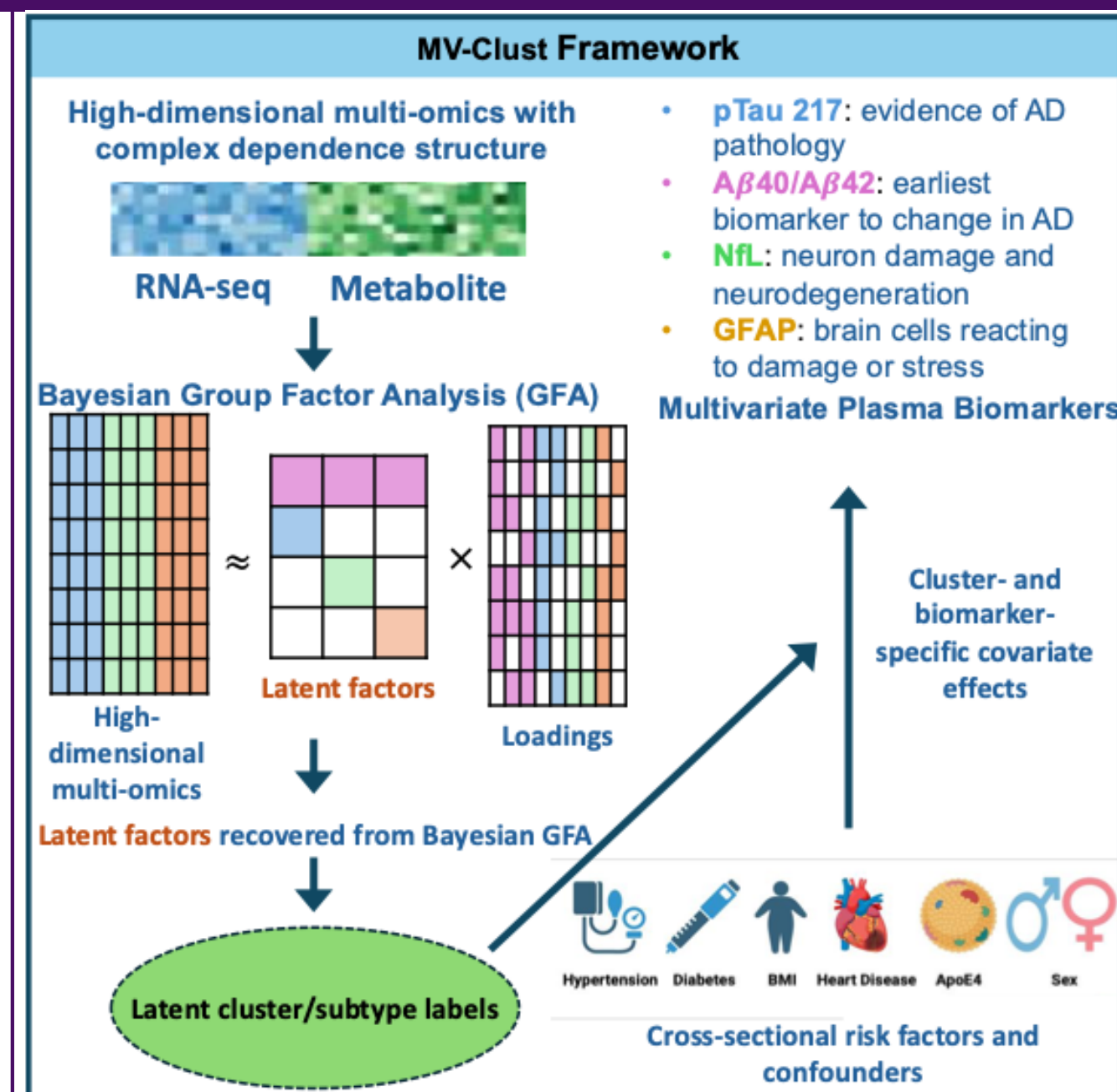
Key Takeaways

- We developed MV-Clust that integrates high-dimensional multi-omics with multivariate plasma biomarkers to identify biologically and clinically meaningful subgroups.
- In the EFIGA study, MV-Clust identified four subgroups with distinct molecular profiles, clinical characteristics, risk factor associations, and dementia prevalence.

INTRODUCTION

- Existing disease subtyping approaches cannot integrate high-dimensional multi-omics data with multivariate clinical biomarkers.
- Aging is characterized by substantial molecular and clinical heterogeneity that cannot be fully captured by a single omics layer or a single biomarker.
- Multiple omics layers provide a more comprehensive representation of molecular profiles.
- Multivariate blood plasma biomarkers provide complementary information on the stage and severity of aging and aging-related conditions.
- Integrating multi-omics data with plasma biomarkers may enable discovery of more interpretable, robust, and clinically meaningful aging subtypes.
- To address this need, we developed **MV-Clust**, a novel integrative subtyping method.
- Applied to data from the **Estudio Familiar de Influencia Genética en Alzheimer (EFIGA)** study, MV-Clust identified four novel aging subtypes.

METHODS



- We use Bayesian group factor analysis to model complex dependence among multi-omics and extract low dimensional latent factor representation denoted as $\hat{\eta}_i$
- y_{ij} : the j-th plasma biomarker
- x_{ip} : covariates such as age, sex, vascular risk factors
- f_k : density function of Y when the subject belongs to subtype k , i.e.
- $$y_{ij} \sim N(\mu_{jk} + \sum_{p=1}^P x_{ip} \beta_{pjk} + \alpha_{ik} \xi_k, \sigma_k^2)$$
- A joint latent class model for Y that consists of k subgroups
- $$f(y_{ij}; x_i, v_i, g_i) = \sum_{k=1}^K \pi_{ik}(\hat{\eta}_i) f_k(y_{ij}; x_i, v_i)$$
- $\pi_{ik}(\hat{\eta}_i)$: probability of subject i belonging to subtype k

Data description:

- Participants:** 616 Hispanic individuals from the EFIGA study.
- Blood plasma biomarkers:** pTau 217, Aβ40/Aβ42, NfL and GFAP.
- Covariates:** : age, sex, education, APOE ε4, diabetes, heart disease, hypertension and BMI.
- Multi-omics:** 3,290 RNA-seq genes, and 2,747 metabolites.

Model fitting:

- We applied **MV-Clust** to the EFIGA data and used **AIC** for model selection.
- AIC selected a model with **four latent subtypes**.

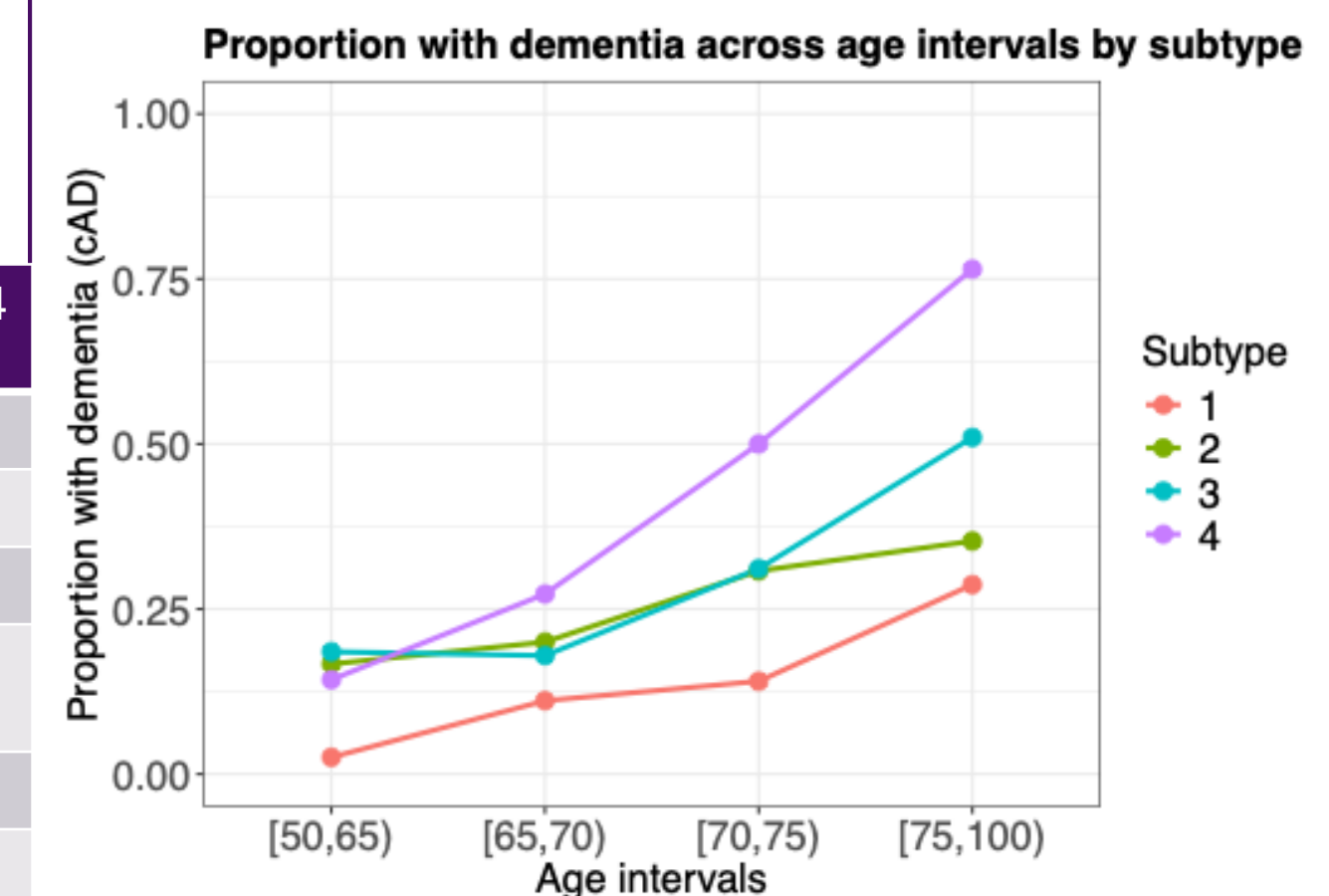
Variable	Subtype 1 (n = 380)	Subtype 2 (n = 39)	Subtype 3 (n = 134)	Subtype 4 (n = 63)
Age at draw	69.52	71.92	71.55	74.32
Hypertension	70.53%	64.10%	76.87%	79.37%
Diabetes	24.47%	23.08%	29.85%	41.27%
Heart disease	5.53%	5.13%	7.46%	6.35%
BMI	27.59	27.40	27.14	25.95
APOE ε4	33.68%	41.03%	38.06%	36.51%
Male	35.00%	35.90%	26.87%	42.86%
Education > 8 years	18.16%	7.69%	15.67%	14.29%
Clinical AD	13.00%	24.30%	32.30%	47.50%
Stroke	0.79%	0.00%	2.24%	3.17%

RESULTS



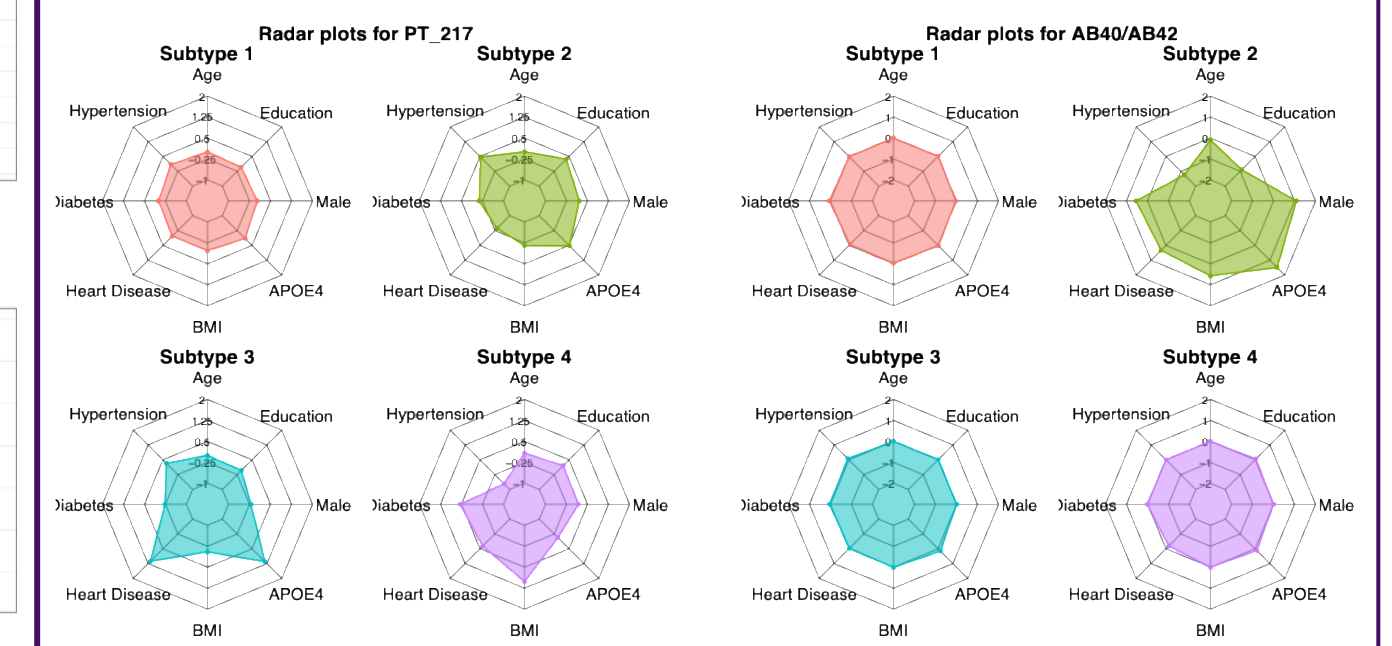
- Subtype 1:** Low biomarker burden
- Subtype 2:** Early amyloid-related change
- Subtype 3:** Brain support responding to stress or damage
- Subtype 4:** Advanced and aggressive disease progression

Dementia proportion across age intervals by identified subtype:



- Subtype 1** showed consistently low dementia prevalence across age intervals, while **Subtype 4** showed the strongest increase with age

Subtype-specific associations between risk factors and biomarkers

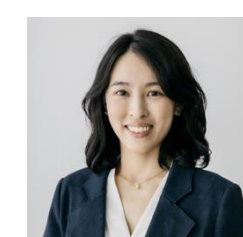


- Subtype 1:** APOE ε4 and hypertension are associated with higher pTau217 ($p < 0.05$), while diabetes is associated with higher Aβ40/Aβ42 ($p < 0.05$).
- Subtype 2:** APOE ε4 and hypertension are associated with higher pTau 217 ($p < 0.05$).
- Subtype 3:** Heart disease and APOE ε4 are associated with higher pTau 217 ($p < 0.05$).
- Subtype 4:** Higher BMI is associated with lower pTau217 ($p < 0.05$).

ACKNOWLEDGEMENT

COLUMBIA UNIVERSITY
MEDICAL CENTER

- Annie J. Lee
- Yuanjia Wang



Dr. Annie J. Lee



Dr. Yuanjia Wang

Contact:
bh2940@cumc.columbia.edu



Grants

- K01AG084849
- NIAP24-1276422
- Carol and Gene Ludwig Pilot Grant Program in Neurodegeneration
- HABS-HD Health Enhancement Scientific Program