Stratifying of Alzheimer's disease leads to the identification of actively protective biologies

C. Navarron Izquierdo, A. Malinowski, M. Sanna, K. Taylor, J. Inchauspe, J. Kozubek, J. Sardell, S. Das, S. Hogg, C. Stubberfield
PrecisionLife Ltd., Oxford, United Kingdom

Abstract #: 8287
Presentation #: PSTR012.23

Introduction

PrecisionLife is a pioneering techbio company, with a unique combinatorial approach that finds more signal in large patient datasets from complex diseases than standard methods. Our high-resolution patient stratification identifies subgroups of patients with similar disease drivers and treatment responses, to make precision medicine possible in complex diseases[1,2].

Alzheimer's disease (AD), like other complex diseases, is characterized by a high degree of heterogeneity across the patient population, reflected in a wide range of disease presentations and therapy responses. Knowledge of the most obvious genetic association has not translated to effective treatments in the clinic.

Applying PL's novel approach to two clinically well-annotated, independent AD datasets, we aim to identify novel disease signatures and introduce our novel actively-protective biology analysis.

Methods

DATASETS:
PrecisionLife analyzed two independent genomic datasets constructed from the UK Biobank (UKBB)[3] and one independent dataset constructed from The Whole Genome Sequence Harmonization Study (WGS Harmonization) including The Religious Orders Study and Memory and Aging Project (ROSMAP)[4], Mount Sinai Brain Bank, and Mayo Clinic studies (ROSMAP/MSBB/Mayo) with the following inclusion/exclusion criteria:
Alzheimer's disease Cases:

- Alzheimer's disease diagnosis (UKBB: ICD-10 code, G30.x; ROSMAP/MSBB/Mayo: MMSE and clinical diagnosis proximate to death)

Healthy Controls:

- No evidence of neurodegenerative disease: UKBB and ROSMAP/MSBB/Mayo using the same criteria described above. Healthy controls from The database of Genotypes and Phenotypes (dbGaP) study number phs001963.v1.p1[5] were included for the ROSMAP/MSBB/Mayo analysis.
- No self-reported cognitive decline
- No family history of Alzheimer's disease

	Causative analysis cohorts			Actively protective analysis cohort
	UKBB 2020	Disjoint UKBB 2023	ROSMAP/MSBB/Mayo	ROSMAP/MSBB/Mayo
Case Count	882	2,287	752	568
Control Count	1,816	5,769	1,839	728

Table 1. Number of cases and controls per cohort included in the analysis shown from Figure 1 to 5.

COMBINATORIAL ANALYSIS:
The datasets were analyzed in the PrecisionLife platform to identify combinations of SNP genotypes that when observed together in a patient are strongly associated with AD.
SNP combinations that have high odds ratios, significant *p*-values, and high prevalence in cases are prioritized. This process undergoes 1,000 cycles of fully randomized permutations and combinations must meet a specified FDR threshold.
SNPs are scored using a Random Forest algorithm in a 5-fold cross-validation framework and prioritized based on their ability to differentiate cases and controls.
The highest-scoring SNPs are then mapped to genes and clustered by the patients they co-occur in to generate a disease architecture.

ACTIVELY PROTECTIVE ANALYSIS:
SNP combinations that have high odds ratios in the ROSMAP/MSBB/Mayo dataset causative run were used to select the controls with high prevalence of disease risk signatures among all the cognitively healthy controls. Combinatorial analysis as described above was then run to identify combinations of SNP genotypes with high prevalence among the "protected" controls.

Results

Figure 1. Combinatorial analysis of genomic data vs GWAS

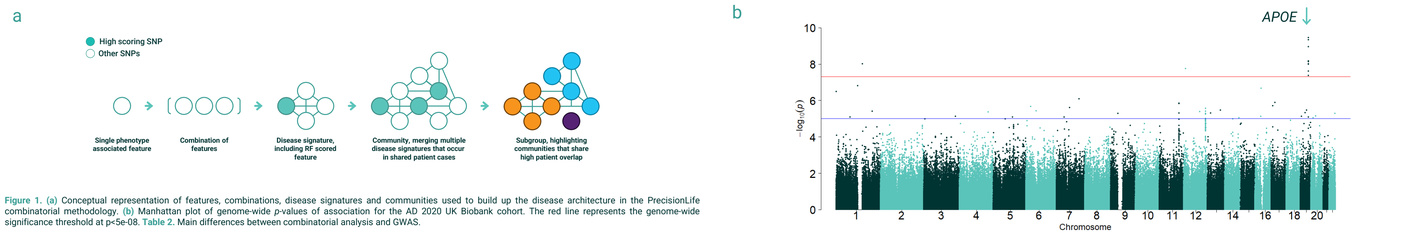

Figure 1. (a) Conceptual representation of features, combinations, disease signatures and communities used to build up the disease architecture in the PrecisionLife combinatorial methodology. (b) Manhattan plot of genome-wide *p*-values of association for the AD 2020 UK Biobank cohort. The red line represents the genome-wide significance threshold at p<5e-08. Table 2. Main differences between combinatorial analysis and GWAS.

Combinatorial Analysis	GWAS
Specific combinations of variants associated with each patient subgroup serve as a genetic stratification biomarker	Single SNP associations must be significant across large groups of patients
Patient subgroups with different causes of disease or even incorrect diagnoses can be distinguished (stratified) by different mechanistic etiology	Limited insights unless disease is likely to be caused by a small number of rare variants with large effect sizes (often in gene coding regions affecting protein 3D structure)
Captures epistatic and non-linear additive effects of all interactions between SNPs, genes, environmental factors and metabolic networks	Does not account for the effects of interactions between SNPs, genes and metabolic networks

Figure 2. AD UKBB 2020 cohort combinatorial analysis reveals 6 patients subgroups associated with distinct biological pathways

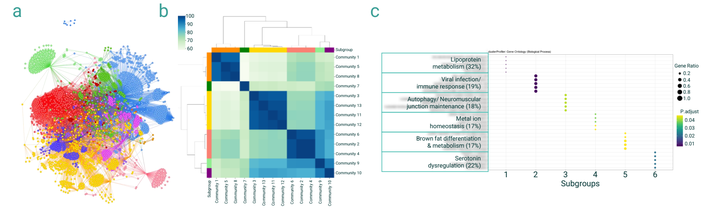

Figure 2. (a) Disease architecture diagram demonstrating the 13 communities of SNPs comprising the structure of the Alzheimer's disease patient subpopulations generated by the PrecisionLife platform. Each circle represents a disease-associated SNP genotype, edges represent co-association in patients, and each color a distinct community of SNPs. (b) Clustered heatmap showing the overlap of AD patients associated with 13 communities. Each border color represents a patient subgroup (n = 6). (c) Pathway enrichment plot for the genes found in the communities associated with different patient subgroups. Gene ratio represents the ratio of genes found in the pathway compared to the genes associated with a community and *p*.adjust represents the *p*-value adjusted for multiple testing. The dots in the plot are color-coded based on their corresponding *p*.adjust values. The percentage of cases represented by each subgroup is displayed in brackets, patients may belong to multiple subgroups.

Figure 3. Disease signatures are replicated in a UKBB 2023 disjoint dataset

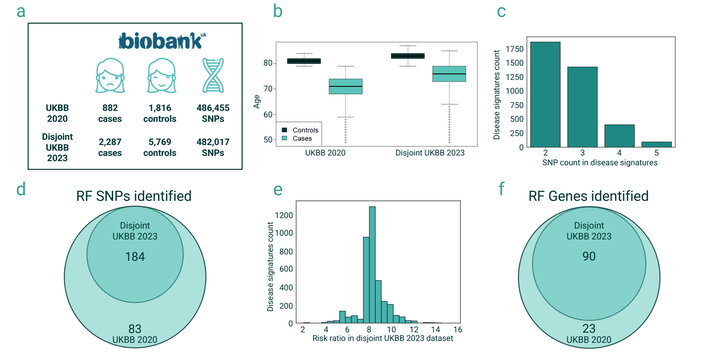

Figure 3. (a) Diagram illustrating the number of independent cases and controls used to generate a new UKBB 2023 disjoint dataset, and the number of SNPs in the datasets following quality controls. (b) Age distribution of first reported Alzheimer's disease in UKBB 2020 and Disjoint UKBB 2023. (c) Number of SNP counts in the disease signatures from UKBB 2020 mapped to the disjoint UKBB 2023 dataset. (d) Number of RF-scored SNPs identified in UKBB 2020 architecture validated in a disjoint UKBB 2023 dataset. (e) Distribution of risk ratio of UKBB 2020 disease signatures validated in UKBB 2023 disjoint dataset. (f) Number of genes mapped to the RF-scored SNPs identified in Figure 3d.

Figure 4. Novel disease risk signatures identified in ROSMAP/MSBB/Mayo cohort

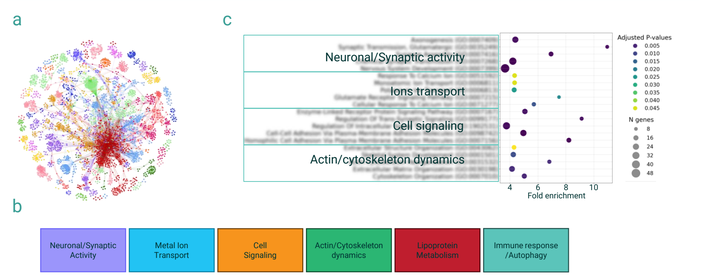

Figure 4. (a) Disease architecture diagram demonstrating the 83 communities of SNPs comprising the structure of the Alzheimer's disease patient subpopulations in the ROSMAP/MSBB/Mayo dataset generated by the PrecisionLife platform. Each circle represents a disease-associated SNP genotype, edges represent co-association in patients, and each color represents a distinct community of SNPs. (b) Colored squares illustrate the different functions representing the genes mapped to RF-scored SNPs from Figure 5a. (c) Pathway enrichment plot for the genes mapped to the RF SNPs found in the architecture shown in Figure 5a. Fold enrichment represents the background frequency of total genes annotated to that term compared to the sample frequency representing the number of genes inputted that fall under the same term. P.adjust represents the *p*-value adjusted for multiple testing. The dots in the plot are color-coded based on their corresponding *p*.adjust values.

Figure 5. Genetic contributions to cognitive resilience/resistance

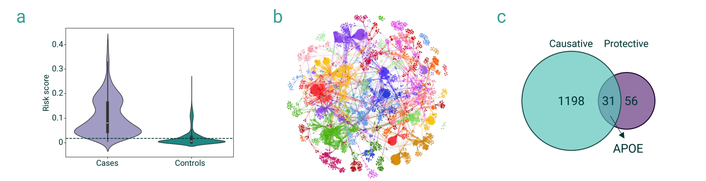

Figure 5. (a) Violin plot showing the distribution of risk scores for AD cases and controls in the ROSMAP/MSBB/Mayo dataset. The dashed line shows the risk score threshold selected to identify the actively protective cohort. (b) Disease architecture diagram demonstrating the 47 communities of SNPs comprising the structure of the Alzheimer's disease patient subpopulations generated by the PrecisionLife platform. Each circle represents a disease-associated SNP genotype, edges represent co-association in patients, and each color a distinct community of SNPs. (c) Venn diagram shows the overlapping genes between the actively protective and causative cohorts from the ROSMAP/MSBB/Mayo cohort.

Discussion

PATIENT STRATIFICATION
SNP-genotype combinations from UKBB 2020 study were clustered based upon the patients in which they were found, generating six major subgroups of patients. Each of these patient groups reflected a specific biological function - lipid metabolism, neuroinflammation, autophagy, serotonin receptor signaling, metal ion homeostasis, and adipose tissue differentiation/fatty acid synthesis. 70% of UKBB 2020 significant SNP-genotype combinations were confirmed in a disjoint UKBB 2023 dataset. Genes linked to biological functions in the UKBB 2020 study were identified in an independent dataset (ROSMAP/MSBB/Mayo). Novel pathways related to neuronal/synaptic activity, cell signaling, and actin/cytoskeleton dynamics were enriched in the ROSMAP/MSBB/Mayo cohort.

NOVEL TARGETS
Our analysis identified combinations of genetic variants which mapped to 113 genes that are significantly associated with increased risk of developing AD. Ninety of these targets were confirmed in the disjoint UKBB 2023 dataset, and 36 were confirmed in the ROSMAP/MSBB /Mayo study.

INDICATION EXTENSION
Genes identified in this analysis that are targeted by drugs in clinical development pipelines may represent potential drug repurposing opportunities. We identify 32 such targets which were evaluated against factors such as prevalence, MoA, safety, route of administration and freedom to operate to assess repositioning potential.

ACTIVELY PROTECTIVE BIOLOGY
Different factors are known to contribute to resistance or resilience in AD, e.g., lifestyle, vascular risk, sleep, sex, and genetics. PL's novel approach makes it possible to evaluate genetics' influence on cognitive resistance and/or resilience by identifying those that present high risk signatures among the healthy controls ("protected" controls). PL combinatorial analysis on the "protected" controls identified 87 genes opening new options for novel target discovery and repurposing opportunities.

The results demonstrate that the PrecisionLife combinatorial analysis is uniquely able to stratify heterogenous patient populations with complex disease pathologies. We can use these insights to identify more effective therapeutic strategies, and accompanying biomarker sets to match them to the patient subgroups that are most likely to demonstrate benefit in downstream clinical trials.

References

1. Das S, Taylor K, Kozubek J, Sardell J, Gardner S. Genetic risk factors for ME/CFS identified using combinatorial analysis. J Transl Med. 2022
2. Taylor K, Pearson M, Das S, Sardell J, Chocian K, Gardner S. Genetic risk factors for severe and fatigue dominant long COVID and commonalities with ME/CFS identified by combinatorial analysis. J Transl Med. 2023
3. Bycroft C, Freeman C, Petkova D, Band G, Elliott LT, Sharp K, Motyer A, Vukcevic D, Delaneau O, O'Connell J, Cortes A, Welsh S, Young A, Effingham M, McVean G, Leslie S, Allen N, Donnelly P, Marchini J. The UK Biobank resource with deep phenotyping and genomic data. Nature. 2018 Oct
4. Bennett DA, Buchman AS, Boyle PA, Barnes LL, Wilson RS, Schneider JA. Religious Orders Study and Rush Memory and Aging Project. J Alzheimers Dis. 2018;64(s1):S161-S189. doi: 10.3233/JAD-179939. PMID: 29865057; PMCID: PMC6380522.
5. The data/analyses presented in the current publication are based on the use of study data downloaded from the dbGaP web site, under phs001963.v1.p1

Acknowledgements

Research described in this study has been conducted using data from the UK Biobank Resource (application number 44289) and from the religious orders study and memory and aging project (ROSMAP), Mount Sinai Brain Bank and Mayo clinic studies available via AD Knowledge Portal. The data available in the AD Knowledge Portal and the Database of Genotypes and Phenotypes (dbGaP) would not be possible without the participation of research volunteers and the contribution of data by collaborating researchers.
Special acknowledgments to the extended PrecisionLife team.

precisionlife
improving health | for everyone

For more information:
www.precisionlife.com