



## Introduction

PrecisionLife is a pioneering techbio company, with a unique approach that finds more signal in complex disease patient data 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 chronic diseases.

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.

## Methods

### DATASET:

PrecisionLife analyzed a dataset constructed from the UK Biobank:

- Alzheimer's disease Cases (n = 882):
  - Alzheimer's disease diagnosis (ICD-10 code, G30.x)
- Healthy Controls (n = 1,816):
  - No reported neurodegenerative disorders
  - No self-reported cognitive decline
  - No family history of Alzheimer's disease

### COMBINATORIAL ANALYSIS:

The dataset was 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, low *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.

**Figure 1. Combinatorial Analysis of Genomic Data vs GWAS**

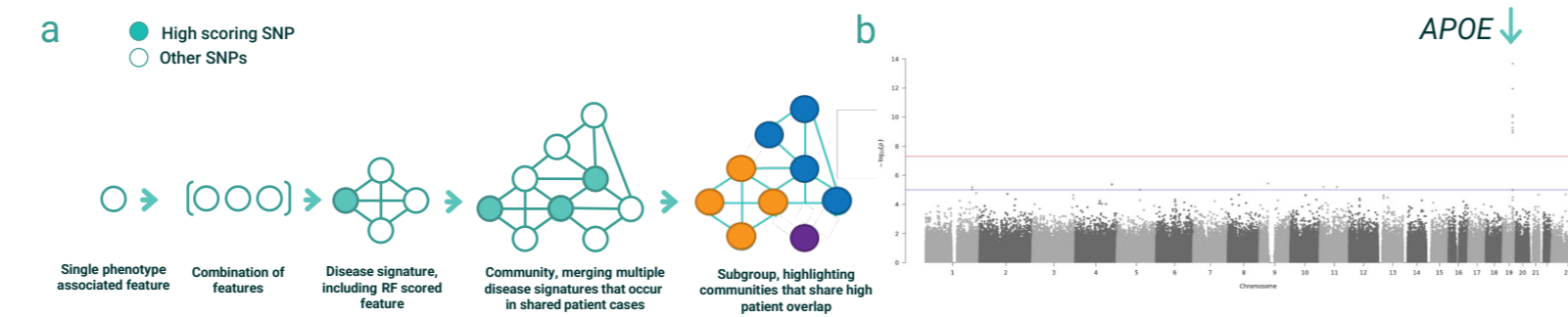


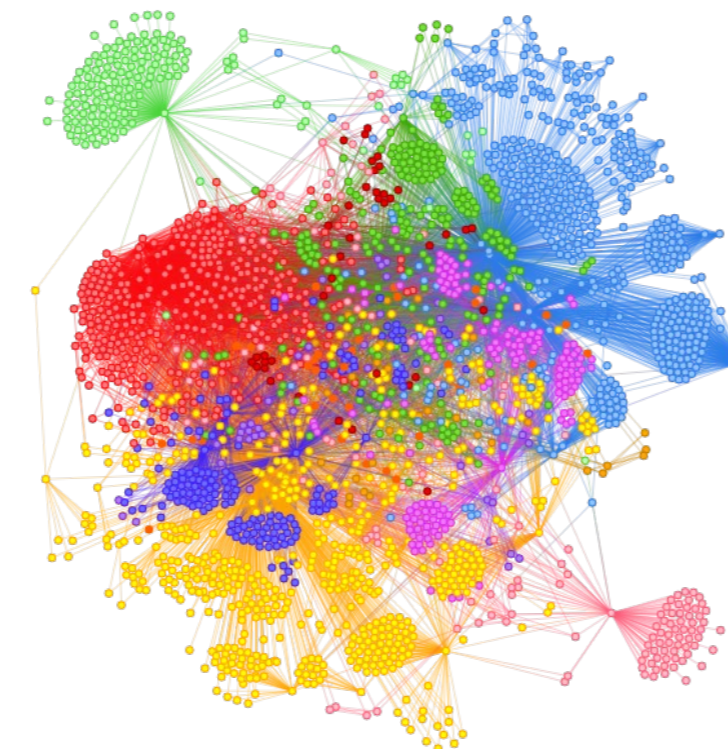
Figure 1a. Conceptual representation of features, combinations, disease signatures and communities used to build up the disease architecture in the PrecisionLife combinatorial methodology. Figure 1b. Manhattan plot of genome-wide *p*-values of association for the AD UK Biobank cohort. The dashed line represents the genome-wide significance threshold at  $p=5e-08$ .

| GWAS  | Combinatorial Analysis   |
|---|--|
| Single SNP associations must be significant across large groups of patients   | Specific combinations of variants associated with each patient subgroup serve as a genetic stratification biomarker                                |
| 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) | Patient subgroups with different causes of disease or even incorrect diagnoses can be distinguished (stratified) by different mechanistic etiology |
| Does not account for the effects of interactions between SNPs, genes and metabolic networks   | Captures epistatic and non-linear additive effects of all interactions between SNPs, genes, environmental factors and metabolic networks           |

**Figure 2. Alzheimer's disease architecture reveals 13 stratified genetic communities**

Figure 2. 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.



**Figure 4. Six patient subgroups are associated with distinct biological pathways and pathological mechanisms**

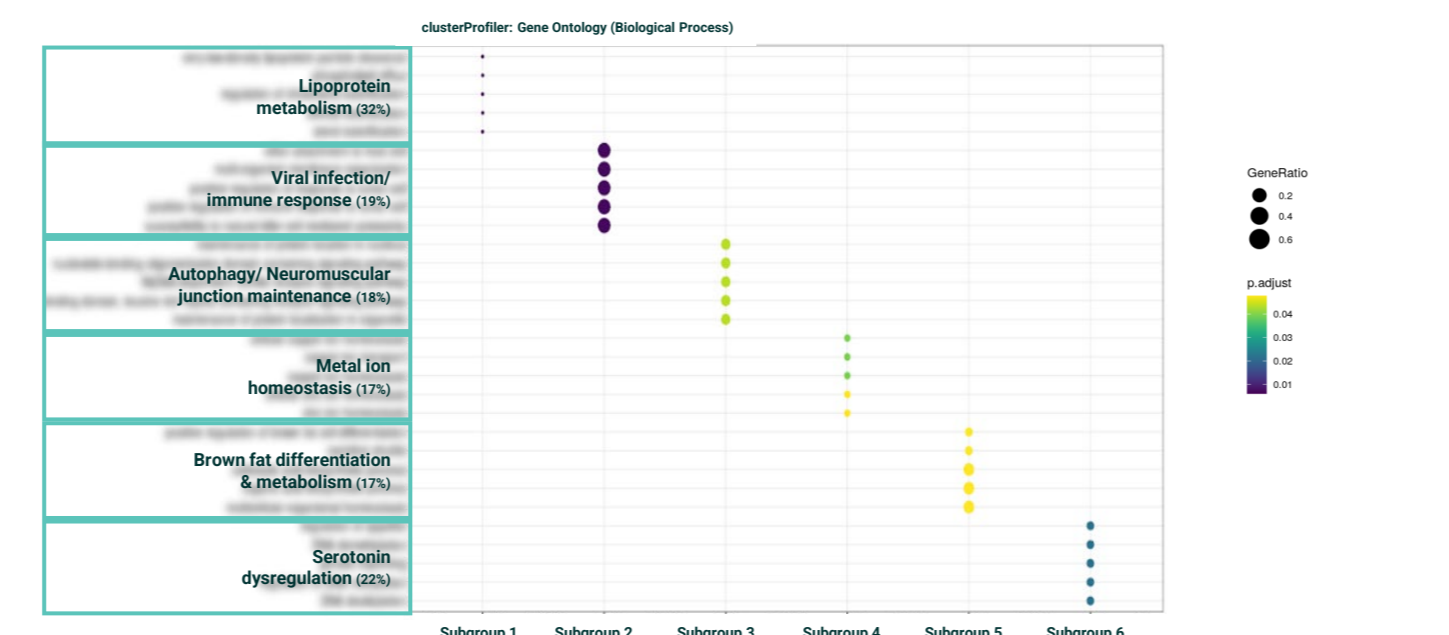


Figure 4. Pathway enrichment plot for the genes found in the communities associated with different patient subgroups. Gene ratio represents the ratio of genes found in the pathway compared to the genes associated with a community and  $p_{adj}$  represents the *p*-value adjusted for multiple testing. The dots in the plot are color-coded based on their corresponding  $p_{adj}$  values. The percentage of cases represented by each subgroup is displayed in brackets, patients may belong to multiple subgroups.

**Table 1. Key Results from PrecisionLife AD Study**

|  |   |
|--|---|
| Validated disease signatures (SNP combinations)              | 4,887                                       |
| Significant RF-scored SNPs associated with AD                | 267   |
| Significant genes associated with AD                         | 113<br>(inc. <i>APOE</i> and <i>APOC1</i> ) |
| Genes targeted by at least one drug in clinical development* | 32  |
| Patient communities identified                               | 13  |
| Patient subgroups (merged communities) identified            | 6   |

\*source: DrugBank and ChEMBL

Table 1. Summary of the results of PrecisionLife's combinatorial analysis of the UK Biobank Alzheimer's disease genomic dataset.

**Figure 3. Patient overlap analysis merges communities into six major genetic subgroups with distinct mechanisms**

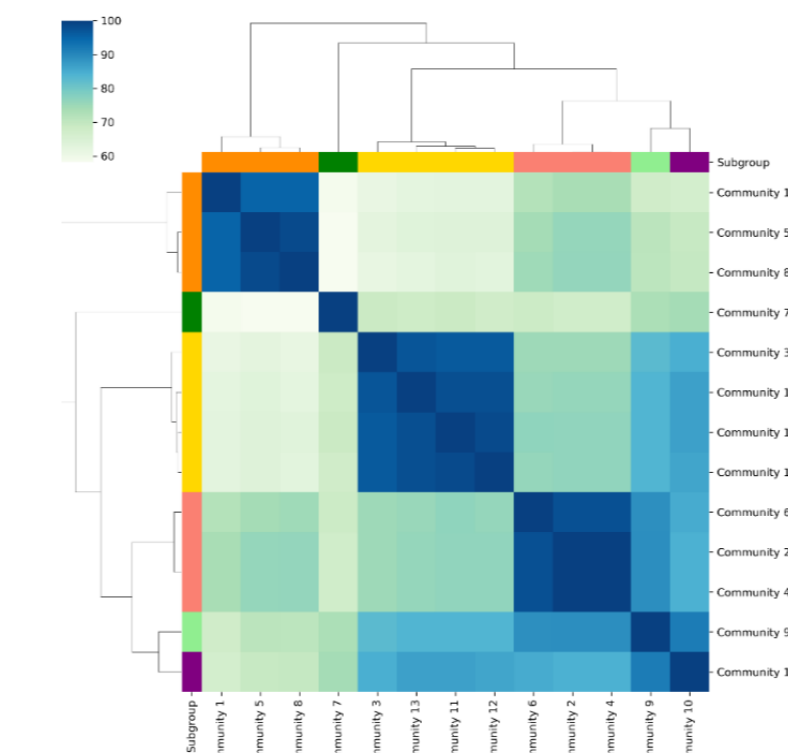


Figure 3. Clustered heatmap showing the overlap of AD patients associated with 13 communities. Each border color represents a patient subgroup (n = 6)

**Figure 5. Pathways identified in AD study are also significant in other CNS diseases**

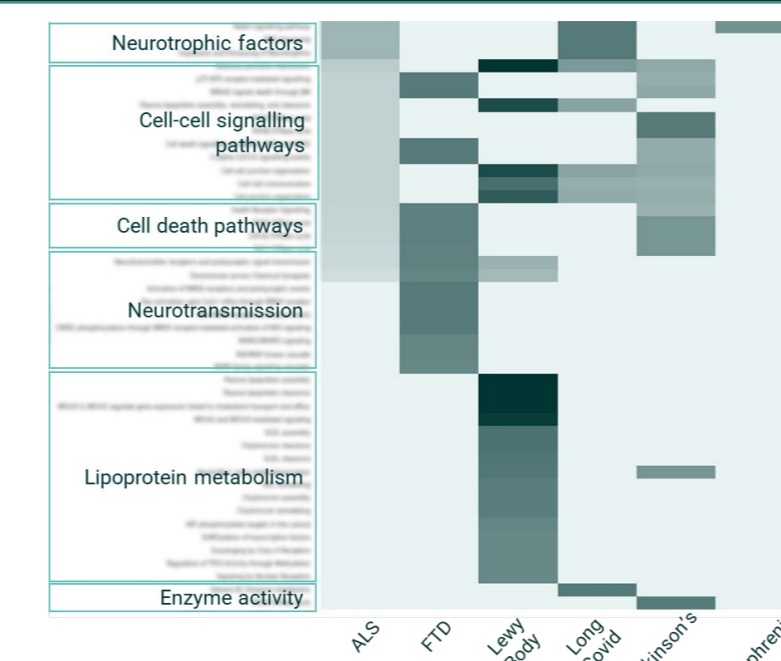


Figure 5. Pathway enrichment for genes identified in AD and other CNS diseases analysed by PrecisionLife. Enrichment analysis was performed using the GProfiler tool to determine pathways significantly enriched in at least one indication ( $p < 0.05$ , *p*-value correction for multiple testing using Benjamini-Hochberg). Scores shown are  $-\log_{10}(p\text{-value})$ .

## Discussion

### PATIENT STRATIFICATION

Clustering these combinations, based upon the patients in which they were found, generated 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.

### NOVEL TARGETS

Our analysis identified combinations of genetic variants which mapped to 113 genes that are significantly associated with AD development. Using further analysis, we shortlisted 26 novel targets to be prioritized based on MoA hypothesis, population prevalence, localization and druggability. Selected targets will be validated in AD-relevant human iPSC-derived neurons using genetic and/or pharmacological approaches. Positively validated targets will be then tested in *in vivo* disease models (mouse/zebrafish). Our cross-disease analysis revealed an overlap between pathways we find enriched in AD and in other CNS indications. This opens an opportunity for identifying biological mechanisms and therapeutic targets relevant to treating multiple related diseases.

### 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 are evaluated against factors such as prevalence, MoA, safety, route of administration and freedom to operate to assess repurposing potential.

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.

### Acknowledgements

Research described in this study has been conducted using data from the UK Biobank Resource (application number 44288).