Using combinatorial analysis to identify actively protective biology, novel causal targets, drug repurposing candidates and mechanistic patient stratification biomarkers in ALS

C. Stubberfield, A. Malinowski, C. Navarron Izquierdo, J. Sardell, J. Kozubek, K. Taylor, S. Das, V. Bouchet, G.L. Møller, S. Gardner
PrecisionLife Ltd, Oxford, UK

Introduction

PrecisionLife (PL) is a pioneer in precision medicine, with a unique approach that finds significantly more signal in complex disease patient data than standard methods.

Our platform uses a hypothesis-free, AI-enabled method to detect combinations of features that together are strongly associated with disease risk, progression rates, treatment response or other clinical phenotypes. This high-resolution mechanistic stratification identifies subgroups of patients with similar disease drivers and treatment responses, to make precision medicine possible in complex diseases[1,2].

Sporadic ALS is characterized by a high degree of heterogeneity across the patient population, reflected in multiple disease etiologies, influences and presentations, while existing familial genetic associations only represent rare sub-populations.

Using independent genomic datasets from sporadic ALS patients, we have stratified the population mechanistically, generating genetic biomarkers that can be used to place an individual in a mechanism-related subgroup. We have identified genes linked to each subgroup that reflect the associated pathology and potentially represent novel targets for therapeutic intervention. We have also provided the first genetic evidence to support the use of repurposed drug treatments in patients from the subgroups most likely to benefit from them.

We have reduced the genetic markers defining ALS mechanistic patient subgroups to a simple saliva based genotyping test that can match patients to existing and/or repurposed treatments. We have also identified targets which may represent actively protective biology providing resilience against ALS pathologies.

Datasets

Sporadic ALS patients' WGS data from Project MinE and Answer ALS was analyzed:

Data Source	Case Numbers	Control Numbers	SNPs Analyzed
Project MinE	1,493	4,479*	938,678
Answer ALS	683	1,185†	650,134

* - UK Biobank was used as a source of additional controls (no evidence of nervous system disease in hospitalisation, self-reported, primary care or death records)

† - Neurologically healthy controls were sourced from dbGaP (phs001963.v1.p1)

Project MinE WGS data (http://databrowser.projectmine.com/) was used in the preliminary causative and actively protective analyses. Answer ALS WGS data (https://dataportal.answerals.org/data-information) was used to replicate the causative findings and to provide additional patient phenotype information.

The additional 'omics data available from Answer ALS was used to support the genomic observations.

Analyses

COMBINATORIAL ANALYSIS

The datasets were analyzed in the PrecisionLife™ combinatorial analytics platform to identify combinations of SNP genotypes that, when observed together in patients, are strongly associated with the risk of developing ALS.

The combinatorial analysis (**Figure 1**) captures the non-linear effects of interactions between different SNPs/genes, such that important biological drivers & insights that are likely to be missed in a GWAS analysis become apparent. It uncovers disease features that may only be significant within subgroups in the disease population. As a result, the approach finds considerable novel disease-related biology.

SNP combinations (disease signatures) 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.

The combinations are clustered together according to the patients in which they co-occur to generate a mechanistic disease architecture (**Figure 2**).

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 mapped to genes for further review of druggability.

The analysis highlighted some genes with prior literature associations to ALS, although most of these had no prior genetic association (**Figure 3**). The majority of the genes found and replicated (19) are however novel.

ACTIVELY PROTECTIVE BIOLOGY AND TARGETS

High-scoring disease signatures from the causative analyses were collated to generate a disease risk score (**Figure 4**). This was used to identify individuals from the control population who had a high prevalence of disease risk signatures and were as old as possible, but had not developed ALS (**Figure 6**). A comparative analysis was then run, comparing patients against these high-risk controls to identify SNP combinations that distinguished the "protected" controls from those with disease.

This approach aims to identify genes which may be active in the background of cellular biology, working to actively resist disease pathophysiology. These actively protective disease resilience targets are a new class that may be considered to be strong candidates for therapeutic intervention, e.g., with therapeutic mRNA vaccines, as well as adding to our understanding of the disease etiology and pathology.

For example, our analysis uncovered a potentially protective gene (*MTTR*) specifically relating cobalamin/vitamin B12 metabolism to ALS (**Figure 7**).

Combinatorial Analysis of Genomic Data vs. GWAS

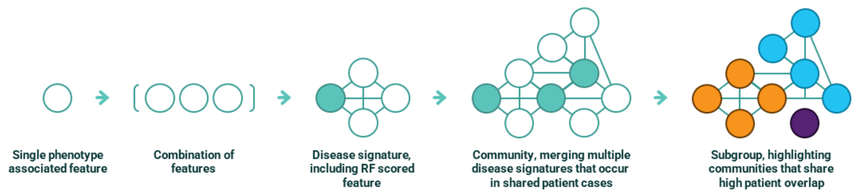

Figure 1. Conceptual representation of features, combinations, disease signatures and communities used to build up the disease architecture in the PrecisionLife combinatorial methodology

GWAS	Combinatorial Analysis
Single SNP associations must be significant across whole populations of patients so struggle with heterogenous and polygenic diseases	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), which limits reproducibility in different ancestries	Patient subgroups with different causes of disease or even incorrect diagnoses can be distinguished (stratified) by different mechanistic aetiology and results replicate better across populations with different ancestries
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

Novel and Known Causal Drug Targets

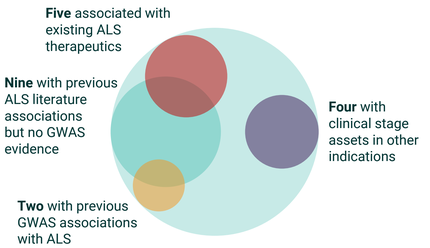

Figure 3. Existing knowledge of potential therapeutic targets identified in causative analyses

From 32 genes identified in causative analyses:

- 13 have some previous link with ALS, but only 2 have previous GWAS evidence
- **5 have been postulated to be involved in the mechanism of existing ALS therapeutics (riluzole, retigabine, baclofen), but only 2 of these genes have previously been linked to the disease**
- The products of 4 of the genes are the subjects of clinical development for other indications

Two of the novel targets are in active validation studies with partners.

Actively Protective Biology Resisting ALS

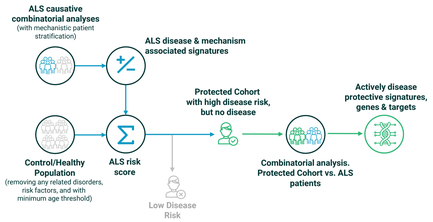

Figure 6. Analysis to find genes that actively prevent disease.

PL uses the predictive accuracy of the risk signatures from causative analyses to define a cohort of people who have all the risk signatures that should lead them to develop disease, have exposure to as many disease triggers and are as old as possible, but who have not been affected by the disease.

This cohort (selected for the longest/maximum exposure to disease risk factors) can be inferred to be protected from developing the disease by some other factors, which can be then identified by a further combinatorial analysis against patients.

Predicting ALS Disease Risk

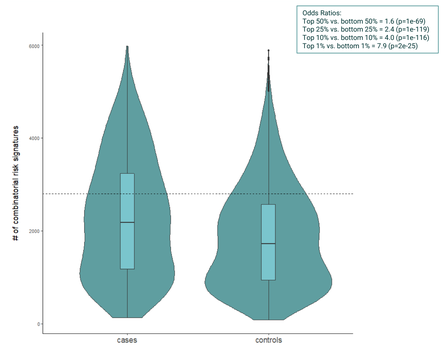

Figure 4. Violin plots for disease risk signatures in ALS patient and control populations

Higher counts of disease risk signatures are significantly associated with risk of developing ALS (ALS patients have significantly more than healthy controls). Dashed lines in both plots represent the 75th percentile signature count.

The distributions shown here were generated by counting the number of risk signatures found in ALS patients and controls. Taking a 25th percentile cut-off, **the odds ratio for this distribution variance is 2.4.**

Example Actively Protective Gene - *MTRR*

Gene	Name	Function
From actively protective analyses		
MTRR	Methionine synthase reductase	• Key enzyme in one-carbon metabolism, supporting methionine and folate homeostasis • Responsible for the reactivation of methionine synthase (MTR/MS) activity by catalyzing the reductive methylation of MTR-bound cobalamin
From causative analyses		
CUBN	Cubilin	• Endocytic receptor which plays a role in lipoprotein, iron and vitamin B_{12} metabolism by facilitating their uptake • Acts together with AMN to mediate endocytosis of the CBLIF-cobalamin complex
LRP2	Megalin	• An endocytic receptor for more than 75 putative ligands, including cubilin-cobalamin • Expressed on neurons, astrocytes, microglia and oligodendrocytes and implicated in diseases of the kidney and brain
CD320	Transcobalamin receptor	• Receptor for transcobalamin carrying cobalamin • CD320-/- mice develop cobalamin deficiency in the nervous system and indicate preferential loss of highly myelinated neurons

Figure 7. Functions of pathway-associated genes identified in actively protective and causative analyses

The actively protective analyses identified a series of 21 genes as significantly associated with individuals who had a high-risk score for developing ALS but nonetheless remained healthy.

This set included *MTRR*, a key enzyme in the role of cobalamin-dependent methionine synthesis. 3 other genes involved in cobalamin/vitamin B12 uptake and function (*CUBN, LRP2, CD320*) had been highlighted in the causative analyses, and hence our attention was drawn to the role of this pathway in ALS.

ALS Disease Architecture

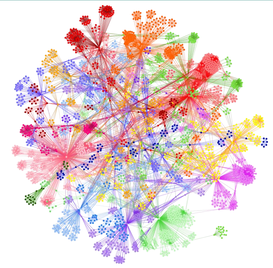

Figure 2. Disease architecture diagram showing the multiple communities of individuals who share specific sets of SNPs, which make up the mechanistic stratification of the ALS patient subgroups generated by the PrecisionLife platform

The complex architecture reflects the high disease heterogeneity.

Each circle represents a disease-associated SNP genotype, edges represent co-association in patients, and each colour a distinct community of SNPs.

Novel Saliva Based Diagnostic Tests

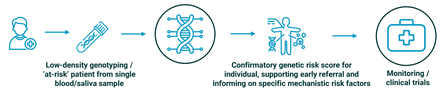

Figure 5. Application of patient stratification biomarkers as a potential early diagnostic and/or therapy selection tool

The risk signatures can be evaluated in a saliva based genotyping test used to support diagnosis and match patients to potential therapies.

Cobalamin / Vitamin B12 Biology

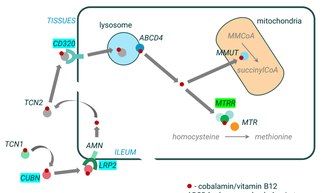

Figure 8. Uptake and function of cobalamin/vitamin B12

Genes highlighted in actively protective (■) and causative analyses (■).

Role of Vitamin B12 in ALS

- Vitamin B12 is required as a cofactor for two reactions in humans, the cytosolic methionine synthase reaction and the mitochondrial methymalonyl CoA mutase reaction[3].
- B12 is critical for hematopoiesis and myelination and deficiency can be associated with progressive tremor, ataxia, and scanning speech[4].
- The vitamin B12 analogue, hydroxocobalamin, protects neuronal cell lines from TDP-43-induced mitochondrial damage and neurotoxicity[5]. Methylcobalamin prevents motor neuron death induced by co-culture with mutant *SOD1* (G93A) expressing astrocytes[6].
- Methylcobalamin ('ultra-high dose', 30 mg/kg) significantly delayed progression of motor and neurological symptoms in the wobbler mouse model of ALS[7].
- Elevated levels of homocysteine, which is known to have neurotoxic effects, are observed in the CSF of ALS patients[8].
- High dose methylcobalamin was efficacious in slowing functional decline in patients with early stage ALS[9].

Figure 9. Literature supporting genetic evidence showing potential ALS protective role of vitamin B12 in some patients

Discussion

Our causative combinatorial analyses generated an architecture for ALS that reflects the complex make-up of the disease and identified genetic/mechanistic disease signatures for the patient subgroups which describe their underlying pathologies.

We can associate disease subgroups and their defining SNPs/genes, with recognised clinical subsets, e.g. progressive muscular atrophy, early onset, fast progressor etc.

As well as identifying a series of potential therapeutic targets novel to ALS, some of which are already being explored in other indications, **the causative studies also found the first genetic support for existing medications that are used in the treatment of ALS and its symptoms, i.e., riluzole, retigabine, baclofen. Hence it would be possible to use these mechanistic biomarkers to select patients most likely to benefit from these therapies.** In the case of baclofen, it may indicate a greater benefit beyond the symptomatic treatment for which it is currently employed.

The disease signatures that emerged from our analyses can also be developed as the basis for patient stratification biomarkers (Mechanostics™). **These offer a route to creating a rapid, cheap early diagnostic tool** (**Figure 5**). A test to count the number of risk signatures in a patient's makeup can be run from a single saliva sample on a low-density genotyping array. These report both the patient's disease risk and their likely mechanistic cause, and hence can match them to the drug most likely to benefit them (including repurposed drugs). It also provides a differential triage tool to suggest the likely diagnosis in the face of of easily confused common symptoms.

Taking all the causative disease signatures together, we were able to build a combined disease risk score that can predict high disease risk. This was used to identify individuals from the control populations who were predicted to have a high risk of developing disease but did not. Comparing these individuals with ALS patients, **we were able to perform an actively protective analysis to identify SNPs/genes which may be acting to promote resilience and prevent disease**.

Review of the actively protective analysis highlighted several genes potentially important in preventing disease, including *GRIP1*, which has previously been associated with ALS reversal[10], and *MTRR*, which codes for an enzyme that is central to the role of vitamin B12 within the cell. This observation is supported by findings from the causative analyses which saw signal from three other genes linked with cobalamin/vitamin B12 uptake/transport – *CUBN*, *LRP2* and *CD320* (***Figure 7 & 8***).

Vitamin B12 is important for the conversion of potentially neurotoxic homocysteine to methionine and its absence can result in defective myelin synthesis and repair. Vitamin B12/methylcobalamin has previously been explored in cellular, animal and clinical studies for its role in ALS, with promising outcomes (**Figure 9**). However there has not been any previous genetic basis to link it to the disease.

Conclusion

PL's combinatorial analytical approach identified mechanistic biomarkers that can distinguish ALS patients most likely to benefit from specific existing / future drugs.

It also identified novel potential actively protective genes & targets, which could potentially become targets of therapeutic mRNA vaccines.

Amongst our findings we report on the first patient-derived, genetic evidence for the benefit of methylcobalamin/vitamin B12 in the treatment of some ALS patients. We hope this will encourage validation of its utility and more widespread use.

References

1. Das S, *et al*, Genetic risk factors for ME/CFS identified using combinatorial analysis. - 2022, *J Transl Med.* **20**(1): 598. https://pubmed.ncbi.nlm.nih.gov/36517845/
2. Taylor K, *et al*, Genetic risk factors for severe and fatigue dominant long COVID and commonalities with ME/CFS identified by combinatorial analysis. 2023, *J Transl Med.* **21**(1): 775. https://pubmed.ncbi.nlm.nih.gov/37915075/
3. McCorvie TJ *et al*, The complex machinery of human cobalamin metabolism. 2023, *J Inherit Metab Dis.* **46**(3): 406-420. https://pubmed.ncbi.nlm.nih.gov/36680553/
4. Pluvinage JV *et al*, Transcobalamin receptor antibodies in autoimmune vitamin B12 central deficiency. 2024, *Sci Transl Med.* **16**(753): eadl3758. https://pubmed.ncbi.nlm.nih.gov/38924428/
5. Jeon YM *et al*, Vitamin B12 Reduces TDP-43 Toxicity by Alleviating Oxidative Stress and Mitochondrial Dysfunction. 2021 *Antioxidants (Basel).* **11**(1): 82. https://pubmed.ncbi.nlm.nih.gov/35052586/
6. Ito S *et al*, Methylcobalamin prevents mutant superoxide dismutase-1-induced motor neuron death *in vitro*. 2017, *Neuroreport.* **28**(2): 101-107. https://pubmed.ncbi.nlm.nih.gov/27922548/
7. Ikeda K *et al*, Neuroprotective effect of ultra-high dose methylcobalamin in wobbler mouse model of amyotrophic lateral sclerosis. 2015, *J Neurol Sci.* **354**(1-2): 70-74. https://pubmed.ncbi.nlm.nih.gov/25982504/
8. Wu Y *et al*, Elevated cerebrospinal fluid homocysteine is associated with blood-brain barrier disruption in amyotrophic lateral sclerosis patients. 2020, *Neurol Sci.* **41**(7):1865-1872. https://pubmed.ncbi.nlm.nih.gov/32086685/
9. Oki R *et al* Efficacy and Safety of Ultrahigh-Dose Methylcobalamin in Early-Stage Amyotrophic Lateral Sclerosis: A Randomized Clinical Trial. 2022, *JAMA Neurol.* **79**(6): 575-583. https://pubmed.ncbi.nlm.nih.gov/35532908/
10. Crayle JI *et al*, Genetic Associations With an Amyotrophic Lateral Sclerosis Reversal Phenotype. 2024, *Neurology* **103**(4):e209696. https://pubmed.ncbi.nlm.nih.gov/39079071/

Acknowledgements

Research described in this study has been conducted using data from Project MinE, King's College London, Answer ALS, Dementia-Seq (NIH dbGAP) and the UK Biobank Resource (application number 44288).

We would like to thank the MNDA, the wider PrecisionLife technical and scientific teams, and all our collaborators for their continuing support, ideas and encouragement.

This data would not be available and this study not possible without the generous participation of research volunteers and the contribution of data by collaborating researchers.

precisionlife
improving health | **for everyone**

For more information, please visit:
www.precisionlife.com/cns