Combinatorial Analysis of ALS Patients Uncovers New Disease Drivers in Genetically Defined Subgroups

A. Malinowski,[1] C. Navarron Izquierdo,[1] J. Kozubek,[1] K. Taylor,[1] J. Sardell,[1] S. Das,[1] A. Iacoangeli,[2] A. Al Khleifat,[3] A. Al-Chalabi,[3] C. Stubberfield[1]

[1]PrecisionLife Ltd, Oxford, UK
[2]Department of Biostatistics and Health Informatics, Institute of Psychiatry, Psychology and Neuroscience, King's College London, London, UK
[3]Maurice Wohl Clinica¹ Neuroscience Institute, Institute of Psychiatry, Psychology and Neuroscience, King's College London, London, UK

Scan to view online

Introduction

PrecisionLife (PL) is a pioneering precision medicine company, with a unique approach that finds 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 characterised by a high degree of heterogeneity across the patient population, reflected in multiple disease etiologies, influences and presentations, while existing familial genetic associations generally represent rare sub-populations.

Using independent genomic datasets from sporadic ALS patients, we have mechanistically stratified the population, generating genetic biomarkers which can be used to place an individual in a subgroup. In each of these subsets, we have identified genes which reflect the associated pathology and potentially represent novel targets for therapeutic intervention.

Amongst these novel targets, we report RBFOX1, an RNA-binding protein involved in alternative splicing, which has not been previously associated with ALS, but has existing links to key processes implicated in the disease.

Datasets

WGS data from patients diagnosed as having sporadic ALS was analysed from Project MinE (http://databrowser.projectmine.com/) and Answer ALS (https://dataportal.answerals.org/data-information).

Data Source	Case Numbers	Control Numbers	SNPs Analysed
Project MinE	1493	4479*	938,678
Answer ALS	683	1185†	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 sourced from dbGaP (phs001963.v1.p1)

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

Methods

The datasets were analysed in the PrecisionLife 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 and insights that are likely to be missed in a GWAS analysis are detected. 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 prioritised. 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 prioritised based on their ability to differentiate cases and controls. The highest scoring SNPs are mapped to genes for further review.

The analysis highlighted some genes with prior literature associations to ALS, although most of these had no prior genetic association (**Figure 4**). The majority of the genes found in the Project MinE dataset and replicated in the Answer ALS dataset are however novel.

Results

Combinatorial Analysis of Genomic Data vs. GWAS

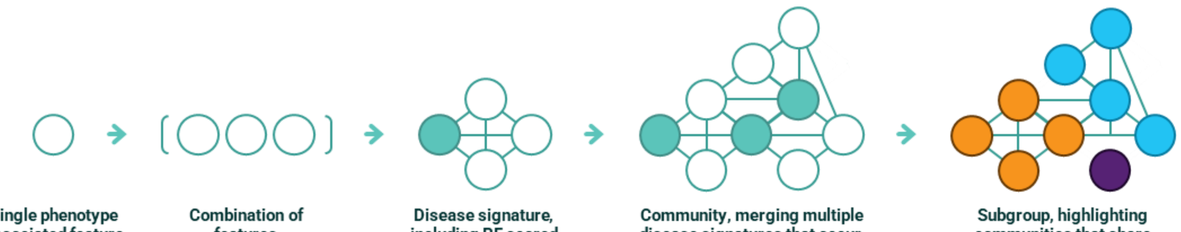


Figure 1. Conceptual representation of features, combinations, disease signatures and communities used to build up the disease architecture in the PrecisionLife combinatorial methodology.

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 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

ALS Disease Architecture

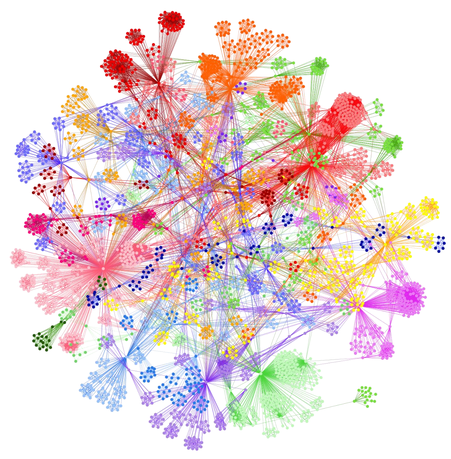


Figure 2. Disease architecture diagram showing the multiple communities of individuals who share specific sets of SNPs, which make up the mechanistic stratification of the ALS patient subgroups generated by the PL platform.

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

Potential Targets

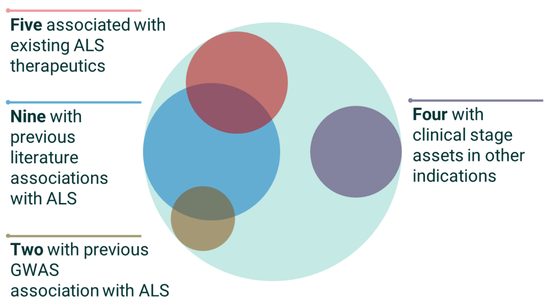


Figure 4. Status of potential therapeutic targets identified in PL analyses.

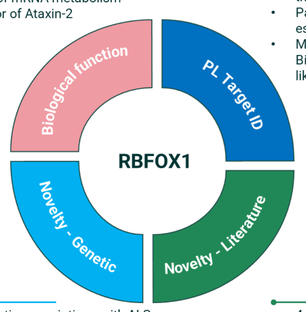
From 32 genes identified in PL analyses,

- 13 have some prior link with ALS, but only 2 have GWAS evidence
- 5 have been linked to existing ALS therapies (riluzole, retigabine, baclofen), but only 2 of these have previously been linked to the disease.
- 4 targets are the subjects of clinical development in other indications.

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

RBFOX1: Novel Target Candidate in ALS

- RNA-binding fox-1 homolog 1 Accession number: Q9NWB1
- RNA-binding protein
- Regulates alternative splicing and other aspects of mRNA metabolism
- Direct interactor of Ataxin-2



- Genetic signal in Project MinE and Answer ALS datasets
- RBFOX1 SNP found in a disease signature containing SNPs in three other genes
- Patient prevalence estimated at 13%
- Mechanistic Patient Stratification Biomarkers to identify patients likely to respond to modulation

- No previous genetic associations with ALS
- Significant GWAS associations with Alzheimer's and amyloid deposition
- Variants associated with autism and psychiatric disorders and phenotypes

- 4 publications broadly linking RBFOX proteins to ALS/MND

Figure 5. RBFOX1 target snapshot.

RBFOX1 is a novel genetic association with ALS and a potential therapeutic target identified by PL. Previously, genetic variants in RBFOX have been linked with AD and neuropsychiatric disorders.

Pathway Enrichment

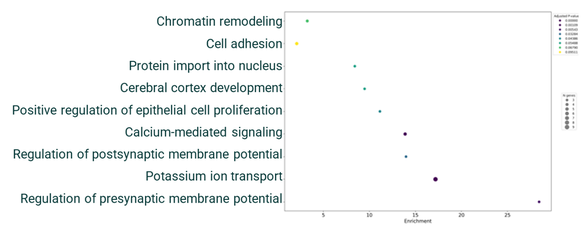


Figure 3. Pathway enrichment plot for prioritised genes associated with ALS.

Gene ratio: genes found in the pathway compared to the genes identified and *p.adjust* represents the p-value adjusted for multiple testing. The dots in the plot are colour-coded based on their corresponding *p.adjust* values.

RBFOX1 Expression Patterns

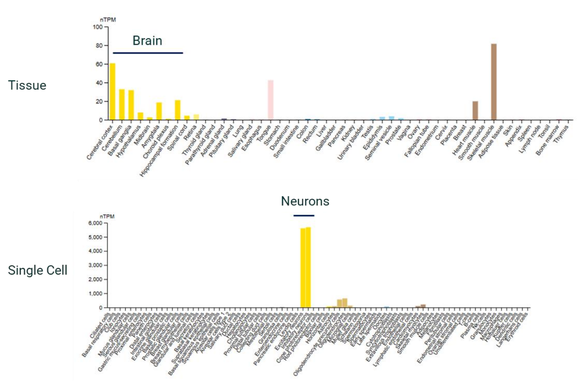


Figure 6. RBFOX1 expression in tissues and cells.

RBFOX1 expression is enhanced in the brain, skeletal muscle and tongue. On a cell type level, RBFOX1 is enriched in inhibitory and excitatory neurons. Image credit: Human Protein Atlas[12], proteinatlas.org/RBFOX1

RNA-Binding Proteins Regulate Key Biological Processes in Neurons

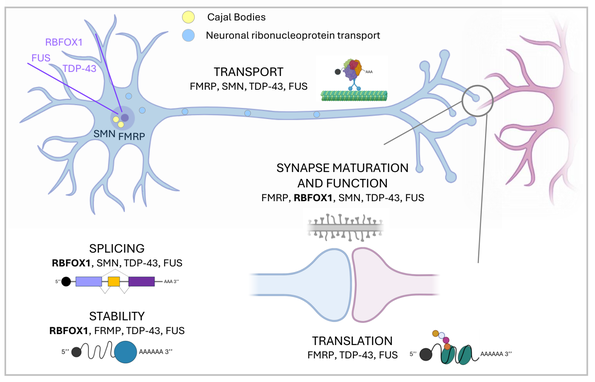


Figure 7. RBFOX1 and other RNA-binding proteins impact neuronal functions that become dysregulated in disease.

Localisation and functions of RBFOX1 and other RBPs implicated in ALS including TDP-43 and FUS. Adapted from[3], created with BioRender.com.

RBFOX1 in ALS disease mechanisms

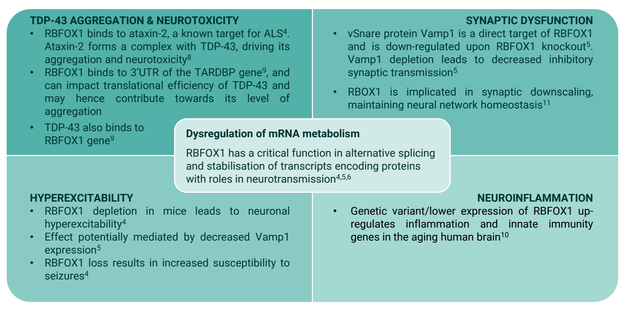


Figure 8. RBFOX1 is implicated in processes impaired in ALS.

Summary of existing evidence supporting potential roles of RBFOX1 in disease mechanisms involved in ALS.

Discussion

PL's combinatorial analyses generated an ALS architecture which reflects the heterogeneity of the disease and identified specific genetic/mechanistic signatures, highlighting disease drivers in patient subgroups.

Our approach enables the association of disease subgroups, and their defining SNPs/genes, with recognised clinical features, e.g. progressive muscular atrophy, early on-set, fast progressor etc.

We identified a series of potential drivers and therapeutic targets in ALS, with varying degrees of novelty and prior evidence linking them to the disease. Interestingly, our studies found the first genetic evidence for targets modulated by existing therapies indicated for MND and its symptoms, such as riluzole, retigabine and baclofen. However, the majority of candidate genes we identified are novel to ALS. This includes therapeutic targets which are already being pursued in other indications as well as proteins that, to our knowledge, have not been explored as drug targets in any disease. Amongst the latter, we identified RBFOX1, an RNA-binding protein, previously linked to neurodegenerative and neuropsychiatric disorders. Existing evidence implicates RBFOX1 in biological mechanisms known to be impaired in ALS, including TDP-43 aggregation, synaptic dysfunction and neuroinflammation.

The disease signatures that emerged from our analyses can also be used as the basis for patient stratification biomarkers (PSBs) which can be used to match RBFOX1 and other potential therapeutic targets with the patients most likely to benefit from their modulation.

Conclusions & Next Steps

PL's combinatorial analyses mechanistically stratified heterogenous population of ALS patients, and identified novel genetic drivers of the disease, which represent potential targets linked to specific patient subgroups. Amongst our findings, we report a candidate therapeutic target RBFOX1, that has been implicated in several biological mechanisms believed to underlie the ALS pathology. We will apply our PSBs to identify patients from the Answer ALS cohort, who have a disease mechanism linked to RBFOX1 and are likely to benefit from its modulation. iPSCs derived from those patients will be developed into ALS-relevant *in vitro* systems for biological validation studies.

References

1. Das *et al*. Genetic risk factors for ME/CFS identified using combinatorial analysis. J Transl Med. 2022;20(1):598. Published 2022 Dec 14. doi:10.1186/s12967-022-03815-8
2. Taylor *et al*. Genetic risk factors for severe and fatigue dominant long COVID and commonalities with ME/CFS identified by combinatorial analysis. J Transl Med. 2023;21(1):775. Published 2023 Nov 1. doi:10.1186/s12967-023-04588-4
3. Prashad & Gopal. RNA-binding proteins in neurological development and disease. RNA Biol. 2021;18(7):972-987. doi:10.1080/15476286.2020.1809186
4. Gehman *et al*. The splicing regulator Rbfox1 (A2BP1) controls neuronal excitation in the mammalian brain. Nat Genet. 2011;43(7):706-711. Published 2011 May 29. doi:10.1038/ng.8414
5. Vuong *et al*. Rbfox1 Regulates Synaptic Transmission through the Inhibitory Neuron-Specific vSNARE Vamp1. Neuron. 2018;98(1):127-141.e7. doi:10.1016/j.neuron.2018.03.008
6. Weyn-Vanhentenryck *et al*. HITS-CLIP and integrative modelling define the Rbfox splicing-regulatory network linked to brain development and autism. Cell Rep. 2014;6(6):1139-1152. doi:10.1016/j.celrep.2014.02.005
7. Becker *et al*. Therapeutic reduction of ataxin-2 extends lifespan and reduces pathology in TDP-43 mice. Nature. 2017;544(7650):367-371. doi:10.1038/nature22038
8. Elden *et al*. Ataxin-2 intermediate-length polyglutamine expansions are associated with increased risk for ALS. Nature. 2010;466(7310):1069-1075. doi:10.1038/nature09320
9. Yang *et al*. CLIPdb: a CLIP-seq database for protein-RNA interactions. BMC Genomics. 2015;16(1):51. Published 2015 Feb 5. doi:10.1186/s12864-015-1273-2
10. Yang *et al*. Genetics of Gene Expression in the Aging Human Brain Reveal TDP-43 Proteinopathy Pathophysiology. Neuron. 2020;107(3):496-508.e6. doi:10.1016/j.neuron.2020.05.010
11. Rajman *et al*. A microRNA-129-5p/Rbfox crosstalk coordinates homeostatic downscaling of excitatory synapses. EMBO J. 2017;36(12):1770-1787. doi:10.15252/embj.201695748
12. Uhlén *et al*. Proteomics. Tissue-based map of the human proteome. Science. 2015;347(6220):1260419. doi:10.1126/science.1260419

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 without the participation of research volunteers and the contribution of data by collaborating researchers.

precisionlife
improving health | for everyone

For more information, visit:
www.precisionlife.com