Combinatorial Analysis of ALS and FTD Patient Genomes to Identify Cross-Disease Mechanisms

A. Malinowski, K. Collins, J. Kozubek, K. Taylor, S. Das, J. Sardell, C. Stubberfield
PrecisionLife Ltd, Oxford, United Kingdom

Scan to view online

Introduction

PrecisionLife® is a computational biology company focusing on precision medicine analytics in complex chronic diseases. Our platform utilises a hypothesis-free method for the detection of combinations of features that are strongly associated with variation in disease risk, progression rates and other clinical phenotypes often observed in patient subgroups. Resulting patient stratification insights allow us to identify mechanistic subgroups with different indications.

Amyotrophic lateral sclerosis (ALS) and frontotemporal dementia (FTD) are complex, progressive neurodegenerative disorders that form a broad disease spectrum. This is reflected in the overlap in clinical symptoms and the fact that mutation of the C9ORF72 gene is a genetic risk factor for both disorders. However, other drivers and mechanisms shared by the two diseases remain largely unknown.

PrecisionLife aimed to identify genetic similarities between ALS and FTD to elucidate common candidate pathways and genes acting as biological drivers of both diseases.

Methods

DATASETS:
FTD: Dementia-Seq NIH dbGAP
936 cases, 1176 controls
ALS (two independent cohorts):
1. UK Biobank (UKBB)
546 cases (ICD10 code G122),
7999 controls (no evidence of nervous system disease, no ICD10 G)
2. Project MinE (UK Cohort) with UKBB controls
1487 cases, 7637 controls

COMBINATORIAL ANALYSIS:
The FTD dataset (lead study) was analysed in the PrecisionLife platform to identify combinations of SNP genotypes that when observed together in a patient are strongly associated with FTD.

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.

Subsequently, the two independent ALS cohorts were screened for presence of the validated SNP combinations (disease signatures) associated with FTD. SNPs identified as part of those disease signatures were then scored using a Random Forest algorithm in a 5-fold cross-validation framework, prioritized based on their ability to differentiate cases and controls, and then mapped to genes.

Genes identified in the FTD-ALS analyses were used for gene expression and gene set enrichment analyses, to identify overlap in affected biological pathways and cell types.

Results

Figure 1. Combinatorial analysis of genomic data vs GWAS

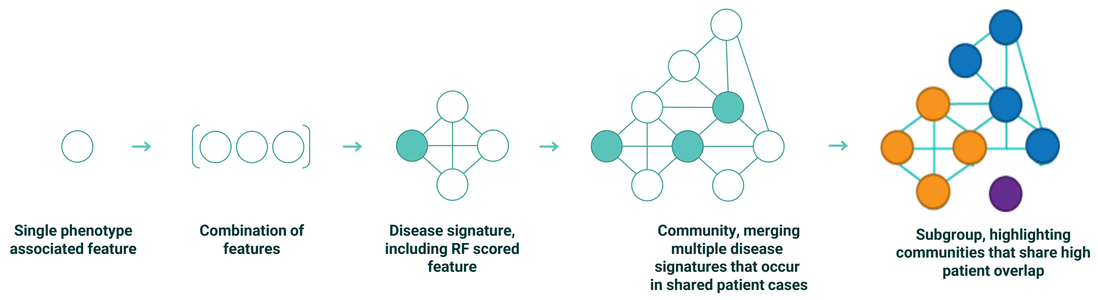

Figure 1. Conceptual representation of features, combinations, disease signatures and communities used to build up the disease architecture in the PrecisionLife combinatorial methodology.

GWAS	Combinatorial Analysis
Single SNP associations must be significant across 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 aetiology
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

Table 1. ALS/FTD overlap analysis using two independent ALS cohorts

	ALS UK Biobank cohort	ALS Project MinE cohort
Disease signatures also associated with increased risk of FTD (p<0.05)	17/688	40/688
Unique SNPs present in disease signatures	50	94
Significant (RF-scored) SNPs mapped to genes	11	17
Genes also associated with increased risk of FTD	12	17

Table 1. Summary of the results of PrecisionLife's analysis to identify FTD-associated genetic risk factors replicating in two independent ALS cohorts.

Figure 2. Genetic risk factors associated with ALS and FTD

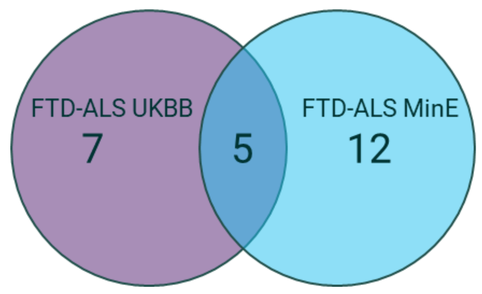

Figure 2. Overlap of the genes identified by screening for FTD-associated risk factors in two independent ALS cohorts.

Figure 3. Gene set enrichment analysis indicates biological processes associated with both ALS and FTD

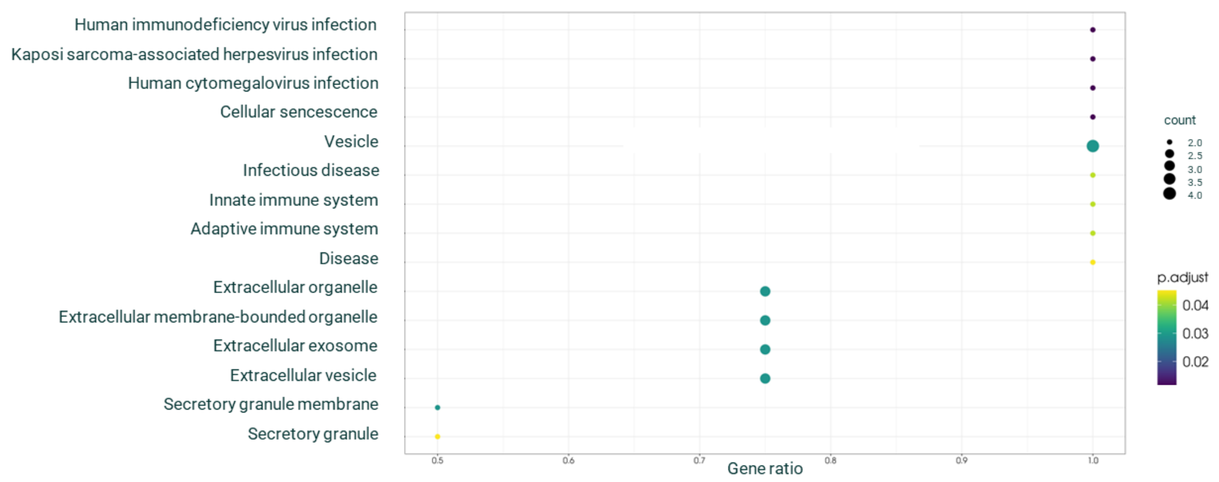

Figure 3. Pathway enrichment plot for genes associated with ALS and FTD, based on genes common between the FTD-ALS UKBB and FTD-ALS MinE analyses. Gene ratio: genes found in the pathway compared to the genes identified; *p.adjust* represents the p-value adjusted for multiple testing. The dots in the plot are color-coded based on their corresponding *p.adjust* values.

Figure 4. Cell type expression of genes associated with ALS and FTD

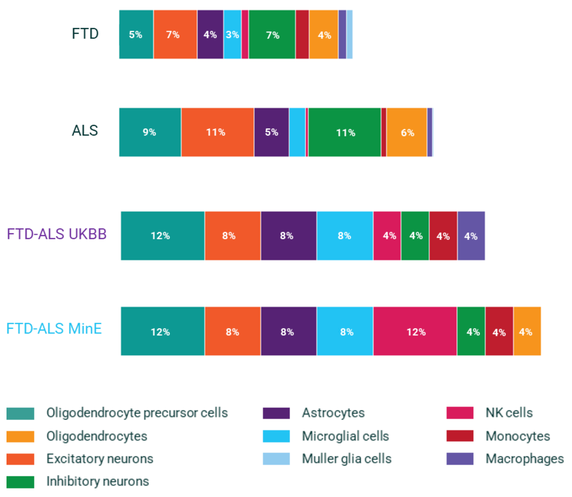

Figure 4. The composition bar plots showing the percentage of genes found in ALS and FTD alone (previous PL studies) and in FTD-ALS analyses, which are expressed (enhanced) in selected neuronal and immune cells (Human Protein Atlas).

Table 2. Candidate genes associated with increased risk of ALS and FTD

	Gene 1	Gene 2	Gene 3	Gene 4
Protein class	Microfilament motor protein	Transmembrane protein	Inositol 1,4,5-triphosphate receptor	Class I Histocompatibility Antigen
Protein function	Muscle contraction	Cellular response to unfolded proteins	Mediator of calcium signaling	Involved in immune response
Brain expression	Not detected	Neurons, microglia, oligodendrocytes	Microglia, astrocytes, oligodendrocyte precursors, neurons	Expressed in glial cells, enriched in immune cells
Genetic evidence in ALS/FTD	No GWAS associations	No GWAS associations	Non-significant GWAS association with sporadic ALS	CNV in coding region associated with ALS, non-significant
Other types of evidence in ALS/FTD	No links	No links	Expression upregulated in spinal cord of ALS patients	Expression upregulated in spinal cord of ALS patients

Table 2. Four candidate genes identified in ALS/FTD overlap analysis. The table provides information regarding their protein class and function, expression patterns in the brain, and existing genetic and literature/expression links to ALS and FTD.

Discussion

Our combinatorial analysis identified genetic disease signatures associated with FTD. Subsequent replication in two independent ALS cohorts revealed SNP combinations shared between the two indications.

SNPs belonging to these overlapping signatures were prioritised and mapped to candidate genes. Subsequent pathway enrichment analysis linked the resulting genes with biological processes previously implicated in the two neurodegenerative diseases. Among those, we found association with vesicular trafficking, which has been reported as one of the central pathways in these forms of neurodegeneration[1]. In addition, our analysis indicated new potential mechanisms underlying both ALS and FTD. These included a link to viral infection. Viruses have been previously implicated as environmental risk factors for ALS, but not for FTD [2].

Finally, we prioritised four genes as candidate risk factors shared by ALS and FTD. Proteins encoded by those genes represent different cellular functions, and tissue/cell type expression patterns, likely reflecting the heterogeneity of both diseases. Further studies will be required to evaluate these genes as starting points for potential therapeutic strategies.

Conclusion

The results indicate that the PrecisionLife combinatorial analysis is able to identify genetic risk factors shared by ALS and FTD. Subsequent analysis confirmed known biological processes disrupted in the two indications, and points to potential novel shared mechanisms. This opens an opportunity for identifying biological pathways and therapeutic targets relevant to treating both ALS and FTD.

References

1. Abramzon et al.,. The Overlapping Genetics of Amyotrophic Lateral Sclerosis and Frontotemporal Dementia. Front Neurosci. 2020 Feb 5:14:42
2. Bellmann et al.,. Viral Infections exacerbate FUS-ALS Phenotypes in iPSC-Derived Spinal Neurons in a Virus Species=Specific Manner. Front. Cell. Neurosci. 2019 Oct 22:13:480

Acknowledgements

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

precisionlife
improving health | for everyone

For more information, visit:
www.precisionlife.com