






BRAIN COMMUNICATIONS

Evidence for *GRN* connecting multiple neurodegenerative diseases

Mike A. Nalls,^{1,2}  Cornelis Blauwendraat,¹  Lana Sargent,¹  Dan Vitale,^{1,2} Hampton Leonard,^{1,2,3} Hirotaka Iwaki,^{1,2} Yeajin Song,^{1,2} Sara Bandres-Ciga,¹  Kevin Menden,³  Faraz Faghri,^{1,2} Peter Heutink,³ Mark R. Cookson,¹ and Andrew B. Singleton¹

Previous research using genome-wide association studies has identified variants that may contribute to lifetime risk of multiple neurodegenerative diseases. However, whether there are common mechanisms that link neurodegenerative diseases is uncertain. Here, we focus on one gene, *GRN*, encoding progranulin, and the potential mechanistic interplay between genetic risk, gene expression in the brain and inflammation across multiple common neurodegenerative diseases. We utilized genome-wide association studies, expression quantitative trait locus mapping and Bayesian colocalization analyses to evaluate potential causal and mechanistic inferences. We integrate various molecular data types from public resources to infer disease connectivity and shared mechanisms using a data-driven process. Expression quantitative trait locus analyses combined with genome-wide association studies identified significant functional associations between increasing genetic risk in the *GRN* region and decreased expression of the gene in Parkinson's, Alzheimer's and amyotrophic lateral sclerosis. Additionally, colocalization analyses show a connection between blood-based inflammatory biomarkers relating to platelets and *GRN* expression in the frontal cortex. *GRN* expression mediates neuroinflammation function related to multiple neurodegenerative diseases. This analysis suggests shared mechanisms for Parkinson's, Alzheimer's and amyotrophic lateral sclerosis.

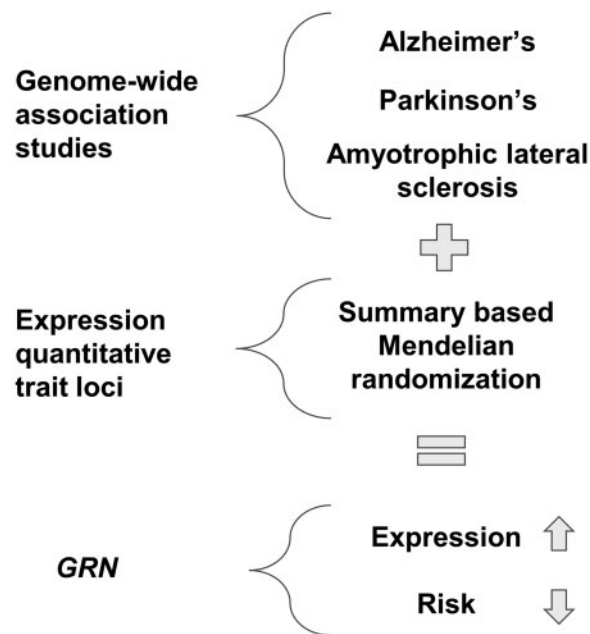
- 1 Center for Alzheimer's and Related Dementias, National Institutes of Health, Bethesda, MD 20892, USA
- 2 Data Tecnica International LLC, Glen Echo, MD 20812, USA
- 3 German Center for Neurodegenerative Diseases (DZNE), Tuebingen, Germany

Correspondence to: Mike A. Nalls,
PhD Center for Alzheimer's and Related Dementias,
National Institutes of Health,
Bethesda, MD 20892, USA
E-mail: nallsm@nih.gov

Keywords: Parkinson's; Alzheimer's; amyotrophic lateral sclerosis; gene expression; target

Abbreviations: AD = Alzheimer's disease; ALS = amyotrophic lateral sclerosis; FTD = frontotemporal dementia; GWAS = genome-wide association studies; MMPs = matrix metalloproteases; NDD = neurodegenerative disease; PD = Parkinson's disease; SNP = single-nucleotide polymorphism; SMR = summary data-based Mendelian randomization

Graphical Abstract



Introduction

Alzheimer's disease (AD), Parkinson's disease (PD) and amyotrophic lateral sclerosis (ALS) are thought to be three distinct disorders representative of the common neurodegenerative disease (NDD) spectrum, with other less common conditions manifesting as combinatorial versions of these symptomatologies. While common lifestyle and environmental factors have been suggested to contribute to all three disorders, well-powered genome-wide association studies (GWAS) generally indicate a distinct rather than shared set of chromosomal loci that contribute to lifetime disease risk.

Mutations in the *GRN* gene lead to reduced function or decreased expression of the encoded protein, progranulin, and an increased risk of frontotemporal dementia (FTD).^{1–5} The *GRN* locus has been nominated as contributing to the lifetime risk of several NDDs to varying extents.^{6–9} This locus is therefore a potential exception to the general rubric that NDDs have distinct genetic architecture.

Prior attempts to confirm single gene risks across NDDs have been limited to simple association studies and meta-analyses in small sample series.¹⁰ Here we apply post-GWAS methods to evaluate whether the *GRN* locus reliably contributes to risk of AD, PD and ALS in order to infer potential shared mechanistic consequences.

Methods

GWAS summary statistics were extracted from the three largest studies of AD, PD and ALS in the public domain for the chromosomal region surrounding *GRN*, defined as 500 kilobases from the 5' and 3' gene boundaries.^{6–8} These studies are in the public domain and include risk estimates derived from genotyping data in millions of genetic variants from over a million study participants spread across dozens of disease relevant cohorts. To test the specific hypothesis that decreased expression of *GRN* is associated with increased risk of NDDs, we located a large dataset for brain gene expression that includes a meta-analysis of brain-derived gene expression across multiple cohorts and brain regions, statistically accounting for sample overlap across studies.¹¹ This dataset is a meta-analysis of all publicly available brain-derived gene expression QTLs (at the time of publication for the initial report) that statistically account for sample overlap across studies and brain regions.

Statistical analysis

To generate functional inferences relating to the role of *GRN* in gene expression for these three archetypes of NDD, we used summary data-based Mendelian randomization (SMR).¹² SMR was carried out using default settings with the linkage disequilibrium reference dataset of

over 30 000 samples described previously.⁶ Initial analyses were limited to testing the association with *GRN* and the three diseases, and *post hoc* analyses were then applied to ascertain any other possible associations within the one megabase region surrounding the gene of interest. Associations were considered valid if *P* from the multi-SNP (single-nucleotide polymorphism) test in SMR was less than 0.01 after multiple test correction. This threshold denotes a significant association between local SNPs and changes in expression that are functionally connected to disease etiology. In this analysis, effect estimates relate to a putative causative effect of increasing genetic risk on gene expression, for example, a negative beta coefficient would indicate increasing risk is associated with decreased expression as part of disease etiology. When expanding past the *GRN* gene itself, we excluded SMR associations if the *P* for the heterogeneity in dependent instruments test was less than 0.01, which would suggest potential violations of the inherent assumptions of the causal inference model.

To fine-map signals, we ran the approximate Bayes factor fine-mapping routine within the R package ‘coloc’ for each trait separately using all *P*-values within the region of interest as an input.^{13–15} In this analysis, we test the possibility of a single causative variant at this locus being associated with disease risk. *P*-values are distributed across the locus and converted to approximate Bayes factors, then these Bayes factors are repeatedly sampled allowing for a posterior probability of a causative association to be generated per variant. We considered a valid result for fine mapping any variant meeting a posterior probability of >0.8. Null results for localizing a putative functional variant led us to not attempt any cross trait colocalization for these diseases. However, to gain further insight, we did mine aggregate colocalization data for *GRN* from the OpenTargets database (accessed 23 November 2020).¹⁶

Data and code availability

Available here (<https://drive.google.com/file/d/1VvYpnhK00-R4Ah9MROcusF1U0hgjdlfx/view?usp=sharing>).

Results

We first examined associations between genomic variation at *GRN* and all three diseases of interest. The strongest effect was with PD, where a single standard deviation increase above the population mean genetic risk for PD at this locus was associated with a 0.1650 standard deviation decrease in expression of *GRN* in the brain (SE = 0.0354, $P=1.67E-07$). A similar direction and magnitude of effect is seen in ALS (beta = -0.1038, SE 0.0411, $P=8.36E-03$), and a smaller but significant effect seen in AD (beta = -0.0230, SE = 0.0064, $P=1.53E-03$). The results are detailed in Table 1 and Fig. 1 below.

Fine-mapping and colocalization across these diseases at *GRN* failed to identify any single variant with posterior probabilities of causality greater than 0.80. Posterior probabilities for the null hypothesis, which relates to the likelihood of no quantifiable causal variant(s) neared this threshold, were 4.92E-06 in PD, 0.8743 in AD and 0.9142 in ALS. The top nominated SNP for PD (rs850738 in *FAM171A2*) only had a posterior probability of 0.1946 for being the causal variant, which is considered extremely weak evidence for causality. Therefore, no conclusive fine-mapping conclusion can be made from our GWAS-derived data. We also failed to support a causal role for the nearby gene *FAM171A2* based on expression quantitative trait locus data. Additionally, the proximal gene *ITGA2B* does not show up significantly in any of the SMR results or as a hit in the fine-mapping. This suggests that while *ITGA2B* may be related to platelet function, its direct impact on NDD is relatively uncertain.¹⁷ The lack of a single putative causal variant identified via colocalization does not invalidate the causal inferences from other analyses, as the putative causal variant may not be captured by the publicly available GWAS summary statistics.

Mining the OpenTargets database shows 55 associations suggesting colocalization evidence between GWAS studies and gene expression estimates at *GRN*. Out of these 55 putative causal associations, the overwhelming majority (80%) were related to platelet-derived GWAS, with an additional 4 of the 55 associations related to

Table 1 Summary of local SMR analyses using brain expression data and GWAS estimates for AD, ALS and PD

Symbol	Gene ID	Disease	N SNPs	Beta	SE	Multi-SNP P	HEIDI P
<i>GRN</i>	ENSG00000030582	PD	14	-0.1650	0.0354	1.67E-07	0.0471
<i>GJC1</i>	ENSG00000182963	PD	20	-0.1925	0.0512	1.09E-04	0.3489
<i>ASB16-AS1</i>	ENSG00000267080	PD	20	0.0805	0.0237	5.57E-04	0.0146
<i>HIGD1B</i>	ENSG00000131097	PD	20	0.1890	0.0579	1.10E-03	0.1037
<i>GRN</i>	ENSG00000030582	AD	18	-0.0230	0.0064	1.53E-03	0.7770
<i>GRN</i>	ENSG00000030582	ALS	18	-0.1038	0.0411	8.36E-03	0.9823

All *P*-values are unadjusted. N SNPs is the number of SNPs in the analysis after removing correlated SNPs. The beta and SE columns relate to the effect estimates of the SMR analysis, with a negative beta suggesting increasing disease risk resulting in decreased expression of the gene. The multi-SNP P and HEIDI P relate to the overall SMR association significance and the test of heterogeneity, respectively.

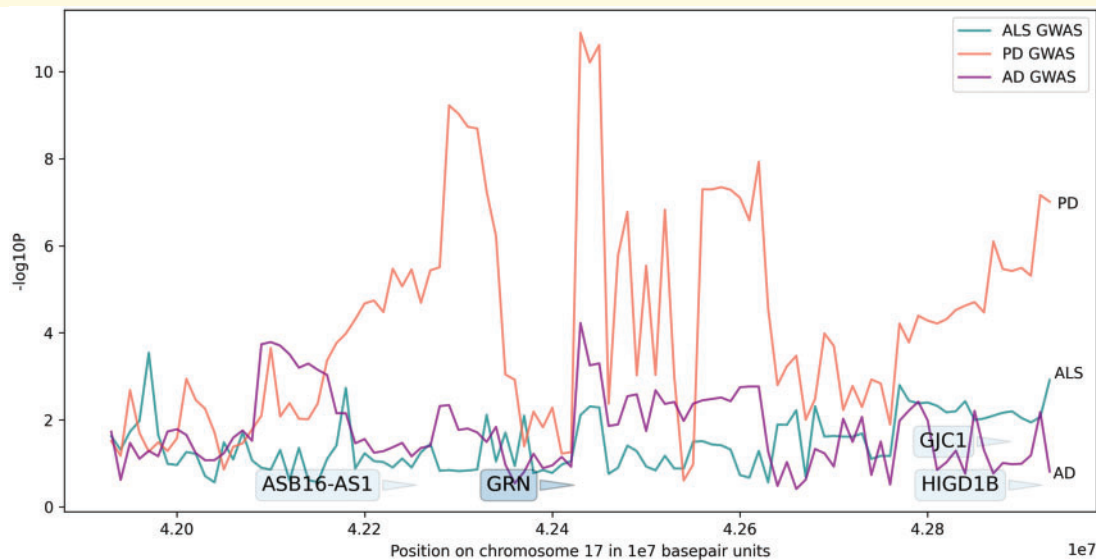


Figure 1 Graphical summary of SMR and GWAS analysis significance in the region per disease. To better display GWAS summary statistics, we only represent the most significant value per 10 kilobase window to avoid overplotting and improve clarity. The annotated boxes represent the SMR results for genes of interest after HEIDI filtering with some vertical spacing to avoid overplotting. The triangle next to the gene name indicates the start of the gene. The y-axis represents negative $\log_{10} P$ -values for each association, the x-axis represents positions on chromosome 17 (in 1E07 base pair units).

white blood cell GWAS. This is a major overrepresentation (44/55 associations) of blood and potential inflammatory traits compared to others at this locus. These inflammatory-related traits from blood (platelet distribution width, platelet count, mean platelet volume and thrombocyte volume) are all colocalized with variants related to changes in brain cortex expression.^{18–20}

Discussion

Here, we report strong causal functional inferences that connect genetic risk of three common NDDs to gene expression changes in the brain for *GRN*. This is the first study to link *GRN* to multiple NDDs and to suggest a gene-level mechanistic connection across a spectrum of NDDs. We additionally used a public resource, OpenTargets, to provide further mechanistic insights suggesting that the risk at this locus may be related to the body's immune response to injury through inflammation and potentially further brain-related etiological effects.

The association between *GRN* and FTD is well documented in the literature. Granulin encodes the progranulin protein which can be cleaved into 7.5 granulin domains. Whereas progranulin protein has anti-inflammatory properties, granulins display pro-inflammatory properties (reviewed in Doi: 10.1007/s12035-012-8380-8). A clear connection between decreased *GRN* expression and increased neuronal inflammation processes such as neutrophil migration and response to interleukins was seen in post mortem brain tissue

of *GRN* mediated FTD and not in *MAPT*-mediated FTD.²¹ Upregulated processes include NF-kappa-B signaling, as well as genes involved in tumor necrosis factor production. In addition, extracellular matrix pathways and matrix metalloproteases (MMPs) are up-regulated. Mediation of the inflammatory response involves proteolytic processing of anti-inflammatory *GRN* into pro-inflammatory granulins by the serine proteases neutrophil elastase, proteinase 3 and some MMPs (reviewed in Doi: 10.1007/s12035-012-8380-8). In mouse models of ALS inhibition of the MMPs, MMP2 and MMP9 could indeed prolong survival and reduce symptoms. This prior data provides evidence for the potential role of inflammatory mediated responses driven by *GRN* with shared etiology for multiple NDDs. Identification of common pathogenic mechanisms across the different NDDs could help explain shared clinical symptoms and establish more accurate genotype-phenotype correlations.

There are limitations to the current analyses that derive largely from data availability. As we used GWAS summary statistics, we did not have sufficient deep sequencing data for *GRN* in these diseases to accurately fine-map the locus. Additionally, due to sample size constraints, we were not able to include data from FTD subtypes or Lewy body dementia in our analyses. Finally, we do not have sufficient data to quantify the associations seen in diverse genetic ancestries but hope to evaluate this in the future. Finally, well-powered GWAS in pathology confirmed samples would benefit future research and allow us to exclude to a degree potential artifacts related to misdiagnosis of NDDs.

Acknowledgements

The funders of the study had no role in the study design, data collection, data analysis, data interpretation or writing of the report. All authors and the public can access all data and statistical programming code used in this project for the analyses and results generation. M.A.N. takes final responsibility for the decision to submit the paper for publication.

Funding

This work was supported in part by the Intramural Research Program of the National Institute on Aging and National Institute of Neurological Disorders and Stroke (project number Z01-AG000949-02).

Competing interests

M.A.N., H.L., D.V., H.I., and F.F. declare that they are consultants employed by Data Tecnica International, whose participation in this is part of a consulting agreement between the US National Institutes of Health and the said company.

References

1. Rademakers R, Baker M, Gass J, et al. Phenotypic variability associated with progranulin haploinsufficiency in patients with the common 1477C→T (Arg493X) mutation: An international initiative. *Lancet Neurol.* 2007;6(10):857–868.
2. Goedert M, Spillantini MG. Frontotemporal lobar degeneration through loss of progranulin function. *Brain.* 2006;129(Pt 11):2808–2810.
3. Cruts M, Van Broeckhoven C. Loss of progranulin function in frontotemporal lobar degeneration. *Trends Genet.* 2008;24(4):186–194.
4. Eriksen JL, Mackenzie IRA. Progranulin: Normal function and role in neurodegeneration. *J Neurochem.* 2008;104(2):287–297.
5. Le Ber I, van der Zee J, Hannequin D, et al.; French Research Network on FTD/FTD-MND. Progranulin null mutations in both sporadic and familial frontotemporal dementia. *Hum Mutat.* 2007;28(9):846–855.
6. Nalls MA, Blauwendraat C, Vallerga CL, et al.; International Parkinson's Disease Genomics Consortium. Identification of novel risk loci, causal insights, and heritable risk for Parkinson's disease: A meta-analysis of genome-wide association studies. *Lancet Neurol.* 2019;18(12):1091–1102.
7. Jansen IE, Savage JE, Watanabe K, et al. Genome-wide meta-analysis identifies new loci and functional pathways influencing Alzheimer's disease risk. *Nat Genet.* 2019;51(3):404–413.
8. Nicolas A, Kenna KP, Renton AE, et al. Genome-wide analyses identify KIF5A as a novel ALS gene. *Neuron.* 2018;97:1268–1283.e6.
9. Bellenguez C et al. Large meta-analysis of genome-wide association studies expands knowledge of the genetic etiology of Alzheimer's disease and highlights potential translational opportunities. *medRxiv* 2020.10.01.20200659 (2020).
10. Noyce AJ, Bandres-Ciga S, Kim J, et al.; 23andMe Research Team5, International Parkinson's Disease Genomics Consortium (IPDGC). The Parkinson's disease Mendelian randomization research portal. *Mov Disord.* 2019;34(12):1864–1872.
11. Qi T, Wu Y, Zeng J, et al.; eQTLGen Consortium. Identifying gene targets for brain-related traits using transcriptomic and methylation data from blood. *Nat Commun.* 2018;9(1):2282.
12. Zhu Z, Zhang F, Hu H, et al. Integration of summary data from GWAS and eQTL studies predicts complex trait gene targets. *Nat Genet.* 2016;48(5):481–487.
13. Plagnol V, Smyth DJ, Todd JA, Clayton DG. Statistical independence of the colocalized association signals for type 1 diabetes and RPS26 gene expression on chromosome 12q13. *Biostatistics.* 2009;10(2):327–334.
14. Wallace C. Statistical testing of shared genetic control for potentially related traits. *Genet Epidemiol.* 2013;37(8):802–813.
15. Giambartolomei C, Vukcevic D, Schadt EE, et al. Bayesian test for colocalisation between pairs of genetic association studies using summary statistics. *PLoS Genet.* 2014;10(5):e1004383.
16. Koscielny G, An P, Carvalho-Silva D, et al. Open Targets: A platform for therapeutic target identification and validation. *Nucleic Acids Res.* 2017;45(D1):D985–D994.
17. Kunicki TJ, Nugent DJ. The genetics of normal platelet reactivity. *Blood.* 2010;116(15):2627–2634.
18. UK Biobank—Neale lab. <http://www.nealelab.is/uk-biobank>. Accessed 23 December 2020.
19. Astle WJ, Elding H, Jiang T, et al. The allelic landscape of human blood cell trait variation and links to common complex disease. *Cell.* 2016;167(5):1415–1429.e19.
20. GTEx Consortium. The GTEx Consortium atlas of genetic regulatory effects across human tissues. *Science.* 2020;369:1318–1330.
21. Menden K, Francescato M, Nyima T, et al. Integrated multi-omics analysis reveals common and distinct dysregulated pathways for genetic subtypes of Frontotemporal Dementia. *Cold Spring Harbor Lab.* 2020;doi:10.1101/2020.12.01.405894.