

DEC 21, 2022

OPEN ACCESS

DOI:

[dx.doi.org/10.17504/protocols.io.3byl4br2zvo5/v1](https://doi.org/10.17504/protocols.io.3byl4br2zvo5/v1)

External link:

<https://doi.org/10.1093/brain/awac325>

Protocol Citation: Karishma D'Sa, Sebastian Guelfi, David Zhang, Alan Pittman, Daniah Trabzuni, Demis A. Kia, Nicholas W Wood, John Hardy, Claudia Manzoni, Mina Ryten 2022. Regulation of mitophagy by the NSL complex underlies genetic risk for Parkinson's disease: Bioinformatic Prioritisation and Hit Validation. **protocols.io** <https://dx.doi.org/10.17504/protocols.io.3byl4br2zvo5/v1>

MANUSCRIPT CITATION:

Soutar MPM, Melandri D, O'Callaghan B, Annuario E, Monaghan AE, Welsh NJ, D'Sa K, Guelfi S, Zhang D, Pittman A, Trabzuni D, Verboven AHA, Pan KS, Kia DA, Bictash M, Gandhi S, Houlden H, Cookson MR, Kasri NN, Wood NW, Singleton AB, Hardy J, Whiting PJ, Blauwendraat C, Whitworth AJ, Manzoni C, Ryten M, Lewis PA, Plun-Favreau H, Regulation of mitophagy by the NSL complex underlies genetic risk for Parkinson's disease at 16q11.2 and MAPT H1 loci. *Brain* 145(12). doi: [10.1093/brain/awac325](https://doi.org/10.1093/brain/awac325)

License: This is an open access protocol distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited

🌐 Regulation of mitophagy by the NSL complex underlies genetic risk for Parkinson's disease: Bioinformatic Prioritisation and Hit Validation

Karishma D'Sa¹, Sebastian Guelfi², David Zhang³, Alan Pittman⁴, Daniah Trabzuni³, Demis A. Kia³, Nicholas W Wood^{3,5}, John Hardy^{3,6,5}, Claudia Manzoni⁷, Mina Ryten^{2,5}

¹Francis Crick Institute, London, UK;

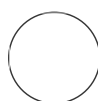
²UCL NIHR Great Ormond Street Hospital, London, UK;

³UCL Queen Square Institute of Neurology, London, UK;

⁴St Georges University, London, UK;

⁵Aligning Science Across Parkinson's (ASAP) Collaborative Research Network, Chevy Chase, MD, 20815.;

⁶UCL Dementia Research Institute, London, UK; ⁷UCL School of Pharmacy, London, UK



Benjamin O'Callaghan

ABSTRACT

This protocol describes the Bioinformatic Prioritisation of PD GWAS candidates for High Content Screening, and Hit Validation by allele-specific expression (ASE) analysis.

Protocol status: Working

Created: May 18, 2022

Last Modified: Dec 21, 2022

PROTOCOL integer ID:
62817

Keywords: High Content Screening, Weighted Protein-Protein Interaction Network Analysis (WPPINA), Parkinson's disease, GWAS, Putamen, Substantia Nigra, Allele-specific expression (ASE), ASAPCRN

Selection of genes for High Content Screening

1

Note

Candidates for High Content Screening were selected based on i) Weighted Protein-Protein Interaction Network Analysis (WPPINA); ii) complex prioritization; and, iii) coloc analysis.

2

WPPINA

WPPINA analysis is reported in [Ferrari et al., 2018](#) where the 2014 PD GWAS reported by [Nalls et al., 2014](#) was analysed; candidate genes were selected among those prioritised and with a linkage disequilibrium (LD) $r^2 \geq 0.8$.

Note

References:

Ferrari R, Kia DA, Tomkins JE, et al. Stratification of candidate genes for Parkinson's disease using weighted protein-protein interaction network analysis. *BMC Genomics*. 2018;19(1):452. doi:10.1186/s12864-018-4804-9 where the 2014 PD GWAS

Nalls MA, Pankratz N, Lill CM, et al. Large-scale meta-analysis of genome-wide association data identifies six new risk loci for Parkinson's disease. *Nat Genet*. 2014;46(9):989-993. doi:10.1038/ng.3043

3

Apply the same pipeline to the 2017 PD GWAS reported by [Chang et al., 2017](#), to update the list of candidate genes.

Note

Reference

Chang D, Nalls MA, Hallgrímsdóttir IB, et al. A meta-analysis of genome-wide association studies identifies 17 new Parkinson's disease risk loci. *Nat Genet.* 2017;49(10):1511-1516. doi:10.1038/ng.3955

4

Briefly, create a protein-protein interaction network based on the Mendelian genes for PD (seeds) using data from databases within the IMEx consortium (<https://www.imexconsortium.org/>).

5

Topologically analyse the network to extract the core network (i.e. the most interconnected part of the network).

This core network should contain the proteins/genes that can connect >60% of the initial seeds and are therefore considered relevant for sustaining communal processes and pathways, shared by the seeds.

6

Evaluate these processes with Gene Ontology Biological Processes enrichment analysis (<http://geneontology.org/>).

7

Use the top single nucleotide polymorphisms (SNPs) of the 2017 PD GWAS reported by [Chang et al., 2017](#), to extract open reading frames (ORFs) in cis-haplotypes defined by LD $r^2 \geq 0.8$.

Note

Reference

Chang D, Nalls MA, Hallgrímsdóttir IB, et al. A meta-analysis of genome-wide association studies identifies 17 new Parkinson's disease risk loci. *Nat Genet.* 2017;49(10):1511-1516. doi:10.1038/ng.3955

8

Match these ORFs to the core network to identify overlapping proteins/genes in relevant/shared pathways.

9

Complex Prioritisation

Obtain results of complex prioritization (neurocentric prioritization strategy) applied to the 2017 PD GWAS from the manuscript by [Chang et al., 2017](#).

Note

Reference

Chang D, Nalls MA, Hallgrímsdóttir IB, et al. A meta-analysis of genome-wide association studies identifies 17 new Parkinson's disease risk loci. *Nat Genet.* 2017;49(10):1511-1516. doi:10.1038/ng.3955

Coloc Analysis

Coloc analysis has been reported in the manuscript by [Kia et al., 2019](#).

Note

Reference:

Kia DA, Zhang D, Guelfi S, et al. Integration of eQTL and Parkinson's disease GWAS data implicates 11 disease genes. *bioRxiv*. Published online 2019:627216. doi:10.1101/627216

- 11 Calculate posterior probabilities for each gene for the hypothesis that both traits, the regulation of expression of a given gene and the risk for PD share a causal variant (PPH4).
Consider genes with $PPH4 \geq 0.75$ to have strong evidence for colocalization.
- 12 Obtain summary statistics from the most recent PD GWAS reported in the manuscript by [Nalls et al., 2019](#), and use these for regional association plotting using LocusZoom (<https://locuszoom.sph.umich.edu/>).

Note

References:

Nalls MA, Blauwendraat C, Vallerga CL, et al. 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. doi:10.1016/S1474-4422(19)30320-5

Pruim RJ, Welch RP, Sanna S, et al. LocusZoom: regional visualization of genome-wide association scan results. *Bioinformatics*. 2010;26(18):2336-2337. doi:10.1093/bioinformatics/btq419

ASEs (Sites of Allele-Specific Expression)

- 13 Identify sites of allele-specific expression (ASE) as described by [Guelfi et al, 2020](#) by mapping RNA-seq data to personalised genomes.
This approach is specifically chosen because it aims to minimise the impact of mapping biases.

Note

Reference

Guelfi S, D'Sa K, Botía JA, et al. Regulatory sites for splicing in human basal ganglia are enriched for disease-relevant information. *Nat Commun* 2020 11. 2020;11(1):1-16. doi:10.1038/s41467-020-14483-x

- 14 Use RNA-seq data generated from 49 putamen and 35 substantia nigra tissue samples from the UK Brain Expression Consortium (<https://ukbec.wordpress.com>) for this analysis.

Note

All samples were obtained from neuropathologically normal individuals of European descent and sites with greater than 15 reads in a sample were tested for ASE.

- 15 Only consider sites present in non-overlapping genes, and consider data from both the tissues together to increase power.
- 16 Mark ASE sites as sites with minimum false discovery rate (FDR) < 5% across samples.
- 17 Generate plots using Gviz3, with gene and transcript details obtained from Ensembl v92 (<https://www.ensembl.org/index.html>)