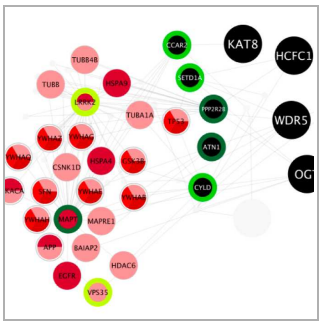

DEC 06, 2022

WORKS FOR ME 1

In Silico analysis links the NSL complex to Parkinson's disease and the mitochondria – Protein-protein interaction data to functional enrichment analysis

COMMENTS 0

DOI
dx.doi.org/10.17504/protocols.io.5qpvorb19v4o/v1

Katie Kelly[1,2], c.manzoni[3,2], Patrick Lewis[4,1,2], Helene Plun-Favreau[1,2]

[1]Department of Neurodegenerative Disease, UCL Institute of Neurology, Queen Square, London WC1N 3BG, UK;
[2]Aligning Science Across Parkinson's (ASAP) Collaborative Research Network, Chevy Chase, MD, 20 815, USA.;
[3]Department of Pharmacy, University of Reading, Reading, RG6 6AX, UK;
[4]The Royal Veterinary College, Royal College Street, London NW1 0TU, UK

Katie Kelly

ABSTRACT

Whilst the majority (~90-95%) of PD cases are sporadic, much of our understanding of the pathophysiological basis of disease can be traced back to the study of rare, monogenic forms of disease. However, in the past decade, the availability of Genome-Wide Association Studies (GWAS) has facilitated a shift in focus, toward identifying common risk variants conferring an increased risk of developing PD across the population.

A recently developed mitophagy screening assay of GWAS candidates, has functionally implicated the non-specific lethal (NSL) complex, a chromatin remodeler, in the regulation of PINK1-mitophagy. Here, a bioinformatics approach has been taken to investigate the interactome of the NSL complex, to unpick its relevance to PD progression. The mitochondrial interactome of the NSL complex has been built, mining 3 separate repositories: PINOT, HIPPIE and MIST, for curated, literature-derived protein-protein interaction (PPI) data. A multi-layered approach has been taken to; i) build the 'mitochondrial' NSL interactome, applying PD gene-set enrichment analysis to explore the relevance of the NSL mitochondrial interactome to PD and, ii) build the PD-oriented NSL interactome, using functional enrichment, to uncover biological pathways underpinning the NSL /PD association.

DOI

dx.doi.org/10.17504/protocols.io.5qpvorb19v4o/v1

EXTERNAL LINK

https://zenodo.org/record/7346957#.Y48jfYfP02x

PROTOCOL CITATION

Katie Kelly, c.manzoni, Patrick Lewis, Helene Plun-Favreau 2022. In Silico analysis links the NSL complex to Parkinson's disease and the mitochondria – Protein-protein interaction data to functional enrichment analysis. **protocols.io** https://dx.doi.org/10.17504/protocols.io.5qpvorb19v4o/v1

FUNDERS ACKNOWLEDGEMENT

The Michael J. Fox Foundation for Parkinson's Research (MJFF) and the Aligning Science Across Parkinson's (ASAP) initiative.
Grant ID: ASAP-000478

Masonic Charitable Foundation (PhD Prize Studentship)
Grant ID: -

KEYWORDS

Parkinson's disease, NSL complex, Mitophagy, In silico, Protein-protein interaction (PPI), Mito-CORE network Interactome

CREATED

Oct 26, 2022

LAST MODIFIED

Dec 06, 2022

OWNERSHIP HISTORY

Oct 26, 2022 Oke Avwenagha

Nov 14, 2022 Katie Kelly

PROTOCOL INTEGER ID

71840

Downloading the Protein-Protein Interaction (PPI) Data

1 All code can be found here : *v1.0.0_W-PPI-NA_NSL*
The pipeline to derive the ***first layer*** interactome can be found in Figure 1.

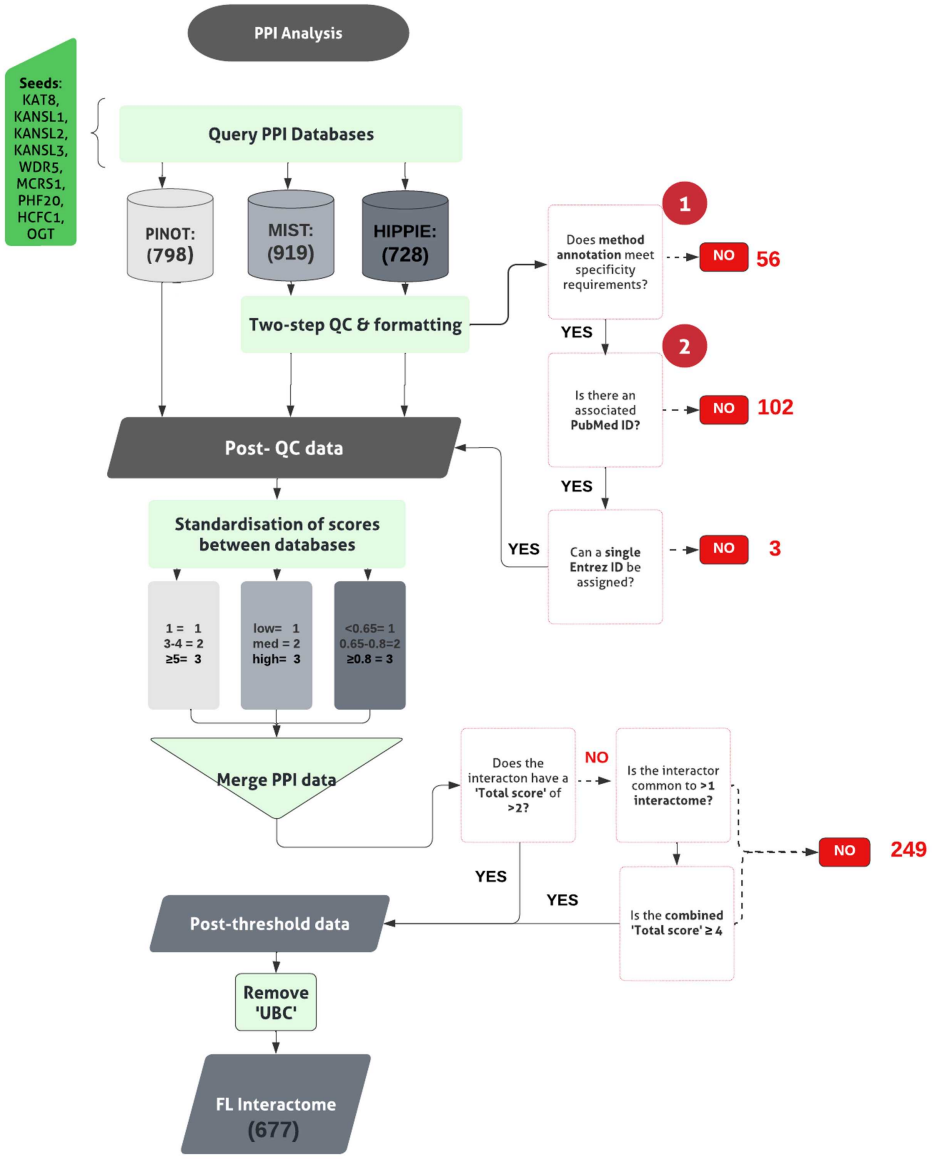

Figure 1. W-PPI-NA pipeline. Generating the first layer interactome of the NSL complex. The '*Seeds*' are the nine members of the NSL complex. Circled numbers (1 & 2) indicate the two stages of quality control (QC) applied. Numbers provided in brackets indicate total number of interactions retained at each stage.

2 Collect PPIs for NSL seeds using 3 different web-based tools;

1) PINOT (Version 1.1 with lenient filter option) (Protein Interaction Network Online Tool) (Tomkins, Ferrari et al. 2020, DOI: http://dx.doi.org/10.1186/s12964-020-00554-5)

2) HIPPIE with no threshold on interaction score (Human Integrated Protein-Protein Interaction rEference) (Alanis-Lobato, Andrade-Navarro et al. 2017 ; DOI: https://doi.org/10.1093/nar/gkw985; RRID:SCR_014651).

Citation: Katie Kelly, c.manzoni, Patrick Lewis, Helene Plun-Favreau In Silico analysis links the NSL complex to ParkinsonÃ¢ÂÂs disease and the mitochondria Ã¢ÂÂ Protein-protein interaction data to functional enrichment analysis https://dx.doi.org/10.17504/protocols.io.5qpvorb19v4o/v1

3) MIST v5.0 (Molecular Interaction Search Tool) (Hu, Vinayagam et al. 2018 ; DOI: 10.1093/nar/gkx1116).

Note

3 PPI data obtained using MIST and HIPPIE are subjected to quality control (QC), QC steps 1 & 2 (already integrated within the PINOT pipeline) to remove low quality data.

Note

Note

4 Formatting between the output files is standardized and interactors' IDs are converted to the approved EntrezID, UniprotID and HGNC gene name.

Note

5 Where 'UBC', a ubiquitin moiety, is identified as an interactor within the *first layer*, review the supporting publication.

Citation: Katie Kelly, c.manzoni, Patrick Lewis, Helene Plun-Favreau In Silico analysis links the NSL complex to ParkinsonÃ¢ÂÂs disease and the mitochondria Ã¢ÂÂ Protein-protein interaction data to functional enrichment analysis
https://dx.doi.org/10.17504/protocols.io.5qpvorb19v4o/v1

Note

6 Where ‘UBC’, a ubiquitin moiety, is identified as an interactor within the *first layer*, review the supporting publication.

Note

7 Merge interaction data, across the 3 databases to generate a single file for each seed’s interactome.

Note

Merging and Thresholding the PPIs

8 Calculate the total score (CS_T) for each interaction the (CS_T) was calculated as:

CS_T=CS_P+CS_M+CS_H

Note

Citation: Katie Kelly, c.manzoni, Patrick Lewis, Helene Plun-Favreau In Silico analysis links the NSL complex to ParkinsonÃ¢ÂÂs disease and the mitochondria Ã¢ÂÂ Protein-protein interaction data to functional enrichment analysis
https://dx.doi.org/10.17504/protocols.io.5qpvorb19v4o/v1

9

Apply an arbitrary score threshold (CS_T >2), to filter and remove lower confidence PPI data lacking reproducibility.

10 Merge interaction data, across the 3 databases to generate a single file for each seed's interactome. For each interactor the (CS_T) was calculated as:

Note

11 If interactions that failed to meet the threshold, interrogate further, to identify those interactors bridging >1 interactome.

12 For those interactors appearing within >1 interactome, apply a multi-interactome threshold represented by a $CS_T \geq 4$ across interactomes. Retain those meeting this multi-interactome threshold.

13 Combine all seed specific interaction lists, to obtain the *first layer interactome.*

14 Generate the list of unique interactors within the *first layer interactome* (code found in file *1.3. Standardisation of Score (GitHub)*).

Note

15 Where 'UBC', a ubiquitin moiety, is identified as an interactor within the *first layer*, review the supporting publication. Unless the interaction being studied is specific, remove.

Citation: Katie Kelly, c.manzoni, Patrick Lewis, Helene Plun-Favreau In Silico analysis links the NSL complex to ParkinsonÃ¢ÂÂs disease and the mitochondria Ã¢ÂÂ Protein-protein interaction data to functional enrichment analysis
https://dx.doi.org/10.17504/protocols.io.5qpvorb19v4o/v1

Note

Generating the Mito-CORE Network

16 The pipeline to derive the ***Mito-CORE network*** can be found in *Figure 2.*

Citation: Katie Kelly, c.manzoni, Patrick Lewis, Helene Plun-Favreau In Silico analysis links the NSL complex to ParkinsonÃ¢ÂÂs disease and the mitochondria Ã¢ÂÂ Protein-protein interaction data to functional enrichment analysis
https://dx.doi.org/10.17504/protocols.io.5qpvorb19v4o/v1

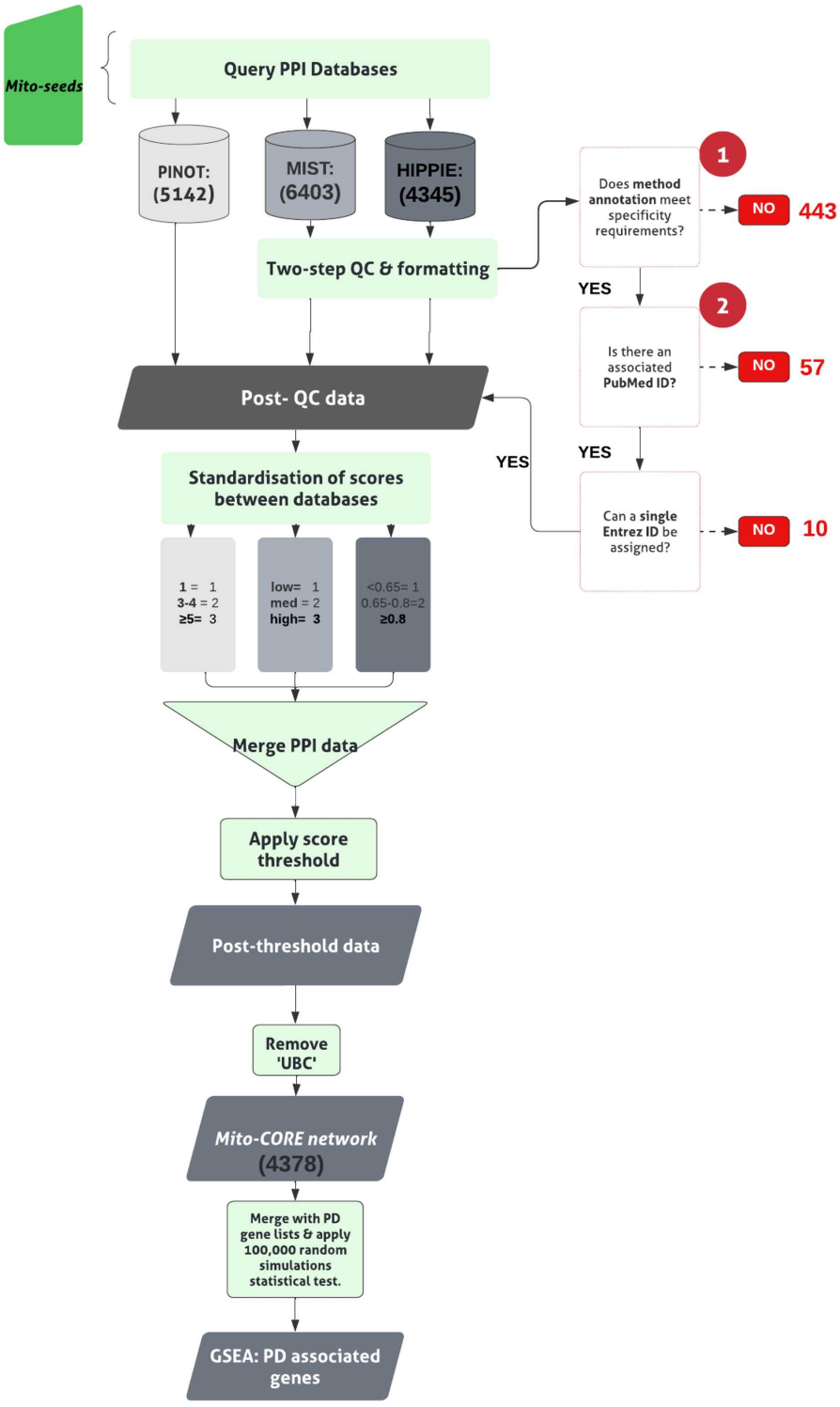

Figure 2. W-PPI-NA pipeline. Building the *Mito-CORE network*, and application of PD Gene-set enrichment analysis (GSEA). '*Mito-seeds*' refers to the mitochondrial first layer members of the NSL interactome. Circled numbers (1 & 2) indicate the two stages of quality control (QC) applied . Numbers provided in brackets indicate total number of interactions retained at each stage. * score threshold is applied as described in the pipeline in Figure 1, after the 'Merge PPI data' step.

Citation: Katie Kelly, c.manzoni, Patrick Lewis, Helene Plun-Favreau In Silico analysis links the NSL complex to ParkinsonÃ¢ÂÂs disease and the mitochondria Ã¢ÂÂ Protein-protein interaction data to functional enrichment analysis
https://dx.doi.org/10.17504/protocols.io.5qpvorb19v4o/v1

17 First, prioritise members of the *first layer* with mitochondrial annotation (- OGT, since it was a seed to derive the *first layer* interactome). Here, these are termed '*Mito seeds'.*

Note

18 Merge each list of mitochondrial proteins with the *first layer* interactome, to find overlaps. The overlaps represent members of the mitochondrial interactome for the NSL complex. (code found in file *1.5. Enrichment Analyses: Mitochondrial Proteins(GitHub)*).

19 Input *mito seeds* into all three PPI tools, to obtain the *second layer*. The *NSL seeds* together with the *Mito seeds*, and *second layer* interactors form the complete ***Mito-CORE network.***

Gene Set Enrichment Analysis (GSEA)

20 Conduct GSEA for PD associated genes by comparing the members of the interactome under investigation (*first layer* alone or *complete Mito-CORE network)* to a list of 180 unique PD associated genes;

Note

Citation: Katie Kelly, c.manzoni, Patrick Lewis, Helene Plun-Favreau In Silico analysis links the NSL complex to ParkinsonÃ¢ÂÂs disease and the mitochondria Ã¢ÂÂ Protein-protein interaction data to functional enrichment analysis
https://dx.doi.org/10.17504/protocols.io.5qpvorb19v4o/v1

21 Merge the list of 180 PD associated genes with the list of unique (*first layer / Mito-CORE network)* interactors, to find overlaps between the two lists. The overlaps represent PD associated proteins within the direct interactome/mitochondrial interactome for the NSL complex (code found in file *1.6. Enrichment Analyses: PD-associated genes (GitHub)*).

22 Repeat the above step with the list of 15 Mendelian PD genes, to ascertain enrichment of this more stringent list.

Note

Statistical Evaluation via Random Networks Simulation

23 Use an '100,000 random simulations' test of significance to validate statistical significance of overlaps of PD genes with the *first layer* and complete ***Mito-CORE network*** (code found in file *100,000 Random Simulations testing (GitHub)*).

Note

Generating the PD-CORE Network

Citation: Katie Kelly, c.manzoni, Patrick Lewis, Helene Plun-Favreau In Silico analysis links the NSL complex to ParkinsonÃ¢ÂÂs disease and the mitochondria Ã¢ÂÂ Protein-protein interaction data to functional enrichment analysis
https://dx.doi.org/10.17504/protocols.io.5qpvorb19v4o/v1

24 The pipeline to derive the ***PD-CORE network*** can be found in *Figure 3.*

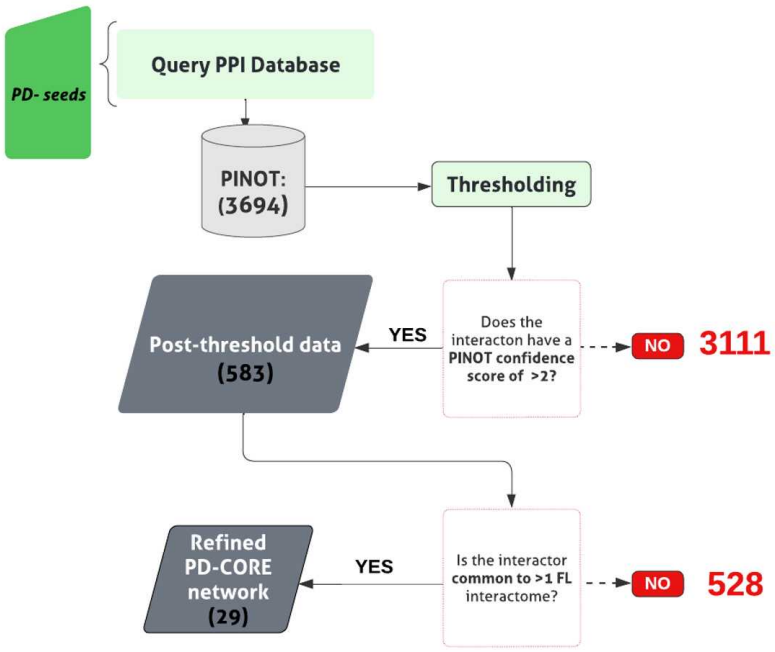

Figure 3. W-PPI-NA pipeline. '*PD-seeds*' refers to the PD associated *first layer* members. Numbers provided in brackets indicate total number of interactions/interactors retained at each stage.

25 Apply an arbitrary confidence threshold of '*CSp* >2', eliminating data with just a single publication and method from the downstream analysis (code found in file *1.7 Functional enrichment analysis (GitHub)*.

26 Once again, convert interactors' IDs to the approved EntrezID, UniprotID and HGNC gene name using the *Gene dictionary*.

27 Remove proteins with nonunivocal conversions to these 3 identifiers.

28 To remove background noise, keep only members of the *second layer* bridging >1 PD seed within the *PD-CORE* network.

Note

Citation: Katie Kelly, c.manzoni, Patrick Lewis, Helene Plun-Favreau In Silico analysis links the NSL complex to ParkinsonÃ¢ÂÂs disease and the mitochondria Ã¢ÂÂ Protein-protein interaction data to functional enrichment analysis
https://dx.doi.org/10.17504/protocols.io.5qpvorb19v4o/v1

29 The NSL seeds together with the PD seeds, and the non-private *second layer* interactors from the complete ***PD-CORE network***.

Functional Enrichment Analysis

30 The general pipeline for this analysis can be found in *Figure 4.*

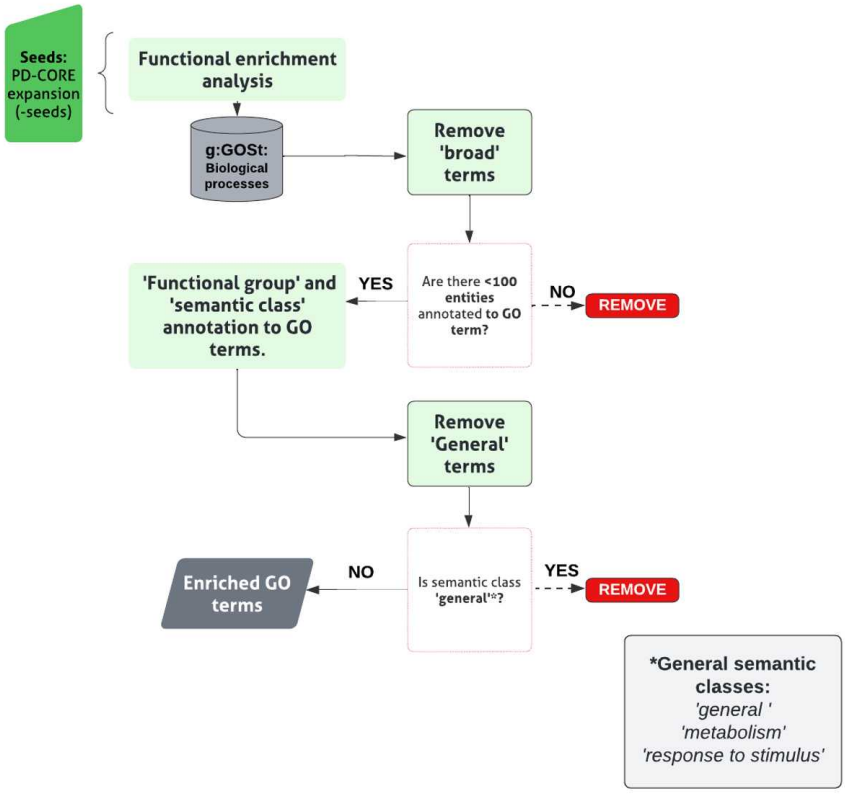

Figure 4. General pipeline for functional enrichment analysis. Grey box indicates Semantic Classes (SCs) removed from the analysis, as they are classified as 'general'.

31 Assess enrichment of particular biological processes within the *PD-CORE* network, members (- NSL seeds), by inputting into the g:Profiler search tool, g:GOSt (G:Profiler ; Ashburner, Ball et al. 2000, Gene Ontology 2021; RRID:SCR_006809).

32 Conduct enrichment for GO terms associated with 'Biological Processes (BPs)' only, with all other analysis

Citation: Katie Kelly, c.manzoni, Patrick Lewis, Helene Plun-Favreau In Silico analysis links the NSL complex to ParkinsonÃ¢ÂÂs disease and the mitochondria Ã¢ÂÂ Protein-protein interaction data to functional enrichment analysis
https://dx.doi.org/10.17504/protocols.io.5qpvorb19v4o/v1

settings left unadjusted, generating a list of enriched GO:BP terms.

33 Apply a threshold to the list of enriched GO:BP terms, to retain those with term size <100 thus effectively removing ‘broad’ GO:BP terms. (code found in file *1.7 Functional enrichment analysis (GitHub)*).

34 Assign remaining terms to custom-made ‘semantic classes’(SC), accompanied by a parent ‘functional group’(FG).

Note

35 Discard generic terms (classified in the semantic classes of: General , Metabolism, and Response to Stimulus) from further analysis.

36 Pool GO:BP terms contributing to each semantic class to identify the list of proteins within the network contributing to the enrichment of that specific semantic class.

Note

37 The final list of semantic classes, within each functional group represents those enriched within the network.

Citation: Katie Kelly, c.manzoni, Patrick Lewis, Helene Plun-Favreau In Silico analysis links the NSL complex to ParkinsonÃ¢ÂÂs disease and the mitochondria Ã¢ÂÂ Protein-protein interaction data to functional enrichment analysis
https://dx.doi.org/10.17504/protocols.io.5qpvorb19v4o/v1