Network based drug repurposing for Alzheimer’s disease

Attila Jones,..., Madhav Thambisetty

Clinical & Translational Neuroscience Section

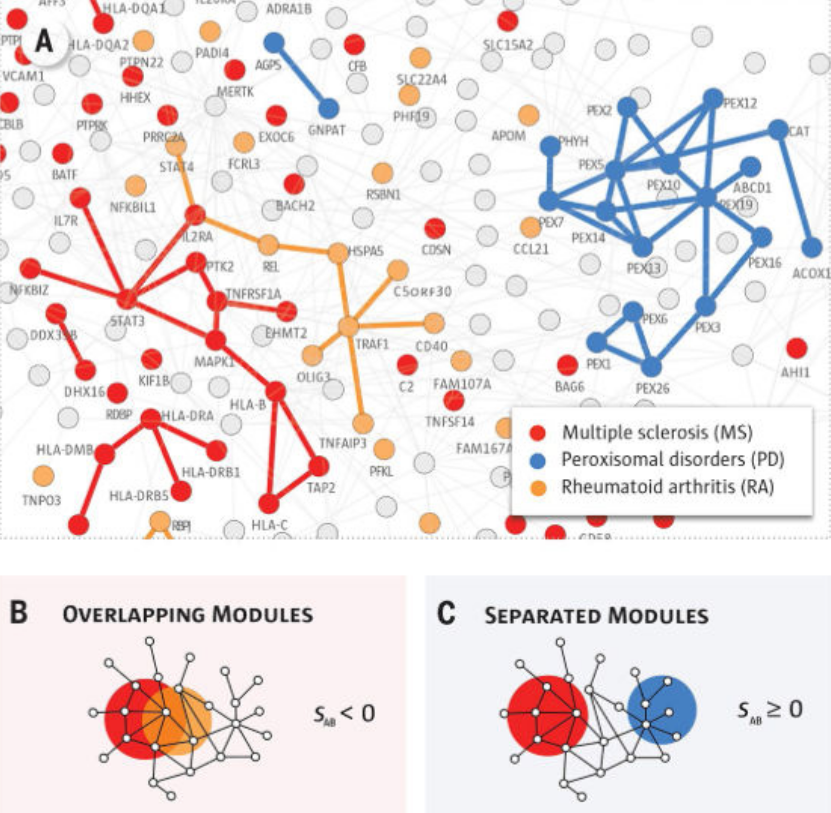
# Introduction

Late onset Alzheimer’s disease (AD) is a slowly progressing, highly heritable complex neurodegenerative disease, whose early and/or late pathological mechanisms are linked to immunity, APP and tau processing, and lipid metabolism (De Strooper and Karran 2016). To this date AD is without efficient treatment so we recently announced the Drug Repurposing for Effective Alzheimer’s Medicines (DREAM) study (Desai et al. 2020), in which we plan to find repurposable drugs based on proteomic, metabolomic and genomic findings by us (Roberts et al. 2021) and many others.

We will take the network approach to drug repurposing of (Cheng et al. 2018), which exploits the general observation that modules of patho-mechanistically related diseases are topologically close to each other in the human interactome i.e the human protein-protein interaction (PPI) network (Menche et al. 2015), Fig. [1](#fig:3modules). This approach identified new drug candidates for cardiovascular diseases that were validated by pharmacoepidemilogical procedures (Cheng et al. 2018).

Crucial to the applicability of the approach to AD is *a priori* knowledge of AD disease genes. However, relatively few AD disease genes (e.g APP, APOE) have been established by functional experiments, while GWAS on AD (Jansen et al. 2019; Kunkle et al. 2019) implicated 400 candidate AD genes in 29 genome-wide significant loci owing to the polygenic nature of AD and the limitations of GWAS in separating functional gene-disease associations from noise. Therefore we will extend the set of established AD genes with genes for which multiple types of evidence suggest involvement in AD pathomechanism. One such source is the *incipient AD proteomic signature*, which we recently discovered with differential proteomic analyses in young APOE4 carriers (Roberts et al. 2021), and which offers an opportunity for efficient, early intervention for AD. Another type of evidence is causality (also known as type II pleiotropy) between transcription level and AD found in statistically rigorous transcriptome-wide association studies (Baird et al. 2021; Gerring et al. 2020; Jansen et al. 2019; Kunkle et al. 2019).

What follows is an outline of our proposed study including data and knowledge sources, methods, tools, expected results and limitations.



Example for three disease modules (red, blue, gold) in the human interactome (Cheng et al. 2018). Successful drug repositioning to the red disease module is more likely from the overlapping gold disease module than from the separated blue module because drug targets in the gold module are topologically closer to and thus have larger impact on the red disease genes.

# Study Design

Fig. [1](#fig:3modules) illustrates disease modules in the human interactome, which for the basis of drug-disease networks (Fig. [3](#fig:drug-disease-net)) and drug repositioning opportunities. A module is a subnetwork defined by a set of disease genes (nodes) and their interactions (edges). (Cheng et al. 2018) constructed modules of cardiovascular diseases and searched for approved drug targets in “nearby” modules of non-cardiovascular diseases, where distance is defined in terms of network topology. We will apply this strategy to AD in the following steps:

1. Obtain a recent human interactome.
2. Collect AD genes. Since the set of AD genes is unknown we will construct it using Endeavour, a gene prioritization tool (Tranchevent et al. 2016) and two input gene sets (Fig. [2](#fig:endeavour)):
   1. High confidence AD genes (seed genes)
   2. Candidate AD genes
3. Construct a drug-AD network (Cheng et al. 2018)

## Obtain a recent human interactome

This step is not straight forward because PPI network databases are incomplete because only a fraction of binary protein-protein interactions have been discovered. (Cheng et al. 2018) used an interactome with 217161 interactions, whereas (Bai et al. 2020) used one with 469993 interactions. Also note that PPI are heterogeneous including binding, phosphorylation, metabolic and other type of interactions so that integration of multiple PPI databases might be necessary.

## Collect AD genes

We will infer the unknown set of AD genes with Endeavour from two gene sets: a small set of high confidence AD genes (seed genes, step 2 in Fig. [2](#fig:endeavour), also see section [2.2.1](#sec:AD-seed)) and a large set of candidate AD genes (step 4 in Fig. [2](#fig:endeavour), also see section [2.3](#sec:AD-candidate)). Endeavour will prioritize the candidate AD genes, which will allow us to include some of the top ranking genes in the AD gene set (of course we will include all the high confidence AD genes as well).

Endeavour uses multiple configurable data sources from public data and knowledge bases (Fig. [2](#fig:endeavour) step 3) when it builds a probabilistic model of a disease given a set of seed disease genes. One type of data source is PPI databases. We will carry out sensitivity analyses by varying seed AD genes and by including or excluding PPI data sources. The rational for the former is that AD is a complex disease that involves several biological processes and cell types (De Strooper and Karran 2016). The latter, on the other hand, will help reveal to what extent certain network topology quantities such as “compactness” of the inferred AD module are encoded in non-network data sources (bio-molecular pathways, chemical information, expression ontologies, expression profiles, gene and protein function, phenotypic information, and sequence based features (Tranchevent et al. 2016)).

### High confidence AD genes

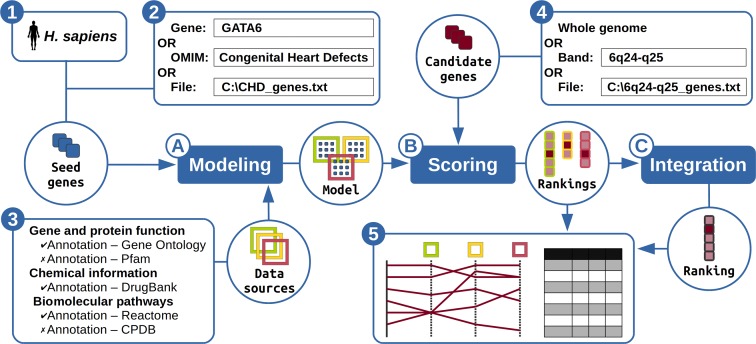
Multiple information sources and procedures will be used:

1. Biomedical literature
   1. Expert knowledge within our group
   2. Text mining with Beegle (ElShal et al. 2016)
2. The incipient AD proteomic signature (Roberts et al. 2021)
3. TWAS: strong statistical evidence for gene expression to be causal to AD
   1. Mendelian Randomization with *post-hoc* colocalization test (Baird et al. 2021; Kunkle et al. 2019)
   2. Transcriptomic imputation based and other TWAS with *post-hoc* colocalization test (Gerring et al. 2020; Jansen et al. 2019)
4. Additional gene mapping procedures like gene based GWAS, chromatin interaction mapping (Jansen et al. 2019), or “annotation and gene-based testing for deleterious coding, loss-of-function and splicing variants” (Kunkle et al. 2019)

We will examine the descriptive statistical relationships among these various information sources to see how they can be integrated to give rise to a single set of high confidence AD genes. If there are multiple plausible ways of integration we will carry out all and replicate our overall work for each of them.

## Candidate AD genes

We will consider 400 candidate AD genes at or near AD GWAS loci.



Gene prioritization with Endeavour (Tranchevent et al. 2016)

## Construct a drug-AD network

Once we infer AD genes they can mapped onto the human interactome to give rise to the AD module. Then we will annotate the AD module and its vicinity in the network with approved drugs to obtain the drug-disease network for AD (Fig. [3](#fig:drug-disease-net)). Computing the network proximity between each of those drugs and the AD module (Cheng et al. 2018) and other information (drugs specificity, target binding affinity,...) will guide us in prioritizing the drugs connected to the AD module.

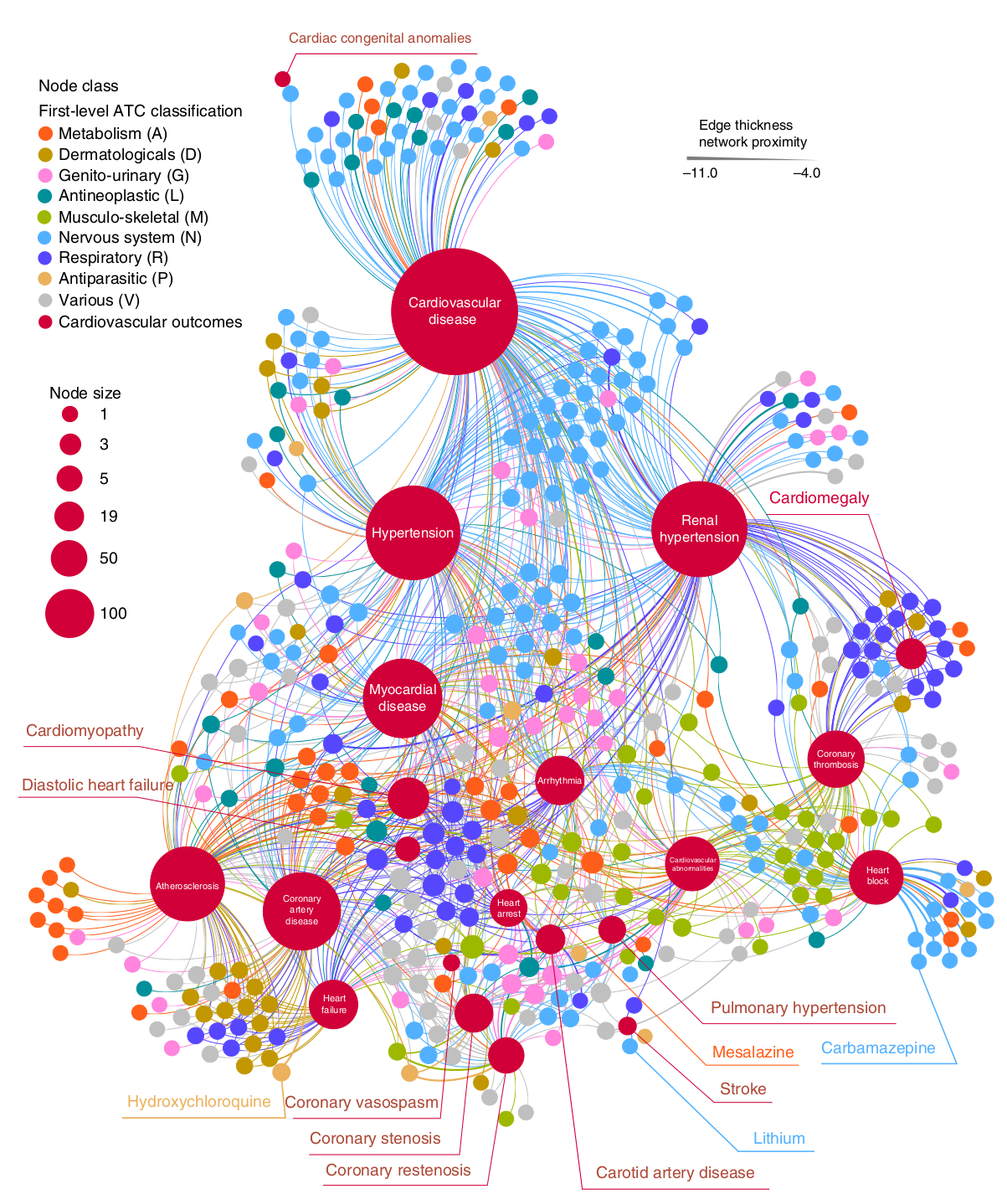
# Expected results

The main goal of the proposed work is a list of drugs prioritized according to their expected therapeutic value for AD. Note that such expectation is difficult to quantify due to limitations of the approach (see next section).

An important intermediate result will be the AD module in the human interactome. Extending this with other disease modules will allow discovering diseases directly, that is: pathomechanistically, related to AD. Comorbidities of AD have been extensively researched to gain insights to the etiology and treatment of AD (Santiago and Potashkin 2021). GWAS have opened up genetic analyses like Mendelian randomization or LDSC regression to quantify pairwise relationships between AD and putatively related traits (Jansen et al. 2019; Kunkle et al. 2019). The interpretation of such pairwise AD-disease relations obtained from genetics, however, is less direct than that of AD-disease relations derived from PPI network based disease modules (Menche et al. 2015). One reason is that proteins, in general, are more directly involved in disease etiology and pathology than their encoding genes. Another is that AD-disease relations form disease *network* (Menche et al. 2015) together with the fact that inferring disease networks from pairwise statistics (such as pairwise genetic correlations) leads to a disease network with many false (“mediated”) edges while network based inference helps eliminate false edges (Marx et al. 2017). With our PPI based Alzheimer’s and other disease modules we will be able to evade the pitfalls of both genetics and pairwise inference.

# Limitations of the approach

There are several sources of uncertainty in the proposed approach, which pertain to either the edges or the nodes of the interactome and disease modules. However, the approach of (Cheng et al. 2018) models disease modules deterministically. This means that it can weight neither edges nore nodes according to the quality and quantity of available evidence. As for edges it means that PPIs supported by relatively weak evidence must either be completely ignored or considered as equally plausible to PPIs supported by strong evidence. Similarly, deterministic treatment of nodes means that low confidence AD genes must be either excluded from the AD module or, similarly to high confidence AD genes, included in it. (This is equivalent to “cutting off” the Endeavour-prioritized candidate AD genes at some arbitrary rank or p-value threshold.) Consequently the deterministic treatment of PPI and AD genes introduces substantial errors that propagate ultimately into the drug-AD network while remaining unquantified. The lack of error quantification (estimation), then, hinders efficient prioritization of repurposing drug candidates.



Drug-disease network for cardiovascular diseases (Cheng et al. 2018). The medium-large red nodes represent cardiovascular disease while the small, non-red, nodes represent drugs approved for non-cardiovascular indications. An edge between a cardiovascular disease and a drug expresses PPI-mediated connection between the drug’s target and the cardiovascular disease’s network module. Thicker edges mean that fewer PPI link a given drug to a given disease hence the repurposed drug is more likely to be effective.

Bai, Bing, Xusheng Wang, Yuxin Li, Ping-Chung Chen, Kaiwen Yu, Kaushik Kumar Dey, Jay M. Yarbro, et al. 2020. “Deep Multilayer Brain Proteomics Identifies Molecular Networks in Alzheimer’s Disease Progression.” *Neuron* 105 (31926610): 975–991.e7. <https://doi.org/10.1016/j.neuron.2019.12.015>.

Baird, Denis A., Jimmy Z. Liu, Jie Zheng, Solveig K. Sieberts, Thanneer Perumal, Benjamin Elsworth, Tom G. Richardson, et al. 2021. “Identifying Drug Targets for Neurological and Psychiatric Disease via Genetics and the Brain Transcriptome.” *PLoS Genetics* 17 (33417599): e1009224–e1009224. <https://doi.org/10.1371/journal.pgen.1009224>.

Cheng, Feixiong, Rishi J. Desai, Diane E. Handy, Ruisheng Wang, Sebastian Schneeweiss, Albert-László Barabási, and Joseph Loscalzo. 2018. “Network-Based Approach to Prediction and Population-Based Validation of in Silico Drug Repurposing.” *Nature Communications* 9 (30002366): 2691–1. <https://doi.org/10.1038/s41467-018-05116-5>.

Desai, Rishi J., Vijay R. Varma, Tobias Gerhard, Jodi Segal, Mufaddal Mahesri, Kristyn Chin, Edward Nonnenmacher, et al. 2020. “Targeting Abnormal Metabolism in Alzheimer’s Disease: The Drug Repurposing for Effective Alzheimer’s Medicines (Dream) Study.” *Alzheimer’s & Dementia (New York, N. Y.)* 6 (33304987): e12095–e12095. <https://doi.org/10.1002/trc2.12095>.

De Strooper, Bart, and Eric Karran. 2016. “The Cellular Phase of Alzheimer’s Disease.” *Cell* 164 (4): 603–15.

ElShal, Sarah, Léon-Charles Tranchevent, Alejandro Sifrim, Amin Ardeshirdavani, Jesse Davis, and Yves Moreau. 2016. “Beegle: From Literature Mining to Disease-Gene Discovery.” *Nucleic Acids Research* 44 (26384564): e18–e18. <https://doi.org/10.1093/nar/gkv905>.

Gerring, Zachary F., Michelle K. Lupton, Daniel Edey, Eric R. Gamazon, and Eske M. Derks. 2020. “An Analysis of Genetically Regulated Gene Expression Across Multiple Tissues Implicates Novel Gene Candidates in Alzheimer’s Disease.” *Alzheimer’s Research & Therapy* 12 (32299494): 43–43. <https://doi.org/10.1186/s13195-020-00611-8>.

Jansen, Iris E., Jeanne E. Savage, Kyoko Watanabe, Julien Bryois, Dylan M. Williams, Stacy Steinberg, Julia Sealock, et al. 2019. “Genome-Wide Meta-Analysis Identifies New Loci and Functional Pathways Influencing Alzheimer’s Disease Risk.” *Nature Genetics* 51 (30617256): 404–13. <https://doi.org/10.1038/s41588-018-0311-9>.

Kunkle, Brian W., Benjamin Grenier-Boley, Rebecca Sims, Joshua C. Bis, Vincent Damotte, Adam C. Naj, Anne Boland, et al. 2019. “Genetic Meta-Analysis of Diagnosed Alzheimer’s Disease Identifies New Risk Loci and Implicates Aβ, Tau, Immunity and Lipid Processing.” *Nature Genetics* 51 (30820047): 414–30. <https://doi.org/10.1038/s41588-019-0358-2>.

Marx, Peter, Peter Antal, Bence Bolgar, Gyorgy Bagdy, Bill Deakin, and Gabriella Juhasz. 2017. “Comorbidities in the diseasome are more apparent than real: What Bayesian filtering reveals about the comorbidities of depression.” Edited by Lilia M. Iakoucheva. *PLOS Computational Biology* 13 (6): e1005487. <https://doi.org/10.1371/journal.pcbi.1005487>.

Menche, Jörg, Amitabh Sharma, Maksim Kitsak, Susan Dina Ghiassian, Marc Vidal, Joseph Loscalzo, and Albert-László Barabási. 2015. “Disease Networks. Uncovering Disease-Disease Relationships Through the Incomplete Interactome.” *Science (New York, N.Y.)* 347 (25700523): 1257601–1. <https://doi.org/10.1126/science.1257601>.

Roberts, Jackson A., Vijay R. Varma, Yang An, Sudhir Varma, Julián Candia, Giovanna Fantoni, Vinod Tiwari, et al. 2021. “A Brain Proteomic Signature of Incipient Alzheimer’s Disease in Young Apoe ε4 Carriers.” *Submitted*.

Santiago, Jose A., and Judith A. Potashkin. 2021. “The Impact of Disease Comorbidities in Alzheimer’s Disease.” *Frontiers in Aging Neuroscience* 13 (33643025): 631770–0. <https://doi.org/10.3389/fnagi.2021.631770>.

Tranchevent, Léon-Charles, Amin Ardeshirdavani, Sarah ElShal, Daniel Alcaide, Jan Aerts, Didier Auboeuf, and Yves Moreau. 2016. “Candidate Gene Prioritization with Endeavour.” *Nucleic Acids Res* 44 (W1): W117–W121. <https://doi.org/10.1093/nar/gkw365>.