Queen Mary University of London

William Harvey Research Institute

Population-level multi-omics across 68 age-related diseases reveals novel shared genomic and proteomic architecture

Krishna Amin[1], Pilar Cacheiro[1], James O'Reilly[2], and Damian Smedley[1]

[1]PhenoGenomics Group, Centre for Clinical Pharmacology and Precision Medicine, William Harvey Research Institute, Queen Mary University of London, London, United Kingdom, EC1M 6BQ. [2]Recursion, The HKX Building, 3 Pancras Square, London N1C 4AG.

INTRODUCTION

- **Age-related diseases (ARDs)** may share common mechanisms, indicated by their shared onset at older age, and their chronic accumulation in individuals.
- Investigating common mechanisms of these complex diseases invites simultaneously analysing **multi-omics data of multiple diseases**.
- Human-based work covering multi-omics or many diseases has been mainly: (1) single-disease, multi-omic[1,2] (2) multi-disease, single-omic[3,4]; leaving a gap for multi-disease, multi-omic.
- Investigating ARD multimorbidity necessitates **longitudinal human-based** data with phenotypes covering **multiple ARDs in a single cohort**.

AIMS

- Our work examines **genomic and proteomic data in 68 ARDs in one cohort at a biobank scale** (423,223 individuals for genomics; 39,337 for proteomics).
- We aim to explore **common mechanisms between sets of ARDs**: what they tell us about ageing and ARDs mechanistically; how they relate to inherent ageing processes; how we can use them for disease risk prediction and therapeutic development.
- Here, we present findings suggesting **similarities between Parkinson's disease (PD; G20), osteoarthritis (OA; M15-17), retinal disorders (RD; H25-26, H35), and interstitial pulmonary disease (IPD; J84)**, a set of diseases with little / no previous relationships.

METHODS

Data used

Genomics: genotyped + imputed
Proteomics: Olink blood proteome (2920 proteins)

Disease definition

Common: < 1:199 case:control. Range: 1:5 – 1:182
Unisex: > 1/1000 males and females
Age-related: Age-of-onset > 50, for ≥ 90% participants
➔ **68 ARDs across 16 ICD-10 chapters**

Association analyses

Genomics: SAIGE GWAS
Proteomics: Firth (prevalent)

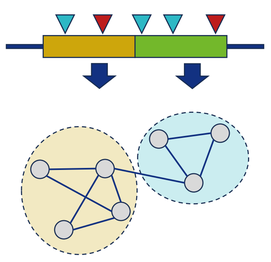
Enrichment analyses

Proteomics: ORA with Reactome

Commonality analyses

Sharing of LD regions with associated variants between diseases

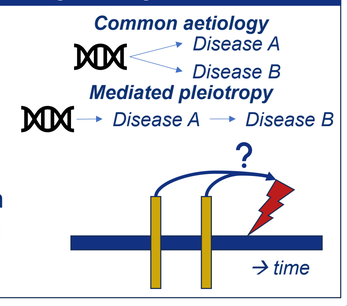
Disease clustering using semantic similarity of high-effect annotations



Causality analyses

Testing for mediated pleiotropy with LCV[5]

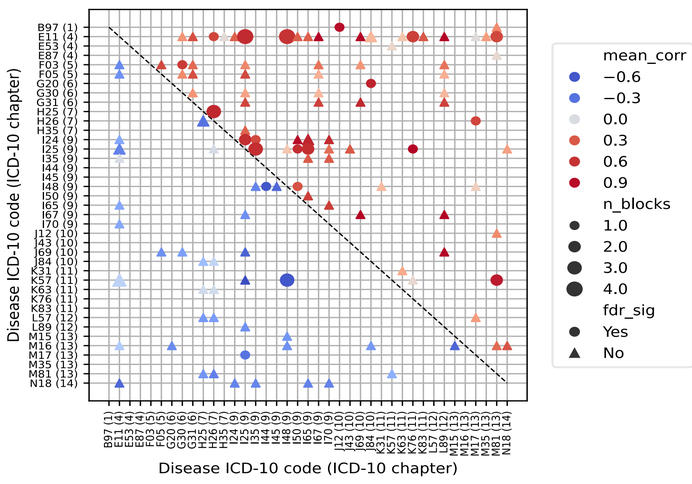
Controlling for prior disease in proteomics Firth regression



RESULTS

Parkinson's disease (PD), osteoarthritis (OA), retinal disorders (RD) and interstitial pulmonary disease (IPD) share multi-level biological similarities

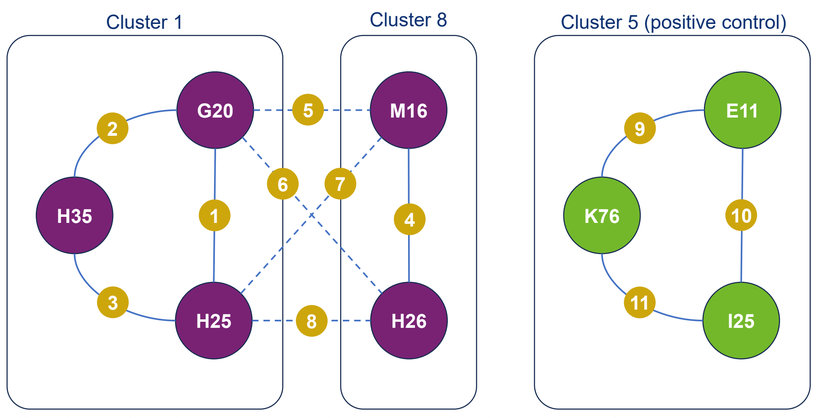
PD, OA, RD and IPD share disease-relevant LD-independent genomic regions (LD blocks), with correlation of effects observed for some disease sets



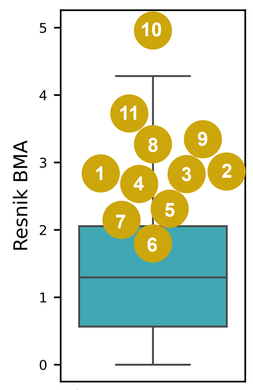
- PD and IPD: sig. pos. correlation, on region with *MAPT, KANSL1, WNT3.*
- RD and OA: sig. pos. correlation, on region with *CLEC18A, WWP2.*
- OA and PD, IPD: LD block sharing, with insig neg. correlation, on region with *MAPT, KANSL1, WNT3.*
- Circulatory diseases (I24-70): LD block sharing, pos. correlation ➔ pos. control.

Spearman correlations of LD blocks shared between disease pairs, averaged over the blocks which show pos. or neg. correlation separately. fdr_sig: FDR < 0.05. n_blocks: number of shared LD blocks. Above the diagonal: pos. correlation. Below the diagonal: neg. correlation. Pos. = positive. Neg. = negative. Sig. = significant (FDR < 0.05). Insig. = insignificant (FDR ≥ 0.05).

Louvain clustering groups ARDs based on the semantic similarity of their enriched Reactome annotations



Selected sections from Louvain clusters using best-match averaged Resnik between diseases' top 20 Reactome annotations by odds ratio. Expected groupings act as a positive control. G20: PD. H25-26: cataracts. H35: 'other retinal disorders'. M16: hip OA. E11: Type 2 diabetes. K76: 'other liver disease'. I25: coronary artery disease.



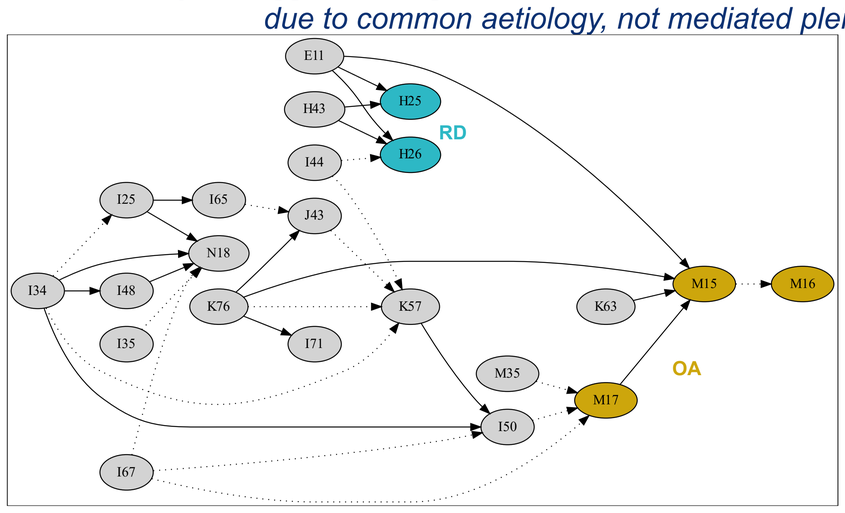
Semantic similarity distribution, with pairwise values superimposed.

Selected pathway detail
Link **(1)** involves different receptor-mediated signalling cascades, TGF-β/BMP signalling, and extracellular matrix (ECM) organisation.

- Shown within-cluster links are in the top 25% of all links, confirming clustering validity.

Observed biological similarities between the four diseases are unlikely due to causal links between diseases

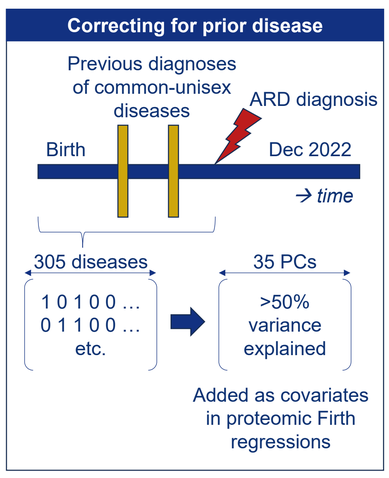
LCV analysis suggests that genetic similarities between PD, OA, RD and IPD are due to common aetiology, not mediated pleiotropy

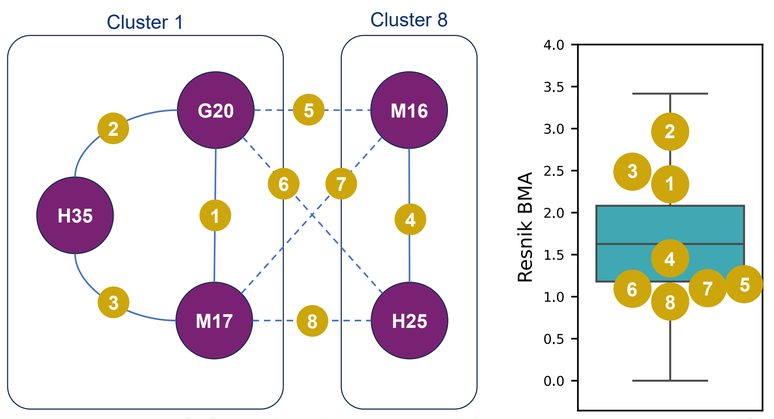


- LCV identifies likely instances of mediated pleiotropy.
- No LCV-derived links exist between PD, OA, RD and IPD ➔ mediated pleiotropy unlikely.

LCV links defined as pairs with FDR < 0.05, mean absolute GCP > 0.6. Solid lines: link direction concordant with the significant (FDR < 0.05) 'alternative' of a Mann-Whitney U rank sum test comparing age-of-onset distributions of those two diseases. Dotted lines: link direction discordant with distribution test. Blue: RD. Gold: OA.

Protein pathway-based biological commonality of PD, OA, RD and IPD remains after correcting for prior disease





As opposite, on LHS. G20: PD. H25: cataracts. H35: 'other retinal disorders'. M16: hip OA. M17: knee OA.

As opposite, on LHS.

Selected pathway detail
Link **(2)** maintains ECM involvement from LHS Link **(1)**.
Links **(1)** and **(2)** share pathway themes in glycan metabolism defects and host defense activation.

- Correcting for prior disease reduces out-of-cluster links, while maintaining within-cluster links.

CONCLUSION

- LD block sharing implies common genetic architecture, with correlation hinting at shared effects.
- For example, PD, IPD and OA share an LD block which contains *MAPT, KANSL1 and WNT3*, but with varying effect correlations. The genes may affect PD and IPD similarly, but OA differently.
- The absence of mediated pleiotropy, and the robustness of the Louvain clusters to controlling for prior disease, suggest observed similarities are likely to be genuine and independent.
- Pathway commonalities indicate potentially shared mechanisms, e.g. ECM organisation in PD and RD; glycan metabolism defects and host defense activation in PD and OA.

LIMITATIONS AND FUTURE DIRECTIONS

- LCV analysis is limited to the 68 ARDs, missing out diseases that may be biologically relevant.
- Data on only 2920 proteins inherently limits what can be captured at the proteomic level.
- Sex specificity, key for ARDs like post-menopausal conditions, is not considered.
- Ethnic diversity is required for representative results. Only Europeans were used here.
- There may be additional relevant lifestyle factors that should have been controlled for.
- Future work could involve the integration of multi-omics at a lower level, e.g. MOFA[6] / multi-modal networks, and the prediction of age-related 'hub' pathways, proteins and protein modules.

AIDD CTP

Recursion®

UKRI Biotechnology and Biological Sciences Research Council

References, contacts, and more… →