Project title:

Modelling multi-omic effects in Parkinson's Disease to validate putative novel PD genes and identify potential biomarkers of disease

Supervision:

Oxford: Prof Jenny Taylor (primary supervisor) and Dr Dianne Newbury (day-to-day supervision)
GSK contact - Robert Lai

Summary:

Parkinson's Disease (PD) has a clear genetic component which may be monogenic, oligogenic or complex (Blauwendraat *et al.* 2019). A handful of genes that cause monogenic forms of PD have been identified allowing a better understanding of the biological aetiology of risk and the identification of underlying pathways (including autophagy-lysosomal networks, membrane trafficking, immune response, synaptic vesicles and microtubule function) (e.g. Masaracchia *et al.* 2018, Fernandez-Mosquera *et al.* 2025). However, the majority of PD cases exhibit incomplete penetrance and are expected to have a complex genetic risk mechanism that includes many genes, common and rare variants that affect multiple processes at different time points and environmental factors (Nalls *et al.* 2019). This makes the identification of universal biomarkers difficult (Pitz *et al.* 2024).

This project proposal aims to cut across these genetic classifications to provide a holistic overview of genetic contributions to PD through **multi-omic analyses of large datasets**. These analyses will focus on 700 unsolved patients in the Genomics England (GEL) 100,000 Genomes Project (100kGP) who are affected with Early Onset PD (mean onset age = 46 years). The proposal builds on a previous GSK-funded project in which *genomic* analyses of the PD cohort in 100kGP, were undertaken, leading to identification of ~70 putative novel candidate PD genes.

Since that time, a range of *other 'omics* data have been made available in 100kGP. 138 PD patients have matched transcriptomic (Illumina), proteomic (SOMAscan/ SOMAlogic), long-range sequencing (ONT), metabolomics (Metabolon) and epigenomic data. **This unique resource will enable us to evaluate the clinical potential of multi-omic analyses and the value added by such characterisation**. Multi-omic modelling can allow the detection of interactive effects within and between molecular levels (Rohart *et al.*, 2017). This latter point is of critical importance in PD where it is likely that combinations of variants will ultimately explain risk. Understanding these combinations will afford better target and biomarker identification through the identification of potential mechanisms for drug intervention.

GEL data is linked to hospital episode statistics (HES), including in/outpatient visits, allowing some assessment of clinical severity. To further investigate links with **clinical progression and disease severity**, the project will utilise PD data from biobanks (e.g. UK-Biobank, All-of-Us), which are likely to include more complex cases of genetic risk, and from clinical cohorts (e.g. Accelerating Medicines Partnership (AMP), Parkinson's Progression Markers Initiative (PPMI).

All individuals across the databases will also be classified in terms of **polygenic risk quartiles** to identify individuals with a high polygenic risk of PD (Kim et al 2024 and Nalls et al 2019, which explains 16-36% of the variance). Patterns of non-polygenic components and environmental exposures will be modelled across high-risk individuals who manifest PD and those who do not to identify risk and protective features. This last step will allow us to move beyond single gene disorders and understand the continuum of PD moving towards individual risk models both in terms of risk and

resilience. Ultimately, this will be essential for the prediction of progression and the identification of biomarkers of wider relevance.

Aims and Objectives

1. Undertake uni-omics analyses of the PD cohort in 100kGP
2. Undertake integrated multi-omics analysis of GEL PD cohort to
    i. identify potential biomarkers linked to PD and
    ii. provide functional assessment of 70 candidate genes previously identified
3. Investigate clinical correlations between 'omics results, polygenic risk profiles and disease severity and progression using a range of PD datasets eg 100kGP, All-of-Us, AMP, PPMI

Alignment with therapeutic area and key scientific theme(s):

- Using data from human samples, identify potential mechanisms for drug intervention that differ based on disease stage.
- Discovery and validation of biomarkers (from human samples) reflecting various pathophysiologies that can be used to predict patient phenotype and track response to therapy.

Project delivery:

Experimental Methodology

Uni-omic analyses will be performed using gold-standard pipelines available in GEL. Integrated multi-omics will use a combination of statistical, machine learning, and tensor decomposition approaches building on the approach taken for multi-omics analysis recently used for the covid COMBAT study (COvid-19 Multi-omics Blood ATlas (COMBAT) Consortium, 2022). Pathway and network enrichment analyses will be applied to interpret the biological relevance of integrative components, and visualization tools developed to facilitate exploration of the multi-omics data.
We envisage that each aim of the project will take 1 year to complete, creating distinct milestones for the Fellow to work to.

Training for Fellow

The fellow will be provided with extensive training in data handling, bioinformatics and linkage to clinical datasets relevant to the project, including multi-omics.
This will enable them to develop an understanding of the different 'omics datasets and their relationship to disease. The Fellow's clinical background will help relate clinical symptoms to underlying biological mechanisms and the identification of potential alternative diagnoses.

The fellow will also model interactions between the genetic risk background (polygenic risk), higher-effect rare variants and demographic and clinical elements to understand how these factors come together to influence progression in PD. The project enables the unique resource of multi-omics data available in GEL to interrogated and presents opportunities to apply the expertise gained to other disease projects of interest to GSK.

Research environment:

The fellow will be embedded within the CHG's Translational Genomics Group, University of Oxford which specialises in genomics analysis of WGS and GEL data. We are collaborating with Prof Heather Harrington (Department of Mathematics) on multi-omics analysis and with Prof Ira Milosevic on functional validation of selected PD targets. Access agreements for GEL are already in place and analyses are currently free to user. We will apply for access to the other cohorts as the project progresses as project agreements are often time-limited in nature.

References:

Blauwendraat C, Nalls MA, Singleton AB. The genetic architecture of Parkinson's disease. Lancet Neurol. 2020 Feb;19(2):170-178. doi: 10.1016/S1474-4422(19)30287-X. Epub 2019 Sep 11. PMID: 31521533; PMCID: PMC8972299.

COvid-19 Multi-omics Blood ATlas (COMBAT) Consortium. A blood atlas of COVID-19 defines hallmarks of disease severity and specificity. Cell. 2022 Mar 3;185(5):916-938.e58. doi: 10.1016/j.cell.2022.01.012. Epub 2022 Jan 21. PMID: 35216673; PMCID: PMC8776501.

Fernandez-Mosquera L, Yambire KF, Couto R, Pereyra L, Pabis K, Ponsford AH, Diogo CV, Stagi M, Milosevic I, Raimundo N. Mitochondrial respiratory chain deficiency inhibits lysosomal hydrolysis. Autophagy. 2019 Sep;15(9):1572-1591. doi: 10.1080/15548627.2019.1586256. Epub 2019 Mar 27. PMID: 30917721; PMCID: PMC6693470.

Kim JJ, Vitale D, Otani DV, Lian MM, Heilbron K; 23andMe Research Team; Iwaki H, Lake J, Solsberg CW, Leonard H, Makarious MB, Tan EK, Singleton AB, Bandres-Ciga S, Noyce AJ; Global Parkinson's Genetics Program (GP2); Blauwendraat C, Nalls MA, Foo JN, Mata I. Multi-ancestry genome-wide association meta-analysis of Parkinson's disease. Nat Genet. 2024 Jan;56(1):27-36. doi: 10.1038/s41588-023-01584-8. Epub 2023 Dec 28. PMID: 38155330; PMCID: PMC10786718.

Masaracchia C, Hnida M, Gerhardt E, Lopes da Fonseca T, Villar-Pique A, Branco T, Stahlberg MA, Dean C, Fernández CO, Milosevic I, Outeiro TF. Membrane binding, internalization, and sorting of alpha-synuclein in the cell. Acta Neuropathol Commun. 2018 Aug 14;6(1):79. doi: 10.1186/s40478-018-0578-1. PMID: 30107856; PMCID: PMC6090819.

Nalls MA, Blauwendraat C, Vallerga CL, Heilbron K, Bandres-Ciga S, Chang D, Tan M, Kia DA, Noyce AJ, Xue A, Bras J, Young E, von Coelln R, Simón-Sánchez J, Schulte C, Sharma M, Krohn L, Pihlstrøm L, Siitonen A, Iwaki H, Leonard H, Faghri F, Gibbs JR, Hernandez DG, Scholz SW, Botia JA, Martinez M, Corvol JC, Lesage S, Jankovic J, Shulman LM, Sutherland M, Tienari P, Majamaa K, Toft M, Andreassen OA, Bangale T, Brice A, Yang J, Gan-Or Z, Gasser T, Heutink P, Shulman JM, Wood NW, Hinds DA, Hardy JA, Morris HR, Gratten J, Visscher PM, Graham RR, Singleton AB; 23andMe Research Team; System Genomics of Parkinson's Disease Consortium; 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 Dec;18(12):1091-1102. doi: 10.1016/S1474-4422(19)30320-5. PMID: 31701892; PMCID: PMC8422160.

Pitz V, Makarious MB, Bandres-Ciga S, Iwaki H; 23andMe Research Team; Singleton AB, Nalls M, Heilbron K, Blauwendraat C. Analysis of rare Parkinson's disease variants in millions of people. NPJ Parkinsons Dis. 2024 Jan 8;10(1):11. doi: 10.1038/s41531-023-00608-8. PMID: 38191580; PMCID: PMC10774311.

Rohart F, Gautier B, Singh A, Lê Cao KA. (2017) mixOmics: An R package for 'omics feature selection and multiple data integration. *PLoS Comput Biol*. 2017 Nov 3;13(11):e1005752. doi: 10.1371/journal.pcbi.1005752.