WP 14A

Biostatistics – genetics
Start date: 1 November 2014. **Completion date: 31 December 2018**

Overall work package objectives:
To develop and apply state-of-the-art stratification methods to DPUK cohorts. It will also provide exemplar statistical analyses to test and demonstrate the utility of the informatics portal for integrated analyses.

1. Develop genetic risk stratification analyses for AD using polygenic score and other analytical techniques, through collation of existing GWAS and new replication data. These analyses will be extended to other forms of dementia and neurodegeneration to include PD, MND and others.
2. Provide exemplar statistical analyses to test and demonstrate the utility of the informatics portal for integrated (across cohort) analyses. Analyses will include linking research results data to routinely collected data within cohorts and data linking individuals between cohorts.
3. To create a results database focusing on genetics, genomics and associated data.

Deliverables	Milestones	Milestone deadline	Work package dependencies	Person(s) responsible
Objective 1:				
D1.1 Report on questionnaire	M1.1.1 List of genetics information issued to WP1 for questionnaire	M1.1.1 Complete	To and from WP1	Georgina Menzies
D1.2 Assess availability of genetic data	M1.2.1 Availability of genetic data from DPUK cohort scoped	M1.2.1 Complete		
D1.3 Polygenic score (PS) for AD	M1.3.1 Develop genetic risk stratification strategy	M1.3.1 Complete	None	Georgina Menzies
	M1.3.2 Validations of AD PS using currently available data	M1.3.2 Complete		
	M1.3.3 Publish AD PS paper	M1.3.3 Complete		
D1.4 Polygenic score (PS) for dementia and neurodegeneration	M1.4.1 Extend PS to Parkinson's Disease	M1.4.1 Complete		
	M1.4.2 Extend PS to other diseases	M1.4.2 Complete		
	M1.4.3 Use AD PS in other sub phenotypes	M1.4.3 Complete		
Objective 2:				
D2.1 Polygenic score analysis strategy developed to model the genetic component to GxE analyses	M2.1.1 Exemplar cross cohort statistical analyses identified	M2.1.1 Complete	To and from WP1	Georgina Menzies
D2.2 First paper submitted to peer review journal	M2.2.1 First exemplar analyses completed	M2.2.1 Complete	None	
D2.3 Second paper submitted to peer review journal	M2.3.1 Second analyses completed: genetic and epidemiology	M2.3.1 Complete		
D2.4 Third paper submitted to peer review journal	M2.4.1 Third analyses completed: genetic and epidemiology	M2.4.1 Complete		
D2.5 Statistical support programme delivered	D2.5.1 Statistical support programme in place	M2.5.1 Complete		
Objective 3:				
	M3.1.1 Conceptual design of database in place	M3.1.1 Complete	None	

D3.1 Create and build database through external contractor	M3.1.2 External supplier appointed	M3.1.2 Complete		Georgina Menzies
	M3.1.3 Create and build database through external contractor	M3.1.3 Complete		
D3.2 Data uploaded, tested, interrogated and reporting available	M3.2.1 Collect genetic, genomic, and associated data from DPUK and other worldwide collaborators	M3.2.1 Complete		

Key updates on delivery against milestones since last report

We have made excellent progresses with all milestones, producing more papers than anticipated.

Summary of plan to deliver on outstanding work (with dates)

- Current focus is on understanding age ranges that are most influenced by polygenic risk score. We anticipate submitting a paper by the end of Dec 19. Initial analysis shows polygenic scores influences the risk of AD beyond 80, which is also beyond the APOE influence risk. We feel this is a very interesting observation that will help cohorts utilise data in a variety of further designs.
- We will continue to support and add data to genetics and genomics databases with DPUK until Dec 2019 when project formally closes.

Risks	**Mitigation**
1) N/A	1)

Team members <u>funded</u> (full or part-time) by DPUK

Julie Williams, Valentina Escott-Price, Georgina Menzies, Janet Harwood, Ganna Leonenko, Detelina Grozeva (previously also Elisa Majounie, Christian Bannister)

Team members involved with the project but <u>not</u> funded by DPUK

Rebecca Sims, Catherine Bresner, Matthew Bareford, Ewen Sommerville, Salha Saad

ECR's: Georgina Menzies, Emily Baker, Detelina Grozeva, Ganna Leonenko

Outcomes

- Escott-Price, V.et al. 2015. Common polygenic variation enhances risk prediction for Alzheimer's disease. *Brain* 138(12), pp. 3673-3684. (10.1093/brain/awv268)

This publication was the first to show polygenic risk in Alzheimer's Disease. Results showed polygenic risk score (PRS) could correctly classify cases and controls 78% of the time. This is something which has now been validated in a number of independent datasets and direct many future research studies.

- Escott-Price, V.et al. 2015. Polygenic risk of Parkinson disease is correlated with disease age at onset. *Annals of Neurology* 77(4), pp. 582-91. (10.1002/ana.24335)

This method of calculating PRS was successfully extended to a number of disorders, including this significant publication which correlated PRS with age-of-onset in Parkinson's Disease.

- Leonenko G, Sims R, Shoai M, et al. Polygenic risk and hazard scores for Alzheimer's disease prediction. *Ann Clin Transl Neurol*. 2019;6(3):456–465. 2019(1) 18. doi:10.1002/acn3.716

This recently submitted evaluation of PRS pits this method against other published statistical methods in the field showing polygenic risk to be the best current tool to evaluate a number of outcomes relating to the subjects' genetic architecture. Additionally, this paper calculates polygenic risk across a number of cohorts.

- Leonenko G, Shoai M, Bellou E, Sims R, Williams J, Hardy J, Escott-Price V (2019) (2) Genetic risk for alzheimer disease is distinct from genetic risk for amyloid deposition Ann Neurol. 2019 Sep;86(3):427-435. doi: 10.1002/ana.25530. Epub 2019 Jul 1.

In this study we tested the prediction accuracy of AD, MCI and amyloid deposition risks with polygenic risk score (PRS) and analysed the most up-to-date biological pathways in the ADNI cohort. Results suggest that APOE mostly contributes for amyloid accumulation and the AD PRS affects risk of further conversion to dementia.

- Baker E, Schmidt KM, Sims R, et al. POLARIS: Polygenic LD-adjusted risk score approach for set-based analysis of GWAS data. *Genet Epidemiol*. 2018;42(4):366–377. doi:10.1002/gepi.22117

Through this project, we continually strive to improve all of our statistical algorithms. One such improvement is the POLARIS algorithm for calculating PRS. This improvement automatically adjusts for linkage disequilibrium and allows set-based analysis of genetic-wide association studies (GWAS) data.

- Grozeva, Detelina, Saad, Salha, Menzies, Georgina E. and Sims, Rebecca 2019. Benefits and challenges of rare genetic variation in Alzheimer's disease. *Current Genetic Medicine Reports* 7 (1), pp. 53-62.10.1007/s40142-019-0161-5

We reviewed the current understanding of the genetic architecture of Alzheimer's disease (AD), focusing on rare susceptibility variants. We synthesised diverse results and provided other scholars with an up-to-date snapshot of the domain, along with offering future research directions in the important emerging field of rare variation linked to AD. Such a review paper of the most current state of the rare genetics field in AD was lacking. Scientists often find such papers very useful and tend to widely cite them in any future manuscripts.

- Guerreiro, R.et al. 2016. Genome-wide analysis of genetic correlation in dementia with Lewy bodies, Parkinson's and Alzheimer's diseases. *Neurobiology of Aging* 38, article number: 214.e7–214.e10. (10.1016/j.neurobiolaging.2015.10.028) (reported in May 2018 GM)
- Escott-Price, V.et al. 2018. Polygenic risk for schizophrenia and season of birth within the UK Biobank cohort. *Psychological Medicine* (10.1017/S0033291718000454) (reported in May 2018 GM)
- Ahmad, S.et al. 2018. Disentangling the biological pathways involved in early features of Alzheimer's disease in the Rotterdam Study. *Alzheimer's and Dementia* (10.1016/j.jalz.2018.01.005)
- Morgan, A.et al. 2017. The correlation between inflammatory biomarkers and polygenic risk score in Alzheimer's Disease. Journal of Alzheimer's Disease 56(1), pp. 25-36. (10.3233/JAD-160889)

Presentations

Emily Baker, PhD Student (Supervisor – Prof Valentina Escott-Price)

- Poster at AAIC Conference – July 2017
- Poster at ASHG Conference – October 2017

Georgina Menzies - ECR

- DPUK Genetics Platform presentation – AAIC July 2017
- Presentations at cohort workshops – May and September 2017
- Presentations at cohort workshops – December 2017 and February 2018
- Presentation at the MRC Oversight Board in London

Detelina Grozeva - ECR

- Poster presentation at European Society of Human Genetics Conference 2019

DPUK Study Requests – Cardiff

- DPUK Study Number 0234 (Glasgow):
- DPUK Study Number 0246 (Oxford)
- Identifying predictors of reversion across the dementia spectrum
- DPUK Study Number 0257
- Stratified medicine approaches to drug repurposing for Alzheimer's disease

Project narrative

The genetic architecture of non-familial forms of dementia is complex and has been difficult to capture, at the genome-wide level, in a form useable to the health sciences. WP14A has developed a polygenic risk score (PRS) algorithm for Alzheimer's disease (AD) that provides an individual-specific genetic risk score. The PRS model has been validated in neuropathologically confirmed cohorts, with up to 97% prediction accuracy at the extremes of polygenic risk (+/- 1.5 standard deviations from the distribution mean), making this the best PRS model presently available. WP14A has established PRS algorithms able to predict age at disease onset and specific to the underlying biology of disease (i.e. Immunity, Endocytosis, Cholesterol Metabolism, Protein Ubiquitination, Abeta processing and Tau). The availability of this genetic risk profile now allows the creation of complex stem cell models, which truly reflect disease activity in common forms of dementia, and participant selection for

neuroimaging, clinical trial and testing of public health risk reduction strategies. Based on the success of the AD PRS model, large-scale projects are planned with Genomics England and UKDRI to study individuals at polygenic extremes.

The data portal, complete with genome browser, provides a simple, but powerful, secure online workspace for searching and identifying genes, single nucleotide polymorphisms (SNPs), and genomic locations of interest or with special relevance to dementia. It is a secure online resource that will provide interactive search interfaces for the genomics results data, where users can easily combine search results by adding steps to an interactive workflow. The portal integrates summary data from DPUK cohorts and from relevant non-DPUK studies for the purposes of aggregate data meta-analysis and informing the potential value of individual participant data meta-analyses. This will enhance the power of DPUK to work with international collaborations to test for genetic and environmental interactions in dementia by integrating multiple cohort studies from around the world. These results of these analyses will be stored, logged, documented, and made accessible to the dementia research community.