Biostatistics – genetics
Start date: 1 November 2014. **Completion date: April 2020**

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:				
D3.1 Create and build database through external contractor	M3.1.1 Conceptual design of database in place	M3.1.1 Complete	None	Georgina Menzies
	M3.1.2 External supplier appointed	M3.1.2 Complete		
	M3.1.3 Create and build database through external contractor	M3.1.3 Complete		

WP 14a Project report

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

Location(s):

Cardiff University

Lessons Learnt

Data sharing is essential for bioinformatics analyses. The current challenges with data sharing delaying the research (data analysis and methodology development) in the AD and dementia field.

Please tell us the most successful outcome and what it means to dementia research

Development and validation of AD polygenic risk scoring. Genomic profiling will become pivotal in selecting therapies based on individual risk. To achieve high diagnostic accuracy, polygenic risk scoring should be linked with biomarkers, clinical and environmental information and tested for accuracy in groups of people targeting disease specific functional mechanisms.

Outcomes

PUBLICATIONS

- 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, 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, et al. (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, 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, (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)

Under review

- V Escott-Price et al. (2020) **Unified mathematical framework for LD-informed polygenic risk scores.** *Frontiers of Genetics*.
- E Bellou, et al. (2020) **Polygenic risk and pleiotropy in neurodegenerative diseases.** *Neurobiology of Disease.*
- E Bellou, (2020) **Age-dependent effect of APOE and polygenic component on Alzheimers' disease.** *Neurobiology of Aging.*

ENGAGEMENT ACTIVITIES

Presentations

- Emily Baker, et al. (2015.) Common polygenic variation enhances risk prediction for Alzheimer's disease. *Brain* 138(12), pp. 3673-3684. (10.1093/brain/awv268)
- 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)
- Leonenko G, et al. (2019) 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
- Leonenko G, et al. (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.
- Baker E, 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

- Grozeva D, (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
- 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)
- Leonenko G, et al. (2019) Genetic risk for Alzheimer's disease and for amyloid deposition is distinct. Annals of Neurology (doi:10.1002/ana.25530)
- Kunkle et al (2019) Meta-analysis of genetic association with diagnosed Alzheimer's disease identifies novel risk loci and implicates Abeta, Tau, immunity and lipid processing. Nature genetics, 51:414–430
- Richards AL, et al. (2019) The relationship between polygenic risk scores and cognition in schizophrenia. Schizophrenia Bulletin. 10.1093/schbul/sbz061
- Hardy J and Escott-Price V (2019) Genes, pathways and risk prediction in Alzheimer's disease. Human Molecular Genetics. 10.1093/hmg/ddz163
- Guerreiro R, et al (2019) Heritability and genetic variance of dementia with Lewy bodies (2019) Neurobiology of Disease. 10.1016/j.nbd.2019.04.004
- Baker E, et al. (2019) Gene-Based Analysis in HRC Imputed Genome Wide Association Data Identifies Three Novel Genes for Alzheimer's Disease. PloS One (doi.org/10.1371/journal.pone.0218111)
- Leonenko G, et al. (2019) Polygenic Risk and Hazard Scores for Alzheimer's disease prediction. Annals of Clinical and Translational Neurology (doi.org/10.1002/acn3.716
- Escott-Price V, et al. (2019) Genetic analysis suggests high misassignment rates in clinical Alzheimer's cases and controls. Neurobiology of Aging; 77:178-182.
- Escott-Price V, et al. (2019) Polygenic Risk Score Analysis of Alzheimer's Disease in Cases without APOE4 or APOE2 Alleles. Journal of Prevention of Alzheimer's Disease dx.doi.org/10.14283/jpad.2018.46
- J Harrison, et al. (2020) From polygenic scores to precision medicine in Alzheimer's Disease: A systematic review. *Journal of Alzheimer's Disease* (in press)

Scientific:

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

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

Ioanna Katzourou, PhD Student (Supervisor – Prof Valentina Escott-Price)

- Abstract at AAIG Conference – July 2019
- Abstract at AAIG Conference – July 2020

Karen Crawford, PhD Student (Supervisor – Prof Valentina Escott-Price)

- Abstract at AAIG Conference – July 2019

Thomas Rowe, PhD Student (Supervisor – Prof Valentina Escott-Price)

- Abstract at AAIG Conference – July 2020

Emily Baker – ECR

- Abstract at ESHG Conference – June 2019

- Abstract at AAIG Conference – July 2020

Eftychia Bellou – ECR

- Abstract at AAIG Conference – July 2019
- Abstract at ICAPD Conference – April 2020
- Abstract at AAIG Conference – July 2020

Ganna Leonenko – ECR

- Abstract at ASHG Conference – October 2018
- Abstract at ASHG Conference – October 2019
- Abstract at ESHG Conference – June 2020

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

Public:

2016

- May 2016 Dementia week, seminar series
- June 2016 Sci Screen presentation Cardiff University
- June 2016 Science Café talk
- Mobile roadshow with interactive showcase space for lived experience and discussion. e.g. Delegates use the dementia goggles in a Virtual Reality system that makes you feel like you are very old. Our field team run the event, delivering memory tests, tests of spatial awareness, and recruit participants. CADR with MRC CNGG/NCMH at Wales Gene Park's Genetics and Genomics Conference (22nd June, 2016).
- Used mobile roadshow to deliver interactive space, as above. British Science Festival 10th & 11th September 11am-4pm
- Used mobile roadshow to deliver interactive space, as above. Young Onset Dementia: Different Impact, Positive Solutions conference 27th September
- Field Team delivered interactive presentation directly to public. Public response to scientific discovery
- CADR presentation to Minister who is visiting Swansea University 15th August
- CADR annual public conference 5th October 2016 City Hall, Cardiff
- Welsh Talk (cardiff, 21/11/2016)

2017

- 25/04/2017 Church Talk, Cwmbran
- June 2017: Talk on ‘Dementia’ for Aberystwyth Probus group. (General public)
- September 2017: Study day for Llandaff Diocese, Church of Wales, on Dementia Friendly Churches. (General public)
- July 2017: NOVEL APPROACH TO GENE-BASED ANALYSIS OF ALZHEIMER’S DISEASE INFORMED BY GENETICS OF PSYCHIATRIC DISORDERS. Baker E and Escott-Price V
- October 2017: POLARIS: Polygenic LD-Adjusted Risk Score Approach for Analysis of GWAS Data. Baker E, Sims R, Williams J and Escott-Price V

- BBC Wales: The work of brain banks and the people who donate to them to help with research.
- October 2017: Foundations Autumn Roadshows (Nottingham 50 people 20 public, Southampton 50 people 20 public, London 60 people 25 public, Manchester 40 people 10 public)
- CADR stand at Ageing Well in Wales workshop Swansea 7th February 2018
- 10/10/2017 CADR Conference

2018

- 08/02/2018 Will members of my family develop Alzheimer's disease?
- February 2018: Will members of my family develop Alzheimer's disease?
- March 2018: Genetics of Dementia Summit. Williams
- March 2018: UK DRI at Cardiff. Williams
- April 2018: UK DRI at Cardiff
- April 2018: Wales This Week: Memory Matters
- 17/05/2018 CADR stand at Dementia Action Week open day
- 18/10/ 2018 CADR Annual Conference: Connect 'for' Ageing and Dementia Research

2019

- 20/01/2019 Buckingham Palace 50th Anniv. Investiture Prince of Wales
- 20/02/2019 Whiteley Retirement Village
- 05/03/2019 Davies C Newton Nottage Mother's Union (Porthcawl) monthly meeting talk about Alzheimer's Research
- 05/03/2019 Ministerial Lab Visit
- 25/03/2019 5 Million Genome Analyses
- 21/03/2019 Russell group tour of laboratories
- 03/04/2019 Fund raising for the dementia revolution
- 21/05/2019 Cardiff University Dementia Research stall at Memoria performance by Re-Live
- 29/05/2019 BDR training day at Cardiff
- 30/05/2019 Cheltenham Festival of Science
- 05/06/2019 ARUK Annual Conf 2019
- 17/06/2019 AAIC Los Angeles - Multiplex Model of Alzheimer's
- 19/06/2019 European Society of Human Genetics Conference "Copy number variation in Alzheimer's Patients"
- 19/06/2019 Kerr Lecture, Walton Centre, Liverpool
- 19/06/2019 European Society of Human Genetics Conference "Next Generation Exome Sequencing in a Large Sample of Alzheimer's Patients"
- 26/06/2019 UK-Korea Neuroscience Symposium
- 18/07/2019 Talk to Mother's Union regarding the research conducted by the Dementia Research Team
- 19/07/2019 MRC CNGG celebration dissemination video
- 31/07/2019 AMS fellowship workshop
- 12/08/2019 RCPsych Neuropsychiatry Conf
- 11/09/2019 MRC Centre Celebratory Event

- 20/09/2019 Royal Society visit
- 22/10/2019 Alzheimer's Society Lab Visit
- 23/10/2019 VIB/KUL Distinguished Lecture Series
- 27/11/2019 JPND Genetic Architecture of Alzheimer's

2020

- 01/01/2020 ARUK sponsored Pint of Science Event
- 23/01/2020 Newspaper Article Western Mail. Multiplex Model of Alzheimer's
- 25/02/2020 From Multiomics to mechanisms

ALGORITHMS

Alzheimer's Disease Polygenic Risk Profiling - This model used data from the powerful dataset comprising 17 008 cases and 37 154 controls obtained from the International Genomics of Alzheimer's Project (IGAP). Alzheimer's disease (AD) Polygenic risk scores were generated for 3177 cases and 7277 controls (GERAD data) and tested whether the alleles identified to associate with disease in IGAP sample are significantly enriched in the cases relative to the controls in the GERAD sample. The disease prediction accuracy was investigated in a sample of 3049 cases and 1554 controls (for whom APOE genotype data was available) by means of sensitivity, specificity, area under the receiver operating characteristic curve (AUC) and positive and negative predictive values. The best prediction accuracy AUC = 78.2% (95% confidence interval 77–80%) was achieved by a logistic regression model with APOE, the polygenic score, sex and age as predictors (Escott-Price V. et al. 2015,2017).

DPUK DATA PORTAL STUDY REQUESTS

- DPUK Study Number 0234 (Glasgow) - Apolipoprotein e genotype and healthy brain ageing: stratification, modification and outcomes
- 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
- DPUK Study Number 0276 - Exploration of the rate of progression in Alzheimer's disease

USE OF FACILITIES & RESOURCES

- GERAD consortium genetic and basic cohort phenotype data has been uploaded to the data portal.
- Data generated as part of EM9 (MRC and CHARIOTPRO cohorts) has been uploaded to the genetics data portal.

FURTHER FUNDING

- Leveraging human genetics to identify target populations for dementia therapeutics (VEP PI), £200,134; 01/03/20-28/02/23; (UKDRI (Eisai/DRI joint initiative))
- Translating Individual Alzheimer Genetic risk into disease phenotypes (VEP PI), £371,564; 01/05/20-30/04/23; JPND (MRC)

NEXT DESTINATION

- Christian Bannister – Digital Health Labs
- Elisa Majounie – British Columbia Cancer Agency
- Aura Frizzati – Swansea University Post doctoral researcher
- Georgina Menzies – Cardiff University Lecturer

WP 14a Project report

- Nandini Badarinarayan – NHS Informatics Service
- Ganna Leonenko – Dementia Research Institute
- Detelina Grozeva – Post-doctoral researcher Division of Psychological Medicine
- Roswithas Hopkins - Dementia Research Institute
- Lauren Luckcuck – Centre for Aging and Dementia Research

OTHER

Our findings implicate new pathways to the development of common Alzheimer's disease. We believe this will have a considerable impact on the future understanding of disease mechanism and therapeutic targets.

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.