Lifebrain

D2.6. GWAS data and harmonization of genetic data across samples for genetic analyses across WP4 tasks 2-6.

Project title:	Healthy minds from 0-100 years: Optimising the use of European brain imaging cohorts
Due date of deliverable:	31st December, 2018
Submission date of deliverable:	21st December, 2018
Leader for this deliverable:	University of Lübeck

This project has received funding from the *European Union's Horizon2020 research and innovation programme* under grant agreement *No 732592.*

Contributors	Name	Organisation	Role/ Title
Deliverable leader	Lars Bertram	University of Lübeck	PI
Contributing authors	William Baaré	REGIONH	WP2 leader
	Ulman Lindenberger	MPIB	PI
	Sandra Düzel	MPIB	Researcher
	Kristine Walhovd	UiO	Coordinator
	Anders Fjell	UiO	Researcher
	Athanasia Monika Mowinckel	UiO	Researcher
	Rik Henson	UCAM	PI
	Lorraine Tyler	UCAM	PI
	Tina Emery	UCAM	CALM and CamCAN cohort manager
	Klaus Ebmeier	UOXF	PI
	Eniko Zsoldos	UOXF	Researcher
	Lars Nyberg	Umeå	PI
	Mikael Stiernstedt	Umeå	Research coordinator
	Brenda Penninx	VUmc	PI
	David Bartres Faz	UB	PI
	Cristina Sole-Padulles	UB	Researcher
Final review and approval	Barbara B. Friedman	UiO	Administrative coordinator

Document history				
Release	Date	Reason for change	Status (Draft/In-review/Submitted	Distribution
1.0	10.12.2018	Draft to the Executive Board	In-review	Email
2.0.	17.12.2018	Revised draft to the coordinator	In-review	Email
3.0	21.12.2018	Final version	Submitted	Participant portal

Dissemination level	
Public	X

Executive summary

UzL has reached out to each relevant site with either available DNA samples (i.e. UiO, BASE-II, CALM; for de novo genotyping and / or DNA methylation profiling) or with already generated available genome-wide genotyping data (i.e. RegionH, NESDA, Whitehall II, CamCan, Betula). For the latter, legal documents are currently being checked by the relevant departments at the respective sites and a full data exchange is expected to occur in Q1 of 2019. As soon as all available data are received, UzL will begin with harmonizing the genetic data across sites and begin with the first GWAS analyses. These will first focus on episodic memory measures as the initial GWAS outcome and then extend to additional neurocognitive as well as imaging traits (as available).

Table of contents

List of acronyms/ abbreviations

Lifebrain	Healthy minds from 0 to 100 years: Optimising the use of European brain imaging cohorts
BASE II	Berlin Aging Study II
CALM	Centre for Attention, Learning and Memory, University of Cambridge
CamCAN	Cambridge Centre for Ageing and Neuroscience, University of Cambridge
DNA	Deoxyribonucleic acid
DTA	Data Transfer Agreement
EB	Executive Board
GA	Grant Agreement
LB	Lifebrain
LCBC	Centre for Lifespan Changes for Brain and Cognition
M	Month
MPIB	Max Planck Institute for Human Development
MTA	Material Transfer Agreement
NESDA	The Netherlands Study of Depression and Anxiety
UB	University of Barcelona
UiO	University of Oslo
UOXF	University of Oxford
UzL	University of Lübeck
VUmc	VU University Medical Centre Amsterdam
WP	Work Package

1. Introduction

1.1. Deliverable description

Task 2.6. Generation of new GWAS (genome-wide association study) data and harmonization of genetic data across samples. Lead: UzL; Participants: UiO, UmU, UOXF, MPIB, UB, REGIONH, MRC, VUMC (M6-M24)

This will include re-calling of genotypes, from existing genome-wide SNP genotype data, quality checking, imputation, determination of population substructure, and producing of analysis-ready genotypes maximally harmonized across the different Lifebrain centers. GWAS data will also allow for optimized data analysis of non-genetic variables across datasets, by allowing to account for population stratification by PCA. Parallel to the DNA methylation profiling, we will also perform genome-wide SNP genotyping using Illumina's newest produce, the Global Screening Array, from the same DNA extracts. This will allow a genetically/epigenetically maximally harmonized dataset while in the 2000 high-priority probands. These results will then be used for the analyses in WP4, alongside the epigenetics data.

1.2. Objectives of the deliverable

Re-calling of genotypes, from existing genome-wide SNP genotype data, quality checking, imputation, determination of population substructure, and producing of analysis-ready genotypes maximally harmonized across the different Lifebrain centres.

2. Description of activities

2.1. GWAS

UzL has reached out to all relevant Lifebrain collaborators with respect to the existence, availability and accessibility of GWAS data for the analyses envisioned in WP4. Bilateral Data Transfer Agreements and Material Transfer Agreements have been prepared between UzL and the respective Lifebrain partners.

The following updates to this task are relevant:

1. CALM: MTA is drafted and currently being reviewed by legal teams at UCAM and UzL
2. NESDA: GWAS analyses will be computed locally at the NESDA site. Appropriate cognitive outcome traits are more limited from NESDA, but are currently being selected. First GWAS results are expected in the first quarter of 2019
3. Whitehall II: DTA has been submitted to the Whitehall Data Access Committee
4. Betula: MTA has been approved by Umeå; currently being reviewed by the UzL legal team
5. RegionH: MTA has been received; answers pending from UzL

The results of this outreach are summarized in the Table below.

Cohort	Number of samples	GWAS data availability	Analysis site	Status
Barcelona (ES)	*n.a.*	*No*	*n.a.*	*Sample excluded*
RegionH (DK)	95	Yes (Illumina 660W Quad)	UzL	DTA pending
NESDA (NL)	2600	Yes (Affy 6.0)	NESDA	Ready to go
Whitehall II (UK)	800	No (Illumina HumanCore)	UzL	MTA in preparation
CamCan (UK)	650	Yes (Illumina Omni Express)	CamCan	Ready to go
CALM (UK)	600	No	UzL	MTA pending
Betula (SWE)	355	Yes (OmniExpress & Exome)	UzL	DTA pending
LCBC (NO)	1200	Yes (GSA [ongoing])	UzL	Genotyping ongoing
BASE-II (DE)	1500	GWAS (Affy 6.0)	UzL	Sampling ongoing
TOTAL (with GWAS)	**7800**			

Fortunately, the CALM dataset has become available for additional GWAS genotyping. By contrast, the previously generated genome-wide genotyping data for Whitehall II was deemed of insufficient coverage for the envisioned analyses, so that these samples will be re-genotyped with Lifebrain funding.

3. Conclusions

In summary, we will have approx. 7800 samples available for the GWAS efforts in Lifebrain. Data from these have either already been generated (UiO) or received (BASE-II), the transfer of GWAS data in the other samples is currently pending (CALM). Approximately 2000 of these will also be subjected to DNA methylation profiling for the purpose of EWAS. In the Lifebrain proposal we had listed 2000 samples for DNA extraction, SNP genotyping and EWAS, so our targets will be met.