5/26/21

Scientific Strategy and Cohort Enhancements Group

Review of 2020-2021 Program

Hakon Hakonarson, MD, PhD, CHOP

International 100K Cohort Consortium

Outline

Outline

- Davos Alzheimer's Collaborative
- Cohort Mapping
- Funded Projects
- Previously Proposed Projects and New Directions

1. Davos Alzheimer's Collaborative

- Extensive foundation work done in 2020 bringing together over 70 experts across all disciplines of AD (Co-Dir: Zerhouni/Hakonarson)
- Culminated in a document presented by DAC management at the World Economic Forum in January 2021 (founded by them in 2020)
- Principal focus to research diversity cohorts globally
- Pilot program being funded by DAC to demonstrate ability of diversity cohort leaders to work together and execute
- DAC has established a 5 year plan to complete the project

Davos Alzheimer's Collaborative

Cohort Foundational Phase

About the Foundational Phase

- Pan-global cohort with emphasis on under-researched populations
- Paired with "gold standard" cohort (e.g., UK Biobank)
- Creates understanding of the ideal cohort sample composition

Pillar-1

Create a uniform, minimum data and biosample set and enable consistent data collection across regionally diverse cohorts

- *12,000 subjects*

Pillar-2

Enrich cohorts to increase collection of measurements highly prioritized by Work Group

- *6,000 subjects*

Pillar-3

Develop a biosample repository and management system

Connected, accessible via the IHCC Data Atlas and (in the future) Alzheimer's Data Work Bench

Component description	N ALL	N 50-60 YO	Must/ Nice to do	Scientific value H,M,L	Feasibility H,M,L	Major Tasks	Comments
Cohort with longitudinal data	2,188,078	~539,570	M	H	H	Data Abstraction, OMOP/HPO	Broadly available
Spans age range incl. 50-60 y/o			M	H	M	Data Abstraction	Available
Active recruitment	~985,000		N	H	M	$300/patient	Available
Re-contact in place	~1,540,954		M	H	H	Covered above	Available
Detailed demogr data (age, sex, ancestry etc.)	~2,188,078	~539,570	M	H	H	Data Abstraction	Broadly available
EHR/Detailed Health Info	~1,502,208	~302,051	M	H	H	Abstraction	Broadly available
Blood/DNA	~1,562,478	~411,070	M	H	M	Repositories shipping	Broadly available
Other biological samples (RNA, protein, Cells)	~859,108	~199,507	N	H	M	Repositories Processing and shipping	Available
GWAS – Calculate PRSs	~1,107,250	~200,520	M	H	H	Data processing	Broadly avail.
Sequencing data	~711,640	~143,540	N	H	M	Data processing	Available
Imaging data	~786,956	~136,705	N	H	M	Data processing	Available
Lab values	~902,651	~199,734	M	M	H	Data processing	Broadly avail.
Biomarker data	~1,107,250	~200,520	N	H	M	Data processing	Available
Cloud data storage analysis and data management			N	M	H	Fee per data	Broadly available
IRB-approved data/sam sharing	~1,711,708	~428,850	M	H	H	Effort	Broadly available
Address/Geocoding available	~1,993,078	~448,820	N	M	H	Effort	Broadly available
Digital Health information	~1,814,178	~391,190	N	H	M	Data process	Available
Family history	~1,226,808	~271,670	M	H	H	abstraction	Broadly avail.
Dementia rating sc	~891,251	~162,270	N	M	M	abstraction	Available
Pre-dementia sxs	~614,266	~128,549	N	M	M	abstraction	Available

Cohort Foundation Phase (Pillar-1)

- Create a uniform, minimum dataset
- Enable consistent data collection across regionally diverse cohorts
- Aim for 12,000+ subjects in this initial PRS pilot

IHCC-DAC Cohort Collaborative

Responders to the IHCC-DAC Cohort Survey of Alzheimer/Dementia Pilot.					
Cohort Name	Cohort N	Active Recruitment	~50-60 YO	Longitudinal Feasibility:	Ancestry
ELSA-Brasil	13,000	2008-2010	6,500	Yes	Admixed European, African and Native American genomic ancestry/
Women's Health Initiative	161,808: ~65,000 in active follow-up	1993-1998	53,500	Yes	83% white, 9% Black, 4% Latina, 2.6% Asian/Pacific Islander, <1 Indigenous
Golestan Cohort Study	50,000	2004-2008	16,000	Yes	Caucasian, Turkmen
Generations cohort	113000	2003-2015	25,000	Yes	Primarily European.
UK Biobank	500,000	2006-2010	100,000	Yes	Primarily European
PLCO	155,000	1993-2001	52,000	No	~85% white
UKLWC	>200,000	2001-2005	95,000	Yes	~96% European
UMeå University	370	1985 - 2014	370	NA	Northern Scandinavia
Estonian Biobank	200000	2002	15,000	Yes	European
HUNT 70+, NTNU	9900	2017-2019	0	Yes	European
Ge0mics England (including 100,000 Ge0mes Project)	89,000	Yes (began 2013)	15,000	No	Mainly white European 9 - 17% from other ancestries
National Center for Geriatrics and Gerontology (NCGG)	11,000	Yes (began 2010)	700	Yes	All of our samples is Japanese.
Korean Ge0me and Epidemiology Study/KNIH	235,000	Yes (began 2001)	130,000	Yes	Primarily East Asian ancestry
SAPRIN	300,000	Yes (began 1992)	21,000	Yes	South Eastern Bantu-speakers
Children's Hospital of Philadelphia	150,000	Yes (began 2006)	10,000	Yes	50% EA; 35% AA; 6% Latin; 5% Asian; 4% admixed/native

Total number of subjects: >2,188,078

Total number of subjects aged ~50-60 YO: ~539,570

Davos Alzheimer's Collaborative

Foundational Phase (pillar-2)

	UK Biobank ("gold standard")	IHCC Member Cohort	IHCC Member Cohort	IHCC Member Cohort	IHCC Member Cohort	Other Cohort(s)
	Europe	**South America**	**Middle East**	**East Asia**	**South/S. East Asia**	**Africa**
			- Number of Subjects -			
			New Data to be Acquired			
Exome Sequencing	Already existing	500	500	500	500	2,000
Imaging	Already existing	500	500	500	500	2,000
Alzheimer's Plasma Biomarkers	1,000	500	500	500	500	2,000
Digital Phenotyping	1,000	500	500	500	500	2,000
Cognitive Assessment	Already existing	500	500	500	500	2,000

Davos Alzheimer's Collaborative

- Polygenic Risk Score Pilot:
  - Underway
  - Focus on non-European Participants
  - Opportunity for other cohorts to join in later phase(s)
  - Will focus on addressing gaps in later phase(s)

IHCC-DAC Objective with PRS

- Identify high-risk individuals with no/minimal AD sxs and invest in omics/imaging/clinical/digital biomarker measures and follow prospectively
- Validate a PRS algorithm and have sites run a transethnic algorithm for this pilot project, using UKB as reference cohort
- The endpoint is elevated PRS for AD in subjects who have had some cognitive testing in the form of AD or dementia rating scale assessment.
- Our goal for the pilot is to demonstrate that PRSs are elevated in diversity cohorts similar to EA (or ideally more) and even more importantly, demonstrate we can get access multiple cohort data.

IHCC-DAC Objective with PRS

- Ultimate goal is to provide sites with PRS to identify those individuals who do have highest PRS and have not developed AD yet (or are minimally affected).
- The proposed plan would then be to follow those subjects prospectively over the next few years and get additional omics work done on them, as well as imaging studies and clinical/digital biomarker measures to try to understand the disease progression.
- These subjects will be the population foundation for future biomarker guided precision medicine trials.

2. Mapping Cohorts to Atlas

Cohort Name	Responded	Have Terminology	Mapped?
ELSA-Brasil	Y	Y	Y
Generations Study	Y	Y	Ready to Map
Northern Sweden Health and Disease Study	Y	N	N
NCGG	N	N	NA
UK Blood Donors Cohort	Y	N	Pending
Korean Genome and Epidemiology Study (KoGES)	N	Y	Y
HUNT	Y	Y	Y
UKLWC	Y	Y	Y
PLCO (Prostate, Lung, Colorectal and Ovarian Cancer Screening Trial, NCI)	Y	Y	Y
Golestan Cohort Study	N	Y	Y
Women's Health Initiative	N	N	N
Estonian Genome Project	N	N	N
CHOP	Y	Y	Y
SAPRIN	Y	Y	Y
CanPath	Y	Y	Y

3. Pilot Project Funding Applications

Project	Contact PI
Exploring the role of genetically determined BMI in infancy, childhood, and early adulthood on colorectal cancer development in later life	David Hughes
High-Throughput Metabolomic Biomarker Measures in Diverse Ancestries	Hakon Hákonarson
Opioid cohort consortium (OPICO) to investigate the effects of regular opioid use on mortality and on cancer development	Paul Brennan
Global Mental Health Impact of the COVID-19 Pandemic	Jordan Smoller
Novel corona virus host susceptibility study in South Africa (COVIGen-SA)	Michele Ramsay
Strengthening biospecimen collection for Global Longitudinal Population Studies in the COVID-19 era	John Chambers

4. Future Directions, Including Revisiting Previously Proposed Projects

- Genetic risk factors in low income countries: To begin addressing the diversity gap in human genetics a pilot project within IHCC that would genetically characterize a large well-phenotyped samples, including:
  - Low-depth whole-genome sequencing: Scientifically optimal and most visionary approach (but also the most costly). This would allow extensive characterization (e.g., ultra-rare variants, structural variants) facilitating ethnic-specific analyses
  - Array genotyping and ethnic-specific imputation

- Loss-of-function variants (LoF): Produce a catalogue of high-confidence loss-of-function variants from existing exome and genome data:
  - Develop a public database of LoF variants, including aggregate-level associated phenotypes and availability of biobanked specimen and health record data
  - Develop (where permissible) a framework for re-contacting individuals worldwide who are heterozygous or homozygous for LoF variants

Future Directions

- Cross-cohort analysis of millions of samples
- Phenotype harmonization across cohorts
- GWAS and CNV analysis prior to sequencing
- WES/WGS analysis with focus on RD/OD samples
- Epigenetic analysis
- pQTL (SomaScan or alike platform) data generation
- Execute on Nightingale lipid and inflammatory biomarkers
- Integrative OMICS/phenotype data analysis with focus on high-impact publications in diversity populations
- Pursue grant funding across continents

Questions/Comments?