

ASHG 2019 (69회) **(American society of human genetics)**

참관기

김 태 형

ASHG19에서 발표된 주요 코호트

- All of US
- gnomad
- UK biobank
- GTEx
- 100,000 genome project(Genomics England)
- 23andme
- MyCode
- DiscovEHR
- ACMG
- eMERGE
- TOPMed
- TCGA
- GIANT
- VA's million veteran program(MVP)

Healthcare cohort

National Institutes of Health, NIH
National Heart, Lung and Blood Institute, NHLBI
National Cancer Institute, NCI
National Human Genome Research Institute, NHGRI
Whole exome sequencing, WES
Whole genome sequencing, WGS

N	Name	By	Genetic data
1,000,000+	The All of Us Research Program (All of US)	US (NIH)	Genotyping, etc.
500,000	UK Biobank	UK	Genotyping & WES
150,000	The Trans-Omics for Precision Medicine (TOPMed)	US (NIH NHLBI)	WGS and omics (e.g., metabolic profiles, epigenomics, protein and RNA expression patterns)
100,000	The 100,000 Genomes Project	UK (Genomics England)	WGS
500,000	FinnGen	Finland	Genotyping
200,000	Estonian Biobank	Tartu University and gene bank of Iceland owned by Amgen	Genotyping
11,000	The Cancer Genome Atlas Program (TCGA)	US (NCI and NHGRI)	Omics (genomic, epigenomic, transcriptomic, and proteomic data)
17,382	The Genotype-Tissue Expression (GTEx)	US (GTEx Consortium)	RNA-seq, WGS, WES



- **Study Designs**
- **Genomics, EHR, EMR, Questionnaire, Lab tests**
- **Healthy cohort, Cancer, Rare disease...**
- **Participant Diversity (Ancestry/Ethnicity)**

All of US

Electronic health data, survey about lifestyle habits (such as smoking),
physical measurements (such as blood pressure)

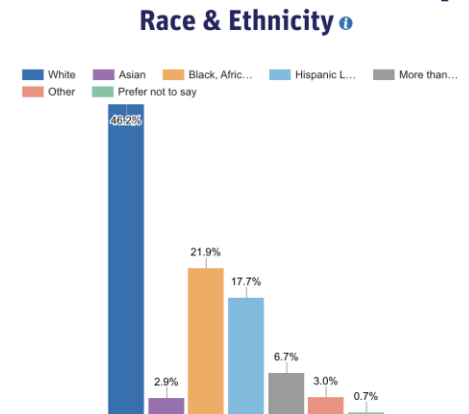
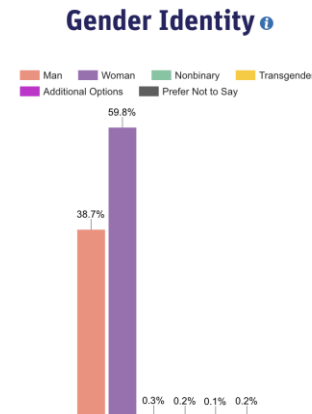
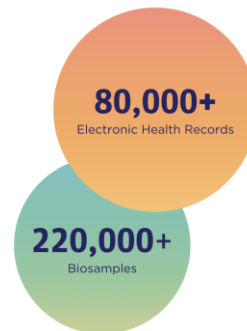
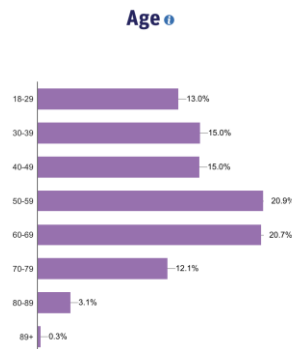
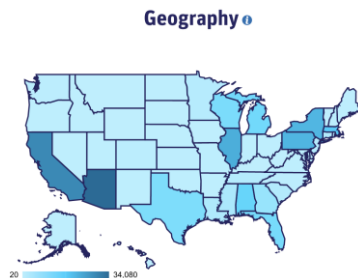


All of UsSM
THE FUTURE OF HEALTH BEGINS WITH YOU

The
Precision
Medicine
Initiative®



1,000,000+ people



<https://www.researchallofus.org/>

<https://allofus.nih.gov/>

EHR Domains: [?](#)

Conditions [?](#)

12,735

medical concepts

36,260 participants in this domain

[View Top Conditions](#)

Drug Exposures [?](#)

14,967

medical concepts

33,440 participants in this domain

[View Top Drug Exposures](#)

Labs and Measurements [?](#)

7,733

medical concepts

32,480 participants in this domain

[View Top Labs and Measurements](#)

Procedures [?](#)

13,202

medical concepts

35,320 participants in this domain

[View Top Procedures](#)

Survey Questions:

The Basics [?](#)

14

survey questions

104,440 participants in this domain

Survey includes participant demographic information.

[View Complete Survey](#)

Overall Health [?](#)

16

survey questions

101,420 participants in this domain

Survey provides information about how participants report levels of individual health.

[View Complete Survey](#)

Lifestyle [?](#)

7

survey questions

100,460 participants in this domain

Survey includes information on participant smoking, alcohol and recreational drug use.

[View Complete Survey](#)

Program Physical Measurements:

Physical Measurements [?](#)

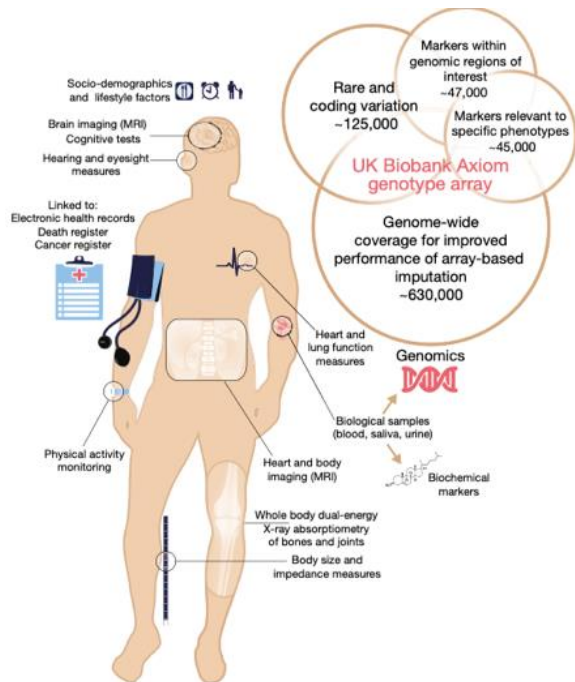
8

physical measurements

81,460 participants in this domain

Participants have the option to provide a standard set of physical measurements as part of the enrollment process ("program physical measurements").

UK Biobank



- Its **500,000** participants have provided information about their **health, well-being and lifestyle**, as well as **blood and other biological samples** for long-term storage and analysis.
- In addition, they have **agreed to have their health followed through medical records** for many years. Scientists from around the world are able to use anonymized data from the resource for research intended to improve the prevention and treatment of a wide range of common disorders.

Bycroft, C., Freeman, C., Petkova, D., Band, G., Elliott, L. T., Sharp, K., et al. (2018). The UK Biobank resource with deep phenotyping and genomic data. *Nature*, 1–25.
<http://doi.org/10.1038/s41586-018-0579-z>

WES of 50,000 UK Biobank participants





Open Targets Genetics

Explore **variant-gene-trait** associations from UK Biobank and GWAS Catalog

Body mass index

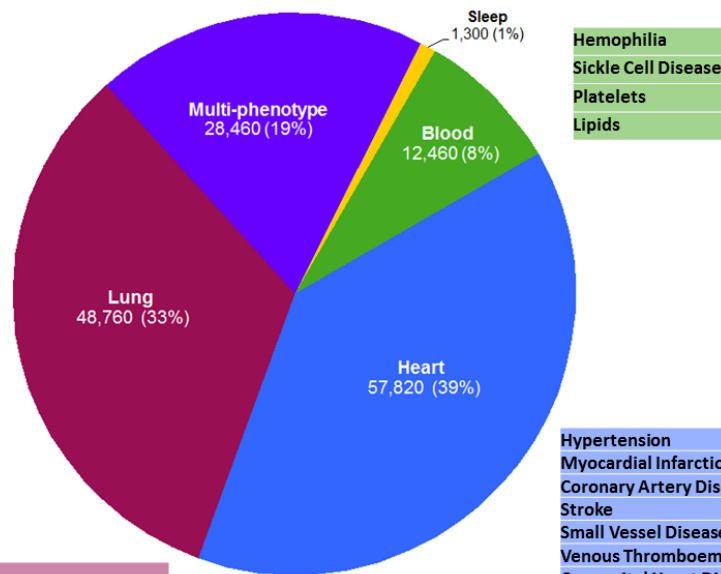


Studies

Body mass index Kichaev G (2018) <i>Am J Hum Genet</i> N Study: 458,000		932 associated loci
Body mass index (BMI) UKB Neale v2 (2018) N Study: 359,983	summary statistics	502 associated loci
Body mass index (BMI) UKB Neale v2 (2018) N Study: 354,831	summary statistics	478 associated loci
Body mass index Hoffmann TJ (2018) <i>Genetics</i> N Study: 334,487		211 associated loci
Body mass index Akiyama M (2017) <i>Nat Genet</i> N Study: 158,284		162 associated loci
Body mass index [EA] Locke AE (2015) <i>Nature</i> N Study: 238,944		78 associated loci

TOPMed

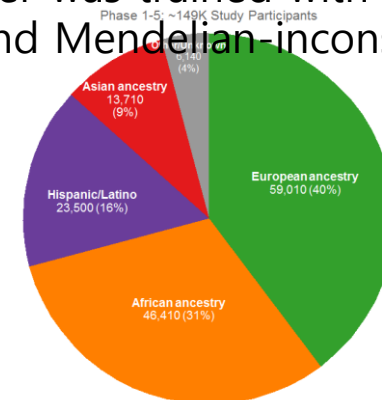
Phase 1-5: ~149K Study Participants



Conditions (Lung)
Asthma
Chronic Obstructive Pulmonary Disease
Idiopathic Pulmonary Fibrosis
Sarcoidosis
Interstitial Lung Disease

Conditions (Heart)
Hypertension
Myocardial Infarction
Coronary Artery Disease
Stroke
Small Vessel Disease
Venous Thromboembolism
Congenital Heart Disease
Atrial Fibrillation
Coronary Artery Calcification
Adiposity
Congestive Heart Failure

- As of October 2019, TOPMed consists of ~149k participants from >80 different studies with varying designs.
- Achieving ancestral and ethnic diversity was a priority in selecting contributing studies. Currently, the 149k participants consist of approximately 60% with substantial non-European ancestry.
- WGS was performed by several sequencing centers to a median depth of 30X using DNA from blood, PCR-free library construction and Illumina HiSeq X technology. A **Support Vector Machine** quality filter was trained with known variants and Mendelian-inconsistent variants.





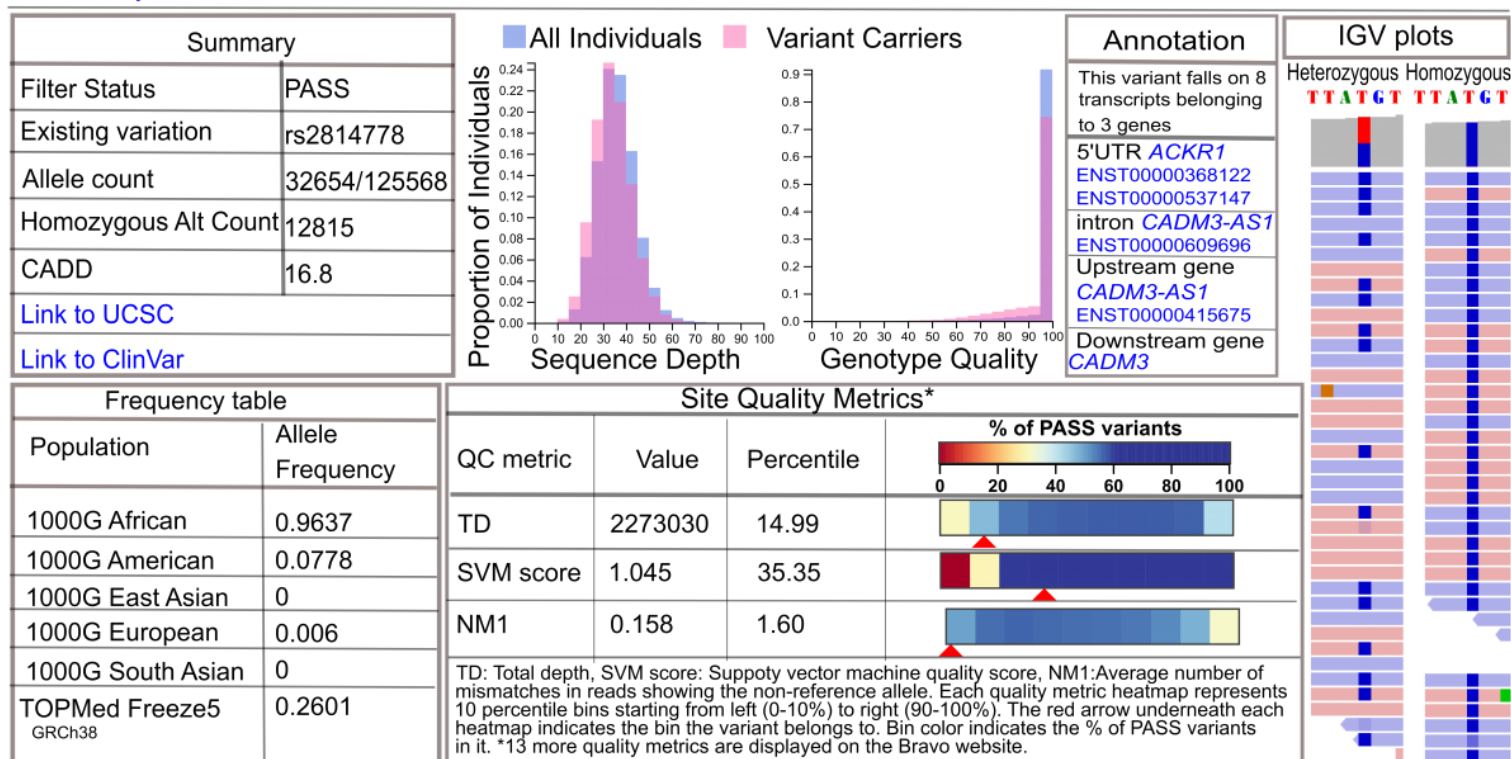
<https://bravo.sph.umich.edu/>

rs2814778

Search

Powered by Freeze5 on GRCh38

The dataset includes 463 million variants on 62784 individuals



100,000 Genomes



- The project was established to sequence 100,000 genomes from around 85,000 NHS patients affected by a **rare disease, or cancer**.
- UK's Department of Health & Social Care => Genomics England (like Ministry of Health and Welfare in Korea)
- In December 2018, the full 100,000 genomes milestone was reached.

- Rare Disease
 - Each Genome: 100Gb
 - Trio is preferred so **300Gb per participant**
 - **x 50,000 participants = 15,000,000Gb total**
- Cancer
 - Germline: 100Gb
 - Tumour: 200Gb
 - **300Gb per patient**
 - **x 25,000 participants = 15,000,000Gb total**
- **10,000,000Gb = 10 Petabytes**
- **Expecting around 30 Petabytes**



10 Billion Photos = 1.5 Petabytes



Data Processed in 1 day = 20 Petabytes

The 100,000 Genomes Project in numbers



100,000 genomes



70,000 patients and family members

110001010101001010100101010000101
110110111010101010001011101000101
110101010001001101010001010100010
001001001110010001000010101010100
100111101100101010110101111001101

21 Petabytes of data.

1 Petabyte of music would take 2,000 years to play on an MP3 player.

FinnGen



DATA FREEZE: 181,821
SAMPLES AVAILABLE: 324,000

- Finland has an internationally unique Biobank Act that makes collections of hospitals and research institutes available for all researchers.

THE FUNDING RESEARCH PARTNERS:

abbvie AstraZeneca Biogen



Genentech
A Member of the Roche Group



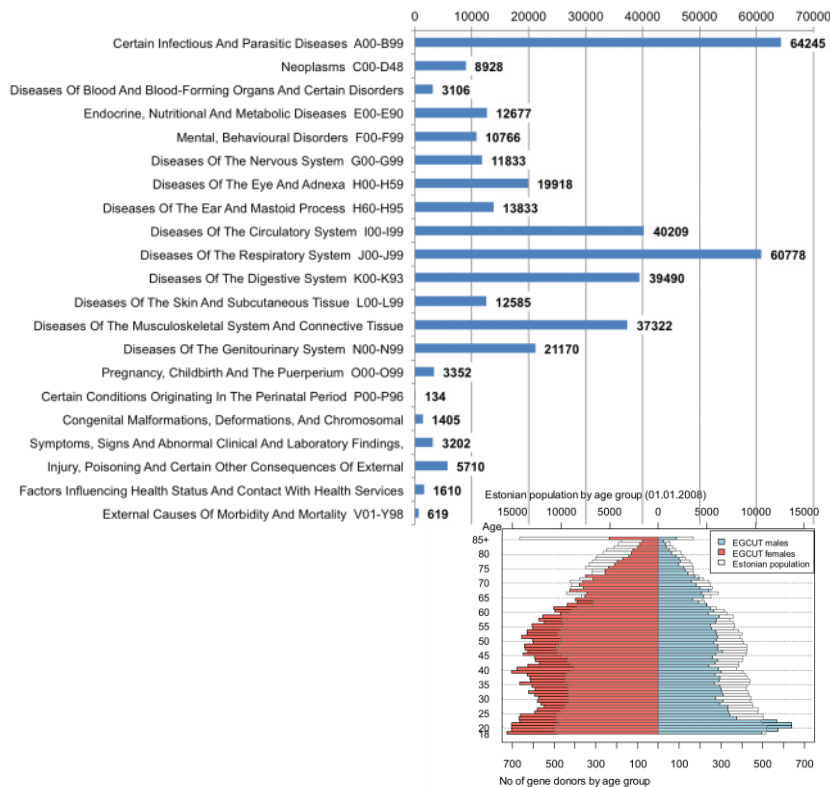
SANOFI

<https://finbb.fi/859-2/>

- Combined with the other strengths of Finland: comprehensive registers, electronic medical records and a research-friendly population

<https://www.finnngen.fi/en/finngenresearchprojectisanexpeditiontothefrontierofgenomicsandmedicine>

Estonian Biobank



- The age, sex and geographical distribution of the **Estonian** population
- Estonians represent 83%, Russians 14%, and other nationalities 3% of all participants
- In addition, there are 40,000 participants with MCTQ (chronotype) data, and 15,000 with both MSTQ and genome-wide microarray (GWAS) data, 3,000 participants have filled the NEO-PI-R questionnaire, including GWAS data available on 2,700 participants.
- Medical history and current health status is recorded according to ICD-10 codes, medication according to ATC

TCGA and GTEX

- **TCGA** is a landmark cancer genomics program, molecularly characterized over 20,000 primary cancer and matched normal samples spanning 33 cancer types.
 - TCGA generated over 2.5 petabytes of genomic, epigenomic, transcriptomic, and proteomic data
- **GTEx** is public resource to study tissue-specific gene expression and regulation.
 - Samples were collected from 54 non-diseased tissue sites across nearly 1000 individuals, primarily for molecular assays including WGS, WES, and RNA-Seq.

AI approach for Genome driven precision care - *Artificial Intelligence Cancer*

