OMICS DATA AVAILABLE IN THL BIOBANK (27.1.2023)

Figures in the tables are estimates of the number of different types of omics data available. For more detailed information, please contact THL Biobank at data.biobank (at) thl.fi.

1. Population collections

Collection name	Imputed genotype data	Chip genotypes	WES (vcf format)	WGS (vcf format)	Metabolomics (NMR)	Other omics
The National FINRISK Study 1992	5130	5200	1190	520	-	-
The National FINRISK Study 1997	7340	7465	3320	1280	7600	-
The National FINRISK Study 2002	7340	7460	5150	2000	7950	-
The National FINRISK Study 2007	5350	5370	2500	660	6030	-
The National FINRISK Study 2012	5460	5560	-	-	5580	-
The National FINRISK Study: Kuusamo Health Examination 2011	240	240	-	-	-	-
The National FINRISK Study: DILGOM 2007	4715 (genotyped from FINRISK 2007 sample)	4730 (genotyped from FINRISK 2007 sample)	2120 (sequenced from FINRISK 2007 sample)	560 (sequenced from FINRISK 2007 sample)	4815	Transcriptomics: 515 Methylomics: 510 Telomeres:4070
The National FINRISK Study: DILGOM 2014	1225 (genotyped from FINRISK 2007 sample)	1230 (genotyped from FINRISK 2007 sample)	-	-	1250 (1230 with data both from 2007 and 2014)	Transcriptomics: 325 (all with data from both 2007 and 2014)
Health 2000 Survey	6800	7000	4960	205	7140	Telomeres: 7400
Health 2011 Survey	4580 (of which 3840 also in H2000)	4700 (of which 4000 also in H2000)	-	-	4750 (3870 with data from both 2000 and 2011)	-
National FinHealth 2017 Study	6150	6250	-	-	5290	-

Note: No omics data is available for Finnish Mobile Clinic Collections.

2. Disease-specific collections

Collection name (collection years)	Imputed genotype data	Chip genotypes	WES (vcf format)	WGS (vcf format)	Metabolomics (NMR)	Other omics
THL Psychiatric Family Collections: schizophrenia (1994-2008)	3170	3180	385	-	-	-
THL Psychiatric Family Collections: bipolar disorder (1994-2004)	560	570	-	-	-	-
SUPER Study (2015-2018)	8860	9020	-	-	-	-
Migraine Family Study (1992-)	8410	8470	490	625	-	-
THL Diabetes Studies (1986-2013)	10 550	10 650	-	-	-	-
The Botnia Study (1990-)	14 440	15 170	-	-	-	-
Corogene Study (2006-2008)	4890	4980	-	-	-	-
FinnishIPF Study (2017-)	270	270	-	-	-	-

3. Other research collections

Collection name (collection years)	Imputed genotype data	Chip genotypes	WES (vcf format)	WGS (vcf format)	Metabolomics (NMR)	Other omics
GeneRISK Study (2015-2018)	7250	7270	-	-	-	-
Twin Study (1975-2012)	14 050	14 100	-	-	-	-
Helsinki Heart Study (1981-1990)	3590	3650	-	-	-	-
ATBC Study / "SETTI" (1984-1993)	15 140	15 280	-	-	-	-
Fin-HIT Study (2011-2014)	(2700 becoming available in August 2023)	(2700 becoming available in August 2023)	-	-	-	-

Data formats:

- Imputed genotype data available as vcf files (build GRCh38/hg38)
- Chip genotypes available mainly as plink files (hg18/hg19/hg38), some as vcf files (build GRCh38/hg38)
- WES data available as vcf files (build GRCh37/hg19 or build GRCh38/hg38)
- WGS data available as vcf files (build GRCh38/hg38)
- Transcriptomics data for DILGOM 2007 and 2014 are in build NCBI36/hg18

THL Biobank imputation reference panel, N=1768 (subset of SISu v3 reference panel) is available as vcf format (build GRCh38/hg38) for solely imputation purposes and is not linkable to any phenotype data.