OMICS DATA AVAILABLE IN THL BIOBANK (18.3.2026)

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 biopankki (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	4930	5200	1220	530	N/A	N/A
The National FINRISK Study 1997	7030	7465	3150	1280	7600	N/A
The National FINRISK Study 2002	7150	7460	5150	1980	7920	N/A
The National FINRISK Study 2007	5200	5370	2420	670	6010	Proteomics: 15 (Olink)
FINRISK: DILGOM 2007 (recall of FINRISK 2007 participants)	4700 ***	4730 ***	2040 ***	560 ***	4815	Transcriptomics: 515 Methylomics: 510 Telomeres:4070
FINRISK: DILGOM 2014 (recall of FINRISK 2007 participants)	1210 ***	1230 ***	N/A	N/A	1250 (1230 with 7-year follow-up)	Transcriptomics: 325 (all with 7-year follow-up)
The National FINRISK Study 2012 **	5460	5800	1915	N/A	5520	Proteomics: 10 (Olink)
Health 2000 Survey	6550	7000	4690	205	7140	Telomeres: 7400
Health 2011 Survey	4400 (of which 3840 also in H2000)	4700 (of which 4000 also in H2000)	2500 (of which all also in H2000)	210	4750 (3870 with 11-year follow-up)	N/A
National FinHealth 2017 Study	6100	6400	N/A	N/A	5290	Proteomics: 30 (Olink)

*All NMR metabolomics are measured with the Nightingale Health platform

** FINRISK 2012 N:s include The National FINRISK Study: Kuusamo Health Examination 2011

*** DILGOM 2007 and 2014 are genotyped and sequenced from FINRISK 2007 sample

2. Other Collections

Collection name (collection years)	Imputed genotype data	Chip genotypes	WES (vcf format)	WGS (vcf format)	Metabolomics (NMR) *	Other omics
SUPER Study (2015-2018)	8570	9020	8350	870	N/A	Proteomics: 125 (Olink) 10 (NULISA)
GeneRISK Study (2015-2018)	6970	7270	N/A	N/A	N/A	Proteomics: 50 (Olink) Lipidomics: 7260
THL Biobank GWAS Cohort	152 400	153 000	33 500	5 800	37 300	Proteomics: 420 (Olink) 180 (NULISA) Lipidomics: 7250

*All NMR metabolomics are measured with the Nightingale Health platform

Data formats:

- Imputed genotype data from FinnGen imputation available as vcf files (build GRCh38/hg38)
- Chip genotypes available either as plink files (hg18/hg19/hg38) or 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=2827 (subset of SISu v4 reference panel) is available as vcf format (build GRCh38/hg38) for solely imputation purposes and is not linkable to any phenotype data.