THL Biobank's SISu v3 WGS Imputation reference panel

The Imputation reference panel

The THL Biobank's SISu v3 reference panel consists of high-coverage whole genome sequence (WGS) data (N= 1768) from THL Biobank's cohorts. The panel can be used to impute any GWAS datasets against a Finnish population-specific reference. The individual sample IDs are pseudonymised, thus the data cannot be combined with any other phenotype data provided from THL Biobank. Use of the reference panel data for any other purposes than as an imputation reference is prohibited.

The reference panel has been prepared in the Sequencing Informatics Team at FIMM, University of Helsinki by Marita Isokallio, Kalle Pärn, Javier Nunez Fontarnau, Samuli Ripatti, Aarno Palotie and Priit Palta.

THL Biobank's cohorts in the reference panel

The THL Biobank's SISu v3 reference panel contains WGS data from the following sample collections:

1. **FINRISK**, N=1160 (National Institute for Health and Welfare/THL Biobank)
   A representative, cross-sectional population surveys from six areas of Finland. Baseline years 1992, 1997, 2002 and 2007.

2. **Health 2000 and 2011 Surveys**, N=205 (National Institute for Health and Welfare/THL Biobank)
   A nationally representative sample of individuals living in Finland in 2000-2001.

3. **Migraine Family Study**, N=403 (University of Helsinki/THL Biobank)
   The sample collection includes persons who have been diagnosed with migraine and their family-members. The samples have been collected since 1992.
   First-degree relatives have been removed from the reference panel data.

All the samples used in the THL Biobank's SISu v3 reference panel have been sequenced at the Broad Institute of MIT and Harvard.

The original SISu reference panel consists of some additional cohorts, see http://sisuproject.fi

The v3 imputation panel is in genome build GRCh38/hg38.

Acknowledging the use of the reference panel

Use of data imputed with the THL Biobank's SISu v3 reference panel and the work of FIMM Sequencing Informatics Team must be acknowledged in publications. A suitable note of acknowledgement is eg: *"The data used for the research was imputed with the THL Biobank's SISu v3 Imputation reference panel obtained from THL Biobank. We thank all study participants for their generous participation in the FINRISK, Health 2000 and Migraine Family studies. We also thank the Sequencing Informatics Team, FIMM Human Genomics, University of Helsinki for the work done in preparation of the reference panel data."*

Additionally, refer to SISu v3 imputation reference panel paper by Pärn *et al.* (manuscript in preparation).

The FIMM Sequencing Informatics Team has also constructed a genotype imputation workflow. If the workflow or some part of it is used, refer to: dx.doi.org/10.17504/protocols.io.nmndc5e