INTERVENE

Report for identification of flagship diseases and secondary phenotypes

Deliverable 2.3

Dissemination level: Public

Part of:

WP2: Data harmonization, integration and evaluation of genetic scores

Project summary	
Project acronym:	INTERVENE
Project full title:	International consortium for integrative genomics prediction
Project Coordinator	Institute for Molecular Medicine Finland FIMM, University of Helsinki; Prof. Samuli Ripatti and Dr. Andrea Ganna
Project start date:	1.1.2021
Project end date:	31.12.2025
Project duration:	60 months
Action type:	RIA
Call identifier:	H2020-SC1-FA-DTS-2018-2020 (Trusted digital solutions and Cybersecurity in Health and Care)
Grant number	101016775
Document descriptors	
Deliverable No.	2.3
Work package	WP2
Deliverable lead	UTARTU and FIMM
Contributors	MGH, NTNU, HUS, GNH
Dissemination level	Public
Expected delivery date	31/12/2021
Submission date	31/12/2021

Change history log

Version	Changes made	Date	Prepared by	PC approved
0.1	**First draft**	**December**	**Kristi Läll, UTARTU**	
1.0	**Technical edits**	**30.12.2021**	**Julius Anckar, UH-FIMM**	**Yes**

Contents

Introduction and purpose

High-quality phenotypes are needed to increase power and accuracy of many genetic analyses, as "cleaner" phenotypes decrease noise and allow us to better characterize the genetic risk factors of different diseases and thus also increase the risk prediction potential. However, obtaining harmonized phenotypes across biobanks is not a straightforward task as biobanks record their data both as coded variables and free text. For coded medical information, several coding systems are simultaneously in use (ICD-9, ICD-10, SNOWMED, etc). Agreeing on phenotypes of interest is the first step, followed by standardized definitions which will result in consistent data structure across biobanks. Therefore, the work done in this deliverable focuses on selecting and standardizing the phenotypes and will facilitate downstream analyses for the entire project.

The flagship project's phenotype selection is based on the following criteria: (**1**) burden on the European healthcare system and potential gain from early prevention and intervention activities (phenotypes where the onset of disease is preventable are preferred); (**2**) strong genetic component and available GWAS summary statistics for creating genetic risk scores or a score available in the PGS catalogue[1]; (**3**) the disease must be frequent enough in contributing biobanks for the analysis; (**4**) the disease must have well defined and good quality endpoints (e.g. blood lipid levels for CAD, fasting glucose and HbA1c for type 2 diabetes). COVID-19 was selected as one of the 'flagship phenotype/disease' analyzed as part of **T3.7** due to the ongoing pandemic. INTERVENE partners have recently collaborated as part of the COVID-19 host genetics initiative[2].

[1] https://www.pgscatalog.org/
[2] https://www.covid19hg.org/

Process description of identification of flagship diseases

Our goal was to select a wide range of diseases, representing several disorder groups, including neurological, endocrinological, rheumatological, ophthalmological, cardiometabolic, pulmonological, oncological, gastroenterological, musculoskeletal, and psychiatric disorders. We constructed a preliminary list of phenotypes, consisting of 54 diseases. To assess the first criteria, we used the Global Burden of Disease tool[3], where we focused on disability-adjusted life years (DALY) as an indicator of overall disease burden, expressed as the number of years lost due to ill-health, disability or early death[4]. We focused on high socio-demographic index countries. Many diseases that contribute to the global burden of disease are not suitable for the type of precision medicine approaches our research aims to create and evaluate. (e.g., injuries or neonatal disorders). The public availability of GWAS summary statistics for creating a genetic risk score or the availability of published scores in the PGS catalogue was then investigated for each trait. We also checked if a curated definition was available in FinnGen clinical endpoint database which was already harmonized with the participating Estonian biobank. Information was aggregated to Supplementary Table 1.

According to the GDB result tool, neoplasms (cancers) impose the greatest overall disease burden (DALYs=51705944). Cardiovascular diseases come second (DALYs=46673687) and musculoskeletal diseases third (DALYs=34405512). Mental and neurological disorders have a similar impact in terms of healthcare burden (DALYs=22126937 and 20016809, respectively). Based on this information, we decided to add “All Cancers” as a flagship endpoint to the final list. For the rest of the diseases under consideration, DALYs are reported separately per disease where possible (Supplementary Table 1). Diseases with no public summary statistics available or diseases with small GWAS case sample size (less than 10,000) were generally excluded from further consideration. For some diseases, partners had access to non-public GWAS results (some cancers, for instance). We shared a survey to let participating partners weigh in on which diseases and traits they were interested in, discussed availability of prevention for each disease and debated possible hypotheses for future analyses, ending up with 40 flagship phenotypes - besides 36 indicated in Supplementary Table 1, All Cancers, BMI as an indicator for overweight and obesity, and two subtypes of epilepsy were added.

[3] http://ghdx.healthdata.org/gbd-results-tool
[4] https://www.who.int/data/gho/indicator-metadata-registry/imr-details/158

To assess the third criteria - phenotype should be frequent enough in biobanks - we conducted a large-scale analysis to define these 38 flagship endpoints in participating cohorts. Endpoint definitions are currently based on ICD-10 and ICD-9 codes simultaneously to accommodate different coding systems among partners. Definitions are derived from FinnGen-Estonian biobank collaboration project (version DF 8[5]) where hundreds of phenotypes have been carefully defined and curated by a large group of experts. From FinnGen's side, key persons in the phenotype definition project have been Elisa Lahtela, Juha Mehtonen, Mitja Kurki, Samuli Ripatti and Mark Daly (please see the full list of clinical groups in Supplement Table 2). From the Estonian side, collaboration was initiated by Tõnu Esko and Priit Palta, phenotype data development was done by Anu Reigo.

After agreeing on phenotype definitions, six partners (FinnGen, Estonian Biobank, Partners Biobank, Genomics England, UK Biobank, Genes and Health) applied given definitions to explore the availability and frequency of agreed diseases/traits, bringing the total sample size to over a million individuals. At the time of this report, HUNT was in process of providing these numbers and some partners needed to account for SNOMED codes as well, so the number of cases and controls will increase. In Table 1, the overall number of cases for each disease is presented based on numbers reported by 6 partners.

Table 1. Definitions of flagship endpoints here reported using ICD-10 codes. Exclusions (with ICD-10 codes when appropriate) and total number of cases is reported. More details about definitions, including other coding systems than ICD-10, are given in Supplementary Table 3.

Endpoints	**ICD-10 Codes**	**Exclude ICD-10 (codes), other exclusions**	**Number of cases**
All Cancers	C00|C01|C02|C03|C04|C05|C06|C07|C08|C09|C10|C11|C12|C13|C14|C15|C16|C17|C18|C19|C20|C21|C22|C23|C24|C25|C30|C31|C32|C33|C34|C35|C36|C37|C38|C39|C40|C41|C43|C44|C45|C46|C47|C48|C49|C50|C51|C52|C53|C54|C55|C56|C57|C58|C60|C61|C62|C63|C64|C65|C66|C67|C68|C69|C70|C71|C72|C73|C74|C75|C76|C77|C78|C79|C80		171582
Malignant neoplasm of colon	C18|C19|C20		22478
Malignant neoplasm of breast	C50	Exclude men	37010
Type 2 diabetes, definitions combined	E11	E10	101784

[5] https://www.finngen.fi/en/researchers/clinical-endpoints

Malignant neoplasm of prostate	C61	Exclude women	33118
Coronary heart disease	I20.0|I21|I22		67438
Subarachnoid haemmorrhage	I60		4975
Melanoma of skin	C43		14715
Asthma	J45|J46		108750
Heart Failure	I50|I11.0|I13.0|I13.2		89463
Stroke excluding SAH	I61|I63|I64	I63.6	35088
Alzheimer's disease (any)	G30|F00		16024
Body mass index	Single time point measurement		887104
Type 1 diabetes	E10	E11	6177
Atrial fibrillation and flutter	I48		84038
Chronic Kidney Disease	N18|Z99.2|Y84.1		31517
Covid-19 severity*	Lab confirmed infection AND hospitalized for COVID.	Confirmed COVID cases not hospitalized	629
Depression	F32|F33		110178
Malignant neoplasm of bronchus and lung	C34		11871
Seropositive rheumatoid arthritis	M05.8|M05.9		12704
Inflammatory Bowel Disease	K50|K51		25026
Venous thromboembolism	I26|I80|O87.1|O88.2	I80.0	40650
Thoracic aortic aneurysm	I71.01|I71.1|I71.2|Q25.43		13511
Abdominal aortic aneurysm	I71.3|I71.4		6296
Hip-Osteoarthritis	M16		59029
Knee-Osteoarthritis	M17		108152
Osteoporosis	M80|M81|M82		36240
Alcohol use disorder	F10|G31.2|G40.51|G62.1|G72.1|I42.6|K29.2|K70|K85.2|K86.0|O35.4|P04.3|Q86.0|Z71.4|E24.4|T51	F100	42856
Hypothyroidism	E03.8|E03.9		103266
Hyperthyroidism	E05.0		8275
Sleep apnoea	G47.3		60363
Idiopathic pulmonary fibrosis	J84.1		3881
Interstitial lung disease	J84		6100
Gout	M10		28636
Primary open-angle glaucoma	H40|H41|H42		39750
Epilepsy	G40		22785
Generalized epilepsy	G403	G40.0|G40.1|G40.2	1475
Focal epilepsy	G40.0|G40.1|G40.2	G40.3	7785
Appendicitis	K35		44696

The majority of the selected diseases have a large combined number of cases. BMI is represented as a continuous variable in the table but can be later used to define binary phenotypes for overweight and obesity. For the COVID-19 severity phenotype, most biobanks were not able to report the exact number of cases, as information about disease status was not widely available yet, but we believe it is likely that they will be in the future. We plan to add another COVID-19 phenotype if data becomes available based on the COVID-19 Host Genetics Initiative's previous work (i.e. defining the case as lab confirmed infection AND hospitalized for COVID AND [death OR respiratory support] AND intensive care, controls would be population without confirmed COVID)

We also visualized the number of cases and grouped them by endpoint family:

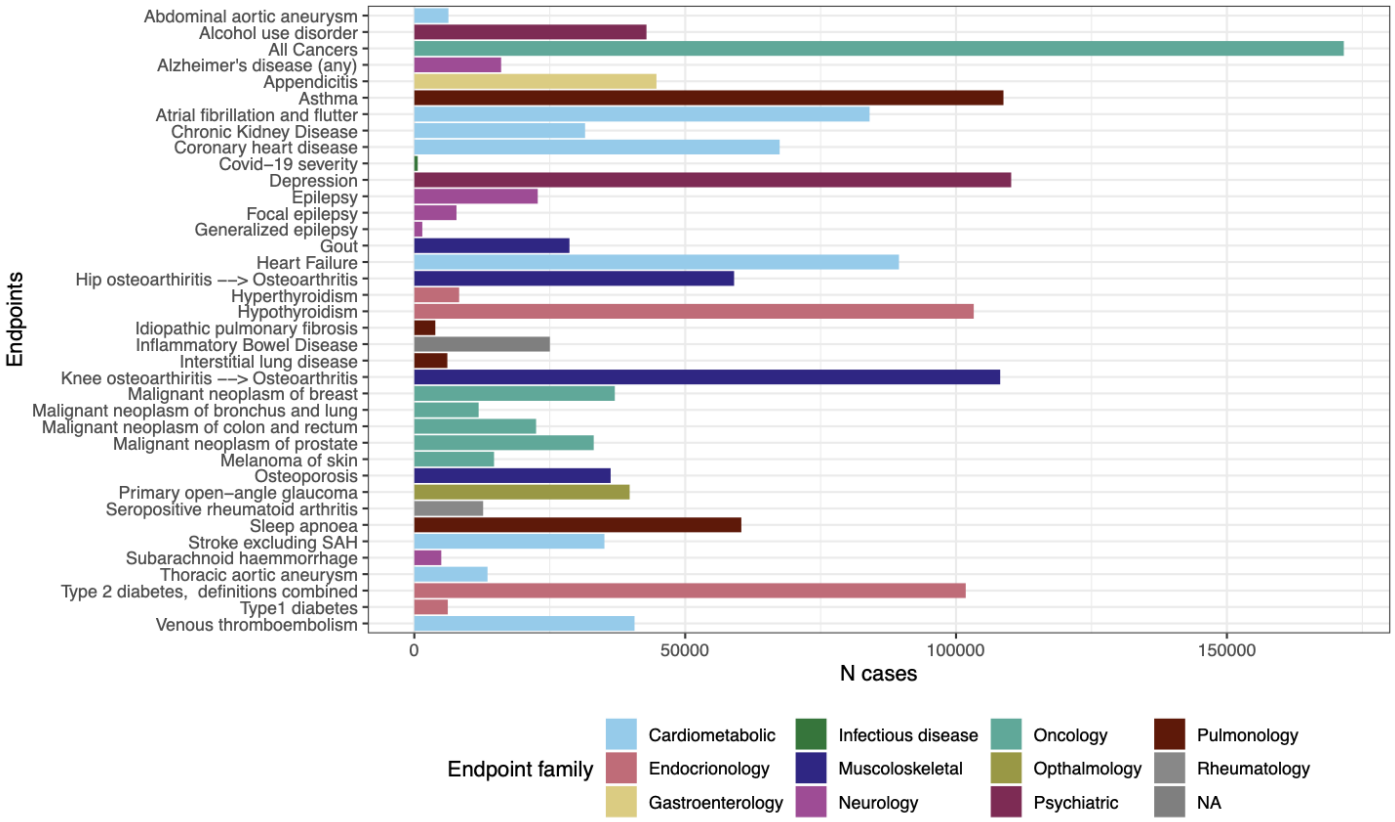

Figure 1. Number of cases visualized. Diseases are grouped by the endpoint family.

We also mapped different risk factors/related endpoints available in the literature for these flagship phenotypes. However, many of them are not available in biobanks, therefore we filtered them to reflect the set of variables which are likely available. This is still an ongoing effort to map out the availability of all flagship disease related endpoints (exact number of values available), but the following variables are most commonly related to diseases and available in most biobanks: BMI, smoking status, sex, education and some lifestyle related variables.

Current and future work with flagship diseases

There are three main areas we plan to focus on next. First, we aim to globally assess the utility of polygenic risk scores across biobanks and ancestries. For that, a pipeline to apply many risk score calculation methodologies at once using harmonized summary statistics was developed by WP3 analysts and it is already tested in some biobanks. Pipeline will be extended in the near future to accommodate ancestry specific analysis and continuous variables as outcomes. Joint predictive ability of polygenic risk scores and publicly available absolute risk algorithms will be studied for some phenotypes. Secondly, it will be investigated how polygenic risk scores vary among demographic and other variables (i.e., to investigate the effect of risk scores in age, sex and other factor groups). Finally, the third direction includes evaluation of the utility of polygenic risk scores for disease progression. In summary, we will evaluate whether the genetic risk factors for the selected diseases can be used for disease risk (alone or jointly with known risk factors) or progression prediction in different ancestries.

Annexes:

1. Supplementary Table 1. Aggregated information about all considered flagship diseases
2. Supplementary Table 2. List of individuals from FinnGen contributing to definitions of flagship diseases
3. Supplementary Table 3. Definitions of flagship diseases using ICD-10 and ICD-9 codes.

Supplementary Table 1. Aggregated information about all considered flagship diseases

Diseases	2019 GBD HIGH-SDI DALYs	Global burden of disease endpoint	Scores in PGS catalog as of August 10,2021	GWAS well powered (> 10,000 cases)	Summary statistics data access link to publicly available sumstats	Finngen clinical endpoint present	Selected
Neurological disorders							
Epilepsy	1281949	idiopathic epilepsy	-	-	+	+	+
Subarachnoid haemmorrhage	3807326	Subarachnoid haemmorrhage	-	+	+	+	+
Late onset Alzheimer’s disease	7655560	Alzheimer's disease and other dementia	+	+	+	+	+
Parkinson's disease	1604208		+	-	+	-	-
Migraine	6085880	Migraine		+	-	+	-
Psychiatric							
Depression	7025129	Depressive disorders	+	+	+	+	+
Bipolar affective disorders	1671828	Bipolar disorder	-	+	+	+	-
Alcohol use disorder	3243926	NA	+	+	+	+	+
Rheumatology							
Seropositive rheumatoid arthritis	705956	rheumatoid arthritis	+	+	+	+	+
Psoriatic arthritis	1123661.978	Psoriasis	+	-	-	+	-
Endocrinology							
Hyperthyroidism	NA	NA	-	+	+	-	+
Hypothyroidism	NA	NA	-	+	+	+	+
Obesity	NA	NA	+	+	+	+	+
Type1 diabetes, definitions combined	750736	Diabetes mellitus type 1	+	+	+	+	+
Type 2 diabetes, definitions combined	10016105	Diabetes mellitus type 2	+	+	+	+	+
Gastroenterology							
Inflammatory bowel disease	603330	Inflammatory bowel disease	+	+	+	+	+
Crohn's disease	NA	NA	-	-	+	+	-
Ulcerative colitis	NA	NA	-	-	+	+	-

Supplementary Table 1. Aggregated information about all considered flagship diseases

Appendicitis	65652	Appendicitis	-	-	-	+	+
Hernia	341845	Inguinal, femoral and abdominal hernia	-	+	+	+	-
Oncology							
Melanoma of skin	743189	Malignant skin melanoma	+	-	+	+	+
Chronic lymphocytic leukemia	263263	Chronic lymphoid leukemia	+	-	+	+	-
Non-Hodgkin lymphoma	1695188	Non-Hodgkin lymphoma	+	-	+	+	-
Malignant neoplasm of colon	6164656	Colon and rectum cancer	+	+	-	+	+
Malignant neoplasm of bronchus and lung	11335290	Tracheal, bronchus, and lung cancer	+	+	-	+	+
Malignant neoplasm of breast	4083850	Breast cancer	+	+	+	+	+
Uterine cancer	596794	Uterine cancer	+	-	+	+	-
Thyroid cancer	204882	Thyroid cancer	+	-	+	+	-
Malignant neoplasm of prostate	2687711	Prostate cancer	+	+	+	+	+
Pulmonology							
Chronic obstructive pulmonary disease (COPD)	10325962	Chronic obstructive pulmonary disease	-	+	-	+	-
Asthma	3002227	Asthma	+	+	+	+	+
Sleep apnoea	NA	NA	-	+	-	+	+
Idiopathic pulmonary fibrosis	NA	NA	-	-	+	+	+
Cardiometabolic							
Atrial fibrillation and flutter	2517229	Atrial fibrillation and flutter	+	+	+	+	+
Heart Failure	NA	NA	+	+	-	+	+
Stroke, includes all strokes	12973818	Stroke	+	+	+	+	-
Stroke, excluding SAH		Ischemic stroke	-	+	+	+	+
Pulmonary embolism	NA	NA	-	-	-		-
Venous thromboembolism	NA	NA	+	+	-	+	+
Coronary heart disease	22126937	Ischemic heart disease	+	+	+	+	+
Myocardial infarction		NA	+	-	-	+	-
Abdominal aortic aneurysm	NA	NA	+	-	-	+	-
Hypertrophic cardiomyopathy	1706484	Cardiomyopathy and myocarditis	+	-	+	+	+
Chronic Kidney Disease	5378054	Chronic kindney disease	+	+	+	+	+

Supplementary Table 1. Aggregated information about all considered flagship diseases

Musculosceletal							
Osteoarthritis	5324252	Osteoarthritis	-	+	+	+	+
Coxarthrosis (hip arthrosis)	398010	Osteoarthritis hip	-	+	+	+	+
Gonarthrosis (knee arthrosis)	2561120	Osteoarthritis knee	-	+	+	+	+
Osteoporosis	NA	NA	-	+	+	+	+
Gout	520272	Gout	+	+	+	+	+
Back pain	15069735	Low back pain	-	+	-	+	-
Opthalmology							
Primary open-angle glaucoma	112534	Glaucoma	+	+	+	+	+
Infectious diseases		NA					
Covid-19 severity	NA	NA	-	-	+	-	+
Covid-19 susceptibility	NA	NA	-	+	+	-	+
Respiratory Syncytial Virus (RSV) severity	4803337	Lower respiratory infections	-	-	-	-	-

Supplementary Table 2 - List of individuals from FinnGen contributing to definitions of flagship diseases

Clinical Groups

Neurology Group

Reetta Kälviäinen	Northern Savo Hospital District, Kuopio, Finland
Valtteri Julkunen	Northern Savo Hospital District, Kuopio, Finland
Hilkka Soininen	Northern Savo Hospital District, Kuopio, Finland
Anne Remes	Northern Ostrobothnia Hospital District, Oulu, Finland
Mikko Hiltunen	Northern Savo Hospital District, Kuopio, Finland
Jukka Peltola	Pirkanmaa Hospital District, Tampere, Finland
Minna Raivio	Hospital District of Helsinki and Uusimaa, Helsinki, Finland
Pentti Tienari	Hospital District of Helsinki and Uusimaa, Helsinki, Finland
Juha Rinne	Hospital District of Southwest Finland, Turku, Finland
Roosa Kallionpää	Hospital District of Southwest Finland, Turku, Finland
Juulia Partanen	Institute for Molecular Medicine Finland, HiLIFE, University of Helsinki, Finland

Gastroenterology Group

Martti Färkkilä	Hospital District of Helsinki and Uusimaa, Helsinki, Finland
Jukka Koskela	Hospital District of Helsinki and Uusimaa, Helsinki, Finland
Sampsa Pikkarainen	Hospital District of Helsinki and Uusimaa, Helsinki, Finland
Airi Jussila	Pirkanmaa Hospital District, Tampere, Finland
Katri Kaukinen	Pirkanmaa Hospital District, Tampere, Finland
Timo Blomster	Northern Ostrobothnia Hospital District, Oulu, Finland
Mikko Kiviniemi	Northern Savo Hospital District, Kuopio, Finland
Markku Voutilainen	Hospital District of Southwest Finland, Turku, Finland
Mark Daly	Institute for Molecular Medicine Finland, HiLIFE, University of Helsinki, Finland

Rheumatology Group

Kari Eklund	Hospital District of Helsinki and Uusimaa, Helsinki, Finland
Antti Palomäki	Hospital District of Southwest Finland, Turku, Finland
Pia Isomäki	Pirkanmaa Hospital District, Tampere, Finland

Supplementary Table 2 - List of individuals from FinnGen contributing to definitions of flagship diseases

Laura Pirilä	Hospital District of Southwest Finland, Turku, Finland
Oili Kaipiainen-Seppänen	Northern Savo Hospital District, Kuopio, Finland
Johanna Huhtakangas	Northern Ostrobothnia Hospital District, Oulu, Finland
Nina Mars	Institute for Molecular Medicine Finland, HiLIFE, Helsinki, Finland

Pulmonology Group

Tarja Laitinen	Pirkanmaa Hospital District, Tampere, Finland
Margit Pelkonen	Northern Savo Hospital District, Kuopio, Finland
Paula Kauppi	Hospital District of Helsinki and Uusimaa, Helsinki, Finland
Hannu Kankaanranta	University of Gothenburg, Gothenburg, Sweden/ Seinäjoki Central Hospital, Seinäjoki, Finland/ Tampere University, Tampere, Finland
Terttu Harju	Northern Ostrobothnia Hospital District, Oulu, Finland
Riitta Lahesmaa	Hospital District of Southwest Finland, Turku, Finland

Cardiometabolic Diseases Group

Teemu Niiranen	Hospital District of Southwest Finland Turku, Finland
Felix Vaura	Hospital District of Southwest Finland Turku, Finland
Veikko Salomaa	The Finnish Institute of Health and Welfare Helsinki, Finland
Kaj Metsärinne	Hospital District of Southwest Finland, Turku, Finland
Mika Kähönen	Pirkanmaa Hospital District, Tampere, Finland
Daniel Gordin	Hospital District of Helsinki and Uusimaa, Helsinki, Finland
Juha Sinisalo	Hospital District of Helsinki and Uusimaa, Helsinki, Finland
Marja-Riitta Taskinen	Hospital District of Helsinki and Uusimaa, Helsinki, Finland
Tiinamaija Tuomi	Hospital District of Helsinki and Uusimaa, Helsinki, Finland
Jari Laukkanen	Central Finland Health Care District, Jyväskylä, Finland
Timo Hiltunen	Hospital District of Helsinki and Uusimaa, Helsinki, Finland
Jussi Hernesniemi	Pirkanmaa Hospital District, Tampere, Finland
Jenni Aittokallio	Hospital District of Southwest Finland, Turku, Finland

Supplementary Table 2 - List of individuals from FinnGen contributing to definitions of flagship diseases

Amanda Elliott	Institute for Molecular Medicine Finland, HiLIFE, University of Helsinki, Finland / Broad Institute, Cambridge, MA, United States
Mary Pat Reeve	Institute for Molecular Medicine Finland, HiLIFE, University of Helsinki, Finland
Sanni Ruotsalainen	Institute for Molecular Medicine Finland, HiLIFE, University of Helsinki, Finland

Oncology Group

Tuomo Meretoja	Hospital District of Helsinki and Uusimaa, Helsinki, Finland
Heikki Joensuu	Hospital District of Helsinki and Uusimaa, Helsinki, Finland
Olli Carpén	Hospital District of Helsinki and Uusimaa, Helsinki, Finland
Johanna Mattson	Hospital District of Helsinki and Uusimaa, Helsinki, Finland
Eveliina Salminen	Hospital District of Helsinki and Uusimaa, Helsinki, Finland
Annika Auranen	Pirkanmaa Hospital District , Tampere, Finland
Peeter Karihtala	Northern Ostrobothnia Hospital District, Oulu, Finland
Päivi Auvinen	Northern Savo Hospital District, Kuopio, Finland
Klaus Elenius	Hospital District of Southwest Finland, Turku, Finland
Johanna Schleutker	Hospital District of Southwest Finland, Turku, Finland
Esa Pitkänen	Institute for Molecular Medicine Finland, HiLIFE, University of Helsinki, Finland
Nina Mars	Institute for Molecular Medicine Finland, HiLIFE, University of Helsinki, Finland
Mark Daly	Institute for Molecular Medicine Finland, HiLIFE, University of Helsinki, Finland

Opthalmology Group

Kai Kaarniranta	Northern Savo Hospital District, Kuopio, Finland
Joni A Turunen	Hospital District of Helsinki and Uusimaa, Helsinki, Finland
Terhi Ollila	Hospital District of Helsinki and Uusimaa, Helsinki, Finland
Hannu Uusitalo	Pirkanmaa Hospital District, Tampere, Finland
Juha Karjalainen	Institute for Molecular Medicine Finland, HiLIFE, University of Helsinki, Finland
Esa Pitkänen	Institute for Molecular Medicine Finland, HiLIFE, University of Helsinki, Finland

Supplementary Table 2 - List of individuals from FinnGen contributing to definitions of flagship diseases

Dermatology Group

Kaisa Tasanen	Northern Ostrobothnia Hospital District, Oulu, Finland
Laura Huilaja	Northern Ostrobothnia Hospital District, Oulu, Finland
Katariina Hannula-Jouppi	Hospital District of Helsinki and Uusimaa, Helsinki, Finland
Teea Salmi	Pirkanmaa Hospital District, Tampere, Finland
Sirkku Peltonen	Hospital District of Southwest Finland, Turku, Finland
Leena Koulu	Hospital District of Southwest Finland, Turku, Finland

Odontology Group

Pirkko Pussinen	Hospital District of Helsinki and Uusimaa, Helsinki, Finland
Aino Salminen	Hospital District of Helsinki and Uusimaa, Helsinki, Finland
Tuula Salo	Hospital District of Helsinki and Uusimaa, Helsinki, Finland
David Rice	Hospital District of Helsinki and Uusimaa, Helsinki, Finland
Pekka Nieminen	Hospital District of Helsinki and Uusimaa, Helsinki, Finland
Ulla Palotie	Hospital District of Helsinki and Uusimaa, Helsinki, Finland
Maria Siponen	Northern Savo Hospital District, Kuopio, Finland
Liisa Suominen	Northern Savo Hospital District, Kuopio, Finland
Päivi Mäntylä	Northern Savo Hospital District, Kuopio, Finland
Ulvi Gursoy	Hospital District of Southwest Finland, Turku, Finland
Vuokko Anttonen	Northern Ostrobothnia Hospital District, Oulu, Finland
Kirsi Sipilä	Northern Ostrobothnia Hospital District, Oulu, Finland

Women's Health and Reproduction Group

Hannele Laivuori	Institute for Molecular Medicine Finland, HiLIFE, University of Helsinki, Finland
Venla Kurra	Pirkanmaa Hospital District, Tampere, Finland
Laura Kotaniemi-Talonen	Pirkanmaa Hospital District, Tampere, Finland
Oskari Heikinheimo	Hospital District of Helsinki and Uusimaa, Helsinki, Finland
Ilkka Kalliala	Hospital District of Helsinki and Uusimaa, Helsinki, Finland
Lauri Aaltonen	Hospital District of Helsinki and Uusimaa, Helsinki, Finland
Varpu Jokimaa	Hospital District of Southwest Finland, Turku, Finland

Supplementary Table 2 - List of individuals from FinnGen contributing to definitions of flagship diseases

Johannes Kettunen	Northern Ostrobothnia Hospital District, Oulu, Finland
Marja Vääräsmäki	Northern Ostrobothnia Hospital District, Oulu, Finland
Outi Uimari	Northern Ostrobothnia Hospital District, Oulu, Finland
Laure Morin-Papunen	Northern Ostrobothnia Hospital District, Oulu, Finland
Maarit Niinimäki	Northern Ostrobothnia Hospital District, Oulu, Finland
Terhi Piltonen	Northern Ostrobothnia Hospital District, Oulu, Finland
Katja Kivinen	Institute for Molecular Medicine Finland, HiLIFE, University of Helsinki, Finland
Elisabeth Widen	Institute for Molecular Medicine Finland, HiLIFE, University of Helsinki, Finland
Taru Tukiainen	Institute for Molecular Medicine Finland, HiLIFE, University of Helsinki, Finland
Mary Pat Reeve	Institute for Molecular Medicine Finland, HiLIFE, University of Helsinki, Finland
Mark Daly	Institute for Molecular Medicine Finland, HiLIFE, University of Helsinki, Finland
Niko Välimäki	University of Helsinki, Helsinki, Finland
Eija Laakkonen	University of Jyväskylä, Jyväskylä, Finland
Jaakko Tyrmi	University of Oulu, Oulu, Finland / University of Tampere, Tampere, Finland
Heidi Silven	University of Oulu, Oulu, Finland
Eeva Slitz	University of Oulu, Oulu, Finland
Riikka Arffman	University of Oulu, Oulu, Finland
Susanna Savukoski	University of Oulu, Oulu, Finland
Triin Laisk	Estonian biobank, Tartu, Estonia
Natalia Pujol	Estonian biobank, Tartu, Estonia

Depression Group

Iiris Hovatta	University of Helsinki, Helsinki, Finland
Chia-Yen Chen	Biogen, Cambridge, MA, United States
Erkki Isometsä	Hospital District of Helsinki and Uusimaa, Helsinki, Finland
Kumar Veerapen	Broad Institute, Cambridge, MA, United States
Hanna Ollila	Institute for Molecular Medicine Finland, HiLIFE, University of Helsinki, Finland
Jaana Suvisaari	National institute of health and welfare (THL)
Thomas Damm Als	Aarhus University, Denmark

Supplementary Table 2 - List of individuals from FinnGen contributing to definitions of flagship diseases

ENT (ear, nose and throath) Group

Antti Mäkitie	Hospital District of Helsinki and Uusimaa, Helsinki, Finland
Argyro Bizaki-Vallaskangas	Pirkanmaa Hospital District, Tampere, Finland
Sanna Toppila-Salmi	University of Helsinki, Helsinki, Finland
Tytti Willberg	Hospital District of Southwest Finland, Turku, Finland
Elmo Saarentaus	Institute for Molecular Medicine Finland, HiLIFE, University of Helsinki, Finland
Antti Aarnisalo	Hospital District of Helsinki and Uusimaa, Helsinki, Finland
Eveliina Salminen	Hospital District of Helsinki and Uusimaa, Helsinki, Finland
Elisa Rahikkala	Northern Ostrobothnia Hospital District, Oulu, Finland
Johannes Kettunen	Northern Ostrobothnia Hospital District, Oulu, Finland

Supplementary Table 3. Definitions of flagship diseases using ICD-10 and ICD-9 codes.

Endpoints	Phenotype Code	ICD 10 Codes (3digit and subtype codes are not separated with ".")	ICD 9 Codes	Exclude ICD 10 (codes)
All Cancers	C3_CANCER	C00|C01|C02|C03|C04|C05|C06|C07|C08|C09|C10|C11|C12|C13|C14|C15|C16|C17|C18|C19|C20|C21|C22|C23|C24|C25|C30|C31|C32|C33|C34|C35|C36|C37|C38|C39|C40|C41|C43|C44|C45|C46|C47|C48|C49|C50|C51|C52|C53|C54|C55|C56|C57|C58|C60|C61|C62|C63|C64|C65|C66|C67|C68|C69|C70|C71|C72|C73|C74|C75|C76|C77|C78|C79|C80	140|141|142|143|144|145|146|147|148|149|150|151|152|153|154|155|156|157|158|159|160|161|162|163|164|165|170|171|172|173|174|180|181|182|183|184|185|186|187|188|189|190|191|192|193|194|195|199	
Malignant neoplasm of colon	C3_COLORECTAL	C18|C19|C20	153|1540|1541	
Malignant neoplasm of breast	C3_BREAST	C50	174	
Type 2 diabetes, definitions combined	T2D	E11	25000|25010	E10
Malignant neoplasm of prostate	C3_PROSTATE	C61	185	
Coronary heart disease	I9_CHD	I200|I21|I22	410|4110	
Subarachnoid haemmorrhage	I9_SAH	I60	430	
Melanoma of skin	C3_MELANOMA_SKIN	C43	172	
Asthma	J10_ASTHMA	J45|J46	493	
Heart Failure	I9_HEARTFAIL_NS	I50|I110|I130|I132	4029B|428	
Stroke excluding SAH	I9_STR	I61|I63|I64	431|4330A|4331A|4339A|4340A|4341A|4349A|436	I636
Alzheimer's disease (any)	G6_AD_WIDE	G30|F00	3310	
Type1 diabetes	T1D	E10	25001|25011	E11
Atrial fibrillation and flutter	I9_AF	I48	4273	
Chronic Kidney Disease	N14_CHRONKIDNEYDIS	N18|Z992|Y841	585	
Depression	F5_DEPRESSIO	F32|F33	2961|2968|3004	
Malignant neoplasm of bronchus and lung	C3_BRONCHUS_LUNG	C34	162	
Seropositive rheumatoid arthritis	RHEUMA_SEROPOS_OTH	M058|M059	7140A	
Inflammatory Bowel Disease	K11_IBD_STRICT	K50|K51	555|556	
Venous thromboembolism	I9_VTE	I26|I80|O871|O882	415|451|6713|6714|6732	I800

Supplementary Table 3. Definitions of flagship diseases using ICD-10 and ICD-9 codes.

Thoracic aortic aneurysm	I9_THAORTANEUR	I7101|I711|I712|Q2543		
Abdominal aortic aneurysm	I9_ABAORTANEUR	I713|I714	4413A|4414A	
Hip-Osteoarthritis	COX_ARTHROSIS	M16	7151E|7152E	
Knee-Osteoarthritis	KNEE_ARTHROSIS	M17	7151F|7152F	
Osteoporosis	M13_OSTEOPOROSIS	M80|M81|M82	7330|7331	
Alcohol use disorder	AUD_SWEDISH	F10|G312|G4051|G621|G721|I426|K292|K70|K852|K860|O354|P043|Q860|Z714|E244|T51	303|291|3050A|3575A|4255|5353A|5710|5711|5712|5713|5770D|5770E|5771C|5771D|7607A|980	F100
Hypothyroidism	E4_HYTHYNAS	E038|E039	2448|2449	
Hyperthyroidism	E4_THYTOXGOITDIF	E05.0	2420	
Sleep apnoea	G6_SLEEPAPNO	G473	3472	
Idiopathic pulmonary fibrosis	IPF	J841		
Interstitial lung disease	ILD	J84	515|516	
Gout	GOUT	M10	2740	
Primary open-angle glaucoma	H7_GLAUCOMA	H40|H41|H42	3650|3651|3652|3653A|3654A|3655|3656A|3656B|3656C|3656E|3656F|3656X|3658|3659	
Epilepsy	G6_EPLEPSY	G40	345	
Generalized epilepsy	GE_STRICT	G403	3450|3451|3452|3453	G400|G401|G402
Focal epilepsy	FE_STRICT	G400|G401|G402	3454|3455	G403
Appendicitis	K11_APPENDACUT	K35	540	