Genomitieto ja lääketieteellinen tutkimus
Majvik, lokakuun 30, 2015
Aarno Palotie M.D., Ph.D.
The technology drives the change
We project 4,500 Tb in 2016 (based on current installed base and yield).
THE COST AND SPEED OF GENOME SEQUENCING IS NO LONGER THE PROBLEM

MOORE’S LAW
BIG DATA

“No other book offers such an accessible and balanced tour of the many benefits and downsides of our continuing infatuation with data.”

— WALL STREET JOURNAL

Viktor Mayer-Schönberger
and
Kenneth Cukier
Tietojen käsittelyn muutos, onko lääketiede samassa murroksessa?
Genome data and other trends
data

development

community

quality

service

services

health

incidence

related

structure

package

functions

incidents

records

system

service

quality

Tourism

priority

pollution

road

finder

GIS

Real

TTC

search

Bike

meeting

public

specific

traffic

complaints

chains

projects

parks

Finder

availability

questions

crime

mobile

Archives

Monitor

monitoring

availability

issues

Emergency

addresses

result

pollutants

air

measures

underground

buses

stations

bus

numbers

land

schools

find

beach

stations

made

open

parking

Raw

apps

voting

Financial

events

routes

mapping

location

councillor

OPEN DATA

FiMM

BROAD INSTITUTE

MASSACHUSETTS GENERAL HOSPITAL
Global IT companies interested in health applications

Google will spend up to $1.5 billion to develop treatments for age-related diseases.
Sensors, >10,000 health apps

Quantified Self

- Heart rate
- Temperature
- Oxygen levels
- Blood pressure
- ECG

- Daily profiles
- Deep wellness

Distance
Duration
Steps
Calories
Geenit ja ympäristö
Elämäntavat

Kuntoilu

Ruokavalió

Sairausriski

Elämäntavat
Shared genes behind immune mediated diseases
Technology driving and supporting Individual organ group approach leads to Systems approach, leading to Change in disease classification. Effect on ICD classification?
The opportunity

• An explosion of genomic information from individuals with known clinical characteristics and disease outcomes

• Learning from these data, we should accelerate progress in:

  • Understanding the basis of inherited disease
  • Cancer outcomes and targeted therapy
  • Identifying targets for drug development
  • Infectious disease

  • Clinical interpretation of individual genome sequences

David Altshuler, Global Alliance
The challenge: we can write down genomes, but we don’t yet know how to read them
To learn, we must compare
Each gene variant has a small effect.

Reference database like reference values in clinical chemistry.
PHARMA R&D SPEND

53 NEW MOLECULAR ENTITIES (NME)

$17 Bn 1996

$45 Bn 2009

18
HIGH COSTS THROUGHOUT THE DRUG DEVELOPMENT PROCESS

- Discovery
- Preclinical Development
- Clinical Development
- Approval
- Market

- Target Identification
- Toxicology
- Efficacy & Tolerability
- Outcomes
- Cost Efficiency
- Surveillance
- New Target Populations
**Could Genomics Help to Control the Rising Costs?**

<table>
<thead>
<tr>
<th>DISCOVERY</th>
<th>PRECLINICAL DEVELOPMENT</th>
<th>CLINICAL DEVELOPMENT</th>
<th>APPROVAL</th>
<th>MARKET</th>
</tr>
</thead>
<tbody>
<tr>
<td>TARGET IDENTIFICATION</td>
<td>TOXICOLOGY</td>
<td>EFFICACY &amp; TOLERABILITY</td>
<td>OUTCOMES</td>
<td>SURVEILLANCE</td>
</tr>
<tr>
<td>TARGET PRIORIZATION</td>
<td></td>
<td></td>
<td>COST EFFICIENCY</td>
<td>NEW TARGET POPULATIONS</td>
</tr>
</tbody>
</table>
LEVERAGING BIG DATA TO CONTROL HEALTHCARE COST INCREASE

COST MANAGEMENT

COLLECTING, COMBINING AND ANALYZING

CUES TO INTERVENE

PATIENT ENGAGEMENT

Imaging files
Published research
Remote monitoring
mHealth apps

Social media
Genomic data
Email, text messages
EHR / EMR

Claims data
Biometric sensors
Clinician notes
Metabolomic data

Photos, video
Microbiomic data
Epidemiological and Clinical Finnish Sample Collections

In these pages we have collected information on epidemiological and clinical Finnish study collections with available DNA samples, though the list is not comprehensive. The pages include short description of the projects, contact information, as well as information on genome-wide SNP genotyping studies involving these study collections. The pages are meant to serve as a resource for investigators and promote collaboration between research groups and institutes.

If you wish to add information on your own study collections, please contact us: Kaisa Silander or Markus Perola (firstname.lastname (at) thi.fi).

Link to Studies Summary Table
TOWARDS NEW INNOVATIONS

ACADEMIC RESEARCH

+ 

INDUSTRY

FIIMM

UNIVERSITY OF HELSINKI

BROAD INSTITUTE

YOUR TESTBED FOR NEXT GENERATION RESEARCH & INNOVATION
From biobanks to personalized medicine

Biobanks
Gene information
Extensive health records

Data storage and integration

Modeling

Implementation in Health Care
<table>
<thead>
<tr>
<th>NAME OF REGISTER</th>
<th>DATA INCLUDED IN REGISTER</th>
<th>KEEPER</th>
<th>ESTABLISHED IN COMPUTER FORMAT</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hospital Discharge Register (HILOMO)</td>
<td>Homes and institutions for the mentally disabled, including information on treatment</td>
<td>THL</td>
<td>1967</td>
</tr>
<tr>
<td>Finnish Registry for Kidney Diseases</td>
<td>Diseases, type of treatment, and laboratory tests</td>
<td>ETK</td>
<td>1964</td>
</tr>
<tr>
<td>Cancer Register</td>
<td>Cancer patient information from hospitals, pathology, laboratory measurements etc.</td>
<td>THL</td>
<td>1953</td>
</tr>
<tr>
<td>Finnish Register of Visual Impairment</td>
<td>Patient’s visual ability</td>
<td>THL</td>
<td>1983</td>
</tr>
<tr>
<td>National Infectious Diseases Register</td>
<td>Detailed information on cases in infectious diseases</td>
<td>THL</td>
<td>1989</td>
</tr>
<tr>
<td>Register of Congenital Malformations</td>
<td>Infants and foetuses</td>
<td>THL</td>
<td>1963</td>
</tr>
<tr>
<td>Drug Reimbursement Registers</td>
<td>Disease that is being treated and medication used</td>
<td>KELA</td>
<td>1967</td>
</tr>
<tr>
<td>Medical Birth Register</td>
<td>Information on all births in Finland, from gestation week 22+0 or birthweight 500g</td>
<td>THL</td>
<td>1987</td>
</tr>
<tr>
<td>Cause-of-Death Register</td>
<td>Intermediate case of death and contributing causes of death</td>
<td>STAT</td>
<td>1969</td>
</tr>
<tr>
<td>Register on Occupational Disease</td>
<td>Diagnosis of occupational disease</td>
<td>FIOH</td>
<td>1964</td>
</tr>
<tr>
<td>Drug Surveillance Register</td>
<td></td>
<td>FIMEA</td>
<td>1982</td>
</tr>
<tr>
<td>National Sickness Insurance</td>
<td>Social benefit information</td>
<td>KELA</td>
<td>1967</td>
</tr>
<tr>
<td>Register on Pensions</td>
<td>Work pension information, age of individual, type of pension</td>
<td>ETK</td>
<td>1962</td>
</tr>
<tr>
<td>Finnish Employment Register</td>
<td>Work in private sector, work as entrepreneur and work without pay</td>
<td>ETK</td>
<td></td>
</tr>
<tr>
<td>Central Population Register</td>
<td>Relations (stillbirths are not registered)</td>
<td>VRK</td>
<td>1973</td>
</tr>
<tr>
<td>Register on Social Assistance</td>
<td></td>
<td>THL</td>
<td>1985</td>
</tr>
<tr>
<td>Child Welfare Register</td>
<td>Individual-level information on children taken into custody</td>
<td>THL</td>
<td>1991</td>
</tr>
</tbody>
</table>

THL = National Institute of Health and Welfare  
ETK = The Finnish Kidney and Liver Association  
KELA = Social Insurance Institution  
STAT = Statistics Finland  
FIOH = Finnish Institute of Occupational Health  
FIMEA = National Agency for Medicine  
ETK = Finnish Centre for Pensions  
VRK = Central Population Register
EXAMPLE OF HEALTH HISTORIES FROM **TWO PERSONS** FROM THE NATIONAL BIOBANKS WITH A 40 YEAR FOLLOW-UP

<table>
<thead>
<tr>
<th>Age</th>
<th>Disease History</th>
</tr>
</thead>
<tbody>
<tr>
<td>20</td>
<td>Neoplasm of unspecified nature</td>
</tr>
<tr>
<td></td>
<td>Benign neoplasm of breast</td>
</tr>
<tr>
<td></td>
<td>Spontaneous vertex delivery</td>
</tr>
<tr>
<td></td>
<td>Cerebral aneurysm, nonruptured</td>
</tr>
<tr>
<td></td>
<td>Spontaneous vertex delivery</td>
</tr>
<tr>
<td></td>
<td>Maternal care for poor fetal growth</td>
</tr>
<tr>
<td></td>
<td>Sterilisation</td>
</tr>
<tr>
<td></td>
<td>Cerebral aneurysm, nonruptured</td>
</tr>
<tr>
<td></td>
<td>Other dystonia</td>
</tr>
<tr>
<td></td>
<td>Severe depressive episode without psychosis</td>
</tr>
<tr>
<td></td>
<td>Recurrent depressive disorder, without psychosis</td>
</tr>
<tr>
<td></td>
<td>Acute cholecystitis</td>
</tr>
<tr>
<td>30</td>
<td>Diarrhea of infectious origin</td>
</tr>
<tr>
<td></td>
<td>Other bursitis</td>
</tr>
<tr>
<td></td>
<td>Internal haemorrhoids with complications</td>
</tr>
<tr>
<td></td>
<td>Essential (primary) hypertension</td>
</tr>
<tr>
<td></td>
<td>Polymyalgia rheumatica</td>
</tr>
<tr>
<td></td>
<td>Guillain-Barré syndrome</td>
</tr>
<tr>
<td></td>
<td>Sleep apnea</td>
</tr>
<tr>
<td></td>
<td>Other spondylitis with radiculopathy</td>
</tr>
<tr>
<td></td>
<td>Nonallergic asthma</td>
</tr>
<tr>
<td></td>
<td>Pilonidal cyst without abscess</td>
</tr>
<tr>
<td></td>
<td>Pain localised to upper abdomen</td>
</tr>
<tr>
<td></td>
<td>Acute pancreatitis</td>
</tr>
<tr>
<td></td>
<td>Calculus of gallbladder without cholecystitis</td>
</tr>
<tr>
<td></td>
<td>Tear of meniscus</td>
</tr>
<tr>
<td></td>
<td>Inguinal hernia, without obstruction or gangrene</td>
</tr>
</tbody>
</table>

**Haplotype A**: CTAAGTA

**Haplotype B**: CTACGTA

**SNP**

**DNA Sequence**

**FIIMM**

**BROAD INSTITUTE**

**MGH**

**Massachusetts General Hospital**

**Finland**

**YOUR TESTBED FOR NEXT GENERATION RESEARCH & INNOVATION**
National Biobanks Finland portal
130 000 individuals from population cohorts
70 000 individuals from disease collections

Type 2 Diabetes
> 10 000 individuals

High Blood pressure

Severe mental health:
- schizophrenia, depression
- > 5000 individuals

Migraine
15 000 individuals

Old age dementia,
~ 5000 individuals

Cancer
> 10 000 cases

Cardiovascular events
stroke, CHD
25 000 individuals

Cardiovascular risk factor data
100 000 individuals

Life course events

Prescription medication data
18 000 statin users
20 000 estrogen substitution th

Cause of death data

National Biobank
201 858 individuals

Life style and socio-economic data
- education, economic state, smoking

National Biobanks Finland portal
130 000 individuals from population cohorts
70 000 individuals from disease collections

Type 2 Diabetes
> 10 000 individuals

High Blood pressure

Severe mental health:
- schizophrenia, depression
- > 5000 individuals

Migraine
15 000 individuals

Old age dementia,
~ 5000 individuals

Cancer
> 10 000 cases

Cardiovascular events
stroke, CHD
25 000 individuals

Cardiovascular risk factor data
100 000 individuals

Life course events

Prescription medication data
18 000 statin users
20 000 estrogen substitution th

Cause of death data

National Biobank
201 858 individuals

Life style and socio-economic data
- education, economic state, smoking
GENETIC ISOLATION

bottleneck: catastrophic reduction in population

original population → chance survivors → new population

Copyright © The McGraw-Hill Companies, Inc. Permission required for reproduction or display.
TARGET EVALUATION

- POPULATION ISOLATE
- HEALTH REGISTERS
- BIOBANKS
- GENOME DATA
YOUR TESTBED FOR NEXT GENERATION RESEARCH & INNOVATION

TARGET EVALUATION

POPULATION ISOLATE
HEALTH REGISTERS
BIOBANKS
GENOME DATA
EARLY SETTLEMENT
• 2000-10 000 years ago
• South and Coast

LATE SETTLEMENT
• 16th century
• multiple bottle necks

EXPANSION
• 18th century – population 250 000
• Today – population 5.4 million
What are the unique opportunities in a bottlenecked population?

- Recessive lethal alleles can exist at or above 1%
- Alleles with OR of 2-5 can exist at low frequencies
- Both of these types of variation are incredibly unlikely to reach even .001 frequency in broader European sample

 Subset of .1-10% alleles are 5-20x more frequent in Finns than all other worldwide populations
TARGET EVALUATION

POPULATION ISOLATE + HEALTH REGISTERS + BIOBANKS + GENOME DATA
POPULATION ISOLATE

HEALTH REGISTERS

BIOBANKS

GENOME DATA

TARGET EVALUATION
TARGET EVALUATION

- POPULATION ISOLATE
- HEALTH REGISTERS
- BIOBANKS
- GENOME DATA
SISU-PROJECT
SEQUENCING INITIATIVE SUOMI (FINLAND)

THE 200K

- Genome wide genotype data: 73,000
- Genome or exome sequences: >26,000

NATIONAL BIOBANK

- 200,000 individuals
- 4% of the population
- Population cohorts
- Extensive health, phenotype, metabolomic data
- Disease specific collections

IMPUTATION

POPULATION SPECIFIC CHIP/GENOTYPING

REFERENCE DATABASE
Sequencing Initiative Suomi - Data resource for the research community

Data resource for the research community

The Sequencing Initiative Suomi (The SISu project) is an international collaboration between research groups aiming to build tools for genomic medicine.

This SISu v3.0 (2015-08-28) release includes:

- 1 037 122 sites with 1 137 703 variants
- 6118 Finnish samples from 10 cohorts that were sequenced in Broad and Sanger Institutes
- Original sequencing was done using three different platforms (Agilent 1.1 refseq plus 3 boosters, Agilent sureselect 50mb, Illumina coding v1)
- Multiallelic sites and indels are now included
- Finnish enrichment and other information fields for custom filtering purposes now available

Search

The SISu data resource currently covers exons only and data is restricted to autosomal SNPs and Indels. Genome build used in this release was GRCh37. Minor allele frequencies before QC and after QC provided.

Information about quality control process

→ Enter search
YOUR TESTBED FOR NEXT GENERATION RESEARCH & INNOVATION

TARGET EVALUATION

+ POPULATION ISOLATE
+ HEALTH REGISTERS
+ BIOBANKS
+ GENOME DATA
HUMAN GENE KNOCK OUTS
NEED FOR HUMAN MODELS
HUMAN GENE KNOCK OUTS
NEED FOR HUMAN MODELS
HUMAN GENE KNOCK OUTS
NEED FOR HUMAN MODELS
PROTECTIVE GENE VARIANTS
Loss-of-function mutations in SLC30A8 protect against type 2 diabetes

Truncating mutations in SLC30A8 seen 3x more often in healthy controls than diabetics

STRONG protection against diabetes !!!
Distribution and Medical Impact of Loss-of-Function Variants in the Finnish Founder Population

Elaine T. Lim¹,²,³,⁴, Peter Würtz⁵,⁶,⁷, Aki S. Havulinna⁶, Priti Palta⁵,⁸, Taru Tukiainen¹,²,³, Karola Rehnström⁸, Tõnu Esko²,³,⁹,¹⁰, Reedik Mägi⁹, Michael Inouye¹¹, Tuuli Lappalainen¹²,¹³, Yingleong Chan²,⁴,¹⁰, Rany M. Salem²,¹⁰, Monkol Lek¹,²,³, Jason Flannick²,³, Xueling Sim¹⁴, Alisa Manning², Claes Ladenvall⁵,¹⁵, Suzannah Bumpstead⁸, Eija Hämäläinen⁵,⁸, Kristiina Aalto¹⁶, Mikael Maksimow¹⁶, Marko Salmi¹⁷, Stefan Blankenberg¹⁸,¹⁹, Diego Ardissino²⁰, Svati Shah²¹, Benjamin Horne²², Ruth McPherson²³, Gerald K. Hovingh²⁴, Muredach P. Reilly²⁵, Hugh Watkins²⁶, Anuj Goel²⁶, Martin Farrall²⁶, Domenico Girelli²⁷, Alex P. Reiner²⁸, Nathan O. Stitziel²⁹, Sekar Kathiresan³⁰, Stacey Gabriel², Jeffrey C. Barrett⁸, Terho Lehtimäki³¹, Markku Laakso³², Leif Groop⁵,¹⁵, Jaakko Kaprio⁵,³³,³⁴, Markus Perola⁵, Mark I. McCarthy³⁵,³⁶,³⁷, Michael Boehnke¹⁴, David M. Altshuler²,³, Cecilia M. Lindgren¹,²,³⁸, Joel N. Hirschhorn²,¹⁰, Andres Metspalu⁹, Nelson B. Freimer³⁹, Tanja Zeller¹⁸,¹⁹, Sirpa Jalkanen¹⁷, Seppo Koskinen⁴⁰, Olli Raitakari⁴¹,⁴², Richard Durbin⁸, Daniel G. MacArthur¹,²,³, Veikko Salomaa⁶, Samuli Ripatti⁵,⁶,⁸,³³,⁴³, Mark J. Daly¹,²,³⁹,⁎, Aarno Palotie¹,²,⁵,⁴⁴,⁴⁹,⁎ for the Sequencing Initiative Suomi (SISu) Project
There are proportionally more LoF variants in Finns.

**Effects of Bottleneck**

1. **Extremely rare variation is depleted:**
   Most rare variants do not make it through.

2. **Increase in low frequency damaging variants:**
   Surviving rare variants get a big frequency boost to (0.5-5%) in FINs.

3. **Boosted variants are more damaging:**
   And bottleneck is recent enough that selection has not eliminated them.

* FINs have significantly more variants than NFEs

Lim and Daly
Targeted LoF genotyping pilot in 35,000 Finns

- LoF SNVs and indels
- 83 LoF variants
- 35,000 population cohort w/ 73 medically relevant quantitative traits

Lim et al, PLoS Genetics, 2014
Traits studied include:

- LDL
- HDL
- TG
- BMI
- SBP
- DBP
- CRP
- HGF
- FGF
- VEGF
- GALECTIN3
- VitB12
- G_CSFR
- IL4, IL6, IL10
- D_DIMER

With reach through into complete medical records:

- Pilot study in 35,000 Finns

ICD-9 & ICD-10 diagnosis count (1986-2010); FINRISK 1992-2007; n= 29,286

ICD-10 converted to ICD-9

Peter Wurtz, March 11, 2013

![Diagram with medical data and disease history: Neoplasm of unspecified nature, Cerebral aneurysm, nonruptured, etc.]
## Two LPA LoF variants

<table>
<thead>
<tr>
<th>LoF variant</th>
<th>Frequency in Finns</th>
<th>Frequency in non Finns</th>
</tr>
</thead>
<tbody>
<tr>
<td>LPA1(4974)</td>
<td>2.8%</td>
<td>0.47%</td>
</tr>
<tr>
<td>LPA2(4289)</td>
<td>4.8%</td>
<td>3.6%</td>
</tr>
</tbody>
</table>

227 Finns LoF homozygotes
Two LPA LoF variants

<table>
<thead>
<tr>
<th>LoF variant</th>
<th>Frequency in Finns</th>
<th>Frequency in non Finns</th>
</tr>
</thead>
<tbody>
<tr>
<td>LPA1(4974)</td>
<td>2.8%</td>
<td>0.47%</td>
</tr>
<tr>
<td>LPA2(4289)</td>
<td>4.8%</td>
<td>3.6%</td>
</tr>
</tbody>
</table>

227 Finns LoF homozygotes

Linking to National Health Records

No sign of increased morbidity, protection for coronary heart disease
PROTECTIVE GENE VARIANTS
Traditional classifications need rethinking

- Bipolar disease
- Schizophrenia
- Epilepsy
- Migraine
- Major depression
- Cognitive impairment

Comorbidity

Shared cellular pathways
Type 2 Diabetes: >10,000 individuals

High Blood pressure

Severe mental health:
- schizophrenia, depression: >5,000 individuals

Migraine: 15,000 individuals

Old age dementia:
- ~5,000 individuals

Cancer: >10,000 cases

Cardiovascular events: stroke, CHD; 25,000 individuals

Cardiovascular risk factor data: 100,000 individuals

Life style and socio-economic data:
- education, economic state, smoking

Cardiovascular events
- Life course events

National Biobank: 201,858 individuals

18,000 statin users

20,000 estrogen substitution therapy

National Biobanks Finland portal
- 130,000 individuals from population cohorts
- 70,000 individuals from disease collections

National Biobank Finland portal:
- 130,000 individuals from population cohorts
- 70,000 individuals from disease collections

Registromics
Nordic population opportunities more generally

National biobank and registry solutions in Finland and Denmark
The Danish Civil Registration System

Unique person identifier used across all records from birth to death or emigration

The National Patient Register

The Psychiatric Central Register

The Register of Causes of Death

The Medical Birth Register

Redeemed prescriptions

First degree relatives, current and past spouses, residences in Denmark, immigration and emigration

Social factors including income, wealth, education, household crowding, marital status, type of job

Life events and stressors as unemployment, criminal records, custodial care etc. etc.

All data available for cases, controls/general population and relatives/spouses

Stanley Center for Psychiatric Research at Broad Institute

Broad Institute

Massachusetts General Hospital
Total dynamic population of Denmark
App 8 million. Can all be followed up in registers

Danish population born since 1955
Can be linked to first degree relatives,
Allowing construction of individual and familial medical
and social history over decades, etc.

Danish Neonatal Screening Biobank
All individuals born in Denmark since 1981. 2 million persons.
Can be linked to first degree relatives, and in many cases also
grandparents, uncles and aunts and cousins.
The neonatal blood spots allow GWAS, sequencing, epigenetics,
gene expression, metabolomics, IgG measurement, etc.

>15,000 diagnosed cases of autism
Different strengths

• Finland
  – Targeted case collection can be costly
  – Recontacting permitted
  – Broad set of registry data available for analysis
  – Targeted phenotyping possible

• Denmark
  – Targeted case collection straightforward
  – No recontacting
  – Broad set of registry data available for analysis
  – Only existing registry and clinical phenotyping
Population isolate

Bottleneck effect

bottleneck: catastrophic reduction in population

original population -> chance survivors -> new population

MAP
SISU-project
Sequencing Initiative Suomi (Finland)
The 200K

- Genome wide genotype data
  73,000
- Genome exome sequences
  >26,000

National Biobank

- 200,000 individuals
  4% of the population
- Population cohorts
- Extensive health, phenotype, metabolomic data
- Disease specific collections

Imputation

Population specific chip/genotyping

Reference database

Institutions involved:
- UK 10K
- Broad Institute
- Lund University Faculty of Medicine
- University of Oxford SPH
- Semel Institute UCLA
- FIMM
- TERVEYDEN JA HYVINVOINNIN LAITOS
- Washington University in St. Louis
- Massachusetts General Hospital
Sequencing Initiative Suomi
- Data resource for the research community

Data resource for the research community

The Sequencing Initiative Suomi (The SISu project) is an international collaboration between research groups aiming to build tools for genomic medicine.

This SISU v3.0 (2015-08-28) release includes:

- 1 037 122 sites with 1 137 703 variants
- 6118 Finnish samples from 10 cohorts that were sequenced in Broad and Sanger Institutes
- Original sequencing was done using three different platforms (Agilent 1.1 refseq plus 3 boosters, Agilent sureselect 50mb, Illumina coding v1)
- Multiallelic sites and indels are now included
- Finnish enrichment and other information fields for custom filtering purposes now available

Search

The SISu data resource currently covers exons only and data is restricted to autosomal SNPs and Indels. Genome build used in this release was GRCh37. Minor allele frequencies before QC and after QC provided.

Information about quality control process

⇒ Enter search
Human gene knockouts
Need for human models
Human gene knockouts
Need for human models
Population isolates

Bottle neck effect and genetic drift

original population

chance survivors

new population

bottleneck: catastrophic reduction in population
Nordic countries

- Similar health care systems
- Similar traditions in epidemiological studies
- Similar health registers
- Similar regulations
- Biobanks
- Positive attitude for research
Nordic countries

- Similar boarders between public service sectors
- Similar protectionism within each governmental institution
- Similar worries
- “It cannot/should not be done” attitude, hardwired in Finns…….
What is quickly needed

• Developing pan-Nordic principals for data sharing
• Resolving unnecessary regulatory hurdles
• Changing the attitude of registry agencies
• Getting the IRB process updated
The train moves persistently

- Do we want to miss the train?
- Would we like to capitalize on your past investments in biobanks and eHealth?
Clustering of the Finnish mutation in a Botnia village around the town of Jakobstad

Loss-of-function mutations in SLC30A8 protect against type 2 diabetes


○ Narpes 0.11%
(N=1,355)
Pilot study in 35,000 Finns

Traits studied include:

- LDL
- HDL
- TG
- BMI
- SBP
- DBP
- CRP
- HGF
- FGF
- VEGF
- GALECTIN3
- VitB12
- G_CS
- IL4, IL6, IL10
- D_DIMER

With reachthrough into complete medical records:

ICD-9 & ICD-10 diagnosis count (1986-2010); FINRISK 1992-2007; n= 29,286
ICD-10 converted to ICD-9
Peter Wurtz, March 11, 2013

Novel associations to:

* Reduced Lp(A) levels – and through this cardioprotection and, oddly, increased diabetes risk
  - Galectin3 levels
  - Triglyceride levels
  - Systolic blood pressure and several immune markers
  - D_DIMER levels
Elevated Lp(a) levels known to associate with elevated CHD risk

(Bennel JAMA In Med 2008)

Newly identified protective LoF, 5.5x enriched in Finland (MAF=2.8%)

1) Lower Lp(a) levels ($p=3 \times 10^{-58}$)
2) Lower CHD risk ($p=0.01$)

Lim et al PLoS Gen
Similar effect in several study samples (heterozygote)

### Association for LPA splice variants with cardiovascular diseases

<table>
<thead>
<tr>
<th>Study</th>
<th>Ncases n/N</th>
<th>Ncontrols n/N</th>
<th>Odds Ratio (95% CrI)</th>
<th>Odds Ratio (95% CrI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>FINRISK (CHD)</td>
<td>1076/25020</td>
<td>23944/25020</td>
<td>-</td>
<td>0.79 (0.72 to 0.86)</td>
</tr>
<tr>
<td>Estonian ExomeChip (IHD+HF)</td>
<td>768/4600</td>
<td>3832/4600</td>
<td>-</td>
<td>0.69 (0.31 to 1.50)</td>
</tr>
<tr>
<td>Estonian Imputed (IHD+HF)</td>
<td>853/7953</td>
<td>7100/7953</td>
<td>-</td>
<td>0.83 (0.51 to 1.36)</td>
</tr>
<tr>
<td>MIGEN ExA (MI)</td>
<td>8890/18176</td>
<td>9286/18176</td>
<td>-</td>
<td>0.88 (0.78 to 0.99)</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>11587/55749</strong></td>
<td><strong>44162/55749</strong></td>
<td>-</td>
<td>0.84 (0.80 to 0.88)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>x</th>
<th>x</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.5</td>
<td>1</td>
</tr>
</tbody>
</table>
Technology driving and supporting

Individual organ group approach

Systems approach

Change in disease classification
In October, we produced 1,350 human genomes (213 Tb total)!
One human genome every 32 minutes
The change is inevitable

- Special opportunities for the Nordic countries
- An opportunity for better stratified medicine
- An opportunity for improved cost efficiency and better use of resources
- An opportunity for new innovations and commercialization
THE COST AND SPEED OF GENOME SEQUENCING IS NO LONGER THE PROBLEM

MOORE’S LAW
ICD-9 & ICD-10 diagnosis count (1986-2010); FINRISK 1992-2007; n= 29,286
ICD-10 converted to ICD-9
Peter Würtz, March 11, 2013

<table>
<thead>
<tr>
<th>Diagnosis</th>
<th>Total hospitalizations</th>
<th>Person hospitalizations</th>
</tr>
</thead>
<tbody>
<tr>
<td>ICD-9 chapters</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Infectious And Parasitic Diseases</td>
<td>5293</td>
<td>3039</td>
</tr>
<tr>
<td>Neoplasms</td>
<td>17207</td>
<td>5295</td>
</tr>
<tr>
<td>Endocrine, Nutritional And Metabolic Diseases, And Immunity Disorders</td>
<td>7318</td>
<td>2805</td>
</tr>
<tr>
<td>Diseases Of The Blood And Blood-Forming Organs</td>
<td>1055</td>
<td>590</td>
</tr>
<tr>
<td>Mental Disorders</td>
<td>10653</td>
<td>2520</td>
</tr>
<tr>
<td>Diseases Of The Nervous System And Sense Organs</td>
<td>14279</td>
<td>5973</td>
</tr>
<tr>
<td>Diseases Of The Circulatory System</td>
<td>38019</td>
<td>8410</td>
</tr>
<tr>
<td>Diseases Of The Respiratory System</td>
<td>13547</td>
<td>5633</td>
</tr>
<tr>
<td>Diseases Of The Digestive System</td>
<td>13782</td>
<td>7270</td>
</tr>
<tr>
<td>Diseases Of The Genitourinary System</td>
<td>13363</td>
<td>6789</td>
</tr>
<tr>
<td>Complications Of Pregnancy, Childbirth, And The Puerperium</td>
<td>19134</td>
<td>5786</td>
</tr>
<tr>
<td>Diseases Of The Skin And Subcutaneous Tissue</td>
<td>2275</td>
<td>1264</td>
</tr>
<tr>
<td>Diseases Of The Musculoskeletal System And Connective Tissue</td>
<td>24108</td>
<td>8881</td>
</tr>
<tr>
<td>Congenital Anomalies</td>
<td>745</td>
<td>432</td>
</tr>
<tr>
<td>Certain Conditions Originating In The Perinatal Period</td>
<td>23</td>
<td>19</td>
</tr>
<tr>
<td>Symptoms, Signs, And Ill-Defined Conditions</td>
<td>13498</td>
<td>6911</td>
</tr>
<tr>
<td>Injury And Poisoning</td>
<td>14726</td>
<td>7009</td>
</tr>
<tr>
<td>Supplementary Classification Of Factors Influencing Health Status And Contact With Health Services</td>
<td>2002</td>
<td>1749</td>
</tr>
<tr>
<td>Supplementary Classification Of External Causes Of Injury And Poisoning</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Intestinal Infectious Diseases</td>
<td>1288</td>
<td>977</td>
</tr>
<tr>
<td>ICD-9 blocks</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tuberculosis</td>
<td>147</td>
<td>78</td>
</tr>
<tr>
<td>Zoonotic Bacterial Diseases</td>
<td>22</td>
<td>19</td>
</tr>
<tr>
<td>Other Bacterial Diseases</td>
<td>1856</td>
<td>1096</td>
</tr>
<tr>
<td>Human Immunodeficiency Virus</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>Poliomyelitis And Other Non-Arthropod-Borne Viral Diseases Of Central Nervous System</td>
<td>51</td>
<td>44</td>
</tr>
<tr>
<td>Viral Diseases Accompanied By Exanthem</td>
<td>204</td>
<td>158</td>
</tr>
<tr>
<td>Arthropod-Borne Viral Diseases</td>
<td>29</td>
<td>24</td>
</tr>
</tbody>
</table>
Example of health histories from two persons from the national biobanks with a 40 year follow-up
Each gene variant has a small effect

Reference database like Reference values in clinical chemistry
National Genome Strategy

The Biobank act

PAREMPPAA TERVEYTTÄ
GENOMITIEDON AVULLA

Kansallinen genomistrategia
Työryhmän ehdotus

SOSIAALI- JÄ
TERVEYSMINISTERIÖ

FI MM

BROAD INSTITUTE

MGH 1811

MASSACHUSETTS GENERAL HOSPITAL
Collaborate. Innovate. Accelerate.
Working together to share knowledge, create networks and accelerate advances in genomics and health.

“Internet of genomics and health”
Distributing and sharing data globally
Common data formats
Connecting parties globally
150 members worldwide
An alternative view for research ethics, moving from the shadow of World War II to the next step

- In particular, it highlights, and is guided by, Article 27 of the 1948 *Universal Declaration of Human Rights*. Article 27 guarantees the rights of every individual in the world “to share in scientific advancement and its benefits” (including to freely engage in responsible scientific inquiry), and at the same time “to the protection of the moral and material interests resulting from any scientific…production of which [a person] is the author”.
Biobanks

Finnish Genome Strategy
Ministry of Health and Sitra

Genomes

SI Su
Sequencing Initiative Suomi

Implementation in health care
Biobank

Genome data

Implementation in health care

Implementation pilots

BBMRI nordic