European populations and the postgenome era

Leena Peltonen
Academy Professor
University of Helsinki and National Public Health Institute, Finland
Genome Era

- Over 160 genomes completely sequenced in databases
- These include human genome
Human Genome Project

22 000 protein coding genes
(www.ensembl.org/Homo_sapiens)

Over 10 million SNPs (> 1% frequency), 7 million catalogued
(www.ncbi.nlm.nih.gov/SNP)

More than 1400 genes correlated directly with the disease
(www.ncbi.nlm.nih.gov/Entrez)
Until now:

Validate hypothesis by serial application of diverse experimental approaches to one or a few genes/proteins.

After genome projects:

Generate hypothesis using one or few parallel high-throughput approaches to obtain data on large group of genes/proteins.
Identification of Mutated Genes

Monogenic Diseases
  1580 disease phenotypes
  1270 mutated genes

Common Diseases
  Mostly rare high impact genes
Genetic Traits

Simplex or monogenic

- Disease gene
- Modifying gene
- Phenotype

Complex or multifactorial

- Susceptibility gene
- Phenotype

- Susceptibility gene
- Phenotype

- Susceptibility gene
- Environment Life Style etc.
Genomics era in Biomedicine

- For the first time in human history we can produce a high-resolution picture of our individual genomes and monitor for changes in diseases.

- For the first time the role of genetic and life-style risk factors can be defined.

- Special European competitive advantage of in biomedical research can be utilized in this historical era.
Ultimately it should be possible

Examine individual’s genetic make-up at any position of the sequence

Deduce functional consequences and relate them to other risk factors

Make a well-informed choices of medical actions
Genetic Information

- The basis of modern medicine
- Diagnostics
  - New diagnostic classification
- Environmental effect
  - Molecular pathogenesis
- Novel therapeutics & prevention
European Strengths

Good, equal education

Equal health care

Developed social infrastructure

Developed information technology and networks
Identification of Complex Disease Genes in Isolated Populations

- Higher degree of genetic homogeneity
  - Fewer mutations in disease genes
  - Mutations technically easier to identify (LD)

- Higher degree of environmental homogeneity
  - Life style, diet, culture
  - Training of physicians, clinical practice
Population Isolates

- Founder effect
- Genetic drift
- Isolation
All affected individuals sharing the same ancestor have a high probability to share the same (founder) disease allele.

The DNA-region flanking the mutation is also identical, reflecting the appearance of the founder chromosome.
Case of Finland

High quality national health care

- Reliable healthcare registers
- Genetically homogeneous population
- High quality epidemiology and mathematics
- Population registers
- Internationally recognized genetic research
- Top expertise in information technology

Equal high quality education
Finland -
One of the best characterized populations for disease mutations

- Founder Effect
- Genetic Drift
- Isolation
- Regional Expansion

- Enrichment of Rare Diseases
- Fin-Major mutation
- Population records since 1634
- Registers of one payer health care system
- "Inbred" training of clinicians
- Favorable attitudes by public
Finnish Disease Database
Finland Chip

- \( \text{HTI}_{\text{Fin}} \) carrier
- \( \alpha_1 \text{AT}_{z}\text{-allele} \) carrier
- \( \text{HFE}_{C282Y} \) carrier
- \( \alpha_1 \text{AT}_{z}\text{-allele/} \) 
  \( \text{CNF}_{\text{major/}} \) 
  \( \text{LCHAD}_{G1528C} \) carrier
- \( \text{DTD}_{\text{Fin/}} \) 
  \( \text{LPI}_{\text{Fin}} \) carrier
- \( \text{GJB2}_{35\Delta G} \) carrier
DNA-Chip for population screening

2400 DNA-samples analyzed for 31 disease mutations on the chip

- Prevalence of recessive mutations
- Regional variations
- Feasibility for large screening programs
Who carries a mutation?

Finnish diseases: 1:10
Any of 31: 1:3

Finnish: 1:8
Any: 1:3
General lessons

- Disease mutations are common (1:3) when monitoring for only 31 mutations.

- If 22,000 mutations would be monitored,

  Every individual carries multiple mutations.
Genetic Traits

Simplex or monogenic

- Disease gene
- Modifying gene
- Phenotype

Complex or multifactorial

- Susceptibility gene
- Susceptibility gene
- Susceptibility gene
- Environment, Life Style etc.
- Phenotype
Genetics can dramatically help in our understanding of common, complex diseases

Identification of key pathways involved

- Find the gene in rare families
- Pick the allelic markers
- Test them for association in case/control samples
- Analyze multiple populations
- Determine the population attributable fraction and significance in epidemiological cohorts
Common Trait Genome Scans in Finnish study samples

- Multiple sclerosis
- Schizophrenia
- Combined hyperlipidemia
- Low-HDL
- Hypertension
- Osteoarthritis, small joint
- Obesity&BMI
- Stature
- Migraine
- Autism
- Lactose intolerance

- Asthma
- Dyslexia
- Crohn’s syndrome & ulcerative colitis
- Familial ventricular tachycardia
- Coronary artery disease
- Type II Diabetes
- Pre-eclampsia
- Psoriasis

Most loci replicated in other populations
Complex Diseases

Identification of a DNA variant in patients / families

Epidemiological studies

Functional studies
Disease Mutations

Fully penetrant monogenic disease
- Coding region mutations
cSNPs

Partially penetrant polygenic disease
- Non-coding and regulatory region mutations
rSNPs
Common disease study samples

Multiplex families
Ascertertainment bias for rare, high impact genes
To identify defective pathways

Case–control samples
“Mixed bag” of rare and common genes,
To verify the relevance of variants

Epidemiological cohorts
To define the size of the effect of genetic versus environmental risk factors
Linkage analysis:

- large pedigrees with several generations
- LOD - Scores
- AP - Analysis

Affected Sib-Pairs
ASP-Analysis

Association analysis with family controls

Case-control analysis with genomic controls
Different study samples meaningful for different aims

For identification of novel genes and metabolic pathways
- Families, sibpairs, ascertained for a given trait (disease)
- Case control samples

For risk impact estimation
- For new diagnostic entities
- Epidemiological study samples with excessive amount of health care and life style information
Necessity of large numbers

Current diagnostic classification of diseases does not reflect the molecular background
Disease alleles of common diseases are probably old and have a wide diversity

We need huge ascertained and non-ascertained populations samples from several populations to:
1) identify disease predisposing variants
2) verify their significance
LARGE BIOBANKS

Estonian Genome project (Egeen)
DeCode (Iceland)
Carthagen (Canada)
UK Biobank
Swedish National Biobank Program (Swegene, Wallenberg Consortium North)
Genomeutwin cohorts
NIH Prospective cohort based population study
Biobanks and Finland

Biobank can be a trash bank without detailed clinical and epidemiological data, DNA:s are worth of nothing

Most biobanks will be useful 10-20 years from now

Finland could start from ”the other end”, not from biobanks but from epidemiological data collections

Benefits from epidemiological samples for society materialize relatively soon
Examples of epidemiological study samples in Finland

<table>
<thead>
<tr>
<th></th>
<th>Size</th>
<th>Consented DNA-samples</th>
</tr>
</thead>
<tbody>
<tr>
<td>For cardiovascular traits</td>
<td></td>
<td></td>
</tr>
<tr>
<td>FINRISK 92</td>
<td>8000</td>
<td>5600</td>
</tr>
<tr>
<td>97</td>
<td>10000</td>
<td>8700</td>
</tr>
<tr>
<td>02</td>
<td>11000</td>
<td>10 000</td>
</tr>
<tr>
<td>For diabetes</td>
<td>6000</td>
<td>2000</td>
</tr>
<tr>
<td>For autism</td>
<td>2000</td>
<td>2000</td>
</tr>
<tr>
<td>For psychosis</td>
<td>3000</td>
<td>3000</td>
</tr>
<tr>
<td>Total:</td>
<td></td>
<td>~ 32 000</td>
</tr>
</tbody>
</table>
### Examples of Finnish Population Cohorts

<table>
<thead>
<tr>
<th>Cohort Type</th>
<th>Size</th>
<th>Consented DNA-samples</th>
</tr>
</thead>
<tbody>
<tr>
<td>Twin cohort</td>
<td>170 000</td>
<td>27 000</td>
</tr>
<tr>
<td>Health 2000 cohort</td>
<td>11 500</td>
<td>11 500</td>
</tr>
<tr>
<td>(National health study)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Northern Finland</td>
<td></td>
<td></td>
</tr>
<tr>
<td>cohort 66</td>
<td>12 000</td>
<td>11 000</td>
</tr>
<tr>
<td>cohort 86</td>
<td>9 500</td>
<td>9 000</td>
</tr>
<tr>
<td><strong>Total:</strong></td>
<td>~ 60 000</td>
<td></td>
</tr>
</tbody>
</table>
Complex Diseases

- Gene-gene interactions
- Environment
- Protein-protein-interactions
- Life style
- Genome studies
Genetic Predisposition of Common Diseases

A common allele
- modest relative risk
- little familial recurrence
- high population attributable fraction

A rare allele
- high relative risk
- much familial recurrence
- low population attributable fraction
The strength of Finnish epidemiological cohorts

- Solid epidemiological criteria have been used to collect the study samples
- Excessive amount of life style and health-related data has been collected
- Possibility for longitudinal studies
- All these features do not exist in current biobanks
European niche in Biomedicine

- Reliable health care infrastructure
- High quality, equal education
- Top level expertise in genetics, epidemiology, clinical medicine and mathematics

Unique possibilities in health care-related genome research and it’s rapid implementation in health care
Key issues

Expertise in epidemiology
Expertise in clinical medicine
High quality biological samples collection, storage and database system
High throughput genotyping and sequencing centers
Expertise in biocomputational analyses
Access to multiple different population samples
Integration of human studies with studies in experimental species
Attractive environment for top scientists worldwide
Nordic Countries: 24 million

- National health care system
- Reliable healthcare registers
- Accurate population registers
- Traditions in genetic research
- Traditions in epidemiology and mathematics
- Expertise in information technology
- Equal, high quality education
- Some isolated populations
Nordic Center of Excellence in Disease Genetics: Partners

Stockholm, Sweden
KI

Helsinki, Finland
UH, NPHI, Folkhälsan

Uppsala, Sweden
University of Uppsala

Lund, Malmö, Sweden
Wallenberg Lab., Univ. f Lund

Århus, Denmark
Århus University Hosp.
GENOMEUTWIN
(genomeutwin@org)

Genome-wide analyses of European twin and population cohorts to identify genes predisposing to common diseases
One of the large EU Genomics Centers, co-ordinated by Finland

- 8 countries, 800 000 twin pairs
- 13.4M €

To use European twin and populations cohorts to study genetic and lifestyle risk factors of common traits

- stature, BMI, CHD, stroke, migraine,
### Twin cohorts
- Australian twins
- Danish twins
- Finnish twins
- Italian twins
- Dutch twins
- Norwegian twins
- UK twins
- Swedish twins

### Intellectual core facilities
- DNA isolation and genotyping (Uppsala, Helsinki)
- Epidemiological expertise (Odense)
- Database expertise (Stockholm)
- Biocomputing expertise (Leiden)
- Ethical and legal expertise (Oslo)
Genetic Information

- The basis of modern medicine
- Diagnostics
  - New diagnostic classification
- Environmental effect
  - Molecular pathogenesis
- Novel therapeutics & prevention
Genome Information Center

Epidemiological databases

Sample Collections

National Registers

Host Institute 1

Founding Members:
Universities, TEKES, Sitra, VTT, KTL, Academy of Finland, ...

(Non-profit) Genome Information Center

National Genome Research Center

EMBL satellite?

Big Pharma Corp.

Biotech Oy / Ltd