PopGen
Population Genetics in Schleswig-Holstein

BMBF 01 GS 0121

Dr. Susanna Nikolaus, Prof. Dr. Stefan Schreiber
Department of Internal Medicine, Christian-Albrechts-University, Kiel, Germany

Prof. Dr. Michael Krawczak
Institute of Medical Informatics and Statistics, Christian-Albrechts-University, Kiel, Germany
Single-Gene Disorders
“rare - single gene - major effect”

genetic variant

trait

environment genetic background

Cystic Fibrosis
Huntington Disease
Duchenne Muscular Dystrophy
Neurofibromatosis
Tuberous Sclerosis
Complex Disorders
“frequent - multiple genes - minor effects”

Diabetes
Coronary Heart Disease
Neuropsychiatric Disease
Cancer
Autoimmune Disease
The Need for Genetic Epidemiology

disease is context-dependent
“gene-environment interaction”

genic structure phenic structure

refined risk assessment
characterisation of (defective) gene products
understanding pathophysiology
cure/therapy
disease-oriented networks
- Cancer
- Heart Disease
- Infection and Inflammation
- Nervous System
- Environmental Disease

technological platform
- Proteomics
- Bioinformatics
- Centres of Expertise in Genetic Epidemiology (GEMs)
Centres of Expertise in Genetic Epidemiology (GEM)

- data analysis
  - execution
  - interpretation
  - improvement

- data basing
  - structuring
  - implementation
  - integration
  - meta analyses

- quality management
  - planning
  - monitoring
  - recruitment
  - data validation
  - data integrity
  - communication

- training and education
PopGen - The Objectives

- to ascertain a geographically confined, representative sample of \( \sim 15,000 \) patients affected by at least one of eight predefined disease phenotypes

- to ascertain a representative sample of \( \sim 10,000 \) randomly selected control individuals of the same geographic and ethnic origin

- to establish a phenotype database (including data from clinical follow-up) comprising all participating individuals

- to establish a DNA bank (600 - 1000 µg DNA) of all participating individuals

- to make all phenotype and genotype information available to other NGFN members studying the genetic basis of the same disease phenotypes
Rationale

obstacles to complex disease genetics
- genetic heterogeneity
- weak marginal effects
- phenocopies

possible solutions
- genetic isolates (e.g. Finland, Iceland, Amish)
- extreme sampling (e.g. extremely concordant sibs)
- subphenotypes (e.g. familiar breast cancer)

i.e. samples useful for genetic analysis are unlikely to be representative of a population-wide patient spectrum!
Rationale (2)

Rationale (3)

BRCA1 (17q21): ~ 400 mutations, breast and ovarian cancer
BRCA2 (13q): ~ 250 mutations, breast and male breast cancer

<table>
<thead>
<tr>
<th>Patient's History</th>
<th>Family History (Includes at least one first or second degree relative)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>No breast cancer &lt;50, or ovarian cancer, in any relative.†</td>
</tr>
<tr>
<td></td>
<td>Breast cancer &lt;50 in one relative; no ovarian cancer in any relative.</td>
</tr>
<tr>
<td></td>
<td>Breast cancer &lt;50 in more than one relative; no ovarian cancer in any</td>
</tr>
<tr>
<td></td>
<td>relative.</td>
</tr>
<tr>
<td></td>
<td>Ovarian cancer at any age in one relative; no breast cancer &lt;50 in any</td>
</tr>
<tr>
<td></td>
<td>relative.</td>
</tr>
<tr>
<td></td>
<td>Ovarian cancer in more than one relative; no breast cancer &lt;50 in any</td>
</tr>
<tr>
<td></td>
<td>relative.</td>
</tr>
<tr>
<td></td>
<td>Breast cancer &lt;50 and ovarian cancer at any age.††</td>
</tr>
</tbody>
</table>

|                                                        | No breast cancer or ovarian cancer at any age | 2.9% | 4.2% | 9.8% | 5.8% | 8.7% | 16.7% |
|                                                        | Breast cancer ≥ 50                          | 3.2% | 8.3% | 11.4%| 7.4% | 9.8% | 19.8% |
|                                                        | Breast cancer <50                           | 7.8% | 17.8%| 31.0%| 16.7%| 31.2%| 44.5% |
|                                                        | Male breast cancer                         | 20.4%| 23.8%| 50.0%| 0%* | None Tested | 100%* |
|                                                        | Ovarian cancer at any age, no breast cancer | 11.9%| 29.3%| 38.8%| 24.7%| 32.2%| 51.4% |
|                                                        | Breast cancer ≥50 and ovarian cancer at any.| 17.6%| 21.1%*| 43.8%*| 18.2%*| 44.4%*| 50.0%* |
|                                                        | Breast cancer <50 and ovarian cancer at any| 32.0%| 56.7%| 72.2%*| 58.8%*| 62.5%*| 81.3% |

† May include families with breast cancer ≥50 (in women or men).
†† Includes family members with either or both diagnoses.

Number of observations in Table 1 is 10231
*N<20

www.myriadtests.com
Other Studies

“... to identify the common factors [...] that contribute to CVD by following its development over a long period of time in a large group of participants who had not yet developed overt symptoms of CVD [...]

“... provides the GSF with the unique opportunity to perform population-based health research in Germany[...] in cooperation with partners within the GSF and from other research institutes and universities.”

“The UK Biobank

“A long term research study aimed at establishing how genes, lifestyle and environmental factors interact to affect people's health.”
Other Studies (2)

The UK Biobank
A study of genes, environment and health

Funding: £60 million

“Picking apart this complexity requires a study to be on a huge scale. In time, it will help us to understand:

- the influence of specific genes or genetic variations on the development or severity of disease
- the number of people who carry a particular genetic variation and hence are at greater or lesser risk of disease
- the way in which interactions between genes and environmental factors are involved in the cause of disease”

www.biobank.ac.uk

“It is the equivalent of planting the shade trees for the next generation, and as such is inevitably going to be criticised by those that think there are more pressing short term needs.”

Dr. John Newton, Manchester, CEO Biobank UK
The Problems of “Bio-Banking”

- all exposures of interest have to be decided upon before commencement of the study
- medical progress implies that incidence rates are likely to decrease during the course of the study
- the advantage of exposure measurement over proband recall depends upon the actual exposure of interest
PopGen Catchment Area

ca. 520,000 inhabitants
6 hospitals
380 private medical practitioners
**Population stratification**

Problem: a population comprises \( n \) subpopulations

\[ f_{i,j} (g_{i,j}): \text{frequency of marker genotype } j \text{ among controls (patients) in subpopulation } i \]

\[ \alpha_i (\beta_i): \text{probability of sampling controls (patients) from subpopulation } i, \sum_i \alpha_i = \sum_i \beta_i = 1 \]

\[ f_j (g_j): \text{frequency of marker genotype } j \text{ among controls (patient) in the total sample, i.e. } f_j = \sum_i \alpha_i f_{i,j} \text{ (} g_j = \sum_i \beta_i g_{i,j} \text{)} \]

\[ f_j \neq g_j \rightarrow f_{i,j} \neq g_{i,j} \text{ or } \alpha_i \neq \beta_i \text{ for at least one } i \]

Example: The “chopstick trait” and ABO

<table>
<thead>
<tr>
<th>trait</th>
<th>A</th>
<th>B</th>
<th>0</th>
<th>AB</th>
</tr>
</thead>
<tbody>
<tr>
<td>no</td>
<td>0.41</td>
<td>0.11</td>
<td>0.43</td>
<td>0.05</td>
</tr>
<tr>
<td>yes</td>
<td>0.32</td>
<td>0.26</td>
<td>0.34</td>
<td>0.08</td>
</tr>
</tbody>
</table>

A “representative” sample from the US American population

Relative risks:
- \( A,0:1.0 \)
- \( AB:2.1 \)
- \( B:3.0 \)
Population stratification (2)


- k=3
- k=4
- k=5
- k=6

377 microsatellites
1056 individuals
52 populations

model: k subpopulations
Minor allele frequency of 67 non-coding SNPs in approximately 720 individuals each from KORA and PopGen

\[ \hat{F}_{ST} = 0.000; \ 95\% \ CI: \ 0.000 - 0.001 \]
\[ \hat{F}_{IT} = \hat{F}_{IS} = 0.007; \ 95\% \ CI: \ -0.001 - 0.015 \]
Minor allele frequency of **68 coding SNPs** in approximately 720 individuals each from KORA and PopGen

\[ \hat{F}_{ST} = 0.000; \text{ 95\% CI: 0.000 - 0.000} \]

\[ \hat{F}_{IT} = \hat{F}_{IS} = 0.006; \text{ 95\% CI: -0.001 - 0.012} \]
## Disease Phenotypes

<table>
<thead>
<tr>
<th>Disease</th>
<th>Age group</th>
<th>Prevalence in age group</th>
<th>Expected no. cases Total</th>
<th>Expected no. cases Recruited</th>
</tr>
</thead>
<tbody>
<tr>
<td>inflammatory bowel disease</td>
<td>18-80</td>
<td>0.3%</td>
<td>1560</td>
<td>780</td>
</tr>
<tr>
<td>irritable bowel syndrome</td>
<td>18-50</td>
<td>3%</td>
<td>8270</td>
<td>2000</td>
</tr>
<tr>
<td>coronary heart disease</td>
<td>18-40</td>
<td>sampling at 4 treatment centres</td>
<td>2500</td>
<td>2000</td>
</tr>
<tr>
<td>juvenile periodontits</td>
<td>18-40</td>
<td>0.4%</td>
<td>730</td>
<td>360</td>
</tr>
<tr>
<td>seizures</td>
<td>18-40</td>
<td>0.3%</td>
<td>550</td>
<td>400</td>
</tr>
<tr>
<td>juvenile atopic disease</td>
<td>18-30</td>
<td>3%</td>
<td>2340</td>
<td>700</td>
</tr>
<tr>
<td>chronic obstructive pulmonary disease</td>
<td>18-80</td>
<td>5%</td>
<td>12200</td>
<td>4100</td>
</tr>
<tr>
<td>tremor</td>
<td>18-80</td>
<td>2%</td>
<td>10400</td>
<td>3000</td>
</tr>
<tr>
<td>+ gallstones + longevity + ...</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Contact and Recruitment

- **patient identification** through public medical infrastructure (i.e. health insurance, treatment centres, self-aid groups)

- **patient contact** through telephone or mailing, addressed to their primary health care provider

- **control individuals** identified through population registry (1000 individuals per sex per one of five age groups)

- **direct response** to sampling team at Kiel University

- **questionnaire** and **blood sampling kit** sent to responders, plus request for permission to obtain detailed clinical data

- medical details documented using **standard schemes**

- telephone **help-line** (0800 number)
Work Plan and Time Schedule

- Staff recruitment, definition of disease phenotypes, questionnaire
- Identification of patients and control individuals
- Collection of clinical data and biological sample
- Enrolment in follow-up program

- Practices and outpatient offices are being visited to identify patients from treatment records and diagnoses on file
- Verification of diagnoses on the basis of available documentation
- Regular follow-up of 50% of patients through yearly contact to their primary health care provider
CHD (Coronary Heart Disease)

Inclusion Criteria
- diagnosis confirmed by coronary angiography (>50% coronary stenoses)
- diagnosis between January 1997 and April 2003
- age $\leq 55$ years at coronary angiography
- males and females
- place of residence north of Nord-Ostsee-Kanal

25,000 patient records scrutinised
3,200 patients identified as matching PopGen criteria
2,100 patients agreed to participate
300 questionnaires and blood samples obtained
CHD (Coronary Heart Disease)

Participating Institutions

- Klinik für Kardiologie, UK-SH Campus Kiel
- Kardiologische Gemeinschaftspraxis Kiel
- Klinik für Innere Medizin Diakonissen Krankenhaus Flensburg
- Klinik für Innere Medizin Kreiskrankenhaus Rendsburg
- Klinik für Innere Medizin Martin-Luther-Krankenhaus Schleswig
- Klinik für Innere Medizin Westküstenklinikums Heide
CHD (Coronary Heart Disease)

Clinical Information Requested

- arterial blood pressure
- lipid status
- glucose parameters
- family history
- smoking history
- cardial function
- ECG
- course of disease (e.g., coronary angioplasty, etc.)
- severity of disease (e.g., infarction, bypass surgery, etc.)
Data Protection and Privacy

PopGen
- personal data (ID1)
- phenotypes (ID1)
- demography (ID1)
- phenotypes (ID2)
- demography (ID2)
- labcode (ID1)
- labcode (ID2)

Laboratory
- labcode (ID1)
- temporarily only
- genotypes (labcode)
- labcode (ID1)

Trustee
- ID1 (ID2)

LIMS
- labcode (ID2)
- phenotypes (ID2)
- genotypes (ID2)
- demography (ID2)

- labcode (ID2)
- phenotypes (ID2)
- demography (ID2)

- IMIS
- 1.Med
- RZ UK-SH
... Think Big!