

# **Nordic twin registries and cohorts**

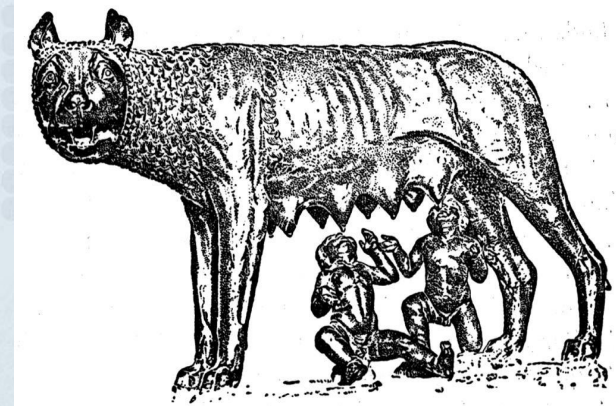
Jaakko Kaprio  
University of Helsinki &  
National Public Health Institute, Helsinki

**Nordic Research Seminar on  
Biobanking**  
**August 26-28 2006, Clinical  
Research Centre, Malmö, Sweden**



# The Classic Twin Study

- Monozygotic (MZ) pairs are genetically alike
- Dizygotic (DZ) pairs, like siblings, share on average half of their segregating genes
- DZ pairs can be same-sexed or opposite-sex (male-female)
- Increased similarity of twin pairs compared to unrelated subjects suggests familial factors
- Increased similarity of MZ pairs compared to DZ pairs provides evidence for genetic factors



# Twin study designs

- Classical MZ–DZ: These studies estimate the contributions of genetic and environmental effects to phenotypic variance, and test, for example, for age, cohort and sex differences in gene expression
- Multivariate analyses to model environmental and genetic correlations between traits to study
  - direction of phenotypic causality
  - causes of co-morbidity of two or more traits
  - multivariate modelling to obtain genotypic (or environmental) values for individuals
  - analysis of longitudinal data to study causes of phenotypic stability and tracking over time
  - testing of  $G \times E$  using measured environmental indices

# Examples

Paul Lichtenstein, Niels Holm, Pia K Verkasalo, Anastasia Illiadou, Jaakko Kaprio, Markku Koskenvuo, Eero Pukkala, Axel Skytthe, Kari Hemminki.  
Environmental and heritable components of cancer: joint analyses of three Nordic twin cohorts. *New Engl J Med* 2000; 343:78-85.

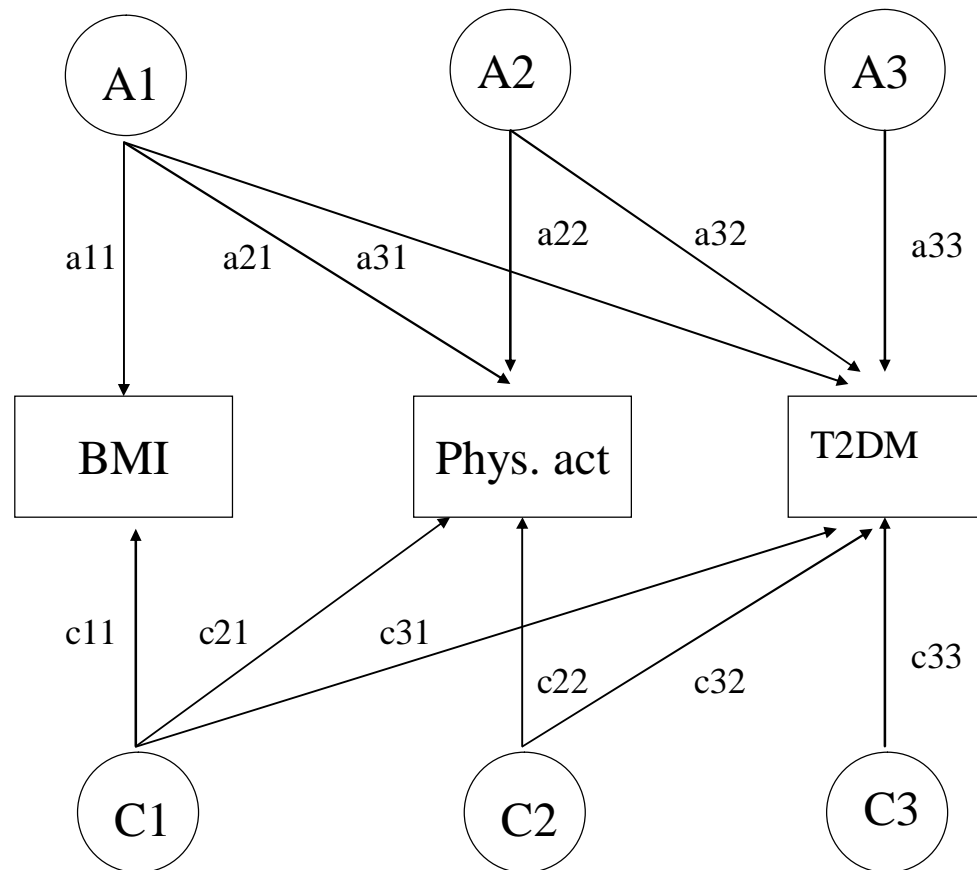
*Hum Genet* (2006)  
DOI 10.1007/s00439-006-0144-y

## ORIGINAL INVESTIGATION

Jacob vB. Hjelmberg · Ivan Iachine · Axel Skytthe  
James W. Vaupel · Matt McGue · Markku Koskenvuo  
Jaakko Kaprio · Nancy L. Pedersen · Kaare Christensen

**Genetic influence on human lifespan and longevity**

# Multivariate (cholesky) model



# Extended twin study

- Parents of twins can be included to study cultural transmission and  $G \times E$  covariance
- Parents of twins can be studied in a quasi-longitudinal design to determine genetic and environmental stability
- Assortative mating can be studied if spouses of twins are included
- Social interactions and special twin effects, such as prenatal hormone transition, the 'private language' of twins and shared prenatal environment, can be studied if siblings of twins are included
- Maternal effects,  $G \times E$  correlation and imprinting can be studied if offspring of MZ twins are included

# Twin epidemiological studies: Co-twin control study

- Case–control studies of MZ twins who are perfectly matched for genes and family background; provides tests of causal hypotheses from epi studies of individuals
- Can also be used to study gene expression and epigenetic effects in discordant twins
- Comparisons of DZ and MZ unaffected cotwins of cases to distinguish between genetic liability and state effects

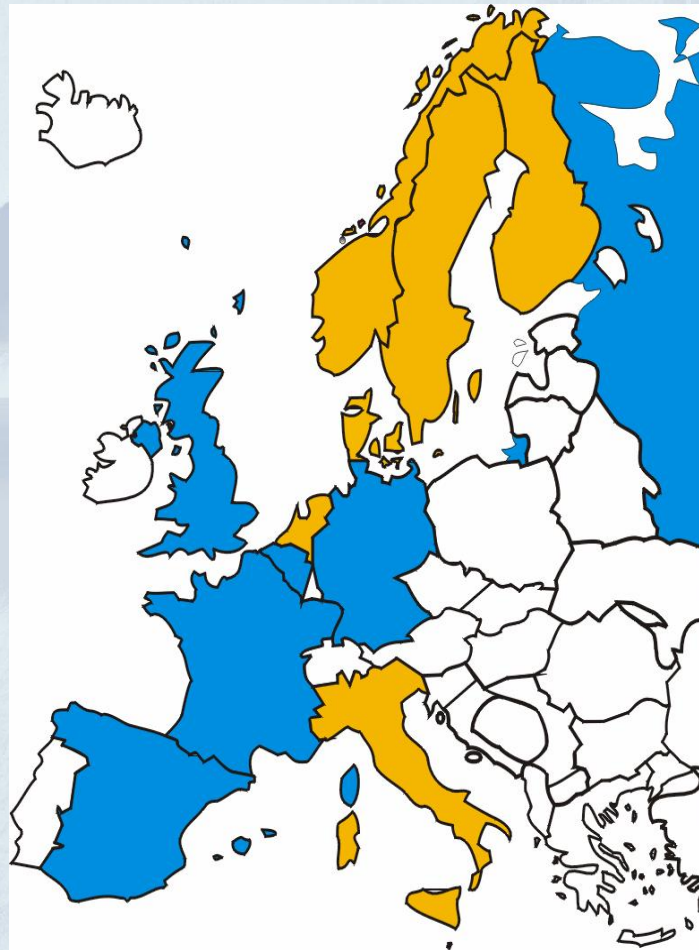
# Height and CHD

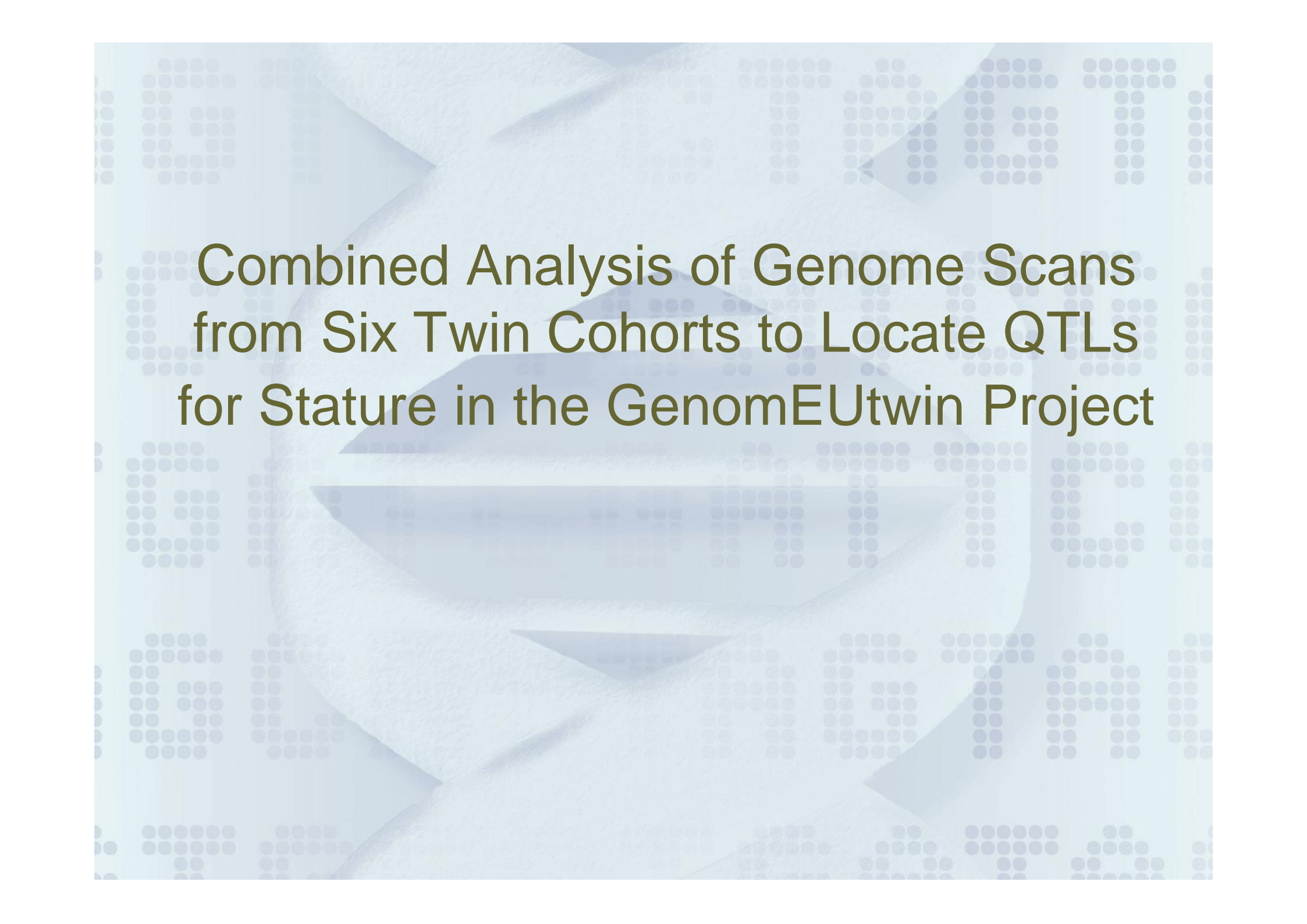
- Association between height and coronary heart disease mortality: a prospective study of 35 000 twin pairs
- Karri Silventoinen et al, Am J Epidemiol 2006
- Finnish, Swedish and Danish twin data with almost 6000 CHD deaths during 2 million person-years.
- An inverse association between height and CHD mortality was found even within monozygotic discordant twin pairs
- Suggests that this association is because of environmental factors which directly affect height and CHD risk.

# Genotyping and twins

- Genotyping at candidate loci
  - genotyping of MZ twins to detect variability genes and to estimate penetrance
  - genotyping of DZ twins to estimate associations within and between families
- Genotyping at marker loci
  - genotyping of DZ twins (and parents) to detect linkage with QTL (quantitative trait locus)
  - selecting informative families from large twin registers to find QTL of small effects

# European Populations studied in GenomEUtwin





Combined Analysis of Genome Scans  
from Six Twin Cohorts to Locate QTLs  
for Stature in the GenomEUtwin Project

# Stature is inherited

- A classical example of a complex trait
- Large geographical differences in mean but not in variance
- DZ taller than MZ
  - Reason? Placenta? Assortative mating?
  - Mothers of DZs taller?

# Stature i.e. Adult Body Height

- Stature is estimated to be at least oligogenic thus it may serve as a model phenotype for other complex phenotypes
- Easily and accurately determined by measurement, also self-reported questionnaire data is highly accurate, correlation  $>0.93$
- Heritability estimates from twin studies 0.87-0.93 and genome scans 0.69-0.98
- Previous genome scans have implicated loci on 3p, 6q and 7q along with many others

(Silventoinen et al. 2003)

# Heritabilities in the GenomEUtwin-project

	Males	Females	
	h <sup>2</sup>	h <sup>2</sup>	Age effect
Australia	0.87	0.71	Yes
Denmark	0.84	0.73	no
Finland	0.89	0.70	no
Italy	0.94	0.95	n.a.
The Netherlands	0.90	0.90	no
Norway	0.77	0.79	n.a.
Sweden	0.77	0.89	yes
UK	n.a.	0.68	yes

Silventoinen et al. Twin  
Research 2003

# Motivation for the Study

- Sample size is one of the most critical determinants of successful gene mapping
  - Pooling data from already conducted genome scans highly cost-efficient
  - Data on stature and BMI are collected for most genome scans and often correlate weakly with original target trait
  - Contemporary bioinformatics tools provide sufficient means for data pooling

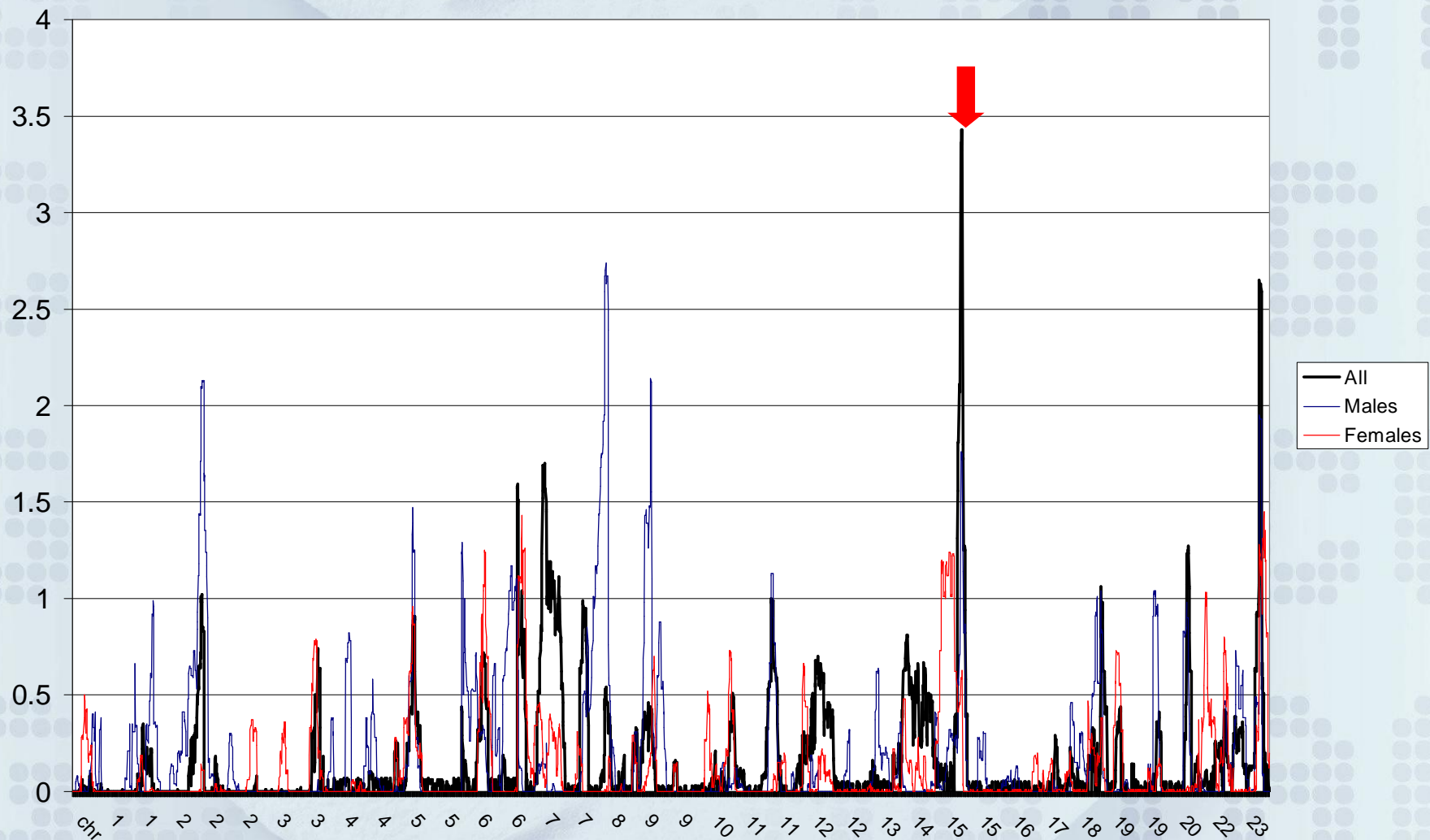
# Samples (with geno+phenotype)

Cohort	Individuals	Families
Australia	2609	1091
Netherlands 1	840	214
Netherlands 2	282	150
Denmark	628	315
Finland	857	387
Sweden	102	51
United Kingdom	2238	1119
Total	7556	3130

# Methods Used

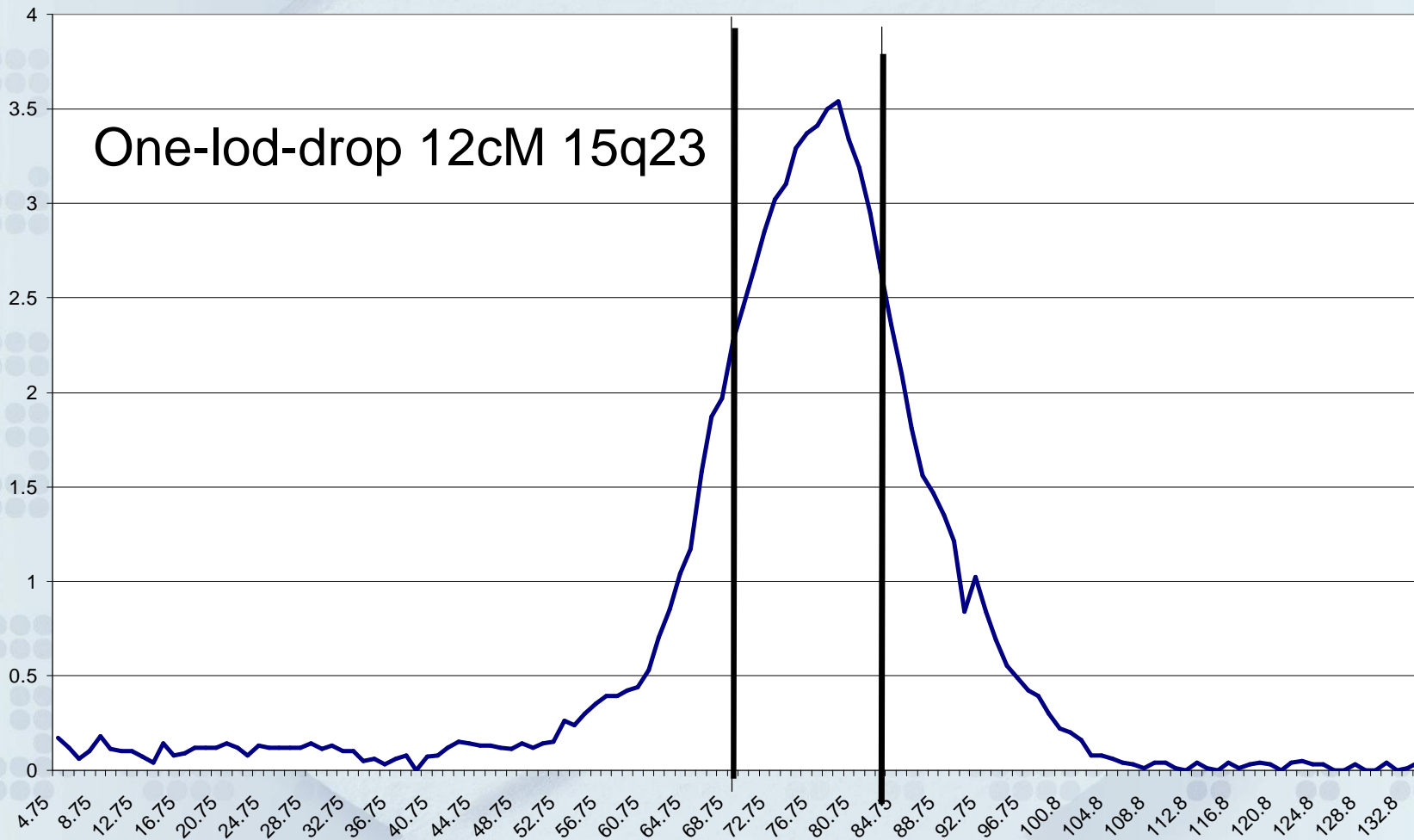
- Phenotypes of under 20-year old individuals discarded since they may still grow in height
- Individuals with phenotype deviating 4 SD from sex-specific mean or more discarded as outliers
- Age, sex and sample label used as covariates
- Variance-components linkage analysis with MERLIN
- In-house software for pooling data and automating analyses

# Whole Genome Scan

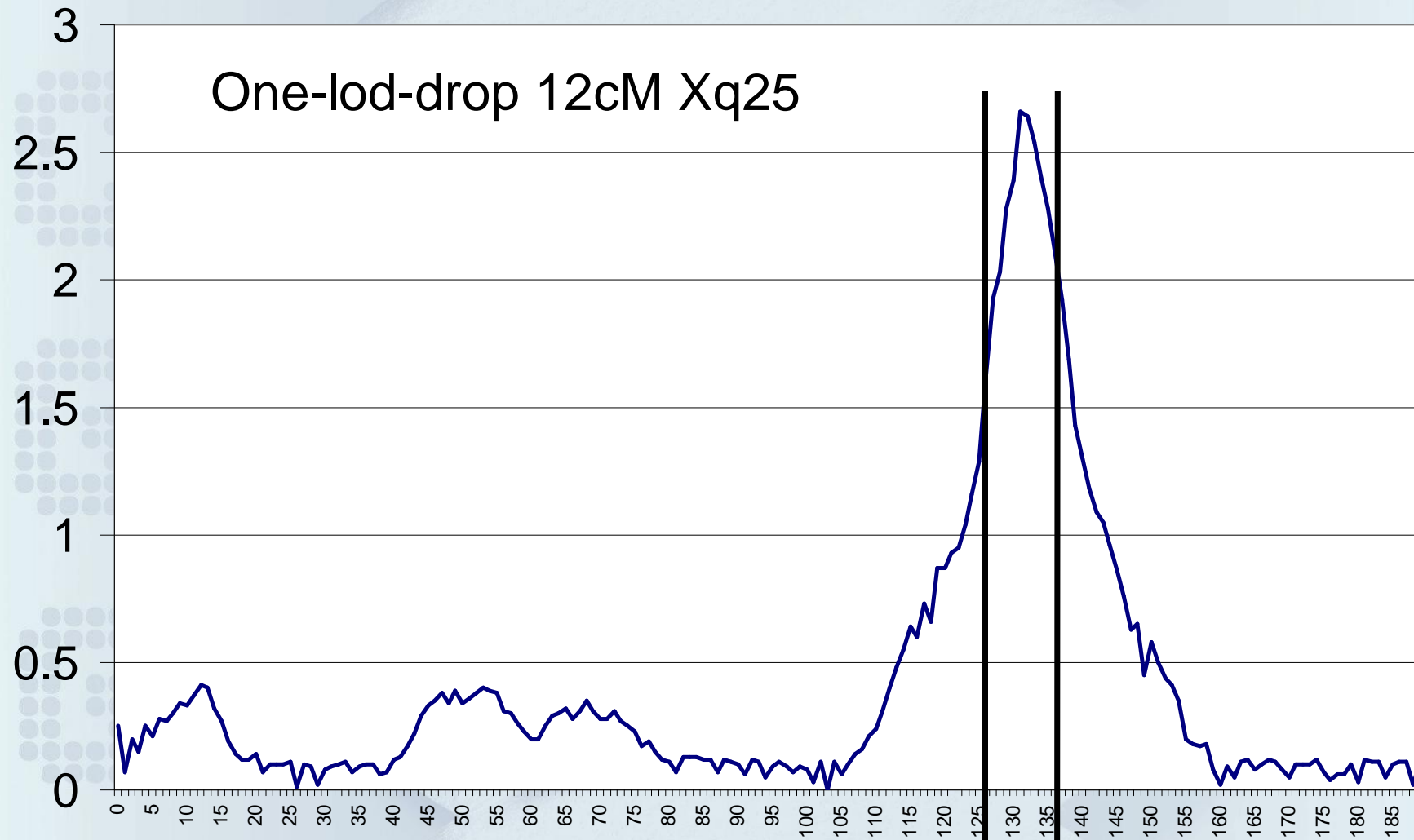


- Highest LOD score 3.43 chromosome 15q (all)

# Results: Stature Chromosome 15



# Results: Stature Chromosome X



# Conclusions

- Chromosome 15 signal comes mainly from the Australian dataset but is supported by some others
- Chromosome X locus builds up from several subcohorts -> common Caucasoid locus for stature?



# The Danish biobank's history

**Started in 1997 in association with LSADT97 (Longitudinal study of Aging Danish Twins). In total 688 blood samples were collected**

**In total 1 molecular geneticist, ½ MD and 3 technicians attached in 2006**

**Biological material are collected from more than 25.000 subjects in 2006. The biobank contains more than 200.000 samples**

**The samples are kept in 7 -80°C freezers (330 boxes with 81 samples in each freezer)**

**Blood spots are kept in 7 filing cabinets**



# Collection of material.

## 2 principal types of collection:

### Population based interview studies

- ✓ Collection of DNA material as cheek swabs or blood spots
- ✓ Lay interviewers employed by the Danish National Institute of Social Research

### Smaller studies based on specific phenotypes – migraine, hand eczema, centenarians etcetera

- ✓ Usually full blood samples
- ✓ Samples are collected by trained phlebotomists (nurses, technicians)



# Projects with DNA material

Project	Age	Number	Samples
The Centenarian study	100	310	Full blood
1905-cohort year 1998	93	1789	Blood spots
year 2000	95	807	Cheek brushes
year 2003	98	325	
year 2005	100	167	Full blood
MADT year 1998	~50	4279	Bloodspot/ Cheek brushes
Twin RNA study	~50	343	Full blood
Unilever	60-80	225	Full blood
GEMINAKAR year 1997	~45	1512	Full blood



# Projects with DNA material

Project	Age	Number	Samples
GenomEUtwin	55	400	Full blood
LSADT	75	688	Blood spot
year 1997	75	2259	Cheek brushes
year 1999	75	2185	
year 2001	75	1235	Full blood
year 2003	75		
Migraine twins	~50	124	Full blood
Danish Oldest Sibs	85	480	Full blood
GEHA	80	306	Full blood
Cleft lip and palate	~45	6061	Cheek brushes



# Primary samples in the Danish Biobank

<b>Sample type</b>	<b>Number</b>
<b>Full blood (plasma, serum, blood cells)</b>	<b>5,365 (Twin samples: 3,346)</b>
<b>Blood spots</b>	<b>13.472 (Twin samples: 10,870)</b>
<b>Cheek brushes</b>	<b>7,538 (Twin samples: 971)</b>
<b>Hair samples</b>	<b>1,875 (Twin samples: 1,559)</b>
<b>Total</b>	<b>28,250 (Twin samples: 16,746)</b>



## Collaborators that use the biological material.

<b>Name</b>	<b>Country</b>	<b>Biological material</b>
<b>Centre of Inflammation and Metabolism, Rigshospitalet</b>	<b>Denmark</b>	<b>1905 Cohort</b>
<b>Jeff Murray, Iowa</b>	<b>USA</b>	<b>Cleft lip and palate</b>
<b>Jeff Gardener/Abraham Aviv</b>	<b>USA</b>	<b>Telomeres</b>
<b>Steen Kølvrå, Århus</b>	<b>Denmark</b>	<b>Telomeres</b>
<b>Uffe Holmsgaard, Medical Biotechnology Center</b>	<b>Denmark</b>	<b>GEMINAKAR</b>
<b>Gabiolo Surdulescu</b>	<b>England</b>	<b>EUROCLOT</b>
<b>Dep. Of Molecular Medicine, Helsinki</b>	<b>Finland</b>	<b>GenomEUtwin. GEHA</b>



## Collaborators that use the biological material.

<b>Name</b>	<b>Country</b>	<b>Biological material</b>
<b>David Gunn</b>	<b>Unilever</b>	<b>GEMINAKAR</b>
<b>Karen Helene Ørstavik</b>	<b>Oslo, Norway</b>	<b>LSADT</b>
<b>Eugenia Wang</b>	<b>USA</b>	<b>Centenarians</b>
<b>Decode Genetics</b>	<b>Iceland</b>	<b>1905 cohort study</b>
<b>Gruppo di Biologica Molecolare</b>	<b>Italy</b>	<b>GEHA, ECHA</b>
<b>Affymetrix, High Wycombe</b>	<b>England</b>	<b>LSADT</b>
<b>Erasmus Medical Center</b>	<b>Netherland</b>	<b>GEMINAKAR</b>



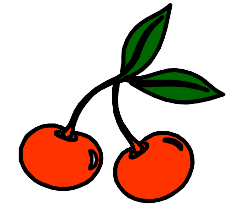
# Why is the biobank so useful?

**The genetic and biochemical information is available from nation-wide longitudinal surveys among twins and elderly in Denmark.**

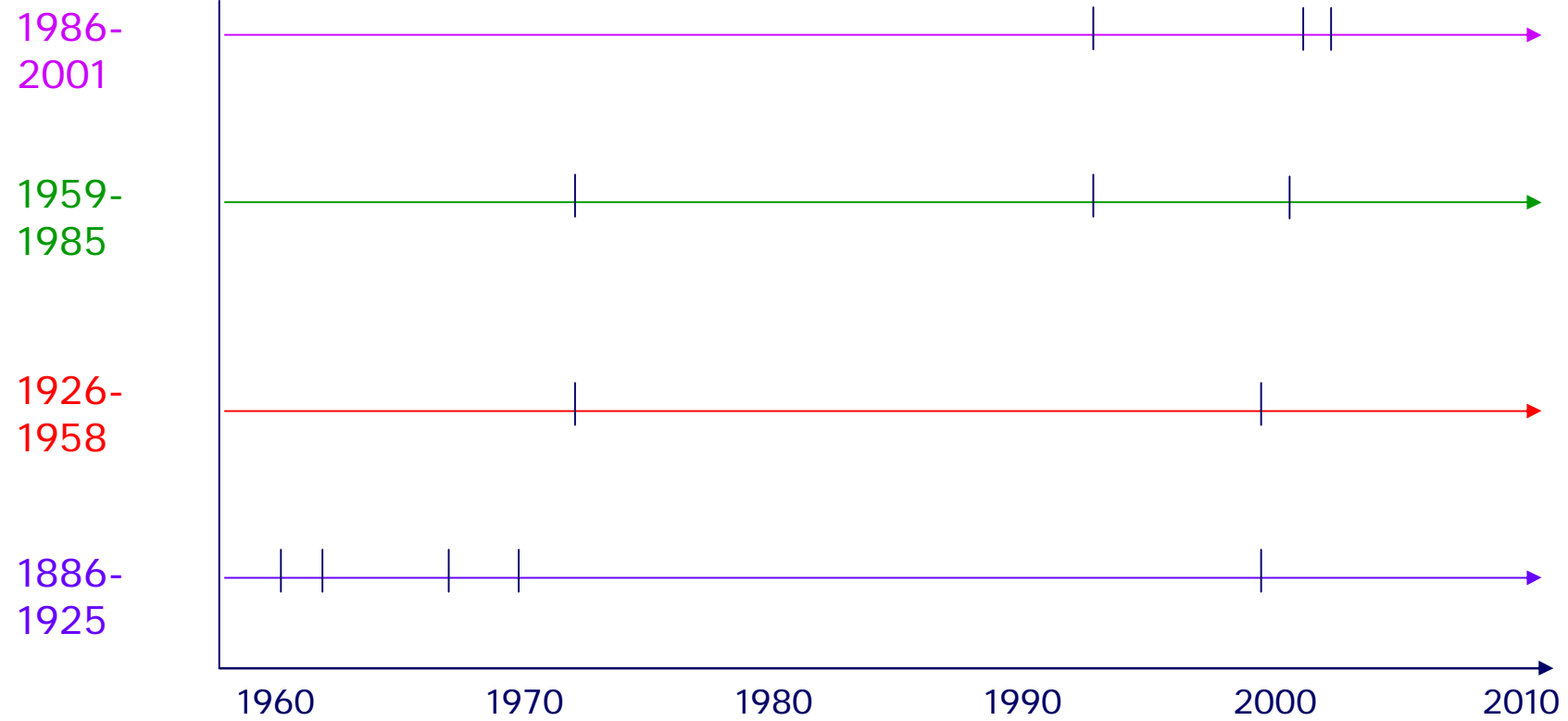
**The genetic and biochemical information can be combined with many parameters obtained from these population based studies such as cognitive and physical functioning, survival and social factors.**

**The genetic and biochemical information can be combined with nationwide health and death registers covering practically all Danish citizens.**

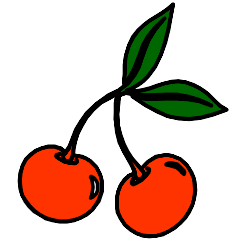
# The Swedish Twin Registry



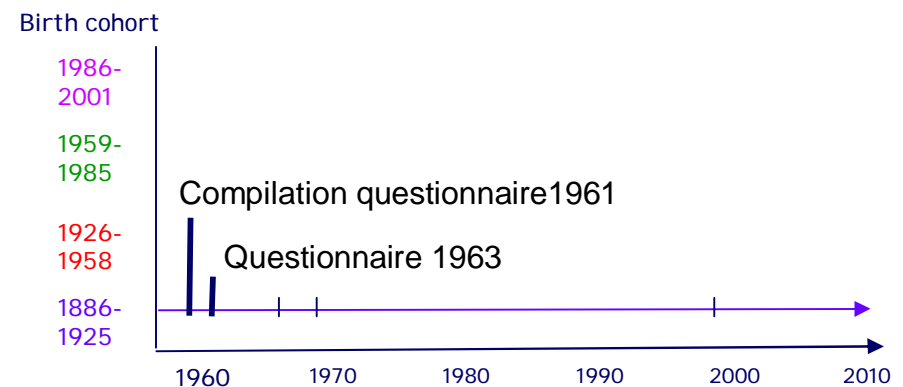
Birth cohort



# The Swedish Twin Registry Cohort 1886-1925

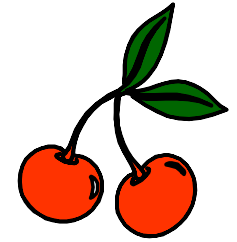


- u Questionnaire for compilation – 1961 (N=21,890)
  - Zygosity; risk factors
- u Q 1963 (N=18,616)
  - Risk factors, diseases
- u Q 1967 (N=16,779)
  - Risk factors, diseases
- u Telephone interview: 1998-2002 (N=45,878)
  - Risk factors, Most common complex disorders

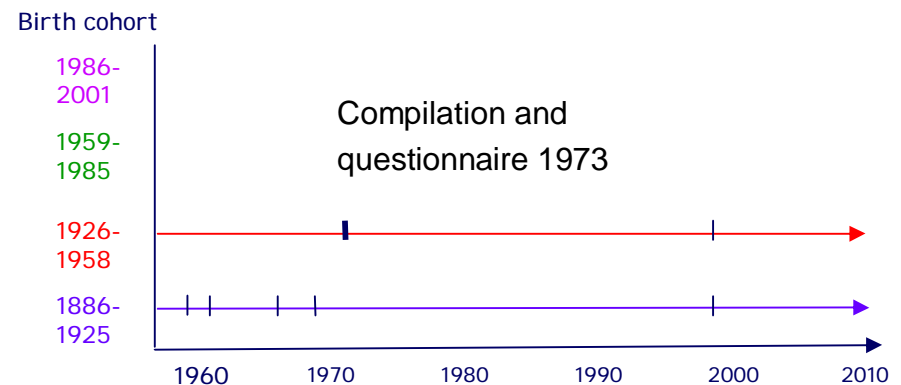


# The Swedish Twin Registry

## Cohort 1926-1958

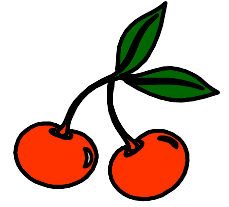


- u Q 1973 (N=30,279)
  - Zygosity, risk factors, diseases
- u Telephone interview: 1998-2002 (N=45,878)
  - Risk factors, Most common complex disorders



# Psychiatric disease of interest?

---



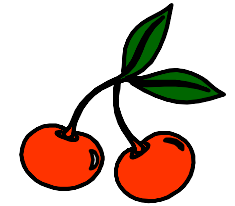
## u Interview data

- Major depression
- GAD
- Parkinson
- Dementia, neurodegenerative & psychiatric diseases
- Fibromyalgia, chronic fatigue, sleeping disorders
- Personality

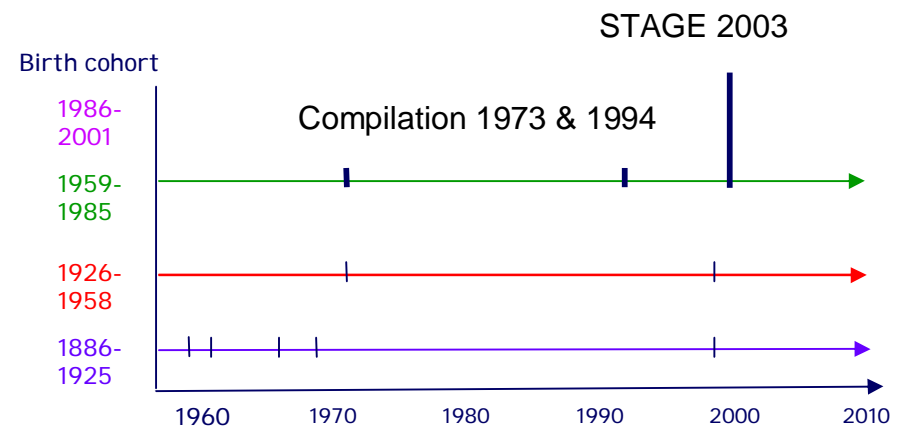
## u Linkage with hospital discharge registry

- Schizophrenia
- Bipolar disorder

# The Swedish Twin Registry Cohort 1959-1985

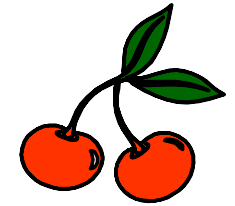


- u Swedish Twin study of Adults – Genes and Environments (STAGE)
- u 43,000 twins
  - Expected response rate N=25,510
- u Web questionnaire
  - telephone interviews + questionnaire
- u Genes and environments for nicotine dependence
- u Somatic and psychiatric diseases
- u Other exposures

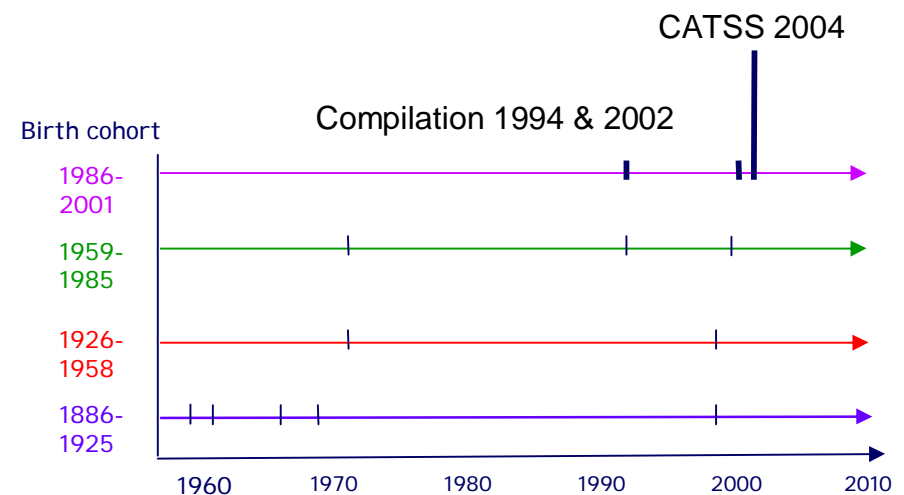


# The Swedish Twin Registry

## Cohort 1986-2001

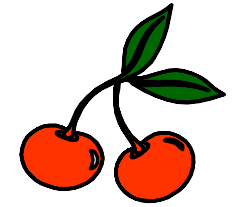


- u Child and Adolescent Twin Study in Sweden (CATSS)
- u 2 entire birth cohorts each year
  - Twin pairs 9 and 12 years
    - » 1,500 + 1,500 twins each cohort
  - Telephone screening (interview with parents)
    - » Psychiatric problems
    - » Asthma/allergies
    - » Somatic diseases



# Contents of registry “base data”

---



## u Administrative information

- National registration number (“personnummer”)
- Age, sex

## u Questionnaire data from 1961, 63, 67, 73

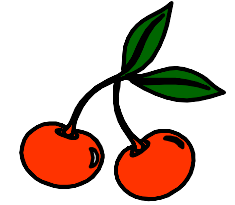
- Zygosity, sociodemographic
- Health/illnesses, health related behaviors
- Environmental exposures

## u Linkage with other population-based registries

- Cause of death
- Marital status and addresses
- Cancer registry
- Medical birth registry
- Hospital discharge registry

# Contents of registry “ancillary data”

---



## u SALT

- Dementia, Parkinson, CFS

## u STAGE

- Smoking cessation

## u CATSS

- Autism, ADHD, Eating disorder

## u TCHAD

- 4 waves of follow up of twins from age 8-20
- Delinquency, psychopathy, CD, ADHD, fears/phobias, etc

## u TOSS (Twin/Offspring Study in Sweden)

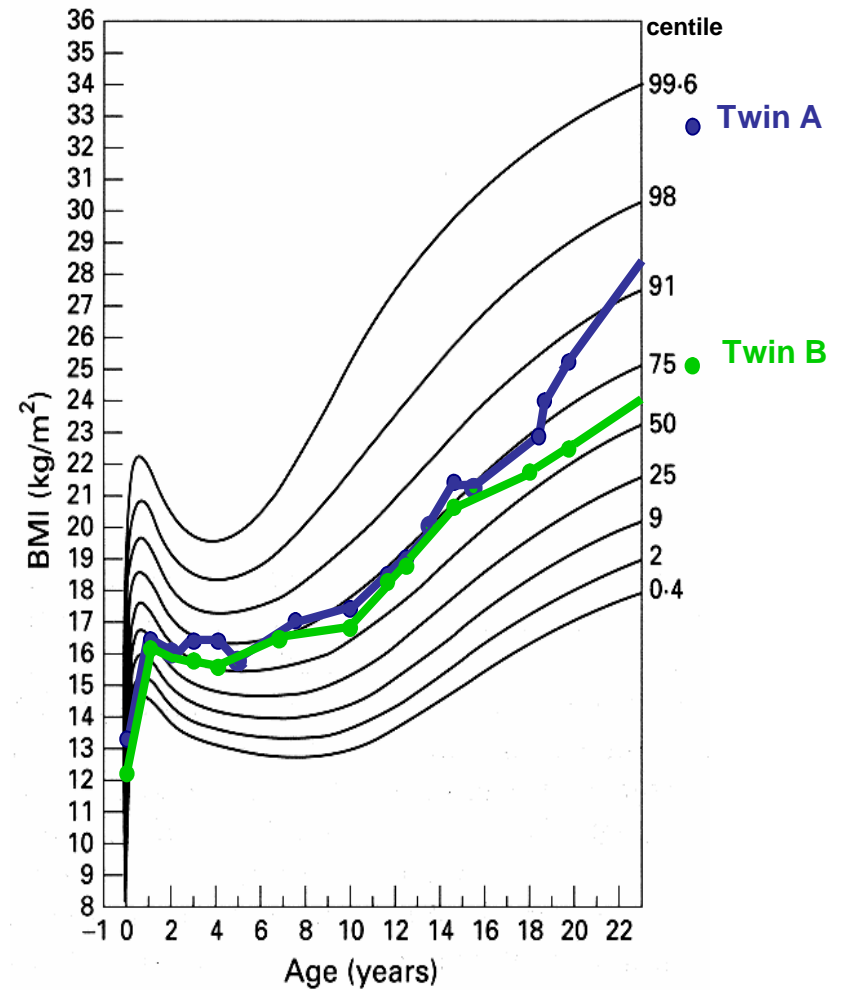
- 1000 pairs of twins, spouses and adolescent child
- Family relationships, depression, well-being etc

# Finnish Twin Cohort study

- u established 30 years ago
- u a unique longitudinal study resource with a richness of phenotypic data collected in repeated surveys, interviews, and by register linkage
- u Cancer registry, hospitalizations, medications, disability pensions and causes of death
- u The older cohort of twins (c. 16500 pairs known zygosity) born before 1958 and studied since 1975 has contributed to longitudinal epidemiological studies, studies of chronic disease endpoints and risk factors with a special emphasis on smoking and nicotine dependence
- u C. 15000 DNA samples collected mostly for specific study projects

# Younger cohorts

- u Twin-family cohorts (FinnTwin16: born 1975-1979 & FinnTwin12: 1983-1987)
- u About 5500 families with 20 000+ participants
- u Multiple measures throughout development
- u Emphasis on behavior and risk factors, especially alcohol use and abuse, smoking and weight gain
- u Deeply phenotyped subsamples with biological samples and extensive clinical and imaging data
  
- u Access to multiple measures over time in a epidemiologically well-defined sample that is also genetically informative, but not selected on the basis of a disease or trait



# Norwegian Twin Registries

- u Historically 3 population-based registries established at different research institutes
- u Recent agreement in place to centralize and coordinate the registries (Still need to secure funding for centralization)
- u The National Twin Registry will be under administration of the Norwegian Institute of Public Health (NI PH)

# The Older Cohorts:

## I. Cohorts born 1895-1945

- Established at Psychiatric Inst. University of Oslo (Kringlen)
- 37,000 pairs (like and unlike sex)

## II. Cohorts born 1915-1960

- Some overlap with I.
- Established at Institute of Medical Genetics, University of Oslo
- 11,175 pairs (like sex only)

Primary research areas for I & II: psychiatric disorders, CHD, epilepsy, cancer & birth weight

# NI PH Twin Program of Research

- u Established in 1992 at the Norwegian Inst. of Public Health (J. Harris)
- u Twins identified through the Medical Birth Registry
- u Cohorts born 1967-1979, like & unlike sex (15374 twin births)
- u New Cohorts recruited over time
  - Q1 sent in 1992 to cohorts born 1967-1974
    - » Responses from 5864 twins (2570 pairs)
  - Q2 sent in 1998 to cohorts born 1967-1979
    - » Responses from 8045 twins (3334 pairs)
- u DNA from more than 4700 twins
- u Have unique Person Number for linkage to other sources
  
- u Multiple sub-projects

Harris et al., *Genet. Epid.*, 1995;  
Harris et al., *Twin Research*, 2002.

# Primary Research Focus

- ✓ mental health & psychological well-being
- ✓ obesity
- ✓ asthma & allergies
- ✓ sex differences & physical health
- ✓ health behaviors
- ✓ health perceptions
- ✓ perinatal influences on health outcomes

# Sub-projects—Mental Health

- I. Mental Health Interview Study: personality and psychiatric disorders 1391 pairs from Q2 interviewed using CIDI and SID-P. (Tambs, Harris, Reichborn-Kjennerud, Kringlen, Torgersen)
- II. Genetics and Personality: genome scan project using data from I (Collaboration with Rockefeller Univ)
- III. Study of genetic and environmental mediation of CO<sub>2</sub> reactivity and anxiety disorders, (346 pairs from I) (M. Battaglia in collaboration with I)
- IV. Pain sensitivity and pain regulation: 188 twins tested examined using multiple pain-measure modalities (Nielsen, Harris, Price, Vassend, Stubhaug)

# Sub-Projects: Health & Disease

- ⌋ GenomEUtwin: identification of genes that modify risk for disease and disease-related factors
- ⌋ Asthma and atopy: biomarkers of inflammation, clinically measured asthma & atopy phenotypes in 170 pairs (Harris, Nystad, Lund, Boe, Kongerud)
- ⌋ Inflammatory Bowel Disease (Vatn, Bengtson & Harris)
- ⌋ Danish and Norwegian Study of Bechterew's (Pedersen & Harris)
- ⌋ Twin-Family Study on the Genetics of Epilepsy (Corey, Magnus, Harris, Nakken, Solaas)

# Potential for Linkages to National Health Registers

- u Medical Birth Registry: perinatal measures, mother's health during pregnancy (already linked to twin data)
- u Cause of Death Registry: ICD codes (from 1951)
- u Cancer Registry: mandatory ICD and TMM coding for all cancers in Norway
- u Prescription Medicine Registry (from 2004)
- u Patient Discharge Registry (under establishment)
- u Disease specific registers (diabetes, MS)