

Using Population Isolates in Disease Genetics

NIASC symposium

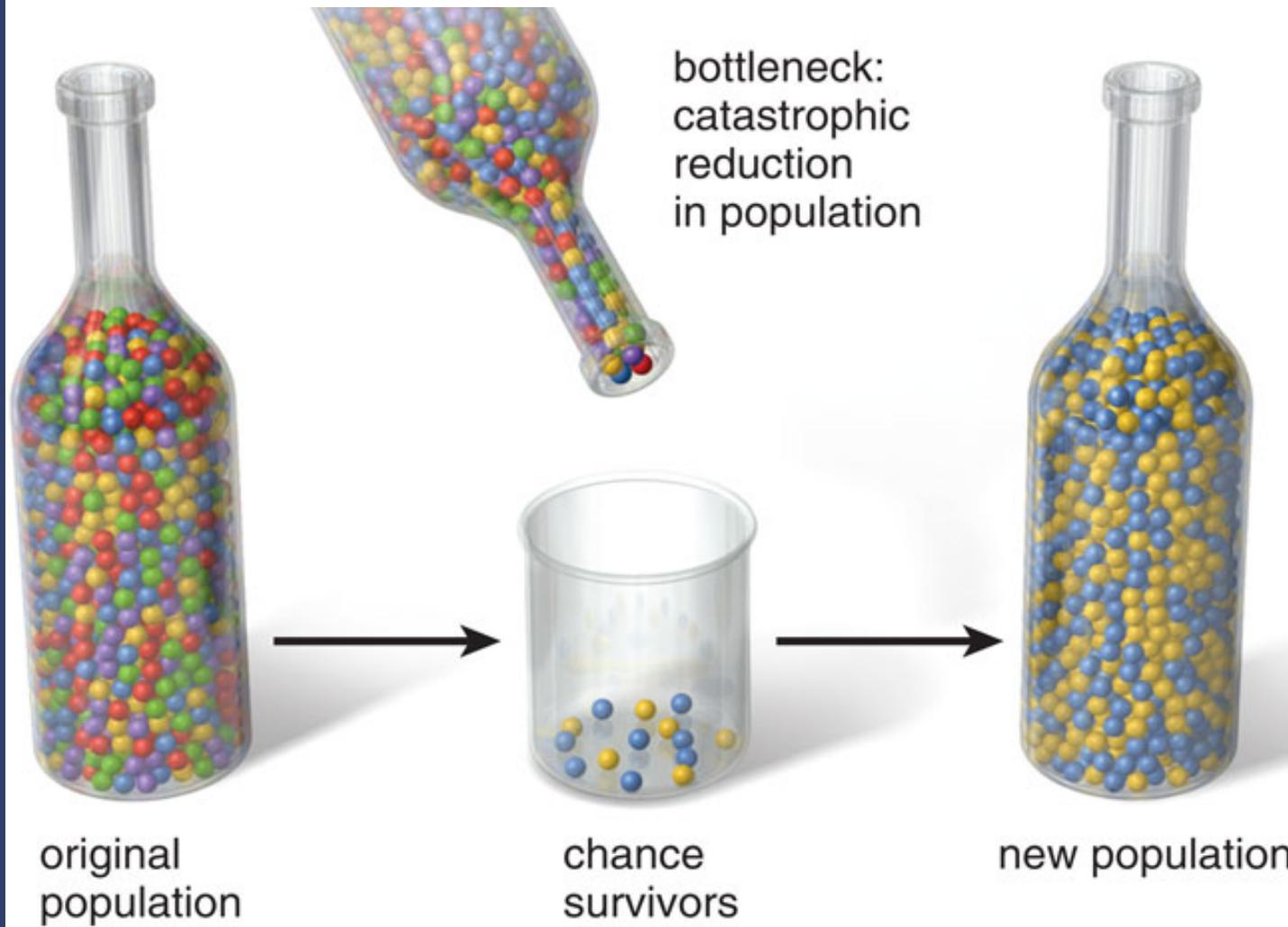
January 16, 2015

Aarno Palotie, M.D., Ph.D.



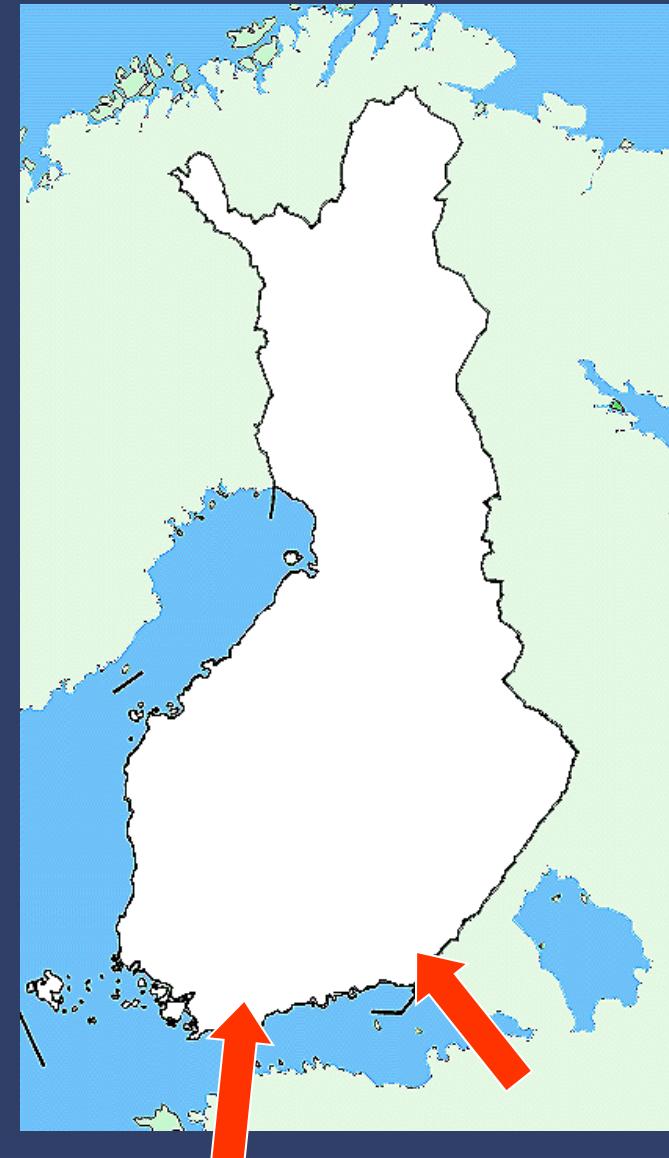
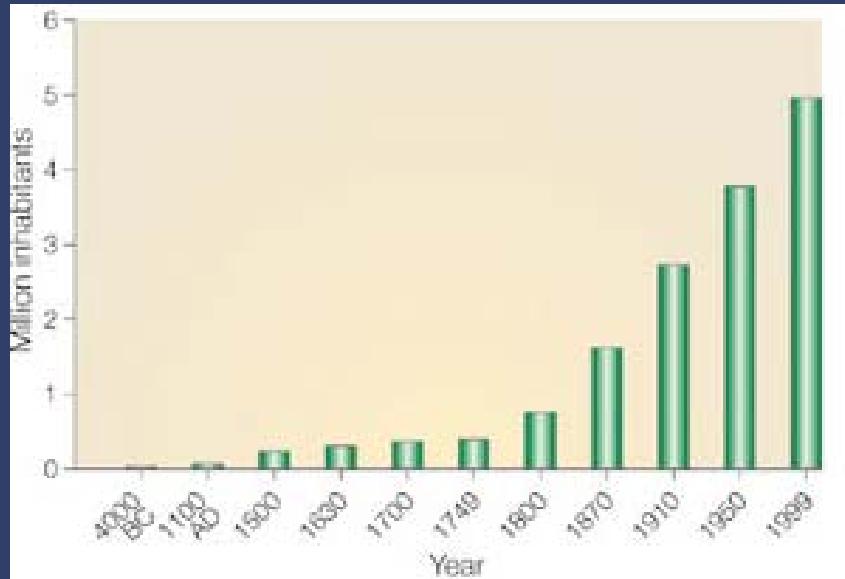
Population isolates

Bottle neck effect and genetic drift



Population History

- Small Number of Founders
- No Immigration
- Isolation
 - Geographical
 - Linguistic, cultural
- Rapid Expansion



Early Settlement

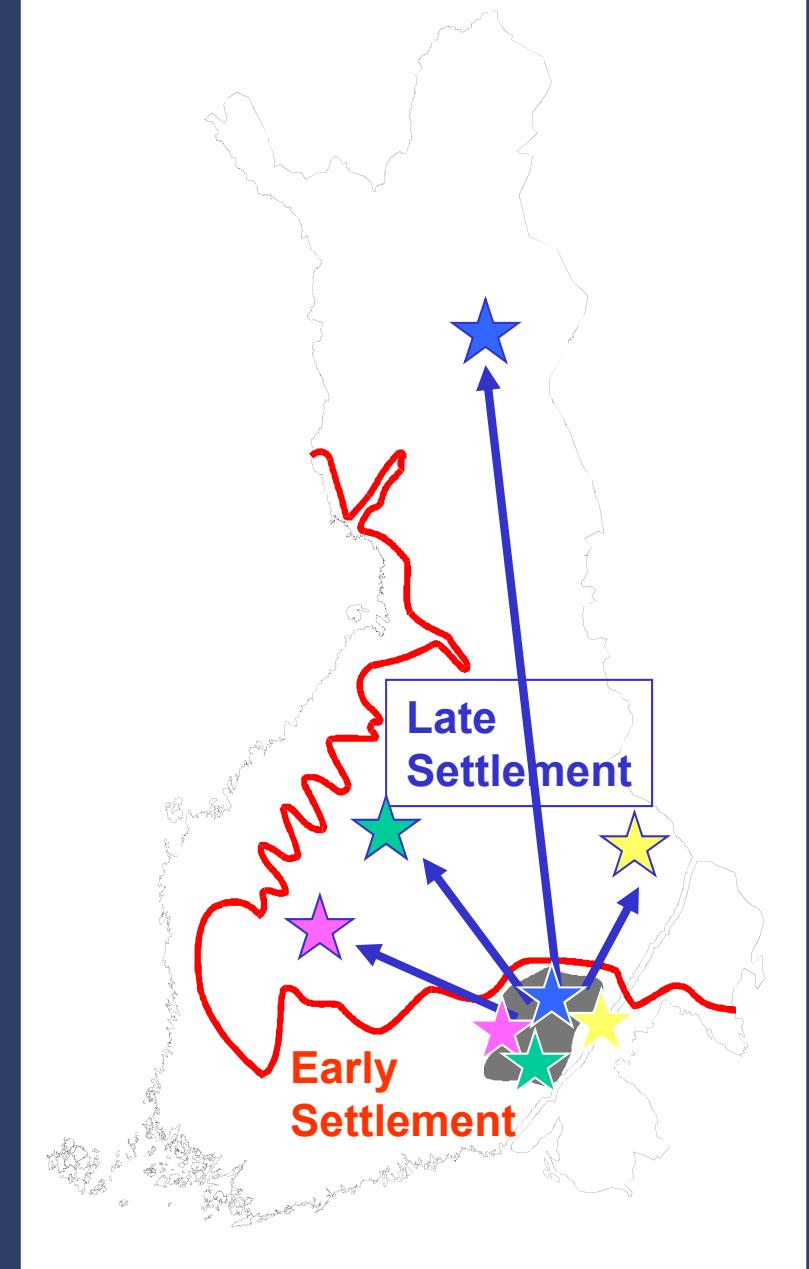
- 2000-10 000 years ago
- South and Coast

Late Settlement

- 16th century
- multiple bottle necks

Expansion

- 18th century -
population 250 000
- Today -
population 5.3 million



Internal Isolates: Kuusamo



1670

40 families

1750

1208 inhabitants

1800

3059 inhabitants

1910

10 599 inhabitants

1995

18 281 inhabitants

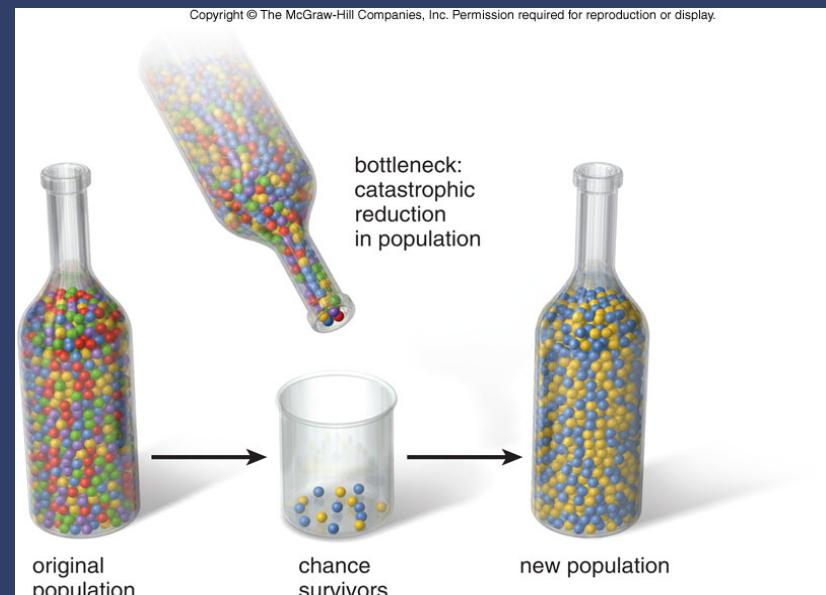


immigration negligible

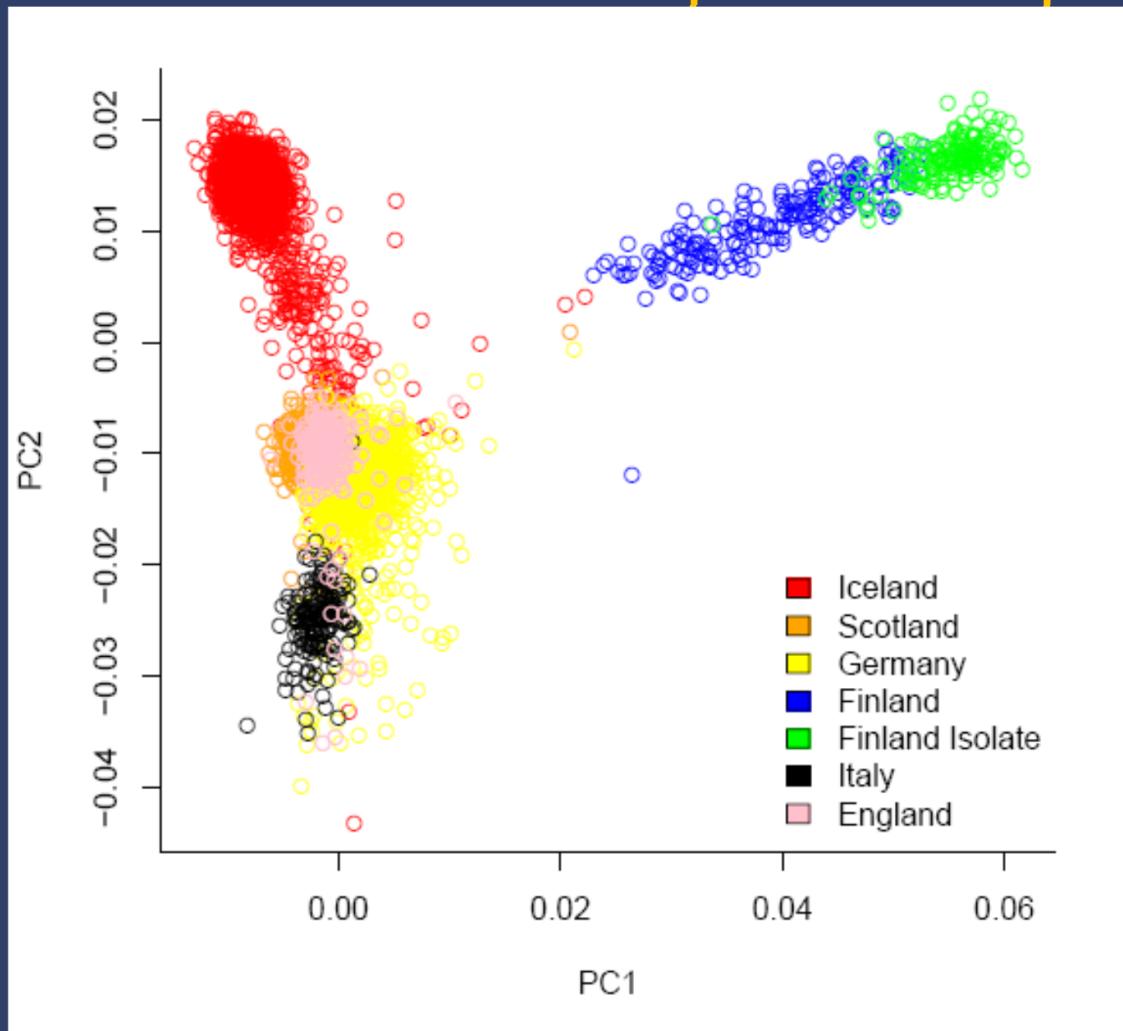
Finland: population history = advantage for genetics

Gene identified- Mutation known

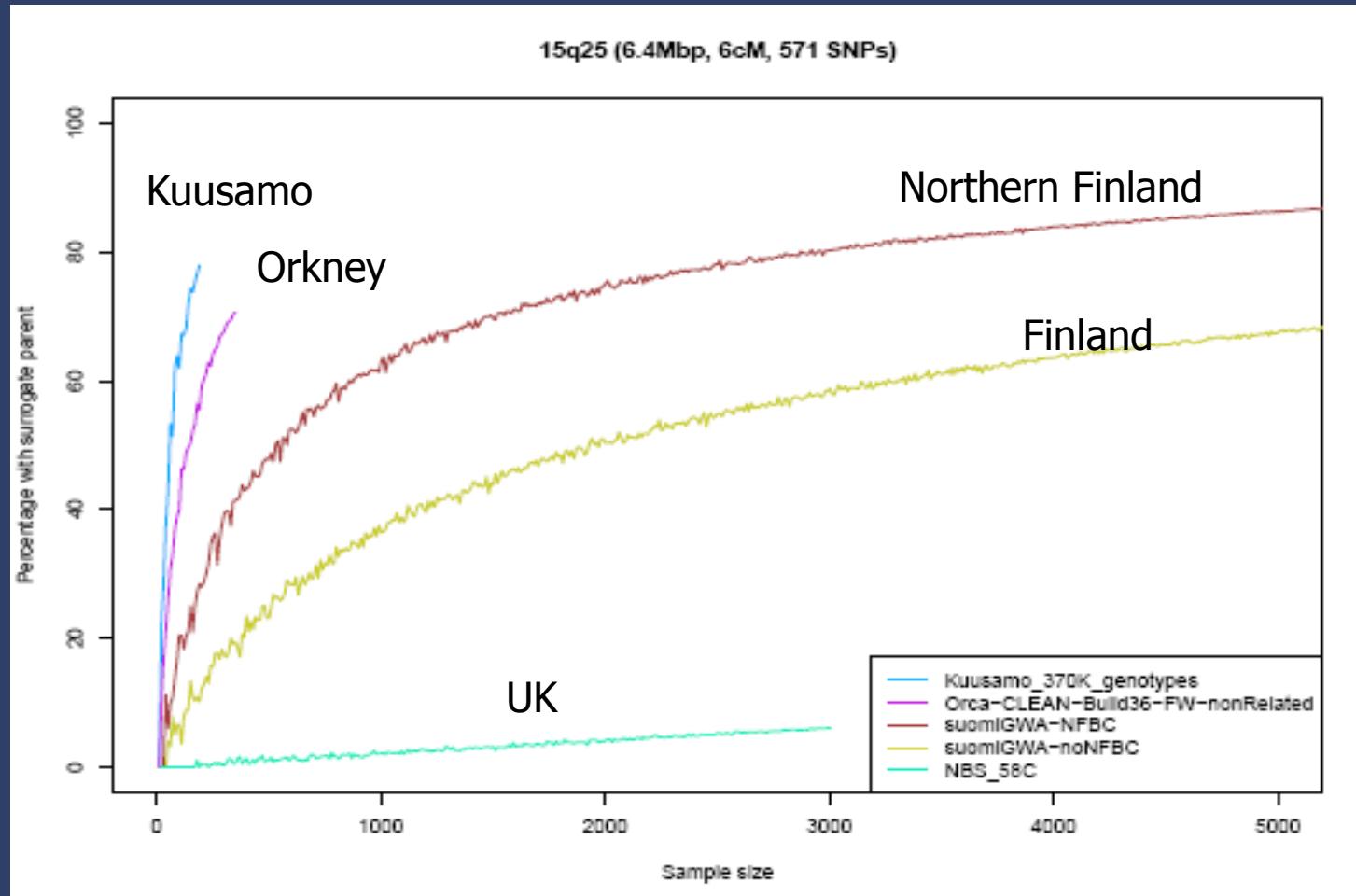
- GRACILE (death in infancy)
- LAAHD (intrauterine death)
- FSH-RO (fertility disturbance)
- EPMR (progressive retardation)
- PEHO (progressive retardation)
- TMD (muscle disease) dominant
- RAPADILINO (growth disturbance with malformations)
- LCCS (intrauterine death)
- IOSCA, OHAHA (progressive retardation)
- CHS (progressive retardation)
- vLINCL (progressive retardation)
- HYDROLET (intrauterine death)
- SALLA (progressive retardation)
- MKS (intrauterine death)
- MEB (severe retardation)
- TCD, CHM (eye disease), X -recessive
- INCL (progressive retardation)
- HOGA (eye disease)
- DTD (growth disturbance)
- JNCL (progressive retardation)
- CHH (growth disturbance)
- MUL (growth disturbance)
- FAF (eye, nerve and skin disease) dominant
- USH3 (ear and eye disease)
- PLOSL (progressive retardation)
- AGU (progressive retardation)
- CLD (watery diarrhea)
- NKH (severe retardation)
- LPI (metabolic disease)
- CCD (watery diarrhea)
- APECED (autoimmune polyendocrinopathy)
- RESCH, RS (eye disease), X- recessive
- PME (neurological disease)
- SMB12 (anemia)
- CNA2 (eye disease)
- CNF (known as)
56... 58... 62... 64... 66... 68... 70... 72... 74... 76... 78... 80... 82... 84... 86... 88... 90... 92... 94... 96... 98

GWA based relationships of Europeans



Sample size needed to construct parental haplotypes: Advantage for Imputation



Kimmo Palin and Richard Durbin, 2011



UNIVERSITY OF
OXFORD



SEMEL
INSTITUTE
UCLA



M SPH



Sequencing Initiative SISU



wellcome trust
sanger
institute

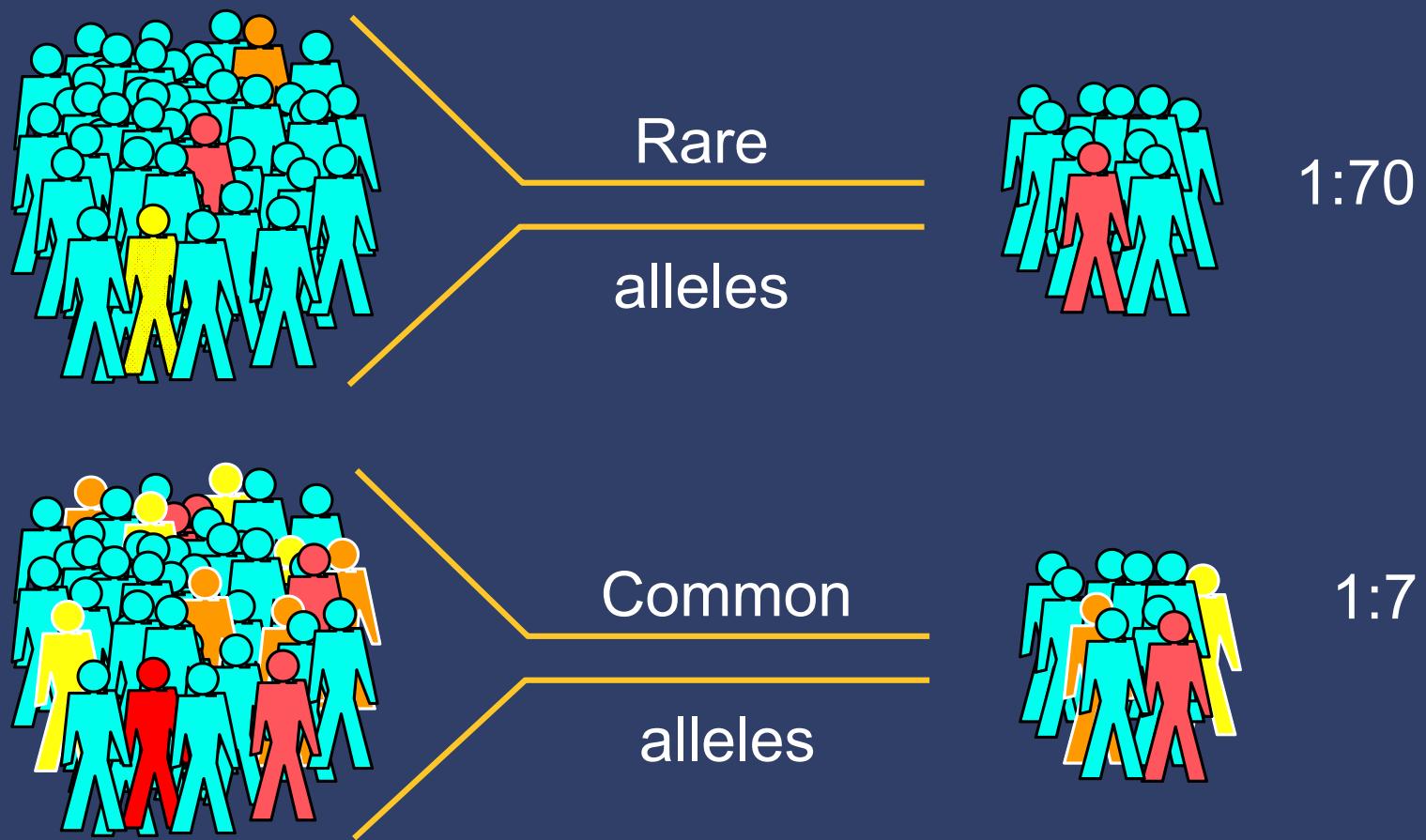
FIMM
Institute for Molecular Medicine Finland
Non-EU EMBI Partnership for Molecular Medicine

THL
FIMM

Washington
University in St. Louis
INSTITUTE

BROAD
INSTITUTE
MGH
HUSSEY
GENERAL HOSPITAL

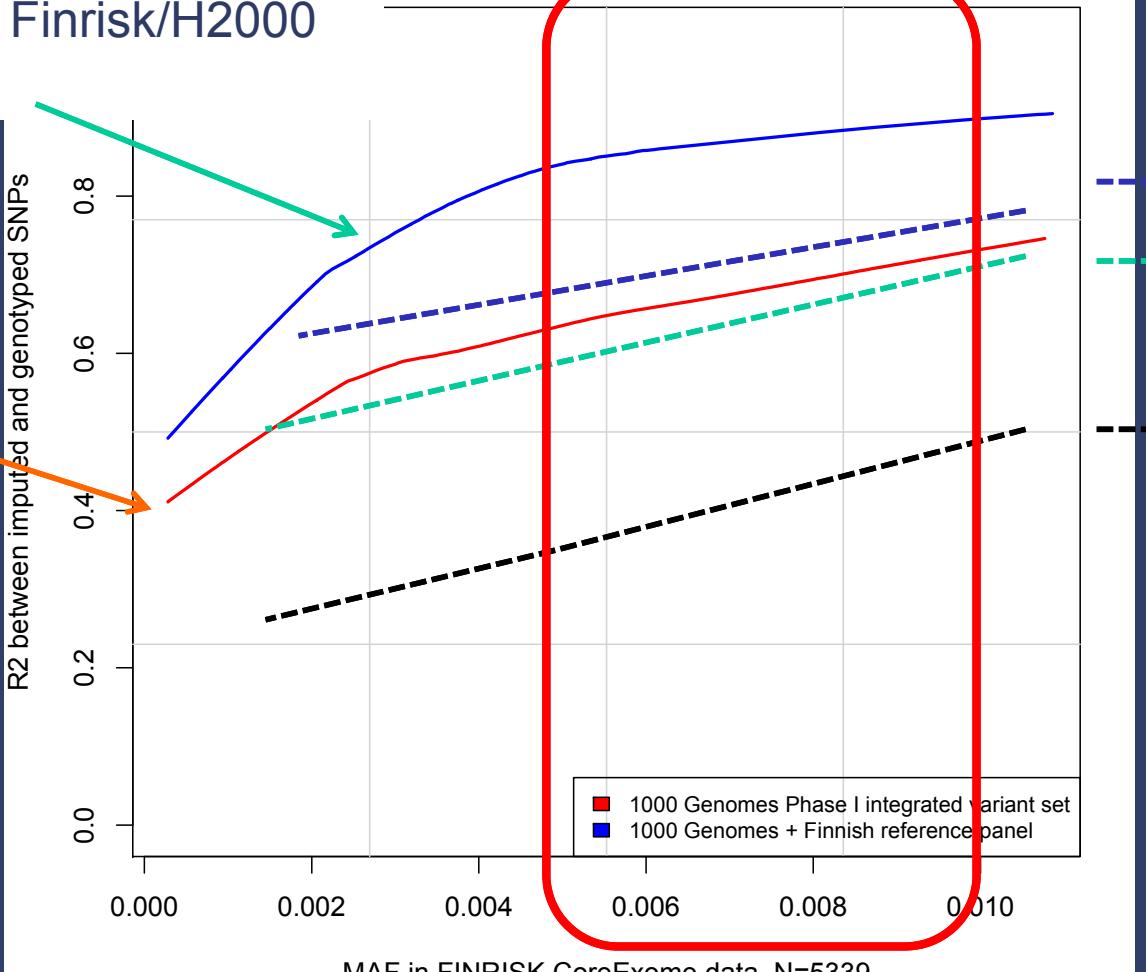
The Effect of Population Bottlenecks to Disease Alleles



Finns with 1900 Finrisk/H2000 panel +1000G

Finns with 1000G panel

Imputation accuracy (chromosome 17)



NL with GoNL + 1000G panels
UK with UK10 + 1000G panels

UK with 1000G Panel

Sarin and Ripatti

Imputation accuracy of SNP with MAF <= 1% in FINRISK HumanCoreExome sample set using two different imputation reference panels. Singletons, doubletons and SNPs with "info" < 0.4 and were filtered out.



MASSACHUSETTS
GENERAL HOSPITAL

Population cohorts

Finrisk
92, 97, 02, 07
Chronic disease trait survey
29 273 participants

Health 2000
Public health examination
8 800 participants

Young Finns cohort
LASERI
3300 participants

Twin cohort
70 000 individuals

Northern Finnish Birth cohort
1966, 1986
10 000 individuals

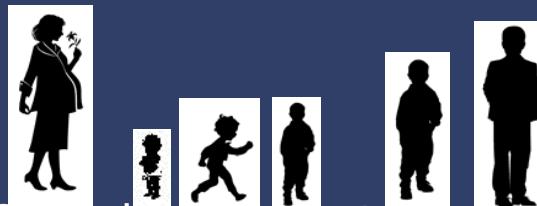
Helsinki Birth cohort
Born in 1933-1944
12 300 individuals



Follow-up using:

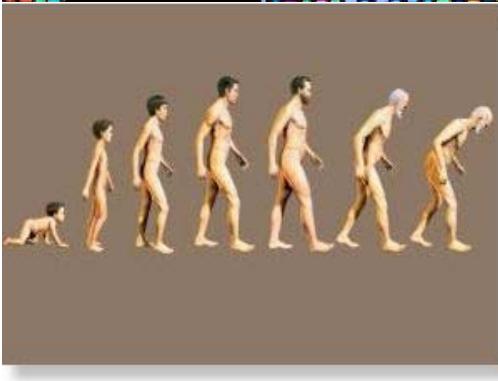
- Causes-of-Death Register
- Hospital Discharge Register
- Drug Reimbursement Register
- Cancer Register

>6000
Cardiovascular
endpoints



- Several recontacts to participants
- Causes-of-Death Register
- Hospital Discharge Register
- Drug Reimbursement Register
- Cancer Register

From biobanks to personalized and stratified medicine



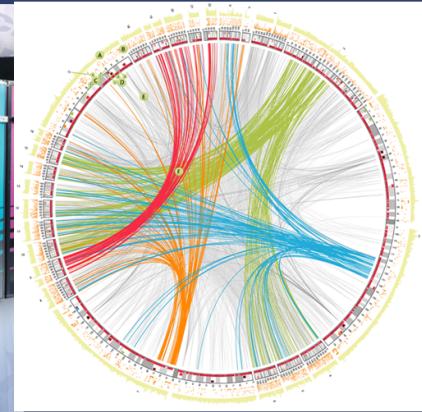
Biobanks

Gene information

Extensive health records

Data storage and integration

Modeling



Implementation in Health Care





www.nationalbiobanks.fi

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.

The pages are currently under development. If you wish to add information on your own study collections, please contact us: Kaisa Silander or Markus Perola (firstname.lastname (at) thl.fi).

[Link to Studies Summary Table](#)

[Link to Studies GWAS Table](#)

Finnish epidemiological study collections with available DNA samples:

[ATBC](#), [Child-sleep](#), [Finnish Twin Cohort](#), [FINRISK](#), [Helsinki Birth Cohort Study](#), [Helsinki Sudden Death Study](#), [Health 2000](#), [METSIM Study](#), [Northern Finland Birth Cohorts 1966 and 1986](#), [Young Finns Study](#)

Finnish disease-specific study collections with available DNA samples:

[Autism spectrum disorder family Study](#), [Botnia Studies](#), [Corogene](#), [FinnDiane](#), [Finnish Hematological Registry and Biobank \(FHRB\)](#), [FUSION](#), [Helsinki Urological Biobank \(HUB\)](#), [Intracerebral aneurysm \(FIARC\)](#), [Migraine Family Study](#), [Multiple Sclerosis Family Study](#), [THL Psychiatric Family Collections](#)

International collaborative projects, involving also Finnish cohorts:

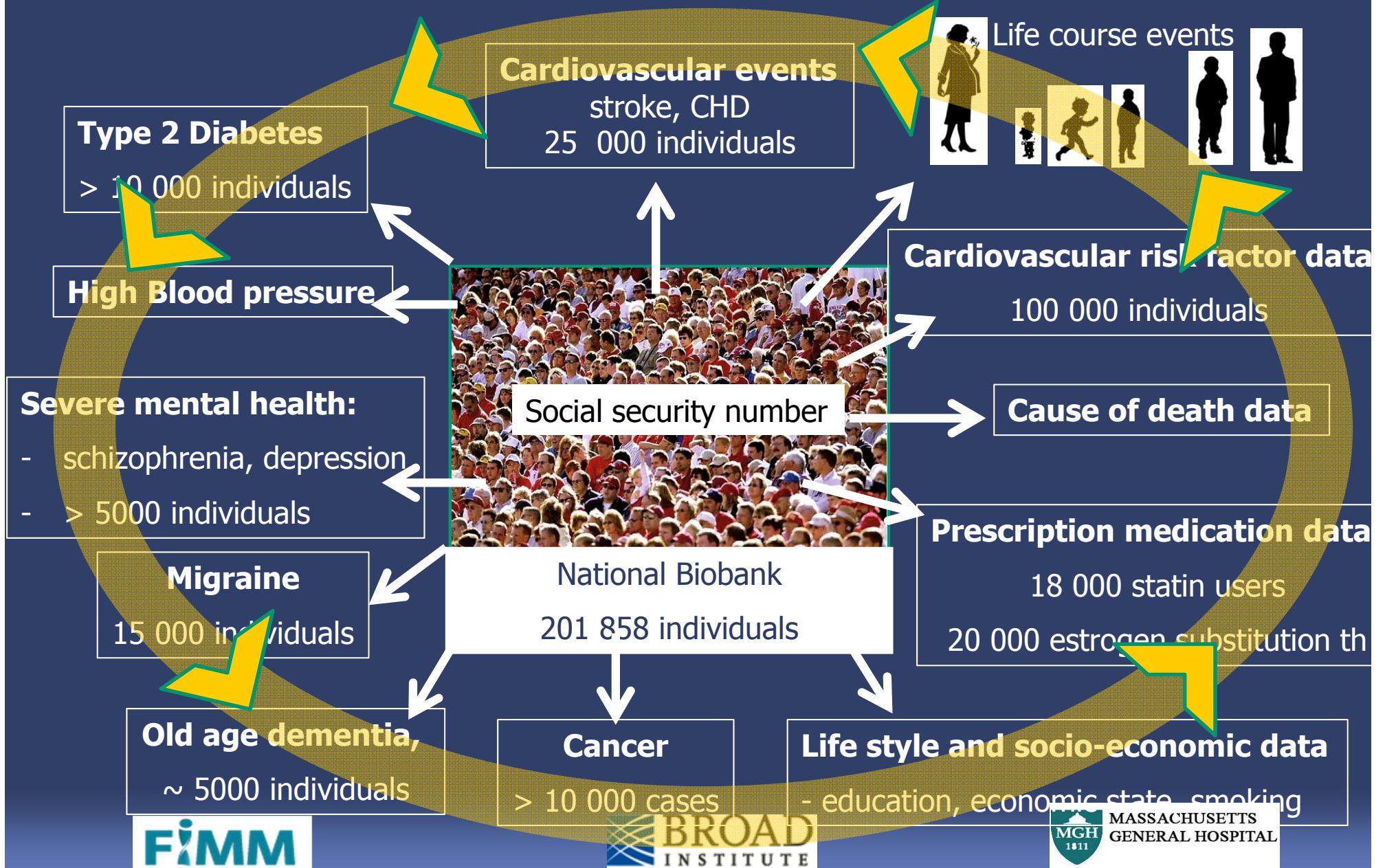
[GEHA](#), [MORGAM](#), [1000 Genomes](#)

<http://www.nationalbiobanks.fi>

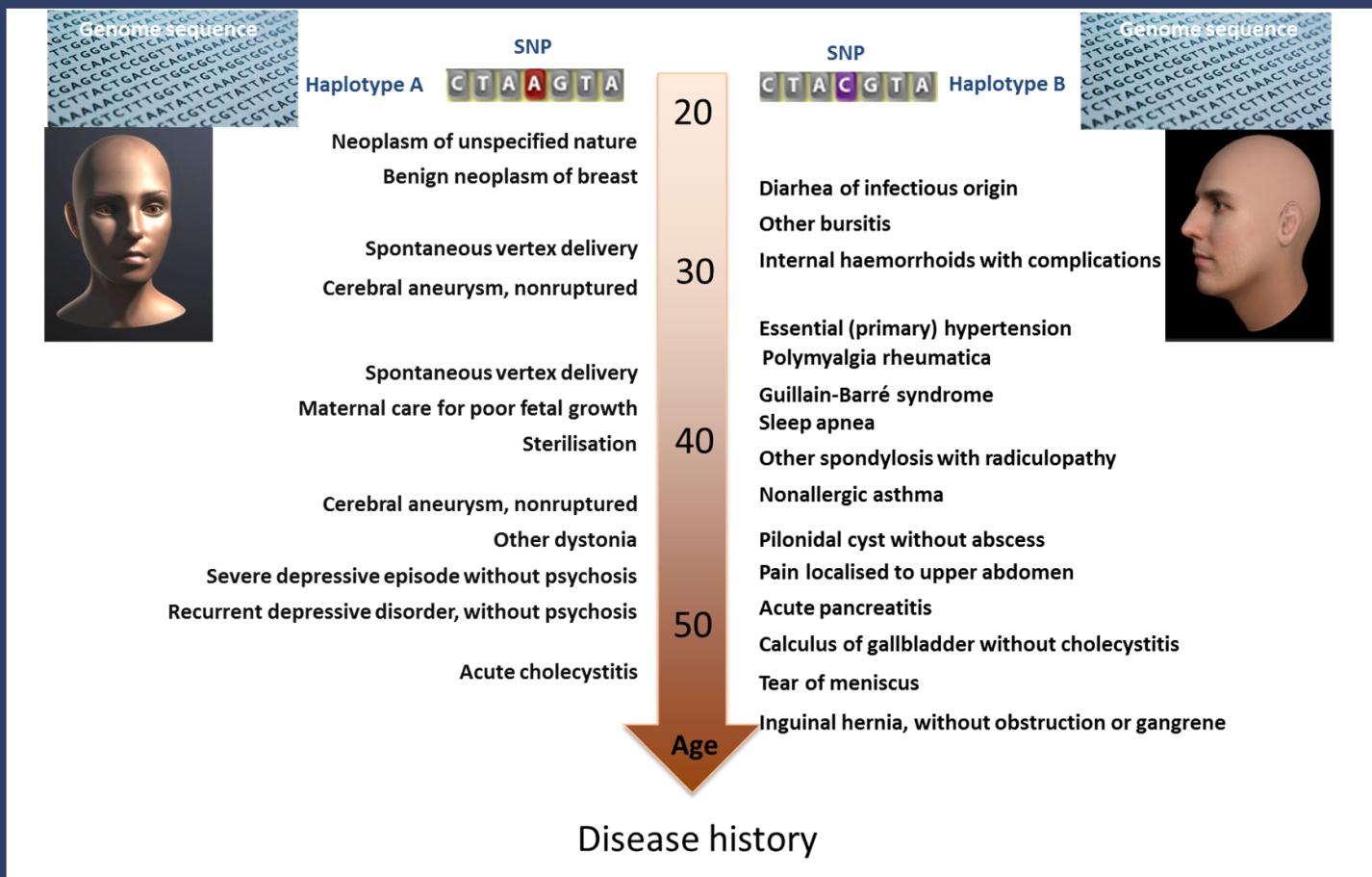
National Biobanks Finland

130 000 individuals from population cohorts

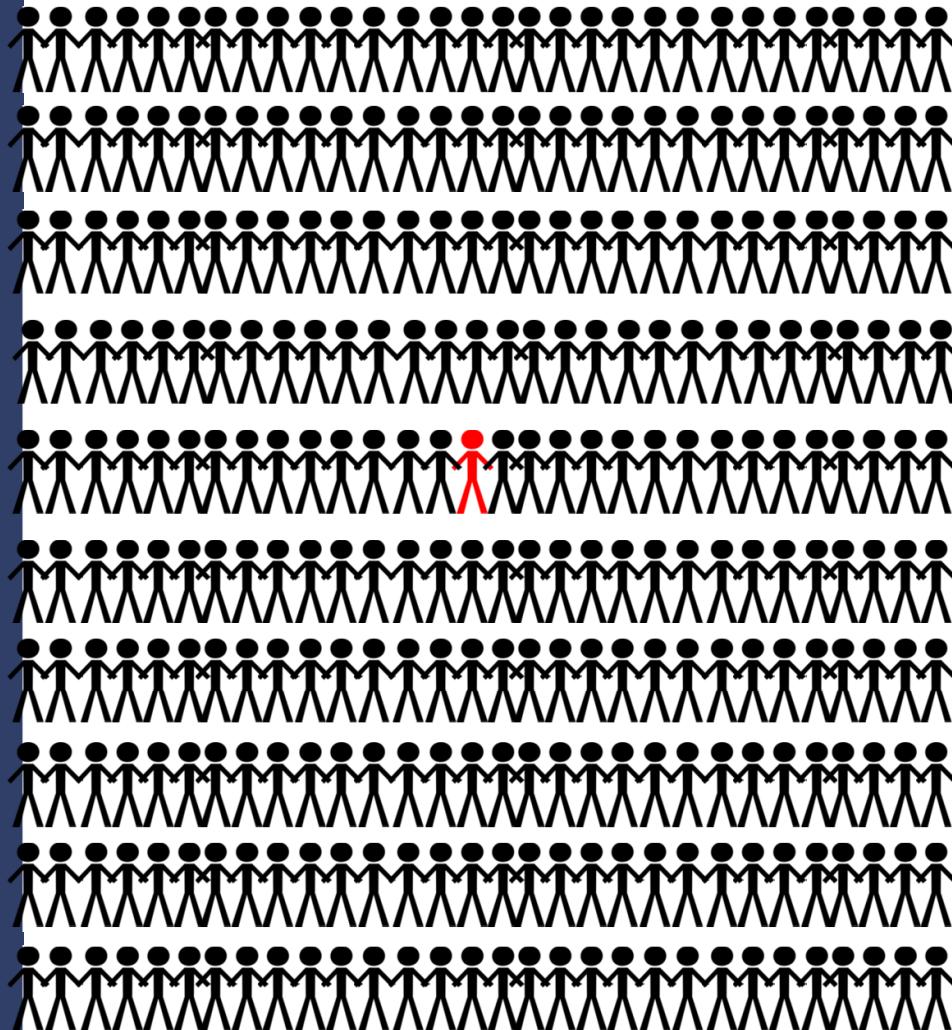
70 000 individuals from disease collections



Example of health histories from two persons from the national biobanks with a 40 year follow-up



Rare and low frequency vs. common variants



Single rare variant dominating the risk ($>5 \times$ risk)

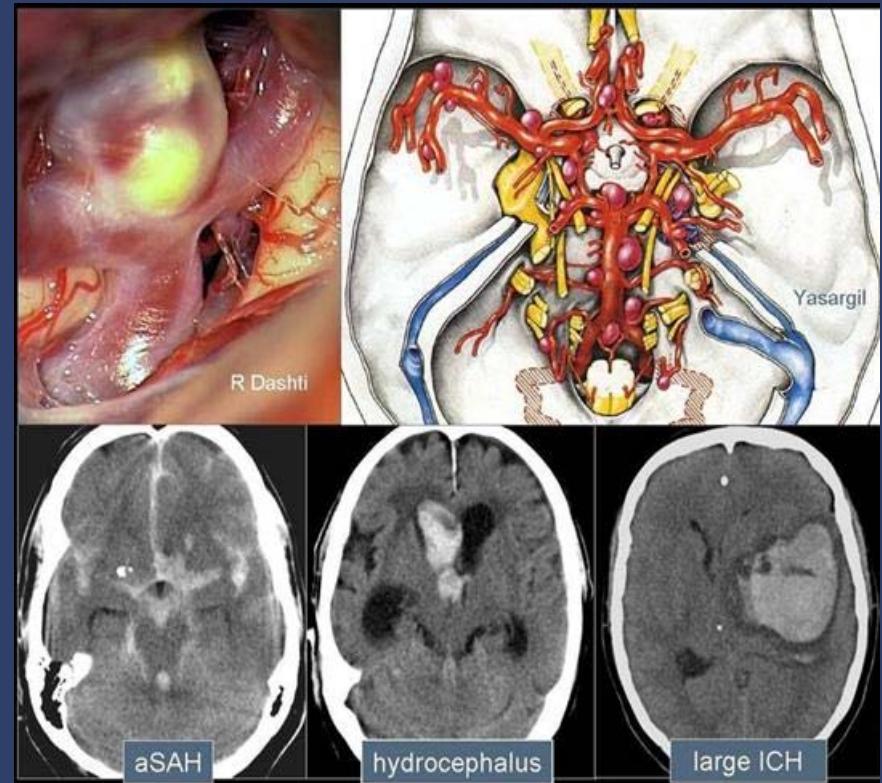


Multiple genetic and other risk factors contributing ($>2 \times$ risk)

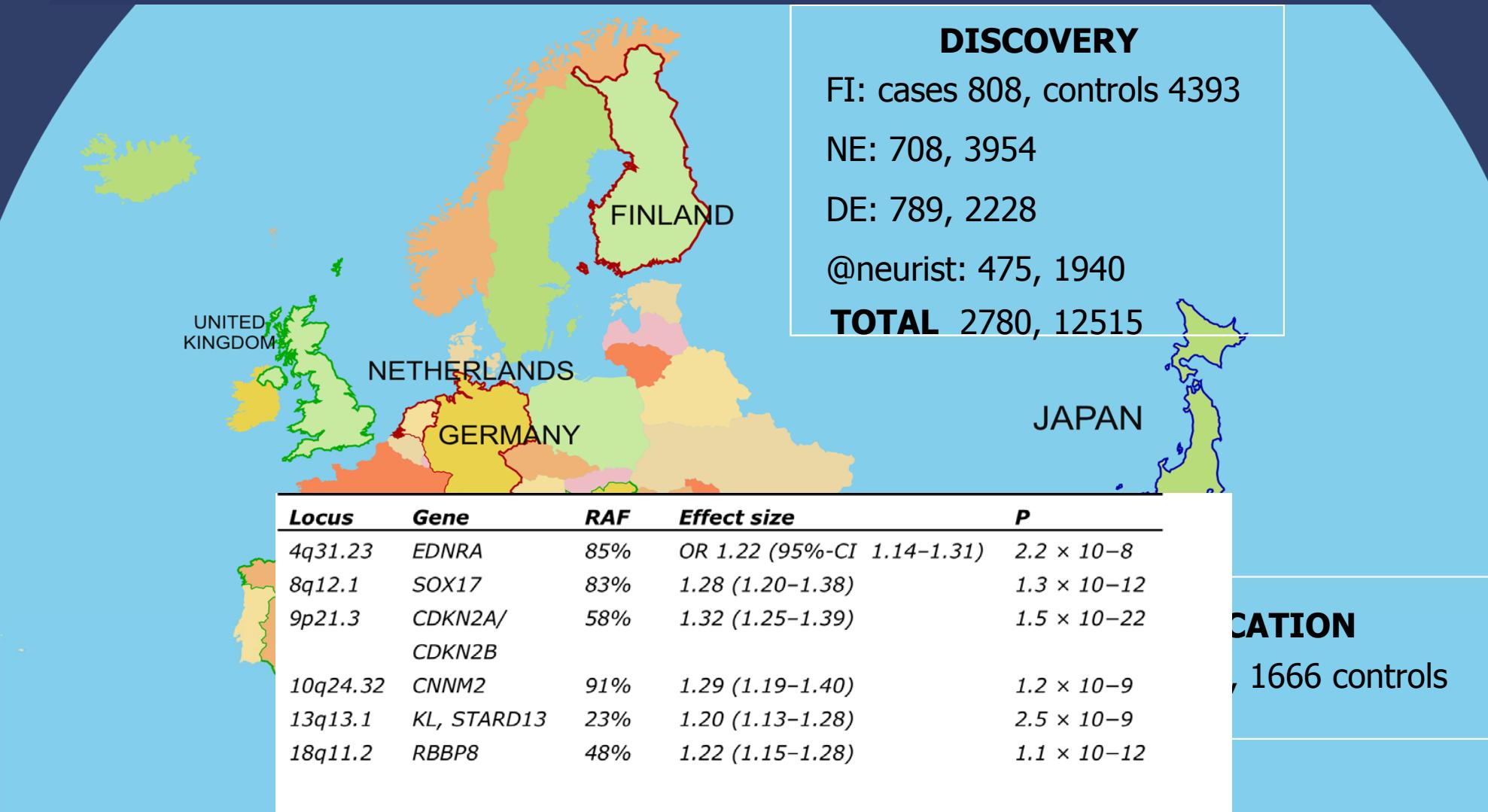
Courtesy of Emmi Tikkanen

Saccular Intracranial Aneurysm

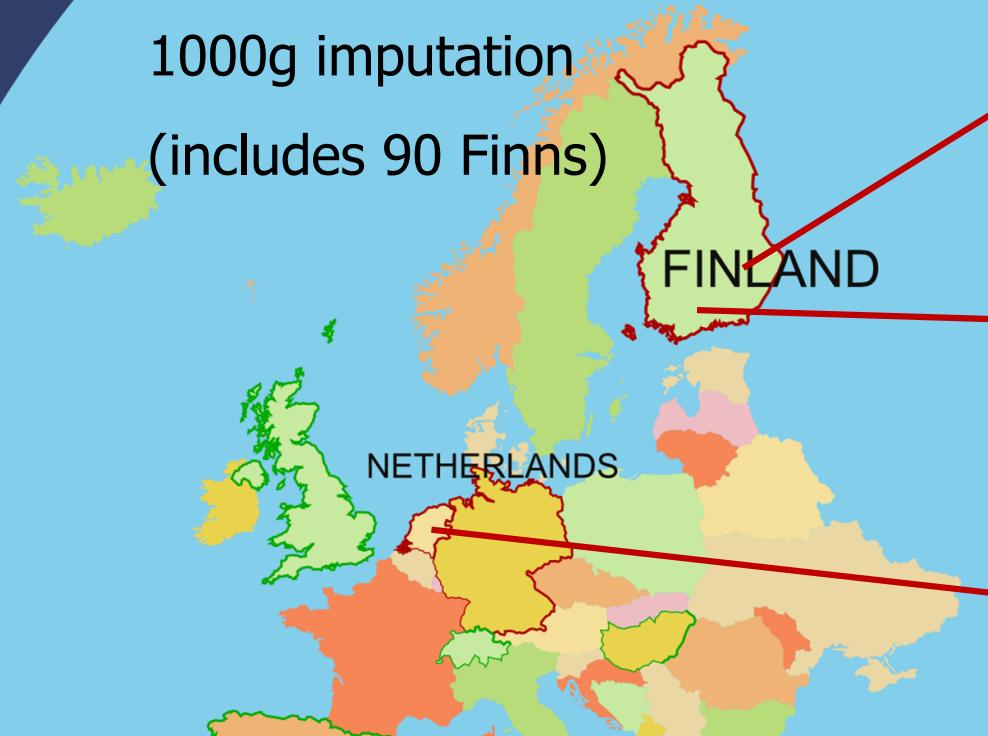
- The rupture of saccular intracranial aneurysm (sIA) is a major cause of subarachnoid hemorrhage (sIA-SAH)
- 2-3% of population carry sIA, most of which never rupture
- Complex disease affected by female sex, excess alcohol, smoking, hypertension
- ~10% of the cases have 1st degree relative with sIA (familial)



Saccular Intracranial Aneurysm (common variants)



Saccular Intracranial Aneurysm (low frequency variants)



1000g imputation

(includes 90 Finns)

DISCOVERY

FI: 760, 2513

REPLICATION

FI: 858, 4048

COMPARISON

DE: 717, 4048

Locus	Gene	RAF	Effect size	P
2q23.3	LYPD6	2%	OR 1.89	1.42×10^{-9}
5q31.1	FSTL4	2%	1.66	3.17×10^{-8}
6q24.2	EPM2A	1.6%	1.87	7.14×10^{-11}
7p22.1	RADIL	1.3%	1.59	6.08×10^{-9}

Kurki et al, in press



MASSACHUSETTS
GENERAL HOSPITAL

Intracranial aneurysm team

University of Eastern Finland
Mitja I. Kurki,
Mikko Hiltunen
Mikael von und zu Fraunberg
Timo Koivisto
Antti Ronkainen
Jaakko Rinne
Seppo Helisalmi
Juha E. Jaaskelainen

University of Helsinki, FIMM
Emlia Ilona Gal
Johannes Kettunen
Mari Kaunisto
Samuli Ripatti
Aarno Palotie
Univeristy of Turku, Finland
Terho Lehtimki
Olli T. Raitakari

University of Helsinki,
Neurosurgery
Mika Niemela
Hanna Lehto
Aki Laakso
Riku Kivilahti
Juha Hernesniemi

THL, Finland
Markus Perola
Arpo Aromaa
Johan G. Eriksson
Veikko Salomaa

University of Geneva
Tuuli Lappalainen
Emmanouil T. Dermitzakis

Univeristy of Utrecht
Femke van 't Hof
Ynte M. Ruigrok
Androniki Menelaou
Gabriel J.E. Rinkel
Paul de Bakker

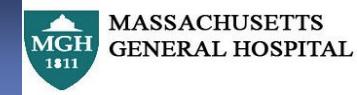
University of Nijmegen
Lambertus A.L.
Kiemeneij
Sita H. Vermeulen

Yale University
Murat Gunel

MGH/Broad
Verner Anttila
Aarno Palotie

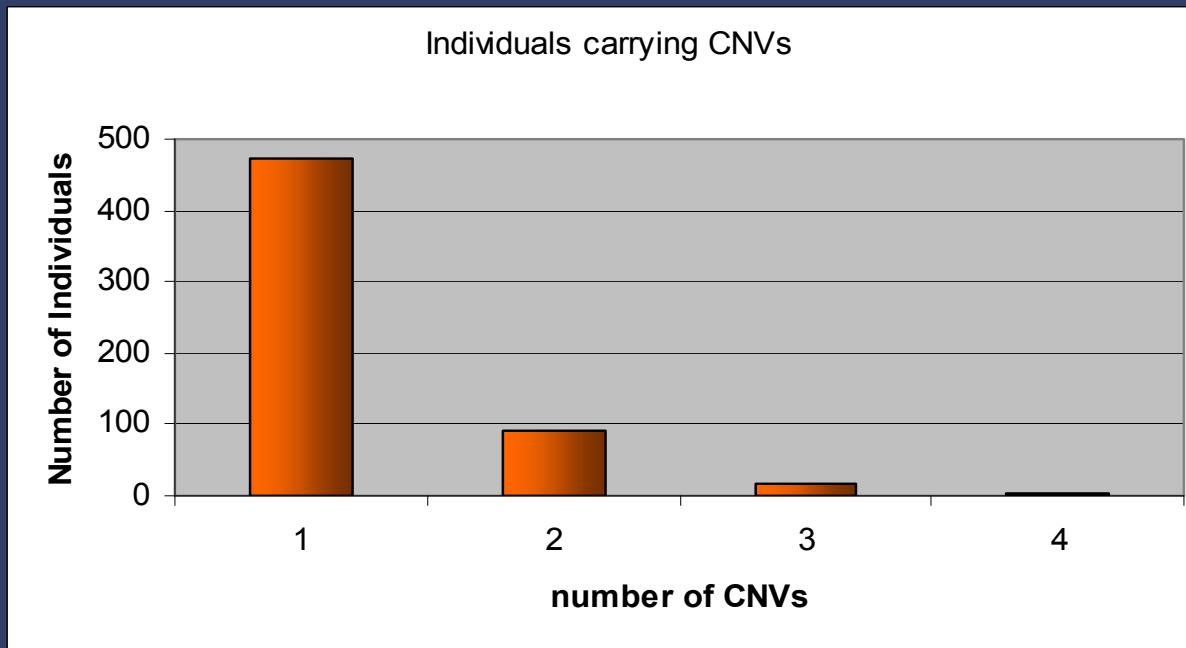


CNVs as models for enriched rare variants



CNV scan, phenotype mining

12% of 4932 individuals have
>500kb CNV

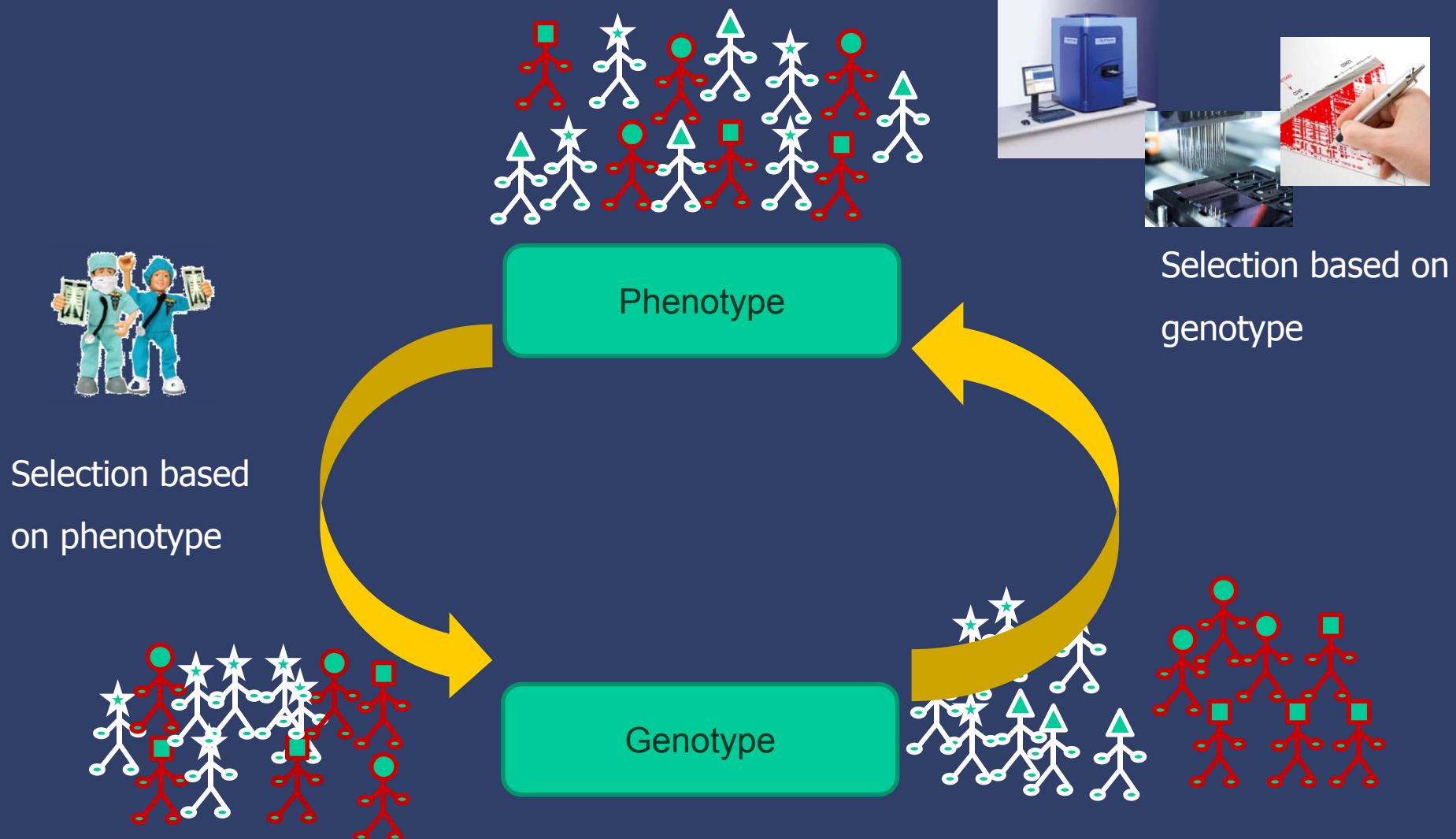


Northern Finnish Birth Cohort



165 CNV regions

From Genotype to Phenotype: strength of population cohorts



Large deletions confer risk to cognitive performance

	Number of "affected"	>500 kb deletion (160)	No CNV (4381)	P-value	OR (95 %- CI)
Low IQ (IQ<86)	74	8 (5.0%)	60 (1.4%)	0.002	3.79 (1.54-8.14)
Repeated grades in school	205	16 (10.0%)	173 (3.9%)	0.0009	2.70 (1.47-4.67)
Psychosis	102	4 (2.5%)	88 (2.0%) ^a	0.57	1.25 (0.33-3.38)
Epilepsy	79	3 (19%)	70 (1.6%)	0.74	1.18 (0.23-3.65)
Cerebral palsy and/or perinatal brain damage	66	2 (1.3%)	56 (1.3%)	1	0.98 (0.11-3.76)
Neonatal Convulsions	216	6 (3.8%)	197 (4.5%)	0.84	0.83 (0.3-1.88)
Impaired hearing	173	10 (6.3%)	141 (3.2%)	0.002	2.72 (1..38-4.95)

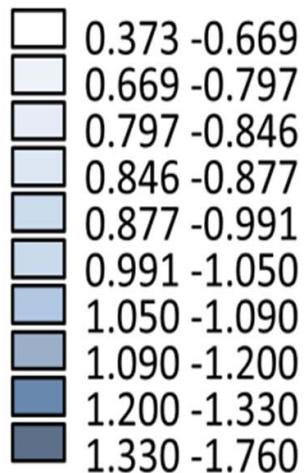
- 12 % of all cognitive defects occur together with a large deletion
- 8.5 % of redos of a grade in school occur together with a large deletion

Pietilainen et al 2011

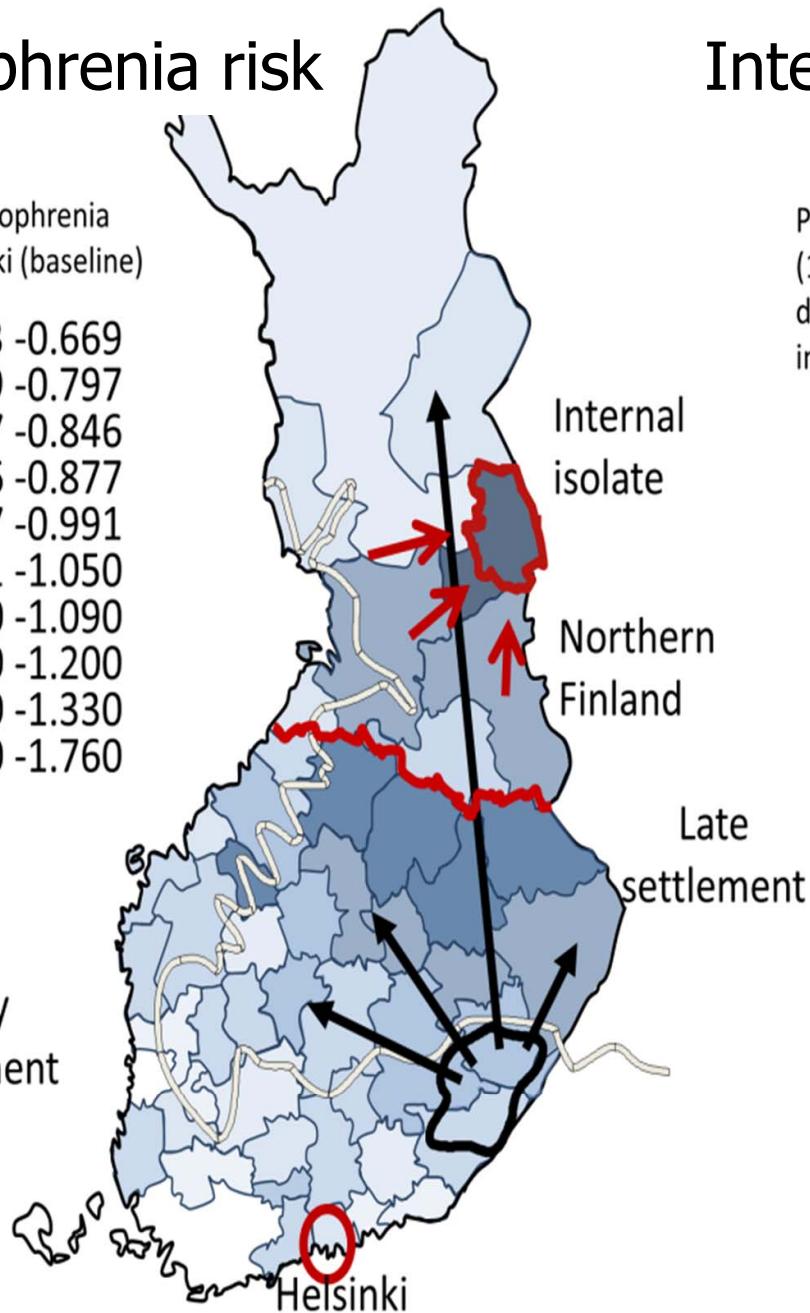


Schizophrenia risk

Relative risk of schizophrenia compared to Helsinki (baseline)

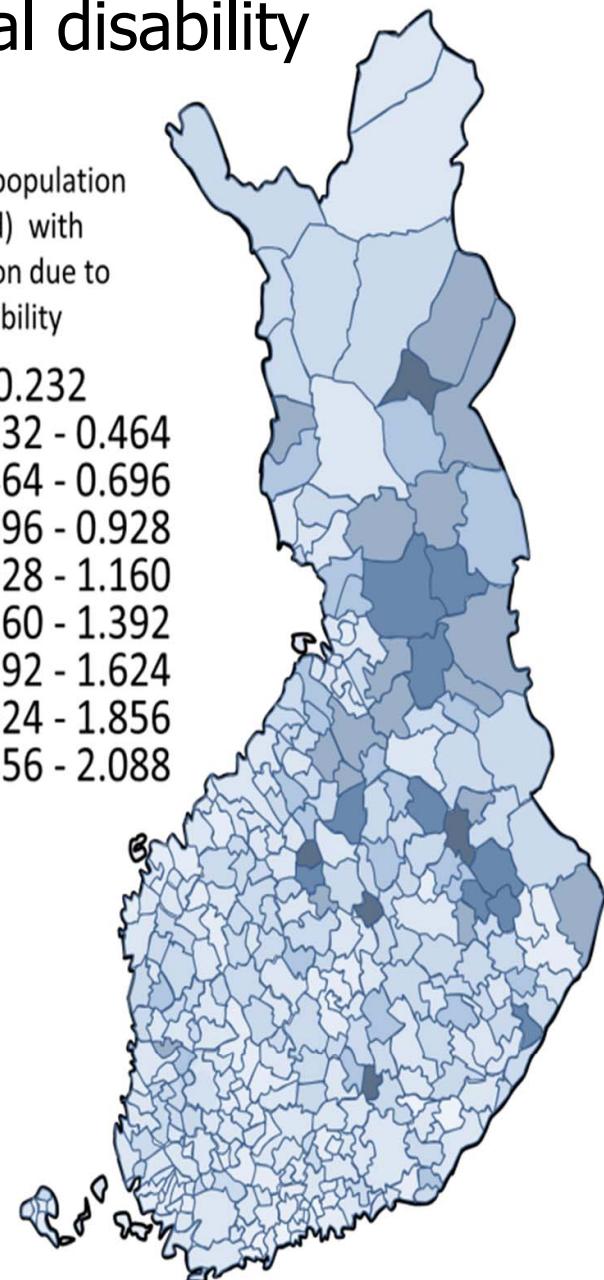
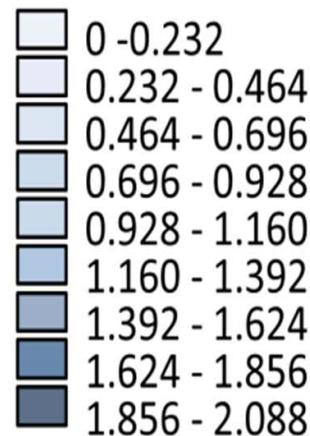


Early settlement

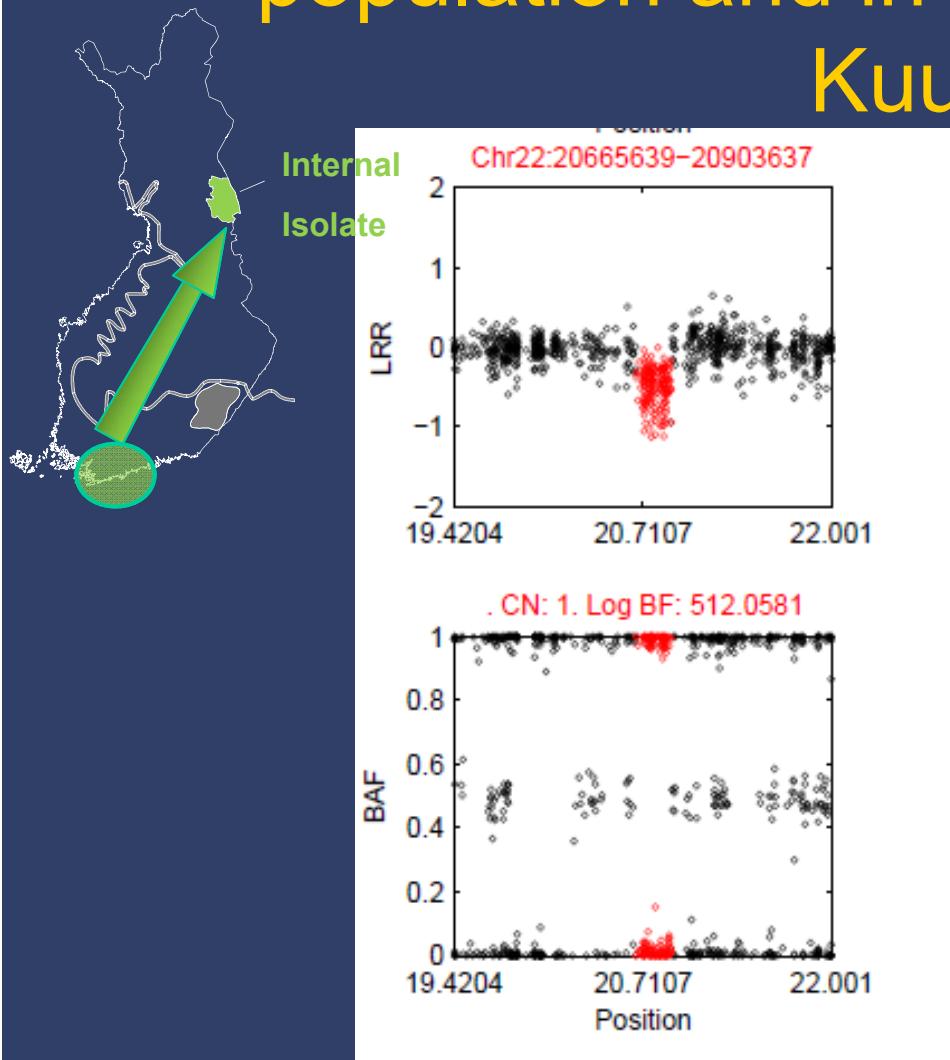


Intellectual disability

Percentage of population (15-64 years old) with disability pension due to intellectual disability



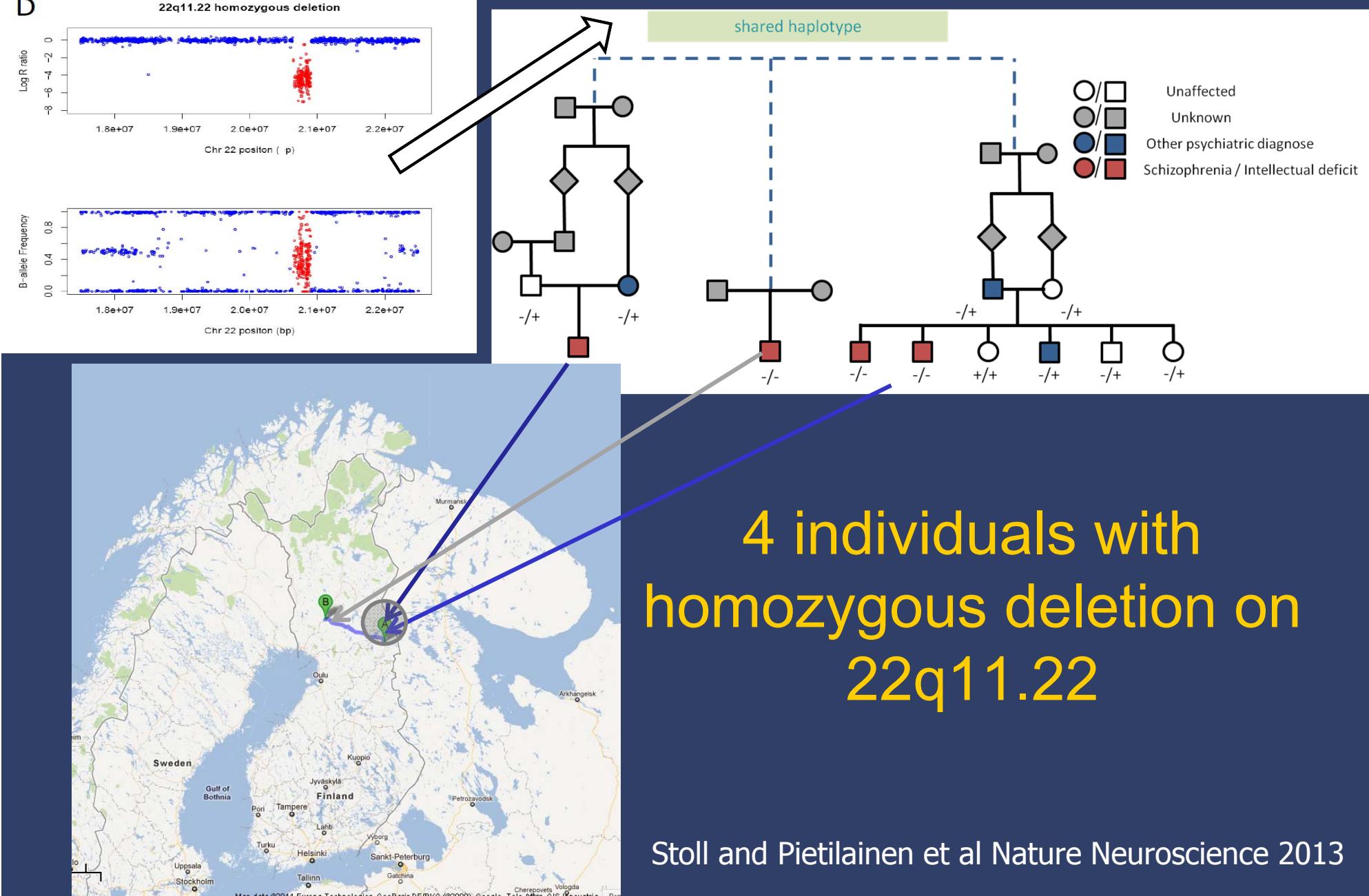
Deletion enriched in the Kuusamo population and in schizophrenics from Kuusamo



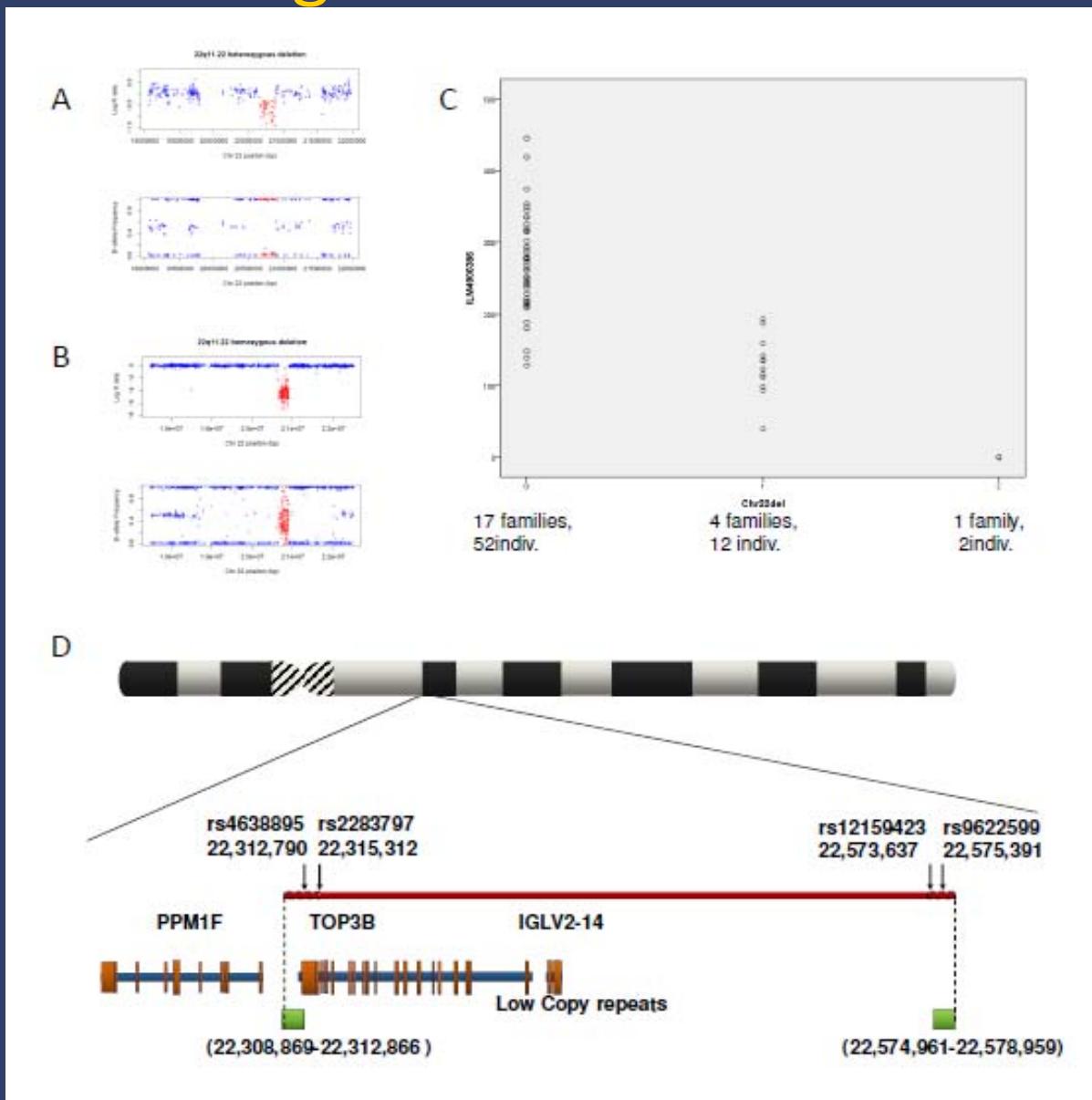
- 174 schizophrenia patients
- 747 Controls

22q11.22 Deletion		
Frequency	OR (95%-CI)	p-value
0.03	1.84 (1.05-3.23)	0.031

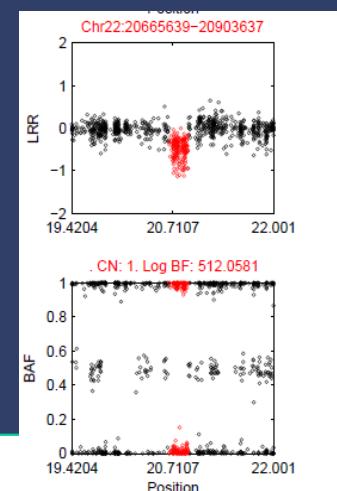
Stoll, Pietilainen et al 2013

D

The *TOP3B* gene is deleted and down regulated



Deletion on 22q11.22 is associated with schizophrenia



Sample	cases	CTRLs	22q11.22 Deletion		
			Frequency	OR (95%-CI)	p-value
Finland, Isolate	185	747	0.03	1.84 (1.05-3.23)	0.031
Finland, non-isolate	467	11124	0.003	2.63 (1.28-5.59)	0.0065
ISC and Sweden	9176	9529	0.0005	2.17 (0.81-5.80)	0.12
Combined	9828	21400	0.002	1.84 (1.18-2.87)	0.0074

Stoll and Pietilainen et al 2013

22q11.22 deletion is associated with impaired cognitive function in non-schizophrenics

NFBC66	Frequency		22q11 deletion alleles**		P	
	Affected	Unaffected	Affected	Unaffected	OR	(Fisher)
Phenotype						
Psychosis*	0.02	0.004974	0/1/24	1/46/4778	4.082	0.2242
Intellectual deficit	0.02174	0.004811	1/1/67	0/46/4735	4.597	0.03227
Repeated grades in school	0.01777	0.004513	1/5/191	0/42/4611	3.99	0.00339
Epilepsy	0	0.005132	0/0/76	1/47/4726	0	1
Neonatal convulsions	0.002392	0.005171	0/1/208	1/46/4594	0.4613	0.7239
Cerebral palsy and/or perinatal brain damage	0.008475	0.005009	0/1/58	1/46/4744	1.698	0.4519
Impaired Hearing at 14 years old	0.005814	0.005024	0/2/170	1/45/4632	1.158	0.6927

*Excluding schizophrenia

**Homozygous carriers/heterozygous carriers/ non-carriers

When 59 schizophrenia patients included:

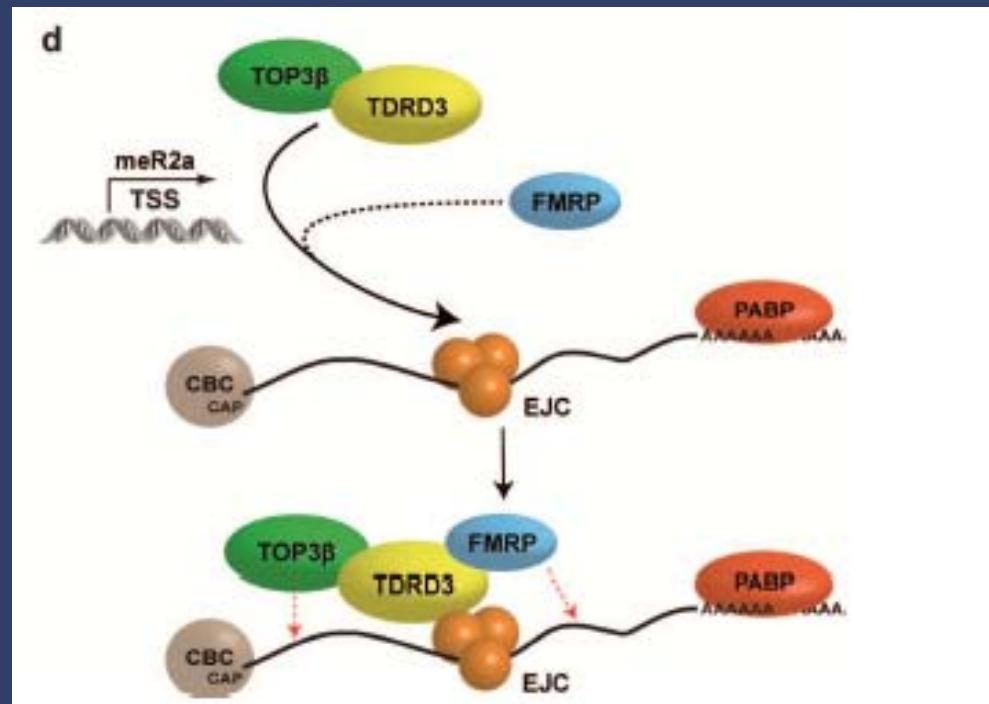
- ID: p 0.003, OR 5.4
- Repeated grades: 0.0008 OR: 4.3

Stoll and Pietilainen et al 2013



Top3B involved in regulation of translation

Top3B forms a complex with TDRD3 and FMRP (FRA(X) protein)



EJC=exon junction complex

Stoll et al Nature Neuroscience 2013

Enriched CNVs



Aarno Palotie
Samuli Ripatti
Jaakko Kaprio
Elisabeth Widen
Eveliina Jakkula
Virpi Leppä



Marius Lahti

Nelson Nelson-Williams
Chiara Sabatti
Eliza Congdon
Susan Service



Mark J Daly



Jouko Lönnqvist

Jaana Suvisaari

Tiina Paunio

A.Tuulio-Henriksson

Teppo Varilo

Johan Eriksson



Hreinn Stefansson

Stacy Steinberg



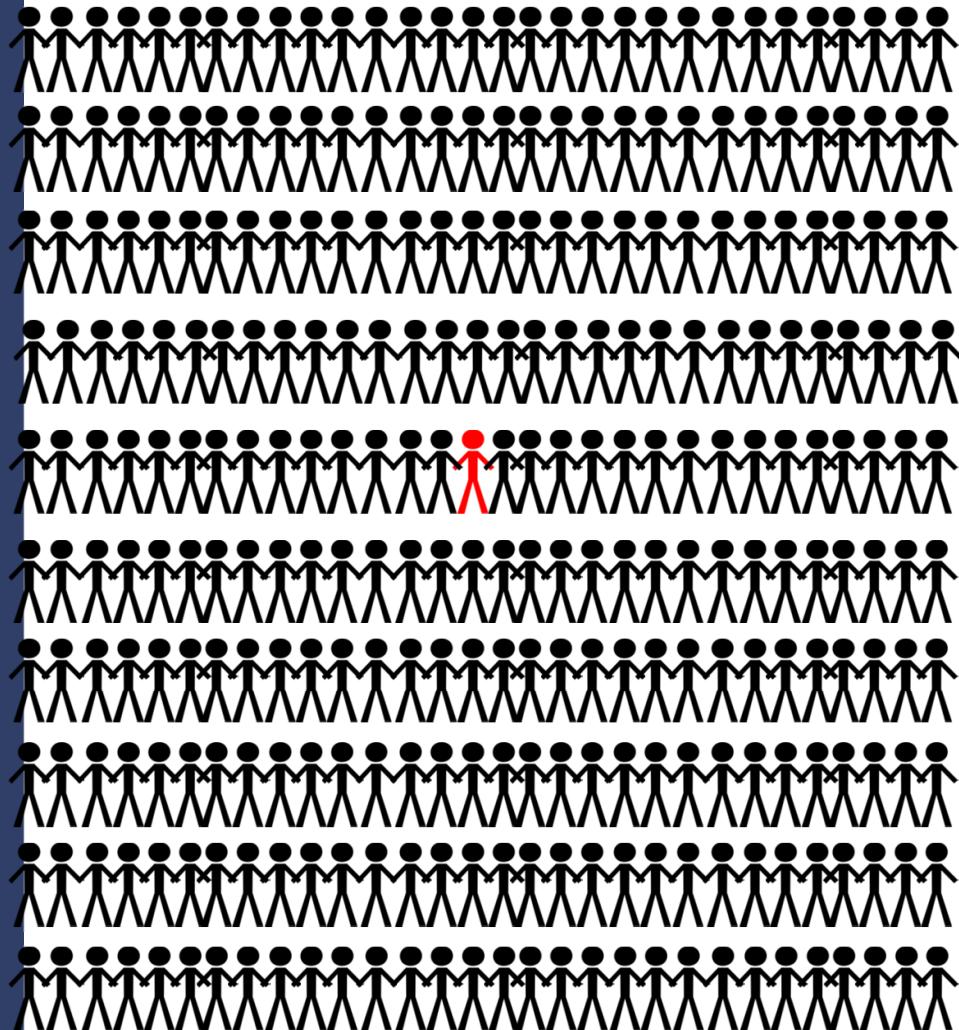
University of Oulu

Matti Isohanni

Marjo-Riitta Järvelin



How about enriched SNP variants?



Single rare variant dominating the risk ($>5 \times$ risk)



Multiple genetic and other risk factors contributing ($>2 \times$ risk)

Courtesy of Emmi Tikkanen

SISU-project

Sequencing Initiative Suomi (Finland)

The 200K

Genome wide genotype data
73 000

Genome exome sequences
 $>16\,000$



200 000 individuals
4% of the population

Population cohorts

Extensive health,
phenotype,
metabolomic data

Disease specific
collections



Imputation



Population specific
chip/genotyping



Reference database



Production sites



Whole genomes (4-5x)
(Finrisk, H2000)

On site calling

Whole exomes



Whole genomes (30x)
Whole genomes 4-5x (T2D)

On site calling

Whole exomes



Whole genomes (30x)

On site calling

Whole exomes



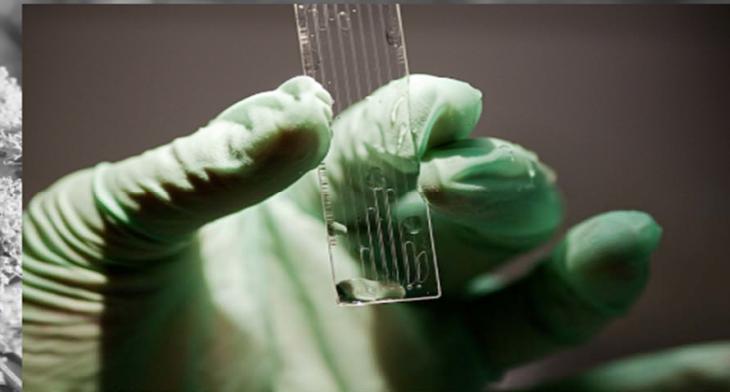
Joint calling



SISu repository



Genomic data from Finnish study collections

[Cohorts](#)

SEQUENCING INITIATIVE SUOMI (SISU) SYMPOSIUM HELSINKI, FINLAND AUGUST 26, 2014

Search

[Search](#)

rs1570248, rs184935153, rs150680234, rs201177049, rs200920925, rs201493772

[Reset](#)

Results

rsID	Chr	Coord	Minor	Major	RefSNP Alleles	N_minor	N_het	N_major	SISu	1000g
rs201493772	5	132015505	T	C	-	0	3	3322	0.0004511	N/A
rs200920925	5	114482990	A	G	-	0	3	3322	0.0004511	0
rs1570248	9	35751221	C	T	-	458	1557	1310	0.3719	0.332168
rs201177049	5	81613844	T	C	-	1	4	3320	0.0009023	N/A
rs150680234	2	233396060	C	A	-	0	5	3320	0.0007519	N/A
rs184935153	2	43958656	A	T	-	0	3	3322	0.0004511	N/A

query params: rs1570248, rs184935153, rs150680234, rs201177049, rs200920925, rs201493772

Elaine Lim



Peter Wurtz

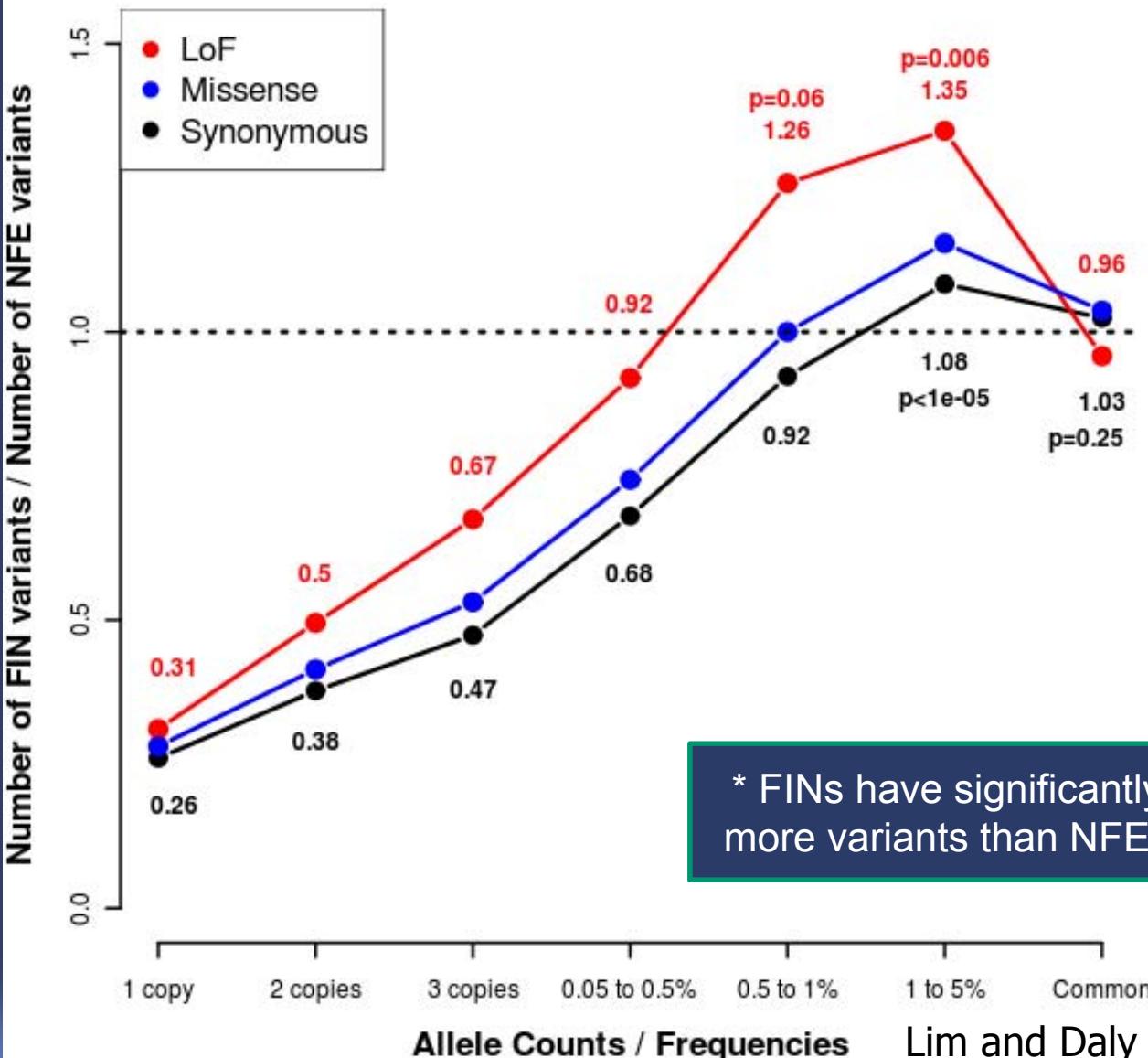
OPEN ACCESS Freely available online

PLOS GENETICS

Distribution and Medical Impact of Loss-of-Function Variants in the Finnish Founder Population

Elaine T. Lim^{1,2,3,4}, Peter Würtz^{5,6,7}, Aki S. Havulinna⁶, Priit Palta^{5,8}, Taru Tukiainen^{1,2,3}, Karola Rehnström⁸, Tõnu Esko^{2,3,9,10}, Reedik Mägi⁹, Michael Inouye¹¹, Tuuli Lappalainen^{12,13}, Yinglong Chan^{2,4,10}, Rany M. Salem^{2,10}, Monkol Lek^{1,2,3}, Jason Flannick^{2,3}, Xueling Sim¹⁴, Alisa Manning², Claes Ladenvall^{5,15}, Suzannah Bumpstead⁸, Eija Hämäläinen^{5,8}, Kristiina Aalto¹⁶, Mikael Maksimow¹⁶, Marko Salmi¹⁷, Stefan Blankenberg^{18,19}, Diego Ardissino²⁰, Svari Shah²¹, Benjamin Horne²², Ruth McPherson²³, Gerald K. Hovingh²⁴, Muredach P. Reilly²⁵, Hugh Watkins²⁶, Anuj Goel²⁶, Martin Farrall²⁶, Domenico Girelli²⁷, Alex P. Reiner²⁸, Nathan O. Stitzel²⁹, Sekar Kathiresan³⁰, Stacey Gabriel², Jeffrey C. Barrett⁸, Terho Lehtimäki³¹, Markku Laakso³², Leif Groop^{5,15}, Jaakko Kaprio^{5,33,34}, Markus Perola⁵, Mark I. McCarthy^{35,36,37}, Michael Boehnke¹⁴, David M. Altshuler^{2,3}, Cecilia M. Lindgren^{1,2,38}, Joel N. Hirschhorn^{2,10}, Andres Metspalu⁹, Nelson B. Freimer³⁹, Tanja Zeller^{18,19}, Sirpa Jalkanen¹⁷, Seppo Koskinen⁴⁰, Olli Raitakari^{41,42}, Richard Durbin⁸, Daniel G. MacArthur^{1,2,3}, Veikko Salomaa⁶, Samuli Ripatti^{5,6,8,33,43}, Mark J. Daly^{1,2,3,39*}, Aarno Palotie^{1,2,5,44}* 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.

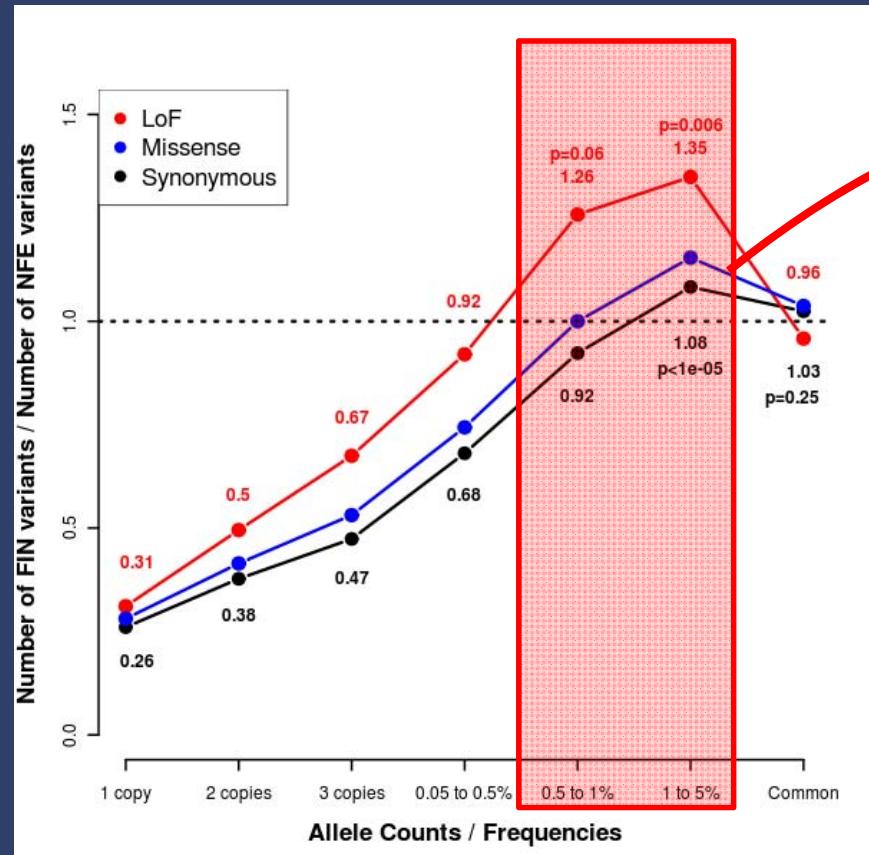
3. Boosted variants are more damaging:

And bottleneck is recent enough that selection has not eliminated them



MASSACHUSETTS
GENERAL HOSPITAL

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



Pilot study in 35,000 Finns

Traits studied include:

LDL

HDL

TG

BMI

SBP

DBP

CRP

HGF

FGF

VEGF

GALECTIN3

VitB12

G_CSF

IL4, IL6, IL10

D_DIM FIMM

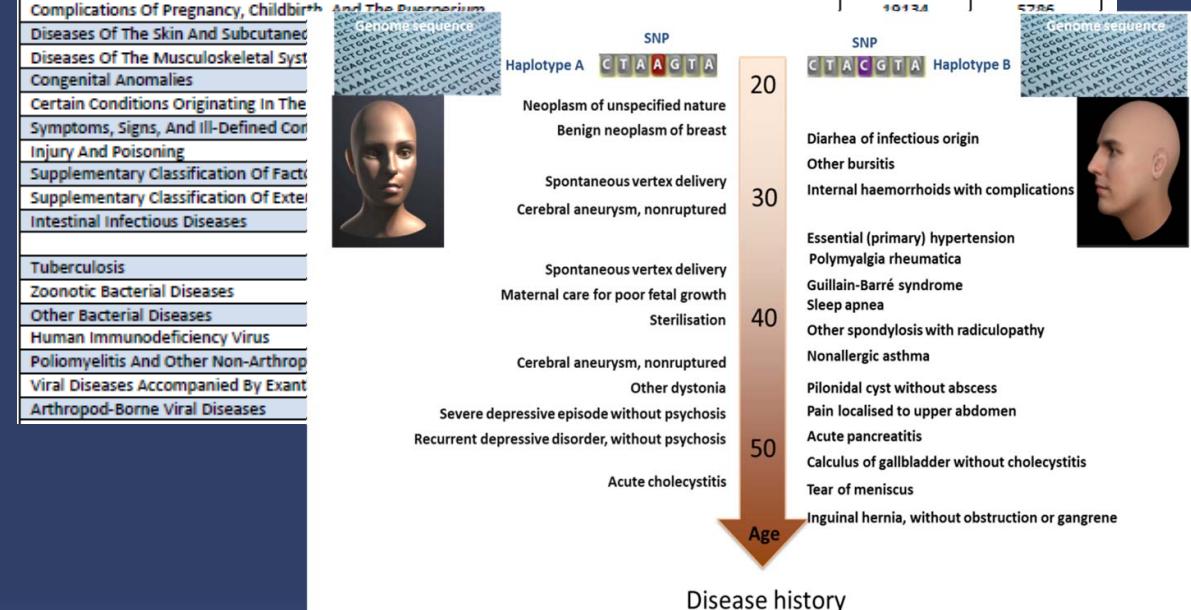
With reach through into complete medical records:

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

ICD-10 converted to ICD-9

Peter Würtz, March 11, 2013

Diagnosis	Total hospitalizations	Person hospitalizations
ICD-9 chapters		
Infectious And Parasitic Diseases	5293	3039
Neoplasms	17207	5295
Endocrine, Nutritional And Metabolic Diseases, And Immunity Disorders	7318	2805
Diseases Of The Blood And Blood-Forming Organs	1055	590
Mental Disorders	10653	2520
Diseases Of The Nervous System And Sense Organs	14279	5973
Diseases Of The Circulatory System	38019	8410
Diseases Of The Respiratory System	13547	5633
Diseases Of The Digestive System	13782	7270
Diseases Of The Genitourinary System	13363	6789
Complications Of Pregnancy, Childbirth And The Puerperium	10124	5786



Pilot study in 35,000 Finns

Traits studied include:

LDL

HDL

TG

BMI

SBP

DBP

CRP

HGF

FGF

VEGF

GALECTIN3

VitB12

G_CSF

IL4, IL6, IL10

D_DIMM

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 Würtz, March 11, 2013

Diagnosis	Total hospitalizations	Person hospitalizations
ICD-9 chapters		
Infectious And Parasitic Diseases	5293	3039
	17207	5295
	7318	2805
	1055	590
	10653	2520
	14279	5973
	38019	8410
	13547	5633
	13782	7270
	13363	6789
	10124	5786

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

Human Immunodeficiency Virus
Poliomylitis And Other Non-Arthrop.
Viral Diseases Accompanied By Exant.
Arthropod-Borne Viral Diseases

Cerebral aneurysm, nonruptured
Other dystonia
Severe depressive episode without psychosis
Recurrent depressive disorder, without psychosis
Acute cholecystitis
50
Age
Disease history
Other spondylosis with radiculopathy
Nonallergic asthma
Pilonidal cyst without abscess
Pain localised to upper abdomen
Acute pancreatitis
Calculus of gallbladder without cholecystitis
Tear of meniscus
Inguinal hernia, without obstruction or gangrene

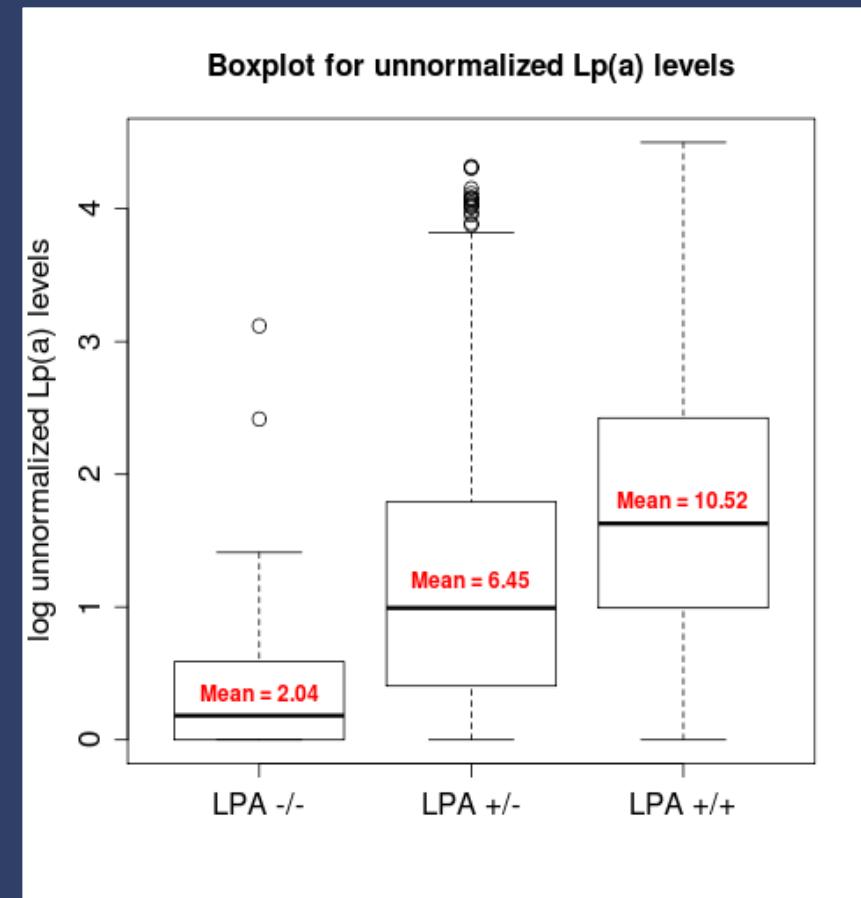
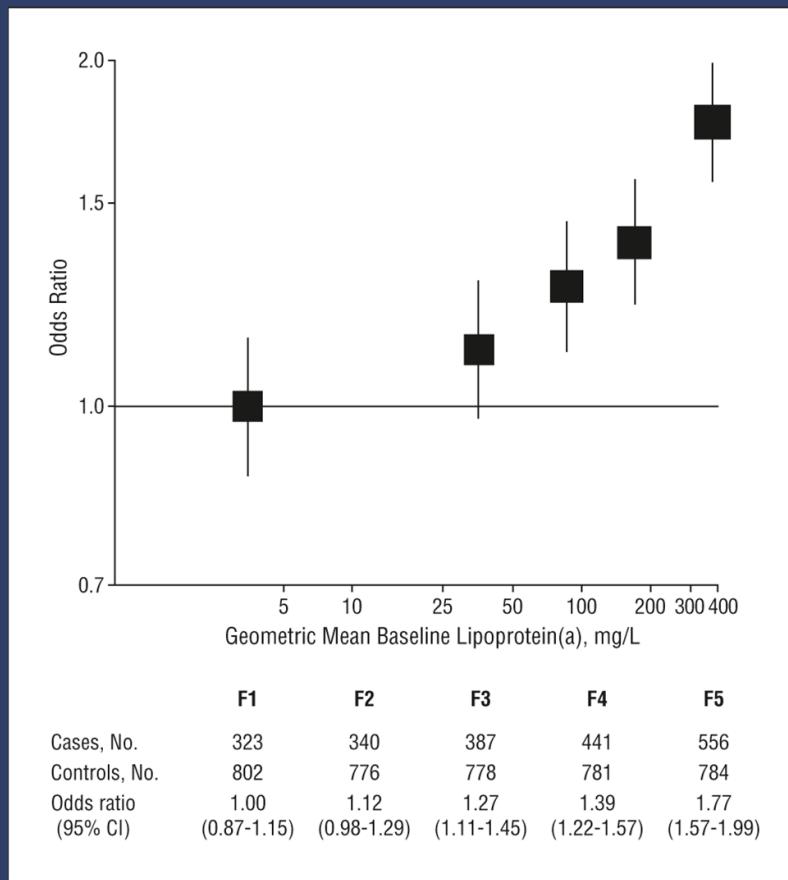
Two LPA LoF variants

LoF variant	Frequency in Finns	Frequency in non Finns
LPA1(4974)	2.8%	0.47%
LPA2(4289)	4.8%	3.6%



227 Finns LoF homozygotes

Data on cardioprotective LoF in *LPA*



Elevated Lp(a) levels known to associate with elevated CHD risk

(Bennel JAMA In Med 2008)

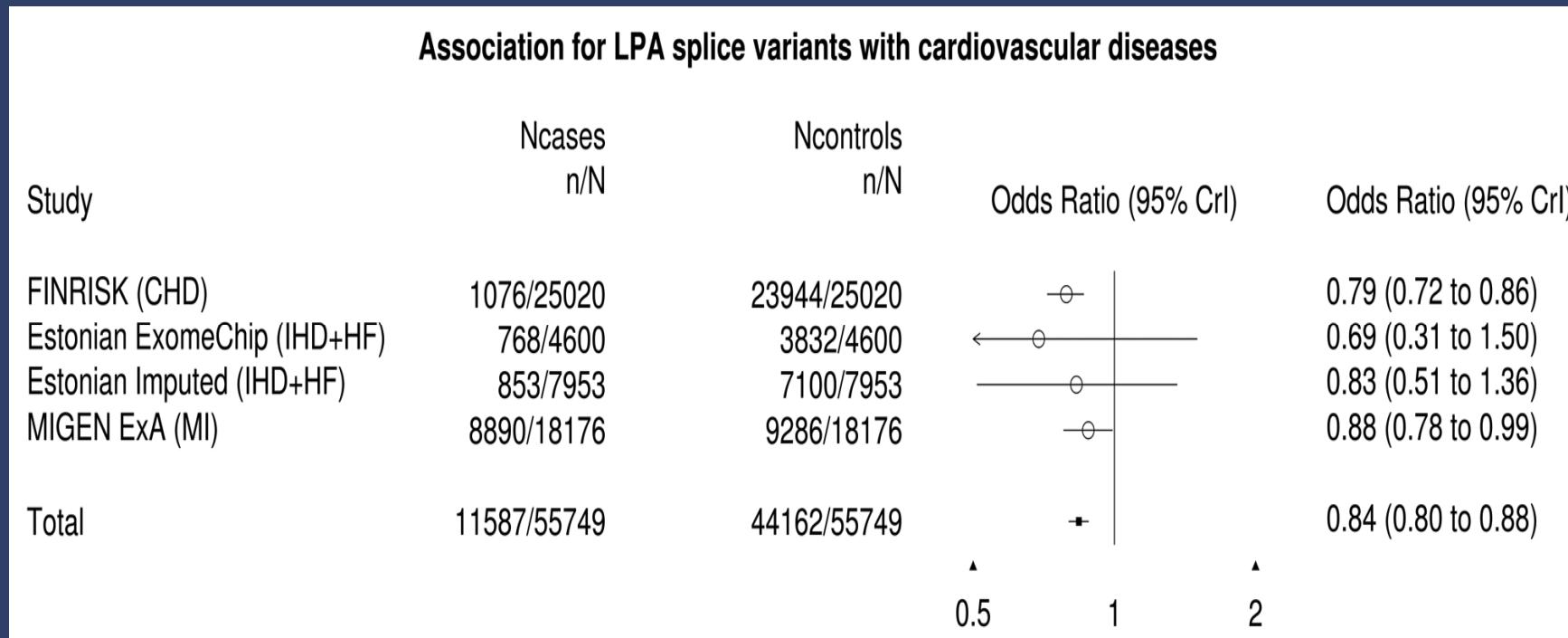
Newly identified protective LoF, 5.5x enriched in Finland (MAF=2.8%)

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

Lim et al PLoS Genetics, 2014



Similar effect in several study samples (heterozygote)



Two LPA LoF variants

LoF variant	Frequency in Finns	Frequency in non Finns
LPA1(4974)	2.8%	0.47%
LPA2(4289)	4.8%	3.6%



227 Finns LoF homozygotes



Linking to National Health Records



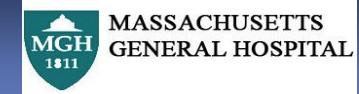
No sign of increased morbidity

SISu PIs

- FIMM: Samuli Ripatti, Jaakko Kaprio, Aarno Palotie
- THL: Veikko Salomaa, Markus Perola
- Lund, FIMM: Leif Groop
- Oxford: Mark McCarthy
- Sanger: Richard Durbin, Jeffrey Barrett
- Broad/MGH: Mark Daly, Joel Hirshhorn, David Altshuler, Daniel McArthur, Cecilia Lindgren, Aarno Palotie
- Michigan University: Michael Boehnke
- UCLA: Nelson Freimer



Same strategy for
Neuropsychiatric diseases?





Konrad Karczewski



Monkol Lek

Enriched LoFs and Neuropsychiatric traits



Autism
Schizophrenia
Cognitive impairment

15 352 LoF

p<0.05

256 LoF
Enriched in Finland

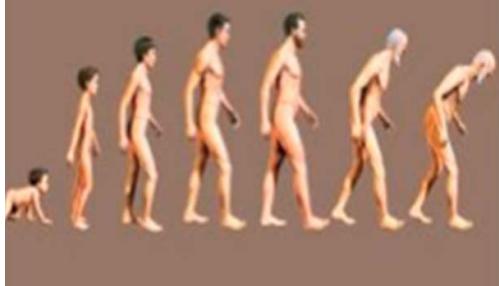
Elements to empower individuals to manage their health



Personal
biobanks



Option to
order
Genome
tests &
biomarkers
Reference
data on
health

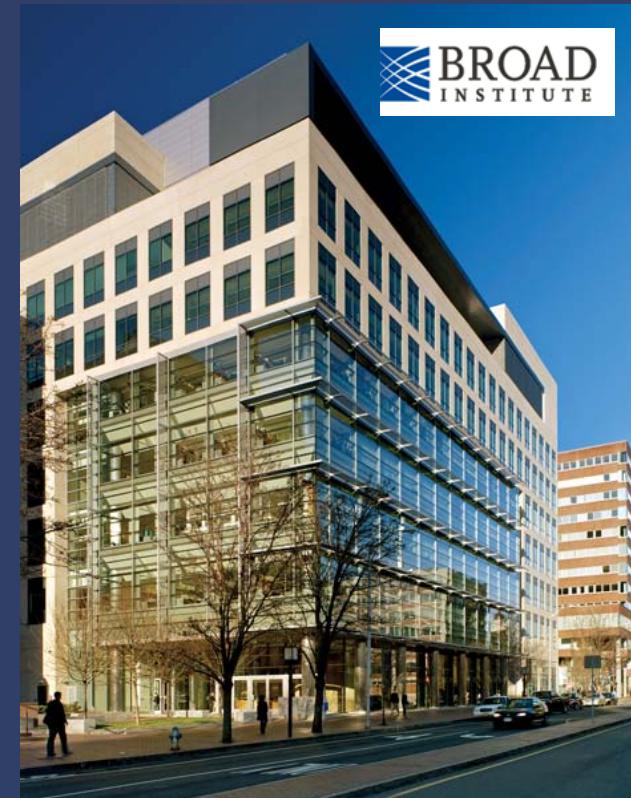


Individual at the center



Mobile health advice
(genome-enhanced)

Health care
system



FIMM

BROAD
INSTITUTE

MASSACHUSETTS
GENERAL HOSPITAL