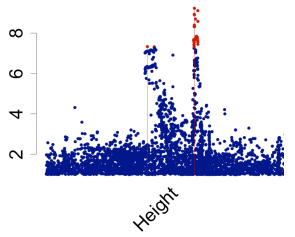
Human genetics and biometry

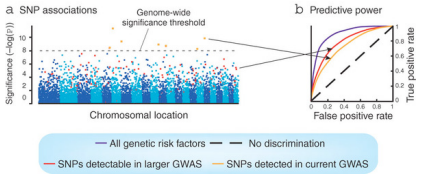
Samuli Ripatti
Hjelt Institute, Faculty of Medicine, Uni Helsinki
Institute for Molecular Medicine Finland (FIMM)
Wellcome Trust Sanger Institute, UK

Genome-wide Association Studies, Nov 26th, 2013

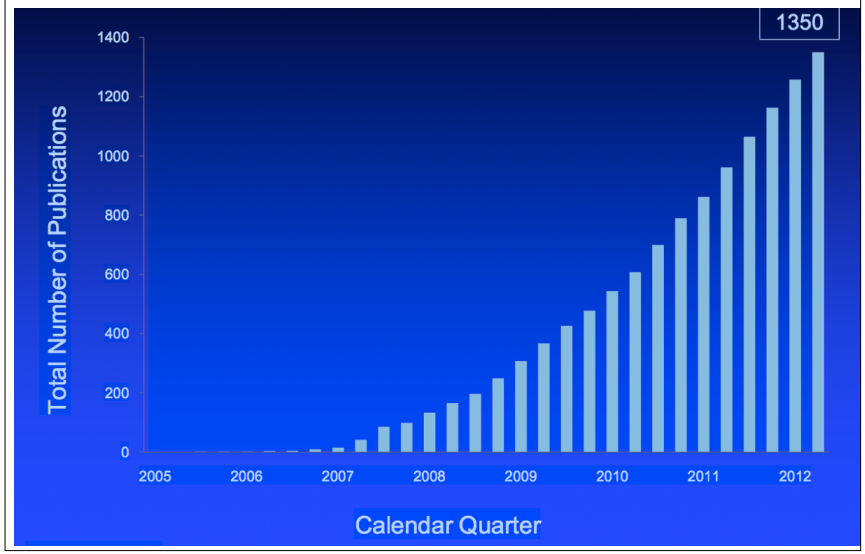
Discovering genetic loci for common diseases/ traits (80%)

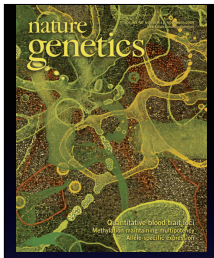


Translation (20%):
Can common SNPs help in predicting heart disease events?

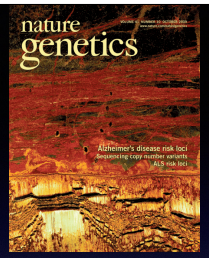


Published GWA studies

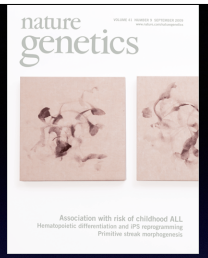




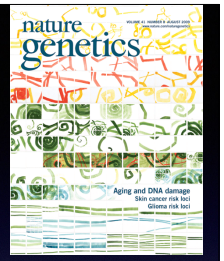
Blood markers
November 2009



Alzheimer's disease
October 2009



Childhood acute
lymphoblastic leukemia
September 2009

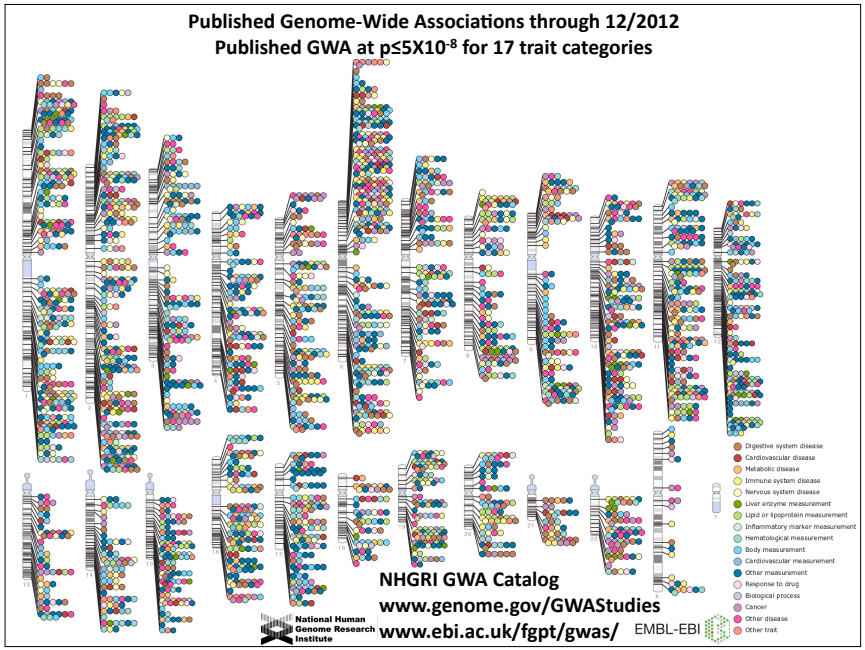


Skin cancer, Glioma
August 2009

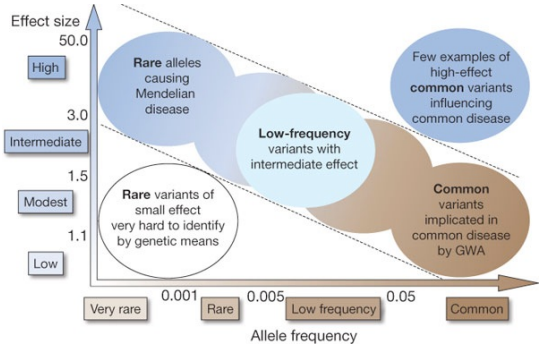
WTCCC
7 common diseases
June 2007

Smoking,
Lung cancer
April 2008

Diabetes Type 2
June 2007



Feasibility of identifying genetic variants by risk allele frequency and strength of genetic effect (odds ratio).



TA Manolio *et al. Nature* **461**, 747-753 (2009) doi:10.1038/nature08494

nature

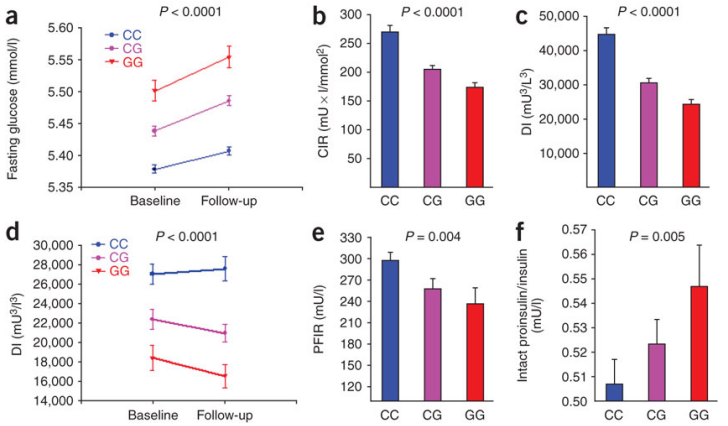
Key advances behind the GWA success

- Technology
  - unbiased and relatively cheap high-throughput genotyping
- Design and analysis
  - large scale population-based study designs, collaboration
  - control for confounding and false positive risks
- Computation
  - reliable algorithms for genotyping
  - haplotype clustering
  - genotype imputation
  - association testing and estimation

Task: Screen for loci associated with disease

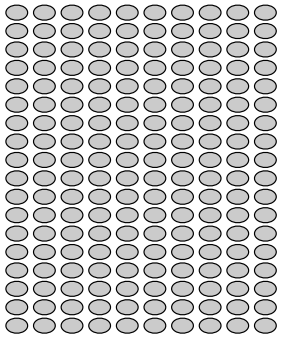
Simple?

1. Collect cases and controls and compare genotype frequencies, or
2. Compare phenotype distributions in genotype classes

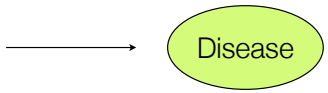


Lyssenko et al Nat Gen 2008

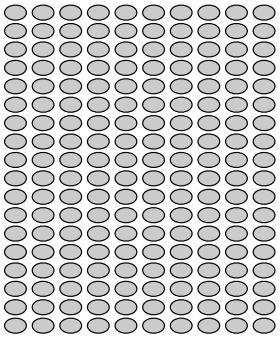
Challenge #1: Number of markers

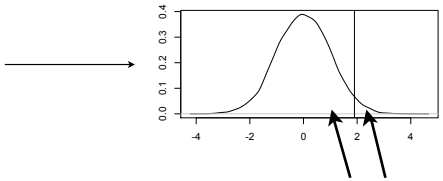


x 5 000



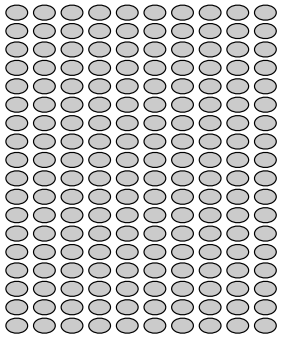
Challenge #2a: Definition of a disease

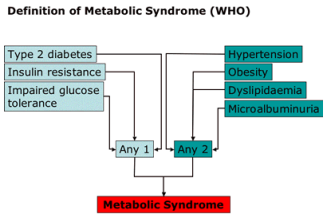




x 5 000

Challenge #2b: Definition of a disease



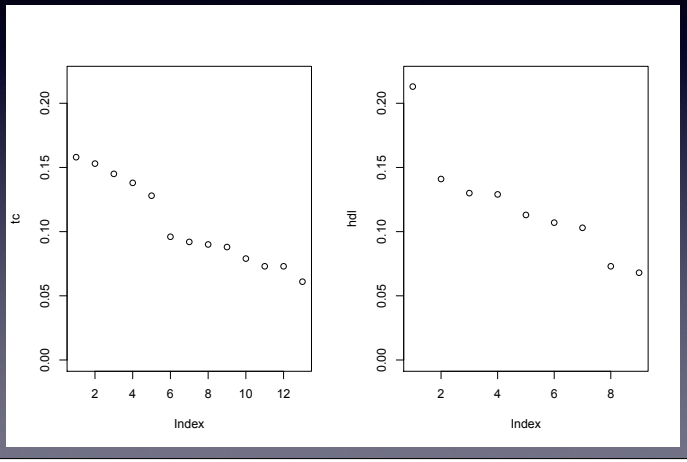


x 5 000

What level of single gene effects can we assume for common alleles?

- OR<1.2
- differences in means < 0.2*sd

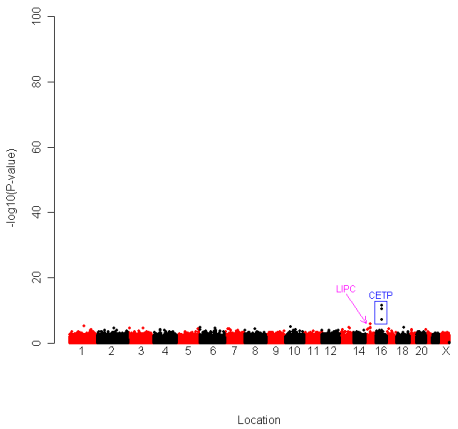
Effect sizes in lipid meta-analysis



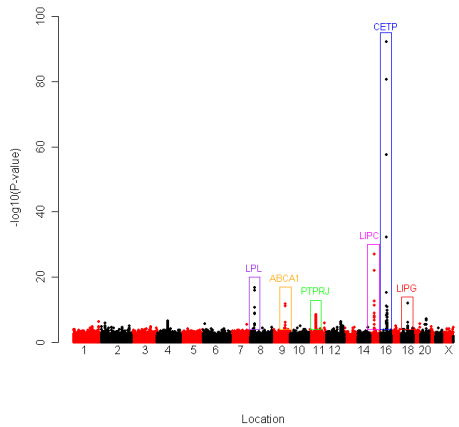
Best case power

- Assuming:
  - effect size=0.2*sd
  - allele frequency = 0.5
  - no genotyping error
  - no confounding
  - GWA-threshold = 10^{-7}
- Power:
  - n=1000: 19%
  - n=1500: 56%

n = 2000



n = 20 000



Simple maths for power

Allele	Case	Control
C	a	b
A	c	d

OR = cb / ad

var(OR) = 1/a + 1/b + 1/c + 1/d

Case common allele MAF=0.5

Allele	Case	Control
C	400	500
A	600	500

OR = 600*500 / 400*500 = 1.5

var(OR) = 1/400 + 1/500 + 1/500 + 1/600 = 0.00817

CI = [1.26,1.79]

Case still a common allele MAF=0.1

Allele	Case	Control
C	865	900
A	145	100

OR = 145*900 / 865*100 = 1.5

var(OR) = 1/865 + 1/900 + 1/100 + 1/145 = 0.019

CI = [1.14,1.97]

Case rare allele, MAF=0.01

Allele	Case	Control
C	985	990
A	15	10

OR = 15*990 / 985*10 = 1.5

var(OR) = 0.985 + 0.990 + 0.10 + 0.063 = 0.169

CI = [0.67,3.36]

Designs enriching rare alleles called for: e.g. study population isolates

SISU-project

Sequencing Initiative Suomi

Genome wide genotype data
50 000

Genome and exome sequences
GOT2D, NFBC, **UK10K**, Sanger
9 000

National Biobank

200 000 individuals
4% of the population

Population cohorts

Extensive health, phenotype, metabolomic data

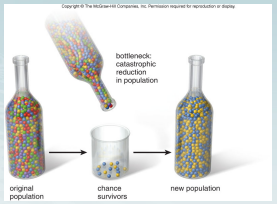
Disease specific collections

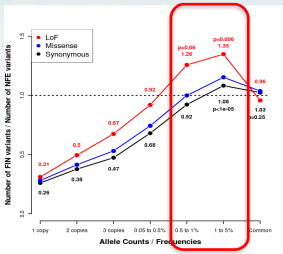
Population specific chip/genotyping

UNIVERSITY OF OXFORD · BROAD INSTITUTE · UK 10K · SPH · FIMM · BRAIN RESEARCH INSTITUTE UCLA · LUND UNIVERSITY Faculty of Medicine · BROAD INSTITUTE · TERVEYDEN JA HYVINVOINNIN LAITOS

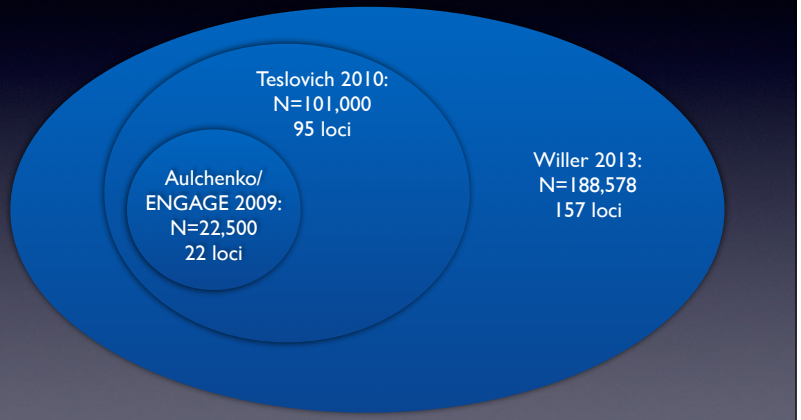
Finland:
Population history and genetic resources

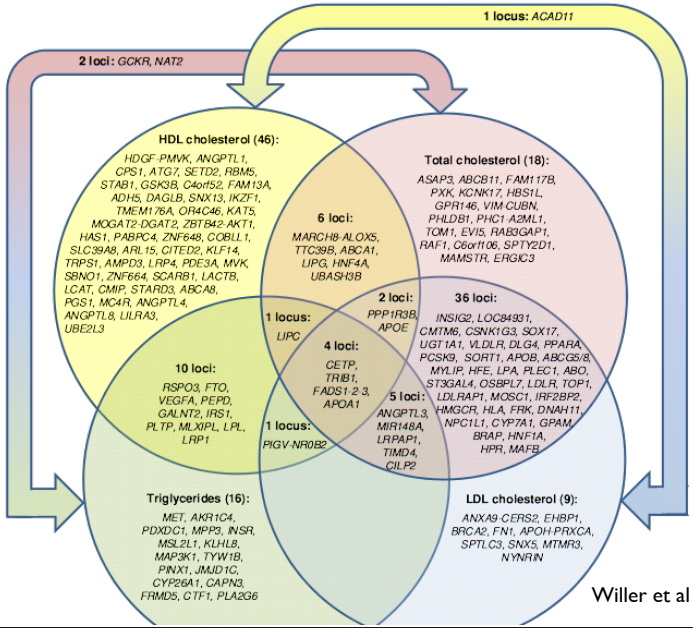
- Population isolate
  - Small number of founders
  - Bottleneck: Less private and rare variants
  - Fast population growth
  - Enrichment of LoF variants in 0.5-5% allele frequency
- Genetic resources:
  - DNA biobank of 200 000 individuals
  - 5000+ WGS & WES samples (Finrisk, Botnia)
  - 50,000 GWAS samples (many cohorts)





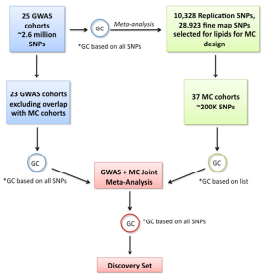
Outcome of GWAS studies: lipids



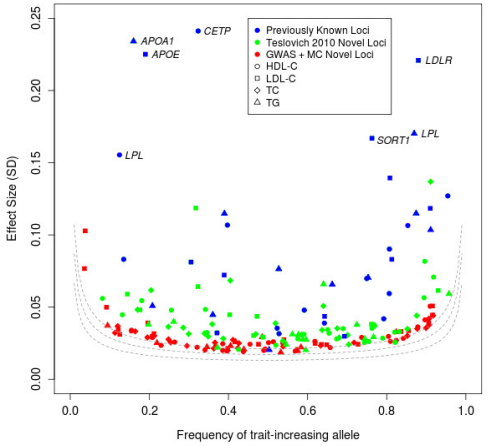


Willer et al 2013 *Nat Genet*

Allele freq and effect size

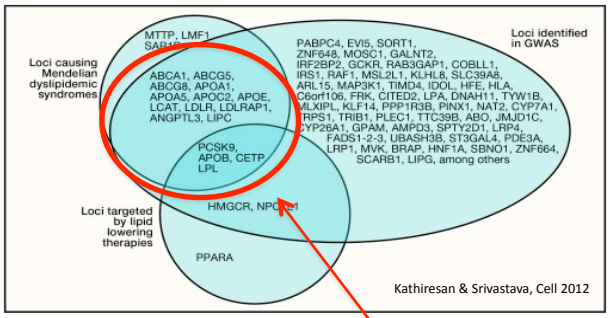


Study design



Willer et al 2013 *Nat Genet*

Overlap with Mendelian lipid genes

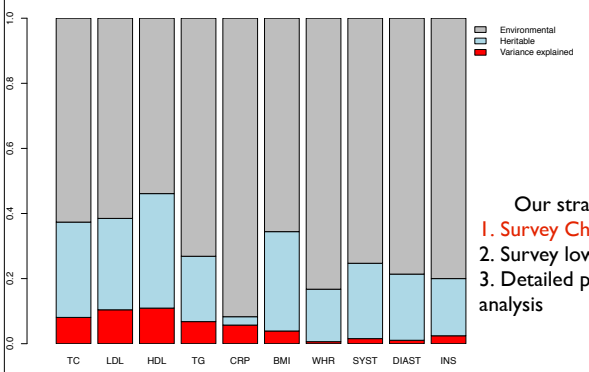


HDL-C
LDL-C
Triglycerides
Total cholesterol

Are these associations driven by the same underlying haplotypes?

Heritability and explained variance

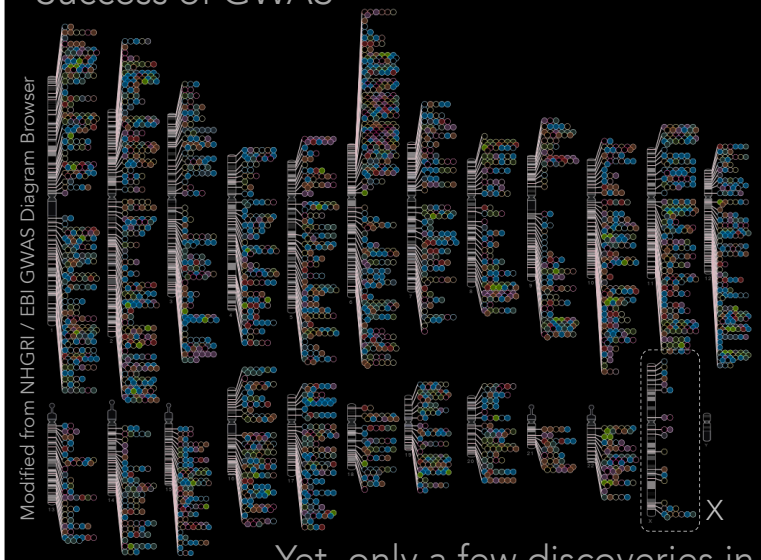
Estimated in NFBC66 cohort



Our strategies to further discovery:

1. Survey ChrX
2. Survey low frequency variants
3. Detailed phenotyping / multivariate analysis

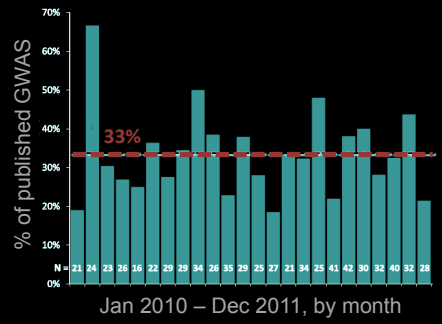
Success of GWAS



Yet, only a few discoveries in chrX

ChrX underanalyzed in GWAS

ChrX included in 33% of GWAS



Modified from
Wise AL et al. AJHG (2013) 92, 643–7

ChrX analyses require special attention

But also hold potential for new discoveries

Aim:
Comprehensively assess the contribution of chrX to complex traits

Key questions

1. How much variability does chrX explain?
2. Can we find any associated loci in chrX?
3. Are the loci fully dosage compensated?

Materials

N = 19,701 +
N = 5,032

7 discovery cohorts:
NFBC1966
YFS
COROGENE
HBCS
GenMets
PredictCVD
DGI

Replication:
FINRISK

404,862 chrX SNPs

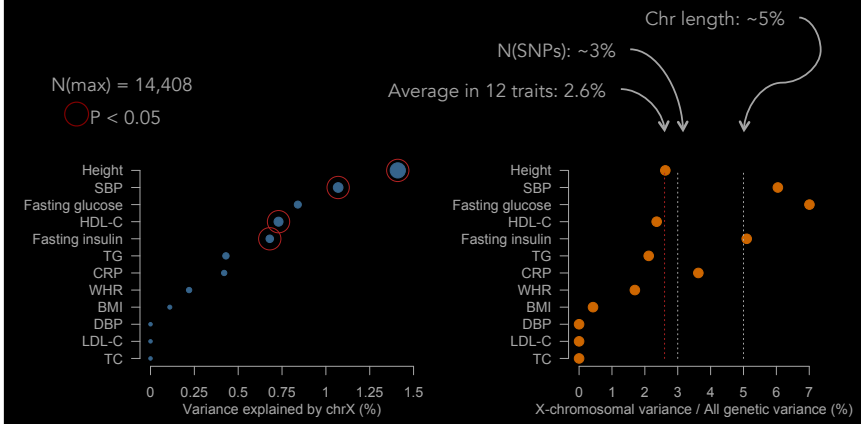
1kG reference (April 2012)

MAC > 3 / cohort
Info > 0.4

12 phenotypes

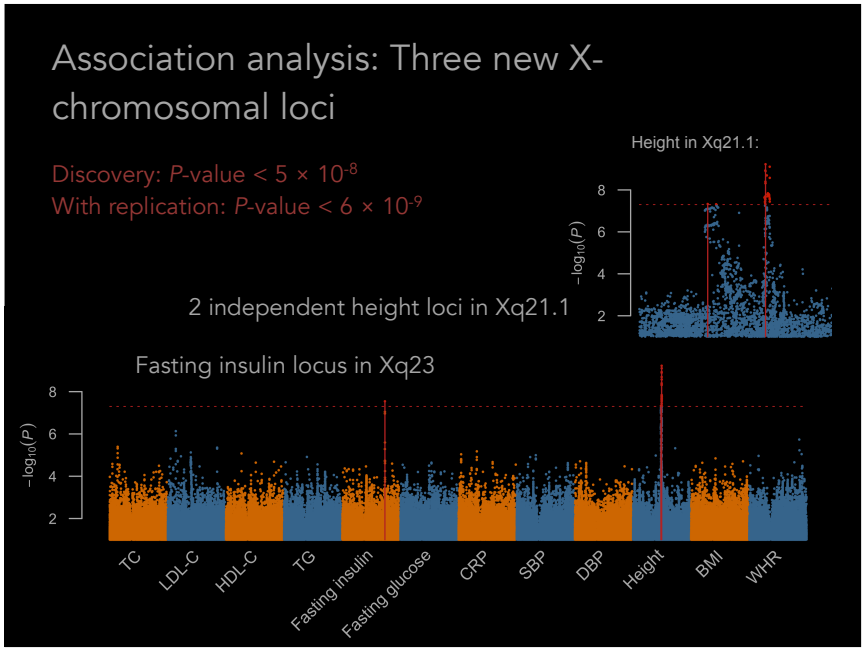
CRP
TC
TG
HDL-C
LDL-C
Glucose
Insulin
Systolic BP
Diastolic BP
Height
BMI
WHR

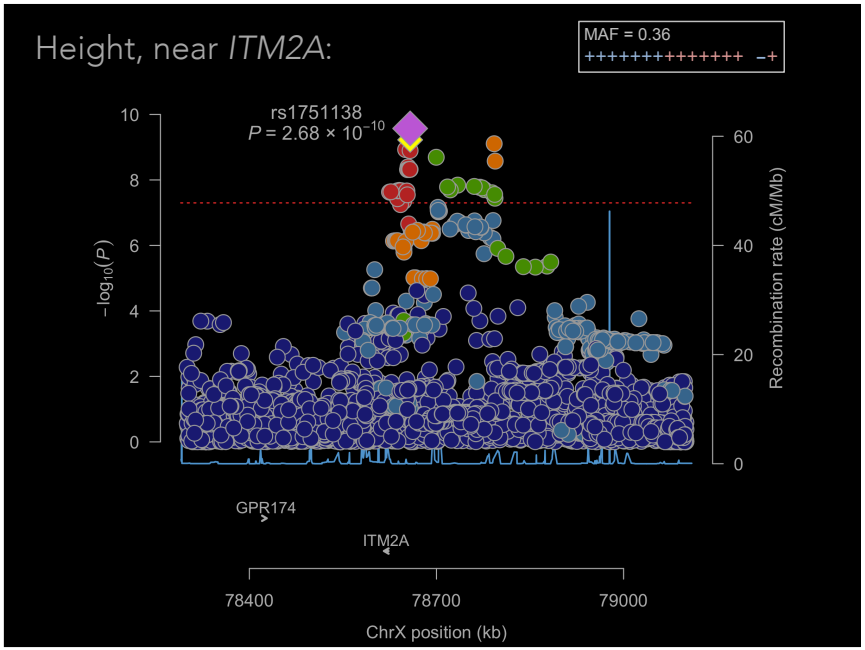
Estimates of the variance attributable to chrX

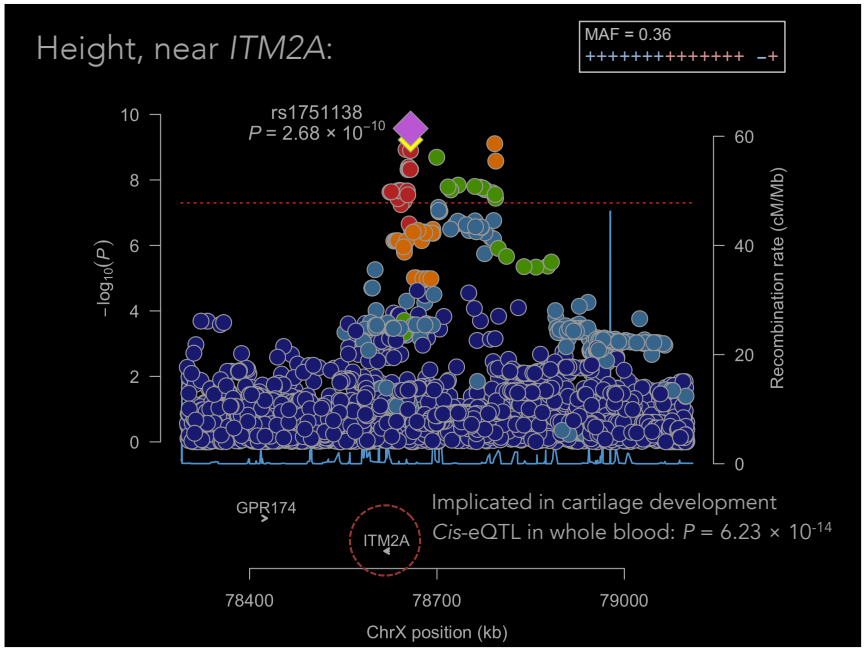


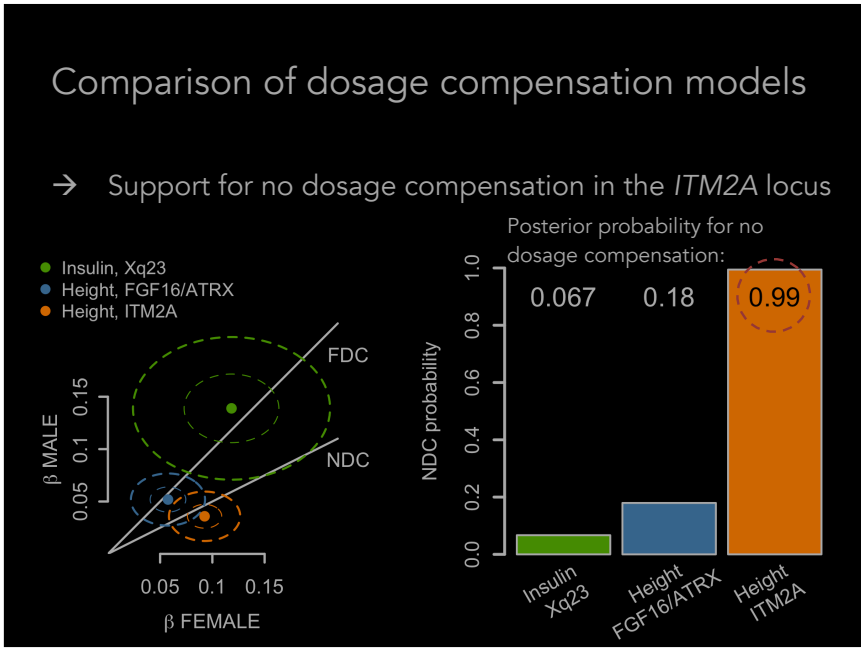
→ ChrX contributes to many complex phenotypes

→ With a proportion equal to the number of SNPs in the chr


Association analysis: Three new X-chromosomal loci
Discovery: P-value < 5 × 10-8
With replication: P-value < 6 × 10-9
Height in Xq21.1:
2 independent height loci in Xq21.1
Fasting insulin locus in Xq23
TC
LDL-C
HDL-C
TG
Fasting insulin
Fasting glucose
CRP
SBP
DBP
Height
BMI
WHR


Height, near ITM2A:
MAF = 0.36
rs1751138
P = 2.68 × 10−10
GPR174
ITM2A
78400
78700
79000
ChrX position (kb)
Recombination rate (cM/Mb)


Height, near *ITM2A*:
MAF = 0.36
rs1751138
$P = 2.68 \times 10^{-10}$
$-\log_{10}(P)$
Recombination rate (cM/Mb)
GPR174
ITM2A
Implicated in cartilage development
Cis-eQTL in whole blood: $P = 6.23 \times 10^{-14}$
78400
78700
79000
ChrX position (kb)


Comparison of dosage compensation models
→ Support for no dosage compensation in the *ITM2A* locus
Insulin, Xq23
Height, FGF16/ATRX
Height, ITM2A
FDC
NDC
β MALE
β FEMALE
Posterior probability for no dosage compensation:
0.067
0.18
0.99
NDC probability
Insulin Xq23
Height FGF16/ATRX
Height ITM2A

Evidence for incomplete dosage compensation in *ITM2A*

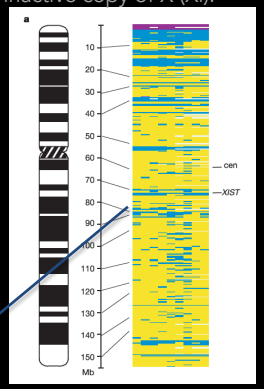
+ Association data consistent with no dosage compensation
+ Higher *ITM2A* expression in women in whole blood (N = 513)
+ *ITM2A* known to variably escape from XCI

Carrel L et al. Nature (2005) 434, 400-4

56% of cells express *ITM2A* from Xi

ITM2A	Integral membrane protein 2A		5 / 9

Gene expression from the inactive copy of X (Xi):



Case: *ITM2A*

- Nearby SNPs associate with height
- Implicated in chodrogenesis
- *Cis*-eQTL: expression ↑ height ↓
- Escapes from X chromosome inactivation

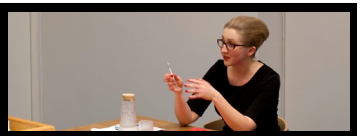
Case: *ITM2A*

- Nearby SNPs associate with height
- Implicated in chodrogenesis
- *Cis*-eQTL: expression ↑ height ↓
- Escapes from X chromosome inactivation

→*ITM2A* contributes to the sexual dimorphism in height?
→yes, explains 1.5% of male-female differences

RunPhoto/Getty Images

Conclusions



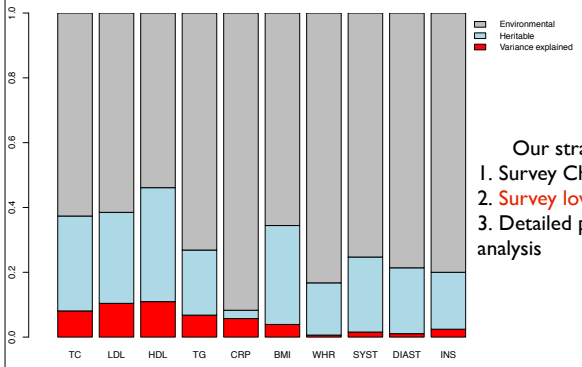
Taru Tukiainen, MGH & Broad

ChrX harbors new loci and interesting biology:

- Variance estimates: ~3% of GWAS loci may be X-chromosomal
- Association analysis: Low-hanging fruits and more to discover in chrX
- Dosage compensation: Clues to sexually dimorphic traits?

Heritability and explained variance

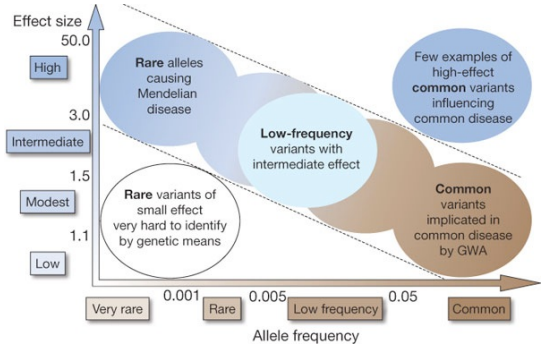
Estimated in NFBC66 cohort



Our strategies to further discovery:

1. Survey ChrX
2. Survey low frequency variants
3. Detailed phenotyping / multivariate analysis

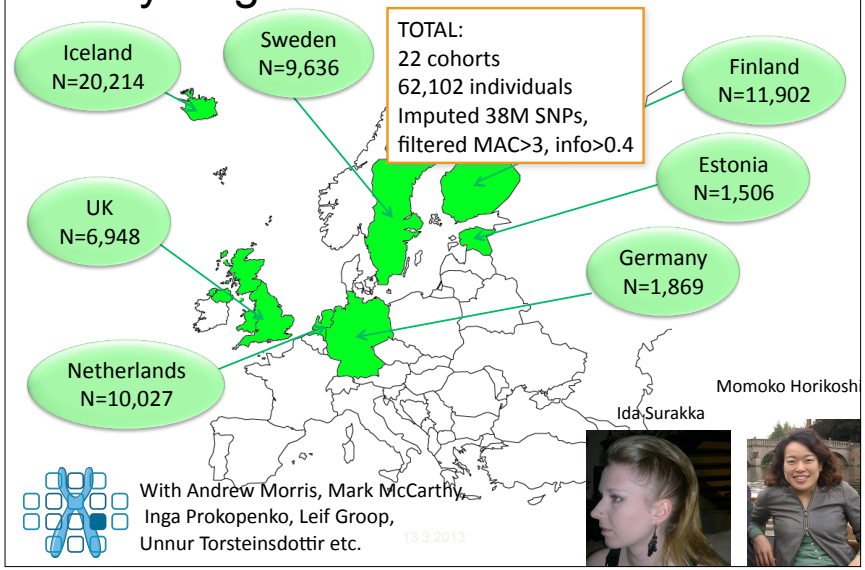
Feasibility of identifying genetic variants by risk allele frequency and strength of genetic effect (odds ratio).



TA Manolio *et al. Nature* **461**, 747-753 (2009) doi:10.1038/nature08494

nature

Recycling the ENGAGE GWAS data



Genotype imputation

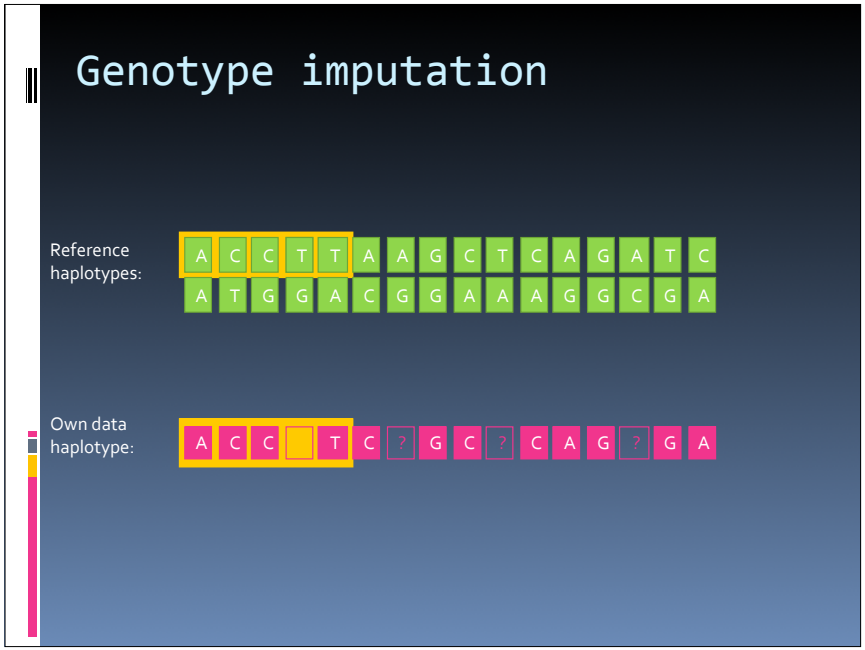
Reference haplotypes:

A	C	C	T	T	A	A	G	C	T	C	A	G	A	T	C
A	A	A	A	A	C	G	G	A	A	A	G	G	C	G	A

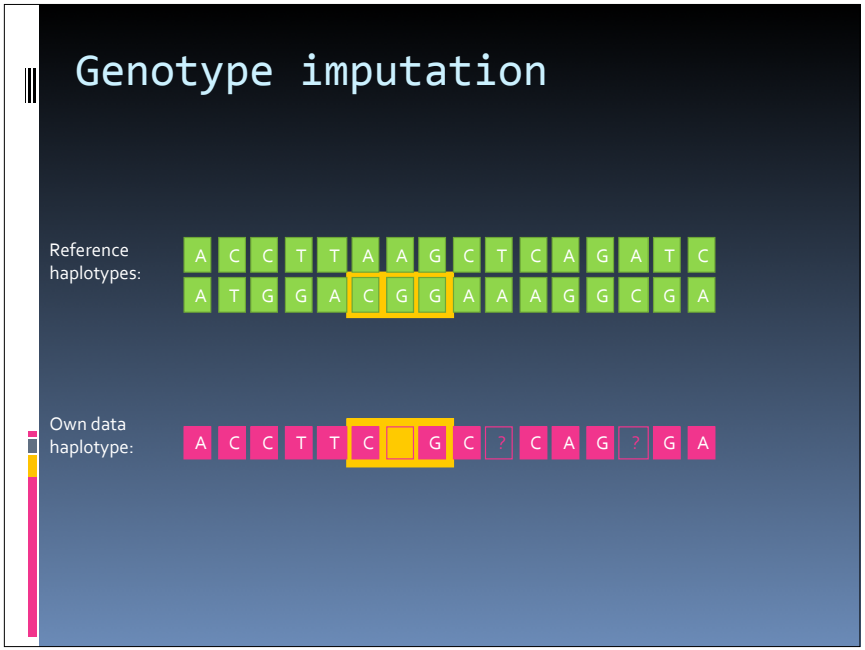
Own data haplotype:

A	C	C	?	T	C	?	G	C	?	C	A	G	?	G	A

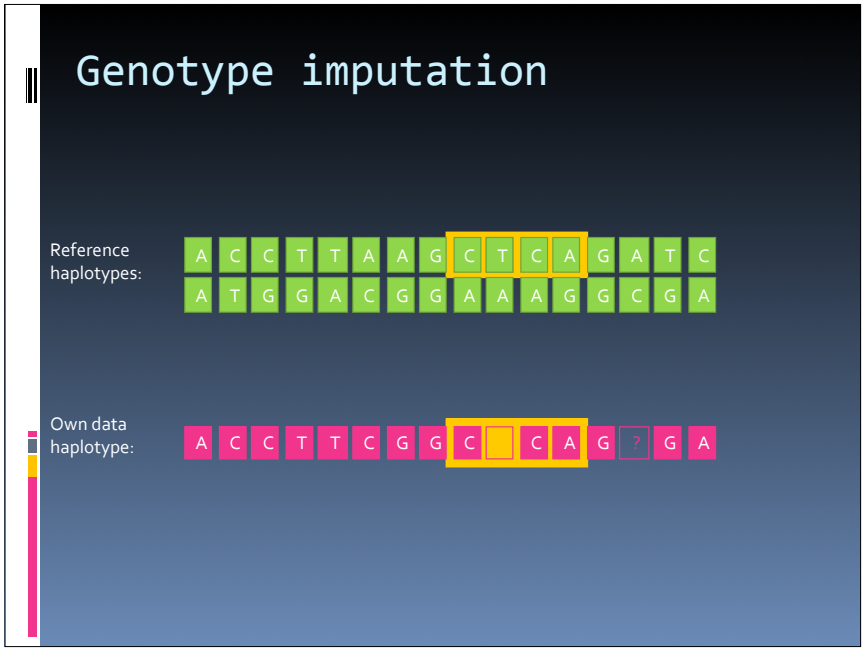
Genotype imputation



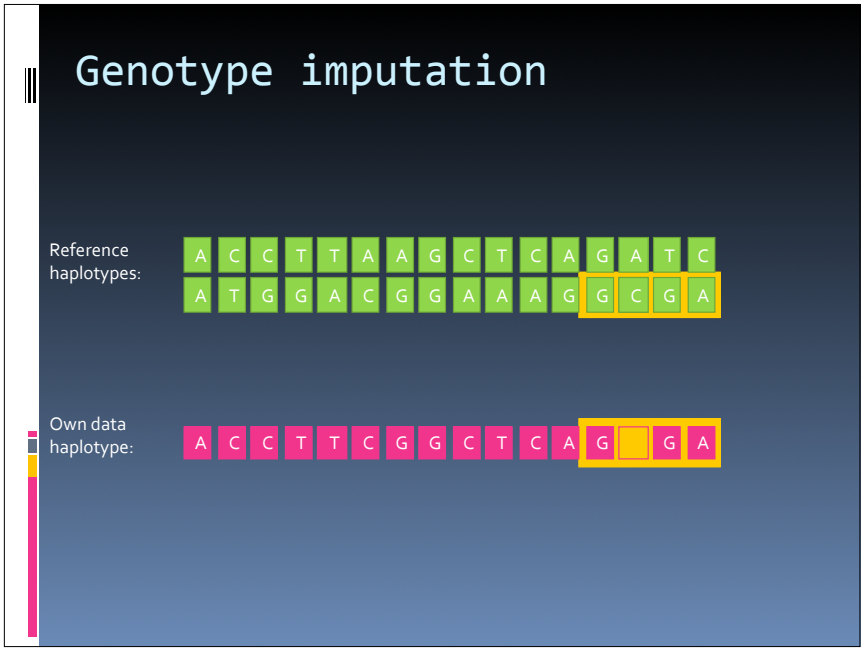
Genotype imputation

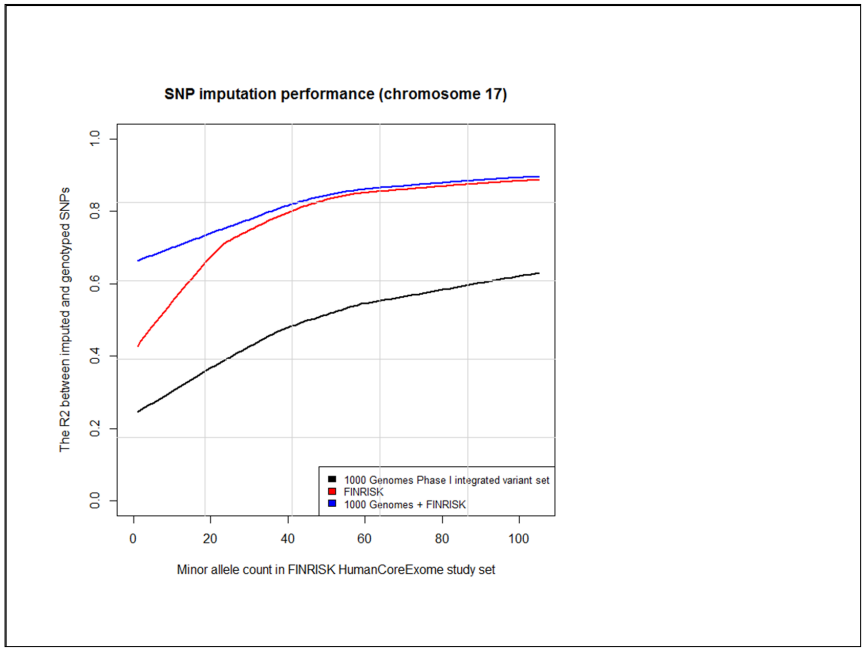


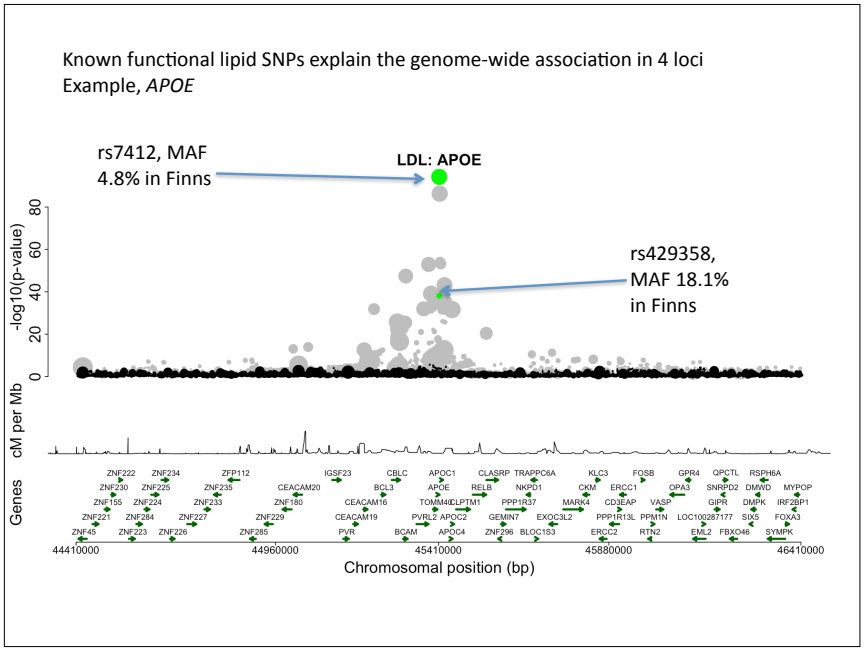
Genotype imputation



Genotype imputation

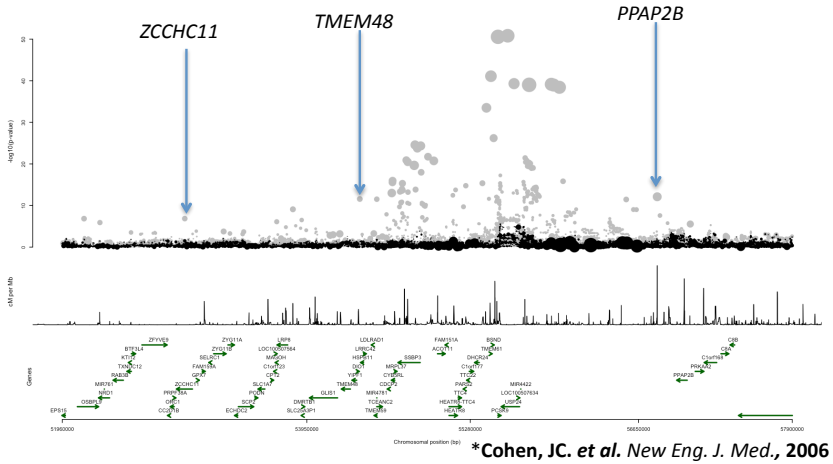



SNP imputation performance (chromosome 17)
The R2 between imputed and genotyped SNPs
0.0
0.2
0.4
0.6
0.8
1.0
0
20
40
60
80
100
Minor allele count in FINRISK HumanCoreExome study set
1000 Genomes Phase I integrated variant set
FINRISK
1000 Genomes + FINRISK


Known functional lipid SNPs explain the genome-wide association in 4 loci
Example, *APOE*
rs7412, MAF
4.8% in Finns
LDL: APOE
rs429358,
MAF 18.1%
in Finns
-log10(p-value)
0
20
40
60
80
cM per Mb
Genes
44410000
44960000
45410000
45880000
46410000
Chromosomal position (bp)

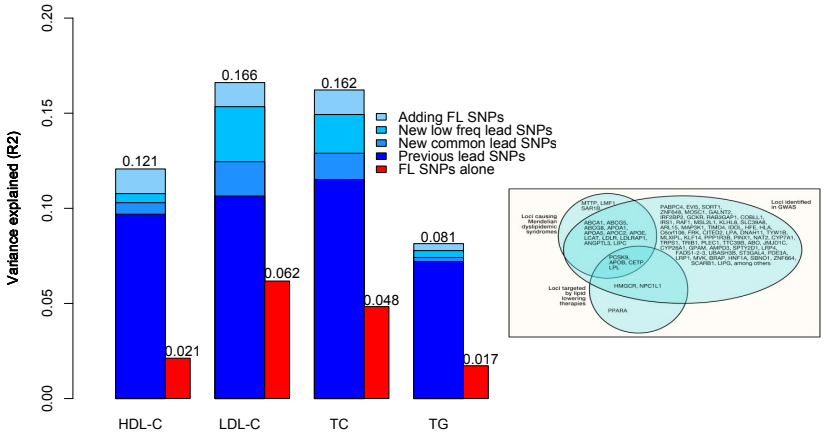
PCSK9 locus for LDL, association explained in conditional analysis.
Area has 4 significant loci at least 1Mb away from each other and in no LD (R^2 < 5%)

PCSK9: rs11591147
(known functional variant*)



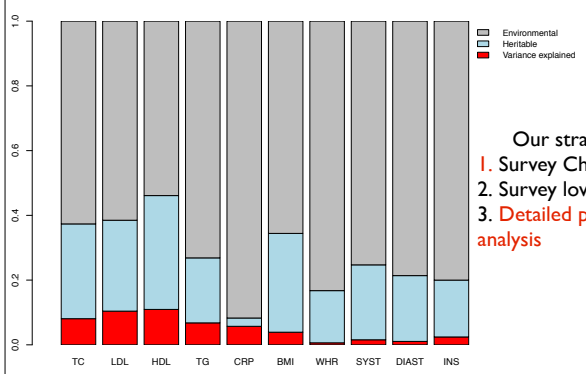
*Cohen, JC. *et al.* New Eng. J. Med., 2006

Further variance explained



Heritability and explained variance

Estimated in NFBC66 cohort



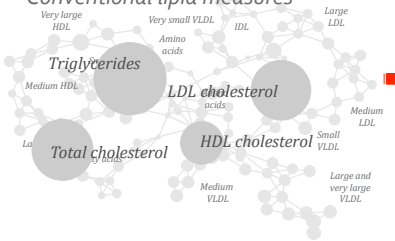
Our strategies to further discovery:

1. Survey ChrX
2. Survey low frequency variants
3. Detailed phenotyping / multivariate analysis

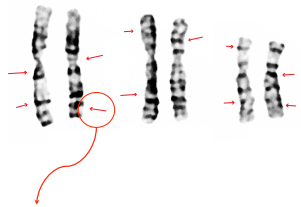
Would comprehensive metabolic and genetic characterization provide further insight?

Detailed metabolomics profile

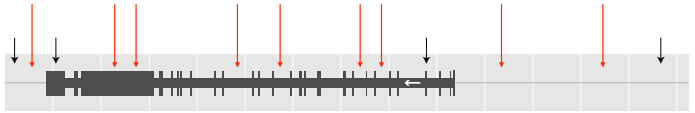
Conventional lipid measures



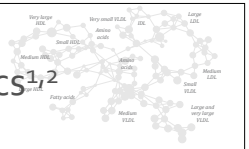
95 lipid loci



Dense coverage of variation in the lipid loci



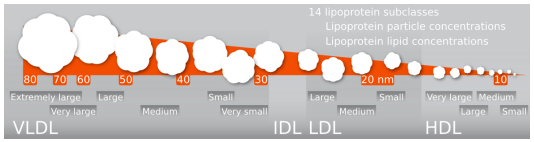
300 µl fasting serum, 15 minutes



Serum NMR metabolomics[1,2]

Lipoprotein subclasses

74 lipoprotein subclass measures



117 metabolites

Amino acids and small molecules

Valine, isoleucine, tyrosine etc.
Glucose, lactate, pyruvate etc.

Lipid measures

Fatty acids, sphingomyelin, phosphocholine etc.

Selected ratios of metabolites and derived measures

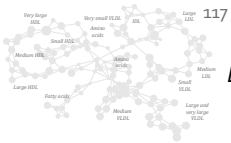
E.g. Alanine:glutamine, average HDL particle size

[1]*Soininen et al. Analyst. 2009; 134(9): 1781-5.*
[2]*Inouye et al. Mol Syst Biol. 2010; 6: 441.*

Materials and methods

N = 8330
NFBC 1966
Young Finns
HBCS
DILGOM
Health 2000

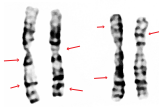
Serum metabolomics profile
117 + 99 metabolite measures



Enzymatic lipid measures
TC, TG, LDL-C and HDL-C

Dense map of variants
440 000 SNPs in the lipid loci

95 genetic loci



➔ *Association analyses*

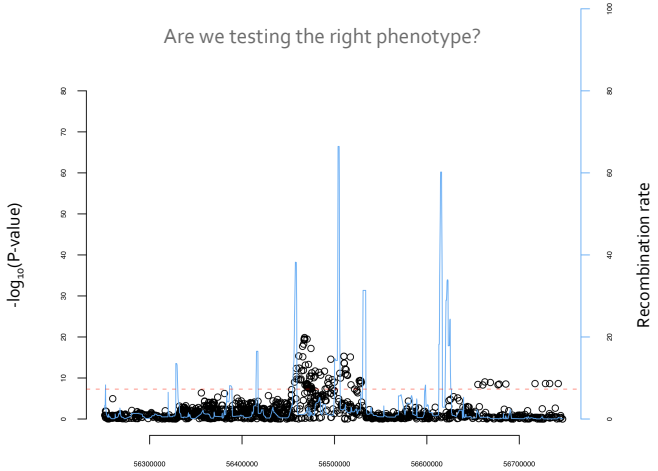
Tukiainen et al Hum Mol Genet 2012

Explained variance

HDL-C: 0.78%

HDL-C

Are we testing the right phenotype?



***LIPC* region in chr 15**
Genomic position (b36)

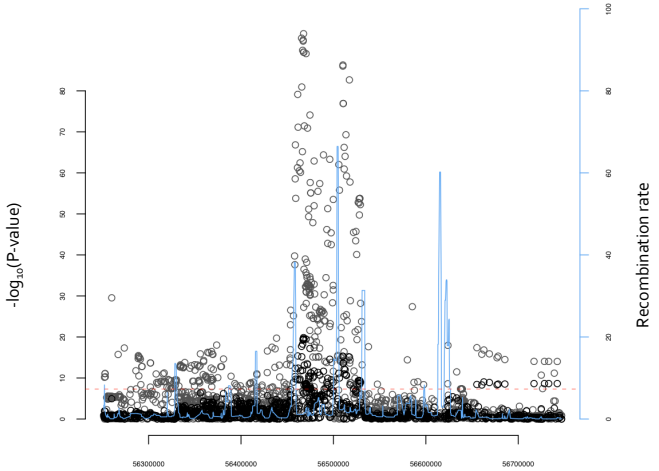
Tukiainen et al Human Mol Genet 2012

Explained variance

HDL-C: 0.78%

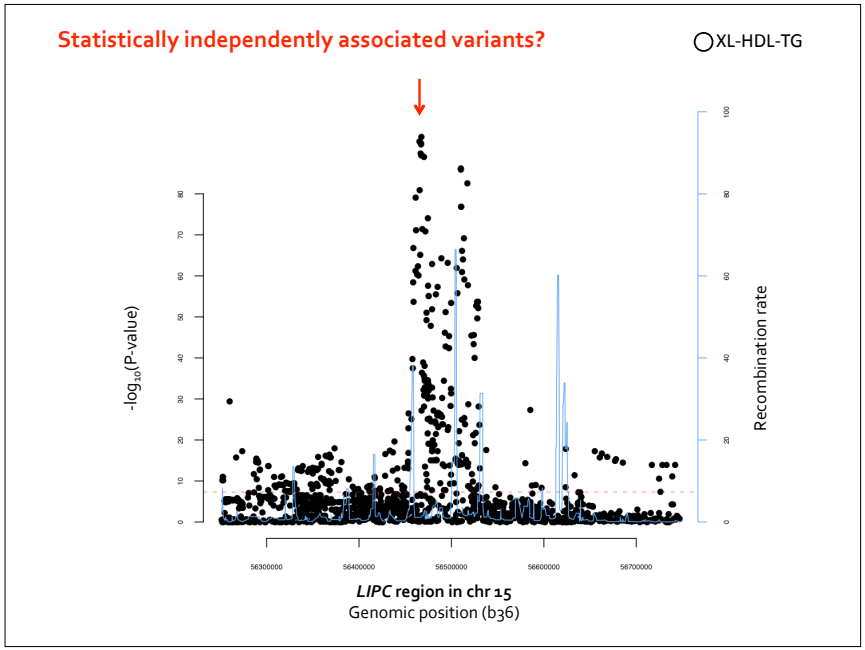
XL-HDL-TG: 3.97%

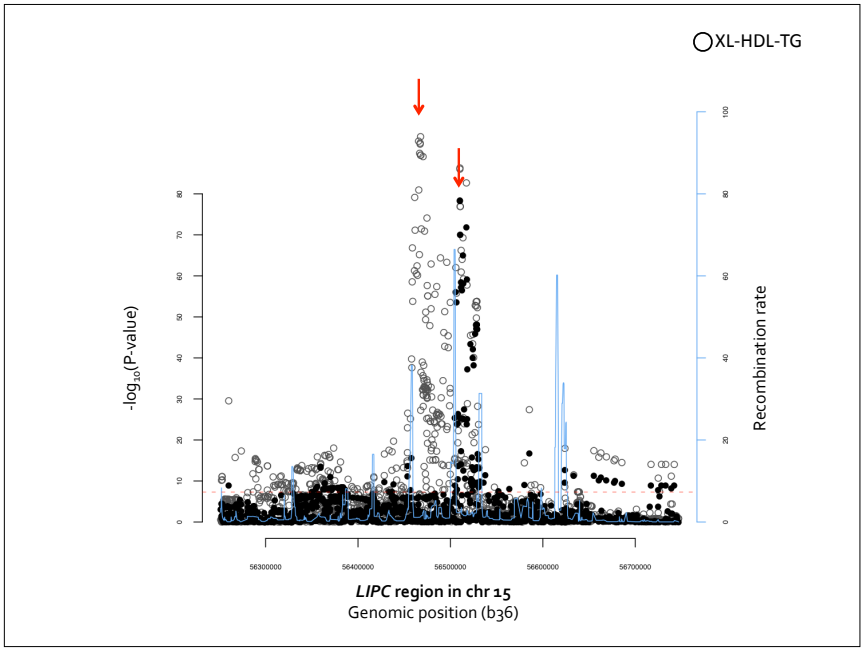
HDL-C
XL-HDL-TG

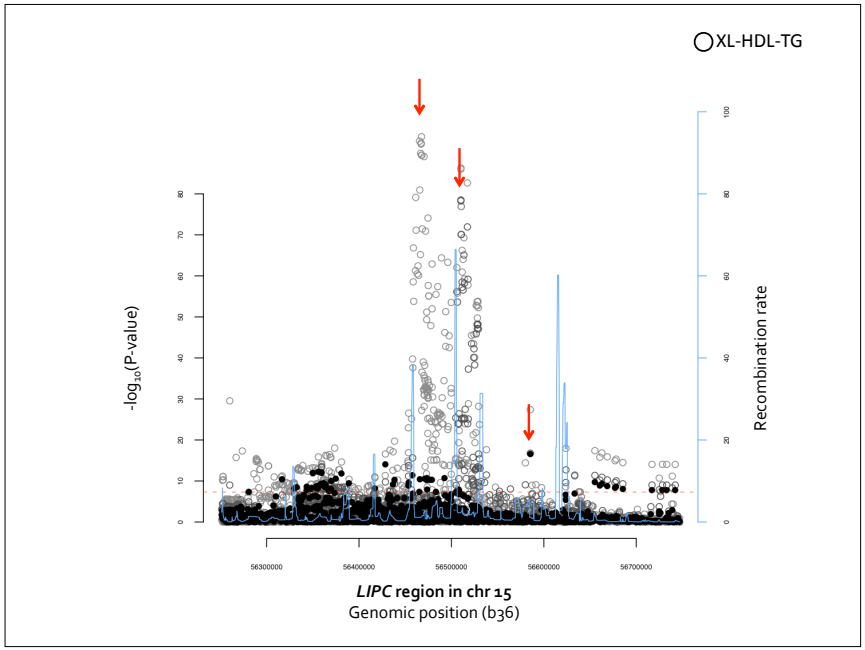


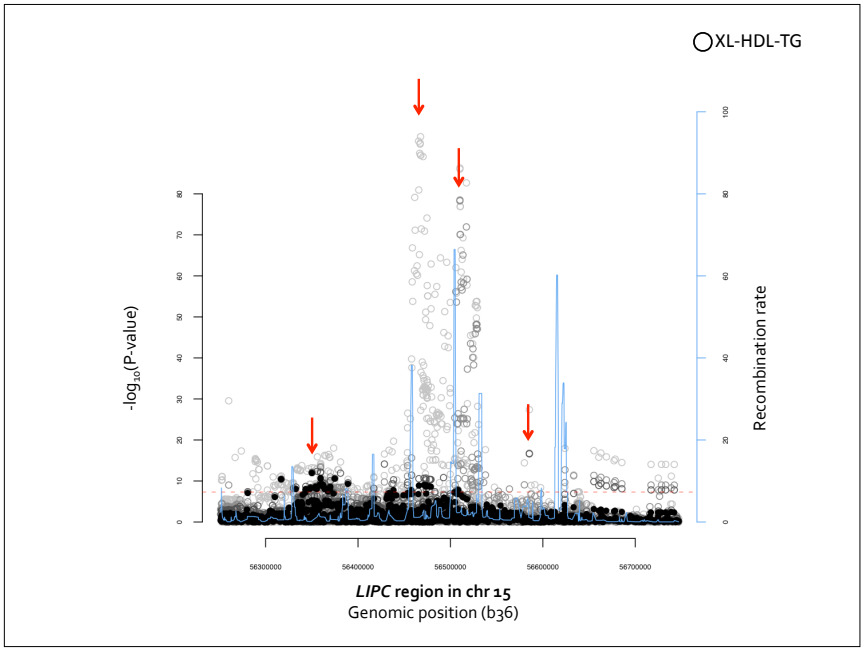
***LIPC* region in chr 15**
Genomic position (b36)

Statistically independently associated variants?



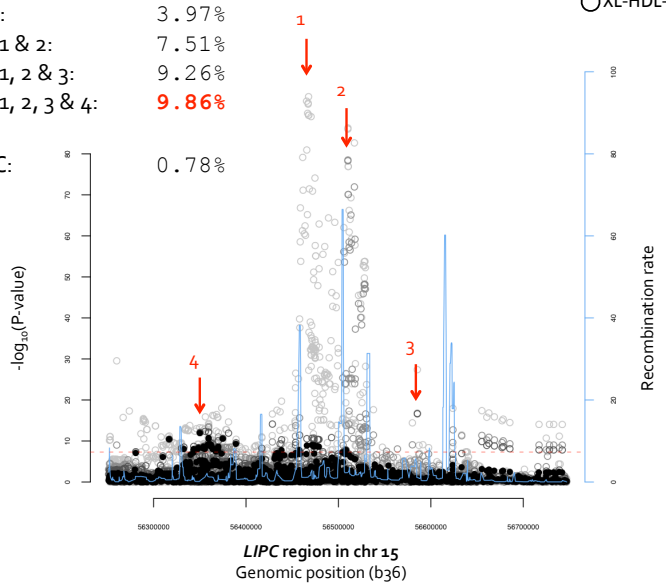



XL-HDL-TG
-log10(P-value)
Recombination rate
56300000
56400000
56500000
56600000
56700000
LIPC region in chr 15
Genomic position (b36)


XL-HDL-TG
-log10(P-value)
Recombination rate
56300000
56400000
56500000
56600000
56700000
LIPC region in chr 15
Genomic position (b36)

Explained variance

SNP 1:	3.97%
SNPs 1 & 2:	7.51%
SNPs 1, 2 & 3:	9.26%
SNPs 1, 2, 3 & 4:	**9.86%**
HDL-C:	0.78%



Johannes Kettunen

Genome-wide association study

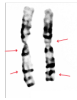
31 associated loci

11 new loci

3 new loci for lipoproteins

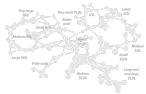
$P < 2.31 \times 10^{-10}$

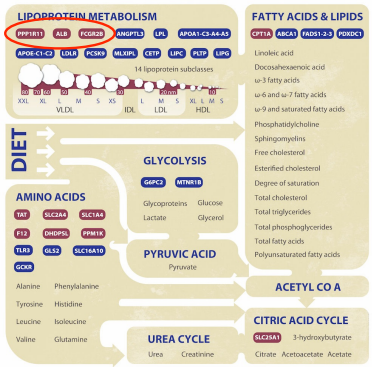
N = 8330



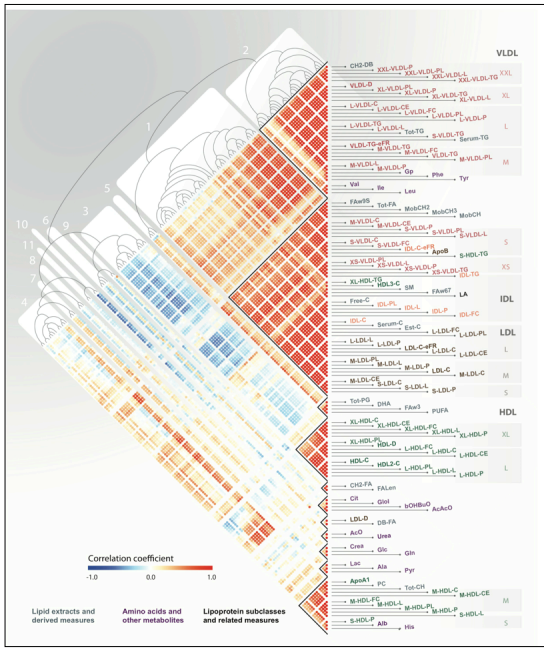
7.7 million variants across the genome

216 metabolite measures





Kettunen J et al. Nat Genet. 2012;44(3):269-76.



Association testing:
Canonical correlation analysis between metabolite clusters and genotype groups

Test statistic Wilk's lambda

Two non-overlapping cohorts

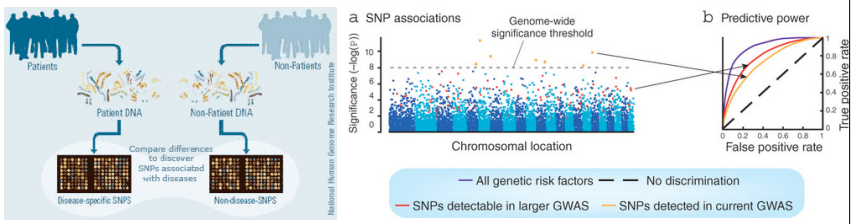
34 associated loci across the 11 metabolite clusters, 9 missed with univariate analysis in Kettunen et al Nat Genet 2012

Challenge in the interpretation of results

Inouye et al 2012, PLoS Genetics

Conclusions: Discovery studies

- There seems to be a lot to harvest from the low frequency variants in the coding regions of the identified GWAS loci, at least for lipids
- On top of detailed marker set, detailed phenotyping may help in fine mapping
- Statistical methods for testing multivariate phenotypes need development
- Large-scale studies are needed to identify low frequency variants, with access to the individual level data



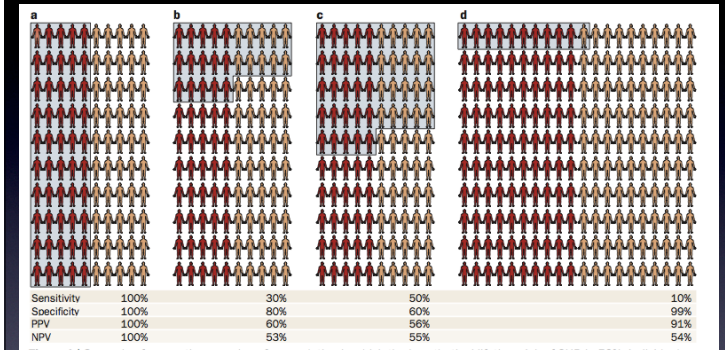
Can common SNPs help in predicting coronary heart disease events?

Outline

- Predicting CHD events with traditional risk factors
- Tens of CHD associated common genetic variants identified using Genome-wide association studies (GWAS)
- Association vs. prediction: results with 13 SNP score, 28 SNP score and 100+ SNP score
- Potential in screening for high risk individuals

Emmi Tikkanen
Thesis defense Nov 8th @ noon
Kytösuontie 9, lh 1

Can we use the CHD loci for prediction?

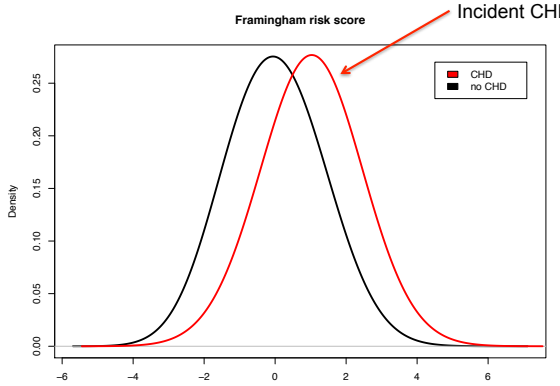


	a	b	c	d
Sensitivity	100%	30%	50%	10%
Specificity	100%	80%	60%	99%
PPV	100%	60%	56%	91%
NPV	100%	53%	55%	54%

Figure 1 | Scenarios for genetic screening of a population in which the hypothetical lifetime risk of CHD is 50%. Individuals who will develop disease are shown in red and those who will not develop disease are shown in yellow. **a** | The ideal test; all people who develop disease are identified and no false positives exist. **b** | A test in which only a small fraction of those who develop disease are detected and a substantial number of those labeled 'test positive' remain disease-free; use of a systolic blood-pressure threshold of 160 mmHg would perform similarly to this hypothetical test. **c** | A test in which half of those who will develop disease and almost half of those who will not develop disease are labeled 'test positive'; a test that uses a common variant of small effect would perform similarly to this hypothetical test and would provide limited discrimination for CHD, as carriage is neither sufficient nor necessary to develop CHD. **d** | A test that screens for a rare allele of large effect detects few cases of CHD in the population as a whole, but might have a role in family-based screening. Abbreviations: CHD, coronary heart disease; NPV, negative predictive value; PPV, positive predictive value.

Holmes et al, Nature Reviews Cardiology, 2011

Risk factor distributions overlapping



Framingham risk score at baseline:
age, sex, total cholesterol, HDL, BMI, systolic blood pressure, blood pressure treatment, current smoking status, diabetes mellitus, family history of CHD

Risk loci identified in GWA studies

- GWA studies have identified >30 loci modifying CHD risk
- > 200 loci associated with major risk factors lipids, blood pressure, BMI, diabetes, smoking
- Causal variants and functional genes often unknown
  - Potential to discover new causal pathways for CHD
    - Particularly as CHD poorly modeled in mice and other model organisms
- But, maybe they could also be used for prediction?

Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls *CDKN2A–CDKN2B–ANRI*

The Wellcome Trust Case Control Consortium*

CELSR2–PSRC1–SORT1

MIA3
CXCL12
PHACTR1
LDLR
WDR12
PCSK9
HNF1A
SH2B3
MRAS
LPA

Newly identified loci that influence lipid concentrations and risk of coronary artery disease

Cristen J Willer[1,18], Serena Sanna[1,2,18], Anne U Jackson[1], Angelo Scuteri[3,4], Lori L Bonnycastle[5], Robert Clarke[6], Simon C Heath[7], Nicholas J Timpson[8], Samer S Najjar[3], Heather M Stringham[1], James Strait[3], William L Duren[1], Andrea Maschio[2], Fabio Busonero[2], Antonella Mulas[2], Giuseppe Albai[2], Amy J Swift[5], Mario A Morken[5], Pierre Meneton[11], Jouko Sundvall[12], Yoav Ben-Shlomo[8], Manuela Uda[2], Jaal Michael Boehnke[1],

Genome-wide association of early-onset myocardial infarction with single nucleotide polymorphisms and copy number variants

Myocardial Infarction Genetics Consortium*

Genomewide Association Analysis of Coronary Artery Disease

SLC5A3–MRPS6–KCNE2

Nilesh J. Samani, F.Med.Sci., Jeanette Erdmann, Ph.D., Alistair S. Hall, F.R.C.P., Christian Hengstenberg, M.D., Massimo Mangino, Ph.D., Bjoern Mayer, M.D., Richard J. Dixon, Ph.D., Thomas Meitinger, M.D., Peter Braund, M.Sc., H.-Erich Wichmann, M.D., Jennifer H. Barrett, Ph.D., David-Alexandre Tregouet, Ph.D., Mark M. Iles, Francois Cambien, M.D., Marcus Fische, Anthony J. Balmforth, Ph.D., Andrea Ingrid Brænne, M.Sc., Christian Gieger, Ph.D., John R. Thompson, Ph.D., and Heribert

Genome-wide haplotype association study identifies the *SLC22A3-LPAL2-LPA* gene cluster as a risk locus for coronary artery disease

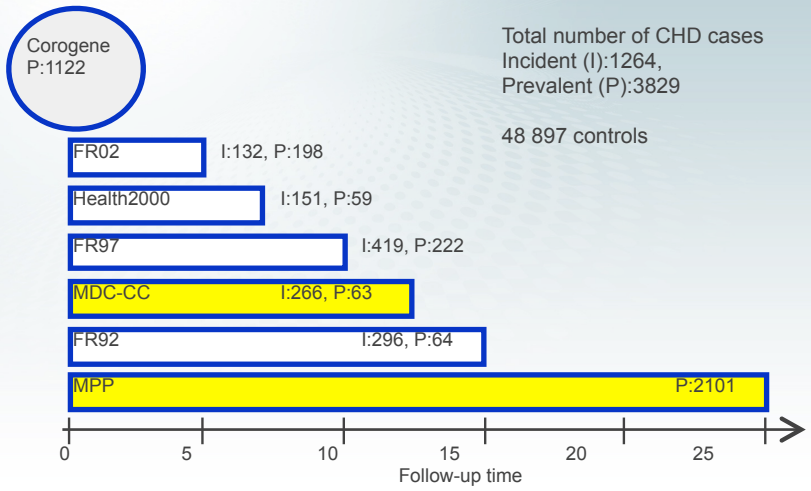
David-Alexandre Trégouët[1], Inke R König[2], Jeanette Erdmann[3], Alexandru Munteanu[1], Peter S Braund[4], Alistair S Hall[5], Anika Großhennig[2,3], Patrick Linsel-Nitschke[3], Claire Perret[1],

Wolfgang Lieb, M.D., ikenberg, M.D., trom, M.D., ndreas Ziegler, Ph.D., ics Consortium*

New susceptibility locus for coronary artery disease on chromosome 3q22.3

Jeanette Erdmann[1], Anika Großhennig[1,2], Peter S Braund[3], Inke R König[2], Christian Hengstenberg[4], Alistair S Hall[5], Patrick Linsel-Nitschke[1], Sekar Kathiresan[6], Ben Wright[7], David-Alexandre Trégouët[8], Francois Cambien[8], Petra Bruse[1],

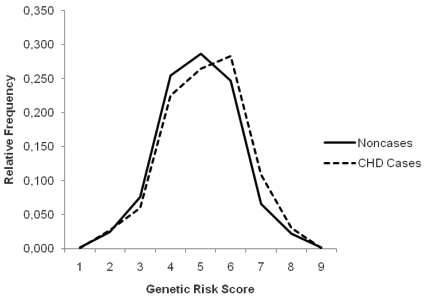
Our study design



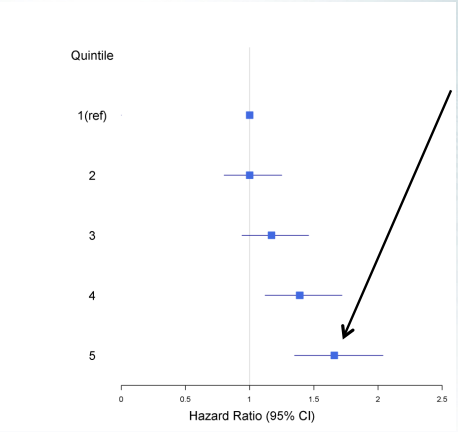
Genetic risk score base on 13 SNPs

SNP	GENE	WEIGHT
rs17465637	MIA3	1.14
rs11206510	PCSK9	1.15
rs646776	CELSR2, PSRCC1, SORT1	1.19
rs6725887	WDR12	1.17
rs9818870	MRAS	1.15
rs3798220	LPA	1.33
rs9349379	PHACTR1	1.12
rs4977574	CDKN2A/2B	1.29
rs1746048	CXCL12	1.14
rs2259816	HNF1a	1.08
rs3184504	SH2B3	1.13
rs1122608	LDLR	1.15
rs9982601	SLC5A3, MRPS6, KCNE2	1.20

Weighted sum



Association of GRS with incident CHD risk



Adjusting for age and sex:
Hazard ratio = 1.72 [1.41-2.10],
$P = 7.3 \times 10^{-11}$

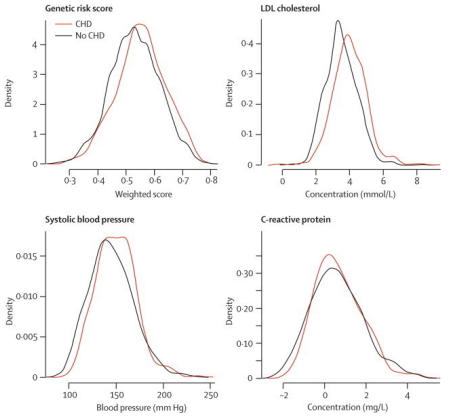
Adjusting for age, sex, and Framingham factors:
Hazard ratio = 1.66 [1.35-2.14]
$P = 7.3 \times 10^{-10}$

Lancet 2010; 376: 1393-400

A multilocus genetic risk score for coronary heart disease: case-control and prospective cohort analyses

Samuli Ripatti, Emmi Tikkanen, Marju Orho-Melander, Aki S Havulinna, Kaisa Silander, Amitabh Sharma, Candace Guiducci, Markus Perola, Antti Jula, Juha Sinisalo, Marja-Liisa Lokki, Markku S Nieminen, Olle Melander, Veikko Salomaa, Leena Peltonen, Sekar Kathiresan*

GRS distribution compared to risk factor distributions



Distributions at baseline of genetic risk score, LDL cholesterol, systolic blood pressure, and log-transformed C-reactive protein by 10-year incident coronary heart disease event status in FINRISK 1992 and 1997 cohorts

Prediction over FRS:
C index: p=0.19
NRI: p=0.18

Conclusions:

Association: strong
Prediction: no(t yet)

Ripatti et al, Lancet 2010

Reclassification: 13 SNP score

Reclassification of individuals in the FINRISK 1992 and 1997 cohorts on the basis of 10-year predicted risk of coronary heart disease with and without genetic risk score

	Genetic risk score 0–5%	Genetic risk score 5–10%	Genetic risk score 10–20%	Genetic risk score >20%
Cases, by predicted risk*				
0–5%	82 (93%)	6 (7%)	0	0
5–10%	10 (8%)	100 (84%)	9 (8%)	0
10–20%	0	9 (5%)	136 (81%)	22 (13%)
>20%	0	0	9 (7%)	129 (93%)
Non-cases, by predicted risk*				
0–5%	9224 (99%)	121 (1%)	0	0
5–10%	179 (12%)	1149 (79%)	131 (9%)	0
10–20%	0	120 (14%)	687 (80%)	49 (6%)
>20%	0	0	54 (13%)	351 (87%)

*10-year predicted risk on the basis of traditional risk factors only.

Cases:
13% reclassified from intermediate to high risk

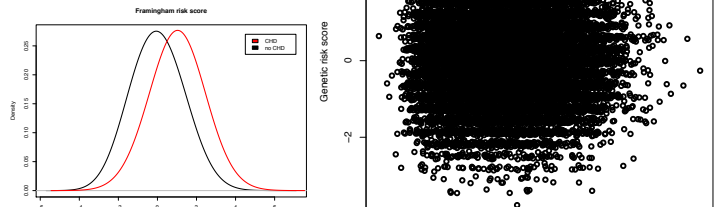
Non-cases:
13% reclassified from high to intermediate risk

Ripatti et al, Lancet 2010

Including further risk factor associated SNPs to GRS

- Women's Genome Health Study (Paynter et al, JAMA 2010):
  - 777 incident CVD events, 101 SNP score
  - – No association, no predictive value over FRS
- CAREMA Study (Vaarhorst et al, Circ Cardiovasc Genet 2012):
  - 742 incident CHD cases, 179 SNP score
  - – Association yes, no predictive value over FRS
- Framingham Heart Study (Thanassoulis et al, Circ Cardiovasc Genet 2012):
  - 537 incident CVD events (182 CHD events), 102 SNP score
  - – Association yes, no predictive value over FRS

Not completely surprising given that
1)The risk factors explain only a fraction of the risk variation
2)The identified risk-factor SNPs together explain only a small proportion of the risk factor variance

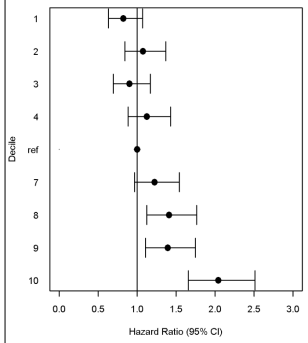


Yet, clear potential for better genetic prediction exists:
1) Framingham risk factors capture only a fraction of the variation in risk
2) FRS and GRS are almost orthogonal
3) More risk variants are being identified through larger GWAS meta-analyses

From 13 to 28 SNP risk score

- 28 CHD SNPs genotyped in FINRISK 92, 97, 02 and Health 2000 cohorts
  - 24,124 individuals
  - 1093 CHD (5%), 1552 CVD (6%) and 731 ACS (3%) cases
  - median follow-up time of 12 years (IQR 8.75–15.25 years)

Association: GRS and risk for incident coronary heart disease events



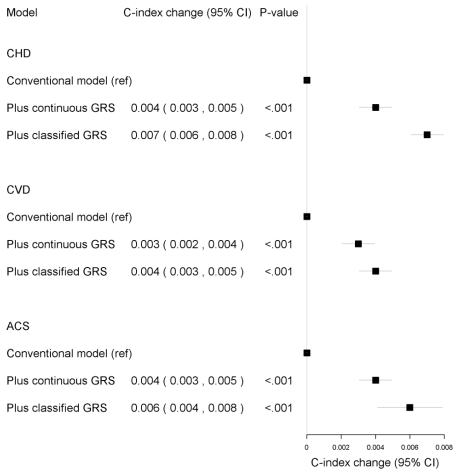
Similar shape of GRS risk distribution for CHD, CVD and ACS

Table 2. Risk for cardiovascular endpoints by genetic risk score

			Top versus middle 20%				
Trait	**HR (95 % CI)***	**P-value**	**Top 20% of GRS (95 % CI)**	**Top 10% of GRS (95 % CI)**	**Top 5% of GRS (95 % CI)**	**N events**	**N**
CHD	1.27 (1.20 , 1.35)	<.001	1.71 (1.42 , 2.06)	2.07 (1.68 , 2.56)	2.12 (1.62 , 2.77)	1093	24124
ACS	1.27 (1.18 , 1.37)	<.001	1.57 (1.25 , 1.97)	1.84 (1.42 , 2.40)	2.00 (1.43 , 2.79)	731	24124
CVD	1.18 (1.12 , 1.24)	<.001	1.59 (1.36 , 1.86)	1.87 (1.56 , 2.24)	1.72 (1.35 , 2.18)	1552	24124

* per SD of GRS. Cox regression models were adjusted for sex, total cholesterol, high-density lipiprotein (HDL) cholesterol, body-mass index, systolic blood pressure, antihypertensive treatment, smoking and type 2 diabetes; age was used as the timescale.
Abbreviations: CHD, coronary heart disease; ACS, acute coronary syndrome; CVD, cardiovascular disease; GRS, genetic risk score; HR, Hazard ratio; CI, confidence interval; N, number of individuals

Prediction: Changes in C-index when adding GRS to FRS prediction



Reclassification: 28 SNP score

eTable 3. Reclassification of individuals in four risk categories after addition of genetic risk score (GRS) to a model with traditional risk factors*

Model without GRS		Model with GRS					
		0-5%	5-10%	10-20%	>20%	NRI	Clinical NRI
0-5%	Events	82	17	0	0	**Events**: 0.04 (P=.03)	**Events**: 0.11 (P=.002)
	Nonevents	7749	205	0	0	**Nonevents**: 0.01 (P=.002)	**Nonevents**: 0.10 (P<.001)
	All	7831	222	0	0	**All**: 0.05 (P=.008)	**All**: 0.22 (P<.001)
5-10%	Events	11	102	25	0		
	Nonevents	277	1109	174	0		
	All	288	1211	199	0		
10-20%	Events	0	15	147	38		
	Nonevents	0	206	781	96		
	All	0	221	928	134		
>20%	Events	0	0	29	177		
	Nonevents	0	0	93	338		
	All	0	0	122	515		

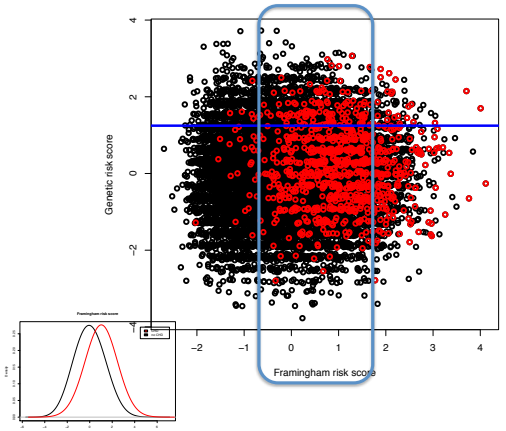
*Traditional risk factors include sex, total cholesterol, high-density lipiprotein (HDL) cholesterol, body-mass index, systolic blood pressure, blood pressure treatment, smoking and type 2 diabetes; age was used as the timescale in the Cox proportional hazards model.

Abbreviations: NRI, net reclassification improvement

19%

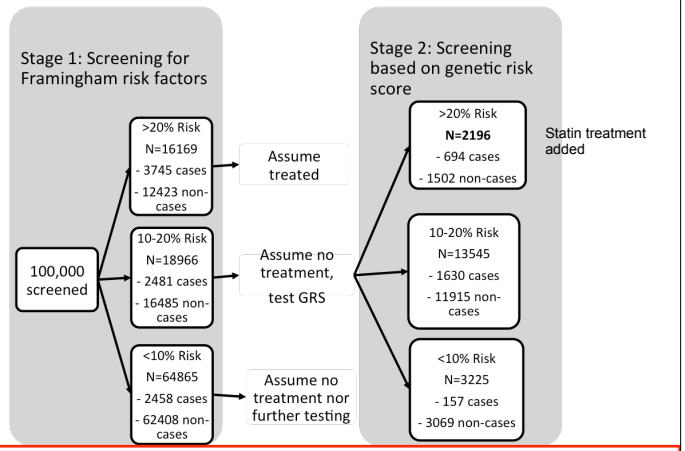
22%

Framingham scores and genetic risk scores for individuals



Incident cases
Non-cases

Would it be beneficial to screen individuals with intermediate risk using GRS?



Assuming:
-20% risk reduction with statin treatment
-Treatment compliance and efficacy similar in Stage 1 and Stage 2 high risk groups

- 694*0.2 = 139 cases prevented in 14 years / 100,000 individuals (119 in 10 years)
- Compared to 17 for lp(a) (Di Angelantonio, JAMA 2012) and 30 for CRP (Kaptoge, NEJM 2012)

Conclusions

- Genetic screens continue to provide new risk variants for CHD, and refine the associations in known risk loci
- These loci will incrementally jointly bring better predictive power over the traditional risk factors
- GRS based on currently known loci offers a small but significant improvement in both CHD risk discrimination and reclassification over and above the traditional risk factors
- GRS screening of individuals in intermediate risk could help to prevent future cases through more accurate statin allocation
- The assumptions and the clinical utility needs to be tested

Acknowledgement: Prediction Study

FIMM
Institute for Molecular Medicine Finland
Nordic EMBL Partnership for Molecular Medicine

Emmi Tikkanen
Kaisa Silander
Aarno Palotie

NATIONAL INSTITUTE FOR HEALTH AND WELFARE

Aki S. Havulinna
Markus Perola
Antti Jula
Veikko Salomaa

LUND UNIVERSITY

Marju Orho-Melander
Amitabh Sharma
Olle Melander

HUCH
Juha Sinisalo
Markku S Nieminen

Haartman Institute
Marja-Liisa Lokki

BROAD INSTITUTE

Candace Guiducci
Elena Gonzalez
Sekar Kathiresan

Acknowledgements:

+ Matti Pirinen: Statistical methods development

+ Antti Sarin: Imputation, sequence QC