Statistical Genetics and Coronary Artery Disease

Robert Davies

October 8th, 2010

CANNeCTIN

Outline

- Part 0 - Rationale
- Part 1 - Genetics
- Part 2 - Ethnicity
- Part 3 - Prediction

- Part 0 – Rationale
- Part 1 - Genetics
- Part 2 - Ethnicity
- Part 3 - Prediction

Rationale

- This group
  - CAnadiaN Network and Centre for Trials Internationally (CANNeCTIN)
- This seminar series
  - Biostatistics Methodology Videoconference
  - Advanced Issues in Clinical Trials Methodology
- This presentation
  - Statistical genetics?

Why?

ReAssessment of Anti-Platelet Therapy Using an InDividualized Strategy Based on GENetic Evaluation (RAPID GENE)

This study is currently recruiting participants.
Verified by University of Ottawa Heart Institute, August 2010

First Received: August 17, 2010 Last Updated: August 23, 2010 History of Changes

Sponsor:	University of **Ottawa** Heart Institute
Collaborator:	Spartan Bioscience Inc.
Information provided by:	University of Ottawa Heart Institute
ClinicalTrials.gov Identifier:	NCT01184300

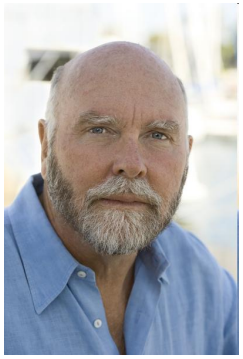
http://clinicaltrials.gov/ct2/show/NCT01184300?term=ottawa+jason+roberts&rank=1

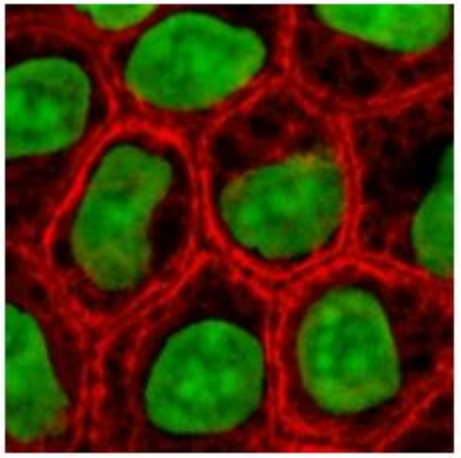
Improved Prediction of Cardiovascular Disease Based on a Panel of Single-Nucleotide Polymorphisms Identified Through Genome-Wide Association Studies

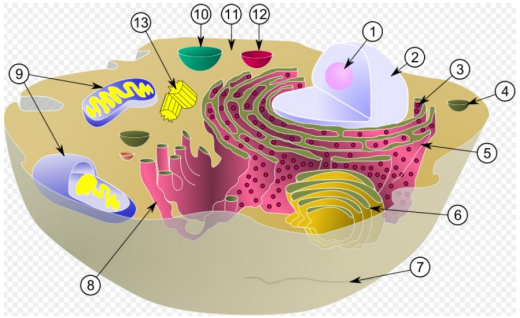
Robert W. Davies, MSc; Sonny Dandona, MD; Alexandre F.R. Stewart, PhD; Li Chen, MSc; Stephan G. Ellis, MD; W.H. Wilson Tang, MD; Stanley L. Hazen, MD, PhD; Robert Roberts, MD; Ruth McPherson, MD, PhD; George A. Wells, MSc, PhD

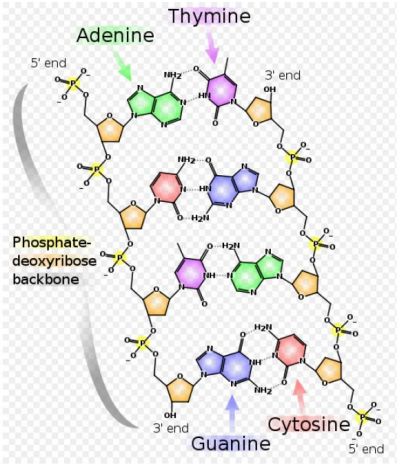
- Part 0 - Rationale
- Part 1 – Genetics
- Part 2 - Ethnicity
- Part 3 - Prediction

People have cells
Cells contain DNA









Refs:

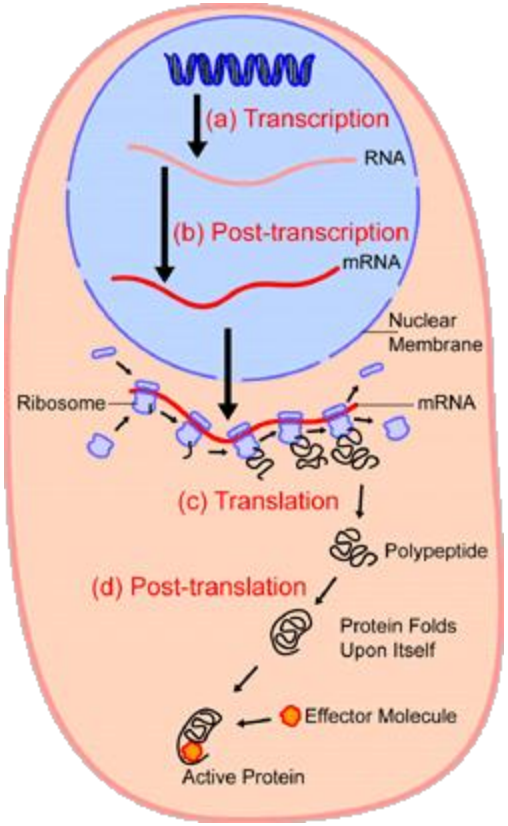
http://www.plosbiology.org/article/slideshow.action?uri=info:doi/10.1371/journal.pbio.0050266

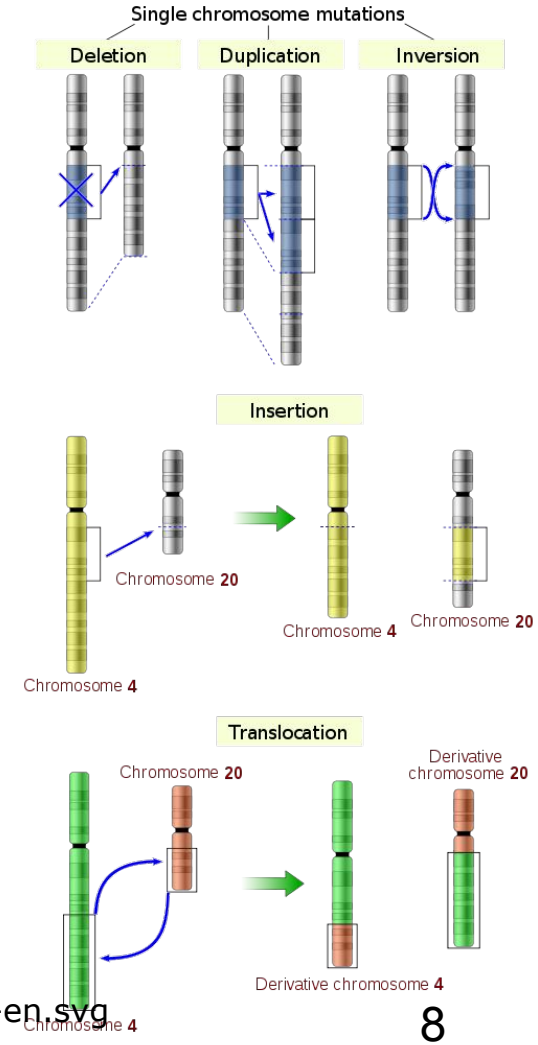
http://en.wikipedia.org/wiki/File:Biological_cell.svg

http://en.wikipedia.org/wiki/File:DNA_chemical_structure.svg

http://en.wikipedia.org/wiki/File:Epithelial-cells.jpg

DNA is important
DNA is susceptible to mutations





And...

http://en.wikipedia.org/wiki/File:Chromosomes_mutations-en.svg
http://en.wikipedia.org/wiki/File:Proteinsynthesis.png

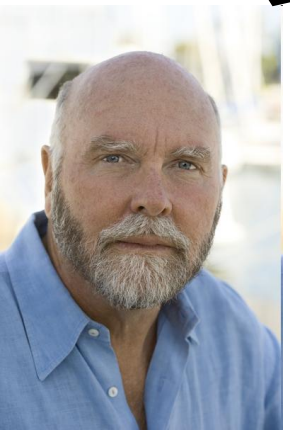
The Single Nucleotide Polymorphism (SNP). Example rs1333049

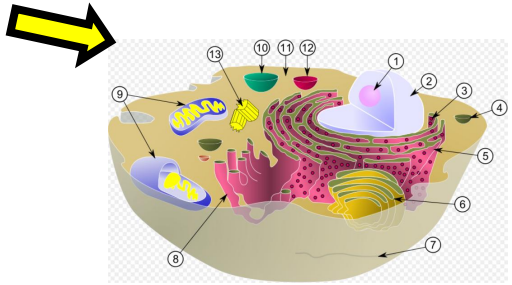
>hg19_gwasCatalog_rs1333049
range=chr9:22125453-22125553
5'pad=50 3'pad=50 strand=+
repeatMasking=none
TGGTCACTACCCTACTGTCATTCCTCA
TACTAACCATATGATCAACAGTT**[G/C**
]AAAAGCAGCCACTCGCAGAGGTAAG
CAAGATATATGGTAAATACTGTGT T

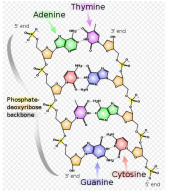
http://genome.ucsc.edu/cgi-bin/hgc?hgsid=170305062&g=htcGetDna2&table=gwasCatalog&i=rs1333049&o=22125502&l=22125502&r=22125503&getDnaPos=chr9%3A22%2C125%2C503-22%2C125%2C503&hgSeq.cdsExon=1&hgSeq.padding5=50&hgSeq.padding3=50&hgSeq.casing=upper&boolshad.hgSeq.maskRepeats=0&hgSeq.repMasking=lower&boolshad.hgSeq.revComp=0&submit=get+DNA

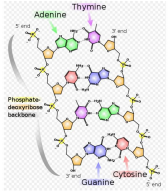
Population Diversity

	Sample Ascertainment				Genotype Detail NEW				Alleles	
ss#	Population	Individual Group	Chrom. Sample Cnt.	Source	C/C	C/G	G/G	HWP	C	G
ss66441130	HapMap-CEU	European	118	GF	0.220	0.542	0.237		0.492	0.508
ss97786460	J. Craig Venter		2	IG		1.000			0.500	0.500









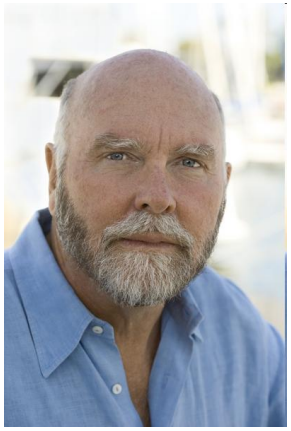
5′ ..AGTT**G**AAAA.. 3′
3′ ..TCAACTTTTT.. 5′

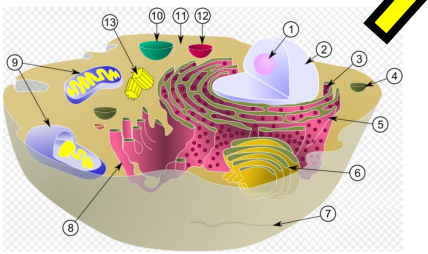
rs1333049
GC = 1

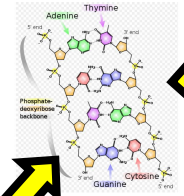
5′ ..AGTT**C**AAAA.. 3′
3′ ..TCAAGTTTT.. 5′

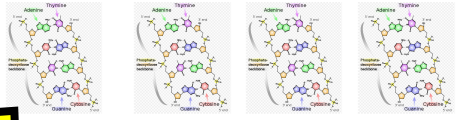
http://www.ncbi.nlm.nih.gov/SNP/snp_ref.cgi?type=rs&rs=rs1333049
http://www.plosbiology.org/article/slideshow.action?uri=info:doi/10.1371/journal.pbio.0050266

DNA microarrays

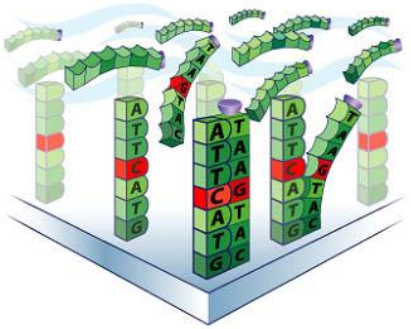


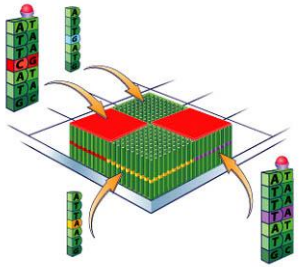


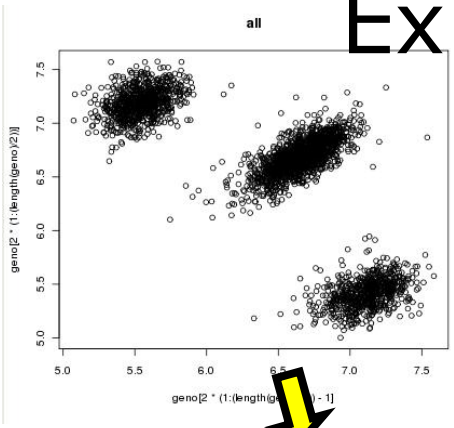




Ex: rs1333049







Person	A	B	CV	D	E	F
Rs1333049	0	1	1	2	2	1
Rs123456	0	0	0	1	1	2

http://en.wikipedia.org/wiki/File:Affymetrix-microarray.jpg

Lots and lots of SNPs

Vol 449|18 October 2007|doi:10.1038/nature06258 nature

ARTICLES

A second generation human haplotype map of over 3.1 million SNPs

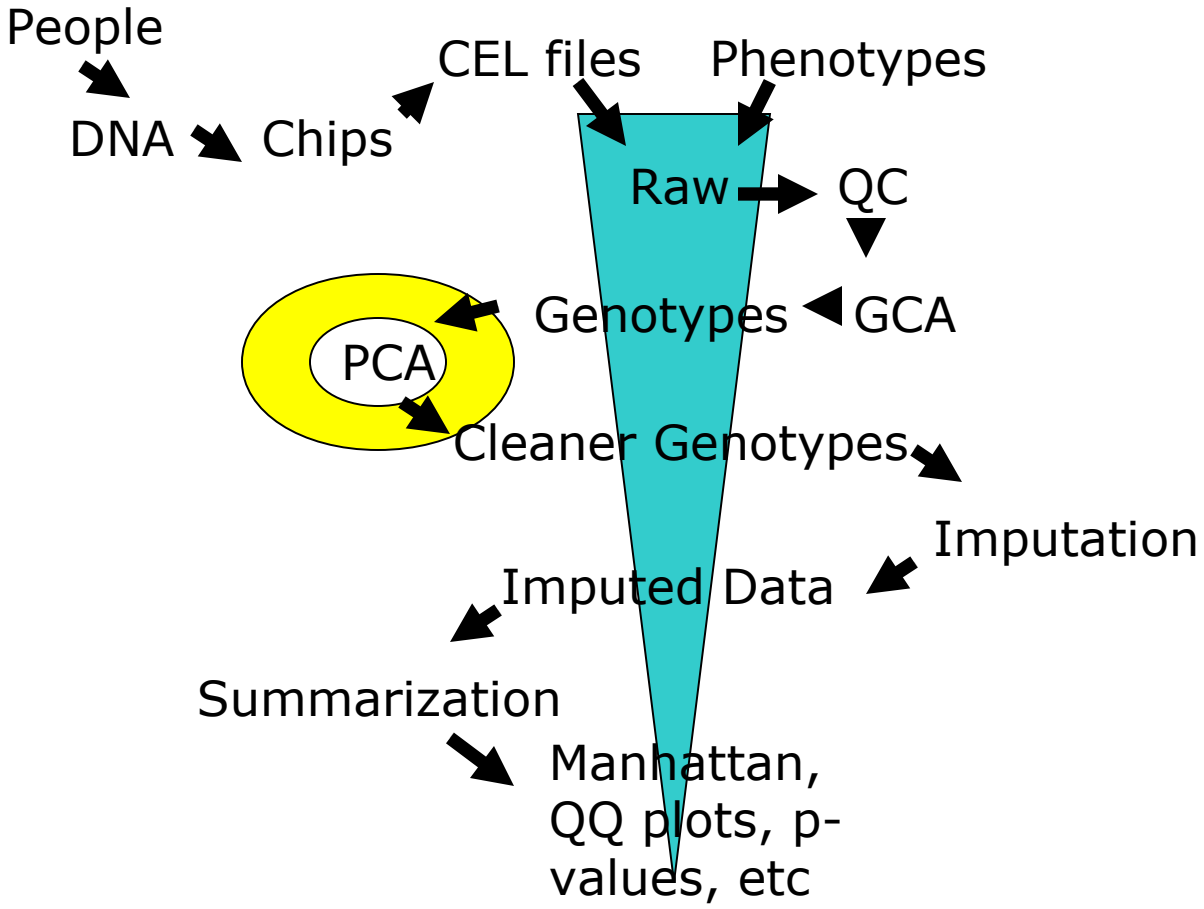
The International HapMap Consortium*

2010 – About 10M SNPs

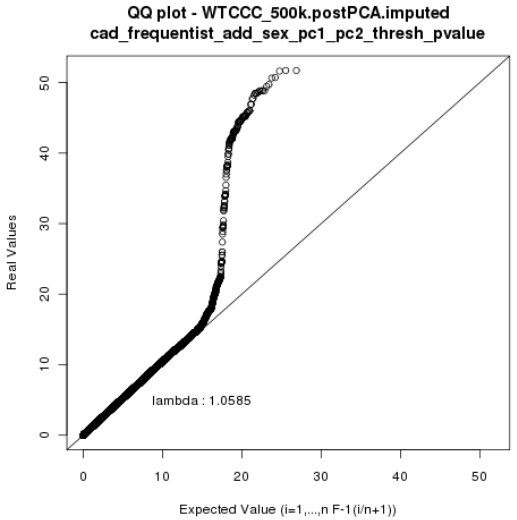
Genome Wide Association Study (GWAS)

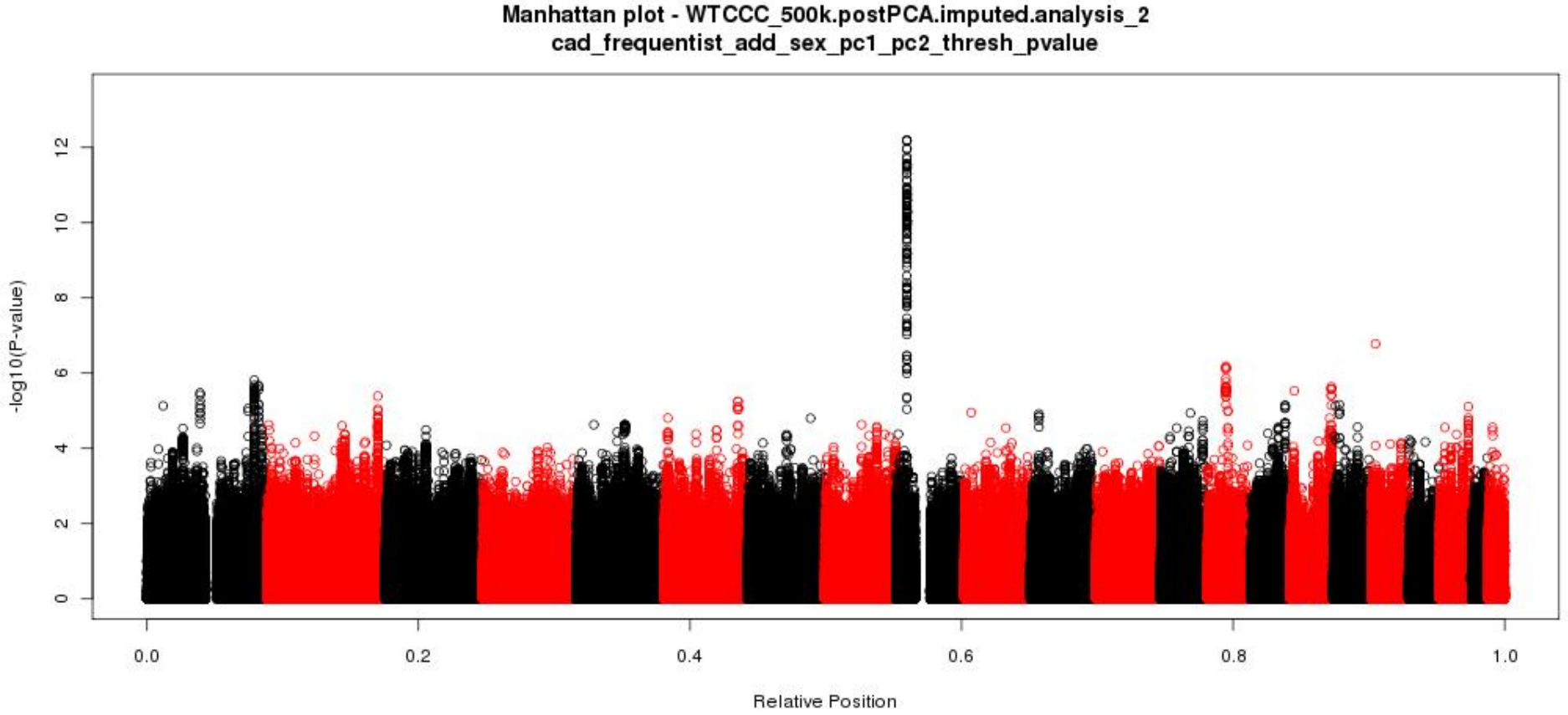
- Choose phenotype and collect subjects – ex: CAD
  - Either continuous, ie HDL, LDL, etc, or binary, ie CAD vs control
- Get DNA and run whole genome microarray
- Search for SNPs which are "statistically significantly" different with respect to phenotype

What's actually done in a GWAS in triangle form



Typical Results

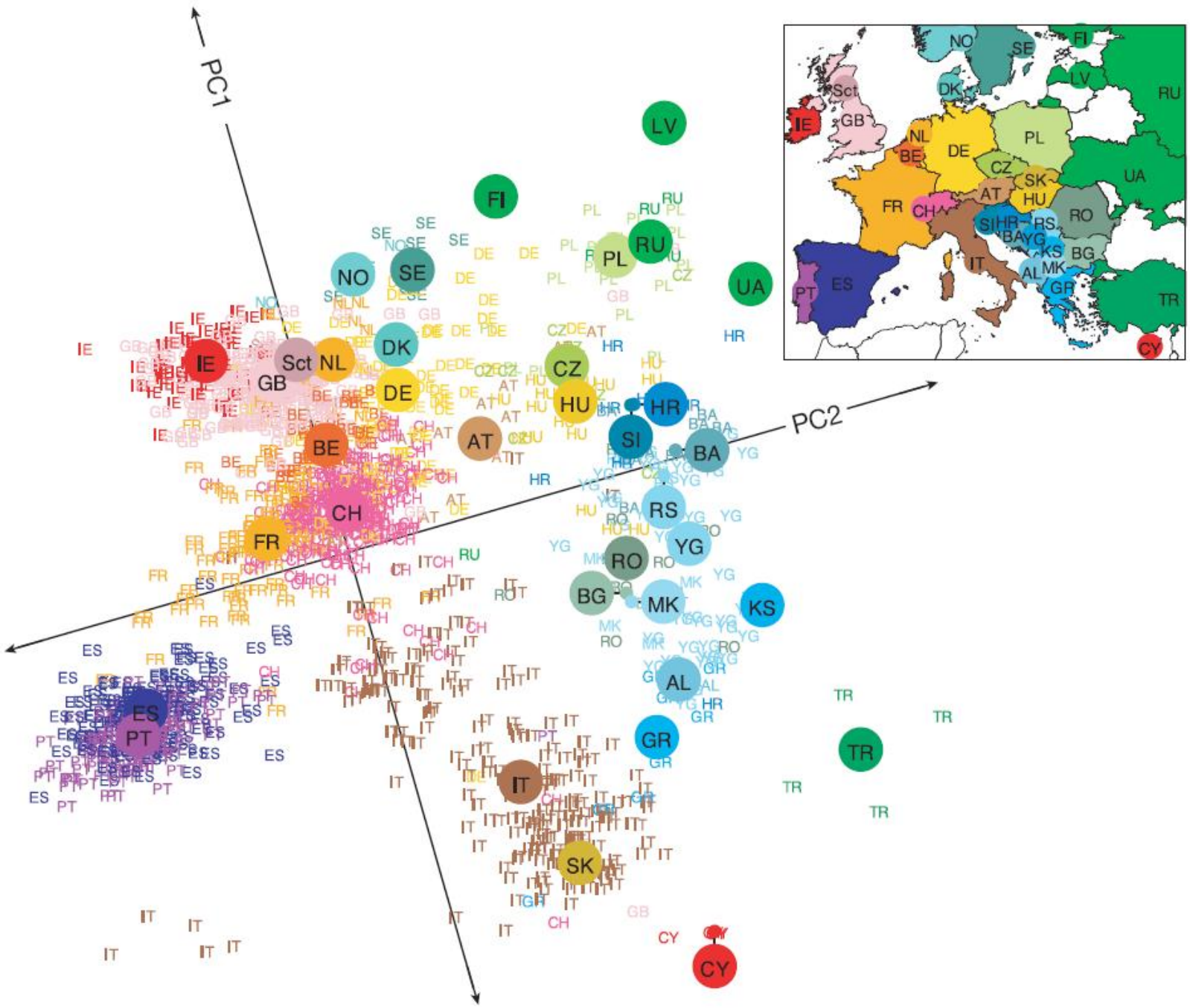




- Part 0 - Rationale
- Part 1 - Genetics
- Part 2 - Ethnicity
- Part 3 - Prediction

Ethnicity

- This section
  - Explain mathematical basis of PCA briefly
  - Show how we use it to identify genetic outliers
  - Explain ancestry informative SNPs
  - Show how to find them
  - Show what they look like in our data
  - Explain how they may be used



Novembre et al 2008, Nature, Genes mirror geography within Europe

Let G be an m by n matrix of genotypes, where m equals the number of SNPs and n the number of subjects. Therefore, $G_{i,j}$ is the genotype of the jth person for the ith SNP.
Consider a zero centered version of G, X, such that

$$X_{i,j} = G_{i,j} - \frac{1}{n}\sum_{j=1}^{n} G_{i,j} \tag{1}$$

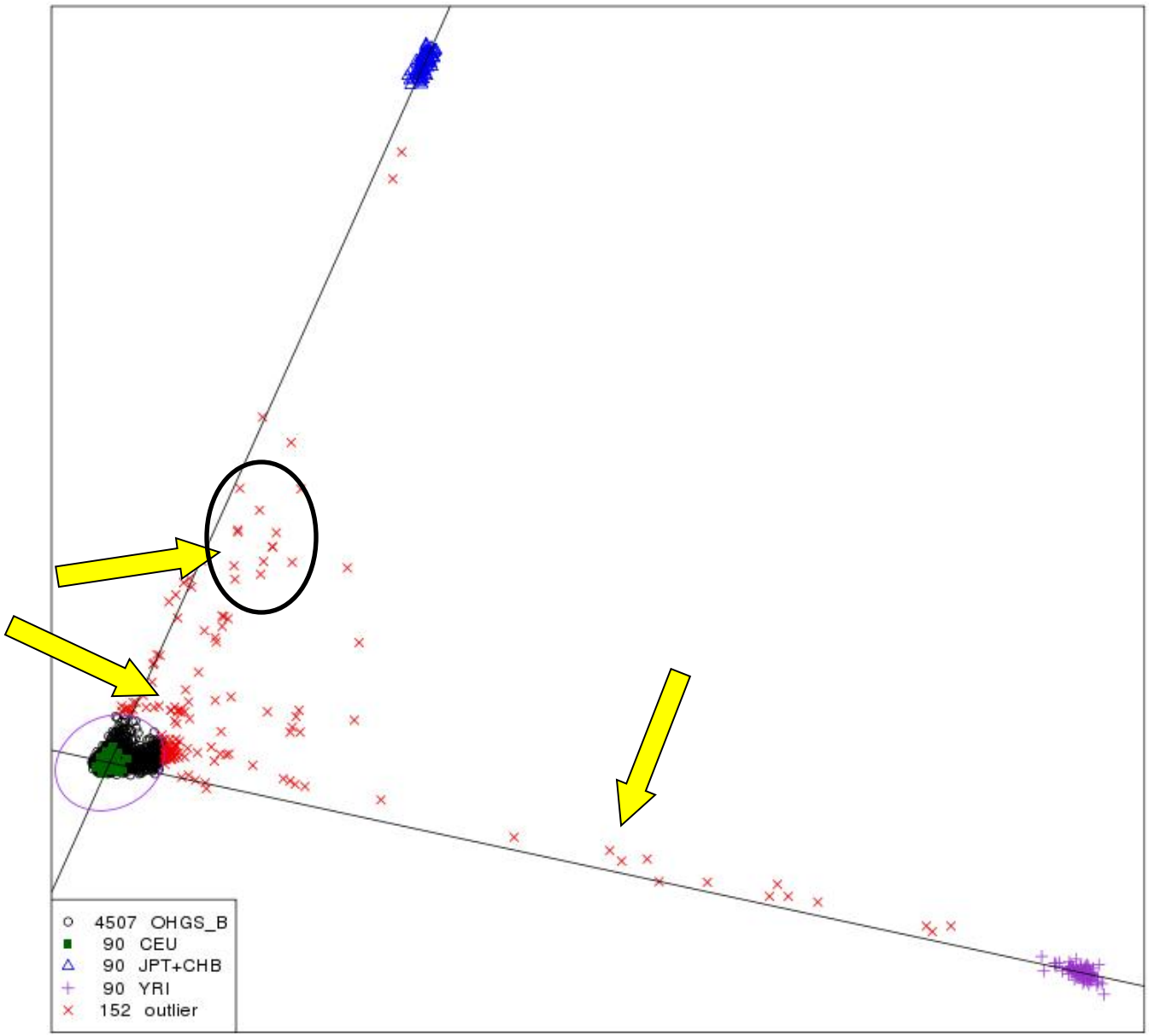
X may or may not be normalized with respect to the observed allele frequency of the SNP i. Principal components analysis allows for the definition of a transformation matrix P of dimension $m \times m$ such that

$$Y = PX \tag{2}$$

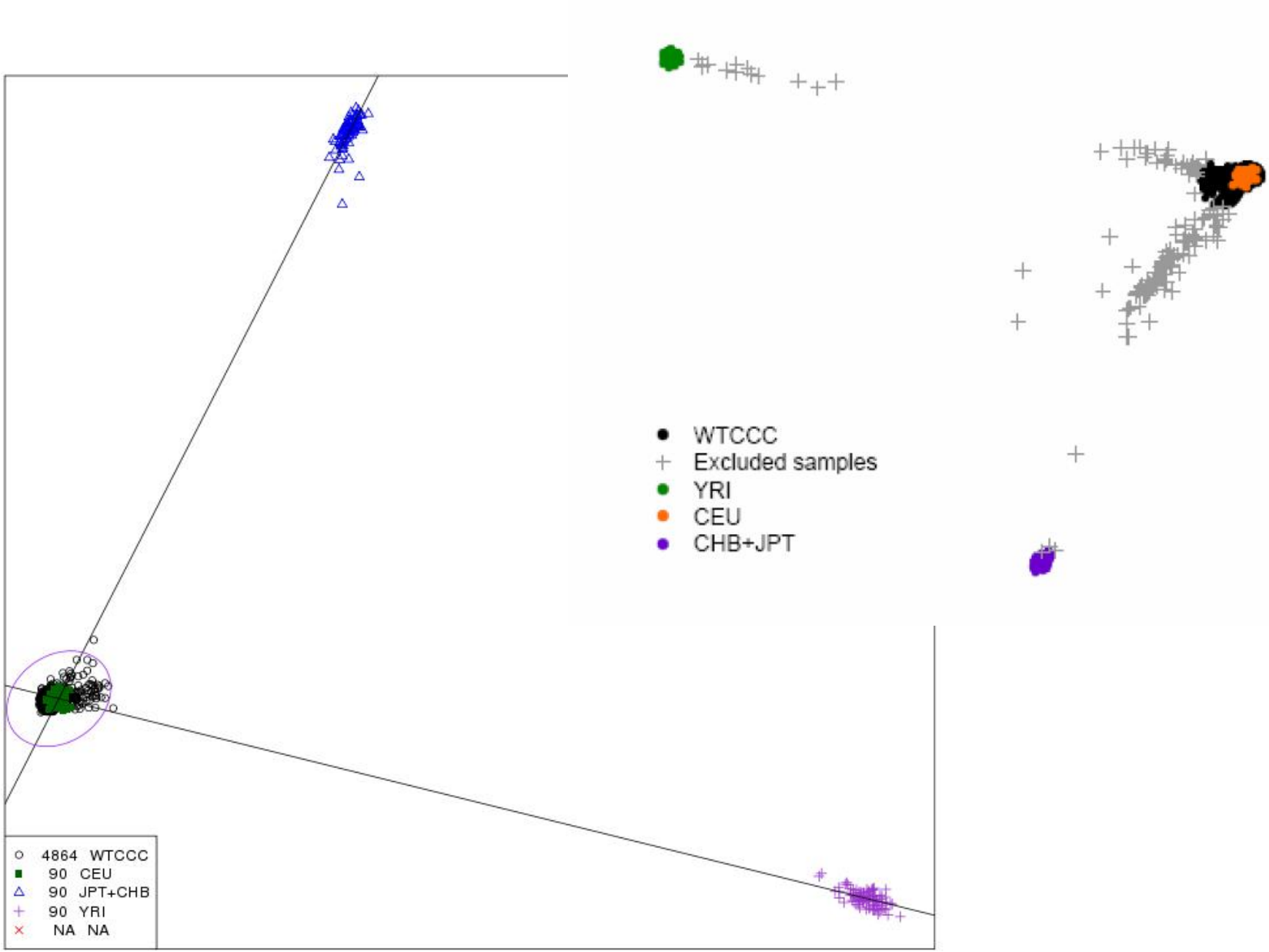
where the i^{th} row of Y represent the i^{th} principal component. One way to calculate the principal components is to first calculate the covariance matrix M between subjects

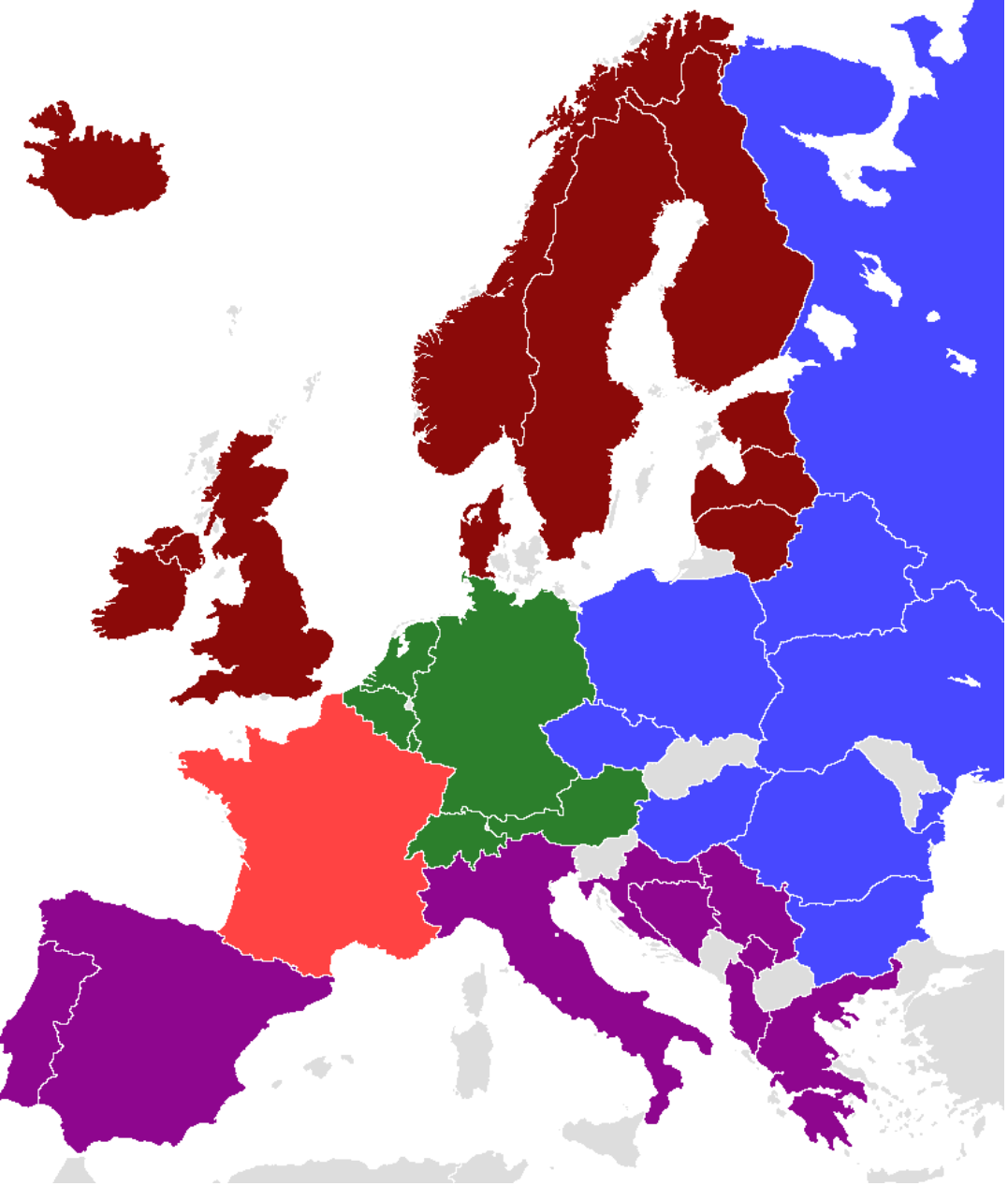
$$M = X^T X \tag{3}$$

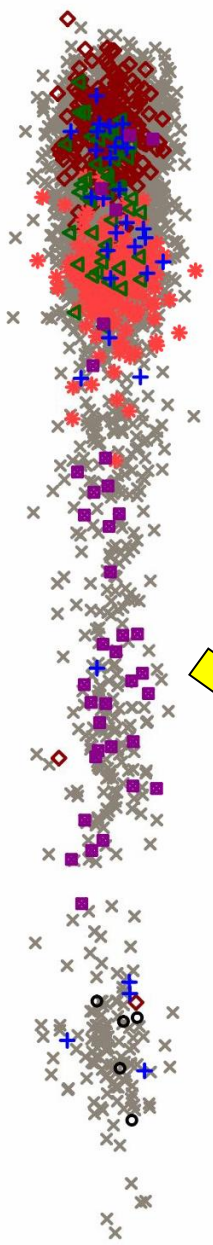
The i^{th} eigenvector of M is equivalent to the i^{th} row of P, which is an easy mathematical calculation. In practice, calculating the eigenvectors of M is dominated by calculating the matrix product of $X^T X$, and for $n << m$, this is $O(mn^2)$ (for m small, it is $O(n^3)$).

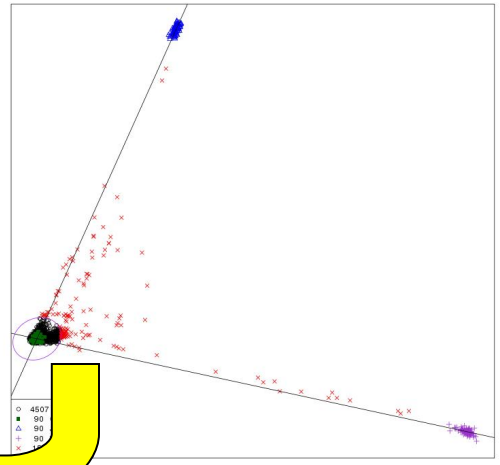

4507 OHGS_B
90 CEU
90 JPT+CHB
90 YRI
152 outlier

WTCCC 2007 Nature 7 June 2007 p661









$\lambda = 1.2$

$\lambda = 1.1$

http://commons.wikimedia.org/wiki/File:Blank_map_europe.png
Colours added post-acquisition

Ancestry Informative SNPs

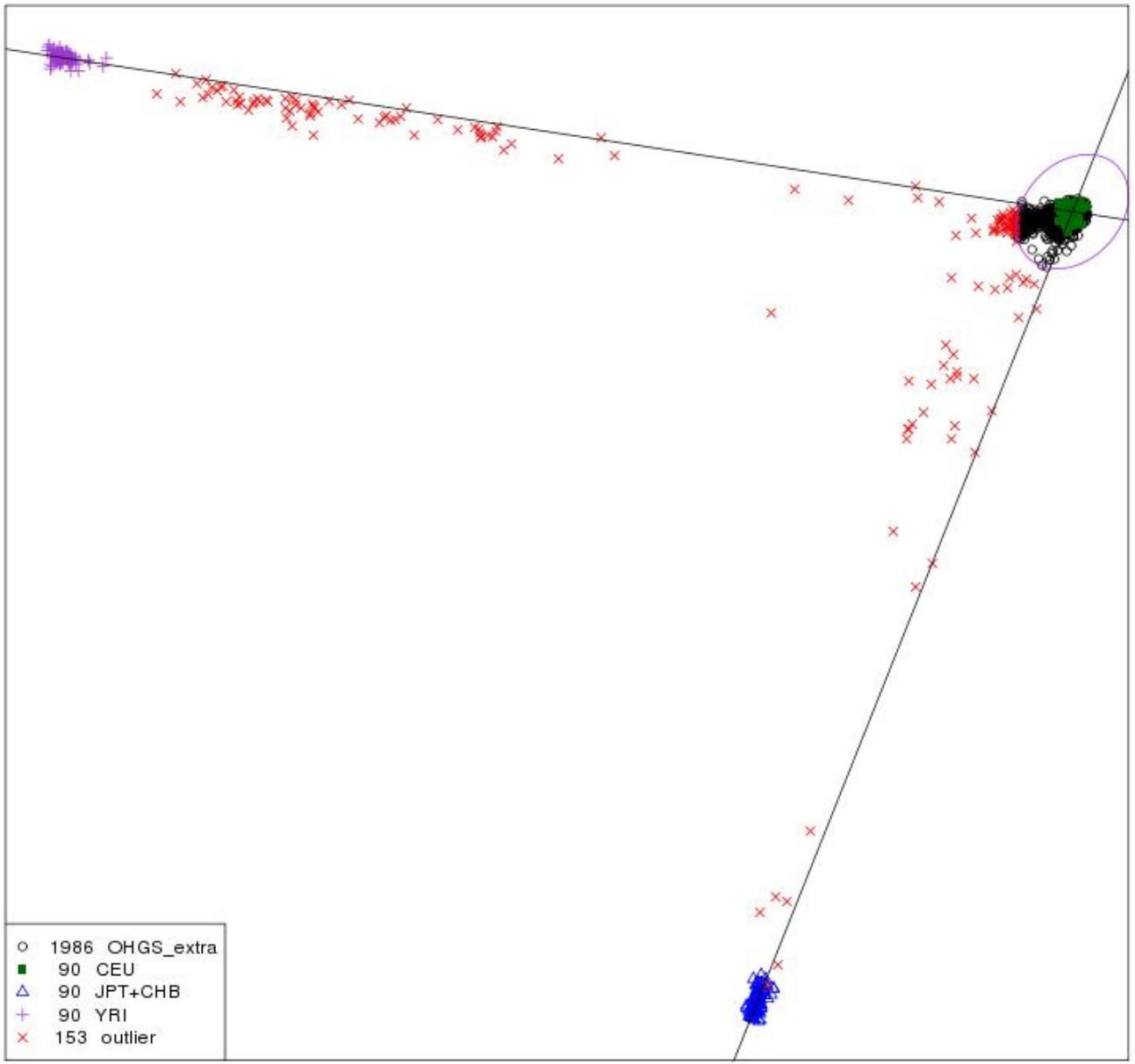
- A lot of SNPs are useless
  - Low frequency, same in different populations, etc.
- How many SNPs do we really need to identify populations?

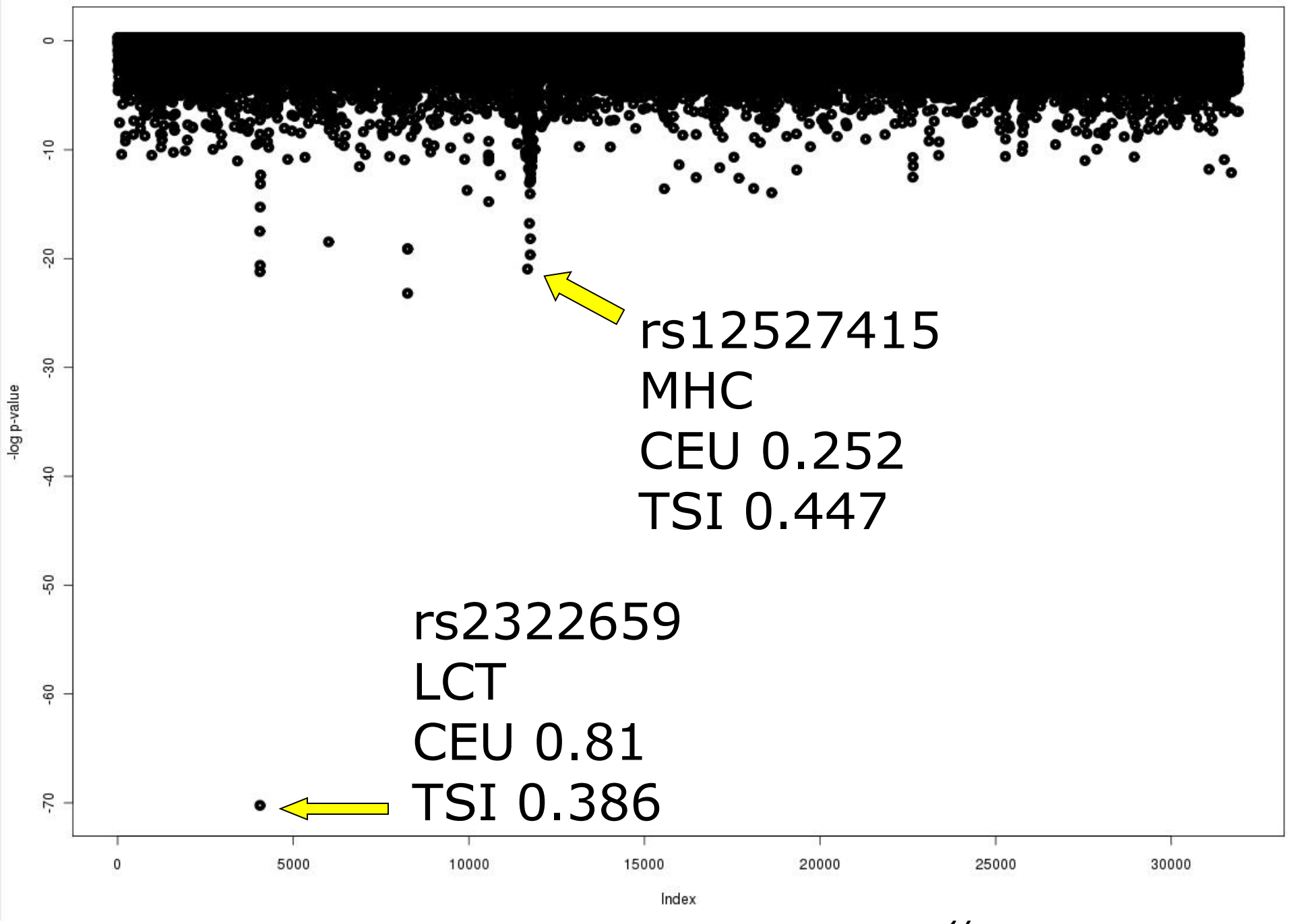
PCA using Ancestry Informative SNPs

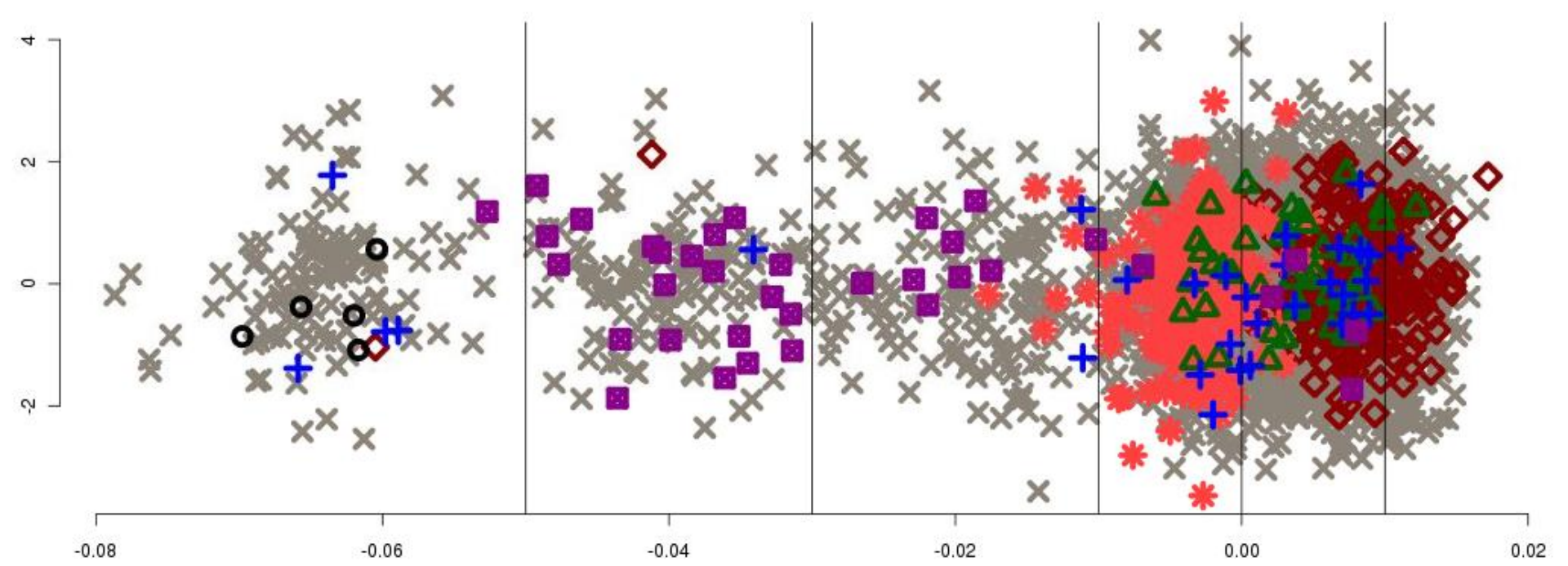
- "We … achieve perfect intercontinental differentiation with **14** PCA-correlated SNPs" – Paschou et al. PLoS Genetics 2007 e160
- "Only **150-200** PCAIMs (PCA-informative markers) suffice to accurately predict fine structure in European Americans" – Paschou et al. PLoS Genetics 2008 e1000114
- Price et al. show that **100** SNPs explains the ancestry of non-Ashkenazi Jewish European American. PLoS Genetics e236

How to select ancestry informative PCA's?

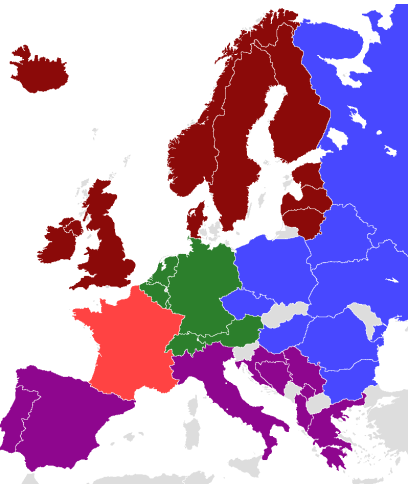
- Take dataset
- Remove outliers (ie non-Caucasians)
- Perform PCA again
- Regress first few PC's against SNPs
- Select SNPs which are of the highest technicaly quality which best explain PC's


1986 OHGS_extra
90 CEU
90 JPT+CHB
90 YRI
153 outlier


rs12527415
MHC
CEU 0.252
TSI 0.447
rs2322659
LCT
CEU 0.81
TSI 0.386
-log p-value
Index
0
-10
-20
-30
-40
-50
-60
-70
0
5000
10000
15000
20000
25000
30000



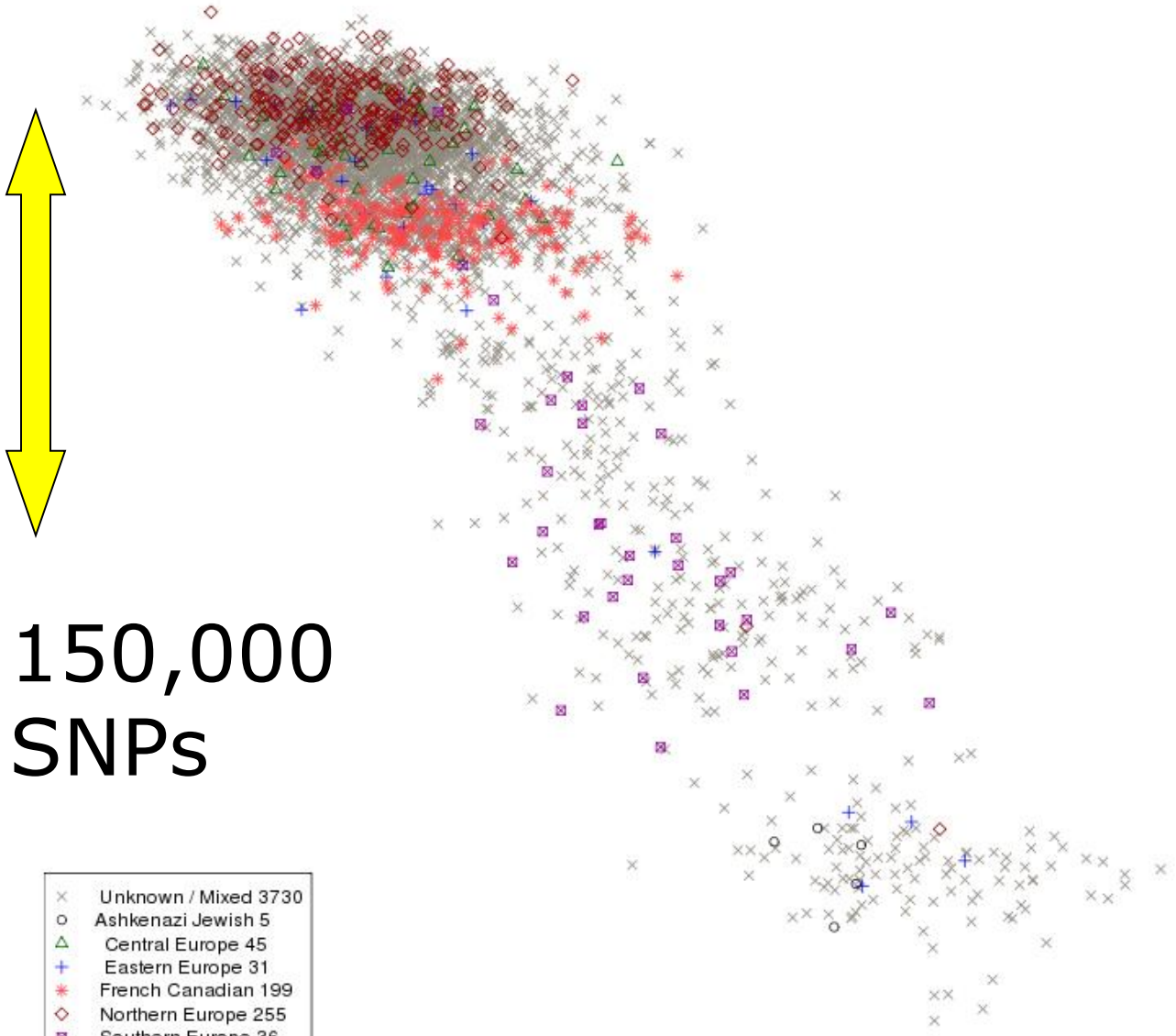
	0	1	2	3	4	5
CC	0.1785714	0.2097902	0.3253012	0.4196332	0.60542169	0.74309979
CT	0.3785714	0.3916084	0.4939759	0.4368932	0.34789157	0.23354565
TT	0.4428571	0.3986014	0.1807229	0.1434736	0.04668675	0.02335456



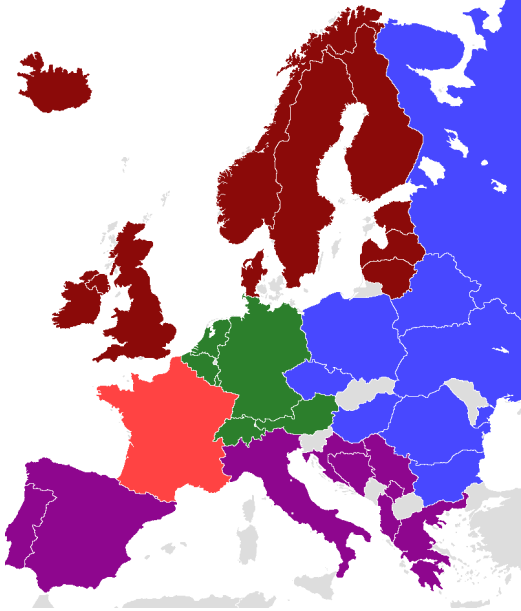
	0	1	2	3	4	5
CC	25	30	54	389	1608	350
CT	53	56	82	405	924	110
TT	62	57	30	133	124	11

LCT gene rs2322659

200 SNPs



150,000 SNPs



R2 ~=0.6

How to use this data?

- Ensure homogenous population
- Use as a covariate in future prediction studies
  - Very interesting – will this matter?
  - Determine ethnic susceptibility
- As a correction in GWAS where sharing subject level whole genome data is not possible

- Part 0 - Rationale
- Part 1 - Genetics
- Part 2 - Ethnicity
- Part 3 - Prediction

Prediction

- This section
  - Describe a paper recently published by us

Improved Prediction of Cardiovascular Disease Based on a Panel of Single-Nucleotide Polymorphisms Identified Through Genome-Wide Association Studies

Robert W. Davies, MSc; Sonny Dandona, MD; Alexandre F.R. Stewart, PhD; Li Chen, MSc; Stephan G. Ellis, MD; W.H. Wilson Tang, MD; Stanley L. Hazen, MD, PhD; Robert Roberts, MD; Ruth McPherson, MD, PhD; George A. Wells, MSc, PhD

Background

- Several CAD GWAS identified 9p21 in 2007

A Common Allele on Chromosome 9 Associated with Coronary Heart Disease

Ruth McPherson,[1]*† Alexander Pertsemlidis,[2]* Nihan Kavaslar,[1] Alexandre Stewart,[1] Robert Roberts,[1] David R. Cox,[3] David A. Hinds,[3] Len A. Pennacchio,[4,5] Anne Tybjaerg-Hansen,[6] Aaron R. Folsom,[7] Eric Boerwinkle,[8] Helen H. Hobbs,[2,9] Jonathan C. Cohen[2,10]†

A Common Variant on Chromosome 9p21 Affects the Risk of Myocardial Infarction

Anna Helgadottir,[1]* Gudmar Thorleifsson,[1]* Andrei Manolescu,[1]* Solveig Gretarsdottir,[1] Thorarinn Blondal,[1] Aslaug Jonasdottir,[1] Adalbjorg Jonasdottir,[1] Asgeir Sigurdsson,[1] Adam Baker,[1] Arnar Palsson,[1] Gisli Masson,[1] Daniel F. Gudbjartsson,[1] Kristinn P. Magnusson,[1] Karl Andersen,[2] Allan I. Levey,[3] Valgerdur M. Backman,[1] Sigurborg Matthiasdottir,[1] Thorbjorg Jonsdottir,[1] Stefan Palsson,[1] Helga Einarsdottir,[1] Steinunn Gunnarsdottir,[1] Arnaldur Gylfason,[1] Viola Vaccarino,[3] W. Craig Hooper,[3] Muredach P. Reilly,[4] Christopher B. Granger,[5] Harland Austin,[3] Daniel J. Rader,[4] Svati H. Shah,[5] Arshed A. Quyyumi,[3] Jeffrey R. Gulcher,[1] Gudmundur Thorgeirsson,[2] Unnur Thorsteinsdottir,[1] Augustine Kong,[1]† Kari Stefansson[1]†

The NEW ENGLAND JOURNAL of MEDICINE

Genomewide Association Analysis of Coronary Artery Disease

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., Inke R. König, Ph.D., Suzanne E. Stevens, M.Sc., Silke Szymczak, M.Sc., David-Alexandre Tregouet, Ph.D., Mark M. Iles, Ph.D., Friedrich Pahlke, M.Sc., Helen Pollard, M.Sc., Wolfgang Lieb, M.D., Francois Cambien, M.D., Marcus Fischer, M.D., Willem Ouwehand, F.R.C.Path., Stefan Blankenberg, M.D., Anthony J. Balmforth, Ph.D., Andrea Baessler, M.D., Stephen G. Ball, F.R.C.P., Tim M. Strom, M.D., Ingrid Brænne, M.Sc., Christian Gieger, Ph.D., Panos Deloukas, Ph.D., Martin D. Tobin, M.F.P.H.M., Andreas Ziegler, Ph.D., John R. Thompson, Ph.D., and Heribert Schunkert, M.D., for the WTCCC and the Cardiogenics Consortium*

Vol 447|7 June 2007|doi:10.1038/nature05911 nature

Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls

The Wellcome Trust Case Control Consortium*

nature genetics

Background

- Several more loci were identified in 2009

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

Myocardial Infarction Genetics Consortium*

A genome-wide meta-analysis identifies 22 loci associated with eight hematological parameters in the HaemGen consortium

Nicole Soranzo[1,2,45*], Tim D Spector[2,45], Massimo Mangino[2,45], Brigitte Kühnel[3], Augusto Rendon[4], Alexander Teumer[5], Christina Willenborg[6,7], Benjamin Wright[8], Li Chen[9], Mingyao Li[10], Perttu Salo[11,12], Benjamin F Voight[13,14], Philippa Burns[4], Roman A Laskowski[15], Yali Xue[1], Stephan Menzel[16], David Altshuler[13,14,17–19], John R Bradley[20], Suzannah Bumpstead[1], Mary-Susan Burnett[21], Joseph Devaney[21], Angela Döring[3], Roberto Elosua[22], Stephen E Epstein[21], Wendy Erber[23], Mario Falchi[2,24], Stephen F Garner[4], Mohammed J R Ghori[1], Alison H Goodall[25], Rhian Gwilliam[1], Hakon H Hakonarson[26], Alistair S Hall[27], Naomi Hammond[1], Christian Hengstenberg[28], Thomas Illig[3], Inke R König[6], Christopher W Knouff[29], Ruth McPherson[9], Olle Melander[30], Vincent Mooser[29], Matthias Nauck[31], Markku S Nieminen[32], Christopher J O'Donnell[18,33], Leena Peltonen[11,12], Simon C Potter[1], Holger Prokisch[34,35], Daniel J Rader[36,37], Catherine M Rice[1], Robert Roberts[9], Veikko Salomaa[11,12], Jennifer Sambrook[4], Stefan Schreiber[38], Heribert Schunkert[7], Stephen M Schwartz[39,40], Jovana Serbanovic-Canic[4], Juha Sinisalo[32], David S Siscovick[39,40], Klaus Stark[28], Ida Surakka[12], Jonathan Stephens[4], John R Thompson[8], Uwe Völker[5], Henry Völzke[41], Nicholas A Watkins[4], George A Wells[9], H-Erich Wichmann[3,42], David A Van Heel[43], Chris Tyler-Smith[1], Swee Lay Thein[16], Sekar Kathiresan[18,33], Markus Perola[11,12], Muredach P Reilly[36,37], Alexandre F R Stewart[9], Jeanette Erdmann[7], Nilesh J Samani[25], Christa Meisinger[3], Andreas Greinacher[44], Panos Deloukas[1,45], Willem H Ouwehand[1,4,45] & Christian Gieger[3,45]

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], Zouhair Aherrahrou[1], Arnika K Wagner[1], Klaus Stark[4], Stephen M Schwartz[9], Veikko Salomaa[10], Roberto Elosua[11], Olle Melander[12], Benjamin F Voight[13], Christopher J O'Donnell[14], Leena Peltonen[15], David S Siscovick[9], David Altshuler[16], Piera Angelica Merlini[17], Flora Peyvandi[18], Luisa Bernardinelli[19,20], Diego Ardissino[21], Arne Schillert[2], Stefan Blankenberg[22], Tanja Zeller[22], Philipp Wild[22], Daniel F Schwarz[2], Laurence Tiret[8], Claire Perret[8], Stefan Schreiber[23], Nour Eddine El Mokhtari[23], Arne Schäfer[23], Winfried März[24–26], Wilfried Renner[25], Peter Bugert[27], Harald Klüter[27], Jürgen Schrezenmeir[28], Diana Rubin[28], Stephen G Ball[5], Anthony J Balmforth[5], H-Erich Wichmann[29,30], Thomas Meitinger[31,32], Marcus Fischer[4], Christa Meisinger[29], Jens Baumert[29], Annette Peters[29], Willem H Ouwehand[33], Italian Atherosclerosis, Thrombosis, and Vascular Biology Working Group[34], Myocardial Infarction Genetics Consortium[34], Wellcome Trust Case Control Consortium[34], Cardiogenics Consortium[34], Panos Deloukas[15], John R Thompson[7], Andreas Ziegler[2], Nilesh J Samani[3] & Heribert Schunkert[1]

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], Maylis DeSuremain[1], Thomas Meitinger[6], Ben J Wright[7], Michael Preuss[2], Anthony J Balmforth[5], Stephen G Ball[5], Christa Meisinger[6], Cécile Germain[8], Alun Evans[9], Dominique Arveiler[10], Gérald Luc[11], Jean-Bernard Ruidavets[12], Caroline Morrison[13], Pim van der Harst[4], Stefan Schreiber[14], Katharina Neureuther[15], Arne Schäfer[14], Peter Bugert[16], Nour E El Mokhtari[14], Jürgen Schrezenmeir[17], Klaus Stark[15], Diana Rubin[17], H-Erich Wichmann[6], Christian Hengstenberg[15], Willem Ouwehand[18], Wellcome Trust Case Control Consortium[19], Cardiogenics Consortium[19], Andreas Ziegler[2], Laurence Tiret[1], John R Thompson[7], Francois Cambien[1], Heribert Schunkert[3] & Nilesh J Samani[4]

Sequence variants affecting eosinophil numbers associate with asthma and myocardial infarction

Daniel F Gudbjartsson*[1], Unnur S Bjornsdottir[1,2], Eva Halapi[1], Anna Helgadottir[1], Patrick Sulem[1], Gudrun M Jonsdottir[1], Gudmar Thorleifsson[1], Hafdis Helgadottir[1], Valgerdur Steinthorsdottir[1], Hreinn Stefansson[1], Carolyn Williams[3–5], Jennie Hui[6,7], John Beilby[6,8], Nicole M Warrington[9], Alan James[10,11], Lyle J Palmer[9], Gerard H Koppelman[12], Andrea Heinzmann[13], Marcus Krueger[13], H Marike Boezen[14], Amanda Wheatley[15], Janine Altmuller[16], Hyoung Doo Shin[17,18], Soo-Taek Uh[19], Hyun Sub Cheong[19], Brynja Jonsdottir[20], David Gislason[20], Choon-Sik Park[21], Linda M Rasmussen[22], Celeste Porsbjerg[22], Jakob W Hansen[22], Vibeke Backer[22], Thomas Werge[23], Christer Janson[24], Ulla-Britt Jönsson[24], Maggie C Y Ng[25], Juliana Chan[25], Wing Yee So[25], Ronald Ma[25], Svati H Shah[26], Christopher B Granger[26], Arshed A Quyyumi[27], Allan I Levey[27], Viola Vaccarino[27], Muredach P Reilly[28], Daniel J Rader[28], Michael J A Williams[29], Andre M van Rij[29], Gregory T Jones[29], Elisabetta Trabetti[30], Giovanni Malerba[30], Pier Franco Pignatti[30], Attilio Boner[31], Lydia Pescollderungg[32], Domenico Girelli[33], Oliviero Olivieri[33], Nicola Martinelli[33], Bjorn R Ludviksson[2,20], Dora Ludviksdottir[20], Gudmundur I Eyjolfsson[34], David Arnar[2,20], Gudmundur Thorgeirsson[2,20], Klaus Deichmann[13], Philip J Thompson[3–5], Matthias Wjst[35,36], Ian P Hall[16], Dirkje S Postma[37], Thorarinn Gislason[2,20], Jeffrey Gulcher[1], Augustine Kong[1], Ingileif Jonsdottir[1,2,20], Unnur Thorsteinsdottir[1,2] & Kari Stefansson[1,2]

Background

- Several studies have commented on the ability of 9p21 to add to risk prediction in prospective cohorts
  - Talmud 2008
    - N=2742 men, n=270 events
    - AUC 0.62 to 0.64 p NS, LR p=0.01
  - Brautbar 2009
    - N=9998 whites, n=1349 events
    - AUC 0.782 to 0.786 CI (0.001, 0.007)
  - Paynter 2009
    - N=22,129 white women, n=615 events
    - AUC 0.807 to 0.809 NS

Background

- One study commented on more than just 9p21's ability to predict CAD
  - Paynter 2010
    - 12 CAD/Stroke SNPs and 101 SNP GRS
    - Neither significant

Association Between a Literature-Based Genetic Risk Score and Cardiovascular Events in Women

Nina P. Paynter, PhD
Daniel I. Chasman, PhD
Guillaume Paré, MD, MS
Julie E. Buring, ScD
Nancy R. Cook, ScD
Joseph P. Miletich, MD, PhD
Paul M Ridker, MD, MPH

Background

- Also, some studies had success with alternative methodologies

From Disease Association to Risk Assessment: An Optimistic View from Genome-Wide Association Studies on Type 1 Diabetes

Zhi Wei[1☯], Kai Wang[2☯], Hui-Qi Qu[3], Haitao Zhang[2], Jonathan Bradfield[2], Cecilia Kim[2], Edward Frackleton[2], Cuiping Hou[2], Joseph T. Glessner[2], Rosetta Chiavacci[2], Charles Stanley[4], Dimitri Monos[5], Struan F. A. Grant[2,6], Constantin Polychronakos[3], Hakon Hakonarson[2,6]*

Objectives

- Using GWAS SNPs for CAD/MI
- 1) See if newer SNPs add to 9p21
- 2) Test out a few different methods

Methods

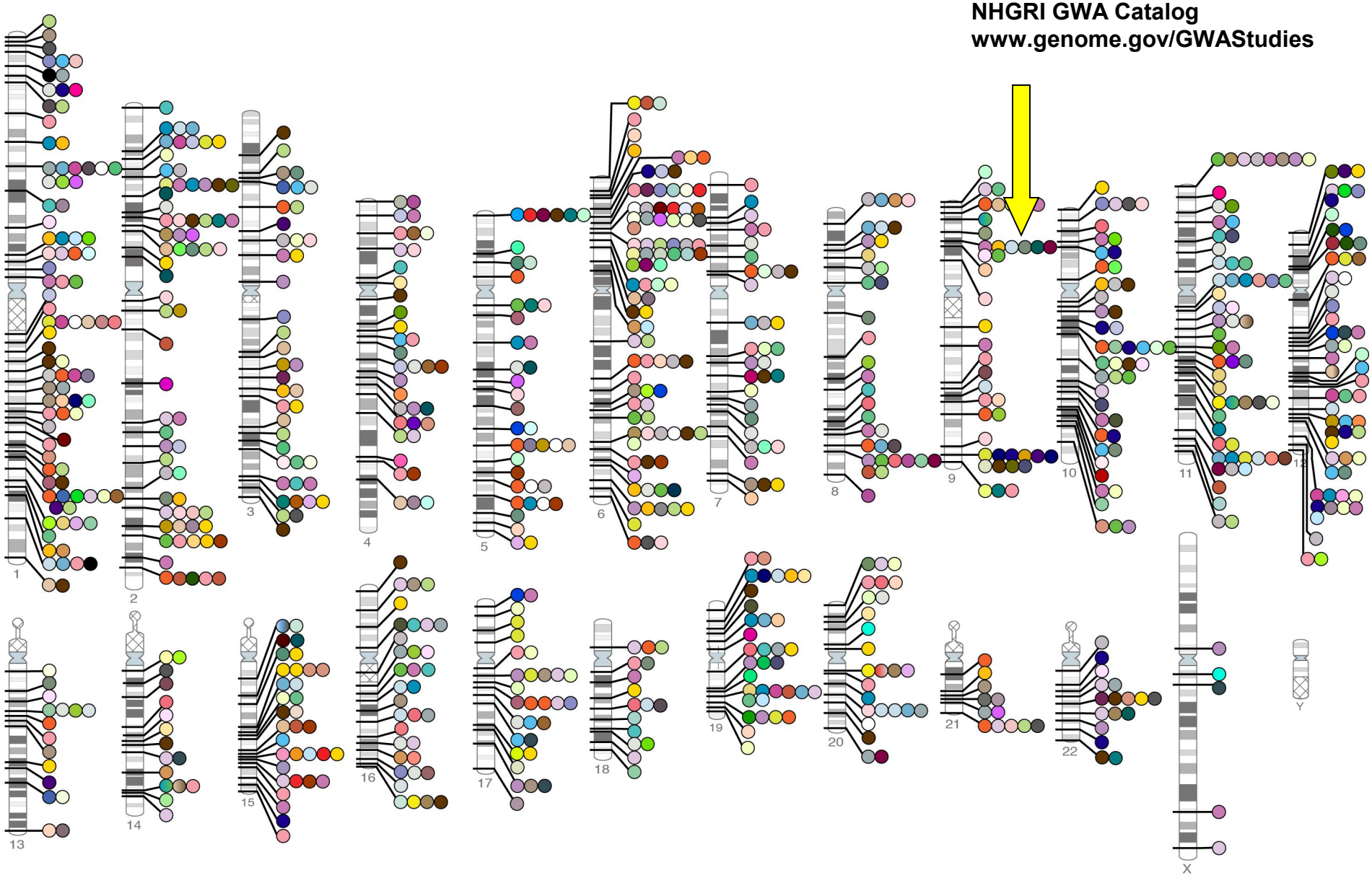
- Identify SNPs from literature
- Get these SNPs, or proxies, in our data ready
- Test to see whether newer SNPs add to 9p21
  - If true, check to see if this holds true when known risk factors are used
- Test to see whether different prediction algorithms affect results

Results 1 – Identifying SNPs from Literature (1/4)

- Resource used – National Human Genome Research Institute catalogue of GWA studies
- http://www.genome.gov/26525384

Published Genome-Wide Associations through 6/2010, 904 published GWA at $p \leq 5 \times 10^{-8}$ for 165 traits

NHGRI GWA Catalog
www.genome.gov/GWAStudies



Results 1 – Identifying SNPs from Literature (3/4)

- Inclusion criteria
  - GWAS study
    - Either primary CAD/MI
    - Or secondary analysis of CAD/MI, where primary analysis was related trait
      - Lipid traits, hypertension, etc
  - Reported p-value <=5e-7

Results 1 – Identifying SNPs from Literature (4/4)

Table 2. Details of Previously Identified SNPs

Studies	Locus	Physical Location, Mb	Original SNP	Genes in Region
5	1p32	55.27	Rs11206510	PCSK9
2, 5	1p13	109.62	Rs646776	CELSR2/PSRC1/SORT1
5	1q41	220.87	Rs17465637	MIA3
5	2q33	203.45	Rs6725887	WDR12
7	3q22	139.60	Rs9818870	MRAS
5	6p24	13.04	Rs12526453	PHACTR1
8	6q26-27	160.88	Rs3798220	SLC22A3/LPAL2/LPA
2–5	9p21	22.09	Rs4977574	CDKN2A/CDKN2B
2, 5	10q11	44.10	Rs1746048	CXCL12
6, 9	12q24	111.36	Rs11066301	SH2B3/ATXN2/PTPN11
7	12q24	119.92	Rs2259816	HNF1A/C12orf43
5	19p13	11.02	Rs1122608	LDLR
5	21q22	34.52	Rs9982601	SLC5A3/MRPS6/KCNE2

Results 2 – Prepare data for analysis (1/2)

- Ottawa Heart Genomics Study
  - Cases M<55, F<65 (n=3323)
  - MI, CABG, PCI, Stenosis >=50%
  - Controls M>65 F>70 (n=2319)
  - Asymptomatic or none of the above
- Wellcome Trust Case Control Consortium
  - Cases M+F<66 (n=1926)
  - MI, CABG, PCI
  - Controls population randoms no phenotypes (n=2938)

Results 2 – Prepare data for analysis (2/2)

Table 2. Details of Previously Identified SNPs

Studies	Locus	Physical Location, Mb	Using SNP	SNP Type	Original SNP	Genes in Region	OHGS, OR (95% CI)
5	1p32	55.27	rs11206510	G	Rs11206510	PCSK9	1.00 (0.91,1.10)*
2, 5	1p13	109.62	rs646776	I	Rs646776	CELSR2/PSRC1/SORT1	1.18 (1.08,1.30)
5	1q41	220.87	rs17465637	G	Rs17465637	MIA3	1.15 (1.06,1.25)
5	2q33	203.45	rs6725887	G	Rs6725887	WDR12	1.28 (1.14,1.43)
7	3q22	139.60	rs9818870	G	Rs9818870	MRAS	1.13 (1.02,1.25)
5	6p24	13.04	rs12526453	G	Rs12526453	PHACTR1	1.11 (1.03,1.21)
8	6q26-27	160.88	CCTC haplo	H	Rs3798220	SLC22A3/LPAL2/LPA	1.79 (1.38,2.31)
2–5	9p21	22.09	rs4977574	G	Rs4977574	CDKN2A/CDKN2B	1.46 (1.35,1.57)
2, 5	10q11	44.10	rs1746049	G-T	Rs1746048	CXCL12	1.17 (1.04,1.31)
6, 9	12q24	111.36	rs11066301	I	Rs11066301	SH2B3/ATXN2/PTPN11	1.17 (1.08,1.26)
7	12q24	119.92	rs2259816	G	Rs2259816	HNF1A/C12orf43	1.13 (1.04,1.22)
5	19p13	11.02	rs1122608	G	Rs1122608	LDLR	1.20 (1.09,1.31)
5	21q22	34.52	rs9978407	G-T	Rs9982601	SLC5A3/MRPS6/KCNE2	1.25 (1.12,1.40)

Original SNP refers to the SNP as identified through the relevant GWAS; using refers to the SNP being used in our analysis. SNP type refers to whether the SNP being used was the original genotype SNP (G), a tag SNP of the genotype SNP (G-T), an imputed SNP (I), or a haplotype (H). Odds ratios (ORs) and 95% confidence intervals (CIs) are given for the risk allele for a logistic regression model containing all 13 SNPs.

*rs1120650 was removed after the model-fitting procedure.

Results 3 – Test whether new SNPs add to 9p21 (1/3)

Table 3. Results for Models Trained Using Sets of SNPs

	OHGS			WTCCC		
	LR	AC	SVM	LR	AC	SVM
9p21.3/rs4977574	0.555	0.555	0.555	0.556	0.556	0.556
All	0.608	0.599	0.581	0.602	0.593	0.579

P=3.59e-14

P=3.50e-11

Results 3 – Test whether new SNPs add to 9p21 (2/3)

- Subset of OHGS with baseline (no drugs) lipids (n=1388 cases, n=2038 controls)
- Variables
  - Smoke current (Y/N)
  - Hypertension (Y/N)
  - Total Cholesterol
  - HDL
  - Sex

Table 1. Clinical Characteristics of Ottawa Heart Genomics Study

	Cases	Control Subjects
No.	3323	2319
Age*	48.6±7.2	75.0±5.2
Men, %	75.9	51.7
Body mass index	29.0±5.2	26.2±4.1
Smoke current, %	21.3	2.5
Hypertension, %	58.9	39.1
Cholesterol, mmol/L†	5.92±1.2	5.67±1.0
TG, mmol/L†	2.07±1.1	1.33±0.7
LDL-C, mmol/L†	3.84±1.1	3.59±0.9
HDL-C, mmol/L†	1.16±0.4	1.48±0.4

TG indicates triglyceride; LDL, low-density lipoprotein; and HDL, high-density lipoprotein.

Values reported are mean±1 SD. All measures are significantly different ($P<0.001$) between cases and control subjects as measured by *t* tests for the continuous variables and χ^2 tests for the binary traits.

*Age refers to age at diagnosis (cases) and age at consent (control subjects).

†All 4 lipid measures were available for 1248 cases and 2016 control subjects at baseline.

Results 3 – Test whether new SNPs add to 9p21 (3/3)

- 1) TRFs (AUC=0.8013)
- 2) TRFs + 9p21 (AUC=0.8044)
- 3) TRFs + 12 (AUC=0.8097)
- p-value 1 vs 2 = 0.097
- p-value 2 vs 3 = 0.037
- p-value 1 vs 3 = 0.0073

Results 4 – Test methods

Table 3. Results for Models Trained Using Sets of SNPs

	OHGS			WTCCC		
	LR	AC	SVM	LR	AC	SVM
9p21.3/rs4977574	0.555	0.555	0.555	0.556	0.556	0.556
All	0.608	0.599	0.581	0.602	0.593	0.579

Values are AUCs for both LR and SVM trained on either 9p21.3 alone or using 12 SNPs.

OHGS

LR vs AC
0.016

LR vs SVM
3.79e-6

Discussion

- 12 SNPs did roughly twice as well as 1
- Somewhat expected, but good to see given uncertainty in previous prospective work

Discussion

- Logistic regression outperformed allele counting, but only very marginally
  - Somewhat surprising result is seen consistently in literature
- SVM did not do very well
  - Only one SNP per locus takes away ability of SVM to detect multiple signals from the same locus
  - Might have allowed too much flexibility in classifier

Discussion

- Other notable future publications
  - Ribatti et al, in press
  - ~30,000 Scandanavians prospective cohort
  - 13 SNP allele counting GRS, should be same 13 loci as identified
  - ~1.7 Hazard ratio for top quintile of score versus lowest quintile (believed to be adjusted for risk factors)
  - ~1.7 HR for LDL, other common factors

Conclusion

- 1000$ genome = 5 to 10 years away
- Better methods are needed to handle sequence data

Thanks

- Colleagues at Cardiovascular Research Methods Centre
  - Dr. George Wells
  - Li Chen
  - Kathryn Williams
- John and Jennifer Ruddy Canadian Cardiovascular Genetics Centre
- Wellcome Trust Case Control Consortium
- R, HapMap, 1kG, Craig Venter, Pubmed, Eigensoft, etc.

Any questions?