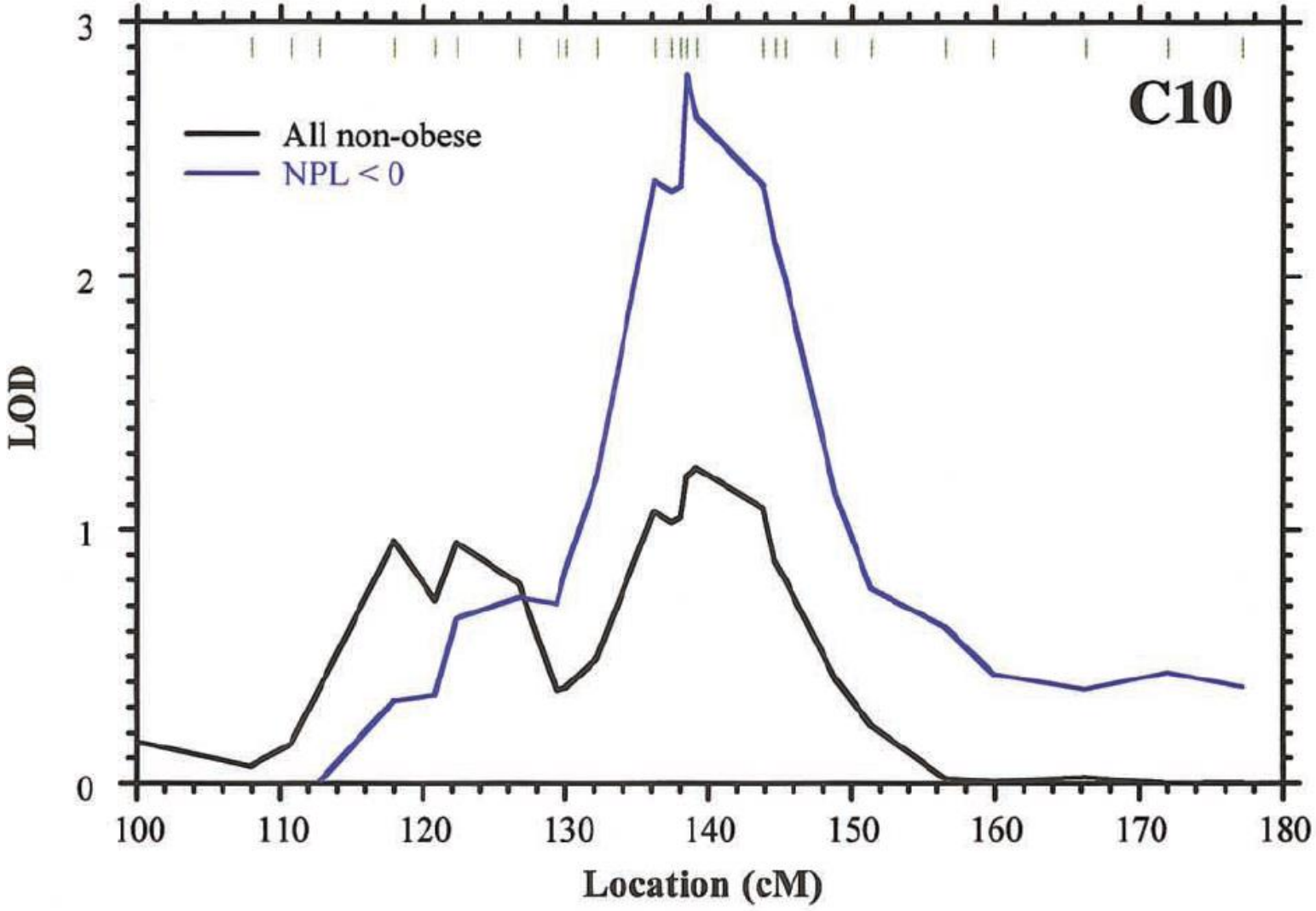
HapMap and Association Studies for Complex Diseases Two Examples

Augustine Kong

deCODE genetics

Mapping of a susceptibility gene for Type 2 Diabetes



- *Also observed in Mexican Americans*

Reynisdottir *et al*, AJHG 2003;73:323-35

Locus-wide association study

- **High density of markers – 10.5 Mb region**
- **Typed 228 microsatellite markers**
- **Average density = one marker every 46 kb**
- **1185 T2D patients and 931 population controls**

DG10S478 : Iceland

Allele	Affected freq (n=1185)	Control freq (n=931)	Relative Risk (multiplicative)	Two sided p-val
0	**0.636**	**0.724**	**0.67**	**2.1×10^{-9}**
4	0.005	0.002	2.36	0.12
8	0.093	0.078	1.21	0.09
12	0.242	0.178	1.48	4.6×10^{-7}
16	0.022	0.015	1.53	0.076
20	0.001	0.003	0.39	0.17

DG10S478 : Iceland

- **DG10S478 genotyped in the CEPH Utah (CEU) HapMap samples**
  - SNP1 - allele G - correlated with allele 0 of DG10S478 ($R^2 = 0.95$, $P = 5.53\times10^{-38}$)
  - SNP1 - allele T - correlated with the other alleles

- **Risk conferred by alleles 8 and 12 of DG10S478 do not differ significantly (*P* = 0.3).**

- **Phylogenetic analysis of haplotypic variation within the LD block where DG10S478 resides**
  - all haplotypes carrying DG10S478 non-0 alleles and SNP1 T in the CEU samples belong to a single, clearly defined monophyletic lineage
  - i.e. they share a relatively recent common ancestor within the phylogeny

- **Reasonable to collapse all the non-0 alleles of DG10S487 into composite allele X.**

DG10S478 : Iceland

Allele	Affected freq (n=1185)	Control freq (n=931)	Relative Risk (multiplicative)	Two sided p-val
0	**0.636**	**0.724**	**0.67**	**2.1×10^{-9}**
4	0.005	0.002	2.36	0.12
8	0.093	0.078	1.21	0.09
12	0.242	0.178	1.48	4.6×10^{-7}
16	0.022	0.015	1.53	0.076
20	0.001	0.003	0.39	0.17
X	**0.364**	**0.276**	**1.50**	**2.1×10^{-9}**

DG10S478 : Denmark

Allele	Affected freq (n=228)	Control freq (n=539)	Relative Risk (multiplicative)	One sided p-val
0	**0.669**	**0.740**	**0.71**	**0.0024**
4	0.002	0.004	0.59	0.310
8	0.070	0.048	1.49	0.046
12	0.239	0.190	1.34	0.016
16	0.020	0.018	1.12	0.390
X	**0.331**	**0.260**	**1.41**	**0.0024**

DG10S478 : USA

Allele	Affected freq (n=361)	Control freq (n=530)	Relative Risk (multiplicative)	One sided p-val
-4	0.001	0.000	-	-
0	**0.615**	**0.747**	**0.54**	**1.7×10^{-9}**
4	0.003	0.004	0.73	0.358
8	0.085	0.049	1.79	0.001
12	0.256	0.180	1.57	6.2×10^{-5}
16	0.040	0.020	2.07	0.006
X	**0.385**	**0.253**	**1.85**	**1.7×10^{-9}**

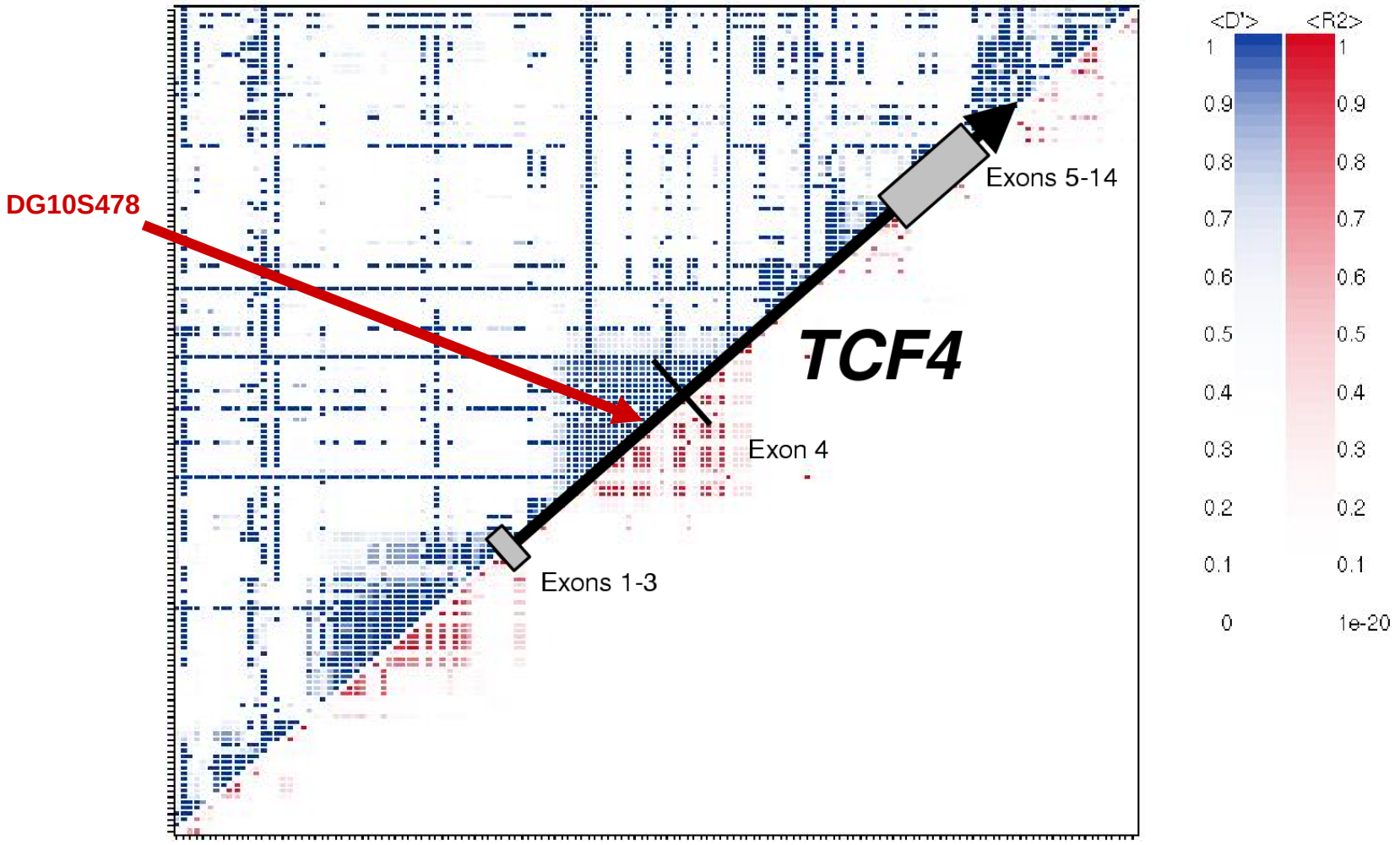
DG10S478 : Estimates of the Genotype relative risks

	Genotype RR			
Cohort	00	0X	XX	PAR
Iceland	1	1.41	2.27	0.21
Denmark	1	1.37	1.92	0.17
USA	1	1.64	3.29	0.28
Combined	**1**	**1.45**	**2.41**	**0.21**

- **Estimated relative risks between cohorts - not significantly different (*P* > 0.05)**

- **Combining the results from all 3 cohorts yields an overall two-sided *P* of 4.6×10^{-18}**
  - Given that the original 228 microsatellite markers tested have a total of 1664 alleles and allele X is the complement of allele 0, applying Bonferonni adjustment gives a *P* of 7.7×10^{-15}

Only one gene in the LD Block
TCF4 (official gene symbol: *TCF7L2*)



Correlation of five selected HapMap SNPs with DG10S478 (with highest R^2 among the Phase I SNPs)

	CEPH Utah HapMap cohort	Combined Icelandic and US cohorts
	R^2	R^2
SNP1	0.95	0.93
SNP2	0.78	0.72
SNP3	0.61	0.65
SNP4	0.43	0.44
SNP5	0.42	0.45

Association of the at-risk alleles of the five selected HapMap SNPs and the composite allele X of DG10S478 to T2D in both Iceland and the US

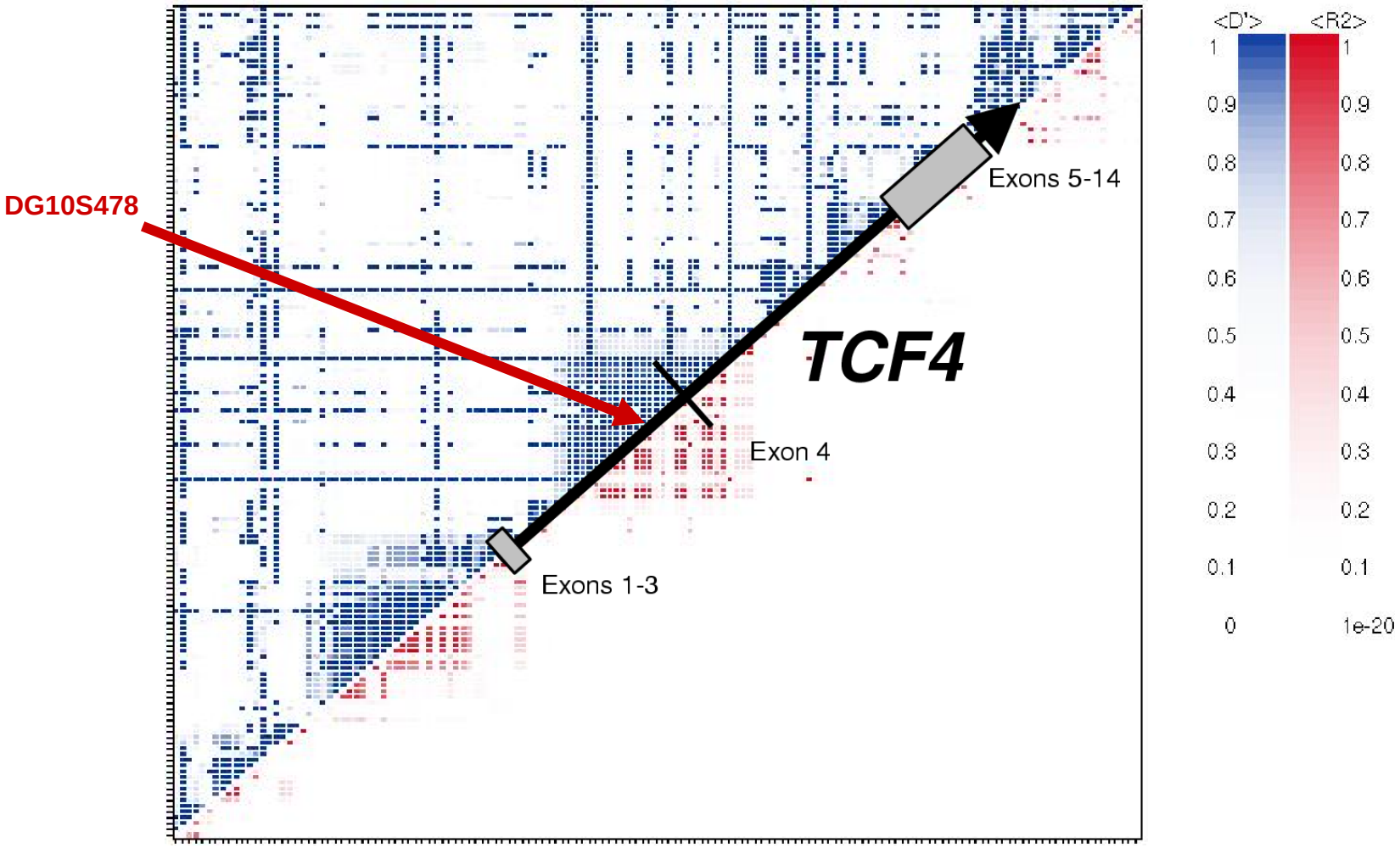
Subsets: Iceland (331 patients, 320 controls) US (226 patients, 210 controls)

		Icelandic cohort				US cohort				Combined	
	Allele	Patients (331)	Controls (320)	RR	P-value	Patients (226)	Controls (210)	RR	P-value	RR	P-value
SNP3	C	0.382	0.289	1.52	0.00040	0.388	0.306	1.44	0.010	1.49	1.3×10^{-5}
SNP2	T	0.380	0.289	1.50	0.00056	0.374	0.268	1.63	0.00083	1.55	1.7×10^{-6}
DG10S478	X	0.377	0.280	1.55	0.00021	0.369	0.243	1.82	5.6×10^{-5}	1.66	5.9×10^{-8}
SNP5	A	0.539	0.458	1.38	0.0034	0.527	0.476	1.22	0.14	1.32	0.0013
SNP4	C	0.543	0.466	1.36	0.0053	0.536	0.479	1.26	0.093	1.32	0.0013
SNP1	T	0.370	0.288	1.45	0.0016	0.376	0.250	1.81	6.0×10^{-5}	1.59	6.3×10^{-7}

- All five SNPs show association to T2D, but none exhibit stronger association to T2D than DG10S478

- Strength of the association to T2D corresponds monotonically to the correlation between each SNP and DG10S478

Further Search of the Causal Variant



- ***Exon 4 mutation ruled out***
- ***All other exonic mutations ruled out***
- ***Pooled sequencing across LD block reveals no better SNP***

Summary

- We did not map the variant/gene/region through genome-wide association, but easily could have
  - medium risk, common variant, population attributable risk not small
- However, genome-wide association focusing only on exonic SNPs might not have worked
- We still have not identified the causal variant yet
  - An unidentified SNP? one of the highly correlated SNPs? Not a SNP? Some Structural polymorphism?
- Still, the HapMap data have substantially speed up our progress in exploring the region
  - The LD structure allowed us to be reasonably confident that we have identified the susceptibility gene
- Maybe the Phase II data will help us further

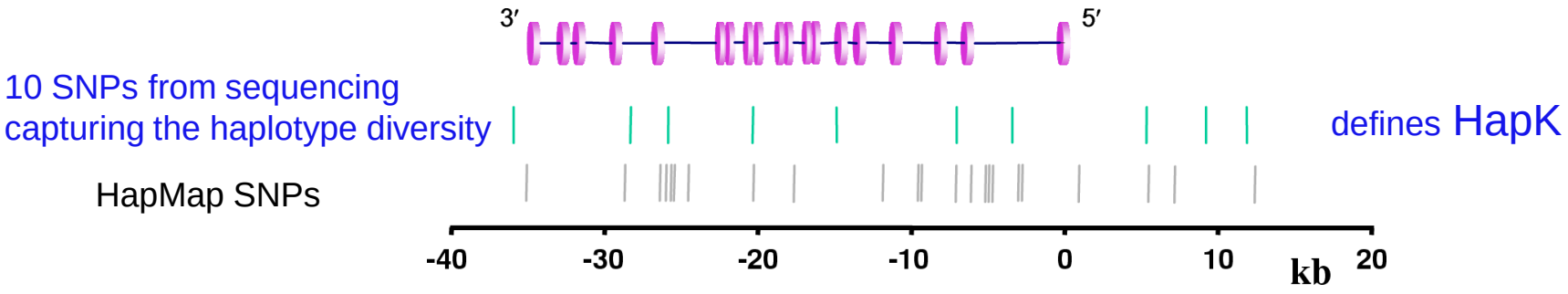
Variant of a gene located on chromosome 10q confers risk of type 2 diabetes mellitus

Saturday Session #63

Struan F.A. Grant[1], G. Thorleifsson[1], I. Reynisdottir[1], R. Benediktsson[2,3], A. Manolescu[1], J. Sainz[1], H. Stefansson[1], V. Emilsson[1], A. Helgadottir[1], U. Styrkarsdottir[1], M.P. Reilly[4], D.J. Rader[4], Y. Bagger[5], C. Christiansen[5], V. Gudnason[2], G. Sigurdsson[2,3], U. Thorsteinsdottir[1], J.R. Gulcher[1], A. Kong[1], K. Stefansson[1]

1) deCODE Genetics, Reykjavik, Iceland; 2) Icelandic Heart Association, Reykjavik, Iceland; 3) Landspitali-University Hospital, Reykjavik, Iceland; 4) University of Pennsylvania Health System, Philadelphia, USA; 5) Center for Clinical and Basic Research A/S, Ballerup, Denmark

Leukotriene A4 Hydrolase (*LTA4H*) gene
Candidate Gene for Myocardial Infarction
Resides in one LD block where there is no other gene



LTA4H structure with exons shown as colored cylinders, and the position of all genotyped SNPs relative to exons shown as green lines. The SNPs and alleles defining **HapK** are SG12S16 (C) (positioned in NCBI human assembly build 34 on chr. 12 94.896055 Mb), rs2660880 (G), rs6538697 (T), rs1978331 (A), rs17677715 (T), rs2247570 (T), rs2660898 (T), rs2540482 (C), rs2660845 (G), and rs2540475 (G), respectively. The relative position of SNPs typed in the HapMap project (Phase I, version 16c.1) are shown as grey lines.

Icelandic Association

Cohorts (n)	Frequency of HapK		RR	*P*-value
	Patients	Controls		
Icelanders				
All MI (1553/863)	0.113	0.104	1.1	0.36
MI and additional CVD (325/863)	0.145	0.104	1.45	0.0091

Additional CVD – Peripheral vascular disease and/or Stroke

P-value of 0.0091 becomes 0.035 after adjusting for multiple haplotyes tested

Marginal significance and very modest risk. Needs replication!

Replication Cohorts: European Americans

Cohorts (n)	Frequency of HapK		RR	*P*-value [b]
	Patients	Controls		
European Americans				
Philadelphia				
All MI (728/430)	0.186	0.143	1.37	0.0051
Cleveland				
All MI (627/792)	0.166	0.151	1.12	0.15
MI and additional CVD (144/792)	0.193	0.151	1.34	0.046
Atlanta				
All MI (236/553)	0.135	0.143	0.94	0.64
MI and additional CVD (39/553)	0.173	0.143	1.25	0.25
Combined				
All MI coh adj (cohort adjustment, Mantel-Haenszel)			1.16	0.018
MI and additional CVD[a] coh adj			1.31	0.037

[a] Additional CVD, Cleveland and Atlanta cohorts only; no information for Philadelphia
[b] *P*-values for replication are one-sided

African Americans

Cohorts (n)	Frequency of HapK		RR	*P*-value
	Patients	Controls		
African Americans				
Philadelphia				
All MI (105/127)	0.103	0.017	6.5	0.000067
Cleveland				
All MI (53/111)	0.122	0.072	1.78	0.11
MI and additional CVD (13/111)	0.152	0.072	2.31	0.14
Atlanta				
All MI (39/149)	0.075	0.015	5.21	0.018
MI and additional CVD (8/149)	0.202	0.015	16.36	0.0039

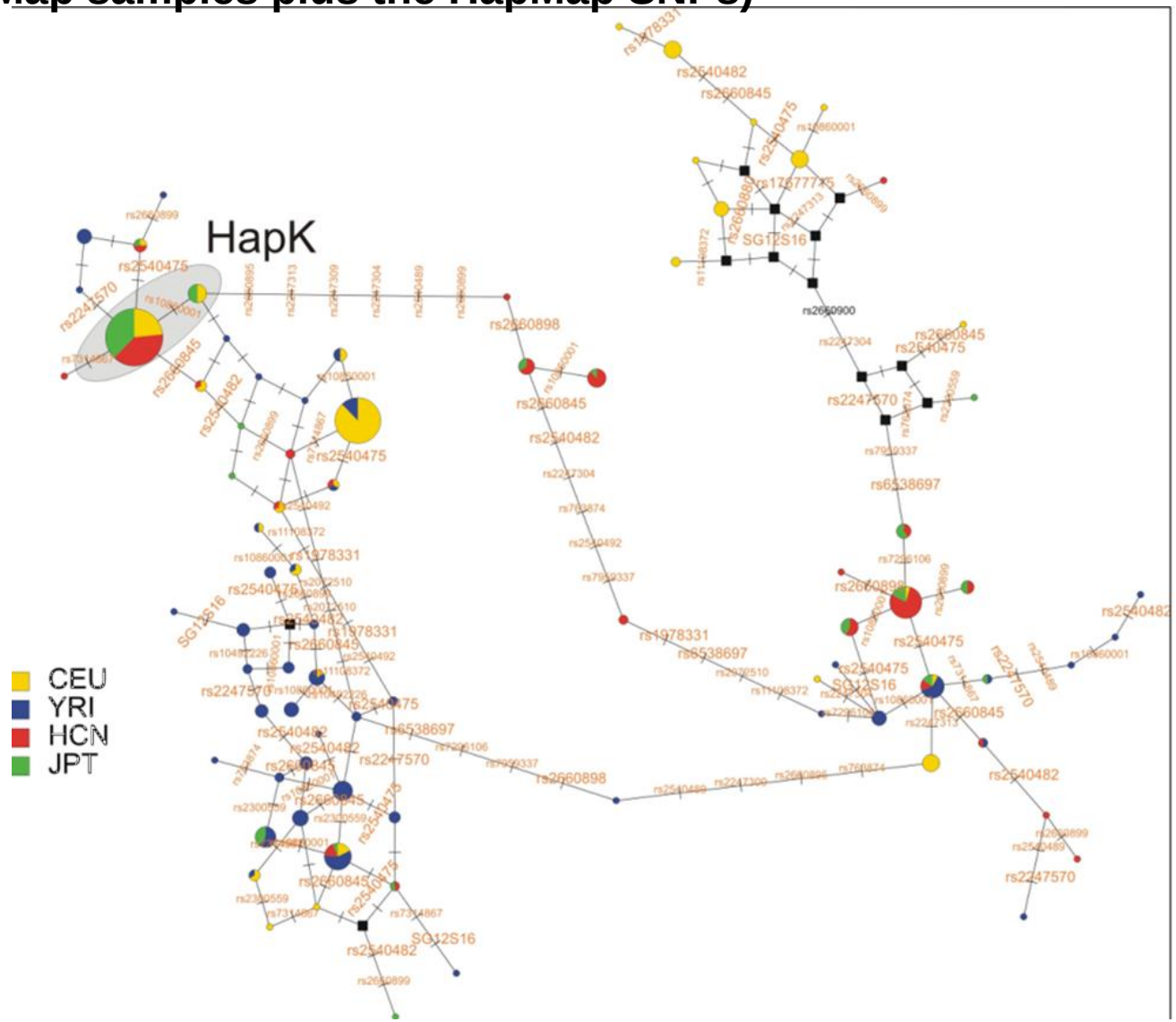
Is this real or is this a consequence of some bias such as imperfect matching of cases and controls? Note that frequency of HapK is substantially lower in the African Americans compared to the European Americans.

Haplotype Diversity in the HapMap Samples

Big differences among populations

	Haplotype (10 SNPs)	CEU (60) Frq	HCB (45) Frq	JPT (45) Frq	YRI (60) Frq
	CGTATTTTAG	37.50%	3.30%	3.00%	8.20%
HapK	**CGTATTTCGG**	**18.30%**	**36.80%**	**50.30%**	-
	CGCGTTGTAG	7.50%	2.00%	4.40%	13.10%
	TGTGCCGTAA	7.00%	1.10%	-	-
	TGTGCCGCGG	5.80%	-	-	-
	TATGTCGTAA	3.50%	-	-	-
	CGTGTCTTGG	3.10%	-	-	8.20%
	CGTGTCTTAG	2.60%	2.30%	10.70%	23.90%
	CGTATTTTAA	2.50%	2.10%	2.80%	1.80%
	CGTGTTTTGG	2.50%	-	-	2.70%
	CATGTCGTAA	1.80%	-	-	-
	CGTATTTCAG	1.70%	1.10%	-	-
	CGTGTCGTGG	1.00%	-	-	-
	TATGTCGTAG	0.90%	-	-	-
	TGTGCCGTAG	0.90%	-	-	-
	CGCGTTGTAA	0.80%	29.50%	17.20%	-
	CGTATTTCGA	0.80%	2.60%	1.80%	0.80%
	CGTGTTTTAG	0.80%	-	-	10.70%
	others	0.00%	19.00%	9.80%	27.00%

A phylogenetic network representing the genealogical relationship between haplotypes in the *LTA4H* region (based on the HapK SNPs which we typed for the HapMap samples plus the HapMap SNPs)



Investigating Ancestry and Admixture Fractions

- Genotyped 75 unlinked microsatellite markers, selected as informative for distinguishing between African and European ancestry
  - --- all the three US cohorts
  - --- 364 Icelanders
  - --- 90 Nigerian Yorubans (HapMap)

- The Structure software was then applied to these data to estimate the fraction of European and African ancestry of individuals.

Distribution of genetically determined European ancestry in MI case-control cohorts

Cohort	Self-Reported Ethnicity	Disease status	WLS group estimate of European ancestry (Std. Err.)[a]	Distribution of estimated individual European ancestry[b]			
				Mean	Std. Deviation	Median	25-75 percentile range
Yoruban Nigerians	**African**	N/A	N/A	**0.036**	0.024	0.03	0.019-0.043
Iceland	**Eur.**	N/A	N/A	**0.991**	0.015	0.994	0.990-0.996
All American	**Eur. Am.**	**Patients**	0.98 (0.0083)	**0.965**	0.083	0.991	0.977-0.995
All American	**Eur. Am.**	**Controls**	0.979 (0.0079)	**0.969**	0.07	0.992	0.979-0.995
All American	**Afr. Am.**	**Patients**	0.243 (0.0138)	**0.223**	0.184	0.178	0.108-0.282
All American	**Afr. Am.**	**Controls**	0.213 (0.016)	**0.199**	0.145	0.174	0.094-0.267
Philadelphia	Afr. Am.	Patients	0.252 (0.0178)	0.235	0.195	0.188	0.121-0.288
Philadelphia	Afr. Am.	Controls	0.213 (0.0217)	0.186	0.137	0.157	0.082-0.257
Cleveland	Afr. Am.	Patients	0.232 (0.0222)	0.21	0.174	0.16	0.096-0.282
Cleveland	Afr. Am.	Controls	0.239 (0.0219)	0.223	0.136	0.191	0.127-0.281
Atlanta	Afr. Am.	Patients	0.226 (0.0246)	0.206	0.166	0.167	0.098-0.283
Atlanta	Afr. Am.	Controls	0.198 (0.0128)	0.193	0.155	0.161	0.086-0.252

Adjusting for Ancestry and Admixture Fractions

- The African American patients do have on average a slightly higher fraction of European ancestry compared to controls
  --- 22.3% versus 19.9%
- Difference can largely be accounted for by a handful of individuals who have a relatively large estimated European ancestry. Removing them
  --- 20.0% versus 19.2%
- Either by excluding potentially misclassified individuals or by using individual ancestry estimates as covariate (Pritchard et al AJHG 2000), the impact on the association results is very modest

Cohorts (n)	Frequency of HapK		RR	*P*-value
	Patients	Controls		
African Americans				
Philadelphia				
All MI sre (105/127)	0.103	0.017	6.5	0.000067
All MI admix adj			6.34	0.0001
Cleveland				
All MI sre (53/111)	0.122	0.072	1.78	0.11
All MI admix adj			1.75	0.11
MI and additional CVD sre (13/111)	0.152	0.072	2.31	0.14
MI and additional CVD admix adj			2.27	0.16
Atlanta				
All MI sre (39/149)	0.075	0.015	5.21	0.018
All MI admix adj			5.08	0.019
MI and additional CVD sre (8/149)	0.202	0.015	16.36	0.0039
MI and additional CVD admix adj			16.67	0.0035

sre: self reported

admix adj: admixture adjustment using estimated European ancestry as covariate

Combining results from the three American cities

Ethnic groups (n)	Frequency of HapK		RR (95% CI)	*P*-value	PAR
	Patients	Controls			
European Americans					
All MI (1591/1775)	0.171	0.148	1.19 (1.04, 1.36)	0.006	
All MI coh adj, admix adj			1.16 (1.01, 1.34)	0.017	0.05
MI and additional CVD (183/1345) [b]	0.192	0.15	1.35 (1.00, 1.81)	0.026	
MI and additional CVD coh adj, admix adj			1.32 (0.98,1.78)	0.035	0.09
African Americans					
All MI (197/387)	0.105	0.032	3.52 (1.96, 6.29)	1.2×10^{-5}	
All MI coh adj, admix adj			3.50 (1.90, 6.43)	2.9×10^{-5}	0.14
MI and additional CVD (21/260) [b]	0.176	0.041	4.94 (1.58, 15.43)	0.003	
MI and additional CVD coh adj, admix adj			4.17 (1.21, 14.30)	0.012	0.22

[b] Cleveland and Atlanta cohorts only; information from Philadelphia not available

Note that for All MI, the RR confidence intervals for the European Americans and African Americans do not overlap ($P < 0.001$)

Summary

- A variant/haplotype apparently European in origin confers much higher risk of MI in African Americans than in European Americans
- An Example of gene-gene(s) interaction?

  --- requires further investigations
- Ethnicity can sometimes be a useful, but imperfect, surrogate for certain genetic variants or combination of genetic variants.

A variant of the gene encoding Leukotriene A4 Hydrolase confers ethnic specific risk of myocardial infarction

Poster # 962

Anna Helgadottir[1], Andrei Manolescu[1], Agnar Helgason[1], Gudmar Thorleifsson[1], Unnur Thorsteinsdottir[1], Daniel F. Gudbjartsson[1], Solveig Gretarsdottir[1], Kristinn P. Magnusson[1], Gudmundur Gudmundsson[1], Andrew Hicks[1], Thorlakur Jonsson[1], Struan F. A. Grant[1], Jesus Sainz[1], Stephen J. O'Brien[2], Sigurlaug Sveinbjornsdottir[3], Einar M. Valdimarsson[3], Stefan E. Matthiasson[3], Allan I. Levey[4], Gudmundur Thorgeirsson[3], Jerome L. Abramson[4], Murdach Reilly[5], Viola Vaccarino[4], Megan Wolfe[5], Vilmundur Gudnason[6], Arshed A. Quyyumi[4], Eric J. Topol[7], Daniel J. Rader[5], Jeffrey R. Gulcher[1], Hakon Hakonarson[1],Augustine Kong[1], Kari Stefansson[1].

[1]deCODE genetics Inc, Reykjavik, Iceland
[2]Laboratory of Genomic Diversity, National Cancer Institute.
[3]National University Hospital, Reykjavik, Iceland
[4]Emory University School of Medicine, Atlanta, GA,USA
[5]University of Pennsylvania School of Medicine, Philadelphia, PA, USA
[6]Icelandic Heart Association, Reykjavik, Iceland
[7]Cleveland Clinic Foundation, Cleveland, OH, USA