BHF-UCL GO annotation team

GO Annotators

Ruth Lovering

Varsha Khodiyar

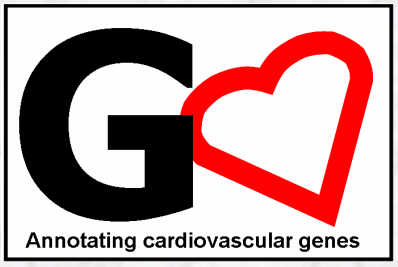
Grant Holders

Philippa Talmud

Rolf Apweiler

Pete Scambler

Mike Hubank



5 year financial support, 2007-2012

www.cardiovasculargeneontology.com

Prioritising genes

Annotating specific genes identified through GWAS or microarray

Genome wide association studies

Microarray

Genome wide association studies

- Genome wide association studies
  - Small list of genes
  - Some previously not known to be involved in disease/trait
  - Researchers require a short summary of gene/protein, ideally with a link to the disease/trait
  - Full annotation too!

Am J Hum Genet. 2009 Nov;85(5):628-42.

Gene-centric association signals for lipids and apolipoproteins identified via the HumanCVD BeadChip.

Talmud PJ, Drenos F, Shah S, Shah T, Palmen J, Verzilli C, Gaunt TR, Pallas J, Lovering R, Li K, Casas JP, Sofat R, Kumari M, Rodriguez S, Johnson T, Newhouse SJ, Dominiczak A, Samani NJ, Caulfield M, Sever P, Stanton A, Shields DC, Padmanabhan S, Melander O, Hastie C, Delles C, Ebrahim S, Marmot MG, Smith GD, Lawlor DA, Munroe PB, Day IN, Kivimaki M, Whittaker J, Humphries SE, Hingorani AD; ASCOT investigators; NORDIL investigators; BRIGHT Consortium.

Collaborators (14)

Centre for Cardiovascular Genetics, Department of Medicine, University College London, London WC1E 6JF, UK.

Abstract

Blood lipids are important cardiovascular disease (CVD) risk factors with both genetic and environmental determinants. The Whitehall II study (n=5592) was genotyped with the gene-centric HumanCVD BeadChip (Illumina). We identified 195 SNPs in 16 genes/regions associated with 3 major lipid fractions and 2 apolipoprotein components at p<10(-5), with the associations being broadly concordant with prior genome-wide analysis. SNPs associated with LDL cholesterol and apolipoprotein B were located in LDLR, PCSK9, APOB, CELSR2, HMGCR, CETP, the TOMM40-APOE-C1-C2-C4 cluster, and the APOA5-A4-C3-A1 cluster; SNPs associated with HDL cholesterol and apolipoprotein AI were in CETP, LPL, LIPC, APOA5-A4-C3-A1, and ABCA1; and SNPs associated with triglycerides in GCKR, BAZ1B, MLXIPL, LPL, and APOA5-A4-C3-A1. For 48 SNPs in previously unreported loci that were significant at p<10(-4) in Whitehall II, in silico analysis including the British Women's Heart and Health Study, BRIGHT, ASCOT, and NORDIL studies (total n>12,500) revealed previously unreported associations of SH2B3 (p<2.2x10(-6)), BMPR2 (p<2.3x10(-7)), BCL3/PVRL2 (flanking APOE; p<4.4x10(-8)), and SMARCA4 (flanking LDLR; p<2.5x10(-7)) with LDL cholesterol. Common alleles in these genes explained 6.1%-14.7% of the variance in the five lipid-related traits, and individuals at opposite tails of the additive allele score exhibited substantial differences in trait levels (e.g., >1 mmol/L in LDL cholesterol [approximately 1 SD of the trait distribution]). These data suggest that multiple common alleles of small effect can make important contributions to individual differences in blood lipids potentially relevant to the assessment of CVD risk. These genes provide further insights into lipid metabolism and the likely effects of modifying the encoded targets therapeutically.

and ABCA1; and SNPs associated with triglycerides in GCKR, BAZ1B, MLXIPL, LPL, and APOA5-A4-C3-A1. For 48 SNPs i previously unreported loci that were significant at p<10(-4) in Whitehall II, in silico analysis including the British Women's Heart and Health Study, BRIGHT, ASCOT, and NORDIL studies (total n>12,500) revealed previously unreported association of SH2B3 (p<2.2x10(-6)), BMPR2 (p<2.3x10(-7)), BCL3/PVRL2 (flanking APOE; p<4.4x10(-8)), and SMARCA4 (flanking LDLR; p<2.5x10(-7)) with LDL cholesterol. Common alleles in these genes explained 6.1%-14.7% of the variance in the five

GCKR
BAZ1B
MLXIPL
SH2B3
BMPR2
BCL3/PVRL2
SMARCA4

- **Provide short gene summary**
- **Link with cardiovascular traits**
  - **cholesterol/triglyceride/CRP**
- **Tissue/cell type expression**
- **Is SNP likely to be causative?**

Evelyn Camon, PSMA6

- SNP rs1048990
- 5’ UTR of PSMA6
- Trait association
  - plasma CRP (C reactive protein)
  - plasma HDL (high density lipoprotein)

Evelyn Camon, PSMA6

- SNP rs1048990
- 5' UTR of PSMA6
  - Transcription factor binding site
    - *C/EBPalpha*
  - Proteosome subunit alpha type 6
- Trait association
  - plasma CRP (C reactive protein)
  - plasma HDL (high density lipoprotein)

PSMA6 - Trait association

- Traits
  - plasma CRP (C reactive protein)
  - plasma HDL (high density lipoprotein)
- Proteosome subunit
  - Activates NF-kappaB (transcription factor)
  - NF-kappaB regulates expression
    - C reactive protein (CRP)
    - Endothelial lipase (EL)
  - EL regulates HDL levels

Future experiments

- Is SNP rs1048990 associated with PSMA6 expression?

GWAS directing gene focus

- 139 cardiovascular trait associated SNPs
  - A genome-wide association study in Europeans and South Asians identifies five new loci for coronary artery disease.
  - Genome-wide association identifies a susceptibility locus for coronary artery disease in the Chinese Han population.
  - Large-scale association analysis identifies 13 new susceptibility loci for coronary artery disease.
  - Biological, clinical and population relevance of 95 loci for blood lipids.
- BHF-Priority
  - 28 priority targets year 1
  - 50 on our 4000 CV associated gene list
  - 61 new targets

Microarray directing gene focus

- Innate immunity
- Derek Gilroy, Melanie Staples, Evelyn Camon
  - Resolving macrophage
  - 350 differentially expressed genes
    - $p<0.05$; fold change>1.5
  - 37 differentially expressed genes
    - $p<0.01$; fold change>2

- Fully annotated 37 genes in 3 months
  - Improved interpretation of dataset
  - Confirmed resolving macrophage hypothesis

Improved microarray analysis

Analysis by Rachael Huntley, GOA, using Ontologizer

		April 2011			December 2010		
GO ID	GO term	p-value (Adj)	Study Count	Population Count	p-value (Adj)	Study Count	Population Count
GO:0001775	cell activation	0	25	390	0.01	19	363
GO:0002376	immune system process	0	39	885	0.016	31	833
GO:0008283	cell proliferation	0	38	862	0.236	27	800
GO:0001816	cytokine production	0.002	16	220	1	8	205
GO:0042221	response to chemical stimulus	0.028	55	1538	1	37	1356
GO:0006928	cellular component movement	0.038	24	531	1	12	483
GO:0051674	localization of cell	0.066	23	507	1	11	457
GO:0032502	developmental process	0.068	81	3006	0.982	65	2832
GO:0071887	leukocyte apoptosis	0.092	7	41	1	2	24

Grey area not significant

Enriched for cell proliferation
cytokine production
response to chemical stimulus
leukocyte apoptosis

Improved PAHSS analysis

Analysis by Rachael Huntley, GOA, using Ontologizer

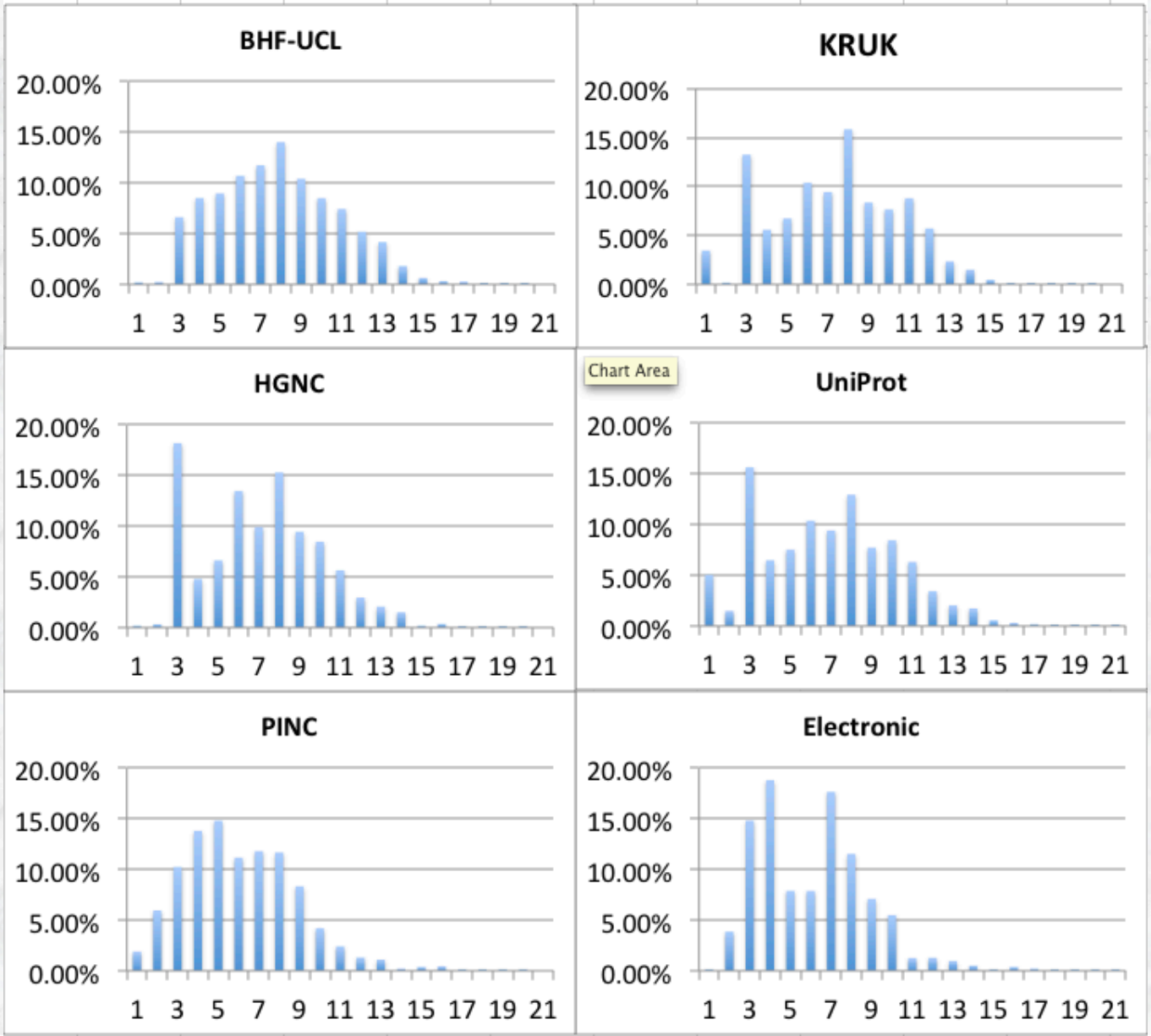
		With BHF-UCL data			Without BHF-UCL data		
GO ID	GO name	p-value (Adj)	Study count	Pop count	p-value (Adj)	Study count	Pop count
GO:0002376	immune system process	0	77	1487	0	71	1406
GO:0016265	death	0	53	1431	0	52	1354
GO:0023052	signaling	0	107	4017	0	106	3898
GO:0001775	cell activation	0.002	32	632	0	27	575
GO:0006950	response to stress	0.002	92	2552	0.004	84	2448
GO:0008283	cell proliferation	0.002	42	1205	0.006	37	1091
GO:0022610	biological adhesion	0.01	30	827	0.186	26	782
GO:0009605	response to external stimulus	0.014	44	1033	0.05	40	952
GO:0001816	cytokine production	0.014	18	289	0.458	13	227
GO:0051674	localization of cell	0.016	29	606	0.37	23	515
GO:0051179	localization	0.038	87	3669	0.078	82	3482
GO:0042221	response to chemical stimulus	0.076	74	2177	0.364	67	2048
GO:0046209	nitric oxide metabolic process	0.094	6	51	1	3	40
GO:0003013	circulatory system process	0.318	12	263	0.068	12	226

Grey area not significant

Systemic scleroderma-related pulmonary arterial hypertension (PAHSS)

Enriched for nitric oxide metabolic process
cytokine production
biological adhesion

Impact of process focused annotation



Impact of process focused annotation

Source	Median	Inter quartile range
BHF_UCL	8.0	6-10
KRUK	8.0	5-10
HGNC	7.0	5-9
UniProt	7.0	4-9
PINC	6.0	4-8
Electronic	6.0	4-8
P value for difference in distribution between 6 groups*		P<0.0001

Impact of process focused annotation

Curators request more GO terms

Curators annotate to more specific GO terms

Curators have time to understand the biology of the field they are annotating