

Comprehensive Multiple eQTL Detection and its Application to GWAS Interpretation

Biao Zeng*, Luke R. Lloyd-Jones[†], Grant W. Montgomery[†], Andres Metspalu[‡], Tõnu Esko[‡], Lude Franke[#], Urmo Vosa[#], Annique Claringbould[#], Kenneth L. Brigham[§], Arshed A. Quyyumi^{§§}, Youssef Idaghdour[^], Jian Yang[†], Peter M. Visscher[†], Joseph E. Powell^{†,+}, and Greg Gibson*¹

Supplementary Information

Table S1. Summary of Illumina probe re-annotation

Detected	# of Mismatches	Probes
HT12v3		48,803
Exome	0	29,452
	1	990
	2	316
Genome	0	13,287
	1	1,486
	2	400
Merge		45,931

The number of probes with 0, 1, or 2 mismatches relative to the reference human genome

CAGE and FHS eQTL Summary Results can be downloaded or searched at:

Excel Table at <https://ggibsongt.wixsite.com/gibsongatech/supplementary-data>

Shiny site at <http://bloodqtlshiny.biosci.gatech.edu:3535/>

Table S2 Citations for GWAS studies used to support eQTL-GWAS colocalization

Height and Weight

Locke, A.E., B. Kahali, S.I. Berndt, A.E. Justice, T.H. Pers, et al., 2015 Genetic studies of body mass index yield new insights for obesity biology. *Nature* 518, 197-206.

Shungin, D., T.W. Winkler, D.C. Croteau-Chonka, T. Ferreira, A.E. Locke, et al., 2015 New genetic loci link adipose and insulin biology to body fat distribution. *Nature* 518, 187-196.

Berndt, S.I., S. Gustafsson, R. Mägi, A. Ganna, E. Wheeler, et al., 2013 Genome-wide meta-analysis identifies 11 new loci for anthropometric traits and provides insights into genetic architecture. *Nat Genet* 45, 501-512.

Wood, A.R., T. Esko, J. Yang, S. Vedantam, T.H. Pers, et al., 2014 Defining the role of common variation in the genomic and biological architecture of adult human height. *Nat Genet* 11, 1173-1186.

Child Growth and Development

van der Valk, R.J., E. Kreiner-Møller, M.N. Kooijman, M. Guxens, E. Stergiakouli, et al., 2015 A novel common variant in DCST2 is associated with length in early life and height in adulthood. *Hum Mol Genet* 24, 1155-1168.

Horikoshi, M., H. Yaghootkar, D.O. Mook-Kanamori, U. Sovio, H.R. Taal, et al., 2013 New loci associated with birth weight identify genetic links between intrauterine growth and adult height and metabolism. *Nat Genet* 45, 76-82.

Horikoshi, M., R.N. Beaumont, F.R. Day, N.M. Warrington, M.N. Kooijman, et al., 2016 Genome-wide associations for birth weight and correlations with adult disease. *Nature* 538, 248-252.

Bradfield, J.P., H.R. Taal, N.J. Timpson, A. Scherag, C. Lecoeur, et al., 2012 A genome-wide association meta-analysis identifies new childhood obesity loci. *Nat Genet* 44, 526-531.

Cousminer, D.L., D.J. Berry, N.J. Timpson, W. Ang, E. Thiering, et al., 2013. Genome-wide association and longitudinal analyses reveal genetic loci linking pubertal height growth, pubertal timing, and childhood adiposity. *Hum Molec Genet.* 22, 2735-2747.

Perry, J.R., F. Day, C.E. Elks, P. Sulem, D.J. Thompson, et al., 2014 Parent-of-origin-specific allelic associations among 106 genomic loci for age at menarche. *Nature* 514, 92-97.

Day, F.R., K.S. Ruth, D.J. Thompson, K.L. Lunetta, N. Pervjakova, et al., 2015 Large-scale genomic analyses link reproductive aging to hypothalamic signaling, breast cancer susceptibility and BRCA1-mediated DNA repair. *Nat Genet* 47, 1294-1303.

Taal, H.R., B. St Pourcain, E. Thiering, S. Das, D.O. Mook-Kanamori, et al., 2012 Common variants at 12q15 and 12q24 are associated with infant head circumference. *Nat Genet* 15, 532-538.

Chronic Kidney Disease, Serum Creatinine and Urate

Pattaro, C., A. Teumer, M. Gorski, A.Y. Chu, M. Li, *et al.*, 2015 Genetic associations at 53 loci highlight cell types and biologic pathways for kidney function. *Nat Comm* 7, 10023

Köttgen, A., E. Albrecht, A. Teumer, V. Vitart, J. Krumsiek, *et al.*, 2013 Genome-wide association analyses identify 18 new loci associated with serum urate concentrations. *Nat Genet* 45, 145-154.

Cognitive Performance and Personality

Okbay, A., J.P. Beauchamp, M.A. Fontana, J.J. Lee, T.H. Pers, *et al.*, 2016 Genome-wide association study identifies 74 loci associated with educational attainment. *Nat Genet* 533, 539-542.

Benyamin, B., B. Pourcain, O.S. Davis, G. Davies, N.K. Hansell, *et al.*, 2014 Childhood intelligence is heritable, highly polygenic and associated with FNBP1L. *Mol. Psychiatry* 19, 253-258.

Rietveld, C.A., S.E. Medland, J. Derringer, J. Yang, T. Esko, *et al.*, 2013 GWAS of 126,559 individuals identifies genetic variants associated with educational attainment. *Science* 314, 1467-1471.

Rietveld, C.A., T. Esko, G. Davies, T.H. Pers, P. Turley, 2015 Common genetic variants associated with cognitive performance identified using the proxy-phenotype method. *Proc Natl Acad Sci USA* 111,13790-13794.

de Moor, M.H., P.T. Costa, A. Terracciano, R.F. Krueger, E.J. de Geus, *et al.*, 2012 Meta-analysis of genome-wide association studies for personality. *Mol Psychiatry* 17, 337-349.

Genetics of Personality Consortium, 2015 Meta-analysis of genome-wide association studies for neuroticism, and the polygenic association with major depressive disorder. *JAMA Psych* 72, 642-650.

van den Berg, S.M., M.H. de Moor, M. McGue, E. Pettersson, A. Terracciano, *et al.*, 2014 Harmonization of neuroticism and extraversion phenotypes across inventories and cohorts in the Genetics of Personality Consortium: an application of Item Response Theory. *Behav Genet* 44, 295-313.

Coronary Artery Disease

Peden, J., J.C. Hopewell, D. Saleheen, J.C. Chambers, J. Hager, *et al.*, 2011 A genome-wide association study in Europeans and South Asians identifies five new loci for coronary artery disease. *Nat Genet* 43, 339-344.

Schunkert, H., I.R. König, S. Kathiresan, M.P. Reilly, T.L. Assimes, *et al.*, 2011 Large-scale association analysis identifies 13 new susceptibility loci for coronary artery disease. *Nat Genet* 43, 333-338.

CARDIoGRAMplusC4D Consortium, 2015 A comprehensive 1000 Genomes-based genome-wide association meta-analysis of coronary artery disease. *Nat Genet* 47, 1121-1130.

Immune and Inflammatory Disease

Liu, J.Z., S. van Sommeren, H. Huang, S.C. Ng, R. Alberts, R., *et al.*, 2015 Association analyses identify 38 susceptibility loci for inflammatory bowel disease and highlight shared genetic risk across populations. *Nat Genet* 47, 979-989.

Anderson, C.A., G. Boucher, C.W. Lees, A. Franke, M. D'Amato, *et al.*, 2011 Meta-analysis identifies 29 additional ulcerative colitis risk loci, increasing the number of confirmed associations to 47. *Nat Genet* 43, 246-252.

Cordell, H.J., Y. Han, G.F. Mells, Y. Li, G.M. Hirschfield, *et al.*, 2015 International genome-wide meta-analysis identifies new primary biliary cirrhosis risk loci and targetable pathogenic pathways. *Nat Commun* 6, 8019.

Liu, J.Z., M.A. Almarri, D.J. Gaffney, G.F. Mells, L. Jostins, *et al.*, 2012 Dense fine-mapping study identifies new susceptibility loci for primary biliary cirrhosis. *Nat Genet* 44, 1137-1141.

Faraco, J., L. Lin, B.R. Kornum, E.E. Kenny, G. Trynka, *et al.*, 2013 ImmunoChip study implicates antigen presentation to T cells in narcolepsy. *PLoS Genet* 9, e1003270.

Tsoi, L.C., S.L. Spain, J. Knight, E. Ellinghaus, P.E. Stuart, *et al.*, 2012 Identification of 15 new psoriasis susceptibility loci highlights the role of innate immunity. *Nat Genet*, 44, 1341-1348.

International Multiple Sclerosis Genetics Consortium, Wellcome Trust Case Control Consortium 2, 2011 Genetic risk and a primary role for cell-mediated immune mechanisms in multiple sclerosis. *Nature* 476, 214-219.

Onengut-Gumuscu, S., W.-M. Chen, O. Burren, N.J. Cooper, A.R. Quinlan, *et al.*, 2015 Fine mapping of type 1 diabetes susceptibility loci and evidence for colocalization of causal variants with lymphoid gene enhancers. *Nat Genet* 47, 381-386.

Roederer, M., L. Quaye, M. Mangino, M.H. Beddall, Y. Mahnke, *et al.*, 2015 The Genetic Architecture of the Human Immune System: A Bioresource for Autoimmunity and Disease Pathogenesis. *Cell* 161, 387-403.

Orrù, V., M. Steri, G. Sole, C. Sidore, F. Virdis, *et al.*, 2013 Genetic variants regulating immune cell levels in health and disease. *Cell* 155, 242-256.

Bentham, J., D.L. Morris, D.S. Cunningham Graham, C.L. Pinder, P. Tombleson, *et al.*, 2015 Genetic association analyses implicate aberrant regulation of innate and adaptive immunity genes in the pathogenesis of systemic lupus erythematosus. *Nat Genet* 47, 1457-1464.

Trynka, G., K.A. Hunt, N.A. Bockett, J. Romanos, V. Mistry, *et al.*, 2011 Dense genotyping identifies and localizes multiple common and rare variant association signals in celiac disease. *Nat Genet* 43, 1193-1201.

Dubois, P.C.A., G. Trynka, L. Franke, K.A. Hunt, J. Romanos, *et al.*, 2010 Multiple common variants for celiac disease influencing immune gene expression. *Nat Genet* 42, 295-302.

Okada, Y., D. Wu, G. Trynka, T. Raj, C. Terao, *et al.*, 2014 Genetics of rheumatoid arthritis contributes to biology and drug discovery. *Nature* 506, 376-381.

Stahl, E.A., S. Raychaudhuri, E.F. Remmers, G. Xie, S. Eyre, et al., 2010 Genome-wide association study meta-analysis identifies seven new rheumatoid arthritis risk loci. *Nat Genet* 42, 508-514.

Hinks, A., J. Cobb, M.C. Marion, S. Prahalad, M. Sudman, et al., 2013 Dense genotyping of immune-related disease regions identifies 14 new susceptibility loci for juvenile idiopathic arthritis. *Nat Genet* 45, 664-669.

Paternoster, L., M. Standl, J. Waage, H. Baurecht, M. Hotze, et al., 2015 Multi-ancestry genome-wide association study of 21,000 cases and 95,000 controls identifies new risk loci for atopic dermatitis. *Nat Genet* 47, 1449-1456.

Psychiatric

Hyde, C.L., M.W. Nagle, C. Tian, X. Chen, S.A. Paciga, et al., 2016 Identification of 15 genetic loci associated with risk of major depression in individuals of European descent. *Nat Genet* 48, 1031-1036.

Cross-Disorder Group of the Psychiatric Genomics Consortium, 2013 Identification of risk loci with shared effects on five major psychiatric disorders: a genome-wide analysis. *Lancet* 381, 1371-1379.

Major Depressive Disorder Working Group of the PGC, 2013 A mega-analysis of genome-wide association studies for major depressive disorder. *Mol Psychiatry* 18, 497-511.

Ruderfer, D.M., A.H. Fanous, S. Ripke, A. McQuillin, R.L. Amdur, et al., 2013 Polygenic dissection of diagnosis and clinical dimensions of bipolar disorder and schizophrenia. *Mol Psychiatry* 19, 1017-1024.

Ripke, S., A.R. Sanders, K.S. Kendler, D.F. Levinson, P. Sklar, et al., 2011 Genome-wide association study identifies five new schizophrenia loci. *Nat Genet* 43, 969-976.

Psychiatric GWAS Consortium Bipolar Disorder Working Group, 2011 Large-scale genome-wide association analysis of bipolar disorder identifies a new susceptibility locus near ODZ4. *Nat Genet* 43, 977-983.

Type 2 Diabetes and related traits from the MAGIC Consortium

Walford, G.A., S. Gustafsson, D. Rybin, A. Stančáková, H. Chen, et al., 2016 Genome-wide association study of the modified Stumvoll Insulin Sensitivity Index identifies BCL2 and FAM19A2 as novel insulin sensitivity loci. *Diabetes* 65, 3200-3211.

Gaulton, K.J., T. Ferreira, Y. Lee, A. Raimondo, R. Mägi, et al., 2015 Genetic fine mapping and genomic annotation defines causal mechanisms at type 2 diabetes susceptibility loci. *Nat Genet* 47, 1415-1425.

DIAGRAM Consortium, AGEN-T Consortium, SAT2D Consortium, MAT2D Consortium, T2D-GENES Consortium, et al., 2014 Genome-wide trans-ancestry meta-analysis provides insight into the genetic architecture of type 2 diabetes susceptibility. *Nat Genet* 46, 234-244.

Prokopenko, I., W. Poon, R. Mägi, B. Prasad, S.A. Salehi, et al. 2014 A central role for GRB10 in regulation of islet function in man. *PLoS Genet* 10, e1004235.

Scott, R.A., V. Lagou, R.P. Welch, E. Wheeler, M.E. Montasser, *et al.*, 2012 Large-scale association analyses identify new loci influencing glycemic traits and provide insight into the underlying biological pathways. *Nat Genet* 44, 991-1005.

Morris, A.P., B.F. Voight, T.M. Teslovich, T. Ferreira, A.V. Segrè, *et al.*, 2012 Large-scale association analysis provides insights into the genetic architecture and pathophysiology of type 2 diabetes. *Nat Genet* 44, 981-990.

Manning, A.K., M.F. Hivert, R.A. Scott, J.L. Grimsby, N. Bouatia-Naji, *et al.*, 2012 A genome-wide approach accounting for body mass index identifies genetic variants influencing fasting glycemic traits and insulin resistance. *Nat Genet* 44, 659-669.

Strawbridge, R.J., J. Dupuis, I. Prokopenko, A. Barker, E. Ahlqvist, *et al.*, 2011 Genome-wide association identifies nine common variants associated with fasting proinsulin levels and provides new insights into the pathophysiology of type 2 diabetes. *Diabetes* 60, 2624-2634.

Saxena, R., M.F. Hivert, C. Langenberg, T. Tanaka, J.S. Pankow, *et al.*, 2010 Genetic variation in GIPR influences the glucose and insulin responses to an oral glucose challenge. *Nat Genet* 42, 142-148.

Soranzo, N., S. Sanna, E. Wheeler, C. Gieger, D. Radke, *et al.*, 2010 Common variants at 10 genomic loci influence hemoglobin A1C levels via glycemic and nonglycemic pathways. *Diabetes* 59, 3229-3239.

Dupuis, J., C. Langenberg, I. Prokopenko, R. Saxena, N. Soranzo, *et al.*, 2010 New genetic loci implicated in fasting glucose homeostasis and their impact on type 2 diabetes risk. *Nat Genet* 42, 105-116.

Metabolic Traits

Kilpeläinen, T.O., J.F.M. Carli, A.A. Skowronski, Q. Sun, J. Kriebel, *et al.*, 2016 Genome-wide meta-analysis uncovers novel loci influencing circulating leptin levels. *Nat Commun* 7, 10494.

Kettunen, J., A. Demirkhan, P. Würtz, H.H. Draisma, T. Haller, *et al.* 2016 Genome-wide study for circulating metabolites identifies 62 loci and reveals novel systemic effects of LPA. *Nat Commun* 7, 11122.

Teumer, A., A. Tin, R. Sorice, M. Gorski, N.C. Yeo, *et al.*, 2015 Genome-wide association studies identify genetic loci associated with albuminuria in diabetes. *Diabetes* 65, 803-817.

Shin, S.Y., E.B. Fauman, A.K. Petersen, J. Krumsiek, R. Santos *et al.*, 2014 An atlas of genetic influences on human blood metabolites. *Nat Genet* 46, 543-550.

Willer, C.J., E.M. Schmidt, S. Sengupta, G.M. Peloso, S. Gustafsson, *et al.*, 2013 Discovery and refinement of loci associated with lipid levels. *Nat Genet* 45, 1274-1283.

Dastani, Z., M.-F. Hivert, N. Timpson, J.R.B. Perry, X. Yuan, *et al.*, 2012 Novel loci for adiponectin levels and their influence on type 2 diabetes and metabolic traits: a multi-ethnic meta-analysis of 45,891 individuals. *PLoS Genet.* 8, e1002607.

Wu, J.H., R.N. Lemaitre, A. Manichaikul, W. Guan, T. Tanak, *et al.*, 2013 Genome-wide association study identifies novel loci associated with concentrations of four plasma phospholipid fatty acids in the de

novo lipogenesis pathway: results from the Cohorts for Heart and Aging Research in Genomic Epidemiology (CHARGE) consortium. *Circ Cardiovasc Genet* 6, 171-183.

Guan, W., B.T. Steffen, R.N. Lemaitre, J.H.Y. Wu, T. Tanaka, *et al.*, 2014 Genome-wide association study of plasma N6 polyunsaturated fatty acids within the cohorts for heart and aging research in genomic epidemiology consortium. *Circ Cardiovasc Genet* 7, 321-331.

Mozaffarian, D., E.K. Kabagambe, C.O. Johnson, R.N. Lemaitre, A. Manichaikul, *et al.*, Genetic loci associated with circulating phospholipid trans fatty acids: a meta-analysis of genome-wide association studies from the CHARGE Consortium. *Am J Clin Nutr* 101, 398-406.

Lemaitre, R.N., T. Tanaka, W. Tang, A. Manichaikul, M. Foy, *et al.*, 2011 Genetic loci associated with plasma phospholipid n-3 fatty acids: a meta-analysis of genome-wide association studies from the CHARGE Consortium. *PLoS Genetics* 7, e1002193.

Teslovich, T.M., K. Musunuru, A.V. Smith, A.C. Edmondson, I.M. Stylianou, *et al.*, 2010 Biological, clinical and population relevance of 95 loci for blood lipids. *Nature* 466, 707-713.

Other Traits

den Hoed, M., M. Eijgelsheim, T. Esko, B.J. Brundel, D.S. Peal, *et al.*, 2013 Identification of heart rate-associated loci and their effects on cardiac conduction and rhythm disorders. *Nat Genet* 45, 621-631.

Lambert, J.C., C.A. Ibrahim-Verbaas, D. Harold, A.C. Naj, R. Sims, *et al.*, 2013 Meta-analysis of 74,046 individuals identifies 11 new susceptibility loci for Alzheimer's disease. *Nat Genet* 45, 1452-1458.

Chami, N., M.H. Chen, A.J. Slater, J.D. Eicher, E. Evangelou, *et al.*, 2016 Exome genotyping identifies pleiotropic variants associated with red blood cell traits. *Am J Hum Genet* 99, 8-21.

van der Harst, P., W. Zhang, I. Mateo Leach, A. Rendon, N. Verweij, *et al.*, 2012 Seventy-five genetic loci influencing the human red blood cell. *Nature* 492, 369-375.

Gieger, C., A. Radhakrishnan, A. Cvejic, W. Tang, E. Porcu, *et al.*, 2011 New gene functions in megakaryopoiesis and platelet formation. *Nature* 480, 201-208.

van Rheenen, W., A. Shatunov, A.M. Dekker, R.L. McLaughlin, F.P. Diekstra, *et al.*, 2016 Genome-wide association analyses identify new risk variants and the genetic architecture of amyotrophic lateral sclerosis. *Nat Genet* 48, 1043-1048.

Zheng, H., V. Forgetta, Y.-H. Hsu, K. Estrada, A. Rosello-Diez, *et al.*, 2015 Whole-genome sequencing identifies EN1 as a determinant of bone density and fracture. *Nature* 526, 112-117.

Saccone, N.L., *et al.*, Tobacco and Genetics Consortium, 2010 Meta-analyses of genome-wide association studies implicate multiple loci for smoking behavior. *Nat Genet* 42, 441-447.

Moffatt, M.F., I.G. Gut, F. Demenais, D.P. Strachan, E. Bouzigon, *et al.*, 2010 A large-scale, consortium-based genomewide association study of asthma. *N Engl J Med* 363, 1211-1221.

Moffatt, M.F., M. Kabesch, L. Liang, A.L. Dixon, D. Strachan, *et al.*, 2007 Genetic variants regulating ORMDL3 expression contribute to the risk of childhood asthma. *Nature* 448, 470-473.

Supplementary Figures

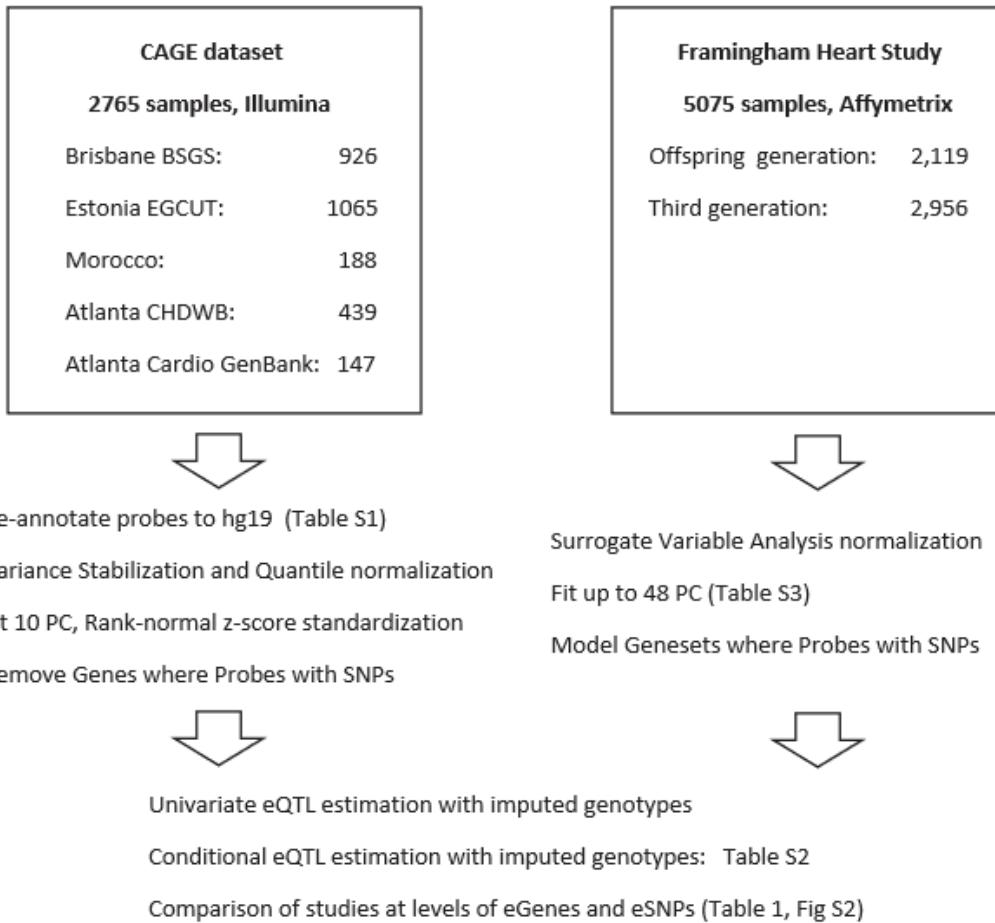


Figure S1. Summary of datasets and analytical workflow, including pointers to location of results.

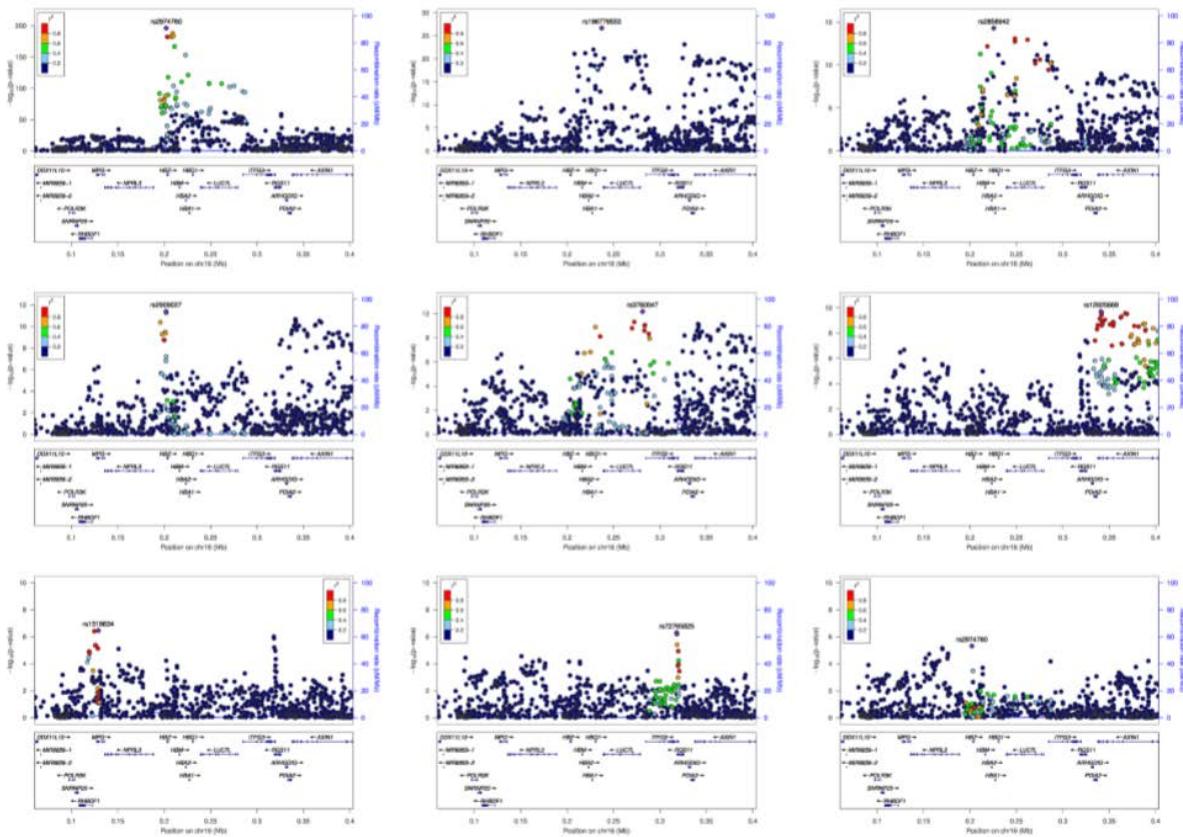


Figure S2. Example of a complex multisite association.

LocusZoom plots show nine sequential conditional association peaks at $p < 10^{-5}$ for HBZ in the CAGE study. Each successive plot shows the profile of univariate association statistics across the locus, conditional on adding all of the previously identified peak SNPs as fixed effects in the linear model.

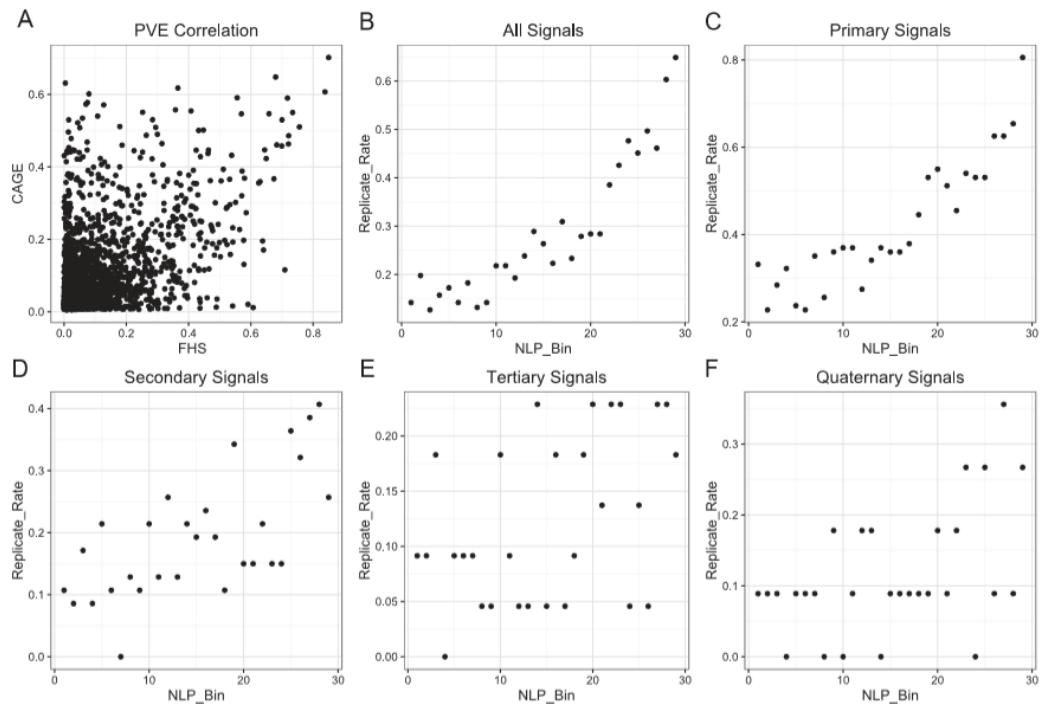


Figure S3. Comparison of Replication Rates in the Framingham and CAGE studies.

(A) Percent variance explained by the sum of effects at each of the discovered independent eQTL at each locus measured in the FHS (x-axis) and CAGE (y-axis) datasets. (B) Replication rate of all FHS eSNPs in the CAGE cohort, expressed as the proportion of FHS eSNPs captured by a SNP with LD $r^2 > 0.8$ in CAGE, in bins of NLP units. (C-F) Replication rates for 5,597 primary, 2,098 secondary, 656 tertiary and 337 quaternary signals.