

## Large-Scale Gene-Centric Meta-analysis across 32 Studies Identifies Multiple Lipid Loci

Folkert W. Asselbergs,<sup>1,2,3,119</sup> Yiran Guo,<sup>4,5,119</sup> Erik P.A. van Iperen,<sup>6,7,119</sup> Suthesh Sivapalaratnam,<sup>8,119</sup> Vinicius Tragante,<sup>1,3,119</sup> Matthew B. Lanktree,<sup>9</sup> Leslie A. Lange,<sup>10</sup> Berta Almqvister,<sup>4</sup> Yolande E. Appelman,<sup>11</sup> John Barnard,<sup>12</sup> Jens Baumert,<sup>13</sup> Amber L. Beitelshoes,<sup>14</sup> Tushar R. Bhangale,<sup>15</sup> Yii-Der Ida Chen,<sup>16</sup> Tom R. Gaunt,<sup>17</sup> Yan Gong,<sup>18</sup> Jemma C. Hopewell,<sup>19</sup> Toby Johnson,<sup>20</sup> Marcus E. Kleber,<sup>21,22</sup> Taimour Y. Langaee,<sup>18</sup> Mingyao Li,<sup>23</sup> Yun R. Li,<sup>4</sup> Kiang Liu,<sup>24</sup> Caitrin W. McDonough,<sup>18</sup> Matthijs F.L. Meijs,<sup>1</sup> Rita P.S. Middelberg,<sup>25,26</sup> Kiran Musunuru,<sup>27,28,29</sup> Christopher P. Nelson,<sup>30</sup> Jeffery R. O'Connell,<sup>31</sup> Sandosh Padmanabhan,<sup>32</sup> James S. Pankow,<sup>33</sup> Nathan Pankratz,<sup>34</sup> Suzanne Rafelt,<sup>30</sup> Ramakrishnan Rajagopalan,<sup>35</sup> Simon P.R. Romaine,<sup>36</sup> Nicholas J. Schork,<sup>37</sup> Jonathan Shaffer,<sup>38</sup> Haiqing Shen,<sup>31</sup> Erin N. Smith,<sup>39</sup> Sam E. Tischfield,<sup>40</sup> Peter J. van der Most,<sup>41</sup> Jana V. van Vliet-Ostaptchouk,<sup>42,43</sup> Niek Verweij,<sup>44,45</sup> Kelly A. Volcik,<sup>46</sup> Li Zhang,<sup>12</sup> Kent R. Bailey,<sup>47</sup> Kristian M. Bailey,<sup>36</sup> Florianne Bauer,<sup>48,2</sup> Jolanda M.A. Boer,<sup>49</sup> Peter S. Braund,<sup>30</sup> Amber Burt,<sup>50</sup> Paul R. Burton,<sup>51</sup> Sarah G. Buxbaum,<sup>52</sup> Wei Chen,<sup>53</sup> Rhonda M. Cooper-DeHoff,<sup>18</sup> L. Adrienne Cupples,<sup>54,55</sup> Jonas S. deJong,<sup>56</sup> Christian Delles,<sup>32</sup> David Duggan,<sup>57</sup> Myriam Fornage,<sup>58</sup> Clement E. Furlong,<sup>35</sup> Nicole Glazer,<sup>54</sup> John G. Gums,<sup>59</sup> Claire Hastie,<sup>32</sup> Michael V. Holmes,<sup>60</sup> Thomas Illig,<sup>61,62</sup> Susan A. Kirkland,<sup>63</sup> Mika Kivimaki,<sup>60</sup> Ronald Klein,<sup>64</sup> Barbara E. Klein,<sup>64</sup> Charles Kooperberg,<sup>65</sup> Kandice Kottke-Marchant,<sup>66</sup> Meena Kumari,<sup>60</sup> Andrea Z. LaCroix,<sup>65</sup> Laya Mallela,<sup>4</sup> Gurunathan Murugesan,<sup>66</sup> Jose Ordovas,<sup>67</sup> Willem H. Ouwehand,<sup>68</sup> Wendy S. Post,<sup>69</sup> Richa Saxena,<sup>70,71</sup> Hubert Scharnagl,<sup>72</sup> Pamela J. Schreiner,<sup>73</sup> Tina Shah,<sup>60</sup> Denis C. Shields,<sup>74</sup> Daichi Shimbo,<sup>38</sup> Sathanur R. Srinivasan,<sup>75</sup> Ronald P. Stolk,<sup>41</sup> Daniel I. Swerdlow,<sup>76</sup> Herman A. Taylor, Jr.,<sup>77</sup> Eric J. Topol,<sup>37</sup> Elina Toskala,<sup>4</sup> Joost L. van Pelt,<sup>78</sup> Jessica van Setten,<sup>3</sup> Salim Yusuf,<sup>79</sup> John C. Whittaker,<sup>80,81</sup> A.H. Zwinderman,<sup>82</sup> LifeLines Cohort Study,<sup>83</sup> Sonia S. Anand,<sup>79</sup> Anthony J. Balmforth,<sup>36</sup> Gerald S. Berenson,<sup>53</sup> Connie R. Bezzina,<sup>56,84</sup> Bernhard O. Boehm,<sup>85</sup> Eric Boerwinkle,<sup>58</sup> Juan P. Casas,<sup>60,86</sup> Mark J. Caulfield,<sup>20</sup> Robert Clarke,<sup>19</sup> John M. Connell,<sup>87</sup> Karen J. Cruickshanks,<sup>64</sup> Karina W. Davidson,<sup>88</sup> Ian N.M. Day,<sup>17</sup> Paul I.W. de Bakker,<sup>2,27,29,89</sup> Pieter A. Doevendans,<sup>1</sup> Anna F. Dominiczak,<sup>32</sup> Alistair S. Hall,<sup>36</sup> Catharina A. Hartman,<sup>90</sup> Christian Hengstenberg,<sup>91</sup> Hans L. Hillege,<sup>44,45</sup> Marten H. Hofker,<sup>92</sup> Steve E. Humphries,<sup>93</sup> Gail P. Jarvik,<sup>35</sup> Julie A. Johnson,<sup>18</sup> Bernhard M. Kaess,<sup>91</sup> Sekar Kathiresan,<sup>27,28,29</sup> Wolfgang Koenig,<sup>94</sup> Debbie A. Lawlor,<sup>17</sup> Winfried März,<sup>22,95</sup> Olle Melander,<sup>96</sup> Braxton D. Mitchell,<sup>31</sup>

<sup>1</sup>Department of Cardiology, Division of Heart and Lungs, University Medical Center Utrecht, 3508 GA Utrecht, The Netherlands; <sup>2</sup>Julius Center for Health Sciences and Primary Care, University Medical Center Utrecht, 3508 GA Utrecht, The Netherlands; <sup>3</sup>Department of Medical Genetics, Biomedical Genetics, University Medical Center Utrecht, 3584 CX Utrecht, The Netherlands; <sup>4</sup>Center for Applied Genomics, Abramson Research Center, The Children's Hospital of Philadelphia, Philadelphia, PA 19104, USA; <sup>5</sup>BGI-Shenzhen, Beishan Industrial Zone, Yantian District, Shenzhen 518083, China; <sup>6</sup>Durrer Center for Cardiogenetic Research, 1105 AZ Amsterdam, The Netherlands; <sup>7</sup>Department of Clinical Epidemiology, Biostatistics and Bioinformatics, Academic Medical Center, University of Amsterdam, 1105 AZ Amsterdam, The Netherlands; <sup>8</sup>Department of Vascular Medicine, Academic Medical Center, 1105 AZ Amsterdam, The Netherlands; <sup>9</sup>Departments of Medicine and Biochemistry, Schulich School of Medicine and Dentistry, University of Western Ontario, London, ON N6A 3K6, Canada; <sup>10</sup>Department of Genetics, University of North Carolina School of Medicine at Chapel Hill, Chapel Hill, NC 27514, USA; <sup>11</sup>Department of Cardiology, VU University Medical Center, 1081 HV Amsterdam, the Netherlands; <sup>12</sup>Department of Quantitative Health Sciences, Lerner Research Institute, Cleveland Clinic, Cleveland, OH 44195, USA; <sup>13</sup>Institute of Epidemiology II, Helmholtz Zentrum München, German Research Center for Environmental Health, D-85764 Neuherberg, Germany; <sup>14</sup>Department of Medicine, University of Maryland School of Medicine, Baltimore, MD 21201, USA; <sup>15</sup>Department of Bioinformatics and Computational Biology, Genentech, South San Francisco, CA 94080, USA; <sup>16</sup>Medical Genetics Institute, Cedars-Sinai Medical Center, Los Angeles, CA 90048, USA; <sup>17</sup>Medical Research Council Centre for Causal Analyses in Translational Epidemiology, School of Social and Community Medicine, University of Bristol, Oakfield House, Oakfield Grove, Bristol BS8 2BN, UK; <sup>18</sup>Department of Pharmacotherapy and Translational Research and Center for Pharmacogenomics, University of Florida, Gainesville, FL 32603, USA; <sup>19</sup>Clinical Trial Service Unit, University of Oxford, Oxford OX3 7BN, UK; <sup>20</sup>Clinical Pharmacology and Barts and the London Genome Centre, William Harvey Research Institute, Barts and the London School of Medicine, Queen Mary University of London, London EC1M 6BQ, UK; <sup>21</sup>LURIC Study, Platz der alten Synagoge 1, 79098 Freiburg, Germany; <sup>22</sup>Mannheim Institute of Public Health, Social and Preventive Medicine, Medical Faculty of Mannheim, University of Heidelberg, Ludolf-Krehl-Strasse 7-11, 68167 Mannheim, Germany; <sup>23</sup>Cardiovascular Institute, the Perelman School of Medicine at the University of Pennsylvania, Philadelphia, PA 19104, USA; <sup>24</sup>Department of Preventive Medicine, Northwestern University Feinberg School of Medicine, Chicago, IL 60611, USA; <sup>25</sup>Genetic Epidemiology, Queensland Institute of Medical Research, Locked Bag 2000, Royal Brisbane Hospital, Queensland 4029, Australia; <sup>26</sup>Department of Medicine, Prince Charles Hospital, Rode Road, Chermiside, Queensland 4032, Australia; <sup>27</sup>Broad Institute of MIT and Harvard, Cambridge, MA 02141, USA; <sup>28</sup>Massachusetts General Hospital, Boston, MA 02114, USA; <sup>29</sup>Harvard Medical School, Boston, MA 02115, USA; <sup>30</sup>Department of Cardiovascular Sciences, University of Leicester, Glenfield Hospital, Leicester LE3 9QP, UK; <sup>31</sup>Division of Endocrinology, Diabetes, and Nutrition, University of Maryland School of Medicine, Baltimore, MD 21201, USA; <sup>32</sup>BHF Glasgow Cardiovascular Research Centre, Institute of Cardiovascular and Medical Sciences, College of Medical, Veterinary, and Life Sciences, University of Glasgow, Glasgow G12 8TA, UK; <sup>33</sup>Division of Epidemiology & Community Health, University of Minnesota, Minneapolis, MN 55455, USA; <sup>34</sup>Department of Laboratory Medicine and Pathology, University of Minnesota, Minneapolis, MN 55455, USA; <sup>35</sup>Division of Medical Genetics,

Grant W. Montgomery,<sup>97</sup> Patricia B. Munroe,<sup>20</sup> Sarah S. Murray,<sup>37</sup> Stephen J. Newhouse,<sup>20</sup> N. Charlotte Onland-Moret,<sup>98,99</sup> Neil Poulter,<sup>100</sup> Bruce Psaty,<sup>101,102</sup> Susan Redline,<sup>103</sup> Stephen S. Rich,<sup>104</sup> Jerome I. Rotter,<sup>16</sup> Heribert Schunkert,<sup>105</sup> Peter Sever,<sup>100</sup> Alan R. Shuldiner,<sup>31,106</sup> Roy L. Silverstein,<sup>107,108</sup> Alice Stanton,<sup>109</sup> Barbara Thorand,<sup>13</sup> Mieke D. Trip,<sup>110</sup> Michael Y. Tsai,<sup>34</sup> Pim van der Harst,<sup>44</sup> Ellen van der Schoot,<sup>111</sup> Yvonne T. van der Schouw,<sup>99</sup> W.M. Monique Verschuren,<sup>49</sup> Hugh Watkins,<sup>112</sup> Arthur A.M. Wilde,<sup>56</sup> Bruce H.R. Wolffenbuttel,<sup>113</sup> John B. Whitfield,<sup>25</sup> G. Kees Hovingh,<sup>8</sup> Christie M. Ballantyne,<sup>114</sup> Cisca Wijmenga,<sup>45</sup> Muredach P. Reilly,<sup>23</sup> Nicholas G. Martin,<sup>25</sup>

Department of Medicine and the Department of Genome Sciences, University of Washington, Seattle, WA 98195, USA; <sup>36</sup>Leeds Institute of Genetics, Health & Therapeutics, University of Leeds, Leeds LS2 9JT, UK; <sup>37</sup>The Scripps Translational Science Institute and The Scripps Research Institute, 3344 N. Torrey Pines Court, Suite 300, La Jolla, CA 92037, USA; <sup>38</sup>Center for Behavioral Cardiovascular Health, Department of Medicine, Columbia University, New York, NY 10032, USA; <sup>39</sup>Department of Pediatrics and Rady's Children's Hospital, University of California at San Diego School of Medicine, La Jolla, CA 92093, USA; <sup>40</sup>Tri-Institutional Training Program in Computational Biology and Medicine, Weill Cornell Medical College, New York, NY 10065, USA; <sup>41</sup>Department of Epidemiology, University Medical Center Groningen, University of Groningen, 9700 RB Groningen, The Netherlands; <sup>42</sup>Molecular Genetics Laboratory, Medical Biology Division, Department of Pathology and Medical Biology, University Medical Center Groningen, Groningen University, 9700 RB Groningen, The Netherlands; <sup>43</sup>Department of Endocrinology, University Medical Center Groningen, Groningen University, 9700 RB Groningen, The Netherlands; <sup>44</sup>Department of Cardiology, University Medical Center Groningen, University of Groningen, Groningen 9700 RB, The Netherlands; <sup>45</sup>Department of Genetics, University Medical Center Groningen, University of Groningen, Groningen 9713 GZ, the Netherlands; <sup>46</sup>Division of Epidemiology, Human Genetics and Environmental Sciences, Human Genetics Center, School of Public Health, University of Texas Health Science Center, Houston, TX 77030, USA; <sup>47</sup>Department of Health Sciences Research, Mayo Clinic, Rochester, MN 55905, USA; <sup>48</sup>Complex Genetics Section, Department of Medical Genetics, University Medical Center Utrecht, 3584 CG Utrecht, The Netherlands; <sup>49</sup>National Institute for Public Health and the Environment, 3721 MA Bilthoven, The Netherlands; <sup>50</sup>Division of Medical Genetics, Department of Medicine, University of Washington, Seattle, WA 98195, USA; <sup>51</sup>Department of Health Sciences, University of Leicester, University Road, Leicester LE1 7RH, UK; <sup>52</sup>Jackson Heart Study and School of Health Sciences, Department of Epidemiology and Biostatistics, Jackson State University, Jackson, MS 39217, USA; <sup>53</sup>Department of Epidemiology, Tulane University, 1440 Canal Street, Suite 1829, New Orleans, LA 70118, USA; <sup>54</sup>Boston University, Boston, MA 02215, USA; <sup>55</sup>The National Heart, Lung, and Blood Institute Framingham Heart Study, Framingham, MA 01702, USA; <sup>56</sup>Heart Failure Research Center, Department of Clinical and Experimental Cardiology, Academic Medical Center, 1105 AZ Amsterdam, the Netherlands; <sup>57</sup>Translational Genomics Research Institute, Phoenix, AZ 85004, USA; <sup>58</sup>The University of Texas Health Science Center at Houston, Houston, TX 77030, USA; <sup>59</sup>Departments of Pharmacotherapy and Translational Research and Community Health and Family Medicine, University of Florida, Gainesville, FL 32607, USA; <sup>60</sup>Research Department of Epidemiology & Public Health, University College London Institute of Epidemiology & Health Care, University College London, 1-19 Torrington Place, London WC1E 6BT, UK; <sup>61</sup>Research Unit of Molecular Epidemiology, Helmholtz Zentrum München, German Research Center for Environmental Health, D-85764 Neuherberg, Germany; <sup>62</sup>Hannover Unified Biobank, Hannover Medical School, 30625 Hannover, Germany; <sup>63</sup>Department of Community Health and Epidemiology and Department of Medicine, Dalhousie University, Halifax NS B3H 1V7, Canada; <sup>64</sup>Department of Ophthalmology & Visual Sciences, University of Wisconsin-Madison, Madison, WI 53706, USA; <sup>65</sup>Division of Public Health Sciences, Fred Hutchinson Cancer Research Center, Seattle, WA 98109, USA; <sup>66</sup>Pathology and Laboratory Medicine Institute, Cleveland Clinic, Cleveland, OH 44195, USA; <sup>67</sup>Jean Mayer USDA Human Nutrition Research Center on Aging, Tufts University, Boston, MA 02111, USA; <sup>68</sup>Department of Haematology, University of Cambridge & Wellcome Trust Sanger Institute, Cambridge CB10 1SA, UK; <sup>69</sup>Division of Cardiology, Department of Medicine, Johns Hopkins University, Baltimore, MD 21205, USA; <sup>70</sup>Center for Human Genetic Research, Massachusetts General Hospital, Boston, MA 02114, USA; <sup>71</sup>Program in Medical and Population Genetics, Broad Institute of MIT and Harvard, Cambridge, MA 02142, USA; <sup>72</sup>Clinical Institute of Medical and Chemical Laboratory Diagnostics, Medical University of Graz, 8036 Graz, Austria; <sup>73</sup>School of Public Health, University of Minnesota, Minneapolis, MN 55455, USA; <sup>74</sup>Conway Institute of Biomolecular & Biomedical Research, University College Dublin, Dublin 4, Ireland; <sup>75</sup>Department of Epidemiology, Tulane University, 1440 Canal Street, Suite 2000, New Orleans, LA 70118 USA; <sup>76</sup>Genetic Epidemiology Group, Research Department of Epidemiology and Public Health, University College London Institute of Epidemiology and Health Care, University College London, London WC1E 6BT, UK; <sup>77</sup>Department of Medicine, University of Mississippi Medical Center, Jackson, MS 39217, USA; <sup>78</sup>Laboratory Center, University Medical Center Groningen, University of Groningen, 9713 GZ Groningen, The Netherlands; <sup>79</sup>Population Health Research Institute, Hamilton Health Sciences, McMaster University, Hamilton, ON L8L 2X2, Canada; <sup>80</sup>GlaxoSmithKline, Stevenage SG1 2NY, UK; <sup>81</sup>Department of Non-communicable Disease Epidemiology, London School of Hygiene and Tropical Medicine, London WC1E 7HT, UK; <sup>82</sup>Department of Clinical Epidemiology, Biostatistics, and Bioinformatics, Academic Medical Center, 1100DE Amsterdam, The Netherlands; <sup>83</sup>LifeLines Cohort Study, University Medical Center Groningen, University of Groningen, 9700 RB Groningen, The Netherlands; <sup>84</sup>Molecular and Experimental Cardiology Group, Academic Medical Center, 1105 AZ Amsterdam, The Netherlands; <sup>85</sup>Department of Internal Medicine, Department of Medicine, University Hospital, 89069 Ulm, Germany; <sup>86</sup>Faculty of Epidemiology and Population Health, London School of Hygiene and Tropical Medicine, London WC1E 7HT, UK; <sup>87</sup>University of Dundee, Ninewells Hospital and Medical School, Dundee DD1 9SY, UK; <sup>88</sup>Departments of Medicine and Psychiatry, Columbia University, New York, NY 10027, USA; <sup>89</sup>Department of Medical Genetics, Biomedical Genetics, University Medical Center Utrecht, 3508 GA Utrecht, The Netherlands; <sup>90</sup>Department of Psychiatry, University Medical Center Groningen, University of Groningen, 9700 AR Groningen, The Netherlands; <sup>91</sup>Klinik und Poliklinik für Innere Medizin II, University of Regensburg, 93042 Germany; <sup>92</sup>Molecular Genetics, Medical Biology Section, Department of Pathology and Medical Biology, University Medical Center Groningen, University of Groningen, 9700 RB Groningen, The Netherlands; <sup>93</sup>Centre for Cardiovascular Genetics, Institute of Cardiovascular Science, Faculty of Population Health Sciences, University College London, 5 University Street, London WC1E 6JF, UK; <sup>94</sup>Department of Internal Medicine II—Cardiology, University of Ulm Medical Centre, 89081 Ulm, Germany; <sup>95</sup>Synlab Academy, Gottlieb-Daimler-Straße 25, 68165 Mannheim, Germany; <sup>96</sup>Clinical Research Center, Malmö University Hospital, Malmö SE-205 02, Sweden; <sup>97</sup>Molecular Epidemiology, Queensland Institute of Medical Research, Locked Bag 2000, Royal Brisbane Hospital, Queensland 4029, Australia; <sup>98</sup>Complex Genetics Section, Department of Medical Genetics, University Medical Center Utrecht, 3508 GA Utrecht, The Netherlands; <sup>99</sup>Julius Center for Health Sciences and Primary Care, University Medical Center Utrecht, 3508 GA Utrecht, The Netherlands; <sup>100</sup>International Centre for Circulatory Health, Imperial College London, London W2 1PG, UK; <sup>101</sup>Cardiovascular Health Research Unit, Departments of Medicine, Epidemiology, and Health Services, University of Washington, Seattle, WA 98195, USA; <sup>102</sup>Group Health Research Institute, Group Health Cooperative, Seattle, WA 98109, USA; <sup>103</sup>University Hospitals, Case Medical Center and Case Western Reserve University, Cleveland, Ohio 44106, USA; <sup>104</sup>Center for Public Health Genomics, University of Virginia, Charlottesville, VA 22908, USA; <sup>105</sup>Medizinische Klinik II, Universität zu Lübeck, 23562 Lübeck, Germany; <sup>106</sup>Geriatric Research and Education Clinical Center, Veterans Administration Medical Center, Baltimore, MD 21201, USA; <sup>107</sup>Department of Cell Biology, Lerner Research Institute, Cleveland Clinic, Case Western Reserve University, Cleveland, OH 44195, USA; <sup>108</sup>Department of Molecular Medicine, Cleveland Clinic Lerner College of Medicine, Case Western Reserve University, Cleveland, OH 44195, USA; <sup>109</sup>Molecular and Cellular Therapeutics, Royal College of Surgeons in Ireland, Dublin 2, Ireland; <sup>110</sup>Department of Cardiology, Academic Medical Center, 1105 AZ Amsterdam, The Netherlands; <sup>111</sup>Department of Experimental Immunology, Sanquin, 1105 AZ Amsterdam, The Netherlands; <sup>112</sup>Department of Cardiovascular Medicine, The Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford OX3 7BN, UK; <sup>113</sup>Department of Endocrinology, University Medical Center Groningen, University of Groningen, 9700 RB Groningen, The Netherlands; <sup>114</sup>Baylor College of Medicine, Houston, TX 77030, USA;

James G. Wilson,<sup>115</sup> Daniel J. Rader,<sup>23</sup> Nilesh J. Samani,<sup>30,116</sup> Alex P. Reiner,<sup>65</sup> Robert A. Hegele,<sup>117</sup> John J.P. Kastelein,<sup>118</sup> Aroon D. Hingorani,<sup>60</sup> Philippa J. Talmud,<sup>93</sup> Hakon Hakonarson,<sup>4</sup> Clara C. Elbers,<sup>2,48,119</sup> Brendan J. Keating,<sup>4,119,\*</sup> and Fotios Drenos<sup>93,119,\*</sup>

Genome-wide association studies (GWASs) have identified many SNPs underlying variations in plasma-lipid levels. We explore whether additional loci associated with plasma-lipid phenotypes, such as high-density lipoprotein cholesterol (HDL-C), low-density lipoprotein cholesterol (LDL-C), total cholesterol (TC), and triglycerides (TGs), can be identified by a dense gene-centric approach. Our meta-analysis of 32 studies in 66,240 individuals of European ancestry was based on the custom ~50,000 SNP genotyping array (the ITMAT-Broad-CARe array) covering ~2,000 candidate genes. SNP-lipid associations were replicated either in a cohort comprising an additional 24,736 samples or within the Global Lipid Genetic Consortium. We identified four, six, ten, and four unreported SNPs in established lipid genes for HDL-C, LDL-C, TC, and TGs, respectively. We also identified several lipid-related SNPs in previously unreported genes: *DGAT2*, *HCAR2*, *GPIHBP1*, *PPARG*, and *FTO* for HDL-C; *SOCS3*, *APOH*, *SPTY2D1*, *BRCA2*, and *VLDLR* for LDL-C; *SOCS3*, *UGT1A1*, *BRCA2*, *UBE3B*, *FCGR2A*, *CHUK*, and *INSIG2* for TC; and *SERPINF2*, *C4B*, *GCK*, *GATA4*, *INSR*, and *LPAL2* for TGs. The proportion of explained phenotypic variance in the subset of studies providing individual-level data was 9.9% for HDL-C, 9.5% for LDL-C, 10.3% for TC, and 8.0% for TGs. This large meta-analysis of lipid phenotypes with the use of a dense gene-centric approach identified multiple SNPs not previously described in established lipid genes and several previously unknown loci. The explained phenotypic variance from this approach was comparable to that from a meta-analysis of GWAS data, suggesting that a focused genotyping approach can further increase the understanding of heritability of plasma lipids.

## Introduction

Cardiovascular disease (CVD) is one of the leading causes of disability and death worldwide.<sup>1</sup> Atherosclerosis is the major underlying pathological process of CVD and is highly prevalent in western societies. Atherogenesis has numerous genetic and environmental risk factors,<sup>2</sup> and abnormalities of plasma lipids and lipoproteins account for ~50% of the population attributable risk of developing CVD.<sup>3,4</sup> Plasma-lipid and lipoprotein levels are themselves highly heritable—estimates range from 40%–60% for total cholesterol (TC), low-density lipoprotein cholesterol (LDL-C), high-density lipoprotein cholesterol (HDL-C), and triglycerides (TGs).<sup>5</sup>

In a large-scale meta-analysis of genome-wide association studies (GWASs), it was shown that plasma-lipid levels are affected by common genetic variants in 95 loci, of which 59 were previously unreported.<sup>6</sup> All together, variation at these loci explains 10%–12% of the total variance and 25%–30% of the genetic variability in plasma-lipid phenotypes.<sup>6</sup> This means that although a portion of the genetic contribution to variation in plasma lipids and lipoproteins has been characterized, there is still variance that remains unattributed.<sup>7</sup>

To further identify genetic associations underlying variation in plasma-lipid phenotypes, we performed a large meta-analysis of 32 studies comprising 66,240 individuals of European ancestry by using the candidate-gene ITMAT-Broad-CARe (IBC) array (Illumina), also known as the CardioChip or the Human CVD BeadArray. The IBC array was designed to capture genetic diversity by using ~50,000 SNPs across ~2,000 candidate-gene regions primarily

related to cardiovascular, inflammatory, and metabolic phenotypes.<sup>8</sup> Prior reports using this array have confirmed previously established associations and identified unreported associations between SNPs and several phenotypes and disease outcomes, including coronary artery disease,<sup>9,10</sup> plasma lipids,<sup>11,12</sup> blood pressure,<sup>13,14</sup> cardiomyopathy,<sup>15</sup> type 2 diabetes (T2D),<sup>16,17</sup> and height.<sup>18</sup> The majority of loci on the IBC array are captured with a marker density equal to or greater than that seen on genome-wide arrays. Compared to the agnostic design of GWAS arrays, gene-centric genotyping with this array might permit a better identification of multiple functional polymorphisms, or their proxies, at each locus. Indeed, this approach has the potential to capture a more detailed genetic architecture in selected high-priority regions and increase the total explained variance.

We sought to contribute to the current literature by using a dense gene-centric approach with the IBC array to identify lipid-trait-associated loci that have not been discovered with more conventional approaches. A flow diagram of the performed analyses is illustrated in Figure 1.

## Material and Methods

### Participating Studies

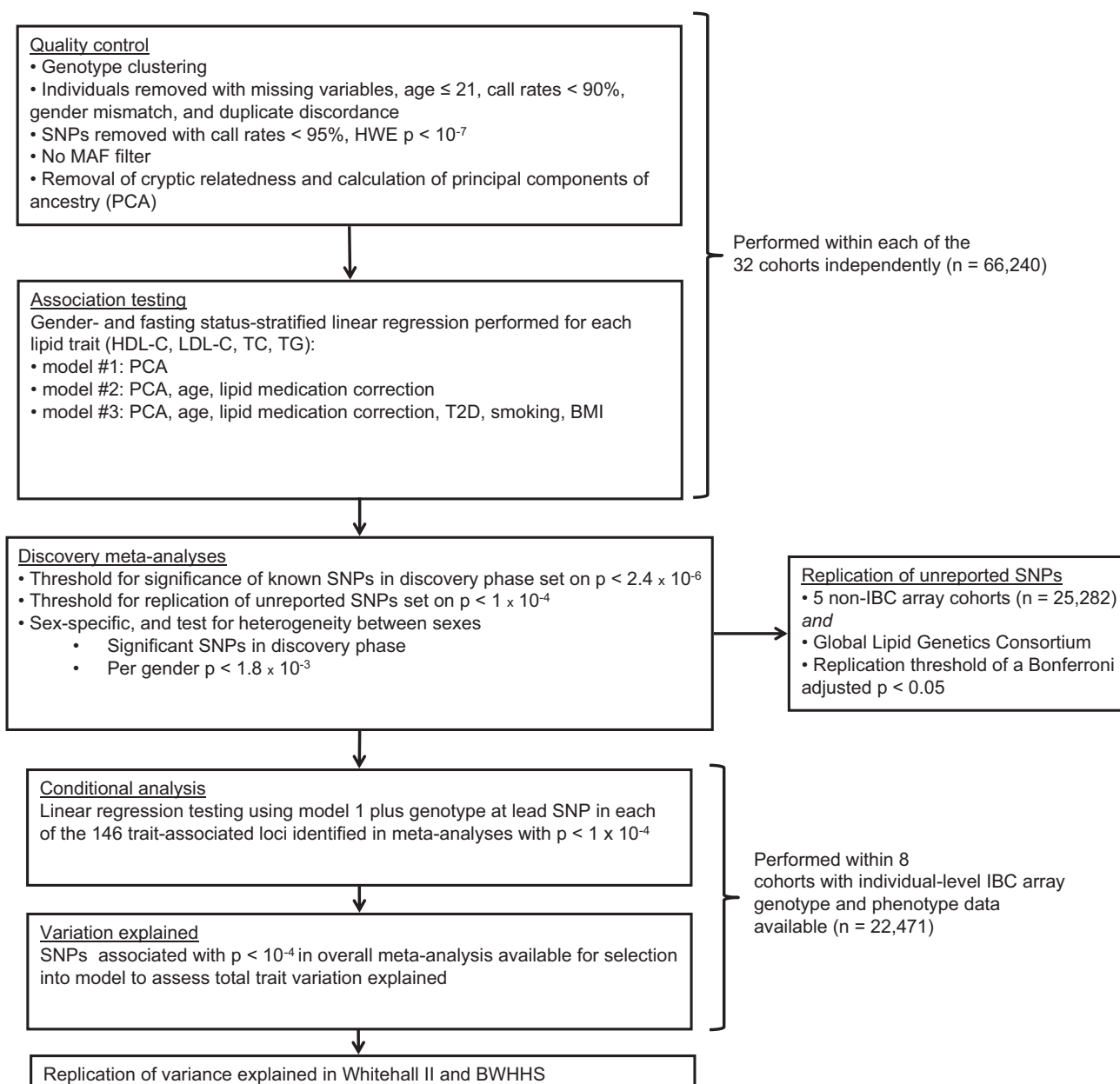
We analyzed individual-level phenotype and genotype data from 22,471 individuals of European descent in seven cohorts, and an additional 25 cohorts contributed summary-level results for 43,769 individuals, yielding a total sample size of 66,240 (Table S1A, available online). Five additional cohorts containing data from a total of 25,282 individuals were used for replication (Table S1B). Further replication was sought through the GWAS

<sup>115</sup>Department of Physiology and Biophysics, University of Mississippi Medical Center, Jackson, MS 39216, USA; <sup>116</sup>National Institute of Health Research Leicester Cardiovascular Biomedical Research Unit, Glenfield Hospital, Groby Road, Leicester LE3 9QP, UK; <sup>117</sup>Robarts Research Institute, University of Western Ontario, London, ON N6A5K8, Canada; <sup>118</sup>Department of Vascular Medicine, Academic Medical Center, 1105 AZ Amsterdam, The Netherlands

<sup>119</sup>These authors contributed equally to this work

\*Correspondence: bkeating@mail.med.upenn.edu (B.J.K.), f.drenos@ucl.ac.uk (F.D.)

http://dx.doi.org/10.1016/j.ajhg.2012.08.032. ©2012 by The American Society of Human Genetics. All rights reserved.



**Figure 1. Summary of the Design Used and the Number of Individuals Involved and p Value Thresholds Used in Each Step**

meta-analysis described by the Global Lipids Genetics Consortium (GLGC).<sup>6</sup> In addition to the genotype data, we obtained data on body mass index (BMI), age, gender, T2D status, smoking history, and, where available, any treatment for dyslipidemia. Informed consent for DNA analysis was received from each respective local institutional and/or national ethical review board.

### Lipid Phenotype Definitions and Correction for the Use of Lipid-Lowering Drugs

Lipid measurements from blood samples collected at baseline or first measurement of each study were used for analysis. We restricted the analyses to those individuals older than 21 years because lipid levels are unstable prior to this age.<sup>19</sup> Lipid samples were categorized as “known fasting,” “nonfasting,” or “undefined.” We converted concentrations from mg/dl to mmol/l by dividing by 38.67 for TC, LDL-C, and HDL-C measurements and

by dividing by 88.57 for TG measurements. With the exception of the PROCARDIS study, where direct LDL-C assay was used (CHOD/PAP assay in an Olympus AU5430<sup>20</sup>), LDL-C concentration was calculated according to Friedewald’s formula (LDL-C = TC – HDL-C – kTGs), where k is 0.45 for mmol/l (or 0.20 if TGs were measured in mg/dl). LDL-C was treated as a missing value if TG values were > 4.51 mmol/l (>400 mg/dl).<sup>21</sup> Prior to analysis, TG levels were transformed with the natural logarithm (ln) for normalizing its distribution.

For individuals receiving lipid-lowering therapy, we multiplied recorded lipid values by a constant: TC was multiplied by 1.271; LDL-C was multiplied by 1.352; HDL-C was multiplied by 0.949, and TGs were multiplied by 1.210 prior to transformation. The multiplicative correction factors were based on analysis of repeatedly measured lipid levels, including levels measured before and after lipid-lowering treatment, in the Whitehall II (WHII)



study<sup>22</sup> as follows. The expected difference between two data-collection time points 5 years apart for each lipid phenotype was estimated among WHII study participants who were not on lipid-lowering therapy. The mean difference between the expected and observed values for those receiving medication at the latter, but not the former, collection phase was calculated and used as the respective correction factor. The correction factors used here are comparable to published estimates for the effects of statins on lipid values from treatment trials.<sup>23</sup>

### Genotyping and Quality Control

Genotyping was performed with the gene-centric IBC array (Illumina HumanCVD).<sup>8</sup> We used genotyping data from the first two versions of the IBC array. Version 1 of the array captures 45,238 SNPs, and version 2 contains an additional 3,989 SNPs, composing a total of 49,227 SNPs. These were clustered into genotypes with Illumina BeadStudio software. Quality-control filters were applied within each cohort at the sample and SNP levels. The filter requirements for the meta-analysis, as sent to each study, required the removal of individuals with a call rate < 90%, gender mismatch, or duplicate discordance. SNPs with call rates < 95% or Hardy-Weinberg equilibrium (HWE) deviation at chi-square  $p < 10^{-7}$  were also removed. To take advantage of the rare variants captured on the array and the large number of samples available, we did not perform any filtering on variants with a low minor allele frequency (MAF) at this stage.

### Evaluation of Cryptic Relatedness

Only founders within cohorts with recorded family structure were included in the analysis; the exceptions were the GRAPHIC, HAPI, and PROCARDIS studies, in which family structure was maintained but adjusted for in the association analysis. To ensure removal of cryptic relatedness and duplicate samples, we estimated pi-hat, a measure of identity by descent (IBD), from the pairwise identity by state (IBS) by using PLINK.<sup>24</sup> PLINK is a computationally efficient open-source analysis toolset for genetic data and is able to perform a series of basic, large-scale analyses. For each set of duplicates or monozygotic twins, and for those with a pairwise pi-hat > 0.3, the sample with the highest genotyping call rate was retained for analysis.

### Evaluation of Population Stratification

For the primary analysis, only individuals of European ancestry were included. Self-reported ethnicity was verified by multidimensional-scaling analysis of IBS distances as implemented in PLINK; HapMap panels were used as reference standards. After SNPs in pairwise linkage disequilibrium (LD) ( $r^2 > 0.3$ ) were removed, EIGENSTRAT software was used for computing principal components on the subset of nonexcluded individuals for use as covariates in the regression analyses. EIGENSTRAT, part of the EIGENSOFT package, corrects for variation in the frequency across ancestral populations through the use of principal components, minimizing potential false-positive signals due to population structure while increasing its power to detect true associations.<sup>25,26</sup> Analysis in participants of non-European ancestry is reported in an accompanying paper (C.C.E., unpublished data).

### Thresholds for Declaration of Statistical Significance

When LD was accounted for, it was previously calculated that genotyping with the IBC array generates ~20,500 independent tests for individuals of European descent.<sup>27</sup> For maintaining the

conventional 5% false-positive rate, the appropriate multiple-testing-corrected threshold for statistical significance was set at  $p = 2.4 \times 10^{-6}$  for the primary analysis.<sup>16,18</sup> When the individual-level data were used in the analysis, as in the conditional analysis and variable selection, or when replication was available, we used a more permissive  $p$  value threshold of  $p < 1 \times 10^{-4}$ . To maintain our statistical power of ~80% for a SNP with an effect size  $R^2$  of 0.05% during the gender-specific analysis, we used a gender-specific threshold of  $p < 1.8 \times 10^{-3}$ . Because the SNPs included in the gender-specific analysis were previously considered significant in the primary main-effect analysis, this choice had little effect on false positives. The GWAS threshold of  $p < 5 \times 10^{-8}$  was referenced as a comparison to common GWAS practice.

Genomic control estimates reflected by lambda ( $\lambda$ ), a method for quantifying and adjusting population stratification from population-based samples,<sup>28</sup> were derived for each study before the meta-analysis. To avoid the problem of  $\lambda$  estimate inflation due to the high proportion of positive variants (based on the selection criteria of the included SNPs and loci), we excluded the upper 10% of the most statistically significant signals during the estimation of  $\lambda$ .<sup>29</sup> METAL used the option to adjust each study with its corresponding  $\lambda$  before the meta-analysis.

### Association Testing

Association analysis was performed with an additive genetic model with one degree of freedom for all cohorts. We performed gender-stratified analysis within each study for the following three models: model 1 corrected only for population stratification to filter out any artificial association related to population differences; model 2 corrected for population stratification, age, and lipid-lowering medication, by using the correction factors described above because those two extra variables are believed to affect the relationship between the traits and the genotypes tested; and model 3 corrected for population stratification, age, T2D, smoking, BMI, and lipid-lowering medication as described earlier to further control for additional variables able to influence the observed associations. The main results and conditional analysis were reported on the basis of model 1. Variable selection used signals from all three models, and these were maintained for the section of explained variance. All three models were also considered in our scan for previously unreported signals and the replication of previously published associations. The three models were also used as means for understanding the associations observed when additional factors were controlled. Meta-analysis was performed with METAL,<sup>30</sup> and the results were verified with MANTEL<sup>31</sup> and the Metafor package in R.<sup>32</sup> METAL was run with the option of using the  $p$  values for the meta-analysis and took into account sample size and direction of effect, whereas MANTEL used the classical approach of meta-analysis with a fixed-effects model<sup>31</sup> and Metafor used a random-effects scheme with the Hunter-Schmidt estimator.<sup>33</sup> Reported  $p$  values are based on METAL unless otherwise stated. The use of the probability-combination option in METAL does not include the meta-analysis of beta coefficients, although it is able to overcome the problems of differences in phenotype distribution and gender between the studies combined.<sup>30</sup> Metafor used a random-effects model that considered differences between studies as part of the heterogeneity adjusted in the model;<sup>34</sup> therefore, given the number of available studies and the difference between them, the beta coefficients from Metafor were considered as the most accurate estimations of the underlying “true” effects of the SNPs, and they are presented throughout.

After the main analysis, we tested for gender-specific signals of association by performing the meta-analysis separately for males and females and combining their results. Only SNPs deemed statistically significant in the overall analysis were compared between genders for evidence of heterogeneity of effect. Heterogeneity of the meta-analysis was assessed with the  $I^2$  statistic, which describes what percentage of total variation in the study estimates is due to the differences between studies. The statistical significance of the heterogeneity was tested by the chi-square heterogeneity statistic.<sup>35</sup> The criteria for selection of SNPs were: a heterogeneity  $p$  value  $< 0.05$  between males and females and a gender-specific  $p$  value  $< 1.8 \times 10^{-3}$ . When SNPs were in LD ( $r^2 > 0.3$ ), only the strongest associated SNP was presented.

### Conditional Analysis

Loci harboring significant evidence of association with  $p < 10^{-4}$  in model 1 were examined for additional signals with conditional analyses in PLINK<sup>24</sup> in data from seven European-ancestry cohorts for which individual-level genotype data were available. A term was added to the regression model including the lead SNP as a covariate, and SNPs within the same genomic region (within 1 Mb of the lead SNP) were evaluated for significance. A locus-specific Bonferroni correction, based on the number of tests performed, was then applied for determining the significance of independent signals.<sup>16</sup> For loci harboring more than one independent signal, we continued the process until no unreported SNP associations were found.

### Variable Selection and Explained Variance

We used variable selection to identify the most informative SNPs to estimate the total phenotypic variance in the lipid phenotype after age and gender adjustment. To avoid removing individuals with missing data from the analysis, we performed variable selection in the individual-level data after imputation of any missing genotypes by using fastPHASE, a package for haplotypic reconstruction and estimation of missing genotypes.<sup>36</sup> All SNPs with lipid associations at  $p < 1.0 \times 10^{-4}$  for any of the meta-analysis algorithms were included in the selection procedure. The previously reported GWAS SNPs for each lipid phenotype were obtained from both the National Human Genome Research Institute (NHGRI) Catalog of Published GWASs<sup>37</sup> and the GLGC publication.<sup>6</sup> All HumanCVD Beadarray SNPs within 500 kb of the reported SNPs were identified with the SNAP<sup>38</sup> web tool, and SNPs with the highest LD for each single reported polymorphism were forced into the model. The stepwise selection scheme with the Akaike's Information Criterion (AIC)<sup>39</sup> was implemented in R separately for each chromosome.

Given that the SNP selection was performed in the available individual-level data, including information on previously reported polymorphisms, an estimate of association in the same sample might lead to overestimation of the true effect. Therefore, unbiased estimates of the true variance explained were obtained in the WHII study and British Women Heart and Health Study (BWHHS), which did not contribute individual-level data. The ratio of phenotypic variance explained by our results, which took into account the number of SNPs used and the number of observations, was further compared to that estimated with only the previously reported SNPs. For comparison with previous GWASs, we also estimated the explained variance by using the top SNP at each locus plus the independent SNPs in the region as identified through conditional analysis.

### Replication of Previously Undescribed Signals and Previously Unreported Signals

We report two categories of associations—previously unreported SNPs at established loci and SNPs at previously undescribed loci—by using the less stringent statistical threshold of  $p < 1 \times 10^{-4}$ . Loci were designated novel if they had not been reported in the NHGRI GWAS database or in the GLGC study,<sup>6</sup> and novel SNPs were those that were not reported in the GLGC study<sup>6</sup> and that were in LD ( $r^2 \leq 0.3$ ) with any of the GLGC lead SNPs. Loci within 500 kb of reported signals were not considered novel. To attain the final list of novel SNPs, we checked for LD between SNPs within the list itself. In groups of SNPs in LD ( $r^2 \leq 0.3$ ), the SNP with the lowest  $p$  value was reported.

### Replication

Independent replication was then sought for all previously unreported associations. Look ups were performed in five additional cohorts containing data from a total of 25,282 individuals. Characteristics and methodological details for cohorts (referred to as the "25K cohort") are listed in Table S1B. Additional replication was sought through the GLGC GWAS meta-analysis.<sup>6</sup> A signal was considered successfully replicated when its Bonferroni-adjusted  $p$  value in the replication sample was lower than 0.05 and when its estimate was directionally consistent with the discovery meta-analysis. Five of the studies (the Cooperative Health Research in the Region of Augsburg Study, the Atherosclerosis Risk in Communities Study, the University of Pennsylvania Catheterization Study, the Cardiovascular Health Study, and the British Genetics of Hypertension Study) used for this meta-analysis had previously contributed data to the GLGC. These studies were thus removed from the meta-analysis of the discovery with both replication studies.

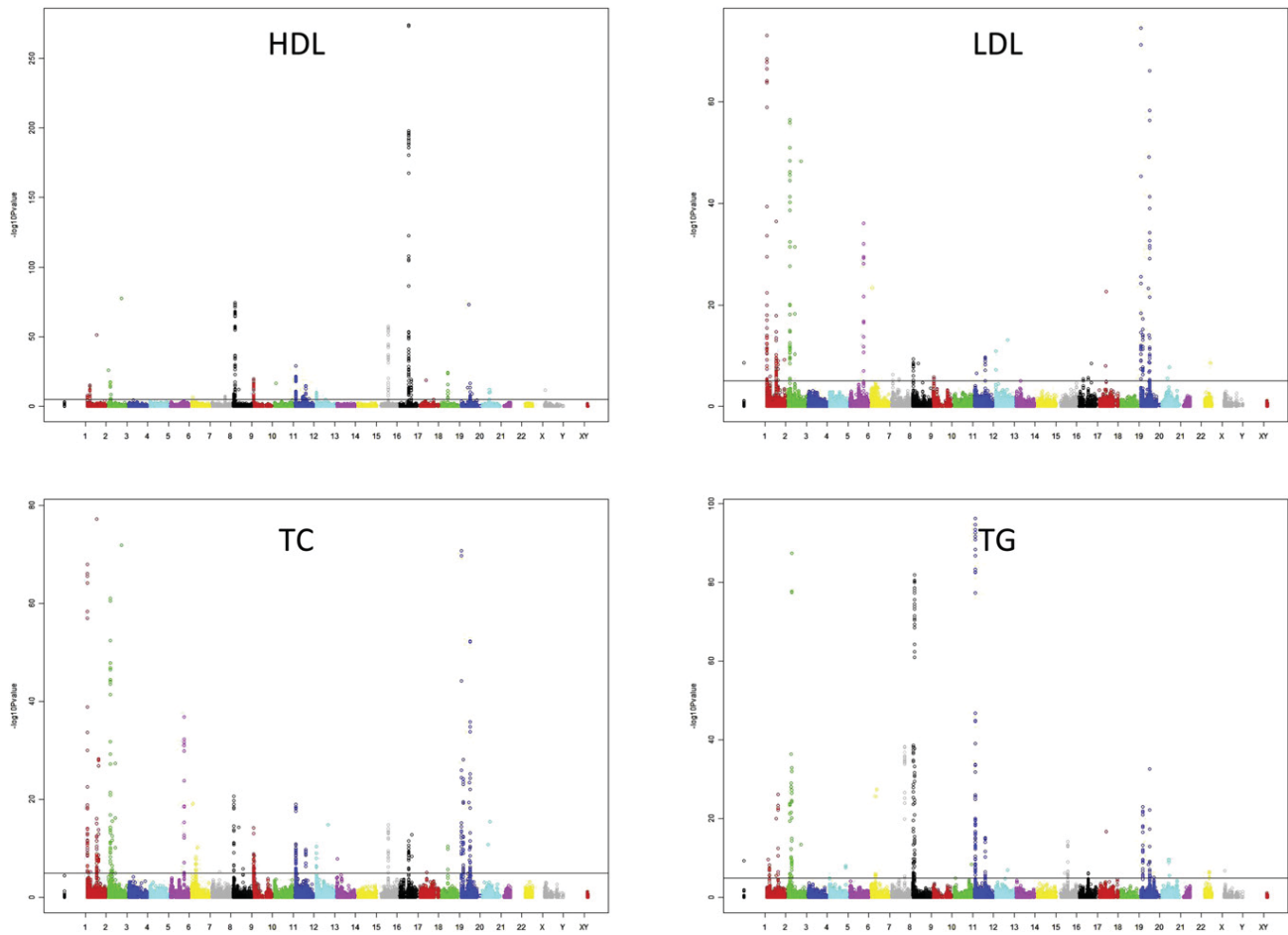
## Results

### Characteristics of Study Samples

A total of 49,227 SNPs were tested in a meta-analysis of 32 cohorts of 66,240 individuals of European ancestry (Table S1A). The ratio of the observed to the null median test statistic,  $\lambda$ , was  $\leq 1.1$  for all studies, except for the GRAPHIC (the Genetic Regulation of Arterial Pressure of Humans in the Community) study, the HAPI (Heredity and Phenotype Intervention) heart study, and PROCARDIS (the Precocious Coronary Artery Disease Study), where related individuals were included. The GRAPHIC study had a  $\lambda$  of 1.2 for all phenotypes considered, and it decreased to  $\sim 1.06$  when rare variants (MAF  $< 0.1\%$ ) were excluded from the data. Both the HAPI study and PROCARDIS had  $\lambda$  values of  $\sim 1.10$  and  $1.12$  for LDL-C and TC, respectively, but these again decreased to  $\leq 1.1$  when rare variants (MAF  $< 0.1\%$ ) were excluded.

### Meta-analysis

We observed 598, 491, 575, and 609 statistically significant ( $p < 2.4 \times 10^{-6}$ ) associations for HDL-C, LDL-C, TC, and TGs, respectively, by using model 1 in METAL (Table S2). After we excluded SNPs present in fewer than 80% of the studies and filtered associations with a meta-analysis



**Figure 2.** Manhattan Plots for HDL-C, LDL-C, TC, and TG Data from the IBC Lipid Meta-analysis Show p Values Based on Those Obtained from the METAL Meta-analysis

$I^2$  value for heterogeneity  $> 35\%$ , the number of statistically significant SNPs was reduced to 276 for HDL-C, 158 for LDL-C, 269 for TC, and 242 for TGs (Table S3). Of the 2,273 statistically significant associations before filtering, 1,094 were with SNPs of MAF  $< 1\%$ , and 1,088 of these had an  $I^2$  value  $> 35\%$  (Table S2). In total, given that several SNPs were associated with more than one phenotype and that SNPs were clustered tightly at certain loci, we identified 549 study-wide significant SNPs in and around 114 different genes. Figure S1 shows the overlap of the identified signals between traits; Manhattan plots for each phenotype are shown in Figure 2.

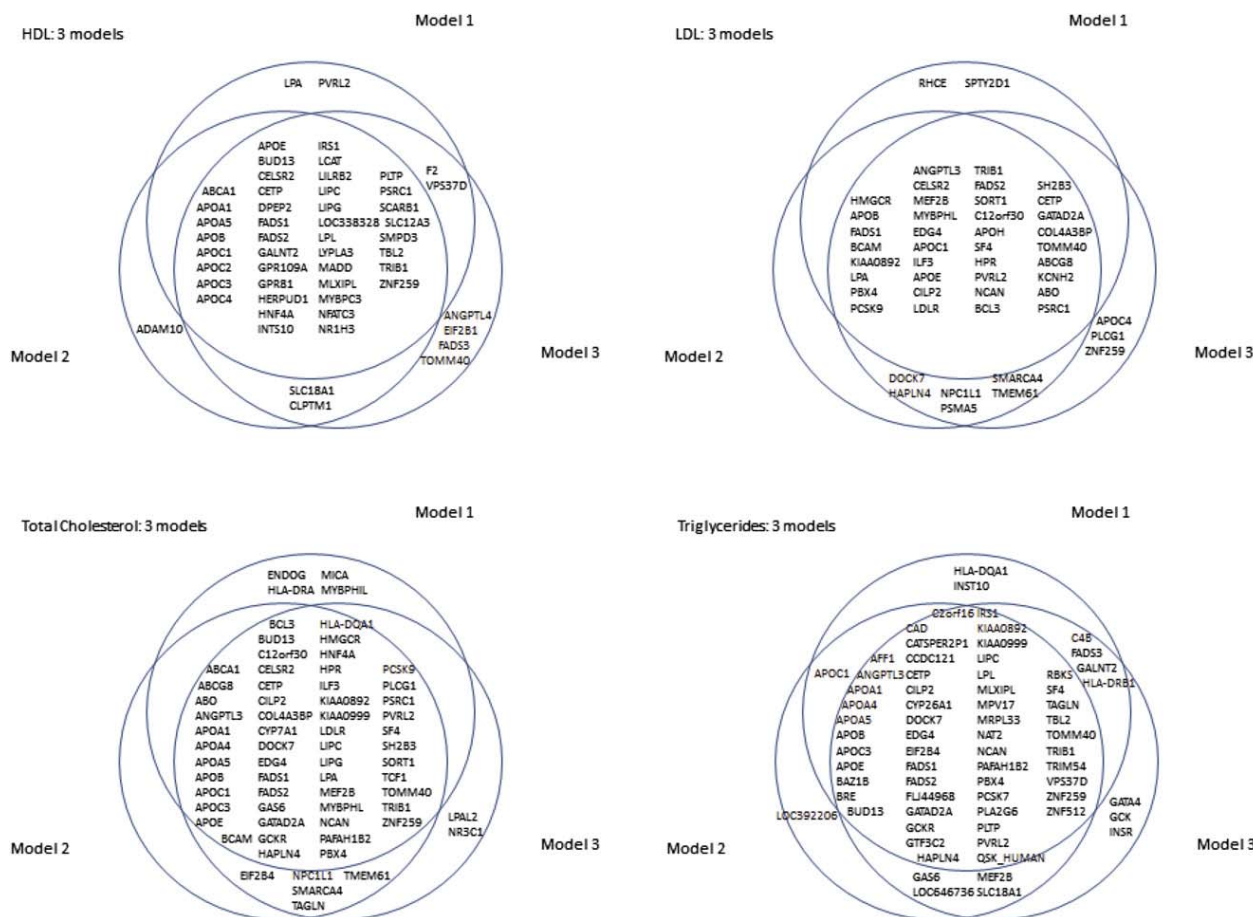
We also analyzed the data by using MANTEL and the Metafor package in R. Although each algorithm used a slightly different method for the meta-analysis, 98% and 95% of the top 100 signals for all phenotypes were also significant in Metafor and MANTEL, respectively. Of the 945 filtered significant associations observed in METAL, 78% were also significant in Metafor and 80% were significant in MANTEL (Table S3). The differences in the results between the three packages were mainly observed in associations with SNPs of low frequency, of high heterogeneity, or with minor statistical sig-

nificance differences, which were close to our cutoff thresholds.

The meta-analysis of plasma-lipid levels, corrected for lipid-lowering medication and adjusted for age (model 2) or for BMI, T2D, and current smoking status (model 3) gave similar results. A summary of the differences (in terms of gene loci identified) between the three models is shown in Figure 3. For the four phenotypes, the Pearson correlation coefficients of the test statistics varied between 0.86 and 0.92 (for models 1 and 2) and between 0.85 and 0.89 (for models 1 and 3). The correlation between models 2 and 3 was 0.94–0.97.

### Conditional Analysis

Because of the dense gene-centric nature of the IBC array, SNPs showing association unsurprisingly formed tight clusters. Using a  $p < 10^{-4}$  threshold, we examined 39 loci for HDL-C, 34 loci for LDL-C, 41 loci for TC, and 32 loci for TGs. Conditioning on the SNP with the strongest p value, we identified independent signals for four traits (Table S4). Although four independent signals were observed in the *LPL* (MIM 609708) locus and three signals were observed in the *BUD13-APOA5* (MIM 606368) cluster



**Figure 3. Venn Diagram per Phenotype for the Comparison of the Three Models Used**

Model 1 corrects only for population stratification; model 2 corrects for population stratification, age, and lipid medication; and model 3 corrects for population stratification, age, T2D, smoking, BMI, and lipid medication. Only signals with  $p < 2.4 \times 10^{-6}$  are included.

for HDL and TGs, only two SNPs (rs268 and rs3289) in the *LPL* locus were overlapping between the phenotypes.

### Gender-Specific Analysis

Of the 66,240 individuals included in the meta-analysis, 31,513 were males and 34,727 were females. The data were analyzed with stratification by gender in each cohort, and the first-stage meta-analysis included both genders. We also tested each gender separately, and the results were compared for concordance between genders. All of the SNPs showing heterogeneity of effect between males and females had the same direction of effect with the overall analysis, but one gender showed a significantly weaker association than the other. Gender-specific differences were found for the individual lipid traits (14 SNPs for HDL-C, 9 SNPs for LDL-C, 14 SNPs for TGs, and 9 SNPs for TC [Table S5]).

### Explained Variance

Variable selection used all unfiltered signals, including 1,156 SNPs for HDL-C, 1,063 SNPs for LDL-C, 1,173 SNPs for TC, and 1,139 SNPs for TGs, identified as significant

at the  $p < 10^{-4}$  threshold from any of the three meta-analysis algorithms. Additionally, previously reported SNPs were forced into the model and included 41 SNPs for HDL-C, 25 SNPs for LDL-C, 22 SNPs for TC, and 35 SNPs for TGs. On the basis of the AIC evaluation, 71 SNPs in HDL-C, 79 SNPs in LDL-C, 120 SNPs in TC, and 75 SNPs in TGs appear to carry additional information beyond the previously reported SNPs. All SNPs, including the previously reported SNPs that were forced into the model, that were retained after variable selection are described in Table S6.

Using the list of SNPs identified by variable selection, we estimated the percentage of explained phenotypic variance in the subset of studies providing individual-level data. After adjustment for age and gender, the identified SNPs explained 9.9% (with 112 SNPs) of the variance in HDL-C, 9.5% (with 104 SNPs) of the variance in LDL-C, 10.3% (with 142 SNPs) of the variance in TC, and 8.0% (with 110 SNPs) of the variance in TGs. Using data derived from previously reported lipid-associated SNPs available in the IBC array, we observed much lower percentages: 6.3% for HDL-C, 4.8% for LDL-C, 4.1% for



TC, and 5.5% for TGs. For comparison, using the common approach of including only the top signal from each locus plus any independent SNP after conditional analysis, we were able to explain 7.9% of the HDL-C variance, 8.4% of the LDL-C variance, 8.2% of the TC variance, and 6.3% of the TG phenotypic variance.

To avoid overestimation resulting from using the same data sets for SNP selection and testing, we also estimated the variance explained in the WHII study and the BWHHS, which did not contribute individual-level data used in the variable selection. For the WHII study, the AIC-selected SNPs explained 11.5% of the variations in HDL-C, 15.6% of the variations in LDL-C, 13.2% of the variations in TC, and 9.8% of the variations in TGs, whereas the previously reported SNPs explained 7.9%, 8.2%, 6.7%, and 7.4% of the phenotypic variance in HDL-C, LDL-C, TC, and TGs, respectively. The corresponding estimates for the BWHHS were 8.2% for HDL-C, 10.7% for LDL-C, 8.1% for TC, and 8.2% for TGs when all the selected SNPs were included in the analysis and 6.0% for HDL-C, 4.2% for LDL-C, 2.6% for TC, and 5.7 for TGs when only the previously reported SNPs were considered. The estimated explained variance, approximate to the heritability due to additive genetic effects, was 10.9% (males) and 12.2% (females) for HDL-C, 12.8% (males) and 11.53% (females) for LDL-C, 12.7% (males) and 12.6% (females) for TC, and 10.7% (males) and 8.6% (females) for TGs (Table S7).

#### Confirmation of Previously Reported Signals

The IBC array covered 57 of the 95 loci reported in the GLGC study<sup>6</sup> and did not include two of the top 20 HDL-C loci (*KLF14* [MIM 609393] and *LILRA3* [MIM 604818]), two of the top 20 LDL-C loci (*TOP1* [MIM 126420] and *ST3GAL4* [MIM 104240]), any of the top 20 TC loci, and two of the top 20 TG loci (*KLHL8* [MIM 611967] and *FRMD5*). Among the directly genotyped SNPs in the IBC array, we were able to replicate the association between HDL-C and 13 out of 18 SNPs, between LDL-C and 11 of 21 SNPs, between TC and 16 of 26 SNPs, and between TGs and 9 of 18 SNPs, each at a threshold of  $p < 10^{-4}$ . Similarly, for the previously reported loci, the lowest  $p$  value SNP in our results replicated 23 of 32 available HDL-C loci, 23 of 32 available LDL-C loci, 30 of 43 available TC loci, and 21 of 29 TG loci. Out of the 57 loci cited above, 31 had specifically the same SNP genotyped by both the GLGC and the IBC array. For those, thus, there was information on directions of effect for both the GLGC and IBC array. In total, these represent 49 signals (given that one SNP can be significant for more than one trait); 13 of those were significant for HDL-C, 11 were significant for LDL-C, 16 were significant for TC, and 9 were significant for TGs. Only one SNP, rs12027135 from *LDLRAP1* (MIM 605747), was found with a non-statistically-significant effect and a direction opposite of that from the GLGC study for two traits (LDL-C and TC).

In addition to the association reported in previous studies, we also identified significant associations for *LCAT* (MIM 606967), *LRP1* (MIM 107770), *LPA* (MIM 152200), *IRS1* (MIM 147545), and *PCSK9* (MIM 607786) loci at the GWAS  $p$  value cutoff of  $p < 5 \times 10^{-8}$  (Table S8).

#### Previously Unreported Signals

Using a  $p < 1.0 \times 10^{-4}$  threshold, we identified 48 significant SNP associations in novel and previously reported genes for HDL-C, and 17 of these are in previously unreported genes. For LDL-C, we identified 38 significantly associated SNPs that were not previously reported in either the NHGRI GWAS database or in the GLGC study.<sup>6</sup> Of these, 18 were located within genes without any previously annotated effect on LDL-C. Similarly, for TC we observed 47 SNPs not previously reported in established genes and 15 SNPs in previously undescribed genes. Finally, for TGs we observed 49 associations, including signals in the 18 genes not previously reported. With the array-wide significance level of  $p < 2.4 \times 10^{-6}$ , there were 11, 5, 12, and 6 novel SNPs for HDL-C, LDL-C, TC, and TGs, respectively. Several loci not previously reported were observed with an array-wide significance of  $p < 2.4 \times 10^{-6}$  (Table S9).

#### Replication

SNPs that showed a significant ( $p < 1.0 \times 10^{-4}$ ) association and that were not previously reported to be associated with lipids were considered for replication in additional studies. These were examined either in our own replication sample of the 25K cohort or with the GLGC data. In total, 23 of the total 69 putative novel gene signals identified in stage 1 were found to be significantly associated in the replication stage. Three of these SNPs reached a GWAS level of significance ( $p < 5 \times 10^{-8}$ ) in the discovery phase and two replicated (67%), nine more reached our array-wide significance and five replicated (56%), and a further 57 surpassed the permissive  $10^{-4}$  cutoff, at which sixteen signals (28%) were replicated.

Of all the signals tested, 11 associations were replicated for HDL-C, 11 were replicated for LDL-C, 17 were replicated for TC, and 12 were replicated for TGs. These replicated signals were in 21 gene regions not previously reported as associated with the lipid phenotypes considered here. A total of 23 signals were replicated in the 25K cohort and/or the GLGC study: five for HDL-C, five for LDL-C, seven for TC, and six for TGs (Table S9). Together with the results of the overall meta-analysis, details of the lead SNPs replicated in each of the novel genes are provided in Table 1.

Additionally, previously unreported SNPs in known loci were also identified. Four SNPs were associated with HDL-C, six were associated with LDL-C, ten were associated with TC, and four were associated with TGs. The results for all SNPs and loci tested for replication are presented in Table S9.

**Table 1. Replicated Previously Unreported Genes**

Gene	Trait	Model	SNP	Chr	Position	Discovery					
						Allele 1	Allele 2	Freq 1	Number of Individuals	p Value	Direction
<i>PPARG</i>	HDL	3	rs12631819	3	12,317,861	T	G	0.0285	53,187	$1.72 \times 10^{-5}$	–
<i>GPIHBP1</i>	HDL	2	rs7388248	8	144,376,728	C	G	0.2878	54,144	$3.30 \times 10^{-7}$	+
<i>DGAT2</i>	HDL	1	rs11236530	11	75,167,052	A	C	0.4246	61,617	$2.44 \times 10^{-6}$	–
<i>HCAR2</i>	HDL	3	rs4759361	12	121,744,233	A	T	0.1584	53,194	$2.05 \times 10^{-9}$	+
<i>FTO</i>	HDL	2	rs1421085	16	52,358,455	T	C	0.6095	54,286	$5.76 \times 10^{-5}$	+
<i>VLDLR</i>	LDL	2	rs7024888	9	2,626,992	T	C	0.9523	48,970	$4.38 \times 10^{-5}$	+
<i>SPTY2D1</i>	LDL	1	rs11024739	11	18,602,419	A	C	0.5514	55,610	$3.09 \times 10^{-7}$	+
<i>BRCA2</i>	LDL	3	rs9534275	13	31,838,345	A	C	0.5149	47,864	$4.61 \times 10^{-6}$	–
<i>SOCS3</i>	LDL	2	rs4082919	17	73,889,077	T	G	0.5148	48,687	$2.33 \times 10^{-5}$	+
<i>APOH</i>	LDL	1	rs1801689	17	61,641,042	A	C	0.9646	55,363	$2.80 \times 10^{-11}$	–
<i>C4B</i>	TGs	1	rs389883	6	32,055,439	T	G	0.6323	57,442	$8.63 \times 10^{-7}$	+
<i>LPAL2</i>	TGs	3	rs3123629	6	160,826,076	A	G	0.3445	47,279	$3.76 \times 10^{-5}$	+
<i>GCK</i>	TGs	3	rs2070971	7	44,164,108	T	G	0.132	49,207	$2.23 \times 10^{-6}$	+
<i>GATA4</i>	TGs	3	rs6983129	8	11,628,545	A	C	0.4918	49,262	$2.30 \times 10^{-6}$	+
<i>SERPINF2</i>	TGs	3	rs2070863	17	1,595,252	T	C	0.2109	49,243	$3.73 \times 10^{-5}$	+
<i>INSR</i>	TGs	1	rs8112883	19	7,130,320	T	G	0.3269	57,525	$8.55 \times 10^{-6}$	–
<i>FCGR2A</i>	TC	3	rs1801274	1	159,746,369	A	G	0.5016	53,200	$2.25 \times 10^{-5}$	+
<i>INSIG2</i>	TC	3	rs12464355	2	118,566,320	A	G	0.9263	53,171	$6.26 \times 10^{-5}$	+
<i>UGT1A1</i>	TC	1	rs11563251	2	234,344,123	T	C	0.0934	65,731	$3.46 \times 10^{-6}$	+
<i>CHUK</i>	TC	3	rs11597086	10	101,943,695	A	C	0.5885	52,928	$3.54 \times 10^{-5}$	–
<i>UBE3B</i>	TC	3	rs7298565	12	108,421,917	A	G	0.5254	48,938	$1.47 \times 10^{-5}$	+
<i>BRCA2</i>	TC	2	rs9534275	13	31,838,345	A	C	0.5157	54,094	$4.16 \times 10^{-6}$	–
<i>SOCS3</i>	TC	2	rs4082919	17	73,889,077	T	G	0.5159	54,065	$1.22 \times 10^{-5}$	+

The following abbreviations are used: chr, chromosome; freq, frequency; het, heterogeneity; mult adj, multiple-variate-adjusted; GLGC, Global Lipids Genetics Consortium; HDL, high-density lipoprotein; LDL, low-density lipoprotein; TGs, triglycerides; and TC, total cholesterol.

## Discussion

To explore association with HDL-C, LDL-C, TC, and TG levels, we used a large-scale locus-centric approach to test 49,227 SNPs carefully prioritized for CVD-related loci in 32 studies with a combined discovery sample size of up to 66,240 individuals of European ancestry. Using an additional sample of 25,282 individuals and the available data derived from the GLGC study,<sup>6</sup> we identified 21 additional loci that have not been associated with lipid levels before and were able to confirm a number of the previously reported associations. We also observed in multiple loci gender-specific differences that were identifiable at the level of explained variance. Finally, although the array covers a smaller proportion of the genome, our heritability estimates were comparable to current GWAS estimates.

Recently, gene-based arrays (such as the IBC array) and the “Metabo-chip,” “Immuno-chip,” and “exome-chip” arrays with content derived from GWASs, next-generation

sequencing, and other plausible sources (such as functional studies) are becoming increasingly popular and offer significant value to individual investigators and consortia.<sup>40–42</sup> They allow the flexibility of incorporating index SNPs, as well as denser probe coverage for finer mapping, across a large number of loci, permitting selective coverage for a range of prioritized findings. The GLGC study<sup>6</sup> provides the most current analysis of the entire genome for common polymorphisms that underpin circulating concentrations of HDL-C, LDL-C, TC, and TGs. Working under the hypothesis that any individual SNP tested is unlikely to have a true effect on the particular phenotype of interest, the GLGC was able to identify SNPs explaining 0.05% of the phenotypic variance with 94.75% power, whereas in our data the same SNPs had 85.07% power to explain the same proportion of the phenotypic variance. This might in part explain our inability to replicate the entire set of SNPs identified in the GLGC study. Nevertheless, these differences in power

Het p Value	25K Replication			Het p Value	Mult Adj p Value	GLGC Replication			Overall Meta-analysis
	Number of Individuals	p Value	Direction			GLGC p Value	Direction	GLGC Mult Adj p Value	
0.2150	10,567	$2.42 \times 10^{-1}$	+	0.6780	1.00	$2.27 \times 10^{-4}$	–	$1.64 \times 10^{-2}$	$8.35 \times 10^{-6}$
0.7494	14,374	$1.42 \times 10^{-2}$	+	0.5845	1.00	$4.03 \times 10^{-4}$	+	$2.90 \times 10^{-2}$	$2.93 \times 10^{-9}$
0.6566	24,732	$3.80 \times 10^{-5}$	–	0.3281	$2.74 \times 10^{-3}$	$4.92 \times 10^{-3}$	–	$3.54 \times 10^{-1}$	$1.40 \times 10^{-9}$
0.8481	13,902	$6.24 \times 10^{-1}$	+	0.2468	1.00	$1.00 \times 10^{-6}$	+	$7.23 \times 10^{-5}$	$2.50 \times 10^{-12}$
0.1952	14,373	$1.08 \times 10^{-1}$	+	0.8410	1.00	$3.31 \times 10^{-5}$	+	$2.38 \times 10^{-3}$	$2.33 \times 10^{-8}$
0.8693	2,998	$1.66 \times 10^{-1}$	+	0.4035	1.00	$4.95 \times 10^{-5}$	+	$3.02 \times 10^{-3}$	$7.51 \times 10^{-8}$
0.3205	24,393	$7.49 \times 10^{-1}$	+	0.5482	1.00	$4.14 \times 10^{-6}$	+	$2.53 \times 10^{-4}$	$6.04 \times 10^{-10}$
0.4505	13,886	$8.26 \times 10^{-1}$	+	0.4184	1.00	$2.48 \times 10^{-7}$	–	$1.51 \times 10^{-5}$	$5.67 \times 10^{-9}$
0.5148	14,347	$1.17 \times 10^{-3}$	+	0.1111	$4.45 \times 10^{-2}$	$3.58 \times 10^{-1}$	+	1.00	$4.46 \times 10^{-4}$
0.6439	2,999	$5.56 \times 10^{-1}$	–	0.3226	1.00	$2.10 \times 10^{-5}$	–	$1.28 \times 10^{-3}$	$9.41 \times 10^{-14}$
0.9931	14,638	$3.81 \times 10^{-2}$	+	0.9599	1.00	$3.95 \times 10^{-15}$	+	$2.85 \times 10^{-13}$	$9.90 \times 10^{-19}$
0.5835	13,935	$5.23 \times 10^{-1}$	+	0.5929	1.00	$1.20 \times 10^{-6}$	+	$8.60 \times 10^{-5}$	$4.40 \times 10^{-8}$
0.2751	13,900	$1.18 \times 10^{-2}$	+	0.8145	$8.46 \times 10^{-1}$	$3.08 \times 10^{-4}$	+	$2.22 \times 10^{-2}$	$2.25 \times 10^{-7}$
0.9860	13,940	$2.02 \times 10^{-1}$	+	0.2548	1.00	$7.54 \times 10^{-6}$	+	$5.43 \times 10^{-4}$	$2.62 \times 10^{-8}$
0.6577	13,903	$4.41 \times 10^{-4}$	+	0.3132	$3.18 \times 10^{-2}$	$5.95 \times 10^{-3}$	+	$4.28 \times 10^{-1}$	$2.41 \times 10^{-5}$
0.2601	14,635	$4.61 \times 10^{-1}$	–	0.8751	1.00	$2.85 \times 10^{-5}$	–	$2.05 \times 10^{-3}$	$4.76 \times 10^{-6}$
0.9812	13,976	$2.54 \times 10^{-3}$	+	0.5643	$2.04 \times 10^{-1}$	$1.58 \times 10^{-4}$	+	$1.26 \times 10^{-2}$	$3.67 \times 10^{-8}$
0.6811	13,851	$7.66 \times 10^{-3}$	+	0.4101	$6.13 \times 10^{-1}$	$8.69 \times 10^{-5}$	+	$6.96 \times 10^{-3}$	$1.19 \times 10^{-8}$
0.0242	14,678	$2.56 \times 10^{-1}$	+	0.0790	1.00	$1.50 \times 10^{-6}$	+	$1.20 \times 10^{-4}$	$2.50 \times 10^{-9}$
0.4075	13,998	$2.05 \times 10^{-1}$	–	0.3244	1.00	$5.81 \times 10^{-5}$	–	$4.65 \times 10^{-3}$	$2.99 \times 10^{-9}$
0.8893	13,997	$7.93 \times 10^{-1}$	+	0.3195	1.00	$1.25 \times 10^{-4}$	+	$1.00 \times 10^{-2}$	$1.98 \times 10^{-6}$
0.6331	14,407	$3.59 \times 10^{-1}$	+	0.3510	1.00	$2.47 \times 10^{-5}$	–	$1.98 \times 10^{-3}$	$5.39 \times 10^{-7}$
0.7658	14,415	$1.61 \times 10^{-4}$	+	0.0394	$1.29 \times 10^{-2}$	$9.56 \times 10^{-3}$	+	$7.65 \times 10^{-1}$	$8.91 \times 10^{-7}$

do not take into account the fact that, compared to those in the GWAS, the SNPs tested here were more likely to be associated with the phenotypes tested because of the selection of SNPs derived from available information concerning a putative role in lipid metabolism. This is apparent from the extreme quantile-quantile plots seen in hypothesis-driven arrays compared to typical GWASs. Such candidate-loci arrays also allow “cosmopolitan tagging” approaches to ensure that sufficient markers across loci of interest for multiple ancestries are achieved. Much of the lipid content of the 2,000 loci on the IBC array was derived from pathway-based candidates.<sup>8</sup> Fifty-seven loci present on the IBC array were associated with lipid traits in the GLGC study. Importantly, however, at the time of array design, very few loci were shown to be robustly associated with lipid phenotypes, showing the clear utility of such candidate-loci approaches for generating putative candidates for validation in large numbers of individuals. One shortcoming of hypothesis-driven

genotyping arrays is limited coverage—38 out of the reported 95 previously reported GLGC loci were not represented on the IBC array. Despite this, the great majority, 74 out of 80, of the strongest associations were covered with greater density than before, highlighting the utility of such approaches. Furthermore, aggregation of data sets such as those presented here has clear utility for conditional analyses—we show that 27 loci have more than one independent signal for the examined lipid traits.

Our most significantly associated locus for HDL-C was *CETP* (MIM 118470). *CETP* is a hydrophobic glycoprotein, which, upon secretion by the liver, is bound mainly to HDL particles in the plasma.<sup>43</sup> *CETP* inhibitors have been shown to significantly increase plasma HDL-C levels and thereby mimic the hyperalphalipoproteinemia encountered in patients with *CETP* deficiency.<sup>44</sup> For both LDL-C and TC, *LDLR* (low density lipoprotein receptor [MIM 606945]) had the lowest p value. *LDLR* encodes the cell-surface LDL receptor, which removes circulating LDL via

receptor-mediated endocytosis. More than 1,600 rare, loss-of-function mutations in *LDLR* have been shown to cause familial hypercholesterolemia.<sup>45–47</sup> Finally, the locus with the strongest association with TG levels was *BUD13* (functional spliceosome-associated protein 71), which is located at the same chromosome 11 locus that contains the *APOA1-C3-A4-A5-ZNF259* (MIM 107680) cluster. In the GLGC GWAS meta-analysis, the top hit for TGs, rs964184 in *APOA1* (MIM 107680), lies within the *BUD13* promoter. *BUD13* is a yeast homolog of an active spliceosome, but little is known about its function in humans. Two of the encoded apolipoproteins, apoA-V and apoC-III, within the cluster influence the activity of lipoprotein lipase (LPL) activity, which is central to hydrolysis of circulating TG-rich lipoproteins. Variants in these genes have long been associated with clinical hypertriglyceridemia.<sup>48,49</sup>

Two of our top signals, *CETP* and *BUD13*, show evidence of gender-specific effects. A wide variety of phenotypes, including CHD, demonstrate sexual dimorphism.<sup>50</sup> Thus, some of the strongest signals we found might be important in one gender alone. An illustrative example is *CETP*, for which SNPs rs17231506 and rs12720922 were both differentially associated with HDL-C levels in men and women. This relationship has been previously suggested, and gender-specific differences in expression levels of the gene product have been hypothesized.<sup>51</sup> Other previously reported SNPs, also shown here to have gender-interactions, include rs531819, rs17398765, rs1367117 in *APOB* (MIM 107730),<sup>52</sup> rs4953023 in *ABCG8* (MIM 605460),<sup>53–55</sup> rs157580 in *TOMM40* (MIM 608061), which is close to *APOE* (MIM 107741),<sup>56</sup> and rs12721109 in *APOC4* (MIM 600745).<sup>57</sup> In addition to gender differences in association with individual SNPs, we observed between-sex differences in trait heritability. Of the four lipid phenotypes examined, LDL-C and TC had minimal between-gender differences in heritability—1.06% and 0.2%, respectively. Females showed higher heritability for HDL-C (1.5% difference), and males showed higher heritability for TGs (1.9% difference). Our results, except those for TGs, are similar to those reported by Weiss et al.<sup>5</sup>—LDL-C showed small narrow-sense heritability differences, and females had higher narrow-sense heritability of HDL-C than did males. In contrast to our findings, Weiss and colleagues<sup>5</sup> showed a stronger but nonsignificant heritability in females compared to men for TGs.

Of the 49,227 SNPs in the array, ~21% had a MAF < 1%, whereas in the 2,273 unfiltered significant associations observed, 48% had a MAF < 1%. After filtering, 0.06% of the significant associations were SNPs of MAF < 1%. Compared to their overall proportion on the array, the higher proportion of rare SNPs passing the array-wide p value threshold can be attributed, in part, to their high heterogeneity. In the majority of cases, SNPs with genotyping errors show high levels of heterogeneity between studies. This might suggest that current methods cannot

easily genotype or call uncommon SNPs. A technical note from Illumina<sup>58</sup> reported that accurate calling of rare variants is possible, although there is an increase in the error rate for rare-allele homozygotes. It is also possible that carriers of rare functional SNPs will have an extreme phenotype, leading, in some cases, to exclusion from the study or to a greater measurement bias in some studies compared to others. At least some of the rare SNPs in our results are known to have functional mutations with large effects. *APOB* (MIM 107730) SNP rs5742904 has a p value of  $1.039 \times 10^{-46}$  with LDL-C in our meta-analysis but has an  $I^2$  of 96.6%. SNP rs5742904 is a known rare mutation (c.10580G>A [p.Arg3527Gln]) (RefSeq accession number NM\_000384.2; ENSG00000084674) involved in hypercholesterolaemia and early CHD.<sup>59,60</sup> The mutation, which has been shown to reduce the affinity for the LDL cholesterol particle, where ApoB is the single protein component for the receptor, is present in 5% of patients with familial hypercholesterolaemia (MIM 143890) in the UK.<sup>61</sup> The identification of rare SNP associations is a substantial challenge, and although we observed a number of strong probable associations, high heterogeneity precludes any firm conclusions.

Our results point toward the existence of multiple independent lipid-associated SNPs in several different loci. One example is *LPL* (lipoprotein lipase), for which the classical view of the primary functional importance of the c.1421C>G (p.Ser447\*) variant (rs328), which causes a premature stop codon, has been modified by the findings that several different polymorphisms at this locus concurrently affect *LPL* expression.<sup>62,63</sup> Interestingly, all of our top signals—*CETP*, *LDLR*, and the *BUD13* cluster—show evidence of the existence of more than one functional SNP. Especially for the cluster around *BUD13*, the risk allele rs9804646-T (MAF = 0.08) is on the same haplotype as the protective allele of the top SNP (rs10750097) in the region, making the former identifiable only after conditional analysis. If this turns out to be the rule for the genetic architecture of lipid loci, any single identified signal at a locus will underestimate the explained variance. Future clinical use in prediction of lipid levels will require more sophisticated approaches for fully capturing information, irrespective of the levels of significance in discovery and replication studies. A number of statistical and computational criteria for selecting the most relevant and informative SNPs are available. Here, we used the AIC criterion as a balance between being inclusive of the SNPs used and avoiding overfitting. It is possible that the exclusive use of only the most significant, not the most informative, SNPs is partly responsible for much of the “missing” heritability that cannot be explained by additional modest-effect common variants.<sup>64</sup> The truly causal polymorphisms are not always included within the genotyped SNPs, making heritability estimates dependent on the LD between causal and observed SNPs.<sup>65</sup> Methods such as selection with AIC or the approach used by Yang et al.,<sup>66</sup> which account for the



total information in the area, will recover some of this missing information, as our results suggest. In addition, use of stringent thresholds of statistical significance will exclude polymorphisms explaining a very small percentage of the variation, despite the potential impact of a great number of such SNPs. Our own results and the work by Yang and colleagues<sup>65</sup> suggest that common SNPs, which do not reach generally acceptable significant levels, are likely to hold additional information. Rare variants still undiscovered might explain some of the “missing” heritability of plasma-lipid phenotypes,<sup>67</sup> but it is not clear how much extremely rare changes can contribute to a population measure such as heritability. Gene  $\times$  gene and gene  $\times$  environment interactions can also play an important role, but statistical constraints hinder their identification.<sup>67</sup> Transgenerational epigenetic alterations have also been suggested as possible sources of heritability,<sup>68</sup> but if they persist for many generations, it is likely that they will acquire LD with a SNP already in the analysis.<sup>69</sup>

For these same reasons, we were also less stringent with our criteria in pursuing potentially novel signals for downstream replication by using a  $p < 1 \times 10^{-4}$  threshold instead of our array-wide significance level of  $p < 2.4 \times 10^{-6}$ . To avoid any increase of false-positive signals, we applied the stringent Bonferroni correction in our replication  $p$  values. We used a very specific definition to declare novelty of a signal. Previously unreported loci were defined as those not described in either the NHGRI GWAS database or by the initial GLGC publication,<sup>6</sup> and previously unreported SNPs were defined as those that were not reported in the GLGC study<sup>6</sup> and that were in LD ( $r^2 < 0.3$ ) with a lead GLGC SNP. This caused some previously characterized SNPs and loci from candidate-gene studies (these SNPs had not been replicated in any GWAS) to be considered novel in our analysis. *GATA4* (MIM 600576) is an example of such. Although its association with TGs was missed by the GWAS, evidence in mice, and recently in humans, reveals that the coded protein is involved in TG absorption from the intestine and underpins plasma TG levels.<sup>70</sup>

The most challenging aspect of evaluating large data sets is that it is extremely difficult to find an ideal replication sample, which is bigger than the discovery set, in such a large meta-analysis setting, where most large studies have been exhausted. We used five previously reported GWASs with and without imputed genotype data. This resulted in an uneven replication in which very few SNPs could be genotyped across the entire replication sample, and most were available only in a fraction of the studies. Considering that we most likely overestimated the true effect size of each SNP, in accordance with the winner’s curse, the power to replicate our signals in a smaller sample is markedly reduced. Nevertheless, a small number of signals, notably *DGAT2* (MIM 606983) for HDL-C, *SOCS3* (MIM 604176) for TC and LDL, and *SERPINF2* (MIM 613168) for TGs, were repli-

cated. The published GLGC data evaluated  $\geq 2.5$  million SNPs (both directly genotyped and imputed) in more than 100,000 individuals and, as such, provided a much more reliable replication set. On the basis of this replication, a further 21 previously unreported loci were confirmed. Moreover, we were also able to identify three SNPs that added additional information to what was previously published. The rs753381 variant is a coding nonsynonymous SNP in *PLCG1* (MIM 172420) and is considered in the GLGC study as part of the LDL-C association with *TOP1* (MIM 126420) (rs6029526). The LD between rs753381 and the previously reported rs6029526 is  $r^2 = 0.82$ , but phospholipase C, gamma 1 (*PLCG1*) was reported to affect cholesterol solubility in bile.<sup>71</sup> Therefore, we speculate that this variant might influence serum-cholesterol levels through interference with the cholesterol cycle and that the relevant locus for the association with LDL-C is *PLCG1* rather than *TOP1*. SNP rs389883, in an intron of *C4B* (MIM 120820), is significantly associated with TGs in both our data and in the GLGC results, but it is not included in the GLGC-reported signals. Similarly, *TCF1* rs2244608, associated with LDL, is only included in the ethnic analysis, but not in the main results of the GWAS meta-analysis.

Well-known genes for other metabolic phenotypes were included in the replicated, previously unreported, signals; one example is *FTO* (MIM 610966) for BMI. *FTO* is believed to be involved in the regulation of food intake and to affect lipolysis in adipose tissue,<sup>72</sup> whereas in our data, *FTO* is also associated with HDL-C, probably through its association with BMI, as the loss of significance in model 3 suggests ( $p = 0.8805$ ). *BRCA2* (MIM 600185), here associated with LDL-C, and *BRCA1* are two of the best known genes in which mutations are associated with breast and ovarian cancers.<sup>73</sup> The precise function of *BRCA2* (MIM 600185) is unclear, but its encoded protein has been implicated in a variety of processes, including DNA repair and recombination, cell-cycle control, and transcription.<sup>74</sup> Some of our other signals are already clinically significant. For example, *HCAR2* (MIM 609163), also known as niacin receptor 1, is an important biomolecular target of niacin, which is a widely prescribed drug for the treatment of dyslipidemia and which acts primarily by inhibiting hepatic DGAT-2 (MIM 606983) and thus lowers secretion of TG-rich lipoproteins and increases HDL-C levels.<sup>75,76</sup>

The evidence from the cumulative meta-analysis of our data, the replication studies, and the published GLGC results suggest that further “true” signals might be found with less stringent  $p$  value thresholds. Given the recent deluge of available genetic data, we propose that a more careful examination is required of common variants of moderate and small effects. This might help explain portions of missing heritability, elucidate the pathways and mechanisms involved in lipid metabolism and CHD, and identify potential loci in which rare SNPs with large effects on the phenotype can be discovered.

## Supplemental Data

Supplemental Data include Supplemental Acknowledgments, one figure, and nine tables and can be found with this article online at <http://www.cell.com/AJHG>.

Received: May 12, 2012

Revised: August 2, 2012

Accepted: August 30, 2012

Published online: October 11, 2012

## Web Resources

The URLs for data presented herein are as follows:

Catalog of Published Genome-Wide Association Studies, <http://www.genome.gov/gwastudies>

GLGC Meta-analysis Data, <http://www.sph.umich.edu/csg/abecasis/public/lipids2010/>

Online Mendelian Inheritance in Man (OMIM), <http://www.omim.org>

SNP Annotation and Proxy Search (SNAP), <http://www.broadinstitute.org/mpg/snap/>

## References

1. Lopez, A.D., Mathers, C.D., Ezzati, M., Jamison, D.T., and Murray, C.J. (2006). Global and regional burden of disease and risk factors, 2001: Systematic analysis of population health data. *Lancet* 367, 1747–1757.
2. Berenson, G.S., Srinivasan, S.R., Bao, W., Newman, W.P., 3rd, Tracy, R.E., and Wattigney, W.A. (1998). Association between multiple cardiovascular risk factors and atherosclerosis in children and young adults. The Bogalusa Heart Study. *N. Engl. J. Med.* 338, 1650–1656.
3. Arsenault, B.J., Boekholdt, S.M., and Kastelein, J.J. (2011). Lipid parameters for measuring risk of cardiovascular disease. *Nat Rev Cardiol* 8, 197–206.
4. Di Angelantonio, E., Sarwar, N., Perry, P., Kaptoge, S., Ray, K.K., Thompson, A., Wood, A.M., Lewington, S., Sattar, N., Packard, C.J., et al.; Emerging Risk Factors Collaboration. (2009). Major lipids, apolipoproteins, and risk of vascular disease. *JAMA* 302, 1993–2000.
5. Weiss, L.A., Pan, L., Abney, M., and Ober, C. (2006). The sex-specific genetic architecture of quantitative traits in humans. *Nat. Genet.* 38, 218–222.
6. Teslovich, T.M., Musunuru, K., Smith, A.V., Edmondson, A.C., Stylianou, I.M., Koseki, M., Pirruccello, J.P., Ripatti, S., Chasman, D.I., Willer, C.J., et al. (2010). Biological, clinical and population relevance of 95 loci for blood lipids. *Nature* 466, 707–713.
7. Manolio, T.A., Collins, F.S., Cox, N.J., Goldstein, D.B., Hindorf, L.A., Hunter, D.J., McCarthy, M.I., Ramos, E.M., Cardon, L.R., Chakravarti, A., et al. (2009). Finding the missing heritability of complex diseases. *Nature* 461, 747–753.
8. Keating, B.J., Tischfield, S., Murray, S.S., Bhargava, T., Price, T.S., Glessner, J.T., Galver, L., Barrett, J.C., Grant, S.F., Farlow, D.N., et al. (2008). Concept, design and implementation of a cardiovascular gene-centric 50 k SNP array for large-scale genomic association studies. *PLoS ONE* 3, e3583.
9. IBC 50K CAD Consortium. (2011). Large-scale gene-centric analysis identifies novel variants for coronary artery disease. *PLoS Genet.* 7, e1002260.
10. Clarke, R., Peden, J.F., Hopewell, J.C., Kyriakou, T., Goel, A., Heath, S.C., Parish, S., Barlera, S., Franzosi, M.G., Rust, S., et al.; PROCARDIS Consortium. (2009). Genetic variants associated with Lp(a) lipoprotein level and coronary disease. *N. Engl. J. Med.* 361, 2518–2528.
11. Lanktree, M.B., Anand, S.S., Yusuf, S., and Hegele, R.A.; SHARE Investigators. (2009). Replication of genetic associations with plasma lipoprotein traits in a multiethnic sample. *J. Lipid Res.* 50, 1487–1496.
12. Talmud, P.J., Drenos, F., Shah, S., Shah, T., Palmen, J., Verzilli, C., Gaunt, T.R., Pallas, J., Lovering, R., Li, K., et al.; ASCOT investigators; NORDIL investigators; BRIGHT Consortium. (2009). Gene-centric association signals for lipids and apolipoproteins identified via the HumanCVD BeadChip. *Am. J. Hum. Genet.* 85, 628–642.
13. Fox, E.R., Young, J.H., Li, Y., Dreisbach, A.W., Keating, B.J., Musani, S.K., Liu, K., Morrison, A.C., Ganesh, S., Kutlar, A., et al.; International Consortium for Blood Pressure Genome-wide Association Studies (ICBP-GWAS). (2011). Association of genetic variation with systolic and diastolic blood pressure among African Americans: The Candidate Gene Association Resource study. *Hum. Mol. Genet.* 20, 2273–2284.
14. Johnson, T., Gaunt, T.R., Newhouse, S.J., Padmanabhan, S., Tomaszewski, M., Kumari, M., Morris, R.W., Tzoulaki, I., O'Brien, E.T., Poulter, N.R., et al.; Cardiogenics Consortium; Global BPgen Consortium. (2011). Blood pressure loci identified with a gene-centric array. *Am. J. Hum. Genet.* 89, 688–700.
15. Cappola, T.P., Li, M., He, J., Ky, B., Gilmore, J., Qu, L., Keating, B., Reilly, M., Kim, C.E., Glessner, J., et al. (2010). Common variants in HSPB7 and FRMD4B associated with advanced heart failure. *Circ Cardiovasc Genet* 3, 147–154.
16. Saxena, R., Elbers, C.C., Guo, Y., Peter, I., Gaunt, T.R., Mega, J.L., Lanktree, M.B., Tare, A., Castillo, B.A., Li, Y.R., et al.; Look AHEAD Research Group; DIAGRAM consortium. (2012). Large-scale gene-centric meta-analysis across 39 studies identifies type 2 diabetes loci. *Am. J. Hum. Genet.* 90, 410–425.
17. Walford, G.A., Green, T., Neale, B., Isakova, T., Rotter, J.I., Grant, S.F., Fox, C.S., Pankow, J.S., Wilson, J.G., Meigs, J.B., et al. (2012). Common genetic variants differentially influence the transition from clinically defined states of fasting glucose metabolism. *Diabetologia* 55, 331–339.
18. Lanktree, M.B., Guo, Y., Murtaza, M., Glessner, J.T., Bailey, S.D., Onland-Moret, N.C., Lettre, G., Ongen, H., Rajagopalan, R., Johnson, T., et al.; Hugh Watkins on behalf of PROCARDIS; Meena Kumari on behalf of the Whitehall II Study and the WHII 50K Group. (2011). Meta-analysis of Dense Gene-centric Association Studies Reveals Common and Uncommon Variants Associated with Height. *Am. J. Hum. Genet.* 88, 6–18.
19. de Ferranti, S.D. (2012). Childhood cholesterol disorders: The iceberg base or nondisease? *Med. Clin. North Am.* 96, 141–154.
20. Kaess, B.M., Tomaszewski, M., Braund, P.S., Stark, K., Rafelt, S., Fischer, M., Hardwick, R., Nelson, C.P., Debiec, R., Huber, F., et al. (2011). Large-scale candidate gene analysis of HDL particle features. *PLoS ONE* 6, e14529.
21. Friedewald, W.T., Levy, R.I., and Fredrickson, D.S. (1972). Estimation of the concentration of low-density lipoprotein cholesterol in plasma, without use of the preparative ultracentrifuge. *Clin. Chem.* 18, 499–502.
22. Marmot, M., and Brunner, E. (2005). Cohort Profile: The Whitehall II study. *Int. J. Epidemiol.* 34, 251–256.

23. Law, M.R., Wald, N.J., and Rudnicka, A.R. (2003). Quantifying effect of statins on low density lipoprotein cholesterol, ischaemic heart disease, and stroke: Systematic review and meta-analysis. *BMJ* 326, 1423.
24. Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M.A., Bender, D., Maller, J., Sklar, P., de Bakker, P.I., Daly, M.J., and Sham, P.C. (2007). PLINK: A tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* 81, 559–575.
25. Price, A.L., Patterson, N.J., Plenge, R.M., Weinblatt, M.E., Shadick, N.A., and Reich, D. (2006). Principal components analysis corrects for stratification in genome-wide association studies. *Nat. Genet.* 38, 904–909.
26. Price, A.L., Butler, J., Patterson, N., Capelli, C., Pascali, V.L., Scarnicci, F., Ruiz-Linares, A., Groop, L., Saetta, A.A., Korkolopoulou, P., et al. (2008). Discerning the ancestry of European Americans in genetic association studies. *PLoS Genet.* 4, e236.
27. Lo, K.S., Wilson, J.G., Lange, L.A., Folsom, A.R., Galarneau, G., Ganesh, S.K., Grant, S.F., Keating, B.J., McCarroll, S.A., Mohler, E.R., 3rd, et al. (2011). Genetic association analysis highlights new loci that modulate hematological trait variation in Caucasians and African Americans. *Hum. Genet.* 129, 307–317.
28. Bacanu, S.A., Devlin, B., and Roeder, K. (2002). Association studies for quantitative traits in structured populations. *Genet. Epidemiol.* 22, 78–93.
29. Pearson, T.A., and Manolio, T.A. (2008). How to interpret a genome-wide association study. *JAMA* 299, 1335–1344.
30. Willer, C.J., Li, Y., and Abecasis, G.R. (2010). METAL: Fast and efficient meta-analysis of genomewide association scans. *Bioinformatics* 26, 2190–2191.
31. de Bakker, P.I., Ferreira, M.A., Jia, X., Neale, B.M., Raychaudhuri, S., and Voight, B.F. (2008). Practical aspects of imputation-driven meta-analysis of genome-wide association studies. *Hum. Mol. Genet.* 17 (R2), R122–R128.
32. Viechtbauer, W. (2010). Conducting meta-analyses in R with the metafor package. *J. Stat. Softw.* 36. Published online August 5, 2010.
33. Hunter, J.E., and Schmidt, F.L. (1990). *Methods of meta-analysis: Correcting error and bias in research findings* (Sage Publications).
34. Thompson, J.R., Attia, J., and Minelli, C. (2011). The meta-analysis of genome-wide association studies. *Brief. Bioinform.* 12, 259–269.
35. Higgins, J.P., and Thompson, S.G. (2002). Quantifying heterogeneity in a meta-analysis. *Stat. Med.* 21, 1539–1558.
36. Scheet, P., and Stephens, M. (2006). A fast and flexible statistical model for large-scale population genotype data: Applications to inferring missing genotypes and haplotypic phase. *Am. J. Hum. Genet.* 78, 629–644.
37. Hindorff, L.A., Sethupathy, P., Junkins, H.A., Ramos, E.M., Mehta, J.P., Collins, E.S., and Manolio, T.A. (2009). Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. *Proc. Natl. Acad. Sci. USA* 106, 9362–9367.
38. Johnson, A.D., Handsaker, R.E., Pulit, S.L., Nizzari, M.M., O'Donnell, C.J., and de Bakker, P.I. (2008). SNAP: A web-based tool for identification and annotation of proxy SNPs using HapMap. *Bioinformatics* 24, 2938–2939.
39. Akaike, H. (1974). A new look at the statistical model identification. *IEEE Trans. Automat. Contr.* 19, 716–723.
40. Cortes, A., and Brown, M.A. (2011). Promise and pitfalls of the Immunochip. *Arthritis Res. Ther.* 13, 101.
41. Buyske, S., Wu, Y., Carty, C.L., Cheng, I.N., Assimes, T.L., Dumitrescu, L., Hindorff, L.A., Mitchell, S., Ambite, J.L., Boerwinkle, E., et al. (2012). Evaluation of the metabochip genotyping array in African Americans and implications for fine mapping of GWAS-identified loci: the PAGE study. *PLoS ONE* 7, e35651.
42. Abecasis Group. (2012). Exome Chip Design ([http://genome.sph.umich.edu/wiki/Exome\\_Chip\\_Design](http://genome.sph.umich.edu/wiki/Exome_Chip_Design)).
43. Barter, P. (2000). CETP and atherosclerosis. *Arterioscler. Thromb. Vasc. Biol.* 20, 2029–2031.
44. Barter, P.J., Brewer, H.B., Jr., Chapman, M.J., Hennekens, C.H., Rader, D.J., and Tall, A.R. (2003). Cholesteryl ester transfer protein: a novel target for raising HDL and inhibiting atherosclerosis. *Arterioscler. Thromb. Vasc. Biol.* 23, 160–167.
45. Austin, M.A., Hutter, C.M., Zimmern, R.L., and Humphries, S.E. (2004). Familial hypercholesterolemia and coronary heart disease: A HuGE association review. *Am. J. Epidemiol.* 160, 421–429.
46. Hegele, R.A. (2009). Plasma lipoproteins: Genetic influences and clinical implications. *Nat. Rev. Genet.* 10, 109–121.
47. Leigh, S.E., Foster, A.H., Whittall, R.A., Hubbard, C.S., and Humphries, S.E. (2008). Update and analysis of the University College London low density lipoprotein receptor familial hypercholesterolemia database. *Ann. Hum. Genet.* 72, 485–498.
48. Holmes, M.V., Harrison, S., Talmud, P.J., Hingorani, A.D., and Humphries, S.E. (2011). Utility of genetic determinants of lipids and cardiovascular events in assessing risk. *Nat Rev Cardiol* 8, 207–221.
49. van Dijk, K.W., Rensen, P.C., Voshol, P.J., and Havekes, L.M. (2004). The role and mode of action of apolipoproteins CIII and AV: Synergistic actors in triglyceride metabolism? *Curr. Opin. Lipidol.* 15, 239–246.
50. Strömberg, A., and Mårtensson, J. (2003). Gender differences in patients with heart failure. *Eur. J. Cardiovasc. Nurs.* 2, 7–18.
51. Anagnostopoulou, K.K., Kolovou, G.D., Kostakou, P.M., Mihas, C., Hatzigeorgiou, G., Marvaki, C., Degiannis, D., Mikhailidis, D.P., and Cokkinos, D.V. (2009). Sex-associated effect of CETP and LPL polymorphisms on postprandial lipids in familial hypercholesterolaemia. *Lipids Health Dis.* 8, 24.
52. Matthan, N.R., Jalbert, S.M., Barrett, P.H., Dolnikowski, G.G., Schaefer, E.J., and Lichtenstein, A.H. (2008). Gender-specific differences in the kinetics of nonfasting TRL, IDL, and LDL apolipoprotein B-100 in men and premenopausal women. *Arterioscler. Thromb. Vasc. Biol.* 28, 1838–1843.
53. Dieter, M.Z., Maher, J.M., Cheng, X., and Klaassen, C.D. (2004). Expression and regulation of the sterol half-transporter genes ABCG5 and ABCG8 in rats. *Comp. Biochem. Physiol. C Toxicol. Pharmacol.* 139, 209–218.
54. Duan, L.P., Wang, H.H., Ohashi, A., and Wang, D.Q. (2006). Role of intestinal sterol transporters Abcg5, Abcg8, and Npc111 in cholesterol absorption in mice: Gender and age effects. *Am. J. Physiol. Gastrointest. Liver Physiol.* 290, G269–G276.
55. Hazard, S.E., and Patel, S.B. (2007). Sterolins ABCG5 and ABCG8: Regulators of whole body dietary sterols. *Pflugers Arch.* 453, 745–752.
56. Kantarci, O.H., Hebrink, D.D., Achenbach, S.J., Pittock, S.J., Altintas, A., Schaefer-Klein, J.L., Atkinson, E.J., De Andrade, M., McMurray, C.T., Rodriguez, M., and Weinschenker, B.G.

- (2004). Association of APOE polymorphisms with disease severity in MS is limited to women. *Neurology* 62, 811–814.
57. Kamboh, M.I., Aston, C.E., and Hamman, R.F. (2000). DNA sequence variation in human apolipoprotein C4 gene and its effect on plasma lipid profile. *Atherosclerosis* 152, 193–201.
  58. Illumina. (2010). Genotyping rare variants: A simulated analysis achieves high cell rates and low error rates from loci containing rare variants. [http://www.illumina.com/Documents/products/technotes/technote\\_genotyping\\_rare\\_variants.pdf](http://www.illumina.com/Documents/products/technotes/technote_genotyping_rare_variants.pdf).
  59. Soria, L.F., Ludwig, E.H., Clarke, H.R., Vega, G.L., Grundy, S.M., and McCarthy, B.J. (1989). Association between a specific apolipoprotein B mutation and familial defective apolipoprotein B-100. *Proc. Natl. Acad. Sci. USA* 86, 587–591.
  60. Tybjaerg-Hansen, A., and Humphries, S.E. (1992). Familial defective apolipoprotein B-100: A single mutation that causes hypercholesterolemia and premature coronary artery disease. *Atherosclerosis* 96, 91–107.
  61. Myant, N.B. (1993). Familial defective apolipoprotein B-100: A review, including some comparisons with familial hypercholesterolaemia. *Atherosclerosis* 104, 1–18.
  62. Smith, A.J., Palmen, J., Putt, W., Talmud, P.J., Humphries, S.E., and Drenos, F. (2010). Application of statistical and functional methodologies for the investigation of genetic determinants of coronary heart disease biomarkers: Lipoprotein lipase genotype and plasma triglycerides as an exemplar. *Hum. Mol. Genet.* 19, 3936–3947.
  63. Deo, R.C., Reich, D., Tandon, A., Akyzbekova, E., Patterson, N., Waliszewska, A., Kathiresan, S., Sarpong, D., Taylor, H.A., Jr., and Wilson, J.G. (2009). Genetic differences between the determinants of lipid profile phenotypes in African and European Americans: The Jackson Heart Study. *PLoS Genet.* 5, e1000342.
  64. Shuldiner, A.R., and Pollin, T.I. (2010). Genomics: Variations in blood lipids. *Nature* 466, 703–704.
  65. Yang, J., Manolio, T.A., Pasquale, L.R., Boerwinkle, E., Caporaso, N., Cunningham, J.M., de Andrade, M., Feenstra, B., Feingold, E., Hayes, M.G., et al. (2011). Genome partitioning of genetic variation for complex traits using common SNPs. *Nat. Genet.* 43, 519–525.
  66. Yang, J., Benyamin, B., McEvoy, B.P., Gordon, S., Henders, A.K., Nyholt, D.R., Madden, P.A., Heath, A.C., Martin, N.G., Montgomery, G.W., et al. (2010). Common SNPs explain a large proportion of the heritability for human height. *Nat. Genet.* 42, 565–569.
  67. Frazer, K.A., Murray, S.S., Schork, N.J., and Topol, E.J. (2009). Human genetic variation and its contribution to complex traits. *Nat. Rev. Genet.* 10, 241–251.
  68. Johannes, F., Porcher, E., Teixeira, F.K., Saliba-Colombani, V., Simon, M., Agier, N., Bulski, A., Albuissou, J., Heredia, F., Audigier, P., et al. (2009). Assessing the impact of transgenerational epigenetic variation on complex traits. *PLoS Genet.* 5, e1000530.
  69. Slatkin, M. (2009). Epigenetic inheritance and the missing heritability problem. *Genetics* 182, 845–850.
  70. Lamina, C., Coassin, S., Illig, T., and Kronenberg, F. (2011). Look beyond one's own nose: Combination of information from publicly available sources reveals an association of GATA4 polymorphisms with plasma triglycerides. *Atherosclerosis* 219, 698–703.
  71. Pattinson, N.R., and Willis, K.E. (1991). Effect of phospholipase C on cholesterol solubilization in model bile. A concanavalin A-binding nucleation-promoting factor from human gallbladder bile. *Gastroenterology* 101, 1339–1344.
  72. Loos, R.J., and Bouchard, C. (2008). FTO: The first gene contributing to common forms of human obesity. *Obes. Rev.* 9, 246–250.
  73. Lancaster, J.M., Wooster, R., Mangion, J., Phelan, C.M., Cochran, C., Gumbs, C., Seal, S., Barfoot, R., Collins, N., Bignell, G., et al. (1996). BRCA2 mutations in primary breast and ovarian cancers. *Nat. Genet.* 13, 238–240.
  74. Venkitaraman, A.R. (2002). Cancer susceptibility and the functions of BRCA1 and BRCA2. *Cell* 108, 171–182.
  75. Viljoen, A., and Wierzbicki, A.S. (2010). Safety and efficacy of laropiprant and extended-release niacin combination in the management of mixed dyslipidemias and primary hypercholesterolemia. *Drug Healthc Patient Saf* 2, 61–71.
  76. Wise, A., Foord, S.M., Fraser, N.J., Barnes, A.A., Elshourbagy, N., Eilert, M., Ignar, D.M., Murdock, P.R., Steplewski, K., Green, A., et al. (2003). Molecular identification of high and low affinity receptors for nicotinic acid. *J. Biol. Chem.* 278, 9869–9874.