UCLA
UCLA Electronic Theses and Dissertations

Title
Identifying and analyzing genetic and epigenetic variation involved in cardiovascular diseases and related metabolic traits

Permalink
https://escholarship.org/uc/item/2xn554hj

Author
Nikkola, Elina

Publication Date
2016

Peer reviewed|Thesis/dissertation

eScholarship.org Powered by the California Digital Library
University of California

UNIVERSITY OF CALIFORNIA

Los Angeles

Identifying and analyzing genetic and epigenetic variation involved in cardiovascular diseases and related metabolic traits

A dissertation submitted in partial satisfaction of the

requirements for the degree Doctor of Philosophy

in Human Genetics

by

Elina Marjaana Nikkola

2016

© Copyright by

Elina Marjaana Nikkola

2016

ABSTRACT OF THE DISSERTATION

Identifying and analyzing genetic and epigenetic variation involved in cardiovascular diseases and related metabolic traits

by

Elina Marjaana Nikkola

Doctor of Philosophy in Human Genetics

University of California, Los Angeles, 2016

Professor Päivi Elisabeth Pajukanta, Chair

Cardiovascular disease (CVD) is the most common cause of death in the U.S. Its risk factors include smoking, hypertension, obesity, and dyslipidemia. High low-density lipoprotein cholesterol (LDL-C) is one of the well-established treatable risk factors for CVD. However, many individuals remain underdiagnosed and current medications do not serve all patients well, mainly due to the side-effects of statins. There are several successful examples of hypercholesterolemia treatments that are developed based on original discoveries in genetic studies, both from dyslipidemic families and population-based association studies. Identification of genes for low LDL-C in individuals with extremely low LDL-C levels is especially attractive as it can help reveal biologically safe mechanisms to lower LDL-C in humans. Chapter 3 describes the exome sequencing analysis we performed in two small Mexican families with familial hypobetalipoproteinemia, characterized by very low levels of LDL-C. We identified a region on chromosome 2p16, segregating with FHBL in the affected family members. In chapter 2, we studied an extended multigenerational Austrian family with familial hypercholesterolemia (FH),

characterized by high LDL-C and premature cardiovascular events. We comprehensively analyzed this family using linkage analysis followed by whole exome sequencing; evaluated their weighted genetic risk scores relative to the general population; and systematically analyzed all previously known FH genes in all family members. We identified a family-specific 10-SNP combination, distinguishing the affected family members from the unaffected ones. In addition, we observed that a subset of family members have rare, previously identified less severe, FH mutation combinations in the *LDLR* and *APOB* genes, likely also contributing to their high levels of LDL-C.

Almost 800,000 individuals suffer a stroke yearly in the United States, and stroke is the leading cause of serious long-term disability. Therefore, it is crucial to develop therapeutic interventions improving outcomes of stroke. One attractive putative treatment option is the limp remote ischemic conditioning (RIC) that is shown to provide neuro-protective effects in previous animal studies, and a trend in human studies. However, mechanisms of RIC are not well understood in humans. In chapter 4, we identified cell cycle and inflammatory changes in our transcriptome and DNA methylation analyses associated with RIC. In summary, this dissertation employs a variety of state-of-the-art massive parallel sequencing methods combined with different study designs to search for genetic risk factors for low and high LDL-C as well as genomic changes associated with the RIC treatment. We identified a region on chromosome 2p16 for FHBL; a family-specific combination of 10 GWAS lipid SNPs for FH; and specific cell cycle and inflammatory changes associated with RIC, a potential new treatment of stroke.

The dissertation of Elina Nikkola is approved.

Rita M. Cantor

Janet S. Sinsheimer

Peter John Tontonoz

Päivi Elisabeth Pajukanta, Committee Chair

University of California, Los Angeles

2016

TABLE OF CONTENTS

LIST OF TABLES

Chapter 2

Chapter 3

Chapter 4

Appendix I

LIST OF FIGURES

Chapter 2

Chapter 3

Chapter 4

Appendix I

ACKNOWLEDGMENTS

I would like to thank my supervisor Professor Päivi Pajukanta for her guidance, for always believing in me, and for our decade-long friendship. It has been a long journey and I am still astonished by her enthusiasm for science; I would be fortunate to take even a small part of that with me - and carry it on - in my future endeavors. My doctoral committee has been invaluable to me as well. I give my sincere thanks and respect to Professors Janet Sinsheimer and Rita Cantor for their exquisite attention to detail and for never laughing at me when I marched into their office to ask (likely very simple for them) statistical questions. Most of the principles and nuances of genetic analyses I have learn from them. I want to acknowledge and thank Professor Peter Tontonoz for his immerse knowledge of cholesterol metabolism and, during our yearly meetings, for always pointing out the big picture. Importantly, I would like to thank all the co-authors for their invaluable contribution to my studies.

I also want to thank the numerous previous and current lab members for always creating a positive and supportive atmosphere. While not always scientific, I have learned a lot from all of you! I would particularly like to thank Drs. Elina Suviolahti, Jenny Lee, Adriana Huertas-Vasquez, Chris Plaisier and Anne Joki-Aho for the warm welcome to the lab and teaching me the basics of genetic techniques, and beyond. In addition, Drs. Daphna Weisglass-Volkov, Kerry Deere, Prasad Linga-Reddy and Blake Haas, and more recently Arthur Ko and Marcus Alvarez, you have all been very patient with me and helped me through sometimes very complex computational genetic analyses. I would also like to extend my gratitude to my colleagues and friends Dr. Anne Beigneux, Dr. Dorota Kaminska and Kristina Garske for proof-reading my dissertation and giving me constructive (sometimes unsolicited) feed-back. I want to thank my friends Goran, Inna, Sandy, Kirsten, Larry and Demetria for climbing other types of mountains with me. Finally, to my parents, sisters and grandparents: thank you for cheering me on all these years. *One of these days, I will be back.*

Chapter 2 is a submitted article entitled “Family-specific aggregation of lipid GWAS variants confers the susceptibility to familial hypercholesterolemia in a large Austrian family” by Elina Nikkola, Arthur Ko, Marcus Alvarez, Rita M. Cantor, Kristina Garske, Elliot Kim, Stephanie Gee, Alejandra Rodriguez, Reinhard Muxel, Niina Matikainen, Sanni Söderlund, Mahdi M. Motazacker, Jan Borén, Claudia Lamina, Florian Kronenberg, Professor, Wolfgang J. Schneider, Professor, Aarno Palotie, Markku Laakso, Marja-Riitta Taskinen and Päivi Pajukanta.

Chapter 3 is a version of an article in preparation entitled “Two Mexican families with familial hypobetalipoproteinemia share a region on chromosome 2p16” by Nikkola *et al*. I would like to thank the following co-authors for their contribution: Arthur Ko, Marcus Alvarez, Daphna Weissglas, Prasad Linga Reddy, Janet S. Sinsheimer, Rita M. Cantor, Ivette Cruz-Bautista, Olimpia Arellano-Campos, Lizeth Gomez-Munguia, Laura Riba, Teresa Tusie-Luna, Carlos Aguilar-Salinas and Päivi Pajukanta.

Chapter 4 is a reprint of ”Remote Ischemic Conditioning Alters Methylation and Expression of Cell Cycle Genes in Aneurysmal Subarachnoid Hemorrhage” by Elina Nikkola, Azim Laiwalla, Arthur Ko, Marcus Alvarez, Mark Connolly, Yinn Cher Ooi, William Hsu, Alex Bui, Päivi Pajukanta, and Nestor R. Gonzalez. Stroke. 2015;46(9):2445-51 and appears with the permission of the Wolters Kluwer Health Lippincott Williams & Wilkins©.

Appendix I is a reprint of “Exome sequencing identifies 2 rare variants for low high-density lipoprotein cholesterol in an extended family” by Mallivara, Reddy, Iulia Iatan, Daphna Weissglas-Volkov, Elina Nikkola, Blake Haas, Miina Juvonen, Isabella Ruel, Janet S. Sinsheimer, Jacques Genest and Päivi Pajukanta. Circ Cardiovasc Genet. 2012;5(5):538-46. And appears with the permission of the Wolters Kluwer Health Lippincott Williams & Wilkins©.

CURRICULUM VITAE

2004	B.S. Bachelor of Biomedical Laboratory Technology, Pirkanmaa Polytechnic, Tampere, Finland
2013	Teaching Assistant, Human Genetics, UCLA

PUBLICATIONS AND PRESENTATIONS

PEER REVIEVED PUBLICATIONS AND SUBMISSIONS

1. **Nikkola E**, Ko A, Alvarez M , Cantor R, Garske K , Kim E, Gee S , Rodriguez A , Muxel R, Matikainen N, Söderlund S , Motazacker M, Borén J, Lamina C , Kronenberg F , Schneider W, Palotie A, Laakso M , Taskinen M-R, Pajukanta P. Family-specific aggregation of lipid GWAS variants confers the susceptibility to familial hypercholesterolemia in a large Austrian family. Submitted

2. Kaminska D, Käkelä P, **Nikkola E**, Venesmaa S, Ilves I, Herzig K-H, Kolehmainen M, Karhunen L, Kuusisto J, Gylling H, Pajukanta P, Laakso M, Pihlajamäki J. Regulation of alternative splicing in human obesity loci. Obesity (Silver Spring). 2016;24(10):2033-7.

3. Rodríguez A, Gonzalez L, Ko A, Alvarez M, Miao Z, Bhagat Y, **Nikkola E**, Cruz-Bautista I, Arellano-Campos O, Muñoz-Hernández LL, Ordóñez-Sánchez ML, Rodriguez-Guillen R, Mohlke KL, Laakso M, Tusie-Luna T, Aguilar-Salinas CA, Pajukanta P. Molecular Characterization of the Lipid Genome-Wide Association Study Signal on Chromosome 18q11.2 Implicates HNF4A-Mediated Regulation of the TMEM241 Gene. Arterioscler Thromb Vasc Biol. 2016;36(7):1350-5.

4. Gusev A, Ko A, Shi H, Bhatia G, Chung W, Penninx BW, Jansen R, de Geus EJ, Boomsma DI, Wright FA, Sullivan PF, **Nikkola E**, Alvarez M, Civelek M, Lusis AJ, Lehtimäki T, Raitoharju E, Kähönen M, Seppälä I, Raitakari OT, Kuusisto J, Laakso M, Price AL, Pajukanta P, Pasaniuc B. Integrative approaches for large-scale transcriptome-wide association studies. Nat Genet. 2016;48(3):245-52.

5. **Nikkola E**, Laiwalla A, Ko A, Alvarez M, Connolly M, Ooi YC, Hsu W, Bui A, Pajukanta P, Gonzalez NR. Remote Ischemic Conditioning Alters Methylation and Expression of Cell Cycle Genes in Aneurysmal Subarachnoid Hemorrhage. Stroke. 2015;46(9):2445-51.

6. Ko A, Cantor RM, Weissglas-Volkov D, **Nikkola E**, Reddy PM, Sinsheimer JS, Pasaniuc B, Brown R, Alvarez M, Rodriguez A, Rodriguez-Guillen R, Bautista IC, Arellano-Campos O, Muñoz-Hernández LL, Salomaa V, Kaprio J, Jula A, Jauhiainen M, Heliövaara M, Raitakari O, Lehtimäki T, Eriksson JG, Perola M, Lohmueller KE, Matikainen N, Taskinen MR, Rodriguez-Torres M, Riba L, Tusie-Luna T, Aguilar-Salinas CA, Pajukanta P. Amerindian-specific regions under positive selection harbour new lipid variants in Latinos. Nat Commun. 2014;5:3983.

7. Iatan I, Choi HY, Ruel I, Reddy MV, Kil H, Lee J, Odeh MA, Salah Z, Abu-Remaileh M, Weissglas-Volkov D, **Nikkola E**, Civelek M, Awan Z, Croce CM, Aqeilan RI, Pajukanta P, Aldaz CM, Genest J. The WWOX gene modulates high-density lipoprotein and lipid metabolism. Circ Cardiovasc Genet. 2014;7(4):491-504.

8. Weissglas-Volkov D, Aguilar-Salinas CA, **Nikkola E**, Deere KA, Cruz-Bautista I, Arellano-Campos O, Muñoz-Hernandez LL, Gomez-Munguia L, Ordoñez-Sánchez ML, Reddy PM, Lusis AJ, Matikainen N, Taskinen MR, Riba L, Cantor RM, Sinsheimer JS, Tusie-Luna T, Pajukanta P. Genomic study in Mexicans identifies a new locus for triglycerides and refines European lipid loci. J Med Genet. 2013;50(5):298-308.

9. Haas BE, Horvath S, Pietiläinen KH, Cantor RM, **Nikkola E**, Weissglas-Volkov D, Rissanen A, Civelek M, Cruz-Bautista I, Riba L, Kuusisto J, Kaprio J, Tusie-Luna T, Laakso M, Aguilar-Salinas CA, Pajukanta P. Adipose co-expression networks across Finns and Mexicans identify novel triglyceride-associated genes. BMC Med Genomics. 2012;5:61.

10. Reddy MV, Iatan I, Weissglas-Volkov D, **Nikkola E**, Haas BE, Juvonen M, Ruel I, Sinsheimer JS, Genest J, Pajukanta P. Exome sequencing identifies 2 rare variants for low high-density lipoprotein cholesterol in an extended family. Circ Cardiovasc Genet. 2012;5(5):538-46.

11. Hosseini M, Ehrhardt N, Weissglas-Volkov D, Lai CM, Mao HZ, Liao JL, **Nikkola E**, Bensadoun A, Taskinen MR, Doolittle MH, Pajukanta P, Péterfy M. Transgenic expression and genetic variation of Lmf1 affect LPL activity in mice and humans. Arterioscler Thromb Vasc Biol. 2012;32(5):1204-10.

12. Haas BE, Weissglas-Volkov D, Aguilar-Salinas CA, **Nikkola E**, Vergnes L, Cruz-Bautista I, Riba L, Stancakova A, Kuusisto J, Soininen P, Kangas AJ, Ala-Korpela M, Tusie-Luna T, Laakso M, Pajukanta P. Evidence of how rs7575840 influences apolipoprotein B-containing lipid particles. Arterioscler Thromb Vasc Biol. 2011;31(5):1201-7.

13. Huertas-Vazquez A, Plaisier C, Weissglas-Volkov D, Sinsheimer J, Canizales-Quinteros S, Cruz-Bautista I, **Nikkola E**, Herrera-Hernandez M, Davila-Cervantes A, Tusie-Luna T, Taskinen MR, Aguilar-Salinas C, Pajukanta P. TCF7L2 is associated with high serum triacylglycerol and differentially expressed in adipose tissue in families with familial combined hyperlipidaemia. Diabetologia. 2008;51(1):62-9.

PRESENTATIONS AT INTERNATIONAL MEETINGS

E. Nikkola, A. Ko, R.M. Cantor, R. Muxel, N. Matikainen, S. Söderlund, M.M. Motazacker , J.A. Kuivenhoven, J. Boren, F. Kronenberg, W. Schneider, A. Palotie, M. Laakso, M.R. Taskinen, P. Pajukanta. Investigation of Multiple Dyslipidemias in a Large Austrian Pedigree by Genetic Risk Scores and Exome Sequencing. The 84th European Atherosclerosis Society Congress 2016, Innsbruck, Austria. (Oral presentation)

E. Nikkola, A. Ko, M. J. Connolly, Y. C. Ooi, P. Pajukanta, and N. Gonzalez. Transcriptome and DNA Methylation Changes in Patients with Subarachnoid Hemorrhage Undergoing Remote Ischemic Preconditioning. International Stroke Conference and State-of-the-Science Stroke Nursing Symposium 2015, Nashville, Tennessee. (Oral presentation, presented by Y. C. Ooi)

E. Nikkola, M. Alvarez, M. V. P. Linga Reddy, A. Ko, D.Weissglas-Volkov, C. Gutierrez-Cirlos, L. Riba, M. L.Ordoñez Sánchez, Y. Segura Kato, T. Tusie-Luna, C.Aguilar-Salinas, P. Pajukanta. Exome Sequencing Identifies a Novel Candidate Gene, NRG1, for Serum Cholesterol Levels in Mexicans. The American Society of Human Genetics (ASHG) Annual Meeting 2013, Boston, Massachusetts. (Poster presentation)

Chapter 1

Introduction

Cardiovascular disease (CVD) is the leading cause of death world-wide with over 17.3 million deaths per year (31% of all deaths) [1]. As many as 85.6 million Americans live with some form of CVD, and by 2030, potentially up to 44% of the US population may have CVD (unpublished AHA tabulation, based on the methodology described by Heidenreich *et al.*[2]). Many forms of CVD are debilitating and lead to a loss of mobility/productivity [1], and such a large number of patients, both in the present day and in the upcoming years (according to projections), also predicts a high financial burden in healthcare costs.

The term CVD covers a continuum of diseases, from mild forms of CVD including dyslipidemias, and hypertension, all the way to severe forms of CVD and disease end-points, such as myocardial infarction (MI), coronary heart disease (CHD) and stroke [3]. Risk factors include environmental factors, especially obesity, poor diet, smoking, and lack of exercise, and genetic factors also contribute to the spectrum of the disease. Heritability can be used to estimate how much of the variation in a phenotypic trait is due to genetics [4]. Using the end points, MI and CHD, the heritability of CVD is 0.38-0.57 [5,6]. CVD heritability can also be estimated for clinical and biochemical measurements such as blood pressure (heritability 0.39-0.42), body mass index (0.37-0.52), waist circumference (0.41), plasma triglycerides (0.48), plasma total cholesterol (TC) and low density lipoprotein cholesterol (LDL-C) (0.57- 0.59) [1].

LDL-C is one of the treatable risk factors for CVD. In fact, the widespread prescription of LDL-C lowering medication, especially statins - or HMG-CoA reductase (HMGCR) inhibitors - has significantly decreased the number of deaths from CVD in the past decade [7] . However, some individuals with high LDL-C cannot tolerate statins, and suffer side effects that are not worth the possible benefits of the treatment [8]. As a result, and supported by significant advancements in sequencing technology, there is a major on-going effort by academic groups and pharmaceutical companies to identify novel genetic risk factors predisposing to CHD and

high LDL-C levels, with the hope of discovering new therapeutic targets and/or biomarkers for CVD.

Brief overview into the genetics of LDL-C metabolism

A classic approach to gain insight into lipid metabolism has been to study dyslipidemic families. The most prominent example is the study of familial hypercholesterolemia by Goldstein and Brown [9], that led to the discovery of the LDL receptor (LDLR), and the feedback regulation of LDLR. The latter discovery is the basis of the mechanism by which statins lower plasma LDL-C levels and reduce heart disease. Over the years, the use of *in vitro* studies combined with the study of dyslipidemic families eventually allowed Goldstein and Brown to discover the LDLR pathway and advance our understanding of cholesterol homeostasis [10,11]. They showed that the LDLR facilitates the uptake of LDL by the cells [10], and that mutations in the *LDLR* gene are associated with elevated plasma LDL-C levels [12]. To this day, over 1,200 LDLR mutations have been identified that contribute to high LDL-C levels [13].

The low density lipoprotein receptor adapter protein 1 (LDLRAP1) plays a critical role in the LDLR-mediated endocytosis of LDL by hepatocytes [14]. Over 10 mutations in the *LDLRAP1* gene have been shown to cause an autosomal recessive form of hypercholesterolemia (ARH) [15].

Following endocytosis, as the pH decreases in endosomes, the LDLR-LDL complex dissociates, and LDLR can either recycle back to the cell-surface or be degraded [16]. PCSK9 is an enzyme encoded by the *PCSK9* gene. PCSK9 binds to LDLR and promotes the rapid degradation of LDLR [17]. Gain-of-function mutations in *PCSK9* are associated with high LDL-C levels and premature CHD by reducing the amount of LDLR at the cell surface and thereby inhibiting the removal of LDL particles from the circulation [18]. Loss-of-function mutations in *PCSK9* increase the number of LDLR on the cell surface; facilitate the uptake of LDL particles;

and are associated with low plasma LDL-C levels [18]. The latter mutations can lower plasma LDL-C by as much as 30-40% and result in an almost 88% reduction of the CHD risk [19].

Apolipoprotein B (ApoB) is encoded by the *APOB* gene and is the primary apolipoprotein of LDL particles [20]. ApoB also functions as a ligand for the LDLR [20]. Nonsense mutations in *APOB* leading to the production of truncated forms of ApoB cause familial hypobetalipoproteinemia (FHBL) [21]. The larger the truncation, the more severe the negative effect on the LDL particle formation, and the lower plasma LDL-C levels are [22]. Conversely, *APOB* missense mutations affecting residues located within the LDLR binding domain in the carboxyl terminus of ApoB - around amino acid 3500 - block the LDLR/ApoB interaction and are associated with increased plasma LDL-C levels and possible premature CHD [23]. In addition, missense mutations that interfere with the proper folding of the carboxyl terminus of ApoB (*i.e.* mutations affecting amino acids 3174–3184, and 4181–4540) indirectly alter the binding properties of ApoB and are associated with increased LDL-C levels [24].

Niemann-Pick C1-like 1 (NPC1L1), a transmembrane protein of the apical membrane of enterocytes and the canalicular membrane of hepatocytes, is required for the intestinal absorption of cholesterol [25]. Loss-of-function mutations in the *NPC1L1* gene are associated with lower plasma LDL-C levels and a reduced risk of having an atherosclerotic CVD event [26].

Familial hypercholesterolemia

Familial hypercholesterolemia (FH) is an autosomal dominant form of dyslipidemia that typically leads to a premature cardiovascular event. The heterozygous form of FH affects as many as 1 in every 67 individuals in some founder populations, and 1 in every 250-600 individuals in the general populations of Europe and the United States [27,28]. The homozygous form of FH is rare (1:1,000,000), with a severe manifestation of the disease [29]. There are no consensus diagnosis criteria for FH. However, the US MedPed Program [30], the Simon Broome

Register Group in the United Kingdom [31], and the Dutch Lipid Clinic Network[28] are widely used as references. Diagnostic is based on the family history; clinical features. including tendon xanthoma, corneal arcus, and premature CHD; biochemical features especially high total cholesterol and LDL-C levels; and genetic testing [32]. The most common genetic causes for FH are loss-of-function mutations in the *LDLR* gene that account for 90% of the FH patients with a known genetic etiology. A smaller number of FH are caused by mutations in the *APOB*, *PCSK9*, and *LDLRAP1* genes. However, only approximately 40% of the FH families have a clear genetic cause [33–36], suggesting that mutations in genes other than the usual culprits may be responsible for FH. Alternatively FH may also be oligogenic or even polygenic in some families. Regardless of the disease etiology, FH diagnosis and treatment are not optimal, to the point that it is estimated that only 25% of individuals with FH are diagnosed [37,38]. The rest of the FH patients are often diagnosed with FH only after their first heart attack [27].

Familial hypobetalipoproteinemia

Familial hypobetalipoproteinemia (FHBL) is an autosomal dominant disorder characterized by reduced levels of plasma TC, LDL-C, and ApoB [39] Heterozygous FHBL affects about 1:500-1:1,000 individuals [22]. However, it is challenging to estimate the exact prevalence of FHBL because most of the subjects with FHBL are heterozygous and asymptomatic, excluding few heterozygous FHBL subjects that develop nonalcoholic fatty liver [40,41]. It is important to note that FHBL is associated with a reduced risk of CVD [22], most likely due to the beneficial effect of life-long low levels of TC and LDL–C [42]. Therefore, it would be highly desirable to identify FHBL-causing mutations and genes, as they might provide new therapeutic targets for lowering plasma LDL-C. Over 60 mutations in the *APOB* gene have been associated with FHBL, and all of these mutations result in the production of truncated ApoB [39,43]. However, FHBL is a genetically heterogeneous disorder that is not always linked to the *APOB* gene [40].

Mutations in *PCSK9* and *ANGPTL3*, and loci on chromosomes 10q25-26 and 3p21 have also been implicated in some FHBL families [44–47].

Examples of treatments of high LDL-C supported by genetics

The main therapeutic treatment for high LDL-C is statins, which are inhibitors of HMGCR, the rate limiting enzyme in cholesterol synthesis. The common HMGCR variant rs12916 has approximately ~2–3% effect on plasma LDL-C, as estimated by genome wide association (GWAS) studies [48]. However, statins can reduce LDL-C by 50% or more [49]. Interestingly, it was shown in a longitudinal study that the HMGCR risk variant carriers have similar serum fatty-acid, and metabolomics profiles as statins users [50].

Ezetimibe is a cholesterol-absorption inhibitor that functions on the brush border of the small intestinal epithelium. NPC1L1 is believed to be the molecular target of ezetimibe [51]. Similarly as the loss-of-function mutations of the *NPC1L1* gene, inhibition of NPC1L1 by ezetimibe reduces the intestinal absorption of cholesterol; lowers plasma LDL-C levels; and reduces the risk of atherosclerotic CVD events [51,52].

Recently, PCSK9 antibodies have become available as a new class of plasma LDL-C lowering drugs. Administration of PCSK9 antibodies can result in a 50-60% reduction in LDL-C levels, even in patients already receiving statin and ezetimibe [53]. Long term end point data, such as reduction in cardiovascular disease, are not available for this particular therapy yet, but are expected in 2017.

Identifying genetic risk factors beyond the known FH and FHBL genes

Dyslipidemic families can be used to identity new variants, and genes that control plasma LDL-C levels in order to design novel therapeutic interventions to lower plasma LDL-C and reduce CVD. Particularly, extended families enable the investigation of co-segregation between the trait and rare variants among the affected and non-affected family members.

Several studies have been carried out using FH families without a known causative mutation, either by focusing on the known FH genes [54–56], or by whole exome-sequencing of all protein-coding genes [57,58]. As a result, exome sequencing has helped identify mutations in the *APOE*, *LIPA* and *STAP1* genes as a likely cause for FH [56,58–60]. In addition, chromosomal loci on 3q25-26 and 21q22 and other loci have been identified using linkage analysis [61–63] . However, these studies were not able to pinpoint the actual causative functional variants underlying these linkage peaks.

Another study first screened 554 individuals from both tails of the plasma LDL-C levels, the <2nd percentile and >98th percentile, using exome sequencing. The candidate genes that emerged from that initial screen were then tested on a follow-up set of 1,302 samples. This approach discovered overrepresentation of rare variants in a novel gene, *PNPLA5* that significantly correlated with high LDL-C using a gene-level burden test [64]. These studies also found suggestive evidence of an association between variants in *ABCG5* and *NPC1L1* and plasma LDL-C levels [64]. Of note, both of these genes had been previously identified in Mendelian forms of high LDL-C [65,66].

We previously successfully identified two rare variants in the well-known lipid genes, *LPL* and *ABCA1*, contributing to low levels of HDL-C in a large French-Canadian family by exome sequencing and linkage analysis [67] (Appendix I). In addition, as described in Chapter 3, we performed exome sequencing on two small Mexican families with FHBL. We began to analyze these two families in 2013, and have been reanalyzing this small set of individuals yearly, as methods for gene annotation (ClinVar [68], ANNOVAR [69]), population frequencies (Exome Aggregation Consortium (ExAC) and the Genome Aggregation Database (gnomAD) [70]), gene expression data in different tissues (GTEx [71]) and genotypes with phenotype data (http://www.type2diabetesgenetics.org/) have been improving. Eventually, we identified a region on chromosome 2p16 shared by all affected individuals in the two small Mexican families, but

we could not pinpoint the functional causative variant/s within that region (Chapter 3). The lack of success can be partially explained by the small family sizes, as it is likely there were not enough informative family members to narrow our search for causal variants.

In addition, large scale GWAS studies have found 58 replicated loci for LDL-C with modest effect sizes [48,72] Some of these "small effect variants" reside close to genes *ABCG5/8*, *SORT1*, *MYLIP*, *INSIG2* and *TM6SF2,* which have strong functional evidence to support their role in LDL-C metabolism [51]. Of note, some of the GWAS SNPs are located within an active enhancer or transcription factor binding sites. This could be a common the mechanism by which these SNPs influence serum LDL-C levels in aggregate.

Futema *et al*. searched for evidence of rare and novel functional variants within the LDL-C GWAS genes in FH patients without finding evidence of over-representation in that particular FH cohort when compared to controls [57].

Overall, unlike originally expected, GWAS studies have not identified variants with a large impact, and whole exome sequencing has not been very successful in finding novel FH or LDL-C genes and variants. For example, Stitziel *et al*. exome-sequenced 42 dyslipidemic kindreds, with potentially a monogenic cause, but found likely pathogenic variants in only 12% of the families [58].

Polygenic causes for high LDL-C and FH

Despite the constantly improving sequencing methods, the causal mutation has not be detected in roughly 60% of individuals who fulfill the clinical diagnosis of FH [73]. Recently it has been observed that some of these unsolved FH cases could be explained by the aggregate of small effects from LDL-C GWAS loci [74]. It is note-worthy that many of the small effect common GWAS LDL-C variants reside in the genes that were originally identified for the monogenic FH

(LDLR, APOB and PCSK9) as well as in other known genes of cholesterol metabolism, such as HMGCR.

Weighted polygenic risk scores have been calculated to estimate the cumulative effect size of each genome-wide significant LDL-C-increasing SNP derived from a lipid GWAS study of a metacohort of approximately 180,000 individuals [48]. A population-specific weighted genetic risk score over the 90th percentile has previously been suggested as a potential polygenic cause for FH [73]. In addition, there are several proposed LDL-C increasing SNP subsets and combinations that have been used to distinguish FH patients from healthy individuals [75].

Chapter 2 describes the comprehensive analysis of a large Austrian FH family. The proband was found to be negative for the known FH variants. We utilized a commonly used approach in which a linkage analysis of a large family is followed by exome sequencing to search for rare variants shared by the affected family members in the linked regions [76]). We first demonstrated that we have adequate power to detect significant two-point lod scores (lod>3.0) with the given family structure using the SLINK software [77]. However, the highest two-point lod score in the FH family for high LDL-C status was only 1.9 on chromosome 17. In total, there were 17 regions with a lod score > 1.0, which we then screened for functional variants (nonsynonymous and splice site) using exome sequencing. We found no evidence for plausible functional variants shared by the affected family members within these regions. Nevertheless, the validity of our approach was confirmed by utilizing a subset of family members that exhibited high lipoprotein (a) (Lp(a)) levels, an independent risk factor for CVD. Lp(a) has a strong genetic component (heritability = 0.91), and is largely regulated by variants at the LPA locus on chromosome 6q27. We identified a well-known Lp(a) variant, rs3798220, at the LPA locus using linkage and subsequent family-based co-segregation analysis in the family members with high Lp(a). Taken together, we had a sufficient power to identify novel locus/loci for high LDL-C.

The results from the linkage analysis followed by an exome sequence analysis suggested a polygenic cause for FH in this extended family. Therefore, we examined the known LDL-C increasing SNPs from GWAS reported by Willer *et al.* using weighted genetic risk score (wGRS) analysis[48]. The average of wGRS in the affected individuals was in the ~75th percentile, thus not completely explaining the high LDL-C. This led us to look more carefully at the GWAS variants predominately shared by the affected family members and not present in the unaffected family members. We identified a combination of 10 family-specific SNPs that can distinguish the affected family members from the unaffected family members. More importantly, the risk scores constructed from this combination resulted in an average risk score of the affected family members of more than the 90th percentile. As a validation, we randomly constructed 100 times the 10-SNP risk scores and could not find a combination leading to the 90th percentile among the affected family members using this permutation analysis. In addition, when we screened the known FH genes in all family members, we identified variant combinations in APOB and LDLR in a subset of the affected individuals, potentially explaining high LDL-C levels. Taken together, this family exhibits very unique etiology for the clinical FH with a specific set of 10 GWAS LDL-C variants, and rare LDLR and APOB variants in a subset of the affected individuals.

Stroke is the leading cause of serious long-term disability

Overall, stroke is the fifth leading cause of death in the United States, with approximately 795,000 deaths yearly [1]. Stroke presents gender differences between women and men, being the third leading cause of death for women versus the fifth in men (Centers for Disease Control and Prevention). In addition to gender, other risk factors include age, genetics, and ethnicity. Also, some medical conditions, such as high blood pressure, high cholesterol, heart disease, diabetes, and obesity, can increase the risk of stroke. Avoiding smoking, drinking alcohol moderately, eating a healthy diet, and exercising can reduce the risk of stroke [1]. Most strokes (87%) are ischemic strokes in which the artery that brings oxygen-rich blood to the brain

becomes blocked. Another type of stroke is the hemorrhagic stroke (3%), occurring when an artery in the brain leaks blood or ruptures [1]. Subarachnoid hemorrhage is one form of the hemorrhagic stroke that can occur at any age, including young adults [78]. Strokes cause brain damage that potentially leads to physical disability, including paralysis, speech and swallowing difficulties, as well as a memory loss [79].

Remote ischemic conditioning

During a heart attack or stroke, the heart or brain can be exposed to a detrimental ischemia-reperfusion injury. Therapeutic interventions that could protect these two tissues from such a damaging effect are in high demand. Przyklenk [80] was the first to describe remote ischemic conditioning (RIC), an intervention that consists of inducing short, non-lethal ischemia in peripheral tissues to protect vital organs such as the heart and the brain from ischemia-reperfusion injury. Since then, many animal studies have shown cardio- and neuro-protective effects of RIC [81,82]. A small number of human studies have been conducted as well, and the majority of these studies were done in patients undergoing coronary artery bypass graft surgery [83,84]. While some of the human studies showed that RIC was associated with a reduction in the post-operative release of biomarkers for myocardial injury, such as troponin I and creatine kinase-myocardial band (CK-MB) [85,86], other studies could not document a positive effect of RIC on the outcome [84]. Thus, it would be highly desirable to firmly establish the therapeutic value of RIC in humans. It would be an attractive intervention due to its low-cost and non-invasive nature.

The specific mechanisms contributing to the cardio- and neuro-protective effects of RIC are not fully understood. They appear to be complex and multifactorial [87]. The most prevalent hypothesis is that, in response to short non-lethal ischemic assaults, some unknown factors are released from the pre-conditioned tissue into the bloodstream, reaching then the heart or brain, and exerting their cardio- or neuro-protective effect either through the nervous system, or

through a systemic anti-inflammatory and anti-apoptotic pathway. The presence of circulating cardio-protective factor(s) after RIC was first shown in blood transfusion experiment performed in rabbits [88] , where the blood from the conditioned rabbit conferred cardio-protection to the recipient, untreated rabbit. Studies in rodents and rabbits have shown that the cardio-protective effect of the circulating humoral factor(s) released during RIC is mediated by the activation of G-protein coupled receptors and the subsequent induction of the intracellular protein kinase pathway [82]. Interestingly, other studies have shown an increase in plasma levels of stromal-derived factor-1α (SDF-1a or CXCL12) in rats undergoing RIC [89]. CXCL12 is a chemokine that regulates multiple physiological processes of organ homeostasis, and exerts its biological effects through the G protein-coupled chemokine receptor CXCR4 [90].

Studies in humans have identified inflammatory gene expression changes in healthy adults undergoing RIC [91,92], as well as changes in the expression of genes involved in innate immune responses, TNF-signaling pathway, leukocyte adhesion, chemotaxis, and exocytosis [91]. More recently, STAT5 signaling has been implicated in the expression changes associated with RIC in humans [93]. In addition, a recent proteomic study reported multiple potential cardio-protective targets released into the bloodstream after a limb RIC protocol [94]. However, these studies are far from being definitive, mostly because human proteomic studies have been controversial, mainly due to the lack of reproducible data [95]. Overall, animal studies have been more successful than human studies at revealing the molecular mechanisms underlying the cardio-protective effect of RIC. Taken together, there have been too few studies on humans undergoing RIC to draw firm conclusions. This lack of definitive data needs to be addressed because identification of the molecules that mediate the potential cardio-protective effect of RIC could help design new treatments, and provide useful biomarkers for a better management of myocardial ischemia and stroke.

Transcriptome profiling can provide molecular insights into human health and disease. Recent advances in sequencing techniques have allowed researchers to tease out gene expression differences between tissues as well as gene expression differences between individuals. By RNA-sequencing "normal" human tissues, the Genotype-Tissue Expression Project (GTEx) has developed a searchable database that catalogs gene expression across different tissues [71]. Mele' *et al*. were able to show that variation in gene expression is greater between the tissues of the same individual (47%) than between individuals (4%) [96]. A lot of the individual-to-individual gene expression variation are accounted by age, sex, and ethnicity, although some could potentially also be accounted by disease phenotypes [96]. Interestingly, Li *et al*. showed that transcriptome analysis can be a powerful tool to observe changes in molecular networks in response to treatments [97].

Epigenetic studies investigate heritable changes in gene function without changes in DNA sequence. It is known to be highly relevant to disease processes, including CVD and its risk factors [98], through the complex interplay between tissues and environment [99]. DNA methylation is one of the main forms of epigenetic modifications and typically occurs in a CG dinucleotide, so that the C is methylated. The majority of the DNA methylation sites are static, but a sizeable 21.8% of the DNA methylation regions are actually dynamic, and could potentially be methylated in response to outside stimuli [100].

Chapter 4 describes our transcriptomics and epigenetics study of 13 patients with aneurysmal subarachnoid hemorrhage (aSAH) undergoing RIC. Gonzalez *et al*. previously showed the safety and practicality of lower limb RIC in aSAH patients combined with modifications in neurovascular and cerebral metabolism [87]. They showed that RIC was associated with a good outcome (odds ratio 5.17; 95% confidence interval: 1.2-25.0) and a trend toward lower incidence of stroke and death in aSAH patients [101]. In these patients, we identified coordinated methylation and expression changes of cell cycle genes, utilizing a

longitudinal study design in which we were able to use intra-individual control. This study design was critical to our success in identifying significant changes because the patients were of different age, sex, ethnicity, and were taking various medications. Furthermore, because integrated genomic methods can strengthen our understanding of disease pathophysiology and provide additional evidence of the mechanisms involved [102], we focused our differential methylation investigation only on the regions residing near the genes that we observed to be differentially expressed genes during RIC. However, our cohort was rather small, and follow-up studies with additional patients undergoing RIC are needed to replicate and robustly further verify our findings, and in order to identify the actual cardio- and neuro-protective factors from the bloodstream, accessible either as a therapeutic target or biomarker.

In summary, we are living in exciting times in genetics and genomics. While the sequencing methods are improving in quality and cost, collaborative efforts between research laboratories, clinicians, and disciplines will advance genetic and genomic research. The ENCODE (Encyclopedia of DNA Elements) project aims to characterize the functional parts of the genome, specifically the transcriptome, DNA and RNA binding sites, DNA accessibility, and DNA methylation [103]. The data of the ENCODE and other large scale projects, such as the String protein-protein network database [104] and ExAC, can already be freely accessed online. This will help researchers with fewer resources to prioritize genomic regions, genes, and variants. While we still have several limitations to overcome (Chapter 5), we are steadily moving towards personalized and family-centered genomic medicine.

References

1. Mozaffarian D, Benjamin EJ, Go AS, Arnett DK, Blaha MJ, Cushman M, Das SR, Ferranti S De, Després JP, Fullerton HJ, Howard VJ, Huffman MD, Isasi CR, Jiménez MC, Judd SE, Kissela BM, Lichtman JH, Lisabeth LD, Liu S, MacKey RH, Magid DJ, McGuire DK, Mohler ER, Moy CS, Muntner P, Mussolino ME, Nasir K, Neumar RW, Nichol G, Palaniappan L, Pandey DK, Reeves MJ, Rodriguez CJ, Rosamond W, Sorlie PD, Stein J, Towfighi A, Turan TN, Virani SS, Woo D, Yeh RW, Turner MB. Heart disease and stroke statistics-2016 update a report from the American Heart Association. *Circulation*. 2016;133(4):e38-e48. doi:10.1161/CIR.0000000000000350.

2. Heidenreich P, Trogdon J, Khavjou O, Ezekowitz M, Finkelstein A, Hong Y, Jonhston C, Khera A, Lloyd-Jones D, Nelson S, Nichol G, Orenstein D., Heidenreich P, Trogdon J, Khavjou O, Ezekowitz M, Finkelstein A, Hong Y, Jonhston C, Khera A, Lloyd-Jones D, Nel. Forecasting the future of cardiovascular disease in the United States: A policy statement from the American Heart Association. *Circulation*. 2011;123(8):933-944. doi:10.1161/CIR.0b013e31820a55f5.

3. Dzau VJ, Antman EM, Black HR, Hayes DL, Manson JE, Plutzky J, Popma JJ, Stevenson W. The cardiovascular disease continuum validated: Clinical evidence of improved patient outcomes: Part I: Pathophysiology and clinical trial evidence (risk factors through stable coronary artery disease). *Circulation*. 2006;114(25):2850-2870. doi:10.1161/CIRCULATIONAHA.106.655688.

4. Visscher PM, Hill WG, Wray NR. Heritability in the genomics era — concepts and misconceptions. *Nat Rev Genet*. 2008;9(4):255-266. doi:10.1038/nrg2322.

5. Fischer M, Broeckel U, Holmer S, Baessler A, Hengstenberg C, Mayer B, Erdmann J, Klein G, Riegger G, Jacob HJ, Schunkert H. Distinct heritable patterns of angiographic coronary artery disease in families with myocardial infarction. *Circulation*. 2005;111(7):855-862. doi:10.1161/01.CIR.0000155611.41961.BB.

6. Zdravkovic S, Wienke A, Pedersen NL, Marenberg ME, Yashin AI, De Faire U. Heritability of death from coronary heart disease: A 36-year follow-up of 20 966 Swedish twins. *J Intern Med*. 2002;252(3):247-254. doi:10.1046/j.1365-2796.2002.01029.x.

7. Nichols M, Townsend N, Scarborough P, Rayner M. Trends in age-specific coronary heart disease mortality in the European Union over three decades: 1980-2009. *Eur Heart J*. 2013;34(39):3017-3027. doi:10.1093/eurheartj/eht159.

8. Maningat P, Gordon BR, Breslow JL. How do we improve patient compliance and adherence to long-term statin therapy? *Curr Atheroscler Rep*. 2013;15(1). doi:10.1007/s11883-012-0291-7.

9. Goldstein JL, Brown MS. Familial hypercholesterolemia: identification of a defect in the regulation of 3-hydroxy-3-methylglutaryl coenzyme A reductase activity associated with overproduction of cholesterol. *Proc Natl Acad Sci U S A*. 1973;70(10):2804-2808. doi:10.1073/pnas.70.10.2804.

10. Goldstein JL, Brown MS. The LDL receptor. *Arterioscler Thromb Vasc Biol*. 2009;29(4):431-438. doi:10.1161/ATVBAHA.108.179564.

11. Brown MS, Goldstein JL. A receptor-mediated pathway for cholesterol homeostasis. *Science*. 1986;232(4746):34-47. doi:10.1126/science.3513311.

12. Tolleshaug H, Hobgood KK, Brown MS, Goldstein JL. The LDL receptor locus in familial hypercholesterolemia: Multiple mutations disrupt transport and processing of a membrane receptor. *Cell*. 1983;32(3):941-951. doi:10.1016/0092-8674(83)90079-X.

13. Usifo E, Leigh SEA, Whittall RA, Lench N, Taylor A, Yeats C, Orengo CA, Martin ACR, Celli J, Humphries SE. Low-Density Lipoprotein Receptor Gene Familial Hypercholesterolemia Variant Database: Update and Pathological Assessment. *Ann Hum Genet*. 2012;76(5):387-401. doi:10.1111/j.1469-1809.2012.00724.x.

14. Sirinian MI, Belleudi F, Campagna F, Ceridono M, Garofalo T, Quagliarini F, Verna R, Calandra S, Bertolini S, Sorice M, Torrisi MR, Arca M. Adaptor protein ARH is recruited to the plasma membrane by low density lipoprotein (LDL) binding and modulates endocytosis of the LDL/LDL receptor complex in hepatocytes. *J Biol Chem*. 2005;280(46):38416-38423. doi:10.1074/jbc.M504343200.

15. Tada H, Kawashiri M-A, Nohara A, Inazu A, Kobayashi J, Mabuchi H, Yamagishi M. Autosomal Recessive Hypercholesterolemia: A Mild Phenotype of Familial Hypercholesterolemia: Insight from the Kinetic Study using Stable Isotope and Animal Studies. *J Atheroscler Thromb*. 2015;22:1-9. doi:10.5551/jat.27227.

16. Brown MS, Herz J, Goldstein JL. LDL-receptor structure: Calcium cages, acid baths and recycling receptors. *Nature*. 1997;388(August):629-630. doi:10.1016/0092-8674(83)90052-1.

17. Kosenko T, Golder M, Leblond G, Weng W, Lagace TA. Low density lipoprotein binds to proprotein convertase subtilisin/kexin type-9 (PCSK9) in human plasma and inhibits PCSK9-mediated low density lipoprotein receptor degradation. *J Biol Chem*. 2013;288(12):8279-8288. doi:10.1074/jbc.M112.421370.

18. Schulz R, Schlüter KD, Laufs U. Molecular and cellular function of the proprotein convertase subtilisin/kexin type 9 (PCSK9). *Basic Res Cardiol*. 2015;110(2). doi:10.1007/s00395-015-0463-z.

19. Zhao Z, Tuakli-Wosornu Y, Lagace TA, Kinch L, Grishin N V, Horton JD, Cohen JC, Hobbs HH. ARTICLE Molecular Characterization of Loss-of-Function Mutations in PCSK9 and Identification of a Compound Heterozygote. *Am J Hum Genet Dallas Am J Hum Genet*. 2006;7979:514-523. doi:10.1086/507488.

20. Young SG. Recent progress in understanding apolipoprotein B. *Circulation*. 1990;82(5):1574-1594. doi:10.1161/01.CIR.82.5.1574.

21. Pulai JI, Neuman RJ, Groenewegen AW, Wu J, Schonfeld G. Genetic heterogeneity in familial hypobetalipoproteinemia: Linkage and non-linkage to the ApoB gene in caucasian families. *Am J Med Genet*. 1998;76(1):79-86. doi:10.1002/(SICI)1096-8628(19980226)76:1<79::AID-AJMG15>3.0.CO;2-M.

22. Linton ME, V R, Young SG. Fam i l ial hypobetal i poprotei nem ia. *J Lipid Res*. 1993;34:521-541.

23. Innerarity TL, Weisgraber KH, Arnold KS, Mahley RW, Krauss RM, Vega GL, Grundy SM.

Familial defective apolipoprotein B-100: low density lipoproteins with abnormal receptor binding. *Proc Natl Acad Sci U S A*. 1987;84(19):6919-6923. doi:10.1073/pnas.84.19.6919.

24. Borén J, Lee I, Zhu W, Arnold K, Taylor S, Innerarity TL. Identification of the low density lipoprotein receptor-binding site in apolipoprotein B100 and the modulation of its binding activity by the carboxyl terminus in familial defective Apo-B100. *J Clin Invest*. 1998;101(5):1084-1093. doi:10.1172/JCI1847.

25. Jia L, Betters JL, Yu L. Niemann-pick C1-like 1 (NPC1L1) protein in intestinal and hepatic cholesterol transport. *Annu Rev Physiol*. 2011;73:239-259. doi:10.1146/annurev-physiol-012110-142233.

26. Myocardial Infarction Genetics Consortium Investigators, Stitziel NO, Won H-H, Morrison AC, Peloso GM, Do R, Lange LA, Fontanillas P, Gupta N, Duga S, Goel A, Farrall M, Saleheen D, Ferrario P, König I, Asselta R, Merlini PA, Marziliano N, Notarangelo MF, Schick U, Auer P, Assimes TL, Reilly M, Wilensky R, Rader DJ, Hovingh GK, Meitinger T, Kessler T, Kastrati A, Laugwitz K-L, Siscovick D, Rotter JI, Hazen SL, Tracy R, Cresci S, Spertus J, Jackson R, Schwartz SM, Natarajan P, Crosby J, Muzny D, Ballantyne C, Rich SS, O'Donnell CJ, Abecasis G, Sunyaev S, Nickerson DA, Buring JE, Ridker PM, Chasman DI, Austin E, Ye Z, Kullo IJ, Weeke PE, Shaffer CM, Bastarache LA, Denny JC, Roden DM, Palmer C, Deloukas P, Lin D-Y, Tang Z, Erdmann J, Schunkert H, Danesh J, Marrugat J, Elosua R, Ardissino D, McPherson R, Watkins H, Reiner AP, Wilson JG, Altshuler D, Gibbs RA, Lander ES, Boerwinkle E, Gabriel S, Kathiresan S. Inactivating mutations in NPC1L1 and protection from coronary heart disease. *N Engl J Med*. 2014;371(22):2072-2082. doi:10.1056/NEJMoa1405386.

27. Goldberg AC, Gidding SS. Knowing the prevalence of familial hypercholesterolemia matters. *Circulation*. 2016;133(11):1054-1057. doi:10.1161/CIRCULATIONAHA.116.021673.

28. Austin MA, Hutter CM, Zimmern RL, Humphries SE. Genetic causes of monogenic heterozygous familial hypercholesterolemia: A HuGE prevalence review. *Am J Epidemiol*. 2004;160(5):407-420. doi:10.1093/aje/kwh236.

29. Hobbs HH, Brown MS, Goldstein JL. Molecular genetics of the LDL receptor gene in familial hypercholesterolemia. *Hum Mutat*. 1992;1(6):445-466. doi:10.1002/humu.1380010602.

30. Williams RR, Hunt SC, Schumacher MC, Hegele RA, Leppert MF, Ludwig EH, Hopkins PN. Diagnosing heterozygous familial hypercholesterolemia using new practical criteria validated by molecular genetics. *Am J Cardiol*. 1993;72(2):171-176. doi:10.1016/0002-9149(93)90155-6.

31. Scientific Steering Committee on behalf of the Simon Broome Register Group. Risk of fatal coronary heart disease in familial hypercholesterolaemia. Scientific Steering Committee on behalf of the Simon Broome Register Group. *BMJ*. 1991;303(6807):893-896. doi:10.1136/bmj.303.6807.893.

32. Hartgers ML, Ray KK, Hovingh GK. New Approaches in Detection and Treatment of Familial Hypercholesterolemia. *Curr Cardiol Rep*. 2015;17(12). doi:10.1007/s11886-015-0665-x.

33. Futema M, Whittall RA, Kiley A, Steel LK, Cooper JA, Badmus E, Leigh SE, Karpe F, Neil HAW, Humphries SE. Analysis of the frequency and spectrum of mutations recognised to cause familial hypercholesterolaemia in routine clinical practice in a UK specialist hospital lipid clinic. *Atherosclerosis*. 2013;229(1):161-168. doi:10.1016/j.atherosclerosis.2013.04.011.

34. Deloukas P, Kanoni S, Willenborg C, Farrall M, Assimes TL, Thompson JR, Ingelsson E, Saleheen D, Erdmann J, Goldstein BA, Stirrups K, König IR, Cazier J-B, Johansson A, Hall AS, Lee J-Y, Willer CJ, Chambers JC, Esko T, Folkersen L, Goel A, Grundberg E, Havulinna AS, Ho WK, Hopewell JC, Eriksson N, Kleber ME, Kristiansson K, Lundmark P, Lyytikäinen L-P, Rafelt S, Shungin D, Strawbridge RJ, Thorleifsson G, Tikkanen E, Van Zuydam N, Voight BF, Waite LL, Zhang W, Ziegler A, Absher D, Altshuler D, Balmforth AJ, Barroso I, Braund PS, Burgdorf C, Claudi-Boehm S, Cox D, Dimitriou M, Do R, Doney ASF, El Mokhtari N, Eriksson P, Fischer K, Fontanillas P, Franco-Cereceda A, Gigante B, Groop L, Gustafsson S, Hager J, Hallmans G, Han B-G, Hunt SE, Kang HM, Illig T, Kessler T, Knowles JW, Kolovou G, Kuusisto J, Langenberg C, Langford C, Leander K, Lokki M-L, Lundmark A, McCarthy MI, Meisinger C, Melander O, Mihailov E, Maouche S, Morris AD, Müller-Nurasyid M, Nikus K, Peden JF, Rayner NW, Rasheed A, Rosinger S, Rubin D, Rumpf MP, Schäfer A, Sivananthan M, Song C, Stewart AFR, Tan S-T, Thorgeirsson G, van der Schoot CE, Wagner PJ, Wells GA, Wild PS, et al. Large-scale association analysis identifies new risk loci for coronary artery disease. *Nat Genet*. 2013;45(1):25-33. doi:10.1038/ng.2480.

35. Sharifi M, Walus-Miarka M, Idzior-Waluś B, Malecki MT, Sanak M, Whittall R, Li KW, Futema M, Humphries SE. The genetic spectrum of familial hypercholesterolemia in south-eastern Poland. *Metabolism*. 2016;65(3):48-53. doi:10.1016/j.metabol.2015.10.018.

36. Taylor A, Wang D, Patel K, Whittall R, Wood G, Farrer M, Neely RDG, Fairgrieve S, Nair D, Barbir M, Jones JL, Egan S, Everdale R, Lolin Y, Hughes E, Cooper JA, Hadfield SG, Norbury G, Humphries SE. Mutation detection rate and spectrum in familial hypercholesterolaemia patients in the UK pilot cascade project. *Clin Genet*. 2010;77(6):572-580. doi:10.1111/j.1399-0004.2009.01356.x.

37. Ned RM, Sijbrands E. Cascade screening for familial hypercholesterolemia (FH). *PLoS Curr*. 2011. doi:10.1371/currents.RRN1238.

38. Nordestgaard BG, Chapman MJ, Humphries SE, Ginsberg HN, Masana L, Descamps OS, Wiklund O, Hegele RA, Raal FJ, Defesche JC, Wiegman A, Santos RD, Watts GF, Parhofer KG, Hovingh GK, Kovanen PT, Boileau C, Averna M, Borén J, Bruckert E, Catapano AL, Kuivenhoven JA, Pajukanta P, Ray K, Stalenhoef AFH, Stroes E, Taskinen MR, Tybjærg-Hansen A. Familial hypercholesterolaemia is underdiagnosed and undertreated in the general population: Guidance for clinicians to prevent coronary heart disease. *Eur Heart J*. 2013;34(45). doi:10.1093/eurheartj/eht273.

39. Schonfeld G, Patterson BW, Yablonskiy D a, Tanoli TSK, Averna M, Elias N, Yue P, Ackerman J. Fatty liver in familial hypobetalipoproteinemia: triglyceride assembly into VLDL particles is affected by the extent of hepatic steatosis. *J Lipid Res*. 2003;44(3):470-478. doi:10.1194/jlr.M200342-JLR200.

40. Tarugi P, Averna M. *Hypobetalipoproteinemia: Genetics, Biochemistry, and Clinical Spectrum*. Vol 54.; 2011. doi:10.1016/B978-0-12-387025-4.00004-2.

41. Sankatsing RR, Fouchier SW, De Haan S, Hutten BA, De Groot E, Kastelein JJP, Stroes ESG. Hepatic and cardiovascular consequences of familial hypobetalipoproteinemia. *Arterioscler Thromb Vasc Biol*. 2005;25(9):1979-1984. doi:10.1161/01.ATV.0000176191.64314.07.

42. Anderson KM, Castelli WP, Levy D. Cholesterol and mortality. 30 years of follow-up from the Framingham study. *JAMA*. 1987;257(16):2176-2180. doi:10.1001/jama.257.16.2176.

43. Schonfeld G, Lin X, Yue P. Familial hypobetalipoproteinemia: Genetics and metabolism. *Cell Mol Life Sci*. 2005;62(12):1372-1378. doi:10.1007/s00018-005-4473-0.

44. Cariou B, Ouguerram K, Zaïr Y, Guerois R, Langhi C, Kourimate S, Benoit I, Le May C, Gayet C, Belabbas K, Dufernez F, Chétiveaux M, Tarugi P, Krempf M, Benlian P, Costet P. PCSK9 dominant negative mutant results in increased LDL catabolic rate and familial hypobetalipoproteinemia. *Arterioscler Thromb Vasc Biol*. 2009;29(12):2191-2197. doi:10.1161/ATVBAHA.109.194191.

45. Martín-Campos JM, Roig R, Mayoral C, Martinez S, Martí G, Arroyo JA, Julve J, Blanco-Vaca F. Identification of a novel mutation in the ANGPTL3 gene in two families diagnosed of familial hypobetalipoproteinemia without APOB mutation. *Clin Chim Acta*. 2012;413(5-6):552-555. doi:10.1016/j.cca.2011.11.020.

46. Sherva R, Yue P, Schonfeld G, Neuman RJ. Evidence for a quantitative trait locus affecting low levels of apolipoprotein B and low density lipoprotein on chromosome 10 in Caucasian families. *J Lipid Res*. 2007;48(12):2632-2639. doi:10.1194/jlr.M700078-JLR200.

47. Yue P, Tanoli T, Wilhelm O, Patterson B, Yablonskiy D, Schonfeld G. Absence of fatty liver in familial hypobetalipoproteinemia linked to chromosome 3p21. *Metabolism*. 2005;54(5):682-688. doi:10.1016/j.metabol.2004.12.013.

48. Willer CJ, Schmidt EM, Sengupta S, Peloso GM, Gustafsson S, Kanoni S, Ganna A, Chen J, Buchkovich ML, Mora S, Beckmann JS, Bragg-Gresham JL, Chang H-Y, Demirkan A, Den Hertog HM, Do R, Donnelly LA, Ehret GB, Esko T, Feitosa MF, Ferreira T, Fischer K, Fontanillas P, Fraser RM, Freitag DF, Gurdasani D, Heikkilä K, Hyppönen E, Isaacs A, Jackson AU, Johansson A, Johnson T, Kaakinen M, Kettunen J, Kleber ME, Li X, Luan J, Lyytikäinen L-P, Magnusson PKE, Mangino M, Mihailov E, Montasser ME, Müller-Nurasyid M, Nolte IM, O'Connell JR, Palmer CD, Perola M, Petersen A-K, Sanna S, Saxena R, Service SK, Shah S, Shungin D, Sidore C, Song C, Strawbridge RJ, Surakka I, Tanaka T, Teslovich TM, Thorleifsson G, Van den Herik EG, Voight BF, Volcik KA, Waite LL, Wong A, Wu Y, Zhang W, Absher D, Asiki G, Barroso I, Been LF, Bolton JL, Bonnycastle LL, Brambilla P, Burnett MS, Cesana G, Dimitriou M, Doney ASF, Döring A, Elliott P, Epstein SE, Eyjolfsson GI, Gigante B, Goodarzi MO, Grallert H, Gravito ML, Groves CJ, Hallmans G, Hartikainen A-L, Hayward C, Hernandez D, Hicks AA, Holm H, Hung Y-J, Illig T, Jones MR, Kaleebu P, Kastelein JJP, et al. Discovery and refinement of loci associated with lipid levels. *Nat Genet*. 2013;45(11):1274-1283. doi:10.1038/ng.2797.

49. Ridker, P.M., Danielson, E., Fonseca F.A. et al. Rosuvastatin to prevent vascular events in men and women with elevated C-reactive protein - An analysis. *N Engl J Med*. 2008;359(21):2195-2207. doi:10.1097/01.ogx.0000344393.60303.0a.

50. Würtz P, Wang Q, Soininen P, Kangas AJ, Fatemifar G, Tynkkynen T, Tiainen M, Perola M, Tillin T, Hughes AD, Mäntyselkä P, Kähönen M, Lehtimäki T, Sattar N, Hingorani AD,

Casas J-P, Salomaa V, Kivimäki M, Järvelin M-R, Davey Smith G, Vanhala M, Lawlor DA, Raitakari OT, Chaturvedi N, Kettunen J, Ala-Korpela M. Metabolomic Profiling of Statin Use and Genetic Inhibition of HMG-CoA Reductase. *J Am Coll Cardiol*. 2016;67(10):1200-1210. doi:10.1016/j.jacc.2015.12.060.

51. Christoffersen M, Tybjærg-Hansen A. Novel genes in LDL metabolism--a comprehensive overview. *Curr Opin Lipidol*. 2015;26(3):179-187. doi:10.1097/MOL.0000000000000175.

52. Cannon CP, Blazing MA, Giugliano RP, McCagg A, White JA, Theroux P, Darius H, Lewis BS, Ophuis TO, Jukema JW, De Ferrari GM, Ruzyllo W, De Lucca P, Im K, Bohula EA, Reist C, Wiviott SD, Tershakovec AM, Musliner TA, Braunwald E, Califf RM. Ezetimibe Added to Statin Therapy after Acute Coronary Syndromes. *N Engl J Med*. 2015;372(25):2387-2397. doi:10.1056/NEJMoa1410489.

53. Blom DJ, Hala T, Bolognese M, Lillestol MJ, Toth PD, Burgess L, Ceska R, Roth E, Koren MJ, Ballantyne CM, Monsalvo ML, Tsirtsonis K, Kim JB, Scott R, Wasserman SM, Stein EA. A 52-Week Placebo-Controlled Trial of Evolocumab in Hyperlipidemia. *N Engl J Med*. 2014;370(19):1809-1819. doi:10.1056/NEJMoa1316222.

54. Radovica-Spalvina I, Latkovskis G, Silamikelis I, Fridmanis D, Elbere I, Ventins K, Ozola G, Erglis A, Klovins J. Next-generation-sequencing-based identification of familial hypercholesterolemia-related mutations in subjects with increased LDL-C levels in a latvian population. *BMC Med Genet*. 2015;16:86. doi:10.1186/s12881-015-0230-x.

55. Futema M, Plagnol V, Whittall RA, Neil HAW, Humphries SE. Use of targeted exome sequencing as a diagnostic tool for Familial Hypercholesterolaemia. *J Med Genet*. 2012;49(10):644-649. doi:10.1136/jmedgenet-2012-101189.

56. Wintjens R, Bozon D, Belabbas K, MBou F, Girardet J-P, Tounian P, Jolly M, Boccara F, Cohen A, Karsenty A, Dubern B, Carel J-C, Azar-Kolakez A, Feillet F, Labarthe F, Gorsky A-MC, Horovitz A, Tamarindi C, Kieffer P, Lienhardt A, Lascols O, Di Filippo M, Dufernez F. Global molecular analysis and APOE mutations in a cohort of autosomal dominant hypercholesterolemia patients in France. *J Lipid Res*. 2016;57(3):482-491. doi:10.1194/jlr.P055699.

57. Futema M, Plagnol V, Li K, Whittall R a, Neil HAW, Seed M, Bertolini S, Calandra S, Descamps OS, Graham C a, Hegele R a, Karpe F, Durst R, Leitersdorf E, Lench N, Nair DR, Soran H, Van Bockxmeer FM, Humphries SE. Whole exome sequencing of familial hypercholesterolaemia patients negative for LDLR/APOB/PCSK9 mutations. *J Med Genet*. 2014;51(8):537-544. doi:10.1136/jmedgenet-2014-102405.

58. Stitziel NO, Peloso GM, Abifadel M, Cefalu AB, Fouchier S, Motazacker MM, Tada H, Larach DB, Awan Z, Haller JF, Pullinger CR, Varret M, Rabès JP, Noto D, Tarugi P, Kawashiri MA, Nohara A, Yamagishi M, Risman M, Deo R, Ruel I, Shendure J, Nickerson DA, Wilson JG, Rich SS, Gupta N, Farlow DN, Neale BM, Daly MJ, Kane JP, Freeman MW, Genest J, Rader DJ, Mabuchi H, Kastelein JJP, Hovingh GK, Averna MR, Gabriel S, Boileau C, Kathiresan S. Exome Sequencing in Suspected Monogenic Dyslipidemias. *Circ Cardiovasc Genet*. 2015;8(2):343-350. doi:10.1161/CIRCGENETICS.114.000776.

59. Fouchier SW, Dallinga-Thie GM, Meijers JCM, Zelcer N, Kastelein JJP, Defesche JC, Hovingh GK. Mutations in STAP1 are associated with autosomal dominant hypercholesterolemia. *Circ Res*. 2014;115(6):552-555. doi:10.1161/CIRCRESAHA.115.304660.

60. Awan Z, Choi HY, Stitziel N, Ruel I, Bamimore MA, Husa R, Gagnon MH, Wang RHL, Peloso GM, Hegele RA, Seidah NG, Kathiresan S, Genest J. APOE p.Leu167del mutation in familial hypercholesterolemia. *Atherosclerosis*. 2013;231(2):218-222. doi:10.1016/j.atherosclerosis.2013.09.007.

61. Marques-Pinheiro A, Marduel M, Rabès J-P, Devillers M, Villéger L, Allard D, Weissenbach J, Guerin M, Zair Y, Erlich D, Junien C, Munnich A, Krempf M, Abifadel M, Jaïs J-P, Boileau C, Varret M. A fourth locus for autosomal dominant hypercholesterolemia maps at 16q22.1. *Eur J Hum Genet*. 2010;18(11):1236-1242. doi:10.1038/ejhg.2010.94.

62. Wang X, Li X, Zhang YB, Zhang F, Sun L, Lin J, Wang DM, Wang LY. Genome-wide linkage scan of a pedigree with familial hypercholesterolemia suggests susceptibility loci on chromosomes 3q25-26 and 21q22. *PLoS One*. 2011;6(10). doi:10.1371/journal.pone.0024838.

63. Varret M, Rabès JP, Saint-Jore B, Cenarro A, Marinoni JC, Civeira F, Devillers M, Krempf M, Coulon M, Thiart R, Kotze MJ, Schmidt H, Buzzi JC, Kostner GM, Bertolini S, Pocovi M, Rosa A, Farnier M, Martinez M, Junien C, Boileau C. A third major locus for autosomal dominant hypercholesterolemia maps to 1p34.1-p32. *Am J Hum Genet*. 1999;64(5):1378-1387. doi:10.1086/302370.

64. Lange LA, Hu Y, Zhang H, Xue C, Schmidt EM, Tang ZZ, Bizon C, Lange EM, Smith JD, Turner EH, Jun G, Kang HM, Peloso G, Auer P, Li KP, Flannick J, Zhang J, Fuchsberger C, Gaulton K, Lindgren C, Locke A, Manning A, Sim X, Rivas MA, Holmen OL, Gottesman O, Lu Y, Ruderfer D, Stahl EA, Duan Q, Li Y, Durda P, Jiao S, Isaacs A, Hofman A, Bis JC, Correa A, Griswold ME, Jakobsdottir J, Smith A V., Schreiner PJ, Feitosa MF, Zhang Q, Huffman JE, Crosby J, Wassel CL, Do R, Franceschini N, Martin LW, Robinson JG, Assimes TL, Crosslin DR, Rosenthal EA, Tsai M, Rieder MJ, Farlow DN, Folsom AR, Lumley T, Fox ER, Carlson CS, Peters U, Jackson RD, Van Duijn CM, Uitterlinden AG, Levy D, Rotter JI, Taylor HA, Gudnason V, Siscovick DS, Fornage M, Borecki IB, Hayward C, Rudan I, Chen YE, Bottinger EP, Loos RJF, S??trom P, Hveem K, Boehnke M, Groop L, McCarthy M, Meitinger T, Ballantyne CM, Gabriel SB, O'Donnell CJ, Post WS, North KE, Reiner AP, Boerwinkle E, Psaty BM, Altshuler D, Kathiresan S, Lin DY, Jarvik GP, Cupples LA, Kooperberg C, Wilson JG, Nickerson DA, et al. Whole-exome sequencing identifies rare and low-frequency coding variants associated with LDL cholesterol. *Am J Hum Genet*. 2014;94(2):233-245. doi:10.1016/j.ajhg.2014.01.010.

65. Lee MH, Lu K, Hazard S, Yu H, Shulenin S, Hidaka H, Kojima H, Allikmets R, Sakuma N, Pegoraro R, Srivastava AK, Salen G, Dean M, Patel SB. Identification of a gene, ABCG5, important in the regulation of dietary cholesterol absorption. *Nat Genet*. 2001;27(1):79-83. doi:10.1038/83799.

66. Cohen JC, Pertsemlidis A, Fahmi S, Esmail S, Vega GL, Grundy SM, Hobbs HH. Multiple rare variants in NPC1L1 associated with reduced sterol absorption and plasma low-density lipoprotein levels. *Proc Natl Acad Sci U S A*. 2006;103(6):1810-1815. doi:10.1073/pnas.0508483103.

67. Reddy M V, Iatan I, Weissglas-Volkov D, Nikkola E, Haas BE, Juvonen M, Ruel I, Sinsheimer JS, Genest J, Pajukanta P. Exome sequencing identifies 2 rare variants for low high-density lipoprotein cholesterol in an extended family. *Circ Cardiovasc Genet*. 2012;5(5):538-546. doi:10.1161/CIRCGENETICS.112.963264.

68. Landrum MJ, Lee JM, Riley GR, Jang W, Rubinstein WS, Church DM, Maglott DR. ClinVar: Public archive of relationships among sequence variation and human phenotype. *Nucleic Acids Res*. 2014;42(D1). doi:10.1093/nar/gkt1113.

69. Yang H, Wang K. Genomic variant annotation and prioritization with ANNOVAR and wANNOVAR. *Nat Protoc*. 2015;10(10):1556-1566. doi:10.1038/nprot.2015.105.

70. Lek M, Tewksbury J, Services H. Analysis of protein-coding genetic variation in 60,706 humans. *Nat Publ Gr*. 2014;536(7616):1-26. doi:http://dx.doi.org/10.1101/030338.

71. Consortium TGte, Ardlie KG, Deluca DS, Segrè A V., Sullivan TJ, Young TR, Gelfand ET, Trowbridge CA, Maller JB, Tukiainen T, Lek M, Ward LD, Kheradpour P, Iriarte B, Meng Y, Palmer CD, Esko T, Winckler W, Hirschhorn JN, Kellis M, MacArthur DG, Getz G, Shabalin AA, Li G, Zhou Y-H, Nobel AB, Rusyn I, Wright FA, Lappalainen T, Ferreira PG, Ongen H, Rivas MA, Battle A, Mostafavi S, Monlong J, Sammeth M, Mele M, Reverter F, Goldmann JM, Koller D, Guigó R, McCarthy MI, Dermitzakis ET, Gamazon ER, Im HK, Konkashbaev A, Nicolae DL, Cox NJ, Flutre T, Wen X, Stephens M, Pritchard JK, Tu Z, Zhang B, Huang T, Long Q, Lin L, Yang J, Zhu J, Liu J, Brown A, Mestichelli B, Tidwell D, Lo E, Salvatore M, Shad S, Thomas JA, Lonsdale JT, Moser MT, Gillard BM, Karasik E, Ramsey K, Choi C, Foster BA, Syron J, Fleming J, Magazine H, Hasz R, Walters GD, Bridge JP, Miklos M, Sullivan S, Barker LK, Traino HM, Mosavel M, Siminoff LA, Valley DR, Rohrer DC, Jewell SD, Branton PA, Sobin LH, Barcus M, Qi L, McLean J, Hariharan P, Um KS, Wu S, Tabor D, et al. The Genotype-Tissue Expression (GTEx) pilot analysis: Multitissue gene regulation in humans. *Science (80-)*. 2015;348(6235):648-660. doi:10.1126/science.1262110.

72. Teslovich, T. Musunuru, K. Smith a. E Al. Biological, clinical and population relevance of 95 loci for blood lipids. *Nature*. 2010;466(7307):707-713. doi:10.1038/nature09270.Biological.

73. Talmud PJ, Shah S, Whittall R, Futema M, Howard P, Cooper JA, Harrison SC, Li K, Drenos F, Karpe F, W. Neil HA, Descamps OS, Langenberg C, Lench N, Kivimaki M, Whittaker J, Hingorani AD, Kumari M, Humphries SE. Use of low-density lipoprotein cholesterol gene score to distinguish patients with polygenic and monogenic familial hypercholesterolaemia: A case-control study. *Lancet*. 2013;381(9874):1293-1301. doi:10.1016/S0140-6736(12)62127-8.

74. Santos RD, Gidding SS, Hegele RA, Cuchel MA, Barter PJ, Watts GF, Baum SJ, Catapano AL, Chapman MJ, Defesche JC, Folco E, Freiberger T, Genest J, Hovingh GK, Harada-Shiba M, Humphries SE, Jackson AS, Mata P, Moriarty PM, Raal FJ, Al-Rasadi K, Ray KK, Reiner Z, Sijbrands EJG, Yamashita S. Defining severe familial hypercholesterolaemia and the implications for clinical management: a consensus statement from the International Atherosclerosis Society Severe Familial Hypercholesterolemia Panel. *Lancet Diabetes Endocrinol*. 2016;8587(16):19-21. doi:10.1016/S2213-8587(16)30041-9.

75. Futema M, Shah S, Cooper JA, Li K, Whittall RA, Sharifi M, Goldberg O, Drogari E, Mollaki V, Wiegman A, Defesche J, D'Agostino MN, D'Angelo A, Rubba P, Fortunato G, Waluš-Miarka M, Hegele RA, Bamimore MA, Durst R, Leitersdorf E, Mulder MT, Van Lennep JER, Sijbrands EJG, Whittaker JC, Talmud PJ, Humphries SE. Refinement of variant selection for the LDL cholesterol genetic risk score in the diagnosis of the polygenic form of clinical familial hypercholesterolemia and replication in samples from 6

countries. *Clin Chem*. 2015;61(1):231-238. doi:10.1373/clinchem.2014.231365.

76. Schunkert H, Bourier F. Deciphering Unexplained Familial Dyslipidemias: Do We Have the Tools? *Circ Cardiovasc Genet*. 2015;8(2):250-252. doi:10.1161/CIRCGENETICS.115.001066.

77. Sch??ffer AA, Lemire M, Ott J, Lathrop GM, Weeks DE. Coordinated conditional simulation with SLINK and SUP of many markers linked or associated to a trait in large pedigrees. *Hum Hered*. 2011;71(2):126-134. doi:10.1159/000324177.

78. van Gijn J, Kerr RS, Rinkel GJE. Subarachnoid haemorrhage. *Lancet*. 2007;369(9558):306-318. doi:10.1016/S0140-6736(07)60153-6.

79. Barker-Collo S, Feigin V. The impact of neuropsychological deficits on functional stroke outcomes. *Neuropsychol Rev*. 2006;16(2):53-64. doi:10.1007/s11065-006-9007-5.

80. Przyklenk K, Bauer B, Ovize M, Kloner RA, Whittaker P. Regional ischemic "preconditioning" protects remote virgin myocardium from subsequent sustained coronary occlusion. *Circulation*. 1993;87(3):893-899. doi:10.1161/01.cir.87.3.893.

81. Bromage DI, Pickard JM, Rossello X, Ziff OJ, Burke N, Yellon DM, Davidson SM. Remote ischaemic conditioning reduces infarct size in animal in vivo models of ischaemia-reperfusion injury: a systematic review and meta-analysis. *Cardiovasc Res*. 2016.

82. Lim S, Hausenloy DJ. Remote ischemic conditioning: From bench to bedside. *Front Physiol*. 2012;3 FEB. doi:10.3389/fphys.2012.00027.

83. Thielmann M, Kottenberg E, Kleinbongard P, Wendt D, Gedik N, Pasa S, Price V, Tsagakis K, Neuhäuser M, Peters J, Jakob H, Heusch G. Cardioprotective and prognostic effects of remote ischaemic preconditioning in patients undergoing coronary artery bypass surgery: a single-centre randomised, double-blind, controlled trial. *Lancet (London, England)*. 2013;382(9892):597-604. doi:10.1016/S0140-6736(13)61450-6.

84. Rahman IA, Mascaro JG, Steeds RP, Frenneaux MP, Nightingale P, Gosling P, Townsend P, Townend JN, Green D, Bonser RS. Remote ischemic preconditioning in human coronary artery bypass surgery: From promise to disappointment? *Circulation*. 2010;122(11 SUPPL. 1). doi:10.1161/CIRCULATIONAHA.109.926667.

85. Hausenloy DJ, Mwamure PK, Venugopal V, Harris J, Barnard M, Grundy E, Ashley E, Vichare S, Di Salvo C, Kolvekar S, Hayward M, Keogh B, MacAllister RJ, Yellon DM. Effect of remote ischaemic preconditioning on myocardial injury in patients undergoing coronary artery bypass graft surgery: a randomised controlled trial. *Lancet*. 2007;370(9587):575-579. doi:Doi: 10.1016/s0140-6736(07)61296-3.

86. Hausenloy DJ, Barrabes JA, Bøtker HE, Davidson SM, Di Lisa F, Downey J, Engstrom T, Ferdinandy P, Carbrera-Fuentes HA, Heusch G, Ibanez B, Iliodromitis EK, Inserte J, Jennings R, Kalia N, Kharbanda R, Lecour S, Marber M, Miura T, Ovize M, Perez-Pinzon MA, Piper HM, Przyklenk K, Schmidt MR, Redington A, Ruiz-Meana M, Vilahur G, Vinten-Johansen J, Yellon DM, Garcia-Dorado D. Ischaemic conditioning and targeting reperfusion injury: a 30 year voyage of discovery. *Basic Res Cardiol*. 2016;111(6):70. doi:10.1007/s00395-016-0588-8.

87. Gonzalez NR, Connolly M, Dusick JR, Bhakta H, Vespa P. Phase i clinical trial for the

feasibility and safety of remote ischemic conditioning for aneurysmal subarachnoid hemorrhage. *Neurosurgery*. 2014;75(5):590-598. doi:10.1227/NEU.0000000000000514.

88. Dickson EW, Reinhardt CP, Renzi FP, Becker RC, Porcaro WA, Heard SO. Ischemic preconditioning may be transferable via whole blood transfusion: Preliminary evidence. *J Thromb Thrombolysis*. 1999;8(2):123-129. doi:10.1023/A:1008911101951.

89. Saxena A, Fish JE, White MD, Yu S, Smyth JWP, Shaw RM, DiMaio JM, Srivastava D. Stromal cell-derived factor-1alpha is cardioprotective after myocardial infarction. *Circulation*. 2008;117(17):2224-2231. doi:10.1161/CIRCULATIONAHA.107.694992.

90. Guyon A. CXCL12 chemokine and its receptors as major players in the interactions between immune and nervous systems. *Front Cell Neurosci*. 2014;8:65. doi:10.3389/fncel.2014.00065.

91. Konstantinov IE, Arab S, Kharbanda RK, Li J, Cheung MMH, Cherepanov V, Downey GP, Liu PP, Cukerman E, Coles JG, Redington AN. The remote ischemic preconditioning stimulus modifies inflammatory gene expression in humans. *Physiol Genomics*. 2004;19(1):143-150. doi:10.1152/physiolgenomics.00046.2004.

92. Shimizu M, Saxena P, Konstantinov IE, Cherepanov V, Cheung MMH, Wearden P, Zhangdong H, Schmidt M, Downey GP, Redington AN. Remote Ischemic Preconditioning Decreases Adhesion and Selectively Modifies Functional Responses of Human Neutrophils. *J Surg Res*. 2010;158(1):155-161. doi:10.1016/j.jss.2008.08.010.

93. Heusch G, Musiolik J, Kottenberg E, Peters J, Jakob H, Thielmann M. STAT5 activation and cardioprotection by remote ischemic preconditioning in humans. *Circ Res*. 2012;110(1):111-115. doi:10.1161/CIRCRESAHA.111.259556.

94. Hepponstall M, Ignjatovic V, Binos S, Monagle P, Jones B, Cheung MHH, d'Udekem Y, Konstantinov IE. Remote Ischemic Preconditioning (RIPC) Modifies Plasma Proteome in Humans. *PLoS One*. 2012;7(11). doi:10.1371/journal.pone.0048284.

95. Helgeland E, Breivik LE, Vaudel M, Svendsen ØS, Garberg H, Nordrehaug JE, Berven FS, Jonassen AK. Exploring the human plasma proteome for humoral mediators of remote ischemic preconditioning--a word of caution. *PLoS One*. 2014;9(10):e109279. doi:10.1371/journal.pone.0109279.

96. Mele M, Ferreira PG, Reverter F, DeLuca DS, Monlong J, Sammeth M, Young TR, Goldmann JM, Pervouchine DD, Sullivan TJ, Johnson R, Segre A V., Djebali S, Niarchou A, Consortium TG, Wright FA, Lappalainen T, Calvo M, Getz G, Dermitzakis ET, Ardlie KG, Guigo R. The human transcriptome across tissues and individuals. *Science (80-)*. 2015;348(6235):660-665. doi:10.1126/science.aaa0355.

97. Li S, Rouphael N, Duraisingham S, Romero-Steiner S, Presnell S, Davis C, Schmidt DS, Johnson SE, Milton A, Rajam G, Kasturi S, Carlone GM, Quinn C, Chaussabel D, Palucka AK, Mulligan MJ, Ahmed R, Stephens DS, Nakaya HI, Pulendran B. Molecular signatures of antibody responses derived from a systems biology study of five human vaccines. *Nat Immunol*. 2014;15(2):195-204. doi:10.1038/ni.2789.

98. Muka T, Koromani F, Portilla E, O'Connor A, Bramer WM, Troup J, Chowdhury R, Dehghan A, Franco OH. The role of epigenetic modifications in cardiovascular disease: A systematic review. *Int J Cardiol*. 2016;212:174-183. doi:10.1016/j.ijcard.2016.03.062.

99. Leenen FAD, Muller CP, Turner JD. DNA methylation: conducting the orchestra from exposure to phenotype? *Clin Epigenetics*. 2016;8(1):92. doi:10.1186/s13148-016-0256-8.

100. Ziller MJ, Gu H, Müller F, Donaghey J, Tsai LT-Y, Kohlbacher O, De Jager PL, Rosen ED, Bennett DA, Bernstein BE, Gnirke A, Meissner A. Charting a dynamic DNA methylation landscape of the human genome. *Nature*. 2013;500(7463):477-481. doi:10.1038/nature12433.

101. Laiwalla AN, Ooi YC, Liou R, Gonzalez NR. Matched Cohort Analysis of the Effects of Limb Remote Ischemic Conditioning in Patients with Aneurysmal Subarachnoid Hemorrhage. *Transl Stroke Res*. 2015. doi:10.1007/s12975-015-0437-3.

102. Topol EJ. Individualized medicine from prewomb to tomb. *Cell*. 2014;157(1):241-253. doi:10.1016/j.cell.2014.02.012.

103. The ENCODE Project Consortium, Dunham I, Kundaje A, Aldred SF, Collins PJ, Davis C a, Doyle F, Epstein CB, Frietze S, Harrow J, Kaul R, Khatun J, Lajoie BR, Landt SG, Lee B-KB-K, Pauli F, Rosenbloom KR, Sabo P, Safi A, Sanyal A, Shoresh N, Simon JM, Song L, Trinklein ND, Altshuler RC, Birney E, Brown JB, Cheng C, Djebali S, Dong X, Ernst J, Furey TS, Gerstein M, Giardine B, Greven M, Hardison RC, Harris RS, Herrero J, Hoffman MM, Iyer S, Kelllis M, Kheradpour P, Lassman T, Li Q, Lin X, Marinov GK, Merkel A, Mortazavi A, Parker SCJSL, Reddy TE, Rozowsky J, Schlesinger F, Thurman RE, Wang J, Ward LD, Whitfield TW, Wilder SP, Wu W, Xi HS, Yip KY, Zhuang J, Bernstein BE, Green ED, Gunter C, Snyder M, Pazin MJ, Lowdon RF, Dillon L a L, Adams LB, Kelly CJ, Zhang J, Wexler JR, Good PJ, Feingold E a, Crawford GE, Dekker J, Elinitski L, Farnham PJ, Giddings MC, Gingeras TR, Guigó R, Hubbard TJTJ, Kellis M, Kent WJ, Lieb JD, Margulies EH, Myers RM, Starnatoyannopoulos J a, Tennebaum S a, Weng Z, White KP, Wold B, Yu Y, Wrobel J, Risk B a, Gunawardena HP, Kuiper HC, Maier CW, et al. An integrated encyclopedia of DNA elements in the human genome. *Nature*. 2012;489(7414):57-74. doi:10.1038/nature11247.

104. Szklarczyk D, Franceschini A, Wyder S, Forslund K, Heller D, Huerta-Cepas J, Simonovic M, Roth A, Santos A, Tsafou KP, Kuhn M, Bork P, Jensen LJ, Von Mering C. STRING v10: Protein-protein interaction networks, integrated over the tree of life. *Nucleic Acids Res*. 2015;43(D1):D447-D452. doi:10.1093/nar/gku1003.

Chapter 2

Family-specific aggregation of lipid GWAS variants confers the susceptibility to familial hypercholesterolemia in a large Austrian family

Family-specific aggregation of lipid GWAS variants confers the susceptibility to familial hypercholesterolemia in a large Austrian family

Elina Nikkola, BS[1], Arthur Ko, BS[1,2], Marcus Alvarez, BS[1], Rita M. Cantor, PhD, Professor[1], Kristina Garske, BS[1], Elliot Kim[1], Stephanie Gee[1], Alejandra Rodriguez, BS[1], Reinhard Muxel, MD[3], Niina Matikainen, MD, PhD[4,5,6], Sanni Söderlund, MD, PhD[5,6], Mahdi M. Motazacker, PhD[7], Jan Borén, MD, PhD, Professor[8], Claudia Lamina, PhD[9], Florian Kronenberg, MD, Professor[9], Wolfgang J. Schneider, PhD, Professor[10], Aarno Palotie, MD, PhD, Professor[11,12], Markku Laakso, MD, PhD, Professor[13], Marja-Riitta Taskinen, MD, PhD, Professor emerita[5,6], Päivi Pajukanta, MD, PhD, Professor[1,2,14]

1. Department of Human Genetics, David Geffen School of Medicine at UCLA, Los Angeles, USA
2. Molecular Biology Institute at UCLA, Los Angeles, USA
3. General Practice Centre, Austria
4. Endocrinology, Abdominal Centre, Helsinki University Hospital, Finland
5. Heart and Lung Center, Helsinki University Hospital
6. Research Programs Unit, Diabetes and Obesity, University of Helsinki, Finland
7. Department of Clinical Genetics, Academic Medical Center at the University of Amsterdam, The Netherlands
8. Department of Molecular and Clinical Medicine/Wallenberg Laboratory, University of Gothenburg and Sahlgrenska University Hospital, Sweden
9. Division of Genetic Epidemiology, Medical University of Innsbruck, Austria
10. Department Medical Biochemistry, Medical University Vienna and Max F. Perutz Laboratories, Austria
11. Institute for Molecular Medicine, University of Helsinki, Finland
12. The Broad Institute of MIT and Harvard, Cambridge, MA, USA and Massachusetts General Hospital, Boston, MA, USA
13. Institute of Clinical Medicine, Internal Medicine, University of Eastern Finland and Kuopio University Hospital, Kuopio, Finland
14. Bioinformatics Interdepartmental Program, UCLA, Los Angeles, California, USA

Running title: Family-specific lipid GWAS SNPs identified for FH

The authors declare no conflict of interest.

Corresponding author:
Paivi Pajukanta, MD, PhD, Professor
Department of Human Genetics
David Geffen School of Medicine at UCLA
Gonda Center, Room 6335B
695 Charles E. Young Drive South
Los Angeles, California 90095-7088, USA
Office phone: 1-310-267-2011
Cell phone: 1-310-435-6422
Fax: 1-310-794-5446
Email: ppajukanta@mednet.ucla.edu

Abstract

Hypercholesterolemia confers susceptibility to cardiovascular disease (CVD). Both serum total cholesterol (TC) and LDL-cholesterol (LDL-C) exhibit a strong genetic component (heritability estimates 0.41-0.50). However, a large part of this heritability remains hidden and cannot be explained by the variants identified in recent large scale genome-wide association studies (GWAS) on lipids. To find genetic causes leading to high LDL-C levels and ultimately CVD, we utilized linkage analysis followed by whole-exome sequencing and genetic risk score analysis in a large Austrian family presenting with autosomal dominant inheritance for familial hypercholesterolemia (FH). We did not find evidence for genome-wide significant linkage for LDL-C or causative variants in the known FH genes, rather we discovered a particular family-specific combination of 10 GWAS LDL-C SNPs (p=0.02 by permutation), and putative less severe familial hypercholesterolemia mutations in the *LDLR* and *APOB* genes in a subset of the affected family members. Separately, high Lp(a) levels observed in one branch of the family were explained primarily by the LPA locus, including short (<23) Kringle IV repeats and rs3798220. Taken together, some forms of FH may be explained by family-specific combinations of LDL-C GWAS SNPs.

Keywords: familial hypercholesterolemia (FH), LDL-C, Genetic risk score (GRS)

Introduction

High levels of serum total cholesterol (TC) and especially low-density lipoprotein cholesterol (LDL-C) predispose to cardiovascular disease (CVD), the major cause of death worldwide (1). Genetics plays a major role in CVD (heritability estimates 0.38-0.57) (2,3). However, variants identified in extensive genome-wide association studies (GWAS) explain only 6-20% of the variance in lipid traits and even less of CVD (4). This missing heritability may partially be explained by rare and private variants, and thus large families with several affected individuals without risk variants in the known familial hypercholesterolemia (FH) genes may help identify new genes causing Mendelian forms of dyslipidemia or other inherited mechanisms contributing to high LDL-C.

FH affects 1 in 200-600 people (5). To date there are only a handful of genes known to cause FH, including *LDLR*, *APOB*, *PCSK9* and *LDLRAP1* (6). However, it is estimated that only approximately 20-60% of FH subjects exhibit a causal variant within these four genes (7–9), suggesting that variants in these genes do not explain all cases of familial hypercholesterolemia.

To find mutations leading to high LDL-cholesterol and ultimately CVD, we systematically screened both rare coding and common genomic variants in a large Austrian dyslipidemic family exhibiting elevated TC and LDL-C levels, in addition to elevated lipoprotein a (Lp(a)) levels in one branch of the family. All affected elderly family members had suffered a cardiovascular event in the past, and the index case did not have known FH variants in *LDLR* or *APOB*.

Combining linkage analysis with whole-exome sequencing has become a common approach to pinpoint candidate chromosomal regions and specific variants for Mendelian diseases (10,11). We first genotyped the family members using a genome-wide SNP array to cover the common variants, and then exome-sequenced the family members to capture their coding variants. We

screened for mutations in the known FH genes, performed a genome-wide linkage analysis, and assessed the coding variants present predominantly in the affected individuals for functional predictions. Since no genome-wide significant linkage peaks or mutations in the known FH genes were found, we estimated genetic risk scores using all common GWAS SNPs previously associated with LDL-C (12) and identified a family-specific combination of 10 LDL-C GWAS variants, contributing to the high LDL-C levels in this family.

Materials and Methods

We first searched for a possible monogenic cause for FH in a large Austrian pedigree using a linkage analysis, followed by an exome sequencing analysis and subsequent variant screening in existing European cohorts. We also comprehensively analyzed all variants identified in the known FH genes (6). We then searched for a potential polygenic cause for FH in this family by performing a genetic risk score analysis of the known LDL-C GWAS variants (12).

Study samples

The study sample consists of 16 individuals from a large Austrian dyslipidemic family (Figure 1; for clinical characteristics, see Table 1). DNA was extracted from the blood, and clinical phenotypes were measured using standard protocols. The number of apolipoprotein (a) Kringle IV (apo(a) KIV) repeats were measured by SDS-agarose electrophoresis followed by immunoblotting, as described previously (13). Lp(a) concentrations were measured by an ELISA, as recently described (14). Phenotypes included age, sex, status of statin medication, total cholesterol (TC), low-density lipoprotein cholesterol (LDL-C), Lp(a), and the number of apo(a) KIV repeats. All family members gave a written informed consent, and the study was approved by the local ethic committees.

Validation cohorts

To validate our findings, we utilized genome-wide genotyping data from the METabolic Syndrome In Men (METSIM) cohort (n=10,197) (15) and the European exome sequencing database of type 2 diabetes concortium (n~13,000) (16) for the association with LDL-C.

Genome-wide SNP genotyping and whole exome sequencing

We performed genotyping using a genome-wide SNP panel (Illumina HumanCoreExome-12v1-1) as well as exome sequencing of all available affected and unaffected family members using the Agilent SureSelect All exon 50-Mb capture with the Illumina Hiseq2500 platform employing 75bp paired-end reads, resulting in a mean coverage of 75X, and capturing ~91% of the targeted regions with ≥10X coverage. We aligned and called the variants using BWA (17) and GATK (18). We checked the data quality, including the call rate of the SNPs, gender check based on X chromosome SNPs, and heterozygosity rate using PLINK (19) as well as pedigree consistency using the Mendel software packages (20).

Linkage analysis

We first estimated the expected maximum LOD score (EMLOD) based on the pedigree structure and binary LDL-C affection status by performing a simulation analysis using the fastSLINK package (21). In this simulation, we employed the same penetrance model that was used in the actual linkage analysis (see below). To identify linked regions for LDL-C and Lp(a), we performed a genome-wide parametric two-point linkage analysis for each trait using the Mendel software (20). Linkage analysis was performed utilizing ~95K high-quality (genotyped in all family members) and informative (MAF > 10% in the family (>3 carriers)) SNPs, spaced ~25 kb apart throughout the genome. For LDL-C, we employed an autosomal dominant model with a penetrance of 0.95 and phenocopy rate of 0.001. As Lp(a) has a strong genetic component, we also performed a linkage analysis for Lp(a) and tested the variants at the *LPA* locus. For this

analysis, we used a highly penetrant autosomal dominant model with penetrance of 0.99 and a phenocopy rate 0.0001.

Variant filtering

We focused on the functional variants (nonsynonymous and splice site variants) fulfilling the following criteria: minor allele frequency (MAF) ≤10%; location in the regions with a LOD score ≥1.0; and present predominantly in the affected individuals (high LDL-C or high Lp(a)).

Genetic risk score analysis

We calculated weighted risk scores of the known 47 GWAS LDL-C variants (12) in the METSIM cohort (n=9,836) and family members. For each LDL-C associated locus, we selected the SNP with the lowest p-value and weighted each risk allele with the beta effect size established by Willer et al. from ~180,000 individuals (12). The selected SNPs, including their risk alleles and weights are listed in Supplementary Table 1. We first calculated the risk scores for each individual in the METSIM cohort, and then compared the risk scores of the affected family members with the estimated population percentiles obtained in the METSIM cohort.

Permutation analyses

We performed a permutation analysis for the 10 family-specific GWAS SNPS by randomly selecting 20 individuals with LDL-C >75th percentile (the LDL-C cut off ≥ 3.9 mmol/l) and 20 individuals with LDL-C <50th percentile (the LDL-C cut off ≤ 3.5 mmol/l) from the METSIM cohort. We calculated how many times the allele frequency difference is larger for all 10 SNPs between the METSIM individuals with high LDL-C and normal LDL-C than the 15% difference observed for all 10 SNPs in the family.

We performed an additional permutation of the risk scores by randomly selecting 100 times a 10-SNP set from the 47 LDL-C increasing SNPs (12). We constructed new risk scores weighted

by beta and estimated the percentiles in the METSIM population for each of the 10-SNP sets. We then calculated how many times the average risk score of the affected individuals would be in the $\geq 90^{th}$ percentile.

Evaluation of the known FH variants and genes

We screened all individuals for variants in the four previously known FH genes. (6)

Results

In this study we aimed to identify the variant(s) for an autosomal dominant type of inheritance of high LDL-C levels in a large Austrian dyslipidemic family. The affected family members had an average pre-statin LDL-C level of 5.56 mmol/l (range 4.20-9.20 mmol/l), and four siblings from the first generation had suffered a cardiovascular event (Figure 1). One branch of the family also exhibited 4 individuals with extremely high Lp(a) levels (66 mg/dl-113 mg/dl, i.e. all above the 90^{th} percentile), a known independent risk factor for CVD (22,23). We first performed a linkage study followed by exome-sequencing analysis to find novel variants co-segregating with high LDL-C status in the family. Lp(a) levels were investigated for variants at the *LPA* locus. We evaluated our identified LDL-C variants for association in existing large European cohorts, and calculated the weighted genetic risk scores for LDL-C using genome-wide significant LDL-C variants from the Willer et al. 2013 meta-GWAS study (12), utilizing the METSIM cohort as a reference panel. Lastly, we systematically screened all variants we identified in the known FH genes, *LDLR, APOB, PSCK9,* and *LDLRAP1*, for co-segregation with high LDL-C status among the family members.

Linkage analysis followed by a variant screening did not pinpoint one causal locus for high LDL-C

We performed a two-point parametric linkage analysis for LDL-C by analyzing ~95K SNPs (~25 kb apart) throughout the genome. The estimated maximum LOD score for this family was 3.3 using a simulation analysis (fastSLINK). We observed 17 loci with a LOD score >1.0, with the highest maximum LOD score of 1.9 on chromosome 17, suggesting a polygenic rather than monogenic origin of the high LDL-C levels in this family. To systematically fine-map the chr17 region that yielded the highest LOD score of 1.9, we used all available SNPs within the region with suggestive LOD scores (~3 Mb) to perform a regional refined two point linkage analysis. We identified a total of 21 SNPs with LOD scores between 1.9 and 2.1 spanning a 1.7 Mb region, but none of the SNP resulted in a LOD score close to the estimated theoretical maximum LOD score of 3.3. We then searched for potential functional variants (nonsynonymous and splice site variants), residing in all of the 17 genomic regions with a LOD >1.0 and predominantly present in the affected individuals. Within the 17 loci, we identified 6 potential functional variants (Figure 2a); however, none of these variants were robustly associated with quantitative LDL-C in the large European cohorts (all p-values >0.006). See Supplementary Table 2 for a list of all potential functional variants predominantly present in affected family members.

Genetic risk score analysis of known LDL-C loci identified a family-specific combination of 10 risk variants

Out of the 60 independent LDL-C associated risk loci identified by Willer et al. 2013 (12), 47 SNPs (or their LD proxies, r^2 >0.95) were successfully genotyped for 9,836 METSIM individuals and 16 family members. We constructed the overall genetic risk scores by calculating the sum of number of risk alleles, weighted by the beta established by Willer et al. 2013 of each of the 47

SNPs for every individual. The weighted LDL-C risk score observed in the METSIM cohort was correlated with serum LDL-C levels (Pearson's correlation = 0.28, $p < 2.2 \times 10^{-16}$), after removing the statin users (n = 2,749). For the calculations of the population percentiles of the genetic risk scores, we included all METSIM participants (n=9,836). The 50th percentile of the LDL-C risk scores in the METSIM cohort was 3.55, whereas the average of the affected family members was 3.72 (~75th percentile), suggesting a stronger predisposition for high LDL-C in the family based on the common LDL-C GWAS variants. The 75th percentile is likely a conservative estimate as all METSIM individuals are men.

To better understand the effect of the each LDL-C GWAS risk variant in this family, we further identified and investigated 10 variants with the highest difference in the average number of risk variants (>0.30, corresponding to a MAF difference of ~15%) between the affected and non-affected family members (Table 2). To evaluate if this 10-SNP combination was family-specific, we performed a permutation analysis by randomly selecting 20 individuals with high LDL-C (>75th percentile) and 20 individuals with LDL-C <50th percentile from the METSIM cohort and observed no similar allele frequency difference between all of the 10 LDL-C GWAS SNPs among these subjects with low and high LDL-C levels using 100 permutations ($p<0.01$). This suggests that the distinct combination of a large difference in allele frequencies with these 10 SNPs is specific for this family, and does not appear commonly in the population by chance alone. Next, we derived the risk scores using the sum of the weighted betas of these 10 SNPs (Table 2). The new 10-SNP-weighted LDL-C risk score of the METSIM participants was correlated with serum LDL-C levels (Pearson's correlation = 0.12, $p < 2.2 \times 10^{-16}$), after removing the statin users. As above, for the calculations of the population percentiles of the genetic risk scores, we included all METSIM participants (n=9,836). The average risk score in the METSIM population sample was 0.58 (50th percentile), whereas the average risk score of the affected family members was 0.76 (>0.90th percentile) and of the unaffected family members 0.42

(<0.20th percentile) (Table 1 and Supplementary Figure 1), respectively, further suggesting that the combination of the 10 SNPs is contributing to the high LDL-C levels in this family.

To determine whether this type of aggregation would appear by chance, selecting any set of LDL-C-raising GWAS SNPs, we first calculated the risk scores using the well-established Global Lipid Genetic Consortium 12-SNP LDL-C gene score calculation (24). Indeed, the average risk score of the affected family members was in the 7th decile and the average of the unaffected family members was in the 4th decile when compared to the percentiles in the Whitehall II controls (24). Three of the SNPs were overlapping between the risk scores. Second, we further performed 100 permutations by randomly selecting 10-SNP combinations from the 47 LDL-C-increasing genome-wide GWAS SNPs and calculated how many times the average risk score of the affected family members is $\geq$ 90th percentile. We observed this phenomenon only with two sets (p=0.02), of which both sets had $\geq$ 5 SNPs overlapping with the 10-SNP family-specific risk score we originally identified. These additional risk score permutations suggest that randomly selecting other sets of LDL-C GWAS SNPs does not present as high a risk as the actual 10 family-specific SNP combination, unless at least 5 or more SNPs are included from the family-specific risk score SNPs.

We also assessed the significance of the individual 10 SNPs included in the new risk score analysis within the family by obtaining empirical p-values derived by swapping the low and high LDL-C status within the family in a permutation analysis using PLINK. We observed that three (rs12916 at *HMGCR*, rs1564348 at *SLC22A1*, and rs646776 near *CELSR2*) of the 10 SNPs are significant (p<0.05) after random label-swapping (Table 1 and Supplementary Figure 1). Furthermore, we observed that rs12916, a GWAS variant that resides in the statin target gene *HMGCR*, was only present in its homozygote form C/C in the affected family members (Table 1 and Supplementary Figure 1).

High Lipoprotein (a) (Lp(a)) levels are likely explained by the known genetic variants in the LPA locus

Lp(a) is one of the most heritable (heritability = 0.91±0.01) blood biomarkers (25,26), and is largely regulated by genetic variants at the *LPA* locus on chromosome 6q27, including the number of Kringle (IV) repeats and two SNPs in LD (rs3798220 and rs10455872) which together explain 30-70% of the Lp(a) variation (23). In the family, the individuals with high Lp(a) also had low number of Kringle (IV) repeats (<23) (Table 1). Furthermore, we identified a well-known Lp(a) variant, rs3798220, in the *LPA* locus using linkage and subsequent variant analysis (Figure 2b). This variant is well known to be associated with high Lp(a) concentrations (26); has a MAF of 2% in the general population; and tags a small proportion of small apo(a) isoforms with less than 23 K-IV repeats (27). Taken together, these data suggest that the high Lp(a) levels observed in a branch of the family are very likely explained by variants at the *LPA* locus.

Variants in the known FH genes may explain high LDL-C levels in one family branch

The index case had been previously screened negative for the known FH variants in FH genes (*LDLR*, *APOB*, *PCSK9* and *LDLRAP1*). We screened all family members for the known FH genes and identified a total of 87 variants, of which 19 were non-synonymous or splice site variants (Supplementary Table 3). None of the variants fully segregated with the high LDL-C status. However, we identified two splice site variants (rs72658867, MAF = 0.0039 and rs72658861, MAF = 0.0052) and one non-synonymous variant (rs45508991 [T726I], MAF = 0.0018) in *LDLR* in one family branch (Supplementary Table 3). All of these variants have been previously implicated in FH, but do not consistently co-segregate with hypercholesterolemia (28), suggesting that another variant must be present for these *LDLR* variants to be pathogenic, as previously proposed (29,30).

Similarly, we identified 3 potentially pathogenic non-synonymous variants (rs1801695 [A4481T], MAF = 0.026, rs61742247 [S1613T], MAF = 0.013 and rs1801701 [R3638Q], MAF = 0.077) (Supplementary Table 3) in *APOB*, of which rs1801701 has been implicated for TC in a previous GWAS (31). Interestingly, these *APOB* variants appeared in the same branch as the *LDLR* variants described above, with 3 affected family members sharing a combination of these *LDLR* and *APOB* variants (Table 1, Supplementary Table 3 and Supplementary Figure 1). We postulate that in order to have an impact on the ApoB metabolism, and furthermore on TC and LDL-C levels, these *LDLR* and *APOB* variants may need to appear as a risk combination or require other GWAS LDL-C variants as a haplotypic background.

Discussion

Our comprehensive analysis of a large Austrian family with phenotypical familial hypercholesterolemia (FH) showed evidence of a specific polygenic contribution to high LDL-C. The linkage analysis did not pinpoint to a single genetic locus for high LDL-C; rather, we found 17 loci with a LOD score >1.0, implying that it is likely several loci contribute to the high LDL-C levels in this family. Consistent with that, our risk score analyses followed by a permutation analysis identified a combination of 10 LDL-C GWAS SNPs specific for polygenic FH in this family. In addition, a systematic examination of the variants in the known FH genes resulted in the identification of possible less severe FH mutations in the *LDLR* and *APOB* genes in a subset of the affected family members, in line with the previous hypothesis (28,29) that specific *LDLR* and *APOB* coding variants may only become pathogenic in the presence of an additional risk variant in these FH genes. Because three of the affected family members carried both *LDLR* and *APOB* risk combinations, we postulate that small functional defects in both genes, whose biological functions are tightly bound, escalate the effects and contribute to the high LDL-C levels in these individuals. For example, one of the *APOB* variants (R3638Q) resides in the C-terminus of apoB100, a region known to be regulating LDL receptor binding (32).

Recent evidence suggests that FH is a heterogeneous disorder that can be caused by monogenic or polygenic mechanisms, including rare variants at the traditional FH genes or multiple common variants at the LDL-C GWAS loci and other genes (7). We did not identify FH-causing mutations in the known FH genes tightly co-segregating with LDL-C, and our linkage study combined with exome sequence analysis did not pinpoint a single monogenic causative variant or gene. When we evaluated the effects of the weighted LDL-C risk scores of 47 LDL-C GWAS SNPs collectively present in the family members on LDL-C levels, we observed that the affected members had a significantly higher average risk score than the reference population ($p = 0.001$). However, the risk scores were not in the top 90^{th} percentile, which has previously been used to distinguish polygenic and monogenic forms of FH (24). Interestingly, however, we found a combination of 10 variants at the LDL-C GWAS loci among the affected members of this particular family ($p < 0.01$ by permutation). The risk scores constructed using only these 10 variants accelerated the affected individuals to the 90^{th} percentile of LDL-C.

Among the 10 family-specific GWAS variants (Table 2), we observed that the homozygous risk genotype C/C of the *HMGCR* variant (rs12916) was exclusively present in the affected family members. We also observed that rs12916 is located in the region resulting in a LOD score of 1.4 in this family. HMGCR is a rate-limiting enzyme in cholesterol biosynthesis and the main target of statin therapy. Given the relatively large effect size of this GWAS variant (beta=0.07) for LDL-C (12) and the previous evidence that the rs12916 is a liver eQTL (33), it is likely that the elevation of plasma LDL-C levels due to the C allele is caused by augmented *HMGCR* expression and the subsequent increased cholesterol synthesis in the liver. The increased cholesterol synthesis in turn activates a feedback mechanism that inhibits the uptake of LDL-C from blood via the LDL receptor. Interestingly, a recent longitudinal metabolomics study observed that the carriers of the protective T allele exhibit a similar lipidomics profile as observed in individuals who have started statin therapy (34).

Our study has several limitations. Analysis of only one family does not provide information that could be directly extrapolated to the entire Austrian population. However, our findings further demonstrate the genetic complexity of FH in individuals without the known FH mutations. This type of presentation can clearly complicate the diagnosis and identification of hypercholesterolemic individuals in early stages of disease, emphasizing the importance of family-based evaluation of FH. We showed a polygenic effect that included variants residing in regions with LOD scores >1.0 and a combination of 10 LDL-C GWAS SNPs aggregating in the affected family members. We hypothesize that most of the FH families without a single known pathogenic mutation will exhibit a specific combination of the LDL-C GWAS variants that can be distinguished if extensive family data is available. We recognize that it is still possible that we missed the causal variant(s) since no whole genome sequencing was performed and the causal variant might reside outside the protein coding regions or be a large copy number variation, not studied here. This scenario is, however, less likely given our negative linkage screening that, based on our simulation, had adequate power to identify a single monogenic variant.

In summary, our linkage study followed by exome sequencing and a GWAS LDL-C risk score analysis supports a specific polygenic cause for hypercholesterolemia in this Austrian family. Potential cascade testing of identified variants in the third generation of this family might provide valuable information regarding who should be followed up for early treatment of hypercholesterolemia. Our study demonstrates the importance of using family-wide genetic data, when available, in future personalized medicine initiatives of complex diseases. For example, in other FH families without the known FH mutations, a similar approach could be used for establishing a family-specific polygenic hypercholesterolemia diagnosis when sufficient numbers of affected and unaffected family members are available for identification of a family-specific set of LDL-C increasing GWAS SNPs that exceed the 90^{th} risk score percentile in the

particular population. Subsequently, the family's younger generations could be tested for these variants to provide an earlier personalized diagnosis and potential treatment.

Acknowledgements

We thank the Austrian family and METSIM individuals who participated in this study. We also thank Helinä Perttunen-Nio and Eija Hämäläinen for the laboratory technical assistance. This work was supported by NIH grants HL-095056 and HL-28481; EU-project RESOLVE (No. 305707); the Sigrid Juselius Foundation; and HUCH Research Foundation, Helsinki, Finland.

Conflict of interest

The authors declare no conflict of interest.

Table 1. Clinical characteristics and genetic findings contributing to high LDL-C in the family members.

Individual ID	TC (mmol/l)	LDL-C (mmol/l)	Statins	Lp(a) mg/dl	KR Allele 1	KR Allele 2	Known *LPA* variant (rs3798220), risk allele C, MAF=0.02	Genetic risk score *	Family specific genetic risk score of 10 SNPs (see Table 2)	*HMGCR* (rs12916), risk allele C, MAF=0.43	*SLC22A1* (rs1564348), risk allele C, MAF=0.15	*APOB* and *LDLR* variant combinations **
7711	(4.32)†	(2.16)†	Yes	4	24 (85%)$	36						
7724	6.8 (5.93)†	5.00 (3.90)†	Yes	6	36	0	T/T	3.83 (86th)	0.89 (>99th)	C/C	C/T	
7725	7.00 (5.57)	5.36 (3.19)	Yes	2	36	0	T/T	3.60 (58th)	0.85 (>99th)	C/C	C/T	
7727	6.64	4.54	No	10	36	0	T/T	3.86 (89th)	0.69 (83th)	C/C	C/T	
7729	9.6	6.38	No	3	30	0	T/T	3.79 (82th)	0.80 (97th)	C/C	T/T	
773	(4.52)	(2.96)	Yes	9	36	0	T/T	3.79 (82th)	0.89 (>99th)	C/C	C/T	
7749	7.60 (4.92)	5.00 (2.87)	Yes	4	23 (80%)	36	T/T	3.61 (59th)	0.73 (90th)	C/T	C/T	
7772	7.57	5.11	No	4	26 (10%)	36	T/T	3.91 (92th)	0.78 (96th)	C/C	C/T	
7775	8.98	5.72	No	77	20	0	T/C	3.74 (77th)	0.78 (96th)	C/C	T/T	Yes
7776	6.80 (6.09)	4.20 (3.56)	Yes	90	20	0	T/C	3.41 (30th)	0.51 (32th)	C/T	T/T	Yes
7777	11.70 (5.06)	9.20 (1.94)	Yes	113	20	0	T/C	3.67 (68th)	0.67 (80th)	T/T	T/T	
7792	7.77	5.45	No	7	23 (80%)	36	T/T	3.70 (72th)	0.78 (96th)	C/C	C/T	Yes
776	7.2	3.32	No	11	30 (85%)	36	T/T	3.66 (66th)	0.58 (52th)	C/T	C/T	
7726	5.92	2.82	No	7	30	0	T/T	3.64 (64th)	0.48 (25th)	C/T	T/T	
778	5.03	2.4	No	5	23 (95%)	26	T/T	3.80 (83th)	0.54 (40th)	C/T	T/T	
7789	4.49	2.34	No	66	20	0	T/C	3.32 (20th)	0.35 (6th)	T/T	T/T	
7793	4.38	2.57	No	0	0	0	T/T	2.82 (2nd)	0.32 (4th)	C/T	T/T	

TC, total cholesterol; LDL-C, low density lipoprotein cholesterol; † (measured while on statins); KR, Kringle IV repeats; $ (percentage of allele 1); *Weighted genetic risk score of 47 LDL-C GWAS SNPs; percentile is in parenthesis after risk scores;**Specific *APOB* and *LDLR* variants listed in supplementary table 3; Grey highlight=High LDL-C status.

Table 2. The 10 family-specific LDL-C GWAS variants. Numbers in the matrix under the individual IDs represents number of risk alleles in that particular SNP.

Chr	Rs #	Risk	Alt	beta	P-value	MAF	Family members with high LDL-C										Average # of risk alleles in affected	Family members with normal LDL-C				Average # of risk alleles in unaffected	Difference in average # of risk alleles	Gene	Type
							7777	7776	7772	7775	7749	7725	7729	7792	7724	773	7727		7726	778	7789	7793				
1	rs646776	t	c	0.160	1.63E-272	0.79	2	2	2	2	2	2	2	2	2	2	2	2.00	2	2	1	1	1.50	0.50	*CELSR2*	Intergenic
5	rs12916	c	t	0.073	7.79E-78	0.43	0	1	2	2	1	2	2	2	2	2	2	1.64	1	1	0	1	0.75	0.89	*HMGCR*	3'-UTR
6	rs1564348	c	t	0.048	2.76E-21	0.15	0	0	1	0	1	1	1	1	1	1	1	0.73	0	0	0	0	0.00	0.73	*SLC22A1*	Intronic
8	rs10102164	a	g	0.032	3.74E-11	0.17	2	1	1	1	0	0	1	2	1	0	0	0.82	0	1	0	0	0.25	0.57	*RP11-53M11.3*	Intergenic
8	rs2954029	a	t	0.056	2.1E-50	0.53	1	1	1	1	2	2	1	1	1	2	2	1.36	1	0	2	0	0.75	0.61	*RP11-136O12.2*	Intergenic
11	rs174583	c	t	0.052	7.0E-41	0.63	1	0	0	1	0	2	1	1	2	1	0	0.82	0	0	1	0	0.25	0.57	*FADS2*	Intronic
11	rs11220462	a	g	0.059	6.61E-21	0.14	1	0	1	2	1	1	1	1	1	2	0	1.00	0	1	0	0	0.25	0.75	*ST3GAL4*	Intronic
17	rs7225700	c	t	0.030	3.56E-13	0.67	1	0	2	0	2	1	1	1	2	1	1	1.09	1	1	0	1	0.75	0.34	*RP11-290H9.4*	Intergenic
19	rs492602	g	a	0.029	9.42E-14	0.43	2	1	1	2	2	0	1	0	1	1	1	1.09	0	1	1	1	0.75	0.34	*FUT2*	Synonymous
20	rs2328223	c	a	0.030	5.63E-09	0.25	1	0	1	0	0	1	1	0	1	1	0	0.55	0	0	0	1	0.25	0.30	*RP5-905G11.3*	Intergenic

LDL-C, Low-density lipoprotein cholesterol; Chr, chromosome; Risk, LDL-C increasing variant; Alt, alternative variant; MAF, minor allele frequency

Figures

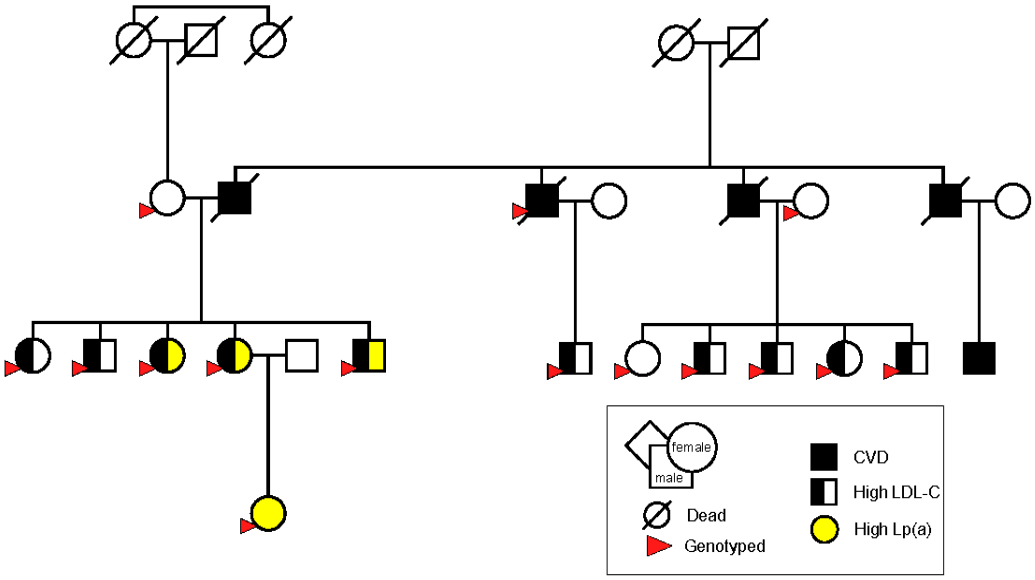


Figure 1. The Austrian hypercholesterolemia family showing an autosomal dominant type of inheritance. The figure includes only the family members who had given the consent for blood sampling and DNA analyses. The pedigree was drawn using CraneFoot (Mäkinen et al. 2005). The box shows the explanations of the drawing scheme.

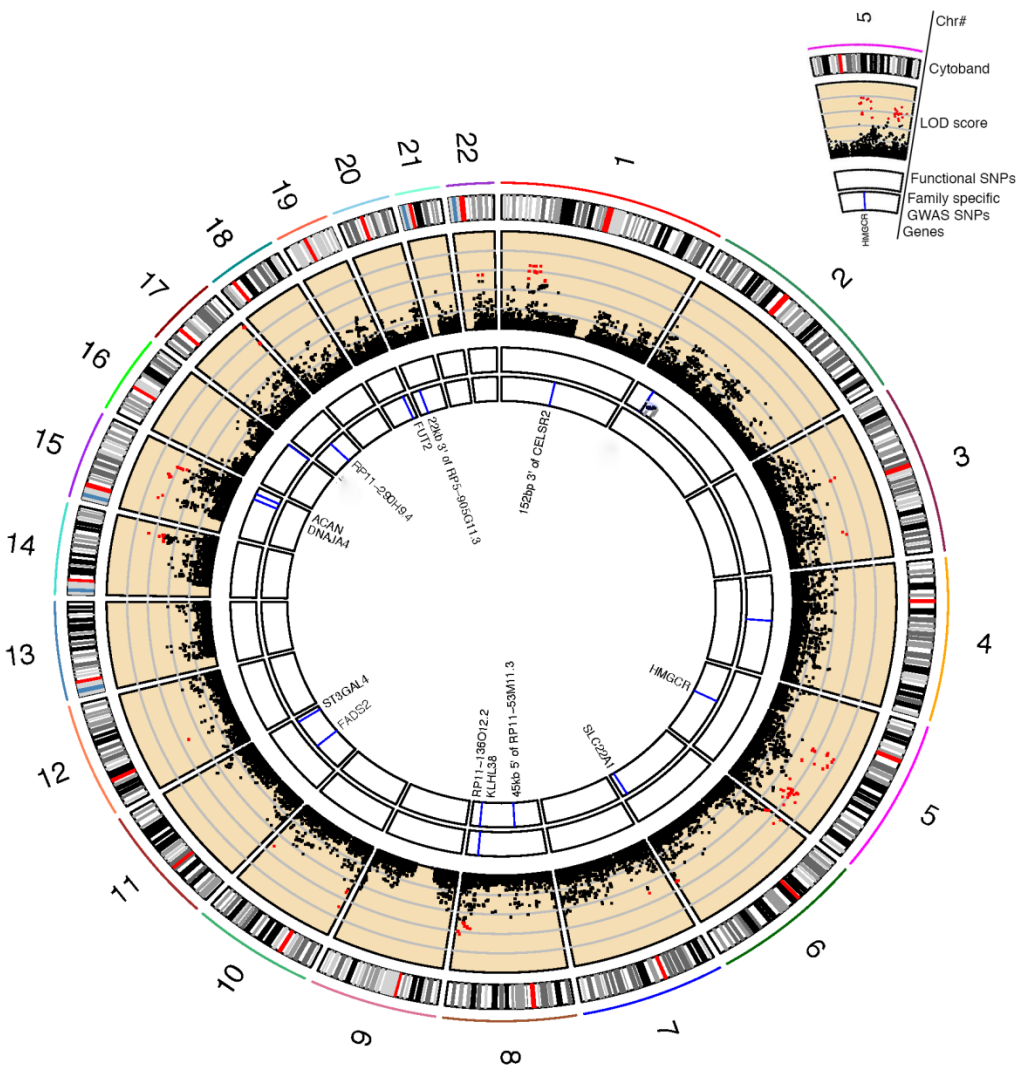


Figure 2a. Overlap between the 17 LDL-C regions with a LOD score > 1.0, exome variants (potentially functional and MAF<10%), and 10 family-specific GWAS variants identified in the Austrian family members, as illustrated by rCircos (Zhang et al. 2013). The outer most track indicates the chromosome number; followed by the cytoband; LOD scores of ~95K SNPs (red indicates LOD > 1.0); exome variants predominantly present in affected family members (Supplementary Table 2); family-specific GWAS SNPs (Table 2); and the gene names of the variants.

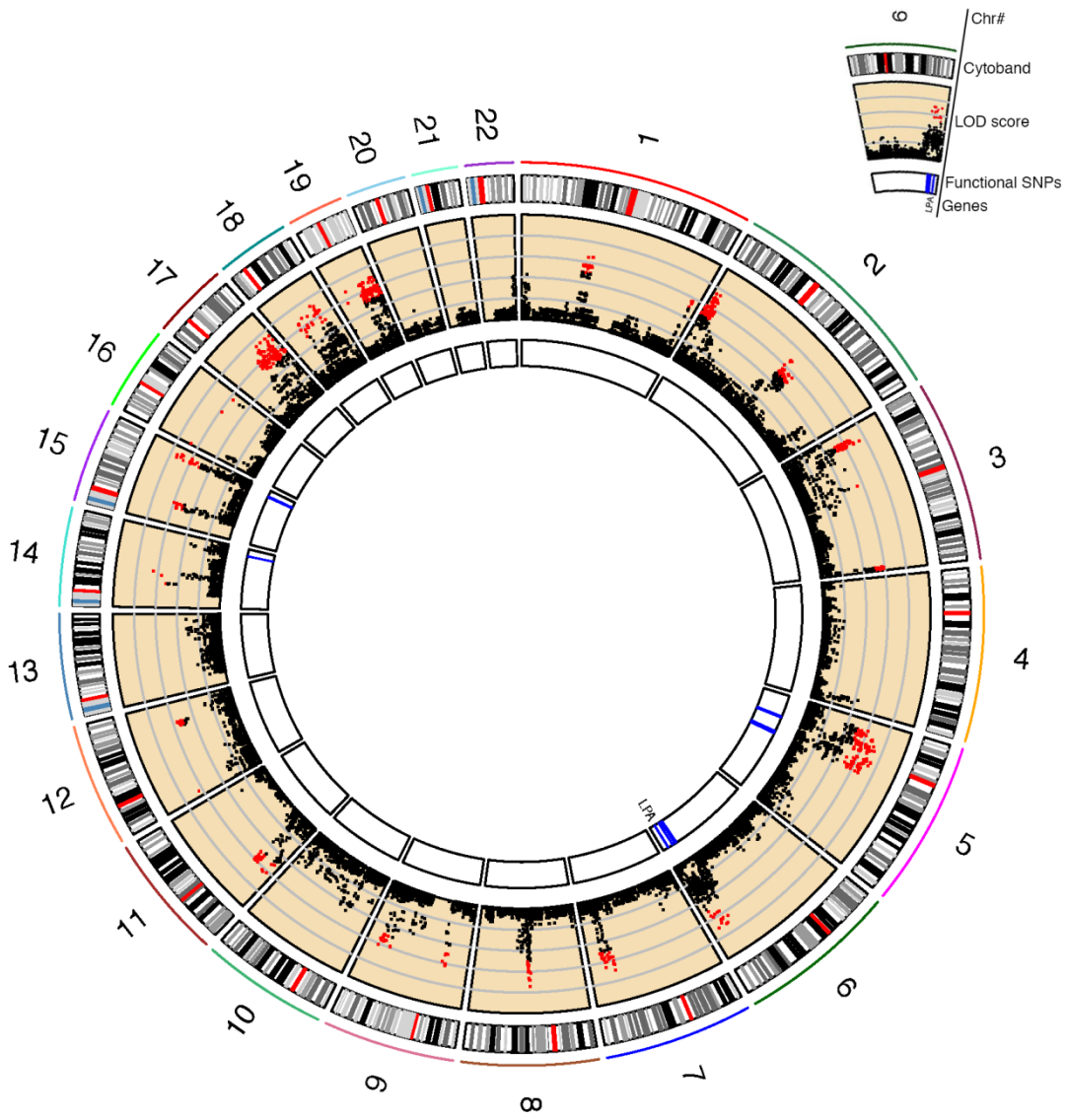


Figure 2b. Overlap between the Lp(a) regions with a LOD score > 1.0 and exome variants (potentially functional and MAF<10%) identified in the Austrian family members, as illustrated by rCircos (Zhang et al. 2013). The outer most track indicates the chromosome number; followed by the cytoband; LOD scores of ~95K SNPs (red indicates LOD > 1.0); and exome variants present only in the family members with high Lp(a).

Supplementary Table 1. The 47 LDL-C GWAS SNPs (or their LD proxies, $r^2 \geq 0.95$) (Willer et al. 2013) that were included in the risk score analysis.

Chr	Location	Rs #	Risk	Alt	Proxy	Risk (proxy)	Beta (weight)	P-value	MAF
1	25768937	rs10903129	g	a			0.033	3.03E-17	0.54
1	27138393	rs12748152	t	c			0.050	3.21E-12	0.07
1	55505647	rs11591147	g	t			0.497	8.58E-143	0.98
1	63133930	rs4587594	g	a	rs1570694	a	0.049	1.63E-32	0.69
1	93269824	rs4847221	c	t			0.038	8.57E-11	0.90
1	109818530	rs646776	t	c			0.160	1.63E-272	0.79
1	150958836	rs267733	a	g			0.033	5.29E-09	0.86
1	234846396	rs6695664	g	a			0.037	9.36E-13	0.16
2	21263900	rs1367117	a	g			0.119	9.48E-183	0.29
2	27742603	rs780093	t	c			0.022	2.36E-08	0.41
2	44073881	rs6544713	t	c	rs4299376	g	0.081	4.84E-83	0.29
2	121309488	rs2030746	t	c			0.021	8.61E-09	0.40
2	135762344	rs16831243	t	c			0.038	9.06E-12	0.18
3	12370737	rs17793951	a	g			0.025	2.63E-09	0.69
3	32533010	rs7640978	c	t			0.039	9.84E-09	0.89
3	132163200	rs17404153	g	t	rs10490862	a	0.034	1.83E-09	0.86
4	3434885	rs6818397	t	g			0.022	1.68E-08	0.41
5	74656539	rs12916	c	t			0.073	7.79E-78	0.43
5	156390297	rs6882076	c	t			0.046	3.31E-31	0.67
6	16127407	rs3757354	c	t			0.038	2.09E-17	0.79
6	160578860	rs1564348	c	t			0.048	2.76E-21	0.15
7	25991826	rs4722551	c	t			0.039	3.95E-14	0.17
7	44581986	rs17725246	c	t			0.047	1.49E-20	0.20
8	55421614	rs10102164	a	g	rs9298506	g	0.032	3.74E-11	0.17
8	116648565	rs2737229	a	c			0.029	3.74E-12	0.72
8	126490972	rs2954029	a	t	rs2954022	c	0.056	2.1E-50	0.53
9	2640759	rs3780181	a	g			0.045	1.76E-09	0.95
9	107664301	rs1883025	c	t			0.030	6.14E-11	0.76
9	136154168	rs579459	c	t			0.067	2.42E-44	0.22
10	113933886	rs2255141	a	g	rs2792751	t	0.030	1.32E-13	0.32
11	18632984	rs10128711	c	t			0.034	9.21E-13	0.72
11	61609750	rs174583	c	t			0.052	7.00E-41	0.63
11	126243952	rs11220462	a	g	rs7940893	c	0.059	6.61E-21	0.14
14	24883887	rs8017377	a	g			0.030	2.52E-15	0.46
16	56989590	rs247616	c	t	rs183130	c	0.055	2.57E-37	0.71
17	7091650	rs314253	t	c			0.024	3.44E-10	0.66
17	45391804	rs7225700	c	t			0.030	3.56E-13	0.67
19	11202306	rs6511720	g	t			0.221	3.85E-262	0.90
19	19407718	rs10401969	t	c			0.118	2.65E-54	0.93
19	45395619	rs2075650	g	a			0.177	1.72E-214	0.13
19	49206417	rs492602	g	a			0.029	9.42E-14	0.43
20	12969400	rs680379	g	a			0.024	7.96E-10	0.63
20	17845921	rs2328223	c	a			0.030	5.63E-09	0.25
20	39797465	rs753381	c	t			0.038	3.57E-25	0.54
20	43042364	rs1800961	c	t			0.069	6.03E-10	0.97
22	30378703	rs5763662	t	c	rs4820821	a	0.077	1.19E-08	0.03
22	46629479	rs4253776	g	a			0.031	3.35E-08	0.12

Chr, chromosome; Risk, LDL-C increasing variant; Alt, alternative variant; MAF, minor allele frequency.

Supplementary Table 2. The potential functional variants predominantly present in the affected family members identified by exome-sequencing and located in the LDL-C regions with a LOD score >1.0.

Chr	Location	Rs #	Type	MAF	Gene	Risk	Family members with high LDL-C											Family members with normal LDL-C				
							7777	7776	7772	7775	7749	7792	7724	7727	773	7725	7729	778	7793	7726	7789	776
2	33482586	rs11686962	SPLICE-SITE	0.071	*LTBP1*	T	C/T	C/T	C/C	C/T	C/T	C/T	C/T	C/T	T/T	C/T	C/C	C/C	C/C	C/C	C/T	C/C
4	87770252	rs17694522	NON-SYNON	0.066	*SLC10A6*	A	G/A	G/A	G/A	G/A	G/A	G/A	G/G	G/A	G/A	G/A	G/A	G/G	G/G	G/G	G/A	G/A
8	124658214	rs11779866	NON-SYNON	0.102	*KLHL38*	T	C/T	C/C	C/T	C/T	T/T	C/T	C/T	C/T	C/T	C/T	C/C	C/C	C/C	C/C	C/C	C/T
15	78558584	rs142025971	FRAMESHIFT	0.065	*DNAJA4*	C	CAT/C	CAT/C	CAT/C	CAT/C	CAT/CAT	CAT/C	CAT/C	CAT/C	CAT/CAT	CAT/C	CAT/C	CAT/CAT	CAT/CAT	CAT/CAT	CAT/CAT	CAT/CAT
15	89398553	rs35430524	NON-SYNON	0.083	*ACAN*	A	C/A	C/A	C/A	C/A	C/C	C/A	C/A	A/A	C/C	C/A	C/A	C/C	C/C	C/C	C/A	C/A
16	89985844	rs1805005	NON-SYNON	0.104	*MC1R*	T	G/T	G/T	G/T	G/T	G/G	G/G	G/T	G/T	G/T	G/T	G/T	G/G	G/G	G/G	G/G	G/T

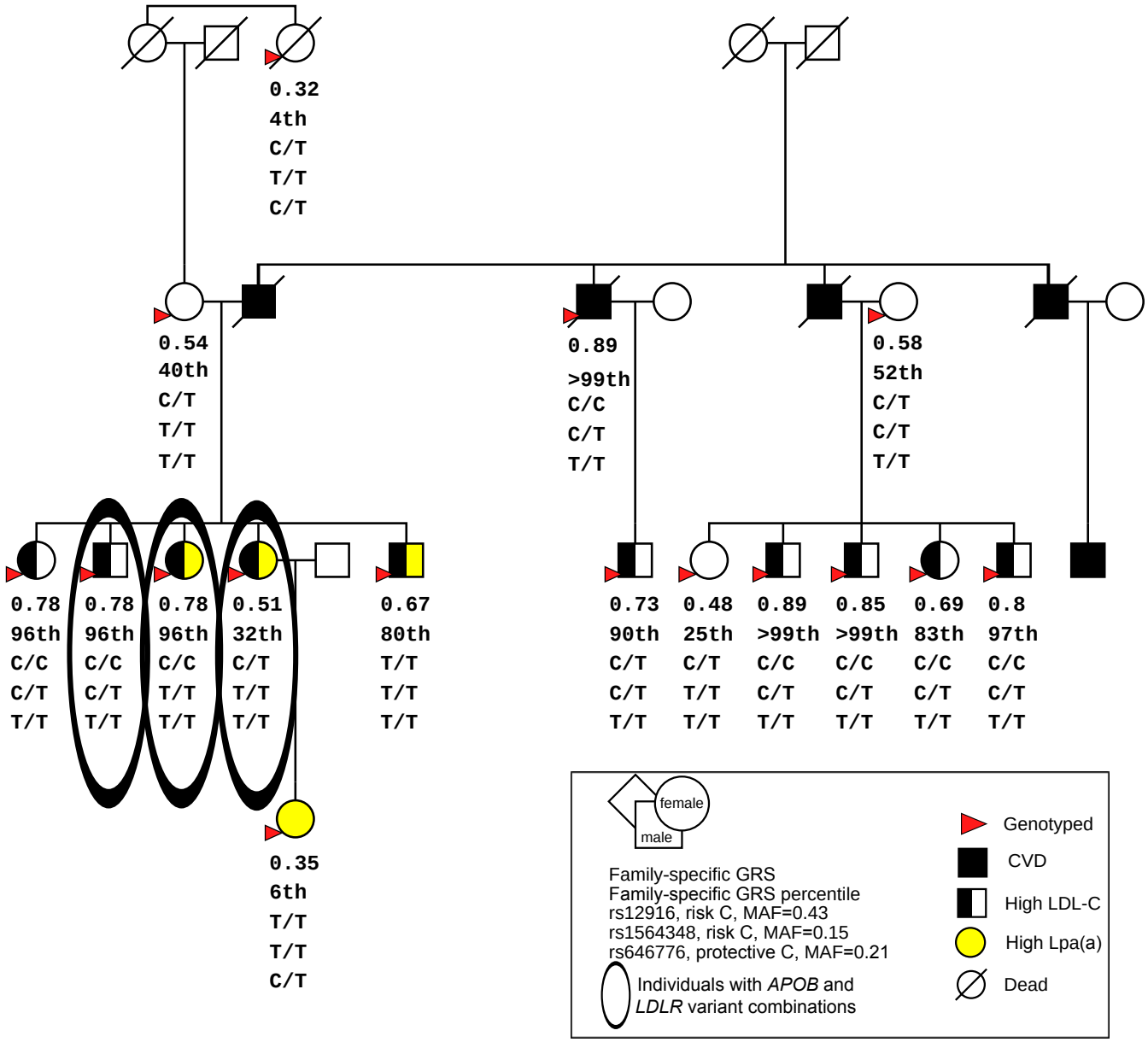
Chr, chromosome; LDL-C, low-density lipoprotein cholesterol; Risk, potential LDL-C increasing allele; MAF, minor allele frequency; and Grey highlight=High LDL-C status.

Supplementary Table 3. The non-synonymous and splice site variants identified in the known FH genes.

			Family members with high LDL-C											Family members with normal LDL-C									
Chr	**Rs #**	**Location**	**7777**	**7776**	**7772**	**7775**	**7792**	**7749**	**7724**	**7727**	**773**	**7725**	**7729**	**778**	**7793**	**7726**	**776**	**7789**	**Variant type**	**Gene**	**MAF**	**FH variant/LDL-C GWAS**	**Change**
Variants MAF < 10%:																							
2	rs1801695	21224853	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/T	C/C	C/T	C/C	C/C	C/T	C/C	C/C	NON-SYN	*APOB*	0.026	FH variant	A4481T
2	rs1801701	21228827	C/T	C/T	C/C	C/T	C/T	C/T	C/C	C/T	C/C	C/C	C/T	C/C	C/C	C/T	C/T	C/T	NON-SYN	*APOB*	0.077	GWAS LDL-C	R3638Q
2	rs61742247	21234902	C/G	C/G	C/G	C/G	C/G	C/C	C/C	C/C	C/C	C/C	C/C	C/G	C/G	C/C	C/C	C/C	NON-SYN	*APOB*	0.013		S1613T
19	rs72658861	11222182	T/T	T/C	T/T	T/C	T/C	T/T	T/T	T/T	T/T	T/T	T/T	T/T	T/T	T/T	T/T	T/C	SPLICE SITE:INTRONIC	*LDLR*	0.0052	FH variant	c.1061-8T>C
19	rs72658867	11231203	G/G	G/A	G/A	G/A	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/A	G/G	G/G	G/G	G/G	SPLICE SITE:INTRONIC	*LDLR*	0.0039	FH variant	c.2140+5G>A
19	rs45508991	11233886	C/C	C/T	C/C	C/T	C/T	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/T	NON-SYN	*LDLR*	0.0018	FH variant	T726I
Common variants:																							
1	rs6687605	25889632	C/C	C/C	C/C	C/C	T/C	T/C	T/C	T/T	T/T	T/C	C/C	T/C	T/C	C/C	T/C	C/C	NON-SYN	*LDLRAP1*	0.51		S202P
1	rs11583680	55505668	C/C	C/C	C/C	C/C	C/C	C/T	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	NON-SYN	*PCSK9*	0.12	GWAS LDL-C	A53V
1	rs2483205	55518316	C/C	C/T	C/C	C/T	C/T	C/T	C/T	C/T	C/C	C/C	C/C	C/T	C/T	T/T	C/T	T/T	SPLICE SITE:INTRONIC	*PCSK9*	0.46	FH variant	c.658-7C>T
1	rs2495477	55518467	A/A	A/G	A/A	A/G	A/G	A/A	A/A	A/G	A/A	A/A	A/A	A/G	A/G	A/G	A/A	A/G	SPLICE SITE:INTRONIC	*PCSK9*	0.41		c.799+3A>G
1	rs562556	55524237	A/A	A/A	A/A	A/A	A/A	A/A	A/A	G/A	A/A	A/A	A/A	A/A	G/A	G/A	A/A	A/A	NON-SYN	*PCSK9*	0.82		V474I
1	rs505151	55529187	A/A	A/A	A/A	A/A	A/A	A/A	A/A	A/A	A/A	A/A	A/A	A/A	A/A	A/A	A/A	A/A	NON-SYN	*PCSK9*	0.96	GWAS LDL-C	c.2009G>A
2	rs1042034	21225281	C/T	C/T	C/T	C/T	C/T	T/T	T/T	T/T	C/T	T/T	C/T	C/T	C/T	C/T	T/T	T/T	NON-SYN	*APOB*	0.77	GWAS LDL-C	S4338N
2	rs676210	21231524	G/A	G/A	G/A	G/A	G/A	G/G	G/G	G/G	G/A	G/G	G/A	G/A	G/A	G/A	G/G	G/G	NON-SYN	*APOB*	0.22	GWAS LDL-C	P2739L
2	rs584542	21232803	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	NON-SYN	*APOB*	1		V2313I
2	rs568413	21235475	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	NON-SYN	*APOB*	1		C1422Y
2	rs679899	21250914	G/A	G/A	G/A	G/G	G/A	G/G	G/G	G/G	G/A	G/G	G/A	G/A	G/A	G/A	G/G	G/G	NON-SYN	*APOB*	0.48	GWAS LDL-C	A618V
2	rs1367117	21263900	G/A	G/A	G/G	G/A	G/A	G/A	G/A	A/A	G/G	G/A	G/A	G/G	G/G	G/A	G/A	A/A	NON-SYN	*APOB*	0.29	GWAS LDL-C	T98I
19	rs2738442	11221454	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	SPLICE SITE:INTRONIC	*LDLR*	1		c.679+7T>C

LDL-C, Low-density lipoprotein cholesterol; Chr, chromosome; MAF, minor allele frequency; FH variant, previously implicated for familial hypercholesterolemia; Grey highlight=High LDL-C status; Red box, risk combination in affected family member; and Blue box, risk combination in unaffected family member.

Supplementary Figure 1. A summary of the genetic findings contributing to high LDL-C in the Austrian family. The pedigree was drawn using CraneFoot (Mäkinen et al. 2005). The box shows the explanations of the drawing scheme. The text and numbers underneath each individual indicate individual's family-specific genetic risk score (GRS); family-specific GRS percentile; and genotypes of three individual SNPs from the family specific GRS that were significant after random label swapping. Individuals with a combination of *APOB* and *LDLR* mutations are circled.



References

1. Mendis S, Puska P, Norrving B. Global atlas on cardiovascular disease prevention and control. World Heal Organ. 2011;2–14.

2. Fischer M, Broeckel U, Holmer S, Baessler A, Hengstenberg C, Mayer B, et al. Distinct heritable patterns of angiographic coronary artery disease in families with myocardial infarction. Circulation. 2005;111(7):855–62.

3. Zdravkovic S, Wienke A, Pedersen NL, Marenberg ME, Yashin AI, De Faire U. Heritability of death from coronary heart disease: A 36-year follow-up of 20 966 Swedish twins. J Intern Med. 2002;252(3):247–54.

4. Deloukas P, Kanoni S, Willenborg C, Farrall M, Assimes TL, Thompson JR, et al. Large-scale association analysis identifies new risk loci for coronary artery disease. Nat Genet. 2013;45(1):25–33.

5. Santos RD, Gidding SS, Hegele RA, Cuchel MA, Barter PJ, Watts GF, et al. Defining severe familial hypercholesterolaemia and the implications for clinical management: a consensus statement from the International Atherosclerosis Society Severe Familial Hypercholesterolemia Panel. Lancet Diabetes Endocrinol. Elsevier Ltd; 2016;8587(16):19–21.

6. Cuchel M, Bruckert E, Ginsberg HN, Raal FJ, Santos RD, Hegele RA, et al. Homozygous familial hypercholesterolaemia: New insights and guidance for clinicians to improve detection and clinical management. A position paper fromthe Consensus Panel on Familial Hypercholesterolaemia of the European Atherosclerosis Society. Eur Heart J. 2014;35(32):2146–57.

7. Futema M, Plagnol V, Li K, Whittall R a, Neil HAW, Seed M, et al. Whole exome

sequencing of familial hypercholesterolaemia patients negative for LDLR/APOB/PCSK9 mutations. J Med Genet. 2014;51(8):537–44.

8. Clarke REJ, Padayachee ST, Preston R, McMahon Z, Gordon M, Graham C, et al. Effectiveness of alternative strategies to define index case phenotypes to aid genetic diagnosis of familial hypercholesterolaemia. Heart. 2013;99(3):175–80.

9. Civeira F, Jarauta E, Cenarro A, García-Otín AL, Tejedor D, Zambón D, et al. Frequency of Low-Density Lipoprotein Receptor Gene Mutations in Patients With a Clinical Diagnosis of Familial Combined Hyperlipidemia in a Clinical Setting. J Am Coll Cardiol. 2008;52(19):1546–53.

10. Bahlo M, Tankard R, Lukic V, Oliver KL, Smith KR. Using familial information for variant filtering in high-throughput sequencing studies. Vol. 133, Human Genetics. 2014. p. 1331–41.

11. Eggers S, Smith KR, Bahlo M, Looijenga LH, Drop SL, Juniarto ZA, et al. Whole exome sequencing combined with linkage analysis identifies a novel 3 bp deletion in NR5A1. Eur J Hum Genet. 2015;23(4):486–93.

12. Willer CJ, Schmidt EM, Sengupta S, Peloso GM, Gustafsson S, Kanoni S, et al. Discovery and refinement of loci associated with lipid levels. Nat Genet. 2013;45(11):1274–83.

13. Kronenberg F, Kuen E, Ritz E, Junker R, König P, Kraatz G, et al. Lipoprotein(a) serum concentrations and apolipoprotein(a) phenotypes in mild and moderate renal failure. J Am Soc Nephrol. 2000;11(1):105–15.

14. Laschkolnig A, Kollerits B, Lamina C, Meisinger C, Rantner B, Stadler M, et al. Lipoprotein (a) concentrations, apolipoprotein (a) phenotypes, and peripheral arterial

disease in three independent cohorts. Cardiovasc Res. 2014;103(1):28–36.

15. Stancakova A, Javorsky M, Kuulasmaa T, Haffner SM, Kuusisto J, Laakso M. Changes in insulin sensitivity and insulin release in relation to glycemia and glucose tolerance in 6,414 Finnish men. Diabetes. 2009;58(5):1212–21.

16. T2D-GENES Consortium, GoT2D Consortium, DIAGRAM Consortium. 2016 March 24. http://www.type2diabetesgenetics.org/.

17. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics. 2009;25(14):1754–60.

18. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, et al. The genome analysis toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res. 2010;20(9):1297–303.

19. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, et al. PLINK: A tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet. 2007;81(3):559–75.

20. Lange K, Papp JC, Sinsheimer JS, Sripracha R, Zhou H, Sobel EM. Mendel: The Swiss army knife of genetic analysis programs. Bioinformatics. 2013;29(12):1568–70.

21. Schäffer AA, Lemire M, Ott J, Lathrop GM, Weeks DE. Coordinated conditional simulation with SLINK and SUP of many markers linked or associated to a trait in large pedigrees. Hum Hered. 2011;71(2):126–34.

22. Nordestgaard BG, Chapman MJ, Ray K, Borén J, Andreotti F, Watts GF, et al. Lipoprotein(a) as a cardiovascular risk factor: Current status. Eur Heart J. 2010;31(23):2844–53.

23. Kronenberg F, Utermann G. Lipoprotein(a): Resurrected by genetics. J Intern Med. 2013;273(1):6–30.

24. Talmud PJ, Shah S, Whittall R, Futema M, Howard P, Cooper JA, et al. Use of low-density lipoprotein cholesterol gene score to distinguish patients with polygenic and monogenic familial hypercholesterolaemia: A case-control study. Lancet. 2013;381(9874):1293–301.

25. Rao F, Schork AJ, Maihofer AX, Nievergelt CM, Marcovina SM, Miller ER, et al. Heritability of Biomarkers of Oxidized Lipoproteins: Twin Pair Study. Arterioscler Thromb Vasc Biol. 2015;35(7):1704–11.

26. Schmidt K, Noureen A, Kronenberg F, Utermann G. Structure, Function, and Genetics of Lipoprotein(a). J Lipid Res. 2016; e-pub ahead of print 13 April 2016; doi: 10.1194/jlr.R067314

27. Kronenberg F. Genetic determination of lipoprotein(a) and its association with cardiovascular disease: Convenient does not always mean better. Journal of Internal Medicine. 2014;276(3):243–7.

28. Brænne I, Kleinecke M, Reiz B, Graf E, Strom T, Wieland T, et al. Systematic analysis of variants related to familial hypercholesterolemia in families with premature myocardial infarction. Eur J Hum Genet. 2015;(April):1–7.

29. Tejedor MT, Cenarro A, Tejedor D, Stef M, Mateo-Gallego R, De Castro I, et al. Haplotype analyses, mechanism and evolution of common double mutants in the human LDL receptor gene. Mol Genet Genomics. 2010;283(6):565–74.

30. Alharbi KK, Aldahmesh MA, Spanakis E, Haddad L, Whittall RA, Chen XH, et al. Mutation scanning by meltMADGE: Validations using BRCA1 and LDLR, and demonstration of the

potential to identify severe, moderate, silent, rare, and paucimorphic mutations in the general population. Genome Res. 2005;15(7):967–77.

31. Teslovich TM, Musunuru K, Smith A V, Edmondson AC, Stylianou IM, Koseki M, et al. Biological, clinical and population relevance of 95 loci for blood lipids. Nature. Nature Publishing Group; 2010;466(7307):707–13.

32. Borén J, Lee I, Zhu W, Arnold K, Taylor S, Innerarity TL. Identification of the low density lipoprotein receptor-binding site in apolipoprotein B100 and the modulation of its binding activity by the carboxyl terminus in familial defective Apo-B100. J Clin Invest. 1998;101(5):1084–93.

33. Swerdlow DI, Preiss D, Kuchenbaecker KB, Holmes M V., Engmann JEL, Shah T, et al. HMG-coenzyme A reductase inhibition, type 2 diabetes, and bodyweight: Evidence from genetic analysis and randomised trials. Lancet. 2015;385(9965):351–61.

34. Würtz P, Wang Q, Soininen P, Kangas AJ, Fatemifar G, Tynkkynen T, et al. Metabolomic Profiling of Statin Use and Genetic Inhibition of HMG-CoA Reductase. J Am Coll Cardiol. 2016;67(10):1200–10.

35. Mäkinen VP, Parkkonen M, Wessman M, Groop PH, Kanninen T, Kaski K. High-throughput pedigree drawing. Eur J Hum Genet. 2005;13(8):987–9.

36. Zhang H, Meltzer P, Davis S. RCircos: an R package for Circos 2D track plots. BMC Bioinformatics. 2013;14:244.

Chapter 3

Two Mexican families with familial hypobetalipoproteinemia share a region on chromosome 2p16

Chapter 3

Two Mexican families with familial hypobetalipoproteinemia share a region on chromosome 2p16

Abstract

Background: To better prevent and treat hypercholesterolemia, one of the most critical risk factors of coronary heart disease, it is important to identify genes and molecular mechanisms decreasing serum total cholesterol (TC) levels. Familial hypobetalipoproteinemia (FHBL) is an autosomal dominant disorder characterized by low levels of plasma total cholesterol (TC), low-density lipoprotein (LDL-C), and apolipoprotein B (apoB). FHBL has previously been shown to be caused by mutations in the APOB, PCSK9, and ANGPTL3 genes, and loci on chromosomes 10 and 3p21 have also been linked to FHBL. However, in most FHBL cases the underlying cause is unknown. To identify genes involved in FHBL and cholesterol metabolism in Mexicans, we explored two Mexican FHBL families without known mutations in the known causal genes. Importantly, the probands of the both families had a similar kinetic profile with an increased apoB-LDL and apoB-VLDL catabolic rate. **Methods and Results:** The affected and unaffected individuals from the Mexican FHBL families were exome sequenced by Illumina Truseq Exome enrichment kit capturing 62Mb coding and regulatory regions. We focused on the variants shared by all affected family members and not present in the family members with normal apoB levels. In addition, we filtered the variants based on their type, frequency, gene expression, and functional predictions using SIFT and PolyPhen. In family 1, filtering for novel missense variants and variants with minor allele frequency (MAF) <5% as well as applying the filters for gene expression and functional predictions, reduced the variant pool to 9 nonsynonymous variants. In family 2, the filtering led to 13 nonsynonymous variants. To further investigate whether any of these 22 genes regulates cholesterol levels largely in Mexicans, we tested their common variants for association with TC in 3,700 Mexican individuals. Variants in one gene, neuregulin 1

(NRG1), passed the Bonferroni correction for association with TC, with rs1383966 providing the strongest signal. However, subsequent replication in ~6,000 Mexicans did not validate these results. We visualized overall chromosomal differences between affected and unaffected family member using rCircos and observed shared locus on chromosome 2p16. **Conclusions:** We did not find novel variants or genes for FHBL, but identified 20-Mb region in chromosome 2 as a new potential FHBL locus. However, further FHBL families from Mexico are needed to confirm these results.

Introduction

Overall, the deaths caused by CVD are declining, reflecting recent improvements in health awareness and medical care [1]. However, this decline is not equally distributed among different ethnic groups and still every third death in the United States is caused by CVD. In about half of these the cause of death is CHD [2]. High levels of blood total cholesterol (TC) and especially of low-density lipoprotein cholesterol (LDL-C) (hypercholesterolemia) are well-established risk factors for CHD [3]. Statins (HMG CoA reductase inhibitors) are the primary treatment to lower LDL-C. However, many of the patients who are on statins are either not responding ideally or suffer side effects [4]. Thus, other therapeutic drugs targeting genes involved in lipid metabolism should be developed for hypercholesterolemia.

Mexicans are an admixed population of European, Native American, and African ancestries [5]. According to previous studies, Mexicans have astoundingly high prevalence of different types of dyslipidemias [6,7]. However, overall the Mexican population is underinvestigated in genomic studies, and particularly in cardiovascular genetic studies [8]. Thus, finding population-specific variants lowering lipids in Mexicans would be very beneficial.

Familial hypobetalipoproteinemia (FHBL) is an autosomal dominant disorder of apolipoprotein B (apoB) metabolism [9]. It is characterized by reduced levels of plasma TC, LDL-C, and apoB [10]. Most of the subjects with FHBL are heterozygous and asymptomatic. However,

some of the heterozygous FHBL subjects develop nonalcoholic fatty liver [11]. It is important to note that FHBL is associated with a reduced risk of CVD [12], most likely because of the life-long low levels of TC and LDL–C [13]. Prevalence of heterozygous FHBL is about 1:500-1:1,000 [12]. FHBL has previously been shown to be caused by mutations in the apolipoprotein B (APOB) gene [10,14], but FHBL is a genetically heterogeneous disorder that is not always linked to the APOB gene [15]. Other genes, including PCSK9 and ANGPTL3 and loci on chromosomes 10 and 3p21 have also been implicated in some FHBL families [16–19]. However, in many FHBL cases the definite genetic cause is not known yet [20].

As an initial step towards identifying novel FHBL genes, we performed exome sequencing in two Mexican families with FHBL. A previous study by Gutiérrez-Cirlos et al. showed that these families do not have causal FHBL variants in the known causative FHBL genes and the probands of these families have a similar kinetic profile for apoB-LDL and apoB-VLDL production and catabolic rate. Furthermore, their LDL and VLDL apoB catabolic rate was significantly increased when compared to controls [20].

We hypothesized that the underlying genetic component of FHBL plays a role in the apoB or LDL pathways. To find additional variants involved in lipid metabolism, we will extend the findings of a familial lipid disorder to the population level. We postulate that genes where these variants reside may regulate apoB, LDL-C, and TC levels in the Mexican general population as well.

Materials and Methods

Exome sequencing

We exome sequenced two Mexican FHBL families (figure 1) without mutations in the known FHBL genes. These families are previously described by Aguilar-Salinas et al [20]. The genomic DNA was extracted using standard protocols. The exome libraries were prepared using Illumina Truseq 2.1 exome capture kit and sequenced using the Illumina HiSeq2000 platform.

We used our standard exome sequence data analysis pipeline, including quality control (QC), sequence alignment, and variant filtering approaches described in detail in our previous exome sequencing paper [21]. Briefly, we aligned the reads using the Burrows-Wheeler Aligner (BWA) software package [22], performed the QC and variant calling using the Genome Analysis Toolkit (GATK) [23] and annotated the variants with the ANNOVAR software [24]. As FHBL is an autosomal dominant disorder, we focused on variants that are shared by the affected individuals and not present in the unaffected family members. Furthermore, because the disorder has a dominant mode of inheritance, we were interested in both homo- and heterozygous variants. Additionally, the variants were filtered by variant type, frequency, and evaluated by gene expression and a functional prediction score by PolyPhen-2 [25] and SIFT [26]. The final candidate genes were selected based on previous evidence in literature related to lipid metabolism, expression in relevant tissues such as liver and adipose.

Association analyses

We analyzed the potential susceptibility gene candidates found by exome sequencing for association with quantitative TC levels in the Mexican population. To avoid multiple testing, we restricted the analysis to the most promising candidate genes in both families. A Bonferroni corrected p-value of $\leq$0.002 (i.e. p=0.05/22 tested genes) was considered statistically significant. We have ~3,700 individuals with genome-wide SNP chip data available from two Mexican case-control GWAS studies, including a dyslipidemic cohort and a T2D cohort. In addition, we imputed the variants with MAF>1% using IMPUTE2 [27] and the 1000 Genomes data as a reference panel. The imputed SNPs with a MAF<1% were excluded from our analysis as their imputation quality is typically low. To ensure accuracy of the imputed genotypes, a posterior probability >0.9 and info value >0.8 were also used asQC criteria of the imputed SNPs. In addition, we excluded SNPs deviating from Hardy-Weinberg Equilibrium (HWE) (P-value <0.0001) [8]. We used SNPtest for association analysis because it can analyze quantitative phenotypes, their covariates, and imputed SNPs with proportional genotypes. Mexican

replication cohort with (n=~6,000) was utilized to further investigate the variants passing the Bonferroni correction in the first study sample for replication.

Circos visualization

The overall difference in genetic variation between affected and unaffected family members was visualized using RCircos [28].

Results

We exome sequenced two Mexican FHBL families: four individuals including two affected family members in family 1 and eight individuals including four affected family members in family 2. Overall, the exome sequencing resulted in on average 64 million uniquely mapped reads per sample. The average coverage was 42X and the average Ti/Tv ratio was 2.4, indicating a good data quality for exome sequencing. We found ~54,000 variants in each individual of which ~23,800 were coding variants of which ~11,900 were nonsynonymous.

Variant filtering

First, we analyzed the two families together and the variant filtering resulted in one nonsynonymous variant predicted to be damaging that was shared by all of the affected family members in both families but not present in any of the individuals with normal lipid levels. However, this variant resides in the TSPYL6 (Testis-specific Y-encoded-like protein 6) gene that is mainly expressed in testis. Furthermore, the variant was not associated with total cholesterol in the Mexican population cohort. Taken together these data made TSPLY6 an unlikely candidate for FHBL.

We then proceeded to analyze the two families separately as the previous evidence suggests that FHBL is a heterogeneous disorder that may be caused by different genes in different families [15]. With the larger family (family 2) consisting of 4 affected family members and 4 unaffected family members, the variant filtering resulted in 13 susceptibility genes for TC (Table 1). We applied more stringent filtering criteria for the smaller family (family 1) by requiring

that the gene where the variant resides in is expressed in relevant tissues to lipid metabolism. This filtering resulted in 9 candidate genes (Table 2).

Population association analyses

We analyzed the susceptibility genes found by exome sequencing for association with TC levels in the Mexican population. We assessed these candidate genes in ~3,700 individuals with genome wide SNP chip data from two Mexican case-control GWAS studies. In most cases, the actual variant identified in the families was missing in our imputed and genotyped Mexican GWAS data sets. Therefore, we investigated the gene for TC levels rather than the actual variant. This was done by testing all of the variants in the candidate gene for association. Using SNPtest for association analysis, variants in NRG1 (neuregulin 1) provided the strongest signals ($p< 2.7e-5$) (Table 3). However, when we further extended the analysis to 6,000 additional Mexican individuals, the most significant NRG1 variant did not show evidence of association with TC, making it a less likely candidate.

RCircos visualization

We recognize that we might have lost the causative variant/gene when applying our relatively strict filtering criteria. To better understand overall genome-wide differences in genetic structure between the affected and unaffected family members, we used RCircos to visualize our exome sequencing data (Figure 2). We observed 25 gene regions throughout the genome shared by all affected family members but not by the controls. Interestingly, 10 of the 25 genes reside in a 20-Mb region on chromosome 2 (Figure 2). Thus, this analysis pinpointed us back to the TSPLY6 gene region on chromosome 2. In more detail, the affected individuals from Family 1 exhibit 39 variants in 23 genes and affected individuals from family 2 exhibit 56 variants in 34 genes in the 20-Mb region not present in the unaffected family members. Importantly, 10 of these genes are shared between both families. Taken together, the RCircos visualization analysis resulted in 25 shared gene regions between the two Mexican FHBL families, the most

interesting one of these being the 20-Mb chromosome 2p16 region with 10 shared genes between the two families.

Discussion

We examined two Mexican families to identify novel causal variants for familial hypobetalipoproteinemia. We identified a large 20-Mb locus on chromosome 2 shared by the affected individuals in both families. While we were not able to pinpoint the exact causal variant within this region, a recent GWAS study of over 180,000 individuals identified a variant residing on EHBP1 (EH domain-binding protein) on chromosome 2 to be associated with LDL-C [29].

Our study has some limitations. As the FHBL is a fairly rare disorder without serious clinical symptoms, finding FHBL families is challenging. Therefore, currently our study sample is small and collecting more families continues gradually. In addition, we potentially lost our susceptibility variant when performing the variant calling only with GATK. However, this data has been subsequently analyzed many times after our initial analysis, always resulting in TSPLY6 as a shared variant between the affected individuals. In addition, we have performed a two-point linkage analysis using data generated by whole exome sequencing as conducted in Park et al. [30] The highest lod score was detected in the same 20-Mb chromosome 2 region (data not shown).

We recognize that FHBL is a heterogeneous disease and that the causal mechanisms might involve two or more genetic loci in these families. Future studies include recruitment of additional FHBL families and construction of Mexican-specific polygenic risk scores, which is currently impossible due to the lack of genome-wide genotypes with detailed phenotypes in the population level in Mexicans.

Tables

Table1. Variants found by exome sequencing and after filtering in Family 2. The brackets indicate that the variant or gene was discovered later than our initial analysis.

Location	Amino acid change	rs #	Gene
Chr2:46711558	W218L	new	LOC388946 (TMEM247)
Chr2:136620315	N361S	rs141448886	MCM6
Chr2:136691500	P162S	rs149170955	DARS
Chr8:110631186	C104W	new	SYBU
Chr9:5462971	K178E	new	CD274
Chr9:117849170	N280K	rs200913956	TNC
Chr13:46935685	N337S	rs151242987	C13orf18
Chr14:94936062	T57I	rs45438596	SERPINA9
Chr16:3707050	V163L	new (rs759204609)	DNASE1
Chr19:3831739	G173R	new (rs755011310)	ZFR2
Chr19:8669976	R119Q	rs3814291	ADAMTS10
Chr19:58370657	Q292K	rs3745132	ZNF587
Chr19:58421128	R173H	rs3826671	ZNF417

Table2. Variants found by exome sequencing and after filtering in Family1. The brackets indicate that the variant or gene was discovered later than our initial analysis.

Location	Amino acid change	rs #	Gene
Chr1:143767544	N102I	new (rs781911564)	PPIAL4G
Chr2:97482987	F325L	new (rs781051911)	CNNM3
Chr3:9970121	P337L	rs115419420	IL17RC
Chr6:3152760	L276S	rs2231370	BPHL
Chr6:32552080	Y59C	rs3175105	HLA-DRB1
Chr8:32611970	V258L	rs74942016	NRG1
Chr13:33703666	R383P	rs34425674	STARD13
Chr22:21354174	A309T	new	THAP7
Chr22:21141193	E447K	rs142451096	SERPIND1

Table 3. The results of the Mexican population sample in candidate genes.

Location	rs#	MAF	p-value	BETA	SE
Chr8:32078675	Deletion	0.16	2.7E-05	-0.13	0.03
Chr8:32475691	rs2466060	0.48	7.5E-05	-0.09	0.02
Chr8:32473401	Insertion	0.47	8.0E-05	-0.09	0.02
Chr8:32490310	rs2439273	0.15	0.0001	-0.12	0.03
Chr8:32099876	rs2347064	0.19	0.0001	-0.12	0.03
Chr8:32079101	rs1383965	0.15	0.0002	-0.12	0.03
Chr8:32088299	Insertion	0.23	0.0002	-0.11	0.03
Chr8:32110257	rs139988270	0.16	0.0002	-0.11	0.03
Chr8:32536454	rs75438610	0.12	0.0002	-0.13	0.03
Chr8:32489868	rs2439274	0.15	0.0002	-0.11	0.03

Chr indicates chromosome; MAF, minor allele frequency; BETA, the effect size; and SE, standard error.

Figures

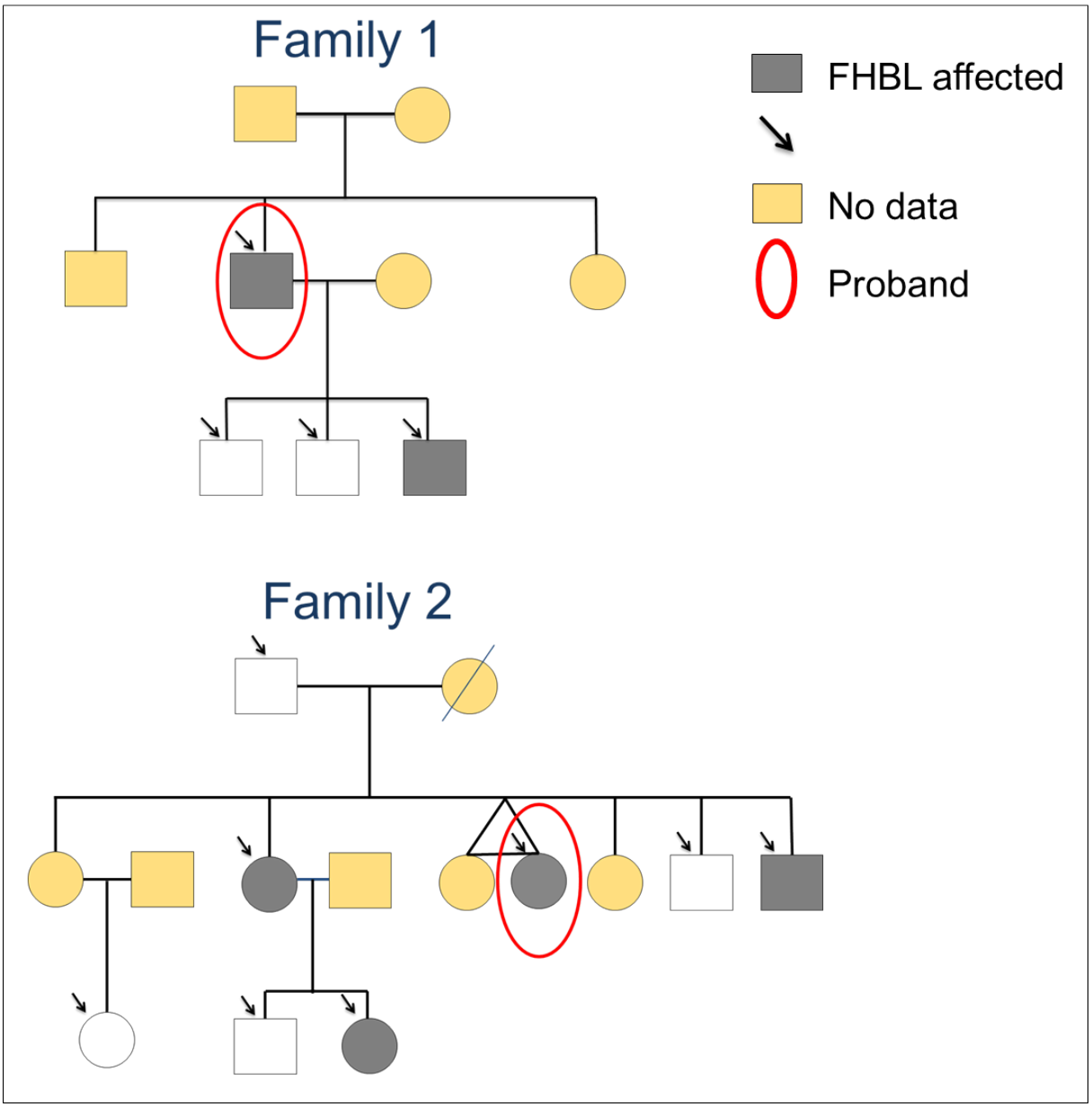


Figure 1. Pedigree structure of the two Mexican FHBL families. Arrow indicates individuals that were exome sequenced.

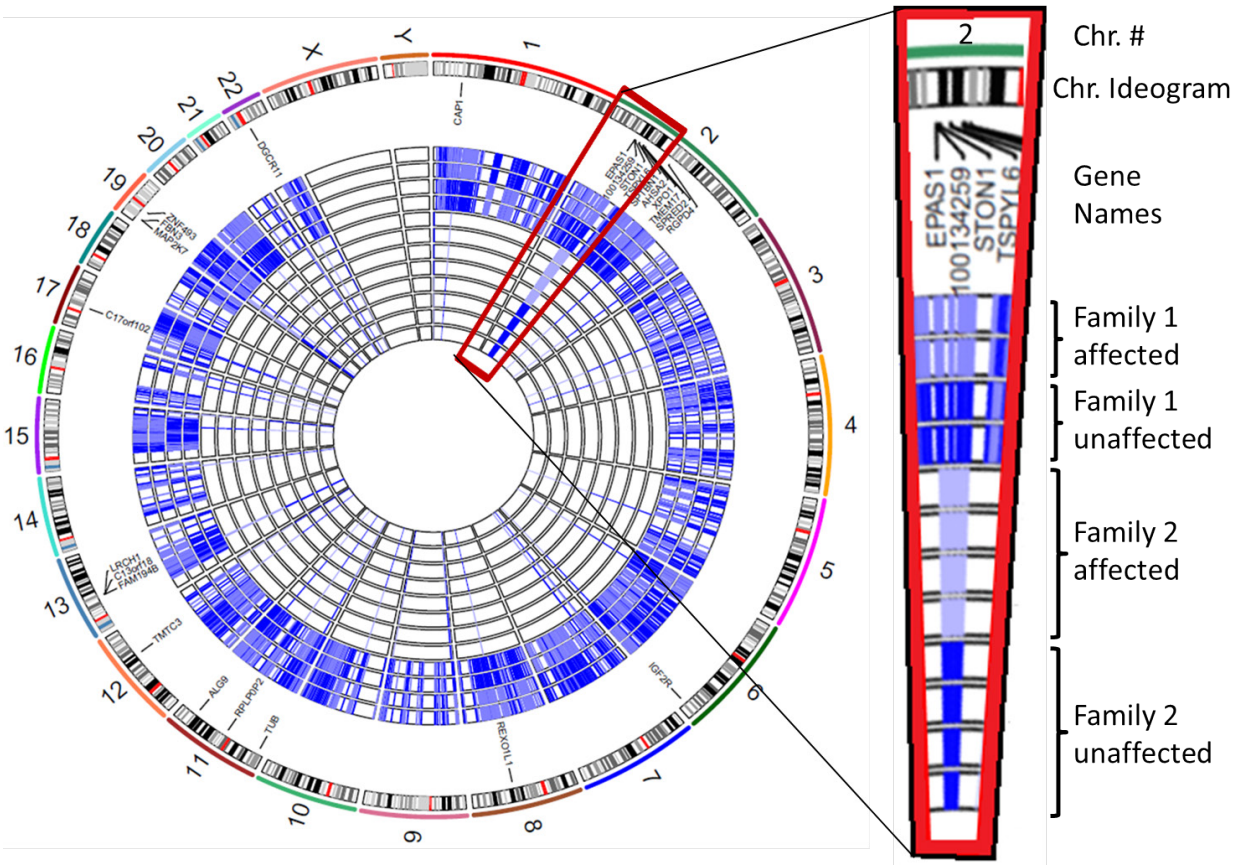


Figure 2. Overall genetic differences between affected and unaffected family members visualized using RCircos. All of the variants presented here are shared by affected family members and not present in unaffected family members, or vice versa. The first four traces after chromosomal g-stain and gene names are from Family 1 and the inner 8 from Family 2. Light blue indicates that the variant is present in the particular individual.

References

1. Nichols M, Townsend N, Scarborough P, Rayner M. Trends in age-specific coronary heart disease mortality in the European Union over three decades: 1980-2009. *Eur Heart J*. 2013;34(39):3017-3027. doi:10.1093/eurheartj/eht159.

2. Roger VL, Go AS, Lloyd-Jones DM, Benjamin EJ, Berry JD, Borden WB, Bravata DM, Dai S, Ford ES, Fox CS, Fullerton HJ, Gillespie C, Hailpern SM, Heit JA, Howard VJ, Kissela BM, Kittner SJ, Lackland DT, Lichtman JH, Lisabeth LD, Makuc DM, Marcus GM, Marelli A, Matchar DB, Moy CS, Mozaffarian D, Mussolino ME, Nichol G, Paynter NP, Soliman EZ, Sorlie PD, Sotoodehnia N, Turan TN, Virani SS, Wong ND, Woo D, Turner MB. Executive summary: Heart disease and stroke statistics-2012 update: A report from the American heart association. *Circulation*. 2012;125(1):188-197. doi:10.1161/CIR.0b013e3182456d46.

3. Lewington S, Whitlock G, Clarke R, Sherliker P, Emberson J, Halsey J, Qizilbash N, Peto R, Collins R. Cholesterol and Vascular Mortality By Age, Sex, and Blood Pressure: a Meta-Analysis of Individual Data From 61 Prospective Studies With 55,000 Vascular. *Lancet*. 2007;370(9602):1829-1839. doi:10.1016/S0140-6736(07)61778-4.

4. Mancini GBJ, Tashakkor AY, Baker S, Bergeron J, Fitchett D, Frohlich J, Genest J, Gupta M, Hegele RA, Ng DS, Pearson GJ, Pope J. Diagnosis, prevention, and management of statin adverse effects and intolerance: Canadian working group consensus update. *Can J Cardiol*. 2013;29(12):1553-1568. doi:10.1016/j.cjca.2013.09.023.

5. Price AL, Patterson N, Yu F, Cox DR, Waliszewska A, McDonald GJ, Tandon A, Schirmer C, Neubauer J, Bedoya G, Duque C, Villegas A, Bortolini MC, Salzano FM, Gallo C, Mazzotti G, Tello-Ruiz M, Riba L, Aguilar-Salinas CA, Canizales-Quinteros S, Menjivar M, Klitz W, Henderson B, Haiman CA, Winkler C, Tusie-Luna T, Ruiz-Linares A, Reich D. A genomewide admixture map for Latino populations. *Am J Hum Genet*. 2007;80(6):1024-1036. doi:10.1086/518313.

6. Aguilar-Salinas CA, Rojas R, Gómez-Pérez FJ, Valles V, Ríos-Torres JM, Franco A, Olaiz G, Rull JA, Sepúlveda J. High prevalence of metabolic syndrome in Mexico. *Arch Med Res*. 2004;35(1):76-81. doi:10.1016/j.arcmed.2003.06.006.

7. Aguilar-Salinas CA, Gómez-Pérez FJ, Rull J, Villalpando S, Barquera S, Rojas R. Prevalence of dyslipidemias in the Mexican National Health and Nutrition Survey 2006. *Salud Publica Mex*. 2010;52(2):S44-S53. doi:10.1590/S0036-36342010000700008.

8. Weissglas-Volkov D, Aguilar-Salinas CA, Nikkola E, Deere KA, Cruz-Bautista I, Arellano-Campos O, Muñoz-Hernandez LL, Gomez-Munguia L, Ordoñez-Sánchez ML, Reddy PMVL, Lusis AJ, Matikainen N, Taskinen M-R, Riba L, Cantor RM, Sinsheimer JS, Tusie-Luna T, Pajukanta P. Genomic study in Mexicans identifies a new locus for triglycerides and refines European lipid loci. *J Med Genet*. 2013;50(5):298-308. doi:10.1136/jmedgenet-2012-101461.

9. Tarugi P, Averna M, Di Leo E, Cefalù AB, Noto D, Magnolo L, Cattin L, Bertolini S, Calandra S. Molecular diagnosis of hypobetalipoproteinemia: An ENID review. *Atherosclerosis*. 2007;195(2). doi:10.1016/j.atherosclerosis.2007.05.003.

10. Schonfeld G, Patterson BW, Yablonskiy D a, Tanoli TSK, Averna M, Elias N, Yue P, Ackerman J. Fatty liver in familial hypobetalipoproteinemia: triglyceride assembly into VLDL particles is affected by the extent of hepatic steatosis. *J Lipid Res*. 2003;44(3):470-

478. doi:10.1194/jlr.M200342-JLR200.

11. Sankatsing RR, Fouchier SW, De Haan S, Hutten BA, De Groot E, Kastelein JJP, Stroes ESG. Hepatic and cardiovascular consequences of familial hypobetalipoproteinemia. *Arterioscler Thromb Vasc Biol*. 2005;25(9):1979-1984. doi:10.1161/01.ATV.0000176191.64314.07.

12. Linton ME, V R, Young SG. Fam i l ial hypobetal i poprotei nem ia. *J Lipid Res*. 1993;34:521-541.

13. Anderson KM, Castelli WP, Levy D. Cholesterol and mortality. 30 years of follow-up from the Framingham study. *JAMA*. 1987;257(16):2176-2180. doi:10.1001/jama.257.16.2176.

14. Schonfeld G, Lin X, Yue P. Familial hypobetalipoproteinemia: Genetics and metabolism. *Cell Mol Life Sci*. 2005;62(12):1372-1378. doi:10.1007/s00018-005-4473-0.

15. Tarugi P, Averna M. *Hypobetalipoproteinemia: Genetics, Biochemistry, and Clinical Spectrum*. Vol 54.; 2011. doi:10.1016/B978-0-12-387025-4.00004-2.

16. Cariou B, Ouguerram K, Zaïr Y, Guerois R, Langhi C, Kourimate S, Benoit I, Le May C, Gayet C, Belabbas K, Dufernez F, Chétiveaux M, Tarugi P, Krempf M, Benlian P, Costet P. PCSK9 dominant negative mutant results in increased LDL catabolic rate and familial hypobetalipoproteinemia. *Arterioscler Thromb Vasc Biol*. 2009;29(12):2191-2197. doi:10.1161/ATVBAHA.109.194191.

17. Martín-Campos JM, Roig R, Mayoral C, Martinez S, Martí G, Arroyo JA, Julve J, Blanco-Vaca F. Identification of a novel mutation in the ANGPTL3 gene in two families diagnosed of familial hypobetalipoproteinemia without APOB mutation. *Clin Chim Acta*. 2012;413(5-6):552-555. doi:10.1016/j.cca.2011.11.020.

18. Sherva R, Yue P, Schonfeld G, Neuman RJ. Evidence for a quantitative trait locus affecting low levels of apolipoprotein B and low density lipoprotein on chromosome 10 in Caucasian families. *J Lipid Res*. 2007;48(12):2632-2639. doi:10.1194/jlr.M700078-JLR200.

19. Yue P, Tanoli T, Wilhelm O, Patterson B, Yablonskiy D, Schonfeld G. Absence of fatty liver in familial hypobetalipoproteinemia linked to chromosome 3p21. *Metabolism*. 2005;54(5):682-688. doi:10.1016/j.metabol.2004.12.013.

20. Gutiérrez-Cirlos C, Ordóñez-Sánchez ML, Tusié-Luna MT, Patterson BW, Schonfeld G, Aguilar-Salinas CA. Familial hypobetalipoproteinemia in a hospital survey: Genetics, metabolism and non-alcoholic fatty liver disease. *Ann Hepatol*. 2011;10(2):155-164.

21. Reddy M V, Iatan I, Weissglas-Volkov D, Nikkola E, Haas BE, Juvonen M, Ruel I, Sinsheimer JS, Genest J, Pajukanta P. Exome sequencing identifies 2 rare variants for low high-density lipoprotein cholesterol in an extended family. *Circ Cardiovasc Genet*. 2012;5(5):538-546. doi:10.1161/CIRCGENETICS.112.963264.

22. Li H, Durbin R. Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics*. 2010;26(5):589-595. doi:10.1093/bioinformatics/btp698.

23. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, Garimella K, Altshuler D, Gabriel S, Daly M, DePristo MA. The genome analysis toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res*. 2010;20(9):1297-1303. doi:10.1101/gr.107524.110.

24. Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res*. 2010;38(16):e164. doi:10.1093/nar/gkq603.

25. Adzhubei IA, Schmidt S, Peshkin L, Ramensky VE, Gerasimova A, Bork P, Kondrashov AS, Sunyaev SR. A method and server for predicting damaging missense mutations. *Nat Methods*. 2010;7(4):248-249. doi:10.1038/nmeth0410-248.

26. Ng PC, Henikoff S. SIFT: predicting amino acid changes that affect protein function. *Nucleic Acids Res*. 2003;31(13):3812-3814. doi:10.1093/nar/gkg509.

27. Howie B, Fuchsberger C, Stephens M, Marchini J, Abecasis GR. Fast and accurate genotype imputation in genome-wide association studies through pre-phasing. *Nat Genet*. 2012;44(8):955-959. doi:10.1038/ng.2354.

28. Zhang H, Meltzer P, Davis S. RCircos: an R package for Circos 2D track plots. *BMC Bioinformatics*. 2013;14(1):244. doi:10.1186/1471-2105-14-244.

29. Willer CJ, Schmidt EM, Sengupta S, Peloso GM, Gustafsson S, Kanoni S, Ganna A, Chen J, Buchkovich ML, Mora S, Beckmann JS, Bragg-Gresham JL, Chang H-Y, Demirkan A, Den Hertog HM, Do R, Donnelly LA, Ehret GB, Esko T, Feitosa MF, Ferreira T, Fischer K, Fontanillas P, Fraser RM, Freitag DF, Gurdasani D, Heikkilä K, Hyppönen E, Isaacs A, Jackson AU, Johansson A, Johnson T, Kaakinen M, Kettunen J, Kleber ME, Li X, Luan J, Lyytikäinen L-P, Magnusson PKE, Mangino M, Mihailov E, Montasser ME, Müller-Nurasyid M, Nolte IM, O'Connell JR, Palmer CD, Perola M, Petersen A-K, Sanna S, Saxena R, Service SK, Shah S, Shungin D, Sidore C, Song C, Strawbridge RJ, Surakka I, Tanaka T, Teslovich TM, Thorleifsson G, Van den Herik EG, Voight BF, Volcik KA, Waite LL, Wong A, Wu Y, Zhang W, Absher D, Asiki G, Barroso I, Been LF, Bolton JL, Bonnycastle LL, Brambilla P, Burnett MS, Cesana G, Dimitriou M, Doney ASF, Döring A, Elliott P, Epstein SE, Eyjolfsson GI, Gigante B, Goodarzi MO, Grallert H, Gravito ML, Groves CJ, Hallmans G, Hartikainen A-L, Hayward C, Hernandez D, Hicks AA, Holm H, Hung Y-J, Illig T, Jones MR, Kaleebu P, Kastelein JJP, et al. Discovery and refinement of loci associated with lipid levels. *Nat Genet*. 2013;45(11):1274-1283. doi:10.1038/ng.2797.

30. Park G, Gim J, Kim AR, Han K-H, Kim H-S, Oh S-H, Park T, Park W-Y, Choi BY. Multiphasic analysis of whole exome sequencing data identifies a novel mutation of ACTG1 in a nonsyndromic hearing loss family. *BMC Genomics*. 2013;14(1):191. doi:10.1186/1471-2164-14-191.

Chapter 4

Remote ischemic conditioning alters methylation and expression of cell cycle genes in aneurysmal subarachnoid hemorrhage

Remote Ischemic Conditioning Alters Methylation and Expression of Cell Cycle Genes in Aneurysmal Subarachnoid Hemorrhage

Elina Nikkola, BS; Azim Laiwalla, MS; Arthur Ko, BS; Marcus Alvarez, BS; Mark Connolly, BS; Yinn Cher Ooi, MD; William Hsu, PhD; Alex Bui, PhD; Päivi Pajukanta, MD, PhD; Nestor R. Gonzalez, MD

Background and Purpose—Remote ischemic conditioning (RIC) is a phenomenon in which short periods of nonfatal ischemia in 1 tissue confers protection to distant tissues. Here we performed a longitudinal human pilot study in patients with aneurysmal subarachnoid hemorrhage undergoing RIC by limb ischemia to compare changes in DNA methylation and transcriptome profiles before and after RIC.

Methods—Thirteen patients underwent 4 RIC sessions over 2 to 12 days after rupture of an intracranial aneurysm. We analyzed whole blood transcriptomes using RNA sequencing and genome-wide DNA methylomes using reduced representation bisulfite sequencing, both before and after RIC. We tested differential expression and differential methylation using an intraindividual paired study design and then overlapped the differential expression and differential methylation results for analyses of functional categories and protein–protein interactions.

Results—We observed 164 differential expression genes and 3493 differential methylation CpG sites after RIC, of which 204 CpG sites overlapped with 103 genes, enriched for pathways of cell cycle ($P<3.8\times10^{-4}$) and inflammatory responses ($P<1.4\times10^{-4}$). The cell cycle pathway genes form a significant protein–protein interaction network of tightly coexpressed genes ($P<0.00001$).

Conclusions—Gene expression and DNA methylation changes in aneurysmal subarachnoid hemorrhage patients undergoing RIC are involved in coordinated cell cycle and inflammatory responses. **(*Stroke*. 2015;46:2445-2451. DOI: 10.1161/STROKEAHA.115.009618.)**

Key Words: aneurysm ■ DNA methylation ■ genomics ■ preconditioning ■ subarachnoid hemorrhage ■ transcriptome

Remote ischemic conditioning (RIC) is a phenomenon where nonlethal ischemic exposure in a peripheral tissue induces a systemic protection of subsequent injuries in distant organs and tissues.[1] RIC has shown encouraging results in animal models by providing cardio- and neuroprotective effects against an ischemic injury, and thus RIC is emerging as an attractive novel therapeutic for clinical trials.[2–4] Recent human studies have confirmed the safety and feasibility of lower limb RIC in patients with aneurysmal subarachnoid hemorrhage (aSAH).[5,6] Based on our separate study (Laiwalla et al, unpublished data), the odds ratio of a good outcome for patients with RIC is 5.17 (95% confidence interval, 1.21–25.02) when compared with matched controls with SAH.

The effectiveness of RIC is likely to be caused by its multifactorial effects, and rodent studies suggest that these are mediated in part by a cascade of transcriptional and translational changes.[7] Activation of basic cell survival responses to transient ischemia causes a shift toward a protective genetic profile, leading to a differential regulation of genes involved in inflammation, neurotransmitter excitotoxicity, apoptosis, and cerebrovascular perfusion.[8–13] Nevertheless, the mechanisms by which RIC provides neuroprotective effects in human are not well understood. The involvement of humoral factors has been demonstrated in animals because the protection can be transferred from an RIC animal to a nonconditioned animal by whole blood transfusion.[14] Thus, genetic and epigenetic studies in human blood could elucidate the humeral processes catalyzed by RIC and furthermore provide potential diagnostic and therapeutic targets for the treatment and prevention of ischemic injury.

Received March 31, 2015; final revision received June 11, 2015; accepted July 2, 2015.

From the Department of Human Genetics (E.N., A.K., M.A., P.P.), Department of Neurosurgery (A.L., M.C., Y.C.O., N.R.G.), and Department of Radiological Sciences (W.H., A.B., N.R.G.), David Geffen School of Medicine at UCLA, Los Angeles, CA; and Department of Human Genetics and Molecular Biology, Molecular Biology Institute at UCLA, Los Angeles, CA (A.K., P.P.).

The online-only Data Supplement is available with this article at http://stroke.ahajournals.org/lookup/suppl/doi:10.1161/STROKEAHA.115.009618/-/DC1.

Correspondence to Nestor R. Gonzalez, MD, Associate Professor of Neurosurgery and Radiology, David Geffen School of Medicine at UCLA, 300 Stein Plaza, Suite 539, Los Angeles, CA 90095. E-mail ngonzalez@mednet.ucla.edu

© 2015 American Heart Association, Inc.

Stroke **is available at http://stroke.ahajournals.org** **DOI: 10.1161/STROKEAHA.115.009618**

In this human pilot study, we performed a prospective longitudinal evaluation in a group of patients with aSAH undergoing RIC to study the induced genomic responses by identifying and comparing blood DNA methylation and gene expression profiles before RIC and 1 week after RIC. Identification of factors altered by a transient limb RIC can provide insights into the mechanisms of neuroprotective action and, ultimately, may yield biomarkers for SAH prognosis and treatment.

Methods

Study Samples

Patients with aSAH were enrolled from the Remote Ischemic Preconditioning in Subarachnoid Hemorrhage Trial (Clinicaltrials.gov No. NCT01158508). The study was approved by the local institutional review board, and all participants gave a written informed consent. Patients 18 to 80 years old with SAH confirmed by computed tomography or lumbar puncture and presence of a ruptured intracranial aneurysm confirmed by computed tomography, magnetic resonance, or catheter angiography were considered for enrollment in this study. Patients who were pregnant or with a history or physical examination findings of peripheral vascular disease, deep venous thrombosis, peripheral neuropathy, or lower extremity bypass were excluded. Clinical characteristics are provided in Table 1.

RIC Protocol

Patients underwent 4 RIC sessions over 2 to 12 days after aneurysm rupture. RIC sessions were performed on the lower limb with a large adult-sized blood pressure cuff. Each session consisted of 4 inflation cycles lasting 5 minutes, followed by 5-minute deflations. Cuff pressure was originally inflated at 20 mm Hg over the patient's baseline systolic blood pressure, then increased until the dorsalis pedis pulse was abolished, as confirmed by a Doppler ultrasonography. This pressure was maintained for 5 minutes throughout the duration of the inflation cycle.

Peripheral blood samples were drawn from aSAH patients at 2 different time points: before RIC (baseline) and after 4 sessions of the RIC treatment. DNA and RNA were isolated according to standard protocols.

Aneurysm Controls

We included 24 control individuals with a history of intracranial aneurysms who never received RIC treatment. The blood collection and sample processing were performed in the same way as described for the aSAH cases above.

RNA Sequencing

We included 13 aSAH sample pairs and 24 aneurysm controls in the study after the initial quality control of the blood RNA (RNA integrity number [RIN] value >7, RNA concentration >10 ng/μL). The blood RNA sequencing libraries were prepared using Illumina TruSeq RNA library kit, and sequencing of the paired-end, 100-bp reads was performed using the Illumina Hiseq2000 platform, resulting in on average 46.1 mol/L reads per sample. We used STAR[15] to align the fastq files to the human GRCh37/hg19 reference genome with the following settings: the maximum intron size was set at 500 kb; the minimum intron size was set at 20; and we allowed for 4 mismatches. We used HTSeq (version HTSeq-0.6.1)[16] to produce raw counts.

Differential Expression Using EdgeR and DESeq2

We used both EdgeR[17] and DESeq2[18] R-packages to identify differentially expressed (DE) genes using the paired sample design and focused on their overlap to obtain a set of highly confident DE genes. First, using EdgeR, we excluded the genes that did not have one count per million reads in at least 50% of the samples. We normalized the read count values using trimmed means of *M* value and estimated common, trended, and tagwise dispersions using R software (version RX64 3.0.2). Together, these quality control steps removed genes with low expression and normalized the libraries for library size and biological variability, resulting in 14 816 genes for our subsequent analyses. We determined DE using the generalized linear model likelihood ratio test using a significance threshold of FDR <0.05.

Second, similarly as in EdgeR, we used a multifactor design with DESeq2. We estimated the size factors and dispersions and performed negative binomial generalized linear model fitting for the sample as a factor and Wald statistics for DE. We used Benjamini–Hochberg–adjusted $P<0.05$ as a threshold for significance.

To compare the aSAH patients with aneurysm controls, we considered only the genes DE between the aSAH baseline and after the treatment. We performed 2 separate DE analyses using negative binomial and determined DE using Wald test for (1) the aSAH baseline group versus the controls and (2) the aSAH RIC treatment group versus the controls (Figure I in the online-only Data Supplement). The genes changing the DE status between the 2 analyses (ie, the genes that were not DE between the aSAH baseline group and controls, but became DE when comparing the

Table 1. Clinical Characteristics of the aSAH Patients

Subject ID	Age	Sex	Smoking	Alcohol	Hypertension	T2D	Vasospasm	Clinical Functional Outcome
SAH 551	61	F	No	No	No	No	N	Improved or no change
SAH 553	77	F	No	No	No	No	Y	Improved or no change
SAH 554	56	M	Former	No	No	No	Y	Improved or no change
SAH 555	53	F	No	No	No	No	Y	Deteriorated
SAH 556	23	F	No	No	No	No	N	Improved or no change
SAH 557	47	F	No	No	Yes	No	Y	Deteriorated
SAH 558	65	F	No	No	Yes	No	Minimal	Improved or no change
SAH 559	43	F	Yes	Yes	Yes	No	Y	Improved or no change
SAH 5510	36	M	Yes	Yes	No	No	Y	Improved or no change
SAH 5511	51	M	Yes	Yes	Yes	Yes	N	Improved or no change
SAH 5512	43	M	Yes	No	Yes	Yes	Y	Improved or no change
SAH 5513	60	F	No	No	Yes	Yes	Y	Deteriorated
SAH 5514	51	M	Yes	Yes	Yes	No	Suspected	Improved or no change

SAH indicates subarachnoid hemorrhage; and T2D, type 2 diabetes mellitus.

aSAH treatment group with the controls) were carried forward to subsequent analyses.

Methylation

We analyzed blood DNA methylation profiles by reduced representation bisulfite sequencing (RRBS). RRBS libraries from human genomic DNA were prepared as previously described.[19] Briefly, we treated blood DNA with sodium bisulfite (Epitecht Illumina), digested it with the MspI enzyme, and selected fragments averaging 100 to 250 bp. We multiplexed 4 samples per lane and sequenced the libraries using single-end 100-bp reads with the Illumina Hiseq2000 platform, resulting in on average 25.1 mol/L reads per sample.

We performed initial QC for fastq files using FastaQC. We used BS-seeker2[20] with Bowtie2[21] for RRBS alignment using hg19 as a reference genome. For alignment, we considered in silico MspI fragments between 40 and 500 bp to cover all possible MspI fragments from the RRBS libraries. We aligned the reads using the Bowtie2 end-to-end alignment mode by allowing 4 mismatches. We called the methylation status of the individual CpG sites (percentage of methylated cells) by requiring at least 10 reads per a CpG site. Pearson correlation coefficient was used to estimate pair-wise correlations in methylation sites between the individuals. Paired Students *t*-tests were conducted to compare between-group and within-group differences. The CpG sites passing a 2-tailed nominal $P<0.01$ were considered significant and carried forward for subsequent analyses. Finally, we used BEDTOOLS[22] to overlap the DE genes with methylated regions.

Functional Annotation and Coexpression of the Pathway Genes

We used DAVID software[23,24] to search for functional categories of the DE genes. To highlight the most relevant gene ontology terms associated with the overlapped DE and differential methylation (DM) gene lists, we performed a batch annotation and gene-GO term enrichment analysis. We searched for protein–protein interaction (PPI) networks using STRING v9.1.[25] We used Pearson correlation coefficient to estimate correlations between the pathway genes and ggplot2 and reshape2 to visualize these results. Reactome[26,27] was used to explore specific pathways.

Results

The overall study design is shown in Figure 1. To identify genomic mechanisms for the effects of RIC in aSAH patients, we used a paired sample design where each patient gave blood samples before and after 4 RIC sessions. Using this longitudinal study design, each individual functions as a control for him-/herself in the DE and DM analyses. Accordingly, we were able to adjust for potential confounding factors, such as age, smoking, medication, and ethnicity, using this intraindividual paired design. We analyzed the blood RNA expression and DNA methylation profiles of each patient before RIC and 1 week after the RIC treatment started. We compared these profiles to the ones of the controls who did not receive any RIC treatments. Finally, we overlapped the DE genes with DM sites and performed functional annotations and PPIs analysis of the overlapping genes (Figure 1).

Differential Expression

We found 451 DE genes after RIC (FDR<0.05) consistently using both EdgeR and DESeq2, of which 205 were upregulated after the RIC treatment and 246 were downregulated, respectively (Figure 2; and Table I in the online-only Data Supplement). Next, to identify genes responding to the RIC treatment, we tested the expression of the 451 genes in the controls for DE against their expression at both the aSAH baseline and after the RIC treatment, considering a Bonferroni-corrected $P<1.1\times10^{-4}$ ($P<0.05/451$ DE genes) significant. We found 164 DE genes (see Table II in the online-only Data Supplement for the list of DE genes) between the controls and aSAH patients before and after RIC treatment, suggesting that these genes may contribute to the response to the RIC treatment.

Differential Methylation

We were able to map on average 66% of reads/sample to the human genome, which is in accordance with previous RRBS studies.[20,28] The resulting methylation profiles per sample covered on average 1 764 402 CpG sites, of which 676 543 were assayed in all individuals. The overall methylation status changed little within an individual (≈98%) and between individuals (≈97.5%; Figure II in the online-only Data Supplement), suggesting that methylation is a stable phenomenon and only a small number of sites are actively responding to environmental factors. We focused on the 403 546 CpG sites that altered by >10% in at least one individual after RIC. To test DM cytosines between the baseline and after treatment, we used 2-tailed paired student's *t* test. A total of 3493 CpG sites were DM ($P<0.01$).

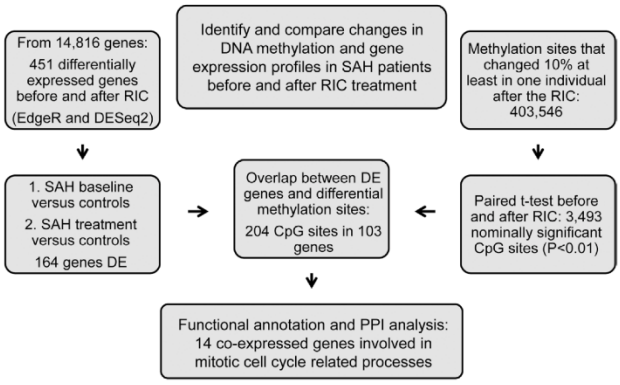


Figure 1. A schematic overview of study design and results. DE indicates differential expression; PPI, protein–protein interaction; RIC, remote ischemic conditioning; and SAH, subarachnoid hemorrhage.

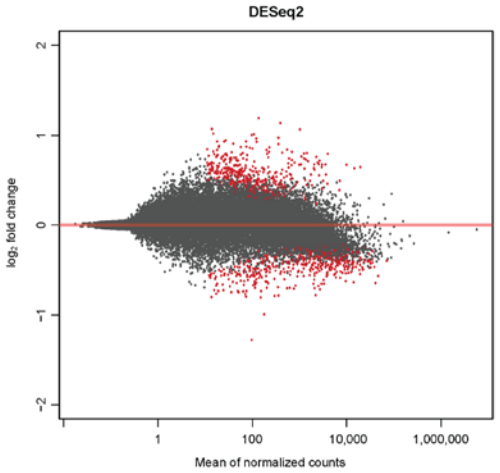


Figure 2. The differentially expressed (DE) genes between the aneurysmal subarachnoid hemorrhage (aSAH) baseline and a week after remote ischemic conditioning (RIC) treatment. Red dots indicate DE genes with an FDR <0.05.

Overlapping the DE and DM Regions

When we overlapped (defined ±250 kb from the each DE gene) the DM CpGs with the DE genes, we found 204 CpG sites corresponding to 103 DE and DM genes, suggesting methylation as a potential mechanism for DE (Table III in the online-only Data Supplement). Furthermore, 52 of the genes had >1 nearby DM site.

Functional Annotation and Coexpression of the Pathway Genes

Functional annotation with DAVID software showed that the overlapping 103 DE and DM genes are enriched for defense and inflammatory responses (Benjamini–Hochberg [B-H]–corrected $P<1.4\times10^{-4}$) and for cell cycle and mitosis (B-H–corrected $P<3.8\times10^{-4}$; Table 2). In addition, we examined the PPIs of the 103 DE genes using String (Figure 3). We found a significant enrichment for PPIs and one large network consisting of 21 DE and DM genes (Figure 3), of which 14 are part of the cell cycle pathway from the functional enrichment analysis (Table 2). We also found 2 smaller PPIs consisting of 3 proteins each: *CEBPB*, *HDAC4*, *PPARG* and *AZU1*, *CTSG*, *MPO* (Figure 3), all present in the significant pathways of defense and inflammatory response mechanisms (Table 2).

Table 2. Functional Annotations of the 103 Identified DE Genes Using the David Pathway Tool

	Gene Count	B-H Corrected *P* Value
Enrichment score 4.35		
Cluster 1		
Defense response	18	1.4×10^{-4}
Inflammatory response	10	1.7×10^{-2}
Enrichment score 3.95		
Cluster 2		
Cell cycle	19	3.8×10^{-4}
M phase	12	1.7×10^{-3}
Cell cycle phase	13	1.9×10^{-3}
Nuclear division	10	1.7×10^{-3}
Mitosis	10	1.7×10^{-3}

B-H indicates *P* value after Benjamini–Hochberg correction for false discovery rate; and DE, differential expression.

Next, we further examined the 14 cell cycle pathway genes for correlations between their gene expressions. These genes exhibited highly dynamic correlation shifts, with substantially tighter correlations after the RIC treatment (Figure 4), suggesting that different phases of cell cycle pathway are turned on as a result of RIC. Interestingly, when we visualized the coexpression of these genes in the control group, we observed a clear difference in their correlations when compared with the aSAH patients at baseline and even more after the treatment (Figure 4), indicating the involvement of these genes in aSAH and the potential influence of the RIC treatment.

Based on a more detailed Reactome pathway analysis (Table IV in the online-only Data Supplement), 8 of the 14 genes ($FDR<1.0\times10^{-5}$) are involved in the cell cycle pathway (*SPC24*, *ESPL1*, *CLSPN*, *CDC45*, *CENPF*, *FOXM1*, *CDK1*, *RAD51*). In the Reactome analysis, *CDK1* acts as a key regulator of specific mitotic cell cycle pathways. For instance, we observed that *CDK1* is involved in G2/M transition and mitotic G2-G2/M phases with *CENPF* and *FOXM1*, regulating the G2/M checkpoints with *CLSPN* and *CDC45*. In addition, *CDK1* is involved in processes such as kinetochore assembly in mitotic prometaphase and M Phase with *SPC24*, *CENPF*, and *ESPL1* (Table IV in the online-only Data Supplement). *CDK1* is also present in numerous activation and signaling pathways within mitotic cell cycle pathway (Table IV in the online-only Data Supplement).

Discussion

We performed the first longitudinal and systematic genome-wide pilot study in humans comparing gene expression and methylation changes after RIC in aSAH. We found 164 DE genes and 3493 DM CpG sites that are modified, potentially as a result of RIC. When we overlapped these regions, we observed 204 DM CpG sites corresponding to 103 DE genes, suggesting methylation as a potential mechanism regulating gene expression. These genes were enriched for cell cycle–related processes, as well as for defense and inflammatory responses. Furthermore, the identified 14 cell cycle genes exhibited highly correlated expression signals after RIC (Figure 4). Overall, these findings provide first insights into the neuroprotective molecular mechanisms underlying RIC in humans.

Our prior work has demonstrated RIC-induced metabolic changes in the preconditioned limb, as well as cerebral tissue.[29,30] Muscle microdialysis during RIC showed an increase in lactate/pyruvate ratio and lactate, without change in glycerol.[29] Cerebral microdialysis during RIC showed a decrease in lactate/pyruvate ratio and glycerol, which persisted after the last RIC session.[30] Identification of markers of the RIC effects beyond local factors is imperative for determining appropriate end points in future RIC clinical studies.

Whole-genome transcriptional analysis has been applied to uncover genetic changes underlying ischemia-induced

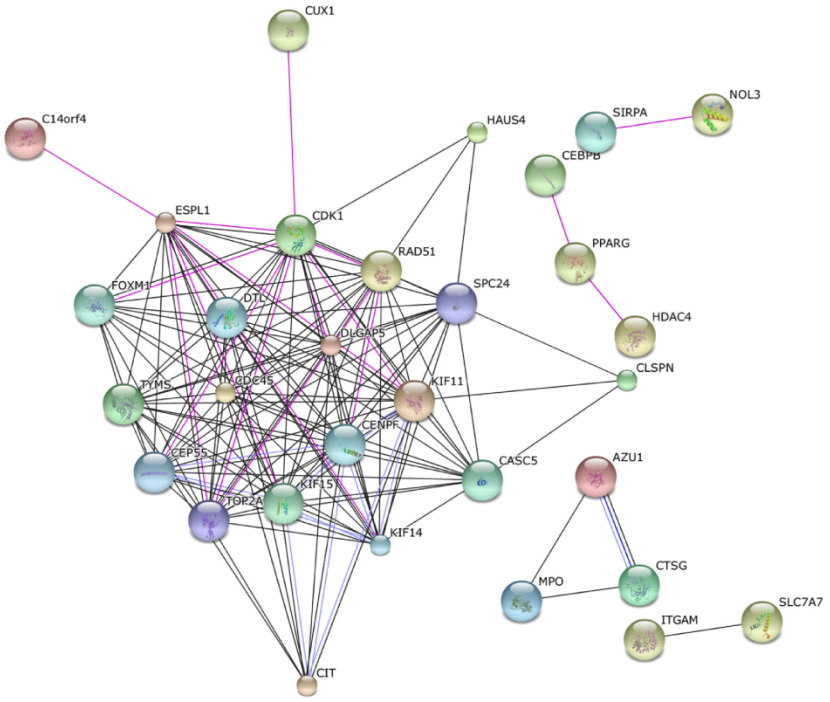


Figure 3. Protein–protein interactions (PPIs) of the 103 differentially expressed (DE) genes using String. We observed a statistically significant (P<0.00001) enrichment of PPIs of the DE genes residing in the cell cycle, defense, and inflammatory response pathways, all passing the Benjamini correction as shown in Table 2.

neuroprotective effects in animal models.[31,32] DNA methylation changes of gene promoter regions have also been investigated to uncover preconditioning-induced epigenetic changes, contributing to neuroprotection in mice.[33] Although proof of concept animal studies have given great insight into the potential mechanisms of RIC, they do not necessarily translate directly to humans, and thus, human studies are essential to evaluate effects in clinical settings.

Cell cycle machinery and related molecules have been previously implicated in ischemic neuronal death,[34,35] and irregular cell cycle activation has been implicated in stroke.[36–38] We show evidence for involvement of genes in cell cycle processes regulated by *CDK1* in the acute stage of aSAH, possibly modified by RIC (Figure 4). We postulate that this pathway may be important in various forms of ischemia. Furthermore, we hypothesize that RIC may induce a release of substances from the ischemic limb muscles to blood. These in turn stimulate white blood cells, such as macrophages, to increase the expression of genes involved in cell cycle and cell proliferation. Subsequently, these white blood cells may stream to the

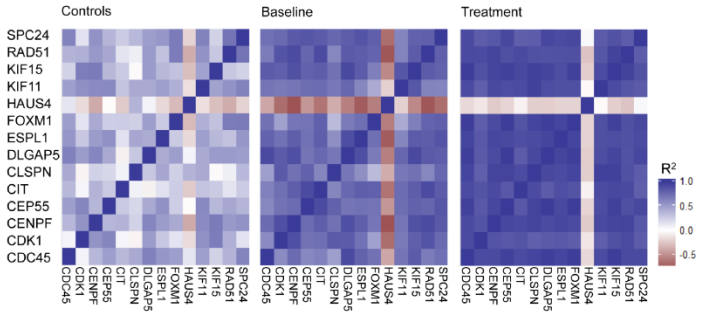


Figure 4. Co-expression analysis of the 14 mitotic cell cycle genes identified in the Protein–protein interactions (PPIs) and functional enrichment analyses.

ischemic location in brain and release substances, including growth factors and other cytokines, to protect brain from further apoptosis. This mechanism could lead to the neuroprotective effects of RIC, although additional functional studies are warranted to verify the underlying mechanisms.

One of the mechanisms proposed for RIC is inflammatory responses.[39,40] In accordance with this, our DAVID pathway analysis implicated a set of 18 both DE and DM genes in defense response pathways (*CEBPB*, *AZU1*, *BPI*, *CTSG*, *CRISP3*, *CYSLTR1*, *HDAC4*, *INHBA*, *IL1R1*, *IL10RB*, *LTF*, *MPO*, *OLR1*, *PPARG*, *PROK2*, *STAT5B*, *STAB1*, and *TLR5*). Six of these genes were also involved in 2 separate PPIs (Figure 3).

A recent study exploring human plasma proteome in RIC found that cysteine-rich secretory protein 3 (CRISP-3) was increased in serum after RIC in 6 adults.[41] This is consistent with our finding of over a 2-fold increase of *CRISP3* gene expression in blood followed by RIC (Table I in the online-only Data Supplement), suggesting its role as a humoral RIC mediator and surrogate marker. CRISP-3 is a glycoprotein present in exocrine secretions, bone marrow, secretory granules of neutrophils, and plasma bound to a1B glycoprotein.[42,43] Although its complete function is unknown, it is thought to act in innate immune response and as a prostate cancer marker.[42,43]

In summary, in this first pilot study, using a longitudinal design to investigate genome-wide expression and methylation changes in aSAH patients after RIC, we found evidence for coordinated expression and methylation changes of a small set of key genes in mitotic cell cycle, defense, and inflammatory responses. We have limitations in this study, and therefore, the results presented here should be further investigated and verified in future considerably larger genomic studies. In addition to the small sample size, we recognize that some of the observed changes in genes expression and methylation are potentially because of other medical treatments these patients received in the hospital, and hence, future studies should comprise a randomization that includes patients not receiving any RIC treatment as controls. We also recognize that differences in blood cell types may contribute to the changes in DNA methylation and gene expression, and thus future RIC studies should include analysis of separate fluorescence activated cell sorting-sorted cells. Nevertheless, longitudinal genome-wide studies of stroke, and especially SAH, integrating expression and methylation changes at the genome-wide level are still sparse, and thus our study provides valuable initial data, starting to elucidate the largely unknown mechanisms underlying RIC in humans.

Sources of Funding

This work was supported by the Ruth and Raymond Stotter Endowement, the National Institutes of Health National Institute of Neurological Disorders and Stroke (NINDS) grant K23NS079477, National Institute of Biomedical Imaging and Bioengineering (NIBIB) grant R01EB000362, and the National Heart, Lung, and Blood Institute (NHLBI) grants HL-28481 and HL-095056.

Disclosures

None.

References

1. Przyklenk K, Bauer B, Ovize M, Kloner RA, Whittaker P. Regional ischemic 'preconditioning' protects remote virgin myocardium from subsequent sustained coronary occlusion. *Circulation*. 1993;87:893–899.
2. Ren C, Gao X, Steinberg GK, Zhao H. Limb remote-preconditioning protects against focal ischemia in rats and contradicts the dogma of therapeutic time windows for preconditioning. *Neuroscience*. 2008;151:1099–1103. doi: 10.1016/j.neuroscience.2007.11.056.
3. Li SJ, Wu YN, Kang Y, Yin YQ, Gao WZ, Liu YX, et al. Noninvasive limb ischemic preconditioning protects against myocardial I/R injury in rats. *J Surg Res*. 2010;164:162–168. doi: 10.1016/j.jss.2009.03.017.
4. Wei D, Ren C, Chen X, Zhao H. The chronic protective effects of limb remote preconditioning and the underlying mechanisms involved in inflammatory factors in rat stroke. *PLoS One*. 2012;7:e30892. doi: 10.1371/journal.pone.0030892.
5. Koch S, Katsnelson M, Dong C, Perez-Pinzon M. Remote ischemic limb preconditioning after subarachnoid hemorrhage: a phase Ib study of safety and feasibility. *Stroke*. 2011;42:1387–1391. doi: 10.1161/STROKEAHA.110.605840.
6. Gonzalez NR, Connolly M, Dusick JR, Bhakta H, Vespa P. Phase I clinical trial for the feasibility and safety of remote ischemic conditioning for aneurysmal subarachnoid hemorrhage. *Neurosurgery*. 2014;75:590–598, discussion 598. doi: 10.1227/NEU.0000000000000514.
7. Sarabi AS, Shen H, Wang Y, Hoffer BJ, Bäckman CM. Gene expression patterns in mouse cortical penumbra after focal ischemic brain injury and reperfusion. *J Neurosci Res*. 2008;86:2912–2924. doi: 10.1002/jnr.21734.
8. Konstantinov IE, Arab S, Kharbanda RK, Li J, Cheung MM, Cherepanov V, et al. The remote ischemic preconditioning stimulus modifies inflammatory gene expression in humans. *Physiol Genomics*. 2004;19:143–150. doi: 10.1152/physiolgenomics.00046.2004.
9. Zhang J, Qian H, Zhao P, Hong SS, Xia Y. Rapid hypoxia preconditioning protects cortical neurons from glutamate toxicity through delta-opioid receptor. *Stroke*. 2006;37:1094–1099. doi: 10.1161/01.STR.0000206444.29930.18.
10. Liu YX, Zhang M, Liu LZ, Cui X, Hu YY, Li WB. The role of glutamate transporter-1a in the induction of brain ischemic tolerance in rats. *Glia*. 2012;60:112–124. doi: 10.1002/glia.21252.
11. Ding ZM, Wu B, Zhang WQ, Lu XJ, Lin YC, Geng YJ, et al. Neuroprotective effects of ischemic preconditioning and postconditioning on global brain ischemia in rats through the same effect on inhibition of apoptosis. *Int J Mol Sci*. 2012;13:6089–6101. doi: 10.3390/ijms13056089.
12. Zhang N, Yin Y, Han S, Jiang J, Yang W, Bu X, et al. Hypoxic preconditioning induced neuroprotection against cerebral ischemic injuries and its cPKCγ-mediated molecular mechanism. *Neurochem Int*. 2011;58:684–692. doi: 10.1016/j.neuint.2011.02.007.
13. Nakamura H, Katsumata T, Nishiyama Y, Otori T, Katsura K, Katayama Y. Effect of ischemic preconditioning on cerebral blood flow after subsequent lethal ischemia in gerbils. *Life Sci*. 2006;78:1713–1719. doi: 10.1016/j.lfs.2005.08.008.
14. Dickson EW, Reinhardt CP, Renzi FP, Becker RC, Porcaro WA, Heard SO. Ischemic preconditioning may be transferable via whole blood transfusion: preliminary evidence. *J Thromb Thrombolysis*. 1999;8:123–129.
15. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics*. 2013;29:15–21. doi: 10.1093/bioinformatics/bts635.
16. Anders S, Pyl PT, Huber W. HTSeq–a Python framework to work with high-throughput sequencing data. *Bioinformatics*. 2015;31:166–169. doi: 10.1093/bioinformatics/btu638.
17. Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*. 2010;26:139–140. doi: 10.1093/bioinformatics/btp616.
18. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol*. 2014;15:550. doi: 10.1186/s13059-014-0550-8.
19. Chen PY, Ganguly A, Rubbi L, Orozco LD, Morselli M, Ashraf D, et al. Intrauterine calorie restriction affects placental DNA methylation and gene expression. *Physiol Genomics*. 2013;45:565–576. doi: 10.1152/physiolgenomics.00034.2013.
20. Guo W, Fiziev P, Yan W, Cokus S, Sun X, Zhang MQ, et al. BS-Seeker2: a versatile aligning pipeline for bisulfite sequencing data. *BMC Genomics*. 2013;14:774. doi: 10.1186/1471-2164-14-774.

21. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. *Nat Methods*. 2012;9:357–359. doi: 10.1038/nmeth.1923.
22. Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics*. 2010;26:841–842. doi: 10.1093/bioinformatics/btq033.
23. Huang da W, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc*. 2009;4:44–57. doi: 10.1038/nprot.2008.211.
24. Huang da W, Sherman BT, Lempicki RA. Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Res*. 2009;37:1–13. doi: 10.1093/nar/gkn923.
25. Franceschini A, Szklarczyk D, Frankild S, Kuhn M, Simonovic M, Roth A, et al. STRING v9.1: protein-protein interaction networks, with increased coverage and integration. *Nucleic Acids Res*. 2013;41(Database issue):D808–D815. doi: 10.1093/nar/gks1094.
26. Milacic M, Haw R, Rothfels K, Wu G, Croft D, Hermjakob H, et al. Annotating cancer variants and anti-cancer therapeutics in reactome. *Cancers (Basel)*. 2012;4:1180–1211. doi: 10.3390/cancers4041180.
27. Croft D, Mundo AF, Haw R, Milacic M, Weiser J, Wu G, et al. The Reactome pathway knowledgebase. *Nucleic Acids Res*. 2014;42(Database issue):D472–D477. doi: 10.1093/nar/gkt1102.
28. Doherty R, Couldrey C. Exploring genome wide bisulfite sequencing for DNA methylation analysis in livestock: a technical assessment. *Front Genet*. 2014;5:126. doi: 10.3389/fgene.2014.00126.
29. Bilgin-Freiert A, Dusick JR, Stein NR, Etchepare M, Vespa P, Gonzalez NR. Muscle microdialysis to confirm sublethal ischemia in the induction of remote ischemic preconditioning. *Transl Stroke Res*. 2012;3:266–272. doi: 10.1007/s12975-012-0153-1.
30. Gonzalez NR, Hamilton R, Bilgin-Freiert A, Dusick J, Vespa P, Hu X, et al. Cerebral hemodynamic and metabolic effects of remote ischemic preconditioning in patients with subarachnoid hemorrhage. *Acta Neurochir Suppl*. 2013;115:193–198. doi: 10.1007/978-3-7091-1192-5_36.
31. Stenzel-Poore MP, Stevens SL, Xiong Z, Lessov NS, Harrington CA, Mori M, et al. Effect of ischaemic preconditioning on genomic response to cerebral ischaemia: similarity to neuroprotective strategies in hibernation and hypoxia-tolerant states. *Lancet*. 2003;362:1028–1037. doi: 10.1016/S0140-6736(03)14412-1.
32. Stenzel-Poore MP, Stevens SL, Simon RP. Genomics of preconditioning. *Stroke*. 2004;35(11 suppl 1):2683–2686. doi: 10.1161/01.STR.0000143735.89281.bb.
33. Zhang S, Zhang Y, Jiang S, Liu Y, Huang L, Zhang T, et al. The effect of hypoxia preconditioning on DNA methyltransferase and PP1γ in hippocampus of hypoxia preconditioned mice. *High Alt Med Biol*. 2014;15:483–490. doi: 10.1089/ham.2014.1042.
34. Rashidian J, Iyirhiaro GO, Park DS. Cell cycle machinery and stroke. *Biochim Biophys Acta*. 2007;1772:484–493. doi: 10.1016/j.bbadis.2006.11.009.
35. Wang W, Bu B, Xie M, Zhang M, Yu Z, Tao D. Neural cell cycle dysregulation and central nervous system diseases. *Prog Neurobiol*. 2009;89:1–17. doi: 10.1016/j.pneurobio.2009.01.007.
36. Osuga H, Osuga S, Wang F, Fetni R, Hogan MJ, Slack RS, et al. Cyclin-dependent kinases as a therapeutic target for stroke. *Proc Natl Acad Sci U S A*. 2000;97:10254–10259. doi: 10.1073/pnas.170144197.
37. Wang F, Corbett D, Osuga H, Osuga S, Ikeda JE, Slack RS, et al. Inhibition of cyclin-dependent kinases improves CA1 neuronal survival and behavioral performance after global ischemia in the rat. *J Cereb Blood Flow Metab*. 2002;22:171–182. doi: 10.1097/00004647-200202000-00005.
38. Rashidian J, Iyirhiaro G, Aleyasin H, Rios M, Vincent I, Callaghan S, et al. Multiple cyclin-dependent kinases signals are critical mediators of ischemia/hypoxic neuronal death *in vitro* and in vivo. *Proc Natl Acad Sci U S A*. 2005;102:14080–14085. doi: 10.1073/pnas.0500099102.
39. Bowen KK, Naylor M, Vemuganti R. Prevention of inflammation is a mechanism of preconditioning-induced neuroprotection against focal cerebral ischemia. *Neurochem Int*. 2006;49:127–135. doi: 10.1016/j.neuint.2006.02.011.
40. Carr-White G, Koh T, DeSouza A, Haxby E, Kemp M, Hooper J, et al. Chronic stable ischaemia protects against myocyte damage during beating heart coronary surgery. *Eur J Cardiothorac Surg*. 2004;25:772–778. doi: 10.1016/j.ejcts.2004.02.011.
41. Helgeland E, Breivik LE, Vaudel M, Svendsen ØS, Garberg H, Nordrehaug JE, et al. Exploring the human plasma proteome for humoral mediators of remote ischemic preconditioning–a word of caution. *PLoS One*. 2014;9:e109279. doi: 10.1371/journal.pone.0109279.
42. Ellias MF, Zainal Ariffin SH, Karsani SA, Abdul Rahman M, Senafi S, Megat Abdul Wahab R. Proteomic analysis of saliva identifies potential biomarkers for orthodontic tooth movement. *ScientificWorldJournal*. 2012;2012:647240. doi: 10.1100/2012/647240.
43. Udby L, Sørensen OE, Pass J, Johnsen AH, Behrendt N, Borregaard N, et al. Cysteine-rich secretory protein 3 is a ligand of alpha1B-glycoprotein in human plasma. *Biochemistry*. 2004;43:12877–12886. doi: 10.1021/bi048823e.

ONLINE SUPPLEMENT

Remote Ischemic Conditioning Alters Methylation and Expression of Cell Cycle Genes in Aneurysmal Subarachnoid Hemorrhage

Elina Nikkola, Azim Laiwalla, Arthur Ko, Marcus Alvarez, Mark Connolly, Yinn Cher Ooi, William Hsu, Alex Bui, Päivi Pajukanta, and Nestor Gonzalez

Supplemental Tables

Table I. The 451 differentially expressed genes passing FDR<0.05. The results shown here are from DESeq2 R-package.

Gene	Mean	log2FC	Stat	B-H Adjusted p-value
HTRA1	98	-1.28	-6.73	$2.47x10^{-7}$
MYB	121	0.96	6.23	$3.38x10^{-6}$
HRH4	123	0.92	5.98	$1.09x10^{-5}$
MAK	621	-0.70	-5.93	$1.10x10^{-5}$
MYO7A	177	-1.00	-5.80	$1.61x10^{-5}$
OLR1	136	1.19	5.82	$1.61x10^{-5}$
CRISP3	394	1.13	5.66	$2.93x10^{-5}$
FBN1	27	0.94	5.65	$2.93x10^{-5}$
CD52	2032	0.63	5.33	$1.46x10^{-4}$
RSAD2	1025	1.07	5.19	$2.68x10^{-4}$
SERPINB10	108	1.01	5.13	$3.24x10^{-4}$
STAT5B	13869	-0.43	-5.04	$4.74x10^{-4}$
SUFU	566	-0.41	-5.04	$4.74x10^{-4}$
MRVI1	2714	-0.72	-5.00	$5.38x10^{-4}$
BCL2L15	74	0.93	4.95	$6.44x10^{-4}$
SSH1	2477	-0.49	-4.92	$7.15x10^{-4}$
IDO1	98	1.00	4.88	$8.44x10^{-4}$
ANLN	39	0.91	4.85	$8.96x10^{-4}$
DNAH17	1640	-0.56	-4.85	$8.96x10^{-4}$
TGFBI	3328	-0.66	-4.83	$9.26x10^{-4}$
HAUS4	1226	-0.57	-4.80	$9.70x10^{-4}$
RP11-473M20.5	362	-0.79	-4.81	$9.70x10^{-4}$
MANSC1	2500	-0.68	-4.78	$1.03x10^{-3}$
MS4A3	355	0.97	4.76	$1.12x10^{-3}$
RP5-968J1.1	49	-0.73	-4.75	$1.15x10^{-3}$
ETS2	2895	-0.65	-4.72	$1.20x10^{-3}$
P2RY14	230	0.86	4.72	$1.20x10^{-3}$
ATP2B4	3792	-0.57	-4.69	$1.36x10^{-3}$
BRIP1	26	0.79	4.66	$1.52x10^{-3}$
CD163	2193	-0.78	-4.60	$1.97x10^{-3}$
ACSL1	40710	-0.65	-4.58	$2.14x10^{-3}$
RP11-1D12.2	26	0.93	4.57	$2.19x10^{-3}$
FAM129A	19351	-0.55	-4.54	$2.28x10^{-3}$
NT5DC4	71	-0.77	-4.55	$2.28x10^{-3}$
ADAM19	4928	-0.48	-4.54	$2.32x10^{-3}$
DGKD	4162	-0.38	-4.52	$2.40x10^{-3}$
ABCA2	2913	-0.48	-4.50	$2.56x10^{-3}$
HINT1	540	0.54	4.46	$2.91x10^{-3}$
TRIM7	122	-0.64	-4.45	$2.94x10^{-3}$
PIWIL4	117	0.63	4.45	$2.96x10^{-3}$
ST14	819	-0.67	-4.44	$2.96x10^{-3}$

IFI44L	357	0.90	4.43	$3.00x10^{-3}$
C9orf123	134	0.56	4.38	$3.66x10^{-3}$
CASC5	57	0.78	4.38	$3.66x10^{-3}$
CYP1B1-AS1	76	-0.73	-4.36	$3.93x10^{-3}$
AL137145.1	205	-0.70	-4.35	$3.98x10^{-3}$
FUT4	360	0.45	4.35	$3.98x10^{-3}$
MKNK1	4681	-0.44	-4.33	$4.17x10^{-3}$
NUCB2	302	0.58	4.33	$4.17x10^{-3}$
INHBA	20	0.88	4.32	$4.30x10^{-3}$
AC061992.1	192	-0.63	-4.31	$4.30x10^{-3}$
CCNJL	2099	-0.70	-4.30	$4.30x10^{-3}$
GBP4	1456	0.80	4.30	$4.30x10^{-3}$
MAD2L1	35	0.75	4.31	$4.30x10^{-3}$
STAB1	3278	-0.61	-4.30	$4.32x10^{-3}$
BMF	1012	-0.33	-4.27	$4.65x10^{-3}$
MRPL32	137	0.50	4.27	$4.65x10^{-3}$
PLK4	27	0.73	4.27	$4.65x10^{-3}$
ACAT1	184	0.46	4.26	$4.76x10^{-3}$
RP11-704M14.1	75	0.87	4.25	$4.93x10^{-3}$
ACOT11	37	0.63	4.23	$4.98x10^{-3}$
CLMN	1191	-0.44	-4.23	$4.98x10^{-3}$
CMPK2	431	0.80	4.23	$4.98x10^{-3}$
IFI44	544	0.84	4.24	$4.98x10^{-3}$
SKA2	133	0.52	4.24	$4.98x10^{-3}$
RFX2	2047	-0.61	-4.23	$4.99x10^{-3}$
BATF2	224	0.81	4.20	$5.27x10^{-3}$
RP11-443B7.1	198	-0.57	-4.21	$5.27x10^{-3}$
RP11-473M20.7	5443	-0.55	-4.21	$5.27x10^{-3}$
AL353791.1	104	-0.67	-4.19	$5.40x10^{-3}$
FABP6	74	-0.80	-4.18	$5.53x10^{-3}$
CES1	755	-0.71	-4.17	$5.57x10^{-3}$
ETV7	89	0.86	4.18	$5.57x10^{-3}$
MRPL13	60	0.55	4.18	$5.57x10^{-3}$
NLRP12	4658	-0.46	-4.17	$5.57x10^{-3}$
WDHD1	41	0.66	4.17	$5.57x10^{-3}$
POLE2	25	0.71	4.14	$6.04x10^{-3}$
COL17A1	101	0.85	4.13	$6.17x10^{-3}$
KIAA1524	34	0.70	4.14	$6.17x10^{-3}$
WDR76	98	0.60	4.11	$6.65x10^{-3}$
RNASE4	146	-0.55	-4.11	$6.72x10^{-3}$
SMTNL1	207	0.79	4.11	$6.72x10^{-3}$
GBP5	3964	0.79	4.09	$7.18x10^{-3}$
ATP8B4	407	0.57	4.08	$7.27x10^{-3}$
CCDC64B	151	-0.81	-4.08	$7.27x10^{-3}$
ELMO2	1763	-0.35	-4.07	$7.46x10^{-3}$
SLFN13	409	0.49	4.06	$7.71x10^{-3}$

ABCA13	197	0.83	4.05	$8.00x10^{-3}$
SNRPG	103	0.65	4.05	$8.00x10^{-3}$
CCT2	429	0.45	4.04	$8.14x10^{-3}$
ATP2C2	77	0.78	4.02	$8.46x10^{-3}$
SLC7A7	3945	-0.52	-4.03	$8.46x10^{-3}$
TAF9	229	0.45	4.00	$9.10x10^{-3}$
HMOX1	1708	-0.51	-4.00	$9.24x10^{-3}$
FANCI	196	0.50	3.99	$9.39x10^{-3}$
KIF15	24	0.76	3.99	$9.39x10^{-3}$
RP11-1334A24.6	556	-0.52	-3.99	$9.39x10^{-3}$
FAM53C	5007	-0.43	-3.98	$9.66x10^{-3}$
RNASE3	701	0.81	3.97	$9.66x10^{-3}$
VSIG4	299	-0.79	-3.97	$9.66x10^{-3}$
GFI1	218	0.65	3.97	$9.71x10^{-3}$
AREL1	2588	-0.33	-3.96	$1.02x10^{-2}$
PI4K2B	275	0.54	3.95	$1.02x10^{-2}$
KIF1B	2186	-0.43	-3.95	$1.03x10^{-2}$
KIAA0101	37	0.81	3.94	$1.06x10^{-2}$
HELLS	72	0.66	3.93	$1.08x10^{-2}$
SIRPA	13345	-0.47	-3.93	$1.09x10^{-2}$
CLC	1257	0.79	3.92	$1.11x10^{-2}$
KDM4B	5004	-0.42	-3.92	$1.11x10^{-2}$
THBS4	49	0.55	3.92	$1.11x10^{-2}$
TNFSF14	1975	-0.38	-3.92	$1.11x10^{-2}$
MAML3	738	-0.49	-3.91	$1.13x10^{-2}$
GBP1	2308	0.76	3.90	$1.16x10^{-2}$
ERG	63	0.80	3.89	$1.19x10^{-2}$
FAM136A	221	0.30	3.88	$1.19x10^{-2}$
RP11-701P16.2	113	-0.65	-3.89	$1.19x10^{-2}$
SAMD4A	72	0.63	3.88	$1.19x10^{-2}$
KB-1208A12.3	93	0.38	3.88	$1.20x10^{-2}$
PFKFB3	6018	-0.58	-3.87	$1.20x10^{-2}$
PRC1	90	0.62	3.88	$1.20x10^{-2}$
CR1	6177	-0.52	-3.86	$1.25x10^{-2}$
FBXO4	60	0.56	3.86	$1.25x10^{-2}$
DLGAP5	28	0.79	3.86	$1.26x10^{-2}$
EAF2	48	0.72	3.84	$1.32x10^{-2}$
HDAC4	2592	-0.37	-3.84	$1.32x10^{-2}$
NCAPG	41	0.75	3.84	$1.32x10^{-2}$
PLXDC2	1377	-0.45	-3.84	$1.32x10^{-2}$
SIRPB2	4171	-0.45	-3.84	$1.32x10^{-2}$
EPSTI1	599	0.74	3.83	$1.33x10^{-2}$
HPN	19	-0.74	-3.83	$1.33x10^{-2}$
IL1R1	1115	-0.52	-3.83	$1.33x10^{-2}$
LAMC1	134	-0.53	-3.83	$1.33x10^{-2}$
TGFA	959	-0.51	-3.83	$1.33x10^{-2}$

APLP2	13367	-0.41	-3.82	$1.33x10^{-2}$
CEP55	29	0.78	3.82	$1.33x10^{-2}$
PPARG	25	-0.73	-3.82	$1.35x10^{-2}$
CIT	46	0.75	3.81	$1.35x10^{-2}$
CTSG	1261	0.74	3.81	$1.35x10^{-2}$
ARHGAP19	870	-0.31	-3.81	$1.38x10^{-2}$
KIF13A	1756	-0.48	-3.80	$1.38x10^{-2}$
SSH2	13437	-0.44	-3.80	$1.38x10^{-2}$
PPT1	7521	-0.38	-3.80	$1.38x10^{-2}$
NSMCE2	92	0.52	3.79	$1.41x10^{-2}$
TLR5	1668	-0.50	-3.79	$1.43x10^{-2}$
CNTNAP3	1973	-0.66	-3.79	$1.44x10^{-2}$
AC108004.3	27	0.63	3.78	$1.45x10^{-2}$
CDCA7	59	0.61	3.78	$1.45x10^{-2}$
MMP14	124	-0.55	-3.77	$1.45x10^{-2}$
MRPL1	38	0.60	3.77	$1.45x10^{-2}$
AMPH	63	-0.74	-3.77	$1.47x10^{-2}$
MGAM	14190	-0.64	-3.77	$1.47x10^{-2}$
FAM219A	584	-0.30	-3.76	$1.51x10^{-2}$
IFIT1	1092	0.76	3.76	$1.51x10^{-2}$
NIF3L1	166	0.35	3.76	$1.51x10^{-2}$
SLC2A9	176	-0.40	-3.75	$1.55x10^{-2}$
HIP1	2520	-0.44	-3.74	$1.60x10^{-2}$
SLC2A5	220	0.75	3.74	$1.60x10^{-2}$
CENPE	33	0.74	3.73	$1.61x10^{-2}$
KDM6B	12490	-0.52	-3.73	$1.62x10^{-2}$
IL4R	10554	-0.49	-3.73	$1.64x10^{-2}$
ARHGEF11	3609	-0.42	-3.72	$1.64x10^{-2}$
CENPF	76	0.73	3.72	$1.64x10^{-2}$
AC004069.2	81	-0.61	-3.72	$1.66x10^{-2}$
LINC00482	120	-0.61	-3.72	$1.66x10^{-2}$
LRP1	3913	-0.55	-3.72	$1.66x10^{-2}$
AATK	4392	-0.47	-3.70	$1.75x10^{-2}$
ZFP36	10959	-0.41	-3.70	$1.76x10^{-2}$
DUSP1	9605	-0.62	-3.70	$1.77x10^{-2}$
NFIL3	2422	-0.47	-3.69	$1.77x10^{-2}$
RP11-20B24.7	183	-0.49	-3.69	$1.79x10^{-2}$
SH3PXD2B	114	-0.70	-3.69	$1.79x10^{-2}$
RP11-76E17.4	36	-0.74	-3.69	$1.79x10^{-2}$
DKFZP667F0711	113	-0.63	-3.68	$1.81x10^{-2}$
TNS3	591	-0.47	-3.68	$1.84x10^{-2}$
YEATS4	84	0.50	3.67	$1.87x10^{-2}$
PC	87	-0.51	-3.66	$1.90x10^{-2}$
MTMR3	7364	-0.38	-3.66	$1.91x10^{-2}$
MERTK	138	-0.59	-3.66	$1.91x10^{-2}$
AC091878.1	107	-0.56	-3.65	$1.95x10^{-2}$

RP11-1137G4.3	34	0.72	3.65	$1.95x10^{-2}$
SVIP	235	0.59	3.65	$1.95x10^{-2}$
KIF11	107	0.55	3.63	$2.09x10^{-2}$
CHST15	9082	-0.46	-3.63	$2.11x10^{-2}$
UGT2B11	22	0.75	3.62	$2.11x10^{-2}$
NFKBIA	4244	-0.48	-3.62	$2.12x10^{-2}$
DAB2	322	-0.47	-3.61	$2.15x10^{-2}$
HRH2	2346	-0.53	-3.61	$2.15x10^{-2}$
PPP1R3B	4579	-0.50	-3.62	$2.15x10^{-2}$
RAB44	280	0.52	3.61	$2.15x10^{-2}$
TECPR2	4937	-0.40	-3.61	$2.15x10^{-2}$
ALOX15B	51	-0.75	-3.61	$2.15x10^{-2}$
CXCR2	47021	-0.57	-3.61	$2.15x10^{-2}$
GPER	153	-0.74	-3.61	$2.18x10^{-2}$
AZU1	2651	0.72	3.60	$2.20x10^{-2}$
PEX3	68	0.51	3.60	$2.22x10^{-2}$
C2orf82	31	-0.60	-3.59	$2.22x10^{-2}$
DCAF13	170	0.39	3.59	$2.22x10^{-2}$
EMILIN2	3082	-0.44	-3.59	$2.22x10^{-2}$
TBL1X	3560	-0.47	-3.59	$2.22x10^{-2}$
TMEM150B	307	-0.51	-3.59	$2.22x10^{-2}$
LSM5	77	0.48	3.58	$2.26x10^{-2}$
TARM1	63	0.74	3.58	$2.26x10^{-2}$
IL6R	14094	-0.41	-3.57	$2.32x10^{-2}$
MARVELD1	480	-0.61	-3.57	$2.32x10^{-2}$
PSMA4	642	0.51	3.57	$2.32x10^{-2}$
RHAG	22	0.73	3.57	$2.32x10^{-2}$
S1PR3	391	-0.54	-3.57	$2.32x10^{-2}$
WDFY4	2382	-0.35	-3.57	$2.32x10^{-2}$
WDR61	267	0.36	3.57	$2.32x10^{-2}$
LAMP3	46	0.72	3.56	$2.37x10^{-2}$
PROK2	9621	-0.54	-3.56	$2.37x10^{-2}$
TIRAP	401	-0.28	-3.56	$2.37x10^{-2}$
TPRKB	74	0.43	3.56	$2.37x10^{-2}$
CDC45	36	0.73	3.55	$2.38x10^{-2}$
CDH26	65	-0.51	-3.55	$2.38x10^{-2}$
LINC00963	1937	-0.49	-3.55	$2.38x10^{-2}$
SUCNR1	20	0.68	3.55	$2.38x10^{-2}$
RP11-4F5.2	888	-0.40	-3.54	$2.44x10^{-2}$
RP11-373D23.3	149	-0.56	-3.54	$2.45x10^{-2}$
ISG15	1380	0.73	3.54	$2.46x10^{-2}$
GAS7	7416	-0.52	-3.53	$2.51x10^{-2}$
ATP6V1C2	30	0.60	3.53	$2.51x10^{-2}$
RASSF2	23485	-0.39	-3.53	$2.54x10^{-2}$
TCN1	688	0.73	3.53	$2.54x10^{-2}$
ZCCHC24	200	-0.38	-3.52	$2.60x10^{-2}$

DOCK2	7416	-0.31	-3.52	$2.61x10^{-2}$
RP11-326C3.11	170	-0.42	-3.51	$2.69x10^{-2}$
HES4	51	0.72	3.50	$2.69x10^{-2}$
KIF14	19	0.65	3.50	$2.72x10^{-2}$
CPAMD8	209	-0.46	-3.50	$2.73x10^{-2}$
PLXND1	1991	-0.42	-3.50	$2.73x10^{-2}$
VWA7	66	-0.53	-3.50	$2.73x10^{-2}$
WIPF2	2153	-0.36	-3.49	$2.76x10^{-2}$
CERS2	3222	-0.33	-3.49	$2.77x10^{-2}$
HAUS1	77	0.54	3.49	$2.77x10^{-2}$
TOP2A	111	0.71	3.49	$2.77x10^{-2}$
MMP8	1533	0.71	3.49	$2.77x10^{-2}$
ARHGEF40	5694	-0.54	-3.48	$2.78x10^{-2}$
ITGAD	57	-0.62	-3.48	$2.81x10^{-2}$
ADAMTS2	552	-0.70	-3.47	$2.90x10^{-2}$
CENPH	38	0.62	3.47	$2.90x10^{-2}$
PHF21A	5812	-0.37	-3.47	$2.90x10^{-2}$
RP11-76E17.3	66	-0.67	-3.47	$2.90x10^{-2}$
SUB1	677	0.52	3.46	$2.91x10^{-2}$
HSPE1	98	0.52	3.46	$2.94x10^{-2}$
PNPLA1	155	-0.50	-3.46	$2.94x10^{-2}$
RP3-525N10.2	33	-0.68	-3.46	$2.94x10^{-2}$
CD3D	476	0.53	3.43	$3.14x10^{-2}$
CEBPB	10987	-0.53	-3.44	$3.14x10^{-2}$
FAM26F	213	0.57	3.43	$3.14x10^{-2}$
FLT3	355	-0.69	-3.43	$3.14x10^{-2}$
IQSEC1	8737	-0.36	-3.43	$3.14x10^{-2}$
RNASEH2B	317	0.39	3.43	$3.14x10^{-2}$
NIPSNAP3A	199	0.39	3.43	$3.14x10^{-2}$
SLA	9171	-0.50	-3.43	$3.14x10^{-2}$
HERC5	794	0.69	3.43	$3.14x10^{-2}$
GCSAML	52	0.57	3.42	$3.16x10^{-2}$
PRMT5	750	-0.44	-3.43	$3.16x10^{-2}$
FBN2	197	-0.53	-3.42	$3.16x10^{-2}$
OAS3	2041	0.68	3.42	$3.16x10^{-2}$
PTX3	72	0.64	3.42	$3.18x10^{-2}$
BAIAP2	369	-0.47	-3.42	$3.19x10^{-2}$
ZWINT	55	0.67	3.42	$3.19x10^{-2}$
BUB3	917	0.28	3.41	$3.21x10^{-2}$
CYFIP1	1090	-0.37	-3.41	$3.21x10^{-2}$
IL17RA	19125	-0.41	-3.41	$3.21x10^{-2}$
IL18RAP	2094	-0.56	-3.41	$3.21x10^{-2}$
NLRP3	1235	-0.48	-3.41	$3.21x10^{-2}$
SRXN1	196	-0.49	-3.41	$3.21x10^{-2}$
STIL	32	0.64	3.41	$3.21x10^{-2}$
MPO	2413	0.69	3.41	$3.22x10^{-2}$

VRK1	122	0.59	3.40	$3.23x10^{-2}$
MRPL39	56	0.48	3.39	$3.35x10^{-2}$
ABAT	965	-0.36	-3.39	$3.35x10^{-2}$
CUX1	2758	-0.41	-3.39	$3.35x10^{-2}$
MBP	9465	-0.42	-3.39	$3.38x10^{-2}$
NDC80	55	0.57	3.39	$3.38x10^{-2}$
RTDR1	47	-0.65	-3.39	$3.38x10^{-2}$
ZSCAN18	211	-0.34	-3.39	$3.38x10^{-2}$
SVIL	3671	-0.45	-3.38	$3.39x10^{-2}$
ZNF503	42	-0.46	-3.38	$3.39x10^{-2}$
ENY2	316	0.45	3.38	$3.42x10^{-2}$
SLC8A1	955	-0.51	-3.38	$3.42x10^{-2}$
UGGT1	1942	-0.29	-3.38	$3.42x10^{-2}$
PRIM2	106	0.36	3.38	$3.43x10^{-2}$
IL5RA	172	0.59	3.37	$3.51x10^{-2}$
RP11-701P16.5	90	-0.57	-3.37	$3.51x10^{-2}$
CASC3	5745	-0.40	-3.36	$3.51x10^{-2}$
FOXM1	107	0.60	3.36	$3.53x10^{-2}$
IFIT3	2993	0.67	3.36	$3.53x10^{-2}$
RP11-802E16.3	317	-0.37	-3.36	$3.53x10^{-2}$
C5AR1	19382	-0.45	-3.36	$3.54x10^{-2}$
BICD2	3275	-0.36	-3.35	$3.55x10^{-2}$
C1orf115	79	-0.45	-3.35	$3.55x10^{-2}$
CEACAM6	578	0.69	3.35	$3.55x10^{-2}$
CEACAM8	1080	0.68	3.36	$3.55x10^{-2}$
ECHDC3	220	-0.65	-3.35	$3.55x10^{-2}$
RP11-298I3.1	96	-0.56	-3.35	$3.55x10^{-2}$
RPA3	93	0.49	3.35	$3.60x10^{-2}$
FKBP3	181	0.38	3.34	$3.60x10^{-2}$
MLF1IP	37	0.59	3.34	$3.60x10^{-2}$
DEFA3	19589	0.64	3.34	$3.60x10^{-2}$
DEFA4	2976	0.64	3.34	$3.60x10^{-2}$
DYRK4	91	0.51	3.34	$3.62x10^{-2}$
SORL1	39212	-0.46	-3.34	$3.64x10^{-2}$
ALDH9A1	1170	-0.26	-3.33	$3.66x10^{-2}$
BAIAP2-AS1	599	-0.42	-3.33	$3.66x10^{-2}$
EPHB2	91	-0.50	-3.33	$3.66x10^{-2}$
BPI	3539	0.67	3.33	$3.72x10^{-2}$
COL8A2	89	-0.47	-3.33	$3.73x10^{-2}$
RNASE1	114	-0.67	-3.33	$3.73x10^{-2}$
TADA2B	1680	-0.31	-3.32	$3.74x10^{-2}$
ARV1	73	0.36	3.32	$3.75x10^{-2}$
NCOA6	1513	-0.30	-3.32	$3.75x10^{-2}$
PSAT1	66	0.49	3.32	$3.75x10^{-2}$
RNF175	461	-0.49	-3.32	$3.75x10^{-2}$
CD1D	738	-0.39	-3.32	$3.77x10^{-2}$

CDCA7L	287	0.49	3.32	$3.77x10^{-2}$
RP11-67C2.2	330	-0.45	-3.32	$3.78x10^{-2}$
C10orf105	295	-0.60	-3.31	$3.79x10^{-2}$
ARHGAP26	7235	-0.43	-3.31	$3.84x10^{-2}$
AQP9	27659	-0.41	-3.31	$3.86x10^{-2}$
MDH1	490	0.44	3.30	$3.89x10^{-2}$
SS18L2	107	0.44	3.30	$3.89x10^{-2}$
NOTCH2	9263	-0.42	-3.30	$3.90x10^{-2}$
CTC-246B18.8	69	-0.49	-3.30	$3.94x10^{-2}$
RFC3	69	0.48	3.30	$3.94x10^{-2}$
RP11-181G12.2	171	-0.36	-3.30	$3.94x10^{-2}$
CDK1	28	0.68	3.29	$3.96x10^{-2}$
TUFT1	132	-0.48	-3.29	$3.96x10^{-2}$
C5orf56	725	0.41	3.29	$4.00x10^{-2}$
ESPL1	61	0.63	3.29	$4.00x10^{-2}$
IL10RB	3218	-0.30	-3.29	$4.00x10^{-2}$
SLC11A1	22240	-0.50	-3.29	$4.02x10^{-2}$
FBXO5	75	0.51	3.28	$4.04x10^{-2}$
IFT57	120	0.48	3.28	$4.04x10^{-2}$
KAT6A	3646	-0.43	-3.28	$4.04x10^{-2}$
KDM5B	1009	-0.34	-3.28	$4.04x10^{-2}$
KIAA0391	418	0.24	3.28	$4.04x10^{-2}$
NOL3	87	-0.41	-3.28	$4.04x10^{-2}$
SLC5A9	68	-0.63	-3.28	$4.04x10^{-2}$
TCF19	193	0.49	3.28	$4.04x10^{-2}$
DTL	44	0.67	3.27	$4.04x10^{-2}$
MIS18A	38	0.50	3.27	$4.04x10^{-2}$
ZNF608	207	-0.64	-3.27	$4.04x10^{-2}$
DHFR	212	0.42	3.27	$4.04x10^{-2}$
IRF2BPL	2936	-0.49	-3.27	$4.05x10^{-2}$
STMN1	432	0.62	3.27	$4.07x10^{-2}$
CYSLTR1	355	0.42	3.27	$4.10x10^{-2}$
IL1B	1836	-0.45	-3.26	$4.10x10^{-2}$
UQCRQ	182	0.50	3.26	$4.12x10^{-2}$
CTNNA1	2898	-0.39	-3.25	$4.16x10^{-2}$
DBN1	642	-0.51	-3.26	$4.16x10^{-2}$
FAM49A	2729	-0.34	-3.26	$4.16x10^{-2}$
RETN	980	0.67	3.26	$4.16x10^{-2}$
UBE4B	1704	-0.33	-3.26	$4.16x10^{-2}$
ZNF367	77	0.52	3.26	$4.16x10^{-2}$
SLC29A3	186	-0.34	-3.25	$4.17x10^{-2}$
CFLAR	14455	-0.35	-3.25	$4.18x10^{-2}$
PTGFRN	47	-0.54	-3.25	$4.19x10^{-2}$
ACADM	201	0.43	3.25	$4.20x10^{-2}$
GBGT1	616	-0.43	-3.25	$4.20x10^{-2}$
MATN2	27	0.54	3.25	$4.20x10^{-2}$

MOCS2	93	0.35	3.25	$4.20x10^{-2}$
SGOL2	40	0.60	3.24	$4.20x10^{-2}$
FOSL2	7657	-0.36	-3.24	$4.24x10^{-2}$
FANCL	75	0.52	3.24	$4.25x10^{-2}$
GPR97	10028	-0.46	-3.24	$4.28x10^{-2}$
PKNOX1	479	-0.26	-3.24	$4.28x10^{-2}$
WASF2	8791	-0.36	-3.24	$4.28x10^{-2}$
IGF1R	2866	-0.43	-3.23	$4.28x10^{-2}$
LTF	14590	0.63	3.23	$4.28x10^{-2}$
NCAPG2	149	0.48	3.23	$4.28x10^{-2}$
ITGAM	12705	-0.44	-3.23	$4.29x10^{-2}$
TIMM10	178	0.62	3.23	$4.29x10^{-2}$
ADH5	436	0.36	3.23	$4.30x10^{-2}$
RPSAP58	124	0.46	3.23	$4.30x10^{-2}$
AC007278.3	212	-0.64	-3.22	$4.35x10^{-2}$
ATP5C1	618	0.39	3.22	$4.35x10^{-2}$
CRISPLD2	4330	-0.50	-3.22	$4.35x10^{-2}$
ITSN1	199	-0.36	-3.22	$4.35x10^{-2}$
RP13-580F15.2	68	0.55	3.22	$4.35x10^{-2}$
COPS4	217	0.40	3.22	$4.36x10^{-2}$
CCNA2	68	0.61	3.21	$4.40x10^{-2}$
ZNF319	1121	-0.36	-3.21	$4.43x10^{-2}$
CIDEB	67	-0.44	-3.21	$4.43x10^{-2}$
MYLIP	1719	-0.36	-3.21	$4.43x10^{-2}$
ATXN1	2498	-0.31	-3.21	$4.44x10^{-2}$
DOK3	14016	-0.47	-3.21	$4.44x10^{-2}$
SHROOM1	141	-0.52	-3.21	$4.44x10^{-2}$
TBC1D30	82	-0.47	-3.20	$4.45x10^{-2}$
TLE4	2706	-0.32	-3.20	$4.45x10^{-2}$
ASPM	46	0.66	3.20	$4.46x10^{-2}$
ZNF823	21	0.59	3.20	$4.46x10^{-2}$
EEF1E1	26	0.54	3.20	$4.47x10^{-2}$
RNF144A	391	0.46	3.20	$4.47x10^{-2}$
SPC24	50	0.66	3.20	$4.47x10^{-2}$
NCOR2	2947	-0.34	-3.20	$4.48x10^{-2}$
SLC8A1-AS1	26	-0.54	-3.20	$4.48x10^{-2}$
SNX27	3312	-0.33	-3.20	$4.49x10^{-2}$
MLLT1	1133	-0.33	-3.19	$4.50x10^{-2}$
MEGF6	891	-0.39	-3.19	$4.50x10^{-2}$
ZNF746	3152	-0.43	-3.19	$4.52x10^{-2}$
BEST1	1992	-0.38	-3.19	$4.52x10^{-2}$
CBL	5966	-0.36	-3.19	$4.52x10^{-2}$
GNS	6114	-0.37	-3.19	$4.52x10^{-2}$
GPX7	90	0.42	3.19	$4.52x10^{-2}$
MFN2	4153	-0.37	-3.19	$4.52x10^{-2}$
RAD51	56	0.59	3.19	$4.52x10^{-2}$

RP11-344B5.2	224	-0.46	-3.19	$4.52x10^{-2}$
PYGL	15676	-0.47	-3.18	$4.54x10^{-2}$
CHEK1	43	0.56	3.18	$4.55x10^{-2}$
ZCCHC9	191	0.29	3.18	$4.55x10^{-2}$
CPSF7	3652	-0.37	-3.18	$4.56x10^{-2}$
ULK1	4884	-0.39	-3.18	$4.57x10^{-2}$
AC096772.6	114	-0.34	-3.18	$4.60x10^{-2}$
IFIT5	664	0.55	3.18	$4.60x10^{-2}$
RPS27L	136	0.43	3.18	$4.60x10^{-2}$
TYMS	163	0.64	3.18	$4.60x10^{-2}$
LINC00211	173	-0.39	-3.17	$4.62x10^{-2}$
NDEL1	3893	-0.37	-3.17	$4.62x10^{-2}$
CLSPN	44	0.54	3.17	$4.64x10^{-2}$
EZH2	173	0.46	3.17	$4.64x10^{-2}$
FAM71F2	105	-0.48	-3.17	$4.64x10^{-2}$
LPCAT3	1538	-0.37	-3.17	$4.64x10^{-2}$
TG	105	-0.52	-3.17	$4.64x10^{-2}$
EXOSC9	265	0.35	3.16	$4.69x10^{-2}$
PSMD10	202	0.36	3.16	$4.70x10^{-2}$
TOMM5	44	0.41	3.16	$4.72x10^{-2}$
LINC00593	37	-0.54	-3.16	$4.74x10^{-2}$
CXCR1	32208	-0.47	-3.16	$4.78x10^{-2}$
ITPRIP	3718	-0.38	-3.15	$4.80x10^{-2}$
NDUFA4	329	0.53	3.15	$4.80x10^{-2}$
PWP1	314	0.26	3.15	$4.80x10^{-2}$
ZNF480	105	0.38	3.15	$4.80x10^{-2}$
HMGN1	1203	0.35	3.15	$4.80x10^{-2}$
MSN	34989	-0.39	-3.15	$4.80x10^{-2}$
NSMCE4A	260	0.31	3.15	$4.80x10^{-2}$
SEMA6A	35	0.48	3.15	$4.80x10^{-2}$
SSB	304	0.39	3.15	$4.80x10^{-2}$
TRIAP1	87	0.38	3.15	$4.83x10^{-2}$
PADI2	8617	-0.45	-3.15	$4.84x10^{-2}$
SLC2A3	16672	-0.43	-3.14	$4.88x10^{-2}$
HAUS2	224	0.28	3.14	$4.89x10^{-2}$
PSMG1	77	0.44	3.14	$4.90x10^{-2}$
DZIP1L	33	0.63	3.14	$4.95x10^{-2}$
ARNTL2	30	0.51	3.13	$4.96x10^{-2}$
TIPIN	35	0.47	3.13	$4.97x10^{-2}$
LDLRAD3	118	-0.48	-3.13	$4.97x10^{-2}$
NRG1	159	-0.43	-3.13	$4.97x10^{-2}$
ASPH	963	-0.51	-3.13	$4.98x10^{-2}$

Log2FC indicates Log2 fold change; Stat indicates the results from Wald statistic; and B-H indicates P-value after Benjamini-Hochberg correction for false discovery rate.

Table II. The 164 differentially expressed genes between controls and aneurysmal SAH (aSAH) patients before and after RIC treatment.

DESeq2	aSAH untreated vs. treated					untreated controls vs. untreated aSAH					untreated controls vs. treated aSAH				
GENE	**Mean**	**Log2FC**	**SE**	**Stat**	**P-value**	**Mean**	**Log2FC**	**SE**	**Stat**	**P-value**	**Mean**	**Log2FC**	**SE**	**Stat**	**P-value**
HTRA1	98	-1.28	0.19	-6.73	$2.47x10^{-7}$	86	2.25	0.27	8.26	$5.00x10^{-13}$*	39	0.63	0.27	2.32	$8.09x10^{-2}$
MYO7A	177	-1.00	0.17	-5.80	$1.61x10^{-5}$	173	1.59	0.20	7.88	$9.69x10^{-12}$*	109	0.34	0.18	1.89	$1.69x10^{-1}$
OLR1	136	1.19	0.20	5.82	$1.61x10^{-5}$	40	0.83	0.34	2.49	$3.59x10^{-2}$	118	2.44	0.31	7.89	$4.65x10^{-12}$*
CRISP3	394	1.13	0.20	5.66	$2.93x10^{-5}$	153	0.46	0.33	1.39	$2.72x10^{-1}$	369	1.85	0.32	5.80	$7.91x10^{-7}$*
SERPINB10	108	1.01	0.20	5.13	$3.24x10^{-4}$	38	0.89	0.30	3.00	$1.01x10^{-2}$	94	2.23	0.30	7.34	$1.26x10^{-10}$*
STAT5B	13869	-0.43	0.08	-5.04	$4.74x10^{-4}$	14359	0.76	0.15	5.27	$4.52x10^{-6}$*	12921	0.34	0.14	2.33	$8.02x10^{-2}$
BCL2L15	74	0.93	0.19	4.95	$6.44x10^{-4}$	45	0.03	0.20	0.16	$9.16x10^{-1}$	80	1.41	0.24	5.85	$6.46x10^{-7}$*
SSH1	2477	-0.49	0.10	-4.92	$7.15x10^{-4}$	2519	0.86	0.15	5.86	$3.48x10^{-7}$*	2221	0.38	0.15	2.46	$6.14x10^{-2}$
DNAH17	1640	-0.56	0.11	-4.85	$8.96x10^{-4}$	1412	1.45	0.22	6.56	$1.27x10^{-8}$*	1121	0.84	0.21	4.02	$1.21x10^{-3}$
ANLN	39	0.91	0.19	4.85	$8.96x10^{-4}$	21	0.64	0.23	2.73	$2.00x10^{-2}$	37	1.79	0.22	8.22	$5.85x10^{-13}$*
RP11-473M20.5	362	-0.79	0.16	-4.81	$9.70x10^{-4}$	394	0.99	0.20	4.95	$1.56x10^{-5}$*	298	0.07	0.20	0.35	$8.43x10^{-1}$
HAUS4	1226	-0.57	0.12	-4.80	$9.70x10^{-4}$	1319	0.80	0.18	4.49	$9.42x10^{-5}$*	1119	0.19	0.17	1.12	$4.59x10^{-1}$
MANSC1	2500	-0.68	0.14	-4.78	$1.03x10^{-3}$	2417	1.14	0.18	6.23	$7.17x10^{-8}$*	1959	0.47	0.20	2.32	$8.12x10^{-2}$
MS4A3	355	0.97	0.20	4.76	$1.12x10^{-3}$	166	0.13	0.28	0.46	$7.47x10^{-1}$	362	1.70	0.28	6.02	$2.63x10^{-7}$*
RP5-968J1.1	49	-0.73	0.15	-4.75	$1.15x10^{-3}$	49	1.15	0.21	5.51	$1.72x10^{-6}$*	37	0.33	0.21	1.61	$2.53x10^{-1}$
ETS2	2895	-0.65	0.14	-4.72	$1.20x10^{-3}$	2937	1.07	0.18	5.86	$3.48x10^{-7}$*	2334	0.32	0.17	1.91	$1.65x10^{-1}$
RP11-1D12.2	26	0.93	0.20	4.57	$2.19x10^{-3}$	7	0.54	0.35	1.54	$2.19x10^{-1}$	23	2.24	0.32	6.93	$1.45x10^{-9}$*
DGKD	4162	-0.38	0.08	-4.52	$2.40x10^{-3}$	4483	0.62	0.10	6.34	$4.20x10^{-8}$*	4143	0.25	0.11	2.34	$7.78x10^{-2}$
TRIM7	122	-0.64	0.14	-4.45	$2.94x10^{-3}$	126	0.92	0.15	6.32	$4.48x10^{-8}$*	105	0.28	0.17	1.62	$2.52x10^{-1}$
ST14	819	-0.67	0.15	-4.44	$2.96x10^{-3}$	886	0.88	0.17	5.06	$1.03x10^{-5}$*	719	0.14	0.18	0.78	$6.26x10^{-1}$
PIWIL4	117	0.63	0.14	4.45	$2.96x10^{-3}$	86	0.42	0.13	3.31	$4.26x10^{-3}$	118	1.21	0.16	7.36	$1.06x10^{-10}$*
CASC5	57	0.78	0.18	4.38	$3.66x10^{-3}$	40	0.29	0.20	1.45	$2.47x10^{-1}$	59	1.27	0.19	6.87	$2.14x10^{-9}$*
CYP1B1-AS1	76	-0.73	0.17	-4.36	$3.93x10^{-3}$	75	1.23	0.25	5.01	$1.27x10^{-5}$*	54	0.34	0.23	1.47	$3.05x10^{-1}$
AL137145.1	205	-0.70	0.16	-4.35	$3.98x10^{-3}$	208	1.12	0.19	5.82	$4.24x10^{-7}$*	160	0.30	0.19	1.57	$2.68x10^{-1}$
FUT4	360	0.45	0.10	4.35	$3.98x10^{-3}$	328	0.25	0.07	3.57	$1.98x10^{-3}$	397	0.75	0.09	8.22	$5.85x10^{-13}$*
MKNK1	4681	-0.44	0.10	-4.33	$4.17x10^{-3}$	4460	1.01	0.19	5.20	$5.95x10^{-6}$*	3895	0.54	0.19	2.91	$2.42x10^{-2}$
INHBA	20	0.88	0.20	4.32	$4.30x10^{-3}$	8	0.75	0.35	2.14	$7.43x10^{-2}$	17	1.84	0.32	5.70	$1.28x10^{-6}$*
AC061992.1	192	-0.63	0.15	-4.31	$4.30x10^{-3}$	158	1.75	0.24	7.30	$2.81x10^{-10}$*	114	1.01	0.22	4.55	$1.97x10^{-4}$
GBP4	1456	0.80	0.19	4.30	$4.30x10^{-3}$	1927	-1.07	0.20	-5.36	$3.14x10^{-6}$*	2517	0.07	0.22	0.34	$8.51x10^{-1}$
STAB1	3278	-0.61	0.14	-4.30	$4.32x10^{-3}$	3351	1.04	0.14	7.20	$5.63x10^{-10}$*	2712	0.32	0.13	2.50	$5.70x10^{-2}$
BMF	1012	-0.33	0.08	-4.27	$4.65x10^{-3}$	1148	0.46	0.08	5.83	$4.03x10^{-7}$*	1081	0.13	0.09	1.52	$2.88x10^{-1}$
CLMN	1191	-0.44	0.10	-4.23	$4.98x10^{-3}$	1292	0.68	0.13	5.41	$2.60x10^{-6}$*	1155	0.21	0.12	1.75	$2.09x10^{-1}$
IFI44	544	0.84	0.20	4.24	$4.98x10^{-3}$	698	-1.45	0.32	-4.49	$9.23x10^{-5}$*	804	-0.73	0.30	-2.42	$6.70x10^{-2}$
RFX2	2047	-0.61	0.14	-4.23	$4.99x10^{-3}$	1849	1.30	0.22	6.00	$1.91x10^{-7}$*	1489	0.67	0.22	3.04	$1.80x10^{-2}$

RP11-473M20.7	5443	-0.55	0.13	-4.21	$5.27x10^{-3}$	5398	0.97	0.18	5.34	$3.35x10^{-6}$*	4601	0.41	0.19	2.16	$1.08x10^{-1}$
NLRP12	4658	-0.46	0.11	-4.17	$5.57x10^{-3}$	4797	0.78	0.16	4.86	$2.29x10^{-5}$*	4292	0.34	0.17	1.98	$1.47x10^{-1}$
COL17A1	101	0.85	0.21	4.13	$6.17x10^{-3}$	30	0.91	0.26	3.48	$2.63x10^{-3}$	89	2.46	0.30	8.14	$8.38x10^{-13}$*
RNASE4	146	-0.55	0.13	-4.11	$6.72x10^{-3}$	146	1.00	0.19	5.33	$3.57x10^{-6}$*	121	0.39	0.18	2.23	$9.61x10^{-2}$
GBP5	3964	0.79	0.19	4.09	$7.18x10^{-3}$	4980	-1.09	0.23	-4.77	$3.13x10^{-5}$*	6743	0.16	0.26	0.63	$7.02x10^{-1}$
ATP8B4	407	0.57	0.14	4.08	$7.27x10^{-3}$	360	0.17	0.14	1.19	$3.57x10^{-1}$	455	0.81	0.14	5.84	$6.47x10^{-7}$*
ABCA13	197	0.83	0.21	4.05	$8.00x10^{-3}$	78	0.81	0.34	2.38	$4.53x10^{-2}$	171	1.96	0.32	6.16	$1.22x10^{-7}$*
SLC7A7	3945	-0.52	0.13	-4.03	$8.46x10^{-3}$	4283	0.77	0.13	5.87	$3.43x10^{-7}$*	3684	0.18	0.12	1.47	$3.05x10^{-1}$
RP11-1334A24.6	556	-0.52	0.13	-3.99	$9.39x10^{-3}$	577	0.81	0.18	4.55	$7.27x10^{-5}$*	505	0.30	0.19	1.61	$2.56x10^{-1}$
KIF15	24	0.76	0.19	3.99	$9.39x10^{-3}$	17	0.13	0.24	0.57	$6.81x10^{-1}$	26	1.19	0.24	4.98	$3.62x10^{-5}$*
VSIG4	299	-0.79	0.20	-3.97	$9.66x10^{-3}$	240	1.94	0.29	6.78	$4.08x10^{-9}$*	109	0.20	0.23	0.87	$5.80x10^{-1}$
RNASE3	701	0.81	0.20	3.97	$9.66x10^{-3}$	88	0.64	0.33	1.97	$1.04x10^{-1}$	163	1.60	0.32	5.03	$3.01x10^{-5}$*
KIAA0101	37	0.81	0.21	3.94	$1.06x10^{-3}$	15	0.52	0.25	2.04	$9.14x10^{-2}$	35	1.97	0.28	7.06	$6.74x10^{-10}$*
SIRPA	13345	-0.47	0.12	-3.93	$1.09x10^{-3}$	13087	0.94	0.18	5.30	$3.99x10^{-6}$*	11456	0.46	0.18	2.56	$5.05x10^{-2}$
ERG	63	0.80	0.21	3.89	$1.19x10^{-2}$	24	0.83	0.27	3.09	$7.94x10^{-3}$	55	2.20	0.28	7.88	$4.65x10^{-12}$*
PFKFB3	6018	-0.58	0.15	-3.87	$1.20x10^{-2}$	4986	1.55	0.26	5.87	$3.33x10^{-7}$*	3844	0.91	0.26	3.55	$4.91x10^{-3}$
PRC1	90	0.62	0.16	3.88	$1.20x10^{-2}$	80	-0.01	0.12	-0.11	$9.41x10^{-1}$	105	0.79	0.16	4.99	$3.36x10^{-5}$*
FBXO4	60	0.56	0.14	3.86	$1.25x10^{-2}$	81	-0.65	0.14	-4.52	$8.22x10^{-5}$*	96	-0.04	0.13	-0.31	$8.65x10^{-1}$
DLGAP5	28	0.79	0.20	3.86	$1.26x10^{-2}$	15	0.56	0.27	2.08	$8.39x10^{-2}$	27	1.79	0.23	7.83	$6.29x10^{-12}$*
HDAC4	2592	-0.37	0.10	-3.84	$1.32x10^{-2}$	2644	0.81	0.16	5.02	$1.21x10^{-5}$*	2366	0.38	0.14	2.73	$3.59x10^{-2}$
NCAPG	41	0.75	0.20	3.84	$1.32x10^{-2}$	26	0.41	0.23	1.81	$1.41x10^{-1}$	41	1.43	0.21	6.75	$4.43x10^{-9}$*
HPN	19	-0.74	0.19	-3.83	$1.33x10^{-2}$	17	1.68	0.25	6.64	$8.73x10^{-9}$*	11	0.71	0.26	2.74	$3.46x10^{-2}$
IL1R1	1115	-0.52	0.14	-3.83	$1.33x10^{-2}$	1038	1.13	0.21	5.27	$4.47x10^{-6}$*	869	0.58	0.21	2.73	$3.59x10^{-2}$
LAMC1	134	-0.53	0.14	-3.83	$1.33x10^{-2}$	144	0.76	0.15	5.18	$6.56x10^{-6}$*	126	0.21	0.16	1.28	$3.85x10^{-1}$
CEP55	29	0.78	0.20	3.82	$1.33x10^{-2}$	16	0.41	0.25	1.63	$1.88x10^{-1}$	29	1.59	0.24	6.72	$5.24x10^{-9}$*
PPARG	25	-0.73	0.19	-3.82	$1.35x10^{-2}$	19	2.13	0.27	7.85	$9.69x10^{-12}$*	12	1.20	0.28	4.36	$3.92x10^{-4}$
CTSG	1261	0.74	0.19	3.81	$1.35x10^{-2}$	74	1.02	0.34	3.01	$9.92x10^{-3}$	162	2.03	0.33	6.21	$9.61x10^{-8}$*
CIT	46	0.75	0.20	3.81	$1.35x10^{-2}$	35	-0.01	0.20	-0.05	$9.76x10^{-1}$	52	1.07	0.19	5.74	$1.03x10^{-6}$*
TLR5	1668	-0.50	0.13	-3.79	$1.43x10^{-2}$	1515	1.16	0.23	5.16	$7.00x10^{-6}$*	1281	0.65	0.22	2.90	$2.42x10^{-2}$
CNTNAP3	1973	-0.66	0.17	-3.79	$1.44x10^{-2}$	1586	1.48	0.30	5.00	$1.30x10^{-5}$*	1265	0.90	0.30	2.98	$2.04x10^{-2}$
MMP14	124	-0.55	0.14	-3.77	$1.45x10^{-2}$	115	1.25	0.18	6.91	$2.16x10^{-9}$*	93	0.62	0.17	3.74	$2.82x10^{-3}$
AMPH	63	-0.74	0.20	-3.77	$1.47x10^{-2}$	53	1.99	0.27	7.46	$1.16x10^{-10}$*	31	0.91	0.25	3.68	$3.35x10^{-3}$
FAM219A	584	-0.30	0.08	-3.76	$1.51x10^{-2}$	647	0.49	0.09	5.35	$3.24x10^{-6}$*	612	0.19	0.09	2.18	$1.04x10^{-1}$
SLC2A5	220	0.75	0.20	3.74	$1.60x10^{-2}$	97	0.54	0.22	2.42	$4.19x10^{-2}$	208	1.94	0.27	7.18	$3.16x10^{-10}$*
IL4R	10554	-0.49	0.13	-3.73	$1.64x10^{-2}$	9961	1.10	0.19	5.74	$6.00x10^{-7}$*	8434	0.57	0.19	3.03	$1.85x10^{-2}$
CENPF	76	0.73	0.20	3.72	$1.64x10^{-2}$	57	0.06	0.20	0.31	$8.30x10^{-1}$	85	1.10	0.19	5.87	$5.89x10^{-7}$*
LINC00482	120	-0.61	0.16	-3.72	$1.66x10^{-2}$	108	1.35	0.22	6.11	$1.17x10^{-7}$*	85	0.67	0.22	2.99	$2.01x10^{-2}$
AATK	4392	-0.47	0.13	-3.70	$1.75x10^{-2}$	4274	0.94	0.15	6.19	$8.69x10^{-8}$*	3805	0.50	0.17	2.87	$2.62x10^{-2}$
ZFP36	10959	-0.41	0.11	-3.70	$1.76x10^{-2}$	11298	0.77	0.15	5.13	$8.11x10^{-6}$*	10170	0.35	0.15	2.35	$7.61x10^{-2}$

DUSP1	9605	-0.62	0.17	-3.70	$1.77x10^{-2}$	9807	1.02	0.19	5.49	$1.83x10^{-6}$*	7874	0.30	0.19	1.56	$2.71x10^{-1}$
NFIL3	2422	-0.47	0.13	-3.69	$1.77x10^{-2}$	2530	0.77	0.17	4.68	$4.59x10^{-5}$*	2240	0.30	0.17	1.71	$2.21x10^{-1}$
RP11-20B24.7	183	-0.49	0.13	-3.69	$1.79x10^{-2}$	168	1.16	0.18	6.40	$2.91x10^{-8}$*	144	0.65	0.19	3.45	$6.37x10^{-3}$
SH3PXD2B	114	-0.70	0.19	-3.69	$1.79x10^{-2}$	101	1.75	0.29	5.95	$2.39x10^{-7}$*	61	0.70	0.26	2.69	$3.87x10^{-2}$
DKFZP667F0711	113	-0.63	0.17	-3.68	$1.81x10^{-2}$	114	1.09	0.23	4.77	$3.13x10^{-5}$*	88	0.33	0.22	1.48	$3.03x10^{-1}$
MTMR3	7364	-0.38	0.10	-3.66	$1.91x10^{-2}$	7804	0.67	0.14	4.74	$3.61x10^{-5}$*	7138	0.28	0.14	1.97	$1.49x10^{-1}$
MERTK	138	-0.59	0.16	-3.66	$1.91x10^{-2}$	141	1.00	0.21	4.63	$5.47x10^{-5}$*	113	0.30	0.21	1.44	$3.18x10^{-1}$
AC091878.1	107	-0.56	0.15	-3.65	$1.95x10^{-2}$	86	1.58	0.28	5.71	$6.72x10^{-7}$*	67	0.98	0.27	3.61	$4.10x10^{-3}$
KIF11	107	0.55	0.15	3.63	$2.09x10^{-2}$	86	0.35	0.15	2.28	$5.69x10^{-2}$	112	1.01	0.16	6.16	$1.24x10^{-7}$*
NFKBIA	4244	-0.48	0.13	-3.62	$2.12x10^{-2}$	4186	0.97	0.14	7.12	$7.27x10^{-10}$*	3635	0.46	0.14	3.24	$1.11x10^{-2}$
TECPR2	4937	-0.40	0.11	-3.61	$2.15x10^{-2}$	4872	0.85	0.17	5.10	$8.89x10^{-6}$*	4416	0.46	0.17	2.69	$3.91x10^{-2}$
GPER	153	-0.74	0.20	-3.61	$2.18x10^{-2}$	128	1.99	0.29	6.93	$1.96x10^{-9}$*	74	0.86	0.28	3.06	$1.71x10^{-2}$
AZU1	2651	0.72	0.20	3.60	$2.20x10^{-2}$	183	0.96	0.33	2.96	$1.11x10^{-2}$	322	1.72	0.32	5.36	$6.56x10^{-6}$*
PEX3	68	0.51	0.14	3.60	$2.22x10^{-2}$	102	-0.82	0.13	-6.11	$1.18x10^{-7}$*	117	-0.26	0.13	-1.94	$1.55x10^{-1}$
TMEM150B	307	-0.51	0.14	-3.59	$2.22x10^{-2}$	294	1.07	0.15	7.01	$1.26x10^{-9}$*	251	0.53	0.17	3.20	$1.20x10^{-2}$
TARM1	63	0.74	0.21	3.58	$2.26x10^{-2}$	14	1.16	0.30	3.86	$8.05x10^{-4}$	26	2.00	0.30	6.67	$6.89x10^{-9}$*
MARVELD1	480	-0.61	0.17	-3.57	$2.32x10^{-2}$	473	1.11	0.16	6.82	$3.59x10^{-9}$*	382	0.41	0.18	2.30	$8.40x10^{-2}$
S1PR3	391	-0.54	0.15	-3.57	$2.32x10^{-2}$	436	0.66	0.14	4.72	$3.80x10^{-5}$*	382	0.11	0.17	0.67	$6.80x10^{-1}$
RHAG	22	0.73	0.21	3.57	$2.32x10^{-2}$	12	0.73	0.33	2.21	$6.52x10^{-2}$	20	1.70	0.28	5.99	$3.04x10^{-7}$*
LAMP3	46	0.72	0.20	3.56	$2.37x10^{-2}$	68	-1.56	0.28	-5.51	$1.67x10^{-6}$*	79	-0.66	0.28	-2.34	$7.82x10^{-2}$
PROK2	9621	-0.54	0.15	-3.56	$2.37x10^{-2}$	8807	1.20	0.25	4.71	$3.99x10^{-5}$*	7119	0.60	0.25	2.45	$6.30x10^{-2}$
LINC00963	1937	-0.49	0.14	-3.55	$2.38x10^{-2}$	1822	1.13	0.17	6.79	$3.97x10^{-9}$*	1544	0.59	0.16	3.57	$4.58x10^{-3}$
CDC45	36	0.73	0.21	3.55	$2.38x10^{-2}$	13	0.70	0.26	2.65	$2.42x10^{-2}$	33	2.14	0.29	7.46	$6.10x10^{-11}$*
SUCNR1	20	0.68	0.19	3.55	$2.38x10^{-2}$	12	0.53	0.23	2.27	$5.71x10^{-2}$	19	1.50	0.24	6.28	$6.80x10^{-8}$*
RP11-373D23.3	149	-0.56	0.16	-3.54	$2.45x10^{-2}$	137	1.32	0.24	5.43	$2.37x10^{-6}$*	105	0.60	0.22	2.76	$3.32x10^{-2}$
GAS7	7416	-0.52	0.15	-3.53	$2.51x10^{-2}$	6704	1.27	0.18	7.19	$5.63x10^{-10}$*	5545	0.70	0.18	3.96	$1.48x10^{-3}$
RASSF2	23485	-0.39	0.11	-3.53	$2.54x10^{-2}$	25670	0.59	0.13	4.55	$7.41x10^{-5}$*	23572	0.20	0.13	1.50	$2.93x10^{-1}$
TCN1	688	0.73	0.21	3.53	$2.54x10^{-2}$	232	0.55	0.24	2.27	$5.72x10^{-2}$	372	1.49	0.26	5.77	$9.20x10^{-7}$*
ZCCHC24	200	-0.38	0.11	-3.52	$2.60x10^{-2}$	204	0.77	0.13	5.83	$4.13x10^{-7}$*	187	0.40	0.14	2.82	$2.96x10^{-2}$
KIF14	19	0.65	0.19	3.50	$2.72x10^{-2}$	12	0.64	0.27	2.37	$4.61x10^{-2}$	17	1.45	0.24	5.93	$4.12x10^{-7}$*
TOP2A	111	0.71	0.20	3.49	$2.77x10^{-2}$	72	0.38	0.25	1.50	$2.31x10^{-1}$	111	1.42	0.20	7.21	$2.86x10^{-10}$*
MMP8	1533	0.71	0.20	3.49	$2.77x10^{-2}$	425	1.05	0.36	2.93	$1.20x10^{-2}$	1275	2.31	0.34	6.84	$2.47x10^{-9}$*
ARHGEF40	5694	-0.54	0.16	-3.48	$2.78x10^{-2}$	5390	1.09	0.20	5.35	$3.33x10^{-6}$*	4535	0.53	0.21	2.50	$5.74x10^{-2}$
ADAMTS2	552	-0.70	0.20	-3.47	$2.90x10^{-2}$	379	2.49	0.38	6.57	$1.25x10^{-8}$*	95	1.03	0.35	2.96	$2.16x10^{-2}$
PHF21A	5812	-0.37	0.11	-3.47	$2.90x10^{-2}$	5832	0.81	0.17	4.77	$3.21x10^{-5}$*	5288	0.42	0.17	2.53	$5.41x10^{-2}$
PNPLA1	155	-0.50	0.15	-3.46	$2.94x10^{-2}$	160	0.86	0.16	5.31	$3.83x10^{-6}$*	138	0.33	0.17	1.89	$1.70x10^{-1}$
CEBPB	10987	-0.53	0.16	-3.44	$3.14x10^{-2}$	10791	1.04	0.18	5.67	$7.96x10^{-7}$*	9016	0.44	0.19	2.37	$7.38x10^{-2}$
FLT3	355	-0.69	0.20	-3.43	$3.14x10^{-2}$	323	1.51	0.28	5.40	$2.65x10^{-6}$*	216	0.54	0.28	1.94	$1.57x10^{-1}$
HERC5	794	0.69	0.20	3.43	$3.14x10^{-2}$	1023	-1.33	0.29	-4.62	$5.69x10^{-5}$*	1205	-0.51	0.29	-1.79	$1.97x10^{-1}$

PTX3	72	0.64	0.19	3.42	$3.18x10^{-2}$	44	0.41	0.14	2.85	$1.47x10^{-2}$	72	1.47	0.22	6.60	$1.03x10^{-8}$*
BAIAP2	369	-0.47	0.14	-3.42	$3.19x10^{-2}$	385	0.79	0.13	5.96	$2.29x10^{-7}$*	340	0.31	0.15	2.07	$1.26x10^{-1}$
ZWINT	55	0.67	0.20	3.42	$3.19x10^{-2}$	35	0.32	0.17	1.86	$1.28x10^{-1}$	56	1.39	0.21	6.48	$2.11x10^{-8}$*
BUB3	917	0.28	0.08	3.41	$3.21x10^{-2}$	1251	-0.42	0.08	-5.46	$2.07x10^{-6}$*	1372	-0.14	0.07	-1.96	$1.52x10^{-1}$
IL17RA	19125	-0.41	0.12	-3.41	$3.21x10^{-2}$	19771	0.75	0.16	4.80	$2.87x10^{-5}$*	17857	0.34	0.16	2.11	$1.17x10^{-1}$
SRXN1	196	-0.49	0.14	-3.41	$3.21x10^{-2}$	192	0.95	0.20	4.79	$2.95x10^{-5}$*	166	0.46	0.21	2.20	$1.00x10^{-1}$
MPO	2413	0.69	0.20	3.41	$3.22x10^{-2}$	368	1.05	0.27	3.87	$7.73x10^{-4}$	636	1.88	0.29	6.56	$1.32x10^{-8}$*
CUX1	2758	-0.41	0.12	-3.39	$3.35x10^{-2}$	3088	0.59	0.12	4.72	$3.87x10^{-5}$*	2787	0.13	0.12	1.12	$4.57x10^{-1}$
RTDR1	47	-0.65	0.19	-3.39	$3.38x10^{-2}$	42	1.46	0.25	5.87	$3.33x10^{-7}$*	31	0.65	0.25	2.55	$5.08x10^{-2}$
UGGT1	1942	-0.29	0.08	-3.38	$3.42x10^{-2}$	2746	-0.07	0.08	-0.91	$4.93x10^{-1}$	2650	-0.37	0.07	-5.38	5.84x10-6*
FOXM1	107	0.60	0.18	3.36	$3.53x10^{-2}$	78	0.30	0.14	2.07	$8.56x10^{-2}$	112	1.17	0.19	6.16	$1.20x10^{-7}$*
BICD2	3275	-0.36	0.11	-3.35	$3.55x10^{-2}$	3504	0.62	0.11	5.64	$9.34x10^{-7}$*	3257	0.27	0.12	2.14	$1.12x10^{-1}$
CEACAM6	578	0.69	0.21	3.35	$3.55x10^{-2}$	184	0.96	0.34	2.87	$1.43x10^{-2}$	493	2.19	0.33	6.66	$7.37x10^{-9}$*
CEACAM8	1080	0.68	0.20	3.36	$3.55x10^{-2}$	306	0.80	0.35	2.30	$5.42x10^{-2}$	944	2.20	0.34	6.55	$1.36x10^{-8}$*
DEFA4	2976	0.64	0.19	3.34	$3.60x10^{-2}$	249	1.26	0.35	3.57	$1.97x10^{-3}$	499	2.07	0.33	6.23	$8.67x10^{-8}$*
BPI	3539	0.67	0.20	3.33	$3.72x10^{-2}$	625	1.09	0.32	3.45	$2.89x10^{-3}$	1122	1.85	0.32	5.83	$6.73x10^{-7}$*
RNASE1	114	-0.67	0.20	-3.33	$3.73x10^{-2}$	87	2.40	0.34	7.06	$1.03x10^{-9}$*	42	1.38	0.30	4.61	$1.57x10^{-4}$
RP11-67C2.2	330	-0.45	0.14	-3.32	$3.78x10^{-2}$	317	1.00	0.20	5.06	$1.05x10^{-5}$*	274	0.51	0.20	2.62	$4.50x10^{-2}$
CTC-246B18.8	69	-0.49	0.15	-3.30	$3.94x10^{-2}$	74	0.74	0.15	4.82	$2.64x10^{-5}$*	64	0.22	0.16	1.33	$3.62x10^{-1}$
CDK1	28	0.68	0.21	3.29	$3.96x10^{-2}$	14	0.50	0.28	1.77	$1.50x10^{-1}$	27	1.76	0.25	7.04	$7.65x10^{-10}$*
IL10RB	3218	-0.30	0.09	-3.29	$4.00x10^{-2}$	3382	0.63	0.14	4.54	$7.81x10^{-5}$*	3166	0.32	0.13	2.43	$6.61x10^{-2}$
ESPL1	61	0.63	0.19	3.29	$4.00x10^{-2}$	36	0.60	0.18	3.38	$3.53x10^{-3}$	58	1.59	0.22	7.20	$2.91x10^{-10}$*
NOL3	87	-0.41	0.12	-3.28	$4.04x10^{-2}$	89	0.81	0.15	5.30	$4.02x10^{-6}$*	79	0.35	0.14	2.56	$5.05x10^{-2}$
ZNF608	207	-0.64	0.19	-3.27	$4.04x10^{-2}$	182	1.38	0.29	4.81	$2.78x10^{-5}$*	137	0.65	0.29	2.23	$9.49x10^{-2}$
DTL	44	0.67	0.20	3.27	$4.04x10^{-2}$	25	0.59	0.23	2.60	$2.75x10^{-2}$	42	1.69	0.23	7.40	$8.58x10^{-11}$*
DHFR	212	0.42	0.13	3.27	$4.04x10^{-2}$	195	0.26	0.12	2.13	$7.61x10^{-2}$	233	0.73	0.12	6.10	$1.67x10^{-7}$*
IRF2BPL	2936	-0.49	0.15	-3.27	$4.05x10^{-2}$	2859	1.01	0.20	5.12	$8.25x10^{-6}$*	2433	0.47	0.20	2.37	$7.43x10^{-2}$
CYSLTR1	355	0.42	0.13	3.27	$4.10x10^{-2}$	521	-0.71	0.14	-4.95	$1.58x10^{-5}$*	592	-0.22	0.15	-1.46	$3.09x10^{-1}$
GBGT1	616	-0.43	0.13	-3.25	$4.20x10^{-2}$	602	0.91	0.15	6.18	$8.92x10^{-8}$*	541	0.49	0.17	2.97	$2.11x10^{-2}$
MOCS2	93	0.35	0.11	3.25	$4.20x10^{-2}$	131	-0.54	0.10	-5.21	$5.78x10^{-6}$*	146	-0.17	0.10	-1.69	$2.26x10^{-1}$
LTF	14590	0.63	0.20	3.23	$4.28x10^{-2}$	2436	0.71	0.35	2.04	$9.25x10^{-2}$	4541	1.58	0.33	4.77	$8.36x10^{-5}$*
ITGAM	12705	-0.44	0.14	-3.23	$4.29x10^{-2}$	12835	0.86	0.18	4.88	$2.09x10^{-5}$*	11307	0.39	0.18	2.21	$9.97x10^{-2}$
ITSN1	199	-0.36	0.11	-3.22	$4.35x10^{-2}$	210	0.66	0.13	5.04	$1.11x10^{-5}$*	193	0.29	0.13	2.21	$9.82x10^{-2}$
CCNA2	68	0.61	0.19	3.21	$4.40x10^{-2}$	53	0.26	0.19	1.38	$2.75x10^{-1}$	73	1.07	0.17	6.35	$4.61x10^{-8}$*
ATXN1	2498	-0.31	0.10	-3.21	$4.44x10^{-2}$	2566	0.70	0.13	5.21	$5.78x10^{-6}$*	2392	0.38	0.13	2.86	$2.65x10^{-2}$
DOK3	14016	-0.47	0.15	-3.21	$4.44x10^{-2}$	13951	0.87	0.18	4.80	$2.86x10^{-5}$*	12377	0.42	0.20	2.12	$1.16x10^{-1}$
ASPM	46	0.66	0.21	3.20	$4.46x10^{-2}$	29	0.43	0.27	1.57	$2.08x10^{-1}$	45	1.49	0.22	6.86	$2.22x10^{-9}$*
RNF144A	391	0.46	0.14	3.20	$4.47x10^{-2}$	596	-0.85	0.13	-6.34	$4.29x10^{-8}$*	686	-0.28	0.15	-1.85	$1.79x10^{-1}$
SPC24	50	0.66	0.21	3.20	$4.47x10^{-2}$	19	0.42	0.23	1.84	$1.32x10^{-1}$	28	1.28	0.21	6.12	$1.55x10^{-7}$*

RAD51	56	0.59	0.18	3.19	$4.52x10^{-2}$	41	0.23	0.15	1.54	$2.16x10^{-1}$	60	1.15	0.20	5.75	$1.01x10^{-6}$*
PYGL	15676	-0.47	0.15	-3.18	$4.54x10^{-2}$	14316	1.11	0.21	5.33	$3.50x10^{-6}$*	12455	0.65	0.22	2.97	$2.10x10^{-2}$
CHEK1	43	0.56	0.18	3.18	$4.55x10^{-2}$	36	0.10	0.14	0.68	$6.21x10^{-1}$	48	0.90	0.18	4.96	$3.96x10^{-5}$*
TYMS	163	0.64	0.20	3.18	$4.60x10^{-2}$	90	0.67	0.22	3.04	$9.14x10^{-3}$	151	1.71	0.23	7.45	$6.25x10^{-11}$*
TG	105	-0.52	0.16	-3.17	$4.64x10^{-2}$	93	1.32	0.19	6.97	$1.64x10^{-9}$*	76	0.73	0.20	3.69	$3.29x10^{-3}$
CLSPN	44	0.54	0.17	3.17	$4.64x10^{-2}$	34	0.46	0.18	2.61	$2.72x10^{-2}$	45	1.14	0.19	6.15	$1.26x10^{-7}$*
ZNF480	105	0.38	0.12	3.15	$4.80x10^{-2}$	153	-0.63	0.13	-4.95	$1.58x10^{-5}$*	171	-0.22	0.12	-1.80	$1.94x10^{-1}$
NSMCE4A	260	0.31	0.10	3.15	$4.80x10^{-2}$	386	-0.62	0.12	-5.10	$8.94x10^{-6}$*	425	-0.29	0.12	-2.38	$7.18x10^{-2}$
SSB	304	0.39	0.12	3.15	$4.80x10^{-2}$	428	-0.60	0.13	-4.62	$5.67x10^{-5}$*	484	-0.16	0.13	-1.19	$4.21x10^{-1}$
SLC2A3	16672	-0.43	0.14	-3.14	$4.88x10^{-2}$	15577	1.00	0.21	4.84	$2.40x10^{-5}$*	13863	0.59	0.22	2.73	$3.53x10^{-2}$
DZIP1L	33	0.63	0.20	3.14	$4.95x10^{-2}$	15	0.94	0.23	4.13	$3.31x10^{-4}$	29	1.96	0.27	7.39	$8.91x10^{-11}$*
ARNTL2	30	0.51	0.16	3.13	$4.96x10^{-2}$	25	0.33	0.18	1.77	$1.49x10^{-1}$	32	0.93	0.17	5.46	$4.11x10^{-6}$*
ASPH	963	-0.51	0.16	-3.13	$4.98x10^{-2}$	923	1.09	0.20	5.44	$2.28x10^{-6}$*	768	0.50	0.20	2.49	$5.78x10^{-2}$

Log2FC indicates Log2 fold change; SE indicates standard error; Stat indicates the results from Wald statistic; and * indicates Bonferroni corrected significant P<1.1X10-4 (P<0.05/451 DE genes).

Table III. Overlap between the differentially expressed genes and differentially methylated CpG sites.

Chr	Gene start position	Gene end position	Gene	CpG site position	CpG site p-value
1	9101426	9129887	SLC2A5	9100067	$3.29x10^{-3}$
1	9101426	9129887	SLC2A5	9147808	$9.31x10^{-3}$
1	9101426	9129887	SLC2A5	9243783	$8.36x10^{-4}$
1	9101426	9129887	SLC2A5	9368498	$6.38x10^{-3}$
1	36197712	36235551	CLSPN	36029011	$2.27x10^{-3}$
1	47023078	47069966	MKNK1	46956771	$7.41x10^{-3}$
1	200520624	200589862	KIF14	200512245	$5.72x10^{-3}$
1	212208894	212278348	DTL	212266836	$3.94x10^{-3}$
1	214776531	214837914	CENPF	214846563	$4.10x10^{-3}$
1	223282747	223316624	TLR5	223487288	$5.21x10^{-3}$
2	7057522	7184309	RNF144A	6922444	$8.81x10^{-3}$
2	7057522	7184309	RNF144A	7148323	$1.56x10^{-4}$
2	7057522	7184309	RNF144A	7148362	$5.43x10^{-3}$
2	7057522	7184309	RNF144A	7213398	$2.32x10^{-3}$
2	102759235	102796334	IL1R1	102647671	$2.18x10^{-3}$
2	102759235	102796334	IL1R1	102647718	$9.45x10^{-3}$
2	128848753	128953249	UGGT1	128973012	$7.59x10^{-3}$
2	128848753	128953249	UGGT1	129033961	$2.77x10^{-3}$
2	128848753	128953249	UGGT1	129037628	$7.08x10^{-3}$
2	234296799	234380743	DGKD	234093536	$5.06x10^{-3}$
2	234296799	234380743	DGKD	234394437	$6.31x10^{-3}$
2	239969863	240322643	HDAC4	239756133	$7.05x10^{-3}$
2	239969863	240322643	HDAC4	239763625	$9.61x10^{-3}$
2	239969863	240322643	HDAC4	240199099	$9.29x10^{-3}$
2	239969863	240322643	HDAC4	240302529	$2.14x10^{-3}$
3	12329348	12475855	PPARG	12110937	$7.86x10^{-3}$
3	12329348	12475855	PPARG	12482963	$1.52x10^{-3}$
3	44803208	44894748	KIF15	44902961	$6.07x10^{-3}$
3	44803208	44894748	KIF15	45106462	$5.07x10^{-3}$
3	46477495	46505161	LTF	46750030	$9.61x10^{-3}$
3	52529355	52558511	STAB1	52389151	$7.61x10^{-4}$
3	52529355	52558511	STAB1	52566526	$7.08x10^{-3}$
3	71820805	71834357	PROK2	71721378	$2.46x10^{-3}$
3	71820805	71834357	PROK2	71932814	$8.91x10^{-4}$
4	89378267	89427319	HERC5	89132710	$5.77x10^{-3}$
5	17130137	17217156	AC091878.1	16950494	$1.59x10^{-4}$
5	17130137	17217156	AC091878.1	17147134	$7.32x10^{-3}$
5	171760502	171881527	SH3PXD2B	172131183	$3.02x10^{-3}$
5	172195092	172198203	DUSP1	172131183	$3.02x10^{-3}$
5	172195092	172198203	DUSP1	172385451	$5.69x10^{-3}$
5	176921996	176930648	RP11-1334A24.6	176759081	$9.88x10^{-3}$
5	176921996	176930648	RP11-1334A24.6	176759115	$9.71x10^{-3}$
5	176921996	176930648	RP11-1334A24.6	176789620	$5.54x10^{-3}$
5	176928905	176937427	DOK3	176832034	$9.25x10^{-3}$
5	176928905	176937427	DOK3	176846745	$5.16x10^{-3}$

5	176928905	176937427	DOK3	176859147	$9.23x10^{-3}$
5	178537851	178772431	ADAMTS2	178913356	$9.18x10^{-3}$
6	49572889	49604587	RHAG	49493945	$8.98x10^{-3}$
6	49695091	49712150	CRISP3	49493945	$8.98x10^{-3}$
7	1127723	1133451	GPER	1095452	$4.51x10^{-3}$
7	1127723	1133451	GPER	1338332	$3.93x10^{-3}$
7	41728600	41742706	INHBA	41502683	$8.89x10^{-3}$
7	48211056	48687091	ABCA13	48231093	$8.56x10^{-3}$
7	101459183	101901513	CUX1	101264700	$5.74x10^{-4}$
7	101459183	101901513	CUX1	101335114	$3.86x10^{-3}$
7	101459183	101901513	CUX1	101365760	$3.77x10^{-3}$
7	101459183	101901513	CUX1	101927584	$9.25x10^{-3}$
7	101459183	101901513	CUX1	102043849	$6.68x10^{-3}$
8	133879204	134147143	TG	134106571	$9.62x10^{-3}$
8	133879204	134147143	TG	134135662	$4.07x10^{-3}$
8	133879204	134147143	TG	134251022	$4.62x10^{-3}$
9	34398181	34458568	FAM219A	34181255	$8.11x10^{-3}$
9	34398181	34458568	FAM219A	34490485	$7.13x10^{-3}$
9	39072763	39288300	CNTNAP3	39033093	$9.08x10^{-3}$
9	95473644	95527083	BICD2	95233776	$8.09x10^{-3}$
9	95473644	95527083	BICD2	95413555	$2.59x10^{-3}$
9	132250938	132275965	LINC00963	132109786	$5.17x10^{-3}$
9	132250938	132275965	LINC00963	132109934	$8.85x10^{-3}$
9	132250938	132275965	LINC00963	132257756	$7.62x10^{-4}$
9	132250938	132275965	LINC00963	132357588	$4.05x10^{-3}$
9	136028334	136039332	GBGT1	136181762	$9.36x10^{-3}$
10	6186842	6277507	PFKFB3	6018769	$7.86x10^{-3}$
10	45940018	45948569	RP11-67C2.2	45817319	$2.22x10^{-4}$
10	45940018	45948569	RP11-67C2.2	45870144	$2.69x10^{-4}$
10	62538088	62546827	CDK1	62751435	$8.54x10^{-3}$
10	94352824	94415152	KIF11	94336562	$3.24x10^{-3}$
10	94352824	94415152	KIF11	94550862	$5.01x10^{-3}$
10	95256368	95288849	CEP55	95051132	$2.50x10^{-3}$
10	95256368	95288849	CEP55	95062681	$7.38x10^{-3}$
10	95256368	95288849	CEP55	95327878	$7.25x10^{-3}$
10	99473464	99477909	MARVELD1	99258396	$9.21x10^{-3}$
10	99473464	99477909	MARVELD1	99446317	$4.06x10^{-3}$
10	105791045	105845638	COL17A1	105801896	$9.89x10^{-3}$
11	45950869	46142985	PHF21A	45803411	$4.33x10^{-3}$
11	76839309	76926286	MYO7A	76796073	$7.66x10^{-3}$
11	76839309	76926286	MYO7A	76900467	$1.02x10^{-3}$
11	94277016	94283064	FUT4	94502805	$6.06x10^{-3}$
11	94300473	94354587	PIWIL4	94502805	$6.06x10^{-3}$
12	2966846	2986321	FOXM1	3053621	$8.55x10^{-3}$
12	8071823	8088892	SLC2A3	7910506	$7.00x10^{-3}$
12	8071823	8088892	SLC2A3	7979933	$1.96x10^{-3}$
12	8071823	8088892	SLC2A3	8070526	$7.36x10^{-3}$

12	10310898	10324790	OLR1	10362476	$3.37x10^{-3}$
12	12482217	12503169	MANSC1	12388557	$8.78x10^{-3}$
12	53662082	53687427	ESPL1	53660942	$6.75x10^{-4}$
12	109185694	109251359	SSH1	109170831	$3.43x10^{-3}$
12	120123594	120315095	CIT	120398191	$8.51x10^{-3}$
12	120123594	120315095	CIT	120549925	$8.00x10^{-3}$
13	28577410	28674729	FLT3	28368126	$9.70x10^{-3}$
13	28577410	28674729	FLT3	28503059	$1.07x10^{-3}$
13	28577410	28674729	FLT3	28659295	$4.69x10^{-3}$
14	23242431	23285101	SLC7A7	23310733	$4.64x10^{-3}$
14	23305741	23316808	MMP14	23310733	$4.64x10^{-3}$
14	23415436	23426351	HAUS4	23310733	$4.64x10^{-3}$
14	25042723	25045466	CTSG	24804985	$9.50x10^{-3}$
14	55614833	55658396	DLGAP5	55584030	$4.28x10^{-3}$
14	77490885	77495042	IRF2BPL	77334528	$5.90x10^{-3}$
14	77490885	77495042	IRF2BPL	77341636	$8.06x10^{-3}$
14	95648275	95786245	CLMN	95983475	$4.06x10^{-3}$
14	95648275	95786245	CLMN	96033927	$6.55x10^{-4}$
14	102829299	102968818	TECPR2	102776206	$3.45x10^{-3}$
14	102829299	102968818	TECPR2	102781393	$8.67x10^{-3}$
14	102829299	102968818	TECPR2	103005143	$4.60x10^{-3}$
14	102829299	102968818	TECPR2	103097135	$6.38x10^{-3}$
14	102829299	102968818	TECPR2	103160813	$9.79x10^{-3}$
15	40380091	40398639	BMF	40544713	$6.33x10^{-3}$
15	40380091	40398639	BMF	40559641	$9.19x10^{-3}$
15	40886446	40954881	CASC5	41150887	$8.20x10^{-3}$
15	40987326	41024356	RAD51	41150887	$8.20x10^{-3}$
16	3082482	3089134	RP11-473M20.5	2840498	$5.84x10^{-3}$
16	3082482	3089134	RP11-473M20.5	3011066	$8.47x10^{-3}$
16	3082482	3089134	RP11-473M20.5	3065902	$2.72x10^{-3}$
16	3101992	3109364	RP11-473M20.7	3236725	$9.36x10^{-3}$
16	3101992	3109364	RP11-473M20.7	3239683	$5.22x10^{-3}$
16	27325229	27376099	IL4R	27367633	$1.35x10^{-3}$
16	31271287	31344213	ITGAM	31540544	$7.98x10^{-3}$
16	67207755	67209640	NOL3	67195610	$2.60x10^{-3}$
17	9813925	9929623	GAS7	9891114	$1.80x10^{-3}$
17	38544772	38574202	TOP2A	38584962	$3.21x10^{-3}$
17	38544772	38574202	TOP2A	38755761	$8.44x10^{-3}$
17	40351194	40428424	STAT5B	40192921	$4.43x10^{-3}$
17	40351194	40428424	STAT5B	40266545	$1.77x10^{-3}$
17	56347216	56358296	MPO	56268785	$7.75x10^{-3}$
17	56347216	56358296	MPO	56405141	$6.24x10^{-3}$
17	76419777	76573476	DNAH17	76373256	$5.16x10^{-3}$
17	76422409	76422834	AC061992.1	76373256	$5.16x10^{-3}$
17	79008946	79091232	BAIAP2	79067113	$3.14x10^{-3}$
17	79091095	79139872	AATK	79097160	$8.14x10^{-3}$
17	79091095	79139872	AATK	79163584	$8.63x10^{-3}$

17	79276623	79283048	LINC00482	79163584	$8.63x10^{-3}$
17	79276623	79283048	LINC00482	79350157	$3.52x10^{-3}$
17	79276623	79283048	LINC00482	79433743	$7.86x10^{-3}$
17	79276623	79283048	LINC00482	79450592	$7.80x10^{-4}$
18	657603	673499	TYMS	559877	$8.32x10^{-3}$
18	657603	673499	TYMS	904650	$4.31x10^{-3}$
18	61582744	61602476	SERPINB10	61559787	$4.73x10^{-3}$
19	827830	832017	AZU1	581656	$5.43x10^{-3}$
19	827830	832017	AZU1	770078	$9.66x10^{-3}$
19	827830	832017	AZU1	788944	$8.25x10^{-3}$
19	827830	832017	AZU1	788951	$8.07x10^{-3}$
19	827830	832017	AZU1	808617	$1.76x10^{-3}$
19	827830	832017	AZU1	823534	$4.37x10^{-3}$
19	827830	832017	AZU1	835209	$1.61x10^{-3}$
19	827830	832017	AZU1	928287	$2.21x10^{-3}$
19	827830	832017	AZU1	999969	$5.73x10^{-3}$
19	827830	832017	AZU1	1036519	$6.03x10^{-3}$
19	827830	832017	AZU1	1060664	$3.30x10^{-3}$
19	827830	832017	AZU1	1079683	$1.73x10^{-3}$
19	827830	832017	AZU1	1079726	$8.22x10^{-3}$
19	827830	832017	AZU1	1079743	$7.55x10^{-3}$
19	827830	832017	AZU1	1079745	$9.42x10^{-3}$
19	827830	832017	AZU1	1079751	$9.46x10^{-3}$
19	5993174	6110664	RFX2	5823791	$4.74x10^{-3}$
19	5993174	6110664	RFX2	6273928	$3.29x10^{-4}$
19	5993174	6110664	RFX2	6274810	$4.50x10^{-3}$
19	11257830	11266484	SPC24	11221291	$9.46x10^{-3}$
19	11257830	11266484	SPC24	11285276	$9.11x10^{-3}$
19	11257830	11266484	SPC24	11367635	$7.42x10^{-3}$
19	11257830	11266484	SPC24	11492648	$6.28x10^{-3}$
19	35531409	35557477	HPN	35490410	$6.66x10^{-3}$
19	35531409	35557477	HPN	35514655	$6.11x10^{-5}$
19	35531409	35557477	HPN	35514678	$6.56x10^{-3}$
19	35531409	35557477	HPN	35514694	$8.94x10^{-3}$
19	35531409	35557477	HPN	35569572	$5.49x10^{-3}$
19	35531409	35557477	HPN	35610338	$6.75x10^{-3}$
19	39805291	39811498	CTC-246B18.8	39575128	$8.90x10^{-5}$
19	39805291	39811498	CTC-246B18.8	39589635	$5.93x10^{-3}$
19	42259427	42276113	CEACAM6	42416578	$5.05x10^{-4}$
19	52800421	52829180	ZNF480	52956986	$7.69x10^{-3}$
19	54296854	54327657	NLRP12	54218164	$4.68x10^{-3}$
19	54296854	54327657	NLRP12	54266550	$5.24x10^{-3}$
19	54296854	54327657	NLRP12	54545252	$2.96x10^{-3}$
19	54573200	54584634	TARM1	54545252	$2.96x10^{-3}$
19	54573200	54584634	TARM1	54607111	$3.44x10^{-3}$
19	55824168	55836708	TMEM150B	55591996	$2.24x10^{-3}$
19	55824168	55836708	TMEM150B	55760810	$6.97x10^{-3}$

20	1784662	1798252	RP5-968J1.1	1876682	$7.03x10^{-3}$
20	1874812	1920540	SIRPA	1920583	$4.29x10^{-3}$
20	4760669	4804291	RASSF2	4969742	$4.66x10^{-3}$
20	36932551	36965905	BPI	36853042	$3.85x10^{-3}$
20	48807119	48809227	CEBPB	48572573	$1.61x10^{-3}$
20	48807119	48809227	CEBPB	48692927	$8.98x10^{-3}$
21	34638664	34669539	IL10RB	34399336	$7.39x10^{-3}$
21	34638664	34669539	IL10RB	34436976	$5.55x10^{-3}$
21	34638664	34669539	IL10RB	34436978	$4.69x10^{-3}$
21	34638664	34669539	IL10RB	34850937	$6.38x10^{-3}$
21	35014783	35210802	ITSN1	34850937	$6.38x10^{-3}$
21	35014783	35210802	ITSN1	34976656	$1.07x10^{-3}$
21	35014783	35210802	ITSN1	35219554	$9.97x10^{-3}$
21	39751949	39956869	ERG	39840765	$9.10x10^{-3}$
22	19467348	19508135	CDC45	19701472	$6.40x10^{-3}$
22	23401593	23484241	RTDR1	23362153	$9.35x10^{-3}$
22	23401593	23484241	RTDR1	23644647	$6.97x10^{-3}$
X	77526968	77583188	CYSLTR1	77359794	$8.10x10^{-3}$

Table IV. Results of the detailed reactome pathway analysis of 14 cell cycle genes. Pathways are sorted by the P-value.

Pathway name	Entities found	Entities total	Entities ratio	Entities P-value	Entities FDR	Reactions found	Reactions total	Reactions ratio	Submitted entities found
Cell Cycle	9	574	0.059	$3.25x10^{-8}$	$9.88x10^{-6}$	89	345	0.046	SPC24; ESPL1; CLSPN; CDC45; CENPF; FOXM1; CDK1; CENPF; RAD51
Cell Cycle, Mitotic	7	496	0.051	$4.12x10^{-6}$	$6.02x10^{-4}$	84	272	0.036	SPC24; ESPL1; CDC45; CENPF; FOXM1; CDK1; CENPF
Polo-like kinase mediated events	3	23	0.002	$5.97x10^{-6}$	$6.02x10^{-4}$	12	15	0.002	CENPF; FOXM1; CENPF
G2/M Transition	4	133	0.014	$4.30x10^{-5}$	0.002	47	62	0.008	CENPF; FOXM1; CDK1; CENPF
Mitotic G2-G2/M phases	4	135	0.014	$4.56x10^{-5}$	0.002	47	64	0.009	CENPF; FOXM1; CDK1; CENPF
G2/M Checkpoints	3	46	0.005	$4.67x10^{-5}$	0.002	4	16	0.002	CLSPN; CDC45; CDK1
G1/S-Specific Transcription	2	17	0.002	$3.19x10^{-4}$	0.014	1	1	$1.33x10^{-4}$	CDC45; CDK1
Cyclin A/B1 associated events during G2/M transition	2	26	0.003	$7.39x10^{-4}$	0.024	21	24	0.003	FOXM1; CDK1
Cell Cycle Checkpoints	3	121	0.012	$7.93x10^{-4}$	0.024	4	38	0.005	CLSPN; CDC45; CDK1
Resolution of Sister Chromatid Cohesion	3	122	0.013	$8.12x10^{-4}$	0.024	5	8	0.001	SPC24; CENPF; CDK1
Mitotic Prometaphase	3	130	0.013	$9.75x10^{-4}$	0.026	7	13	0.002	SPC24;CENPF;CDK1
M Phase	4	306	0.032	0.001	0.026	19	63	0.008	SPC24; ESPL1; CENPF; CDK1
E2F mediated regulation of DNA replication	2	33	0.003	0.001	0.027	3	6	$7.99x10^{-4}$	CDC45; CDK1
Activation of ATR in response to replication stress	2	39	0.004	0.002	0.032	2	9	0.001	CLSPN; CDC45
Kinesins	2	44	0.005	0.002	0.032	4	14	0.002	KIF15; KIF11
Separation of Sister Chromatids	3	179	0.018	0.002	0.032	4	8	0.001	SPC24; ESPL1; CENPF
Mitotic Anaphase	3	193	0.020	0.003	0.032	4	11	0.001	SPC24; ESPL1; CENPF
Mitotic Metaphase and Anaphase	3	194	0.020	0.003	0.032	4	12	0.002	SPC24; ESPL1; CENPF
Assembly of the RAD51-ssDNA nucleoprotein complex	1	5	$5.16x10^{-4}$	0.008	0.032	3	3	$3.99x10^{-4}$	RAD51
Phosphorylation of proteins involved in the G2/M transition by Cyclin A:Cdc2 complexes	1	5	$5.16x10^{-4}$	0.008	0.032	2	2	$2.66x10^{-4}$	CDK1

Presynaptic phase of homologous DNA pairing and strand exchange	1	6	$6.20x10^{-4}$	0.009	0.032	4	7	$9.32x10^{-4}$	RAD51
G2/M DNA replication checkpoint	1	7	$7.23x10^{-4}$	0.011	0.032	2	2	$2.66x10^{-4}$	CDK1
Homologous DNA pairing and strand exchange	1	7	$7.23x10^{-4}$	0.011	0.032	4	8	0.001	RAD51
Chk1/Chk2(Cds1) mediated inactivation of Cyclin B:Cdk1 complex	1	7	$7.23x10^{-4}$	0.011	0.032	1	5	$6.66x10^{-4}$	CDK1
Phosphorylation of Emi1	1	8	$8.26x10^{-4}$	0.012	0.032	1	2	$2.66x10^{-4}$	CDK1
G1/S Transition	2	113	0.012	0.013	0.032	7	32	0.004	CDC45; CDK1
Cyclin B2 mediated events	1	9	$9.29x10^{-4}$	0.014	0.032	2	2	$2.66x10^{-4}$	CDK1
Activation of NIMA Kinases NEK9, NEK6, NEK7	1	9	$9.29x10^{-4}$	0.014	0.032	1	4	$5.33x10^{-4}$	CDK1
E2F-enabled inhibition of pre-replication complex formation	1	10	0.001	0.015	0.032	2	2	$2.66x10^{-4}$	CDK1
Unwinding of DNA	1	11	0.001	0.017	0.032	1	4	$5.33x10^{-4}$	CDC45
G2/M DNA damage checkpoint	1	11	0.001	0.017	0.032	1	7	$9.32x10^{-4}$	CDK1
MHC class II antigen presentation	2	135	0.014	0.018	0.032	1	23	0.003	KIF15; KIF11
MASTL Facilitates Mitotic Progression	1	12	0.001	0.018	0.032	1	4	$5.33x10^{-4}$	CDK1
Mitotic G1-G1/S phases	2	139	0.014	0.019	0.032	8	53	0.007	CDC45; CDK1
ERK1 activation	1	13	0.001	0.020	0.032	1	6	$7.99x10^{-4}$	CDK1
Factors involved in megakaryocyte development and platelet production	2	150	0.015	0.022	0.032	4	42	0.006	KIF15; KIF11
Condensation of Prometaphase Chromosomes	1	15	0.002	0.023	0.032	1	4	$5.33x10^{-4}$	CDK1
ERK activation	1	15	0.002	0.023	0.032	1	11	0.001	CDK1
Phosphorylation of the APC/C	1	17	0.002	0.026	0.032	1	2	$2.66x10^{-4}$	CDK1
Golgi Cisternae Pericentriolar Stack Reorganization	1	17	0.002	0.026	0.032	1	6	$7.99x10^{-4}$	CDK1
Homologous recombination repair of replication-independent double-strand breaks	1	19	0.002	0.029	0.032	4	24	0.003	RAD51
Homologous Recombination Repair	1	19	0.002	0.029	0.032	4	24	0.003	RAD51
APC/C:Cdc20 mediated degradation of Cyclin B	1	21	0.002	0.032	0.032	3	3	$3.99x10^{-4}$	CDK1
RAF/MAP kinase cascade	1	21	0.002	0.032	0.032	1	19	0.003	CDK1

Depolymerisation of the Nuclear Lamina	1	23	0.002	0.035	0.035	2	6	$7.99x10^{-4}$	CDK1
Double-Strand Break Repair	1	25	0.003	0.038	0.038	4	32	0.004	RAD51
SOS-mediated signalling	1	26	0.003	0.040	0.040	1	21	0.003	CDK1
GRB2 events in EGFR signaling	1	26	0.003	0.040	0.040	1	22	0.003	CDK1
Recruitment of NuMA to mitotic centrosomes	1	27	0.003	0.041	0.041	1	3	$3.99x10^{-4}$	CDK1
G0 and Early G1	1	27	0.003	0.041	0.041	1	10	0.001	CDK1
SHC-mediated signalling	1	27	0.003	0.041	0.041	1	21	0.003	CDK1
SHC1 events in EGFR signaling	1	27	0.003	0.041	0.041	1	23	0.003	CDK1
Signalling to p38 via RIT and RIN	1	28	0.003	0.042	0.042	1	22	0.003	CDK1
ARMS-mediated activation	1	29	0.003	0.044	0.044	1	24	0.003	CDK1
SHC-related events	1	30	0.003	0.045	0.045	1	24	0.003	CDK1
Signaling by Leptin	1	30	0.003	0.045	0.045	1	38	0.005	CDK1
Frs2-mediated activation	1	32	0.003	0.048	0.048	1	25	0.003	CDK1
Activation of the pre-replicative complex	1	33	0.003	0.050	0.050	4	8	0.001	CDC45
SHC-related events triggered by IGF1R	1	33	0.003	0.050	0.050	1	23	0.003	CDK1

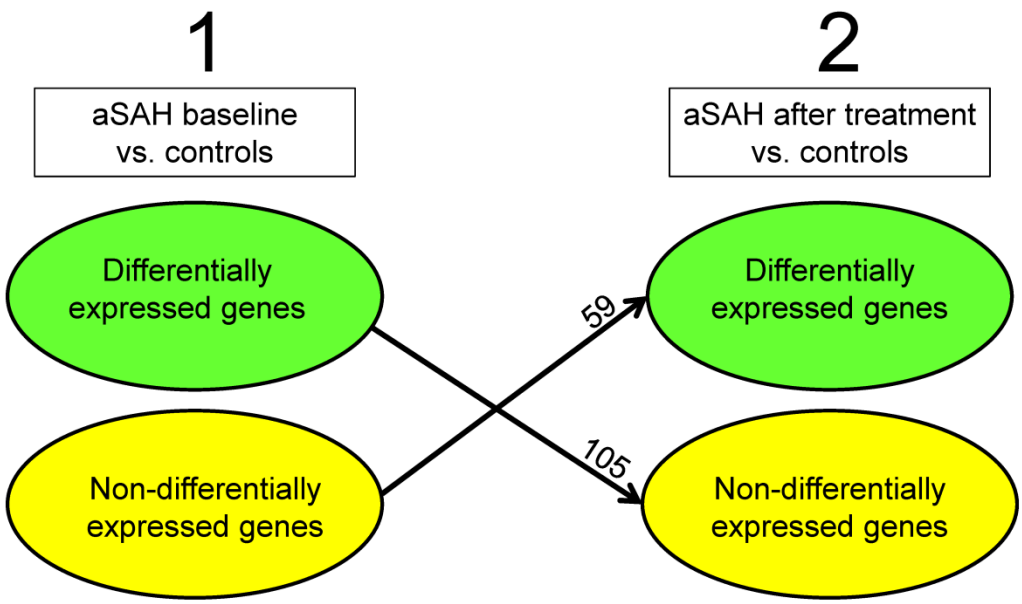


Figure I. Schematic overview of the analysis of the 451 differentially expressed (DE) genes using the controls who did not receive the treatment. We subsequently focused on the 164 genes that changed the DE status between the two analyses.

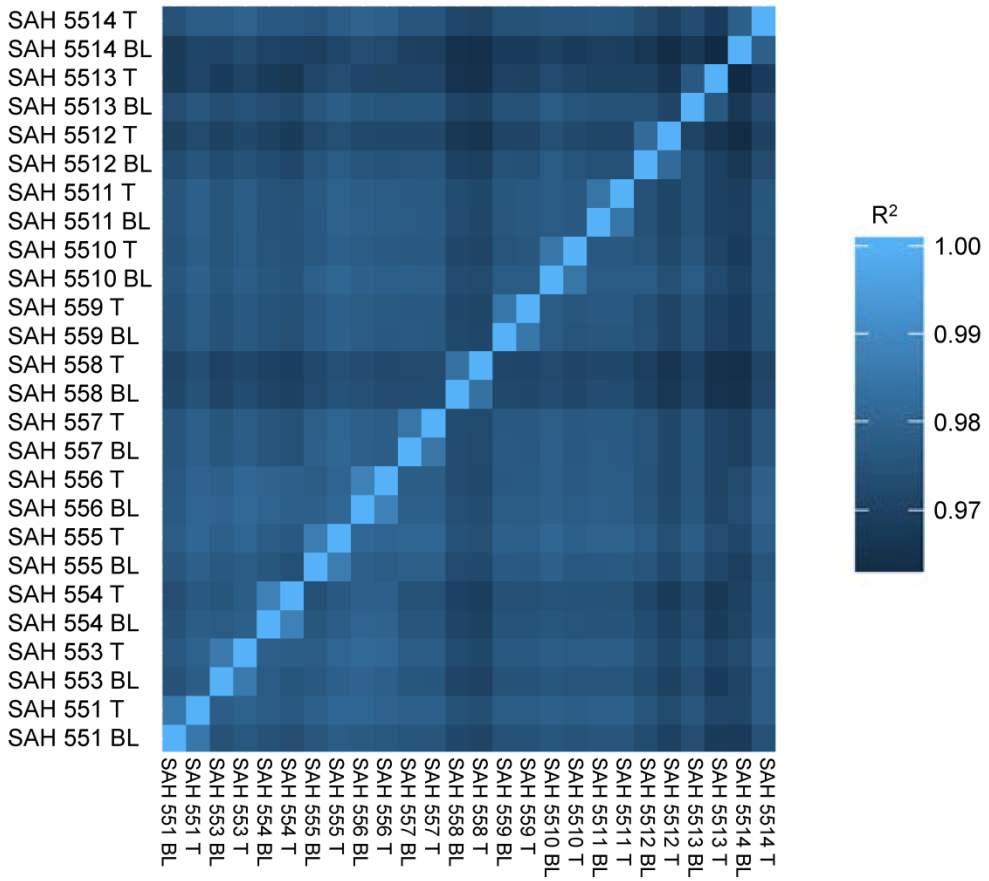


Figure II. The correlation of 676,543 CpG sites assayed in all individuals between the samples. BL indicates baseline and T indicates treatment. The overall methylation status changed very little within an individual (~98%) and between the individuals (~97.5%).

Chapter 5

Conclusions and future directions

Chapter 5

Conclusions and future directions

My graduate work focused on two main themes: 1) identifying variants and genes underlying low and high plasma LDL-C levels, and 2) identifying dynamic genomic signatures associated with RIC, a putative treatment for preventing ischemic reperfusion injury in the heart and the brain. We have used a variety of state-of-the-art massive parallel sequencing methods and study designs in which we employed pedigree information to identify novel genes or variants associated with dyslipidemia. We also generated intra-individual longitudinal genomic data to study effects of RIC on gene expression and DNA methylation. Our studies illustrate the possibilities of the latest genomic techniques that ultimately should help move towards family-centered, personalized medicine.

When I started working on these projects, the field of genetics and genomics was transitioning. GWAS had identified several loci for different complex diseases, but most of the variants in these loci had only small effect sizes. For example, even though Teslovitz *et al*. extended the lipid-associated GWAS loci to a total of 95 [1], researches had difficulties to identify clear candidate genes or functional variants within these mapped regions. The reasons for this lack of success include large linkage disequilibrium blocks, lack of fine mapping (i.e. the researchers often reported only the closest “bookmark” gene), and the difficulty and high-cost of re-sequencing. Massive parallel whole exome-sequencing and targeted re-sequencing of these previously identified loci were meant to overcome some of these limitations. Indeed, exome-sequencing of families resulted in promising results for monogenic diseases [2]. Also, we were one of the first laboratories to identify rare variants for a complex disease by exome sequencing [3] (Appendix I). This initial success gave us the confidence to take on exome-sequencing projects in dyslipidemic families with the goal of identifying novel genes and variants

for lipid traits. We hypothesized that the affected family members would share the same functional variant, contributing to the dyslipidemia.

We first analyzed two Mexican families exhibiting familial hypobetalipoproteinemia (FHBL). FHBL is characterized by low plasma levels of Apo B and LDL-C. As identification of novel genes for low LDL-C could lead to the identification of new therapeutic drug targets, we decided to exome sequence these two families (Chapter 3). We identified a region on chr2p16 shared by all affected individuals, but could not identify the actual causal variants that could account, with confidence, for the low LDL-C and Apo B levels. Because it is still possible that we missed the functional variant, we periodically re-analyze the data, especially since sequence analysis and genome annotation tools improve all the time. Since it is difficult to publish negative data, we suspect that, as was the case with our study, a lot of variants of unknown function are not referenced. It is unfortunate because the identification of the variants and genes would gain further support if they were repeatedly seen in multiple independent dyslipidemic families.

We suspect that the fairly small pedigree sizes of the two Mexican families we studied might have been the major reason for our lack of success since it complicated our analysis and decreased our statistical power. As the families share a large proportion of the genome, there are too many variants to evaluate and filter using the annotation databases. Filtering out the "non-functional" variants might be cumbersome considering the overwhelming number of variants, and these filtering techniques can be rather arbitrary. For example, filters like "minor allele frequency" or "novelty" have shortcomings when searching public databases, because most of these databases only retain genotypic data, but do not provide any phenotypic data. Furthermore, no individual level data is available in these public data bases. The recently launched Genome Aggregation Database (gnomAD) offers aggregate data on 60,706 exomes from unrelated individuals [5], and most likely includes people with dyslipidemia as well. In the

case of FHBL, with a prevalence estimate of 1 in 1,000 [6], one should expect about 60 individuals in this database to have FHBL. This is probably an underestimation, as the prevalence of FHBL might be even higher since many individuals are asymptomatic. If individual level genotype and phenotypic data were available, it would greatly facilitate the identification of novel rare variants.

In the second project (Chapter 2), we studied a large Austrian family with familial hypercholesterolemia (FH). This family was particularly interesting because the proband did not carry any mutation known to cause FH, and because several individuals in this family had both high LDL-C levels and premature CHD. We took this challenge as it provided us an opportunity to find novel genes for hypercholesterolemia utilizing a large multigenerational family. The use of large families offers the power to detect significant linkage regions (lod>3.0), which can be very useful prior to exome sequencing [7]. Employing the mapping data together with exome sequencing analysis, we could focus on portions of the genome that segregated with the disease. Our analysis did not, however, result in the identification of one single significant locus for LDL-C (lod>3.0). Even when we focused on regions with a lod score of >1.0, we could not identify potentially functional variants in these regions. However, when we characterized the genomic differences between the affected and unaffected family members in more detail, we identified rare and putative variant combinations in the known FH genes in three affected family members. We also uncovered a family-specific polygenic risk score, distinguishing the affected family members from the unaffected family members. Specifically, we identified a 10-SNP combination of LDL-C increasing GWAS variants that separated the affected from the unaffected family members. This family illustrates the fact that different disease etiologies might contribute to clinical FH. It also highlights the necessity to consider each variant not just separately, but within the genetic background and familial environment of the carrier. A given variant might be pathogenic in certain genetic and environmental background, but the negative

effects of that single variant could be out-weighted by protective variants or lifestyle choices in other individuals [8,9].

In line with this idea of different disease etiologies that are family-specific, a recent study by Ripatti *et al*. aimed to find high-impact Mendelian variants followed by a search for polygenic cause in dyslipidemic families. They found a known Mendelian variant in 3% of the affected individuals and 35 % had an aggregate of either known LDL-C or TG elevating variants (polygenic score > 90th percentile in the population). However, the genetic etiology could not be identified in 62% families using this approach [10]. To continue the search of the genetic cause for dyslipidemia in these families, one approach could be to try to find family-specific combinations of GWAS variants similarly as we did in the study described in chapter 2. Another approach would be to whole genome sequence these families to rule out the possibility that the disease is caused by intronic or intergenic copy number variations (CNV), small indels, or rare variants outside of coding regions. However, whole genome sequencing still presents some significant challenges. One of the challenges is our incomplete knowledge on annotations of non-coding rare and low-frequency variants. Another one is the fact that GC rich regions are still unreachable with the current methods (using short sequences). Finally, it is still very costly to sequence the whole genome. Despite these challenges, for example, the proband of the Austrian FH family would be interesting to sequence through the whole genome. Since we only found subtle changes in this patient's genome by SNP array and exome-sequencing, it is tempting to hypothesize that he might have some structural variation(s) (indel, CNV) that could only be discovered by whole genome sequencing.

Regardless of the genomic location of the susceptibility variants, it remains challenging to functionally test the variants and provide evidence of their pathogenicity. There is a need for more systematic large scale testing of the pathogenic potential of multiple variants. A promising example of systematic testing of variants is the recent study by Majithia *et al*. on PPARG

variants. The authors developed a high-throughput screen to test the biological activity of all possible missense variants in PPARG (Nature Genetics ref 2016). Using that screen and prior knowledge from previously reported variants known to diminish PPARγ activity, Majithia *et al*. were able to establish a variant classifier that evaluates the likelihood of a functional defect of PPARG missense mutations. The authors went on to use this classifier as a molecular diagnosis tool to predict the pathogenicity of 55 PPARG missense variants recently identified in population-based and clinical sequencing. Among these 55 variants, six were predicted to be pathogenic, and were indeed shown to be defective in subsequent biological assays [11]

Through efforts from the ENCODE [12] and the Roadmap epigenomics projects[13], evidence from recently developed massively parallel reporter assays [14] and eQTL analyses [15] have shown that variants outside of coding regions are functional, and can be potentially pathogenic as well. The CRISPR/Cas9 system is rapidly changing the way we approach functionality of the identified variants [16]. The impact of a single variant can readily be evaluated after CRISPR/Cas9-based DNA editing of cells. One can introduce a variant using this technique, and then, for example, measure enhancer activity or target gene expression in the modified cells [16]. Ideally, in the future we will be able to screen all of the variants identified by exome or whole genome sequencing in a clinical or research setting using this technique. However, even if the molecular effect of genetic variation can be established in the laboratory, it will not necessarily translate into a clinical impact. Human diseases are complex and variation that augments a disease in one individual and/or family background may be benign in another individual and/or family [9,17].

It is well established that introducing lifestyle changes, such as following a healthy diet, exercising regularly, and quitting smoking, can significantly reduce CVD events by 13.7% for smoking; 13.2% for poor diet; 11.9% for insufficient physical activity [18], respectively. Thus the importance of improving lifestyle should never be underestimated when considering the

treatments to prevent CVD. One recent study concluded that a favorable lifestyle can substantially attenuate the high genetic risk of incidence for coronary events and the prevalence of subclinical atherosclerosis burden [8]. On the other hand, this study also showed that a high genetic risk, measured by weighted genetic risk scores for CVD events, cannot be completely overcome by a healthy lifestyle [8].

In theory, clinicians could use weighted polygenic risk scores to predict the risk for a given disease. For example, Ripatti et al. showed that a polygenic CAD risk score can predict the probability of a CAD event as accurately as LDL-C plasma levels [19]. However, one of the limitations of using genetic risk scores is that these scores serve mainly populations of European decent, because the majority of the GWAS studies have been conducted in Caucasians. Ideally, GWAS studies should be conducted in different ethnicities, so that each population would have its own genetic score for a given phenotype as a reference. With our Austrian FH family, we were able to use a Finnish cohort as a reference only because they shared the same allele frequencies for the common variants that we used in the weighted genetic risk score analysis. If the weighted genetic risk scores are planned to be used for predicting disease, it will be critical to obtain accurate effect size estimates with the correct set of replicated and well-validated disease susceptibility variants in each population.

The study of genetic variants can be very useful to evaluate disease risk. However, understanding gene expression regulation and epigenetic effects during disease progression or treatment can also yield insight into molecular mechanisms of disease and related quantitative traits. As an example, the chapter 4 describes our findings on gene expression and DNA methylation changes in aneurysmal subarachnoid hemorrhage (aSAH) patients undergoing RIC. We identified coordinated cell cycle and inflammatory responses in these patients' blood [20]. These changes had been previously implicated in other models of RIC, stroke, or ischemic neuronal death. Even though it is difficult to determine whether these findings can be accounted

for the medications or disease progression of aSAH, it is possible that some are directly related to RIC. There are several ways we could improve on our initial study. One obvious way would be to use larger cohorts and include control patients who do not receive the RIC treatment in follow up studies. Another way to improve our study would be to take advantage of the recent developments in methods to correct for cell type heterogeneity in the blood. Blood transcriptome has advantages over other tissues because it is an easily accessible diagnostic specimen. However, one of the limitations using blood, as was the case in of our study, is that blood is a mixture of many different cell types and the fluctuation in cell populations alone may cause large variations in transcriptomics and DNA methylation [21]. This problem has become more manageable with the progress in flow cytometry methods. It is now possible to study transcriptomics of isolated cell populations from the blood. Computational approaches and cell counting can also help correct for cell composition. Unfortunately, at the time of our analysis, we did not have these methods readily available. It is also noteworthy, that fluctuations in cell populations might be relevant for the treatment. For example, the RIC can potentially lead to the release of substances from the ischemic limb into the bloodstream, and this substance could in turn stimulate white blood cells, such as macrophages, and result in the increase of expression of genes involved in cell cycle and cell proliferation.

As the technology to study transcriptomes (RNAseq analysis) becomes cheaper, faster and more detailed (*i.e*. detection of splicing patterns), it has emerged as the standard method for researchers to evaluate when and where genes are turned on or off in cells and tissues. Transcriptome analysis can also provide information on gene co-expression networks [20,22] and thus provide information on which pathways are turned on or off in different cellular and environmental contexts. However, there are still many technical and analytical limitations in RNAseq analyses. For example, RNA degradation is still an issue when handling RNA samples, and this greatly influences the overall data quality. In addition, we have noticed that batch

effects can cause spurious results, and thus, these types of hidden and known technical factors should be accounted for in the RNAseq data analysis. Many of the analytical choices, such as methods for mapping and counting the RNA reads as well as normalization methods, will influence the downstream analyses. Thus, any RNAseq analysis may be shaped by a series of subjective decisions that may or may not capture the biological relevance of the transcriptome data. For that reason, replicating the transcriptome findings in an independent study sample is essential. One of our future goals is to collect a larger cohort of patients undergoing RIC to validate our transcriptome findings.

Another limitation of our study was that we were not able to cover the whole genome when we performed our methylation analysis. For cost reasons, we chose to use the reduced representation bisulfate sequencing (RRBS) method. The RRBS method was attractive to us because it offers enriched coverage of CpGs representing approximately 10% of all CpG dinucleotides in the human genome. This allowed us to capture many of the promotor regions known to be relevant for gene expression. On the other hand, we might have lost some of the GpC sites because we required a conservative coverage (10X) in both of the time points for every sample. We wanted to be rather conservative since, at the time of our analysis, the RRBS had not yet been very widely used in humans.

Ideally, our studies on transcriptome and DNA methylation during RIC should have been conducted on brain tissue. For obvious practical and ethical reasons that was simply not possible. As a substitute though, blood should have served us well, since the presence of humoral factors had previously been shown in RIC [23]. In respect to different disease etiologies, it is important that the tissue in which the transcriptome and DNA methylation analyses are performed is relevant to the particular disease. For example, studies have shown that many of the GWAS variants reside in the long noncoding RNAs, and those are very tissue-specific [24].

This further reinforces the idea that GWAS variants should always be examined in a relevant tissue for a given disease.

We have currently on-going transcriptome studies in the human liver and adipose tissues of patients undergoing bariatric surgery to treat obesity. For the adipose tissue, we also have follow-up samples, and therefore we should be able to use a longitudinal statistical analysis, similar to the one we used in our study described in chapter 4. Our goal is to identify weight loss-related networks and pathways, with the hope of uncovering molecular mechanisms underlying changes in body weight. This should help better understand genomic factors that influence biological parameters such as insulin sensitivity and lipid levels during weight loss.

In summary, even though the findings in the dyslipidemic families presented in this dissertation may have been modest, they highlight the genetic complexity of lipid traits in humans. Massive efforts are currently made to standardize the exome-sequencing workflow in clinical setting so that variants with unknown function could be better interrogated and integrated into the diagnosis and prevention [25,26]. Our approach, consisting of looking for combinations of GWAS variants that contribute to a given disease in a pedigree, may help the particular family under study. We recognize that these combinations are likely family-specific, and that they may be less relevant to the general population. Nevertheless, approaches utilizing both family structure and genome-wide sequence data, as well as phenotype data in all family members, will likely become more common as a diagnostic tool in the future, and may even potentially become the basis for personalized medicine. In addition, it is noteworthy that the projects presented here were completed through worldwide collaborations. I believe that continuation of collaborative efforts and overall data sharing will be crucial to better understand human health and disease. Open data sharing will allow the scientific community to tackle disease etiologies together, which will eventually lead to major advances in the genomic medicine via careful analysis of the aggregated data. Ideally, technological advances in every

field (genomics, electronics, health devices, proteomics, etc.) will converge towards better medicine. For example, in the future, smart phones could be used to track health parameters, such as blood pressure, glucose levels, and food intake. This information could then be integrated into an individual's medical file to help improve diagnosis and refine therapeutic treatments. We are living in exciting times!

References

1. Teslovich TM, Musunuru K, Smith A V, Edmondson AC, Stylianou IM, Koseki M, Pirruccello JP, Ripatti S, Chasman DI, Willer CJ, Johansen CT, Fouchier SW, Isaacs A, Peloso GM, Barbalic M, Ricketts SL, Bis JC, Aulchenko YS, Thorleifsson G, Feitosa MF, Chambers J, Orho-Melander M, Melander O, Johnson T, Li X, Guo X, Li M, Shin Cho Y, Jin Go M, Jin Kim Y, Lee J-Y, Park T, Kim K, Sim X, Twee-Hee Ong R, Croteau-Chonka DC, Lange LA, Smith JD, Song K, Hua Zhao J, Yuan X, Luan J, Lamina C, Ziegler A, Zhang W, Zee RYL, Wright AF, Witteman JCM, Wilson JF, Willemsen G, Wichmann H-E, Whitfield JB, Waterworth DM, Wareham NJ, Waeber G, Vollenweider P, Voight BF, Vitart V, Uitterlinden AG, Uda M, Tuomilehto J, Thompson JR, Tanaka T, Surakka I, Stringham HM, Spector TD, Soranzo N, Smit JH, Sinisalo J, Silander K, Sijbrands EJG, Scuteri A, Scott J, Schlessinger D, Sanna S, Salomaa V, Saharinen J, Sabatti C, Ruokonen A, Rudan I, Rose LM, Roberts R, Rieder M, Psaty BM, Pramstaller PP, Pichler I, Perola M, Penninx BWJH, Pedersen NL, Pattaro C, Parker AN, Pare G, Oostra BA, O'Donnell CJ, Nieminen MS, Nickerson DA, Montgomery GW, Meitinger T, et al. Biological, clinical and population relevance of 95 loci for blood lipids. *Nature*. 2010;466(7307):707-713. doi:10.1038/nature09270.

2. Ng SB, Bigham AW, Buckingham KJ, Hannibal MC, McMillin MJ, Gildersleeve HI, Beck AE, Tabor HK, Cooper GM, Mefford HC, Lee C, Turner EH, Smith JD, Rieder MJ, Yoshiura K-I, Matsumoto N, Ohta T, Niikawa N, Nickerson DA, Bamshad MJ, Shendure J. Exome sequencing identifies MLL2 mutations as a cause of Kabuki syndrome. *Nat Genet*. 2010;42(9):790-793. doi:10.1038/ng.646.

3. Reddy M V, Iatan I, Weissglas-Volkov D, Nikkola E, Haas BE, Juvonen M, Ruel I, Sinsheimer JS, Genest J, Pajukanta P. Exome sequencing identifies 2 rare variants for low high-density lipoprotein cholesterol in an extended family. *Circ Cardiovasc Genet*. 2012;5(5):538-546. doi:10.1161/CIRCGENETICS.112.963264.

4. Gutiérrez-Cirlos C, Ordóñez-Sánchez ML, Tusié-Luna MT, Patterson BW, Schonfeld G, Aguilar-Salinas CA. Familial hypobetalipoproteinemia in a hospital survey: Genetics, metabolism and non-alcoholic fatty liver disease. *Ann Hepatol*. 2011;10(2):155-164.

5. Lek M, Tewksbury J, Services H. Analysis of protein-coding genetic variation in 60,706 humans. *Nat Publ Gr*. 2014;536(7616):1-26. doi:http://dx.doi.org/10.1101/030338.

6. Linton ME, V R, Young SG. Fam i l ial hypobetal i poprotei nem ia. *J Lipid Res*. 1993;34:521-541.

7. Ott J, Wang J, Leal SM. Genetic linkage analysis in the age of whole-genome sequencing. *Nat Rev Genet*. 2015;16(5):275-284. doi:10.1038/nrg3908.

8. Khera A V., Emdin CA, Drake I, Natarajan P, Bick AG, Cook NR, Chasman DI, Baber U, Mehran R, Rader DJ, Fuster V, Boerwinkle E, Melander O, Orho-Melander M, Ridker PM, Kathiresan S. Genetic Risk, Adherence to a Healthy Lifestyle, and Coronary Disease. *N Engl J Med*. 2016:NEJMoa1605086. doi:10.1056/NEJMoa1605086.

9. Harper AR, Nayee S, Topol EJ. Protective alleles and modifier variants in human health and disease. *Nat Rev Genet*. 2015;16(December):689-701. doi:10.1038/nrg4017.

10. Ripatti P, Rämö JT, Söderlund S, Surakka I, Matikainen N, Pirinen M, Pajukanta P, Sarin

AP, Service SK, Laurila PP, Ehnholm C, Salomaa V, Wilson RK, Palotie A, Freimer NB, Taskinen MR, Ripatti S. The Contribution of GWAS Loci in Familial Dyslipidemias. *PLoS Genet*. 2016;12(5). doi:10.1371/journal.pgen.1006078.

11. Majithia AR, Tsuda B, Agostini M, Gnanapradeepan K, Rice R, Peloso G, Patel KA, Zhang X, Broekema MF, Patterson N, Duby M, Sharpe T, Kalkhoven E, Rosen ED, Barroso I, Ellard S, Kathiresan S, O'Rahilly S, Chatterjee K, Florez JC, Mikkelsen T, Savage DB, Altshuler D. Prospective functional classification of all possible missense variants in PPARG. *Nat Genet*. 2016;(October). doi:10.1038/ng.3700.

12. The ENCODE Project Consortium, Dunham I, Kundaje A, Aldred SF, Collins PJ, Davis C a, Doyle F, Epstein CB, Frietze S, Harrow J, Kaul R, Khatun J, Lajoie BR, Landt SG, Lee B-KB-K, Pauli F, Rosenbloom KR, Sabo P, Safi A, Sanyal A, Shoresh N, Simon JM, Song L, Trinklein ND, Altshuler RC, Birney E, Brown JB, Cheng C, Djebali S, Dong X, Ernst J, Furey TS, Gerstein M, Giardine B, Greven M, Hardison RC, Harris RS, Herrero J, Hoffman MM, Iyer S, Kelllis M, Kheradpour P, Lassman T, Li Q, Lin X, Marinov GK, Merkel A, Mortazavi A, Parker SCJSL, Reddy TE, Rozowsky J, Schlesinger F, Thurman RE, Wang J, Ward LD, Whitfield TW, Wilder SP, Wu W, Xi HS, Yip KY, Zhuang J, Bernstein BE, Green ED, Gunter C, Snyder M, Pazin MJ, Lowdon RF, Dillon L a L, Adams LB, Kelly CJ, Zhang J, Wexler JR, Good PJ, Feingold E a, Crawford GE, Dekker J, Elinitski L, Farnham PJ, Giddings MC, Gingeras TR, Guigó R, Hubbard TJTJ, Kellis M, Kent WJ, Lieb JD, Margulies EH, Myers RM, Starnatoyannopoulos J a, Tennebaum S a, Weng Z, White KP, Wold B, Yu Y, Wrobel J, Risk B a, Gunawardena HP, Kuiper HC, Maier CW, et al. An integrated encyclopedia of DNA elements in the human genome. *Nature*. 2012;489(7414):57-74. doi:10.1038/nature11247.

13. Consortium RE, Kundaje A, Meuleman W, Ernst J, Bilenky M, Yen A, Heravi-Moussavi A, Kheradpour P, Zhang Z, Wang J, Ziller MJ, Amin V, Whitaker JW, Schultz MD, Ward LD, Sarkar A, Quon G, Sandstrom RS, Eaton ML, Wu Y-C, Pfenning AR, Wang X, Claussnitzer M, Liu Y, Coarfa C, Harris RA, Shoresh N, Epstein CB, Gjoneska E, Leung D, Xie W, Hawkins RD, Lister R, Hong C, Gascard P, Mungall AJ, Moore R, Chuah E, Tam A, Canfield TK, Hansen RS, Kaul R, Sabo PJ, Bansal MS, Carles A, Dixon JR, Farh K-H, Feizi S, Karlic R, Kim A-R, Kulkarni A, Li D, Lowdon R, Elliott G, Mercer TR, Neph SJ, Onuchic V, Polak P, Rajagopal N, Ray P, Sallari RC, Siebenthall KT, Sinnott-Armstrong N a., Stevens M, Thurman RE, Wu J, Zhang B, Zhou X, Beaudet AE, Boyer L a., De Jager PL, Farnham PJ, Fisher SJ, Haussler D, Jones SJM, Li W, Marra M a., McManus MT, Sunyaev S, Thomson J a., Tlsty TD, Tsai L-H, Wang W, Waterland R a., Zhang MQ, Chadwick LH, Bernstein BE, Costello JF, Ecker JR, Hirst M, Meissner A, Milosavljevic A, Ren B, Stamatoyannopoulos J a., Wang T, Kellis M, Pfenning A, ClaussnitzerYaping Liu M, et al. Integrative analysis of 111 reference human epigenomes. *Nature*. 2015;518(7539):317-330. doi:10.1038/nature14248.

14. Melnikov A, Murugan A, Zhang X, Tesileanu T, Wang L, Rogov P, Feizi S, Gnirke A, Callan CG, Kinney JB, Kellis M, Lander ES, Mikkelsen TS. Systematic dissection and optimization of inducible enhancers in human cells using a massively parallel reporter assay. *Nat Biotechnol*. 2012;30(3):271-277. doi:10.1038/nbt.2137.

15. Lappalainen T, Sammeth M, Friedländer MR, 't Hoen P a C, Monlong J, Rivas M a, Gonzàlez-Porta M, Kurbatova N, Griebel T, Ferreira PG, Barann M, Wieland T, Greger L, van Iterson M, Almlöf J, Ribeca P, Pulyakhina I, Esser D, Giger T, Tikhonov A, Sultan M, Bertier G, MacArthur DG, Lek M, Lizano E, Buermans HPJ, Padioleau I, Schwarzmayr T, Karlberg O, Ongen H, Kilpinen H, Beltran S, Gut M, Kahlem K, Amstislavskiy V, Stegle O,

Pirinen M, Montgomery SB, Donnelly P, McCarthy MI, Flicek P, Strom TM, Lehrach H, Schreiber S, Sudbrak R, Carracedo A, Antonarakis SE, Häsler R, Syvänen A-C, van Ommen G-J, Brazma A, Meitinger T, Rosenstiel P, Guigó R, Gut IG, Estivill X, Dermitzakis ET. Transcriptome and genome sequencing uncovers functional variation in humans. *Nature*. 2013;501(7468):506-511. doi:10.1038/nature12531.

16. Sander JD, Joung JK. CRISPR-Cas systems for editing, regulating and targeting genomes. *Nat Biotechnol*. 2014;32(4):347-355. doi:10.1038/nbt.2842.

17. Erikson GA, Bodian DL, Rueda M, Molparia B, Scott ER, Scott-Van Zeeland AA, Topol SE, Wineinger NE, Niederhuber JE, Topol EJ, Torkamani A. Whole-Genome Sequencing of a Healthy Aging Cohort. *Cell*. 2016;165(4):1002-1011. doi:10.1016/j.cell.2016.03.022.

18. Yang Q, Cogswell ME, Flanders WD, Hong Y, Zhang Z, Loustalot F, Gillespie C, Merritt R, Hu FB. Trends in Cardiovascular Health Metrics and Associations With All-Cause and CVD Mortality Among US Adults. *JAMA J Am Med Assoc*. 2012;307(12):1273. doi:10.1001/jama.2012.339.

19. Ripatti S, Tikkanen E, Orho-Melander M, Havulinna AS, Silander K, Sharma A, Guiducci C, Perola M, Jula A, Sinisalo J, Lokki M-L, Nieminen MS, Melander O, Salomaa V, Peltonen L, Kathiresan S. A multilocus genetic risk score for coronary heart disease: case-control and prospective cohort analyses. *Lancet*. 2010;376(9750):1393-1400. doi:10.1016/S0140-6736(10)61267-6.

20. Nikkola E, Laiwalla A, Ko A, Alvarez M, Connolly M, Ooi YC, Hsu W, Bui A, Pajukanta P, Gonzalez NR. Remote Ischemic Conditioning Alters Methylation and Expression of Cell Cycle Genes in Aneurysmal Subarachnoid Hemorrhage. *Stroke*. 2015;46(9):2445-2451. doi:10.1161/STROKEAHA.115.009618.

21. Li S, Todor A, Luo R. Blood transcriptomics and metabolomics for personalized medicine. *Comput Struct Biotechnol J*. 2016;14:1-7. doi:10.1016/j.csbj.2015.10.005.

22. Haas BE, Horvath S, Pietiläinen KH, Cantor RM, Nikkola E, Weissglas-Volkov D, Rissanen A, Civelek M, Cruz-Bautista I, Riba L, Kuusisto J, Kaprio J, Tusie-Luna T, Laakso M, Aguilar-Salinas CA, Pajukanta P. Adipose co-expression networks across Finns and Mexicans identify novel triglyceride-associated genes. *BMC Med Genomics*. 2012;5(1):61. doi:10.1186/1755-8794-5-61.

23. Jensen RV, Støttrup NB, Kristiansen SB, Bøtker HE. Release of a humoral circulating cardioprotective factor by remote ischemic preconditioning is dependent on preserved neural pathways in diabetic patients. *Basic Res Cardiol*. 2012;107(5). doi:10.1007/s00395-012-0285-1.

24. Kumar V, Westra HJ, Karjalainen J, Zhernakova D V., Esko T, Hrdlickova B, Almeida R, Zhernakova A, Reinmaa E, Võsa U, Hofker MH, Fehrmann RSN, Fu J, Withoff S, Metspalu A, Franke L, Wijmenga C. Human Disease-Associated Genetic Variation Impacts Large Intergenic Non-Coding RNA Expression. *PLoS Genet*. 2013;9(1). doi:10.1371/journal.pgen.1003201.

25. Amendola LM, Jarvik GP, Leo MC, McLaughlin HM, Akkari Y, Amaral MD, Berg JS, Biswas S, Bowling KM, Conlin LK, Cooper GM, Dorschner MO, Dulik MC, Ghazani AA, Ghosh R, Green RC, Hart R, Horton C, Johnston JJ, Lebo MS, Milosavljevic A, Ou J, Pak CM, Patel RY, Punj S, Richards CS, Salama J, Strande NT, Yang Y, Plon SE, Biesecker

LG, Rehm HL. Performance of ACMG-AMP Variant-Interpretation Guidelines among Nine Laboratories in the Clinical Sequencing Exploratory Research Consortium. *Am J Hum Genet*. 2016;98(6):1067-1076. doi:10.1016/j.ajhg.2016.03.024.

26. Bowdin S, Gilbert A, Bedoukian E, Carew C, Adam MP, Belmont J, Bernhardt B, Biesecker L, Bjornsson HT, Blitzer M, D'Alessandro LCA, Deardorff MA, Demmer L, Elliott A, Feldman GL, Glass IA, Herman G, Hindorff L, Hisama F, Hudgins L, Innes AM, Jackson L, Jarvik G, Kim R, Korf B, Ledbetter DH, Li M, Liston E, Marshall C, Medne L, Meyn MS, Monfared N, Morton C, Mulvihill JJ, Plon SE, Rehm H, Roberts A, Shuman C, Spinner NB, Stavropoulos DJ, Valverde K, Waggoner DJ, Wilkens A, Cohn RD, Krantz ID. Recommendations for the integration of genomics into clinical practice. *Genet Med*. 2016;(January):1-10. doi:10.1038/gim.2016.17.

Appendix I

Exome sequencing identifies 2 rare variants for low high-density lipoprotein cholesterol in an extended family

Exome Sequencing Identifies 2 Rare Variants for Low High-Density Lipoprotein Cholesterol in an Extended Family

M.V. Prasad Linga Reddy, PhD*; Iulia Iatan, MSc*; Daphna Weissglas-Volkov, PhD; Elina Nikkola, BSc; Blake E. Haas, PhD; Miina Juvonen; Isabelle Ruel, PhD; Janet S. Sinsheimer, PhD; Jacques Genest, MD, PhD; Päivi Pajukanta, MD, PhD

Background—Exome sequencing is a recently implemented method to discover rare mutations for Mendelian disorders. Less is known about its feasibility to identify genes for complex traits. We used exome sequencing to search for rare variants responsible for a complex trait, low levels of serum high-density lipoprotein cholesterol (HDL-C).

Methods and Results—We conducted exome sequencing in a large French-Canadian family with 75 subjects available for study, of which 27 had HDL-C values less than the fifth age-sex-specific population percentile. We captured ≈50 Mb of exonic and transcribed sequences of 3 closely related family members with HDL-C levels less than the fifth age-sex percentile and sequenced the captured DNA. Approximately 82 000 variants were detected in each individual, of which 41 rare nonsynonymous variants were shared by the sequenced affected individuals after filtering steps. Two rare nonsynonymous variants in the ATP-binding cassette, subfamily A (ABC1), member 1 (ABCA1), and lipoprotein lipase genes predicted to be damaging were investigated for cosegregation with the low HDL-C trait in the entire extended family. The carriers of either variant had low HDL-C levels, and the individuals carrying both variants had the lowest HDL-C values. Interestingly, the ABCA1 variant exhibited a sex effect which was first functionally identified, and, subsequently, statistically demonstrated using additional French-Canadian families with ABCA1 mutations.

Conclusions—This complex combination of 2 rare variants causing low HDL-C in the extended family would not have been identified using traditional linkage analysis, emphasizing the need for exome sequencing of complex lipid traits in unexplained familial cases. **(*Circ Cardiovasc Genet*. 2012;5:538-546.)**

Key Words: genetics ■ HDL cholesterol ■ exome sequencing ■ rare variants

Low high-density lipoprotein cholesterol (HDL-C) is the most common lipoprotein abnormality and established risk factor of coronary heart disease (CHD). Low HDL-C is caused by multiple genetic factors, common and rare, interacting with one another and with the environment and behavior. In the last 2 decades, significant effort has been devoted to the identification of low HDL-C susceptibility genes. This was initially done using the genome-wide linkage analysis.[1,2] However, progress in identification of the actual disease genes was very slow despite the discovery of many linked intervals. More recently, genome-wide association studies (GWAS) have successfully identified multiple common variants associated with decreased levels of HDL-C.[3]

Clinical Perspective on p 546

However, the sum of common variants identified so far through GWAS explains only a small fraction (10%–15%) of the variance in the HDL-C levels.[3] Hence, it has become evident that other types of DNA variants must contribute substantially to HDL-C levels as well. To identify new rare and low-frequency variants underlying low HDL-C, massive parallel sequencing technologies can be used. The whole-genome sequencing is the most complete approach, but it remains significantly more expensive than exome sequencing that only analyzes coding and transcribed regions, which constitutes <5% of the whole genome sequence.[4] It is estimated that the protein-coding regions of the human genome constitute about 85% of the disease-causing mutations.[4]

We used whole-exome sequencing to search for rare variants conferring susceptibility to low HDL-C. We sequenced the exomes of closely related family members with low HDL-C from a large multigenerational French-Canadian family with 75 subjects available for study and followed up the candidate variants by examining the cooccurrence patterns in the entire extended family.

Methods

Study Samples

The study sample consists of a large multigenerational French-Canadian family collected in the Cardiovascular Genetics Laboratory,

Received January 4, 2012; accepted July 23, 2012.

From the Departments of Human Genetics (M.V.P.L.R., W-V., B.E.H., M.J., J.S.S., P.P.), Biomathematics (J.S.S.), David Geffen School of Medicine at UCLA, Los Angeles, CA; and Division of Cardiology, McGill University Health Center, Royal Victoria Hospital, Montréal, Canada (I.I., I.R., J.G.).

*These authors contributed equally to this article.

The online-only Data Supplement is available at http://circgenetics.ahajournals.org/lookup/suppl/doi:10.1161/CIRCGENETICS.112.963264/-/DC1.

Correspondence to Päivi Pajukanta, MD, PhD, Department of Human Genetics, UCLA, Gonda Center, Room 6335B, 695 Charles E. Young Drive South, Los Angeles, CA 90095–7088. E-mail ppajukanta@mednet.ucla.edu

© 2012 American Heart Association, Inc.

***Circ Cardiovasc Genet* is available at http://circgenetics.ahajournals.org** **DOI: 10.1161/CIRCGENETICS.112.963264**

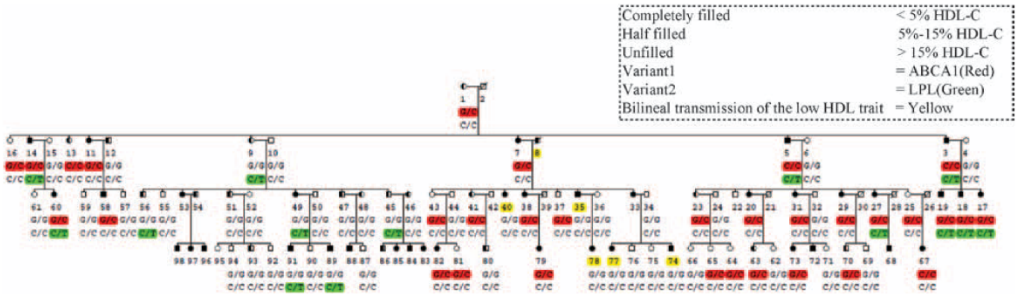


Figure 1. ABCA1 and LPL variants cosegregate with low HDL-C in the multigenerational, French-Canadian, low HDL-C family with 75 (35 males and 40 females) genotyped family members. All of the affected subjects who have HDL-C less than the fifth age-sex specific population percentile comprise risk alleles for either 1 of the 2 variants or both, except, in 1 separate branch, the low HDL-C traits appears to be inherited from the affected spouse's side (indicated in yellow). The subjects with both variants have a lower HDL-C than the subjects with only 1 variant. The subjects whose samples were exome sequenced are indicated by an arrow. HDL-C indicates high-density lipoprotein cholesterol; LPL, lipoprotein lipase; and ABCA1, ATP-binding cassette, subfamily A (ABC1), member 1.

McGill University Health Centre, Royal Victoria Hospital, Montreal, Canada, as described previously.[5] There are 75 family members (35 males and 40 females) with both DNA and extensive demographic and clinical phenotype information available for study in this family. We selected 3 closely related family members with HDL-C levels less than the fifth age-sex percentile from the uppermost generations (Figure 1) for exome sequencing to focus on most severe cases and avoid genetic heterogeneity typical for complex lipid traits.

For a gene–sex interaction analysis, 10 additional French-Canadian families with previously identified mutations in ABCA1[6–8] comprising 125 individuals were also included in the study. The affection status in all families was determined using the fifth age-sex specific population percentile of HDL-C.[5] Family members were sampled (blood collection for lipoprotein analyses, DNA isolation for genetic studies, and skin biopsy for culture of skin fibroblasts used in cellular cholesterol efflux assays) after a 12-hour fast and discontinuation of lipid modifying medications for >4 weeks. Lipids and lipoproteins were measured using standardized techniques as described previously.[6,9] The research protocol was approved by the Research Ethics Board of the McGill University Health Center, and all subjects gave informed consent.

Library Construction and Sequencing

Library construction was performed using 3 µg of genomic DNA and Agilent SureSelect All Exon Kit (50-Mb design) according to the manufacturer's instructions. Further details of library construction and sequencing are given in the online-only Data Supplement Methods.

Data Analysis

Exome Sequencing

We converted the qseq files into a Sanger-formatted FASTQ files that were aligned to a reference sequence (hg19) using the default options of the Burroughs-Wheeler Aligner.[10] Duplicates were removed and a pileup file was generated using SAM tools.[11] The pileup file was used to run the quality control metrics including: a minimum read depth of 4, a maximum read depth of 600, a maximum of 2 single nucleotide polymorphisms (SNPs) per a window size of 10 bases, and a minimum indel score of 25 for filtering nearby SNPs and Phred quality >40. The BED file supplied by Agilent was used to filter only those reads corresponding to the 50 Mb targets.

Annovar was used for functional annotation, dividing the variants into coding and noncoding variants.[12] The coding variants were further divided into synonymous, nonsynonymous (missense), and stop gain or stop loss variants. The synonymous variants were subsequently discarded because they are less likely to be causal. The variants were filtered against the variants present in the HapMap,[13] The 1000 Genomes Project,[14] and dbSNP132[14] databases. Along with novel variants, we selected known rare variants with a minor allele frequency <5%. These variants were classified into damaging and benign based on their predicted protein effect using PolyPhen[15] and SIFT.[16]

Parametric Linkage and Association Analysis

Two-point parametric linkage analysis was performed in the extended family using the Location-Score option of the Mendel software[17] as described in detail in the online-only Data Supplement Methods. Association analysis was performed using a measured genotype approach utilizing the Polygenic-QTL option of Mendel,[18] using continuous HDL-C levels with age and sex as covariates and allele counts of either the ABCA1 variant, lipoprotein lipase (LPL) variant, both variants or none (ie, null model). The heritability and variance explained were calculated as the percent change in total and genetic variance between the null model and the models, including the genotypes as covariates. The LPL variant was further tested for association with log transformed triglyceride (TG) values in a similar fashion.

Genotype by Sex Interaction

We included the extended family together with 10 additional families with previously identified mutations in ABCA1[6–8] in a gene–sex interaction analysis, comprising 200 individuals and 9 different mutations in ABCA1 (DelED1893, G616V, K776N, N1800H, Q2210H, R1851X, R2084X, R909X, and S1731C). Genotype by sex interaction was tested by the SOLAR program[19] using variance-component analysis for discrete traits. We compared models with and without the gene–sex interaction term while keeping the ABCA1 genotypes in both the null and interaction model. We assumed a dominant genetic inheritance, classifying carriers of a mutation as 1 and 0 otherwise, and a multiplicative interaction term, multiplying the genotype score by sex (men=1 and women=0). We also coded a sex-interaction term in which men and postmenopausal women (≥50 years of age) were coded as 1 and premenopausal women (<50 years of age) were coded as 0.

Table 1. Summary of Reads Mapped to the Human Reference Genome (hg19)

Samples	L Range*	H Range†	Average
Total reads	87816621	103315979	95566300
Mapped reads	80911740(82.45%)	85185325(92.14%)	83048533(87.29%)
Duplicate reads	24513472(27.91%)	34633913(33.52%)	29573693(30.72%)
Unmapped reads	6904881(7.86%)	18130654(17.55%)	12517768(12.70%)
Uniquely mapped reads	50551412(48.93%)	56398268(64.22%)	53474840(56.58%)

*The lowest range limit of the total reads.
†The highest range limit of the total reads.

Subjects with HDL-C levels less than the age-sex specific 10th percentiles were classified as affected and subjects with HDL-C levels more than the age-sex specific 20th percentiles as unaffected. *P* values were generated by comparing the 2 models using a likelihood ratio statistic with 1 degree of freedom. Because the affection status is adjusted for sex, the inclusion of the main effect of sex in the model was no longer necessary. The binary HDL-C affection was tested because the variance of HDL-C levels in these ascertained families is reduced and thus limited for effective quantitative analysis.[1]

Cell Culture

Human skin fibroblasts were obtained from 3.0-mm punch biopsies of the forearm of a healthy control subject and the affected proband homozygous for the ABCA1 S1731C variant. The fibroblasts were cultured in Dulbecco's modified Eagle's medium (DMEM) supplemented with 0.1% nonessential amino acids, penicillin (100 U/mL), streptomycin (100 μg/mL), and 10% fetal bovine serum.

Cellular Cholesterol Efflux Assays

Cholesterol efflux was performed as described previously[20,21] with minor modifications. Further details of cellular cholesterol efflux assay are given in the online-only Data Supplement Methods.

Results

Exome Sequencing

To identify rare genetic variants underlying low HDL-C, we sequenced the entire exomes (≈50 Mb) of 3 family members with HDL-C less than the fifth age-sex specific population percentile from a large multigenerational French-Canadian family. The 3 sequenced family members were closely related: 1 affected, his sibling, and his child were sequenced. Exome capture and sequencing were performed using the Agilent Sure Select in-solution method and Illumina Hiseq2000 platform as described in the Methods. We obtained an average of 90 million reads per person and successfully mapped ≈90% of these reads to the reference sequence (Table 1). After quality control the mean coverage was 50X.

Filtering of Identified Variants

On average, 82 000 single-nucleotide variants were detected in each individual. We focused on variants shared by all 3 exome sequenced subjects and filtered the variants based on their type, frequency, and functional predictions. Filtering for missense and stop gain or stop loss variants that were shared by all 3 affected individuals resulted in 3428 nonsynonymous variants and 31 stop gain or loss variants (Table 2). The transition/transversion (Ti/Tv) ratio of the coding variants was 3.4, whereas the Ti/Tv ratio of the noncoding was 2.5, in good agreement with the expected ratios.[22]

The identified variants were further filtered against variants present in the HapMap,[13] 1000 Genomes Project,[14] and dbSNP132[14] databases, resulting in 332 novel variants and known variants with minor allele frequency <5%. These variants were further filtered by selecting variants predicted to affect protein function using PolyPhen[15] and SIFT[16] and expressed in a relevant tissue including liver, adipose, and heart, resulting in 41 shared potentially functional variants that were either novel or known but relatively rare (Table 2). Among the shared variants there were 2 rare functional variants in the ABCA1 and LPL genes that are excellent susceptibility candidates as their key role in HDL-C metabolism is well established.[23] We confirmed their presence by both Sanger sequencing and genotyping.

Table 2. Number of Variants Shared by the 3 Sequenced Affected Family Members After a Series of Filtering Steps

Variant Filter	Variants	Nonsynonymous	Stop Gain or Loss	Total
Shared by all 3*	Known†	3389	30	3419
	Novel	39	1	40
	Total	3428	31	3459
MAF<5% and new‡	Known	293	3	296
	Novel	39	1	40
	Total	332	4	335
Damaging§	Known	52	3	55
	Novel	24	1	25
	Total	76	4	80
Functional#	Known	18	3	21
	Novel	19	1	20
	Total	37	4	41

MAF indicates minor allele frequency; GWAS, genome-wide association studies.
*Nonsynonymous and stop gain or loss variants shared by all 3 sequenced individuals.
†Present in the HapMap, The 1000 Genomes Project and/or dbSNP132 databases.
‡Novel variants (not present in the HapMap, The 1000 Genomes Project and/or dbSNP132 databases) and known rare variants with a MAF<5%.
§Novel and known rare (MAF<5%) variants predicted to be damaging either by PolyPhen and/or SIFT.
#The novel and known rare (MAF<5%) variants located in genes functionally relevant to lipids; or in genes expressed in relevant tissue; or present within 500 kb of the 95 known lipid GWAS loci.[9]

Table 3. The Lipid Levels and Other Clinical Characteristics of the 3 Individuals That Were Exome Sequenced

IND ID	ABCA1	LPL	TC	TG	HDL-C	HDL %	BMI	LDL-C	AGE	SEX
Ind5	C/C	C/T	4.61	1.3	0.67	<5	22.02	3.36	66	Male
Ind14	G/C	C/T	4.6	6.1	0.62	<5	25.81	NA	55	Male
Ind19	G/C	C/T	2.43	2.4	0.62	<5	27.22	0.7	21	Male

LPL indicates lipoprotein lipase; TG, triglyceride; HDL-C, high-density lipoprotein cholesterol; BMI, body mass index; LDL, low-density lipoprotein; and ABCA1, ATP-binding cassette, subfamily A (ABC1), member 1 and TC, total cholesterol.

The lipid levels are shown in millimoles per liter.

Rare Missense Variant in ABCA1

The ABCA1 (S1731C) variant is not present in dbSNP132[14] or The 1000 Genomes Project[14] data, and is located in exon 38. This previously reported rare variant is changing a conserved amino acid from serine to cysteine, and is known to result in decreased cholesterol efflux.[7,24–26]

To further determine the effect of the S1731C variant on cholesterol efflux, we used human fibroblasts from the affected proband homozygous for the S1731C variant, and compared these cells to a normal control. Assays were performed in 22OH/9CRA stimulated fibroblasts (to induce ABCA1 expression), and unstimulated cells, in the presence or absence of lipid free ApoA-I (Figure 2A). We observed a significant decrease (≈40%) in apoA-I-mediated cellular cholesterol efflux in the proband, compared with the control without the ABCA1 variant (P=1.23×10–^{4} using Student t-test and P=0.0495 using a nonparametric 2-sample Wilcoxon rank sum test). These results are in agreement with previously documented findings.[7,25,26] Low efflux levels were also observed in unstimulated cells, presumably due to basal levels of ABCA1 expression and the presence of other apoA-I binding sites at the cell surface. Also, as expected, background basal conditions of passive diffusion of cellular cholesterol were not affected by mutations at the ABCA1 gene locus.

Because the lipid levels of the ABCA1 S1731C variant carriers suggested a possible gene-sex effect (Tables 3 and 4), we further investigated whether exposure to 17β-estradiol steroid hormone endogenously expressed in females, possibly

Table 4. The Mean Lipid Levels and Other Clinical Characteristics of Individuals With Different ABCA1 (S1731C) and LPL (P234L) Genotypes

	All Individuals	TC	TG	HDL-C	AGE	BMI	LDL-C
ABCA1/LPL							
CC/CC	2	3.76(1.29)	0.72(0.01)	0.94(0.30)	35(33.94)	19.45(7.35)	1.67(0.71)
CC/CT	2	3.60(1.44)	3.81(3.61)	0.52(0.22)	61(7.07)	23.17(1.63)	2.19(1.65)
GC/CC	22	5.11(1.84)	1.22(0.69)	0.95(0.21)	37(20.04)	20.65(3.27)	3.53(1.92)
GC/CT	6	4.13(1.25)	2.97(1.99)	0.68(0.07)	34.3(11.36)	24.36(3.18)	2.03(1.90)
GG/CC	37	4.88(1.24)	1.60(0.90)	1.16(0.35)	34.6(17.03)	22.64(5.67)	2.94(1.13)
GG/CT	6	3.91(1.21)	2.18(1.37)	0.78(0.12)	35(17.37)	22.07(3.38)	2.06(0.87)
Total	75	4.74(1.47)	1.68(1.21)	1.01(0.33)	36(17.86)	22.12(4.75)	2.95(1.51)
ABCA1/LPL	All Males	TC	TG	HDL-C	AGE	BMI	LDL-C
CC/CC	0						
CC/CT	2	3.59(1.44)	3.81(3.61)	0.52(0.22)	61(7.07)	23.17(1.63)	2.19(1.65)
GC/CC	3	4.59(1.59)	0.94(0.45)	0.89(0.21)	30(17.44)	22.08(3.03)	3.28(1.56)
GC/CT	3	3.59(1.09)	4.43(1.84)	0.65(0.06)	35(17.78)	25.28(2.25)	0.35(0.49)
GG/CC	22	4.73(1.35)	1.45(0.89)	1.06(0.26)	34.3(17.44)	22.88(3.62)	2.97(1.31)
GG/CT	4	4.02(1.54)	2.62(1.51)	0.73(0.08)	26.2(11.67)	23.39(3.30)	2.04(1.08)
Total	34	4.47(1.36)	1.94(1.55)	0.94(0.29)	34.6(17.25)	23.12(3.30)	2.68(1.41)
ABCA1/LPL	All Females	TC	TG	HDL-C	AGE	BMI	LDL-C
CC/CC	2	3.76(1.29)	0.72(0.01)	0.94(0.30)	35(33.94)	19.46(7.36)	1.67(0.00)
CC/CT	0						
GC/CC	19	5.19(1.91)	1.27(0.72	0.96(0.22)	38.1(20.63)	20.49(3.33)	3.57(2.01)
GC/CT	3	4.67(1.36)	1.51(0.33)	0.71(0.09)	33.7(2.31)	23.46(4.22)	3.15(1.54)
GG/CC	15	5.09(1.06)	1.83(0.89)	1.30(0.43)	35.1(17.00)	22.27(7.92)	2.89(0.85)
GG/CT	2	3.69(0.23)	1.32(0.52)	0.88(0.16)	52.5(13.44)	19.46(1.92)	2.09(0.49)
Total	41	4.97(1.53)	1.47(0.78)	1.06(0.36)	37.2(18.47)	21.29(5.58)	3.17(1.57)

LPL indicates lipoprotein lipase; TG, triglyceride; HDL-C, high-density lipoprotein cholesterol; BMI, body mass index; LDL, low-density lipoprotein; and ABCA1, ATP-binding cassette, subfamily A (ABC1), member 1 and TC, total cholesterol.

*The value in the parenthesis indicates standard deviation.

corrects the cholesterol efflux defect in fibroblasts from the S1731C male ABCA1 carrier during the 22OH/9CRA ABCA1 stimulation phase of 17 hours (Figure 2B). Interestingly, after adjusting for basal cholesterol diffusion, we observed that upon treatment with elevated doses of estradiol (>20 nM), efflux in the S1731C proband significantly increased (P=7.2×10–^{6}, r=0.78 using a nonparametric Spearman trend test), whereas that in the wild-type control remained constant (P=0.2, r=0.25) (Figure 2B). Taken together, these results support a genotype–sex interaction effect, as hormonal regulation with 17β-estradiol partially restored the low efflux observed in the S1731C male proband but had no significant effect on the efflux of a wild-type control.

Rare Missense Variant in LPL

The identified LPL variant rs118204060 is present in the dbSNP132[14] and The 1000 Genomes Project[14] data with an unknown frequency. The rs118204060 located in exon 5 changes a conserved amino acid from proline to leucine (P234L). This variant was initially identified in familial chylomicronemia and was reported as P207L[27–29] because of the differences in genome builds. Upon sequence comparisons, we confirmed that they are indeed the same variant.

Investigation of the ABCA1 and LPL Variants in the Entire Family

We examined the pedigree members for cooccurrence of nonsynonymous ABCA1 and LPL variants with low HDL-C. By stratifying individuals by their HDL percentiles, we can see that all the affected family members with HDL-C less than the fifth percentile carry a risk allele for either 1 or both of the variants (3 P234L, 11 S1731C, and 8 P234L/S1731C), except in 1 separate branch of the extended family in which the low HDL-C traits appears to be inherited from the affected spouse's side (Figure 1). No family member with an HDL-C value greater than the fifth percentile had both LPL and ABCA1 variants. Furthermore, no family member with HDL-C values greater than the 15th percentile had the LPL variant. Seven subjects (1 male and 6 females) had the ABCA1 variant with the HDL-C percentile of 22% for the male and with an average HDL-percentile of 35% for the 6 females (Figure 1). Two of the 3 exome sequenced subjects were heterozygous for both variants and 1 was homozygous for the ABCA1 variant and heterozygous for the LPL variant. There were 4 homozygous subjects for the ABCA1 variant, 2 of which were also heterozygous for the LPL variant, whereas the LPL variant was heterozygous in all 14 family members it was observed (Figure 1). Thus, a heterozygous, milder form of LPL deficiency exists in this family. Accordingly, the LPL variant P234L is also associated with elevated levels of TGs (1.65±0.27 mmol/L, P=6.14×10–^{3}). In addition, we observed that the subjects with both variants have a lower HDL-C than the subjects with only 1 variant, and that the subjects heterozygous or homozygous for the ABCA1 variant do not differ in the HDL-C levels (Table 4).

Explained Variance and Heritability

We estimated that the effect of the ABCA1 and LPL variants on continuous HDL-C measurements in the extended family is −0.17±0.08 mmol/L (P=0.025) and −0.27±0.09 mmol/L (P=0.006), respectively. Together, these 2 variants explain 60% of the genetic variance in this family and 26% of the total (genetic + environment) variance in this family, which amounts to 46% of the heritability explained as assessed in a measured genotype analysis.[18] We also repeated the analysis while excluding the 3 affected subjects that were exome sequenced to reduce the potential for ascertainment bias. In this analysis, the effect sizes of ABCA1 and LPL remained the same (−0.18±0.08 and −0.27±0.10 mmol/L, respectively), and the additive and total variance explained were 50% and 24%, respectively, with 34% of the heritably explained. Importantly, if the subfamily with the bilineal introduction of the low HDL-C trait through the affected spouse is excluded from these analyses, virtually all of the additive variance of HDL-C and virtually all of the heritability of HDL-C is explained by the ABCA1 and LPL variants, suggesting that the ABCA1 and LPL variants can explain the low HDL-C in the non-bilineal part of the extended family.

Genome-Wide Linkage Analyses

To further investigate that we did not miss a major susceptibility variant, we performed a whole-genome 2-point linkage analysis for low HDL-C using a dominant mode of inheritance. We first estimated using the SLINK simulation program[30] that under the assumption of homogeneity, the maximum lod score this family can provide is 4.34. However, none of the actual 553 microsatellite markers reached this lod score, most probably due to the existence of multiple low HDL-C variants in the family (ie, heterogeneity). In more detail, no lod scores >3 were observed anywhere in the genome. The only lod score >2.0 was observed on chromosome 21 for marker D21S1255. However, we noticed that this signal on chr 21 seems to arise from the bilineal branch (Figure 1), as the signal diminishes to lod score of 0.6 when we excluded this subfamily from the analysis and increases to 2.5 when we analyzed this bilineal branch of the family alone. Hence the genome-wide linkage data suggest that there might be another susceptibility variant on chr 21 that accounts for the low HDL-C in the bilineal subfamily branch of the extended family. However, because we did not sequence any family members from this branch, none of the 3459 filtered-out variants would be good candidates. Importantly, we observed lod scores >1 near the LPL and ABCA1 genes (lod scores of 1.62 and 1.28 10.3 Mb and 5.8 Mb from LPL and ABCA1, respectively). Without the bilineal branch these lod scores increased to 2.14 and 1.45, respectively.

Genotype by Sex Interaction

The effect of the ABCA1 S1731C variant on low HDL-C levels appears more profound in the males than in females in the extended family (Table 4). Furthermore, our efflux study also suggested a gene × sex interaction (Figure 2B). Although the frequency of the S1731C variant may be individually too rare for testing genetic interactions (as large sample sizes are necessary), rare variants with large phenotypic effects are collectively common in low HDL-C families.[26] We hypothesized that the apparent sex effect may not be restricted to the S1731C allele, but rather it may generally extend to ABCA1

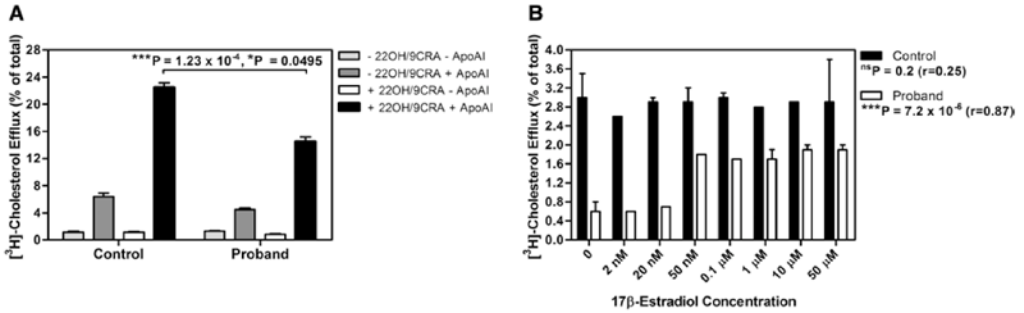


Figure 2. A, Effect of the ABCA1 variant on cholesterol efflux in fibroblasts from a proband homozygous for S1731C and a healthy control. Fibroblasts were isolated by taking a biopsy from the forearm of the proband and a healthy individual, plated in 12-well plates and radiolabeled with [^{3}H]-cholesterol for 48 hours. Cholesterol efflux was performed as described in Methods under background diffusion conditions (–22OH -ApoA-I), unstimulated (–22OH +ApoA-I) and stimulated (+22OH -ApoA-I, +22OH +ApoA-I) conditions, with or without ApoA-I. The proband had a significantly reduced ApoA-I-mediated efflux compared with the control without the variant. Values represent the mean±SD. from triplicate wells. Results shown are a representative of 3 independent experiments. ***P=1.23×10–^{4} by Student t test; and *P=0.0495 using a nonparametric 2-sample Wilcoxon rank sum test. 22OH indicates 22(R)-hydroxycholesterol; 9CRA, 9-*cis*-retinoic acid; and ApoA-I, apolipoprotein A-I. **B**, Elevated concentrations of 17β-estradiol improve cholesterol efflux in the male proband with the ABCA1 S1731C variant. Fibroblasts from a male proband with the ABCA1 S1731C variant and a healthy male control were isolated by taking a biopsy from the forearm of plated in 24-well plates and radiolabeled with [3H]-cholesterol for 24 hours. Cholesterol efflux was performed as described in Methods, with addition of increasing concentrations (2 nM, 20 nM, 50 nM, 0.1 µM, 1 µM, 10 µM, and 50 µM) of 17β-estradiol while stimulating ABCA1 expression with 22OH/9CR for 17 hours. As in "A", experiments were done under 4 conditions (–22OH -ApoA-I, -22OH +ApoA, +22OH -ApoA-I, +22OH +ApoA-I). Efflux results were subsequently adjusted for background basal conditions of passive diffusion. The final stimulated ApoA-I mediated efflux condition is shown. Upon exposure to increasing estradiol concentrations (>20 nM), cholesterol efflux in the S1731C proband significantly increases. Of note, the overall [3H]-cholesterol efflux counts were lesser in magnitude than those observed in "A" given the shorter labeling time period (24 hours). In addition, the difference in efflux between the control and proband was greater than in "A", given the different basal diffusion of the selected conditions (data not shown), which was now removed from the net ApoAI-mediated efflux. Values represent the mean ±SD. from triplicate wells. Results shown are representative of 3 independent experiments. 22OH indicates 22(R)-hydroxycholesterol; 9CRA, 9-*cis*-retinoic acid; and ApoA-I, apolipoprotein A-I. ***P=7.2×10–^{6} (r=0.78) using a nonparametric Spearman trend test for the dose effect on efflux in the S1731C proband; and ns$P$=0.2 (r=0.25) using a nonparametric Spearman trend test for the dose effect on efflux in the wild-type control.

alleles with major phenotypic effects. Thus, to further investigate this intriguing relationship between ABCA1 and sex, we examined the collective effect of multiple rare variants in ABCA1 by sex on HDL-C affection. All in all, 10 additional low HDL-C French-Canadian families with known mutations in ABCA1[6–8] were included in the sex-interaction analysis using the SOLAR program,[19] comprising to a total of 93 males and 107 females. The percentage of mutation carriers was 42% and 53% in males and females, respectively (Figure 3). The S1731C variant was present in 3 of these additional low HDL-C pedigrees,[7] and together with the exome sequenced family, the association signal for the main effect of S1731C on low HDL-C status resulted in a P value of 0.008. In all 11 families we observed, as expected, a highly significant main effect for ABCA1 genotypes (P=1×10–^{09}), as well as a significant ABCA1 genotype × sex interaction on the qualitative HDL-C affection (P=0.03) (Figure 3). Furthermore, the interaction effect appeared to be more pronounced when comparing premenopausal women (aged <50 years) to men and postmenopausal women (P=0.003).

Discussion

By using exome sequencing we identified 2 functional rare variants in the ABCA1 and LPL genes, cosegregating with low HDL-C and explaining a major proportion of the HDL-C variance and heritability in an extended family. We also observed a sex effect for ABCA1 variants, male carriers exhibiting significantly lower HDL-C levels than females. Furthermore, none of the unaffected family members had the LPL variant or both variants. Our study exemplifies how use of exome sequencing was critical to reveal the complex combination of 2 variants of which 1 is less severe in females. Traditional linkage analysis was unable to elucidate this type of complex pattern of variants in this extended family,[31] suggesting that many such combinations have been missed in previous linkage analyses of complex traits.

ABCA1 and LPL are major players of lipid metabolism. The ABCA1 is a key protein involved in reverse cholesterol transport that transports cellular cholesterol to lipid-poor acceptor apolipoproteins, such as apolipoproteinA-I.[20,32] As a result, the apolipoprotein is released with the extracted phospholipid and cholesterol, forming nascent HDL particles. Mutations disrupting the normal function of ABCA1 result in little or no circulating HDL.[33] Previous studies have shown that cell lines with the identified ABCA1 variant S1731C exhibit low levels of protein expression,[25] and that cells transfected with the S1731C allele express abundant ABCA1 mRNA but fail to generate significant amounts of ABCA1 protein.[25] Furthermore, the cholesterol efflux of S1731C has been shown to be reduced to 12.3% to 68.0% of the wild-type.[7,25,26] Here, we observed a ≈40% cholesterol efflux reduction in the proband homozygous for the S1731C variant as compared with a normal control, in line with the earlier findings.[7,25,26] In our previous article,[7] we showed that 3 heterozygous subjects with the S1731C variant have cholesterol efflux values of 63%, 66%, and 68% of the wild type. Thus, about the same 40%

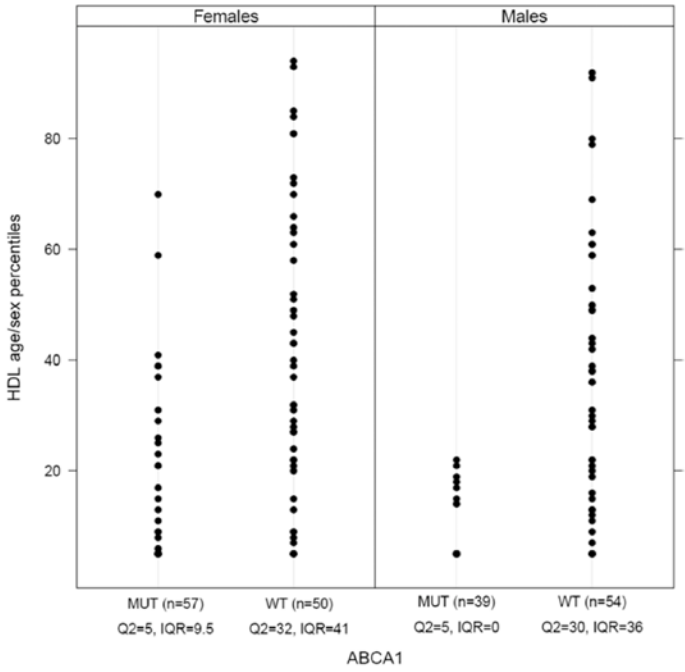


Figure 3. Sex-dependent effect of ABCA1 variants. Figure 3 shows the age–sex specific population HDL-C percentiles by ABCA1 genotypes and sex in 200 French-Canadian family members from 11 French-Canadian families with different ABCA1 mutations (DelED1893, G616V, K776N, N1800H, Q2210H, R1851X, R2084X, R909X, and S1731C). MUT indicates carriers of a mutation in the ABCA1 gene, and WT stands for wild-type genotype (ie, noncarriers). Q2 stands for the median quartile and IQR for interquartile range (Q3–Q1) of HDL-C percentiles. The distributions of the age–sex percentiles are similar between genders in the WT genotype group, whereas in the MUT genotype group, the distribution is more restricted to the lower tail in males than in females. This difference describes the significant result of the genotype–sex interaction analysis. It should be noted that the displayed age–sex percentiles are not adjusted for relatedness, whereas the family relation was taken into account in the genotype–sex interaction analysis performed using SOLAR. HDL-C indicates high-density lipoprotein cholesterol.

decrease is observed in the heterozygous subjects as in the homozygous subject with the S173C variant, in line with their similar HDL-C values (Table 4). Although the phenotype data suggest that the S1731C variant does not have a gene dose effect, this conclusion warrants additional functional studies in future, because there are only 4 homozygotes in the family 2 of which are also heterozygous for the LPL variant, and furthermore the variant has a large range (12.3%–68.0% of the wild-type) in its effect on the cholesterol efflux.[7,25,26]

The main function of LPL is to hydrolyze TGs to deliver fatty acids to the tissue. LPL also hydrolyzes very-low-density lipoproteins. Sequence variation in LPL has been reported to be associated with the risk of coronary heart disease, TGs, and HDL-C.[3] An efficient LPL function is associated with lower TG and low-density lipoproteins and higher HDL. Regarding the identified P207L variant, individuals with this mutation have reduced HDL particles compared with the control subjects.[34] Previous studies have also shown that missense mutations in exon 5 of the LPL gene where the P207L variant resides are the most common cause of LPL deficiency.[35,36] Importantly, Ma et al reported that upon site-directed in vitro mutagenesis this variant produces a catalytically inactive LPL protein, which is the cause of the LPL deficiency in the patients.[29] Taken together, these previous data, along with our PolyPhen[15] and SIFT[16] predictions, show that it is highly likely that both identified variants S1731C and P207L affect protein function.

The 2 identified variants, S1731C and P207L, have been reported previously in French-Canadian dyslipidemic individuals but not in normal controls.[7,24–29] The S1731C ABCA1 variant was present in 3 French-Canadian dyslipidemic families with low HDL-C levels,[7] but not in 528 chromosomes from French-Canadian subjects with normal HDL-C levels.[24] It was also absent in 108 French-Canadian subjects with high HDL-C.[26] The P207L LPL variant was previously observed in 37 unrelated French-Canadian patients with LPL deficiency with 54 mutant alleles present in that study sample.[29] In the same study, the variant was also genotyped in 34 unrelated patients with LPL deficiency from ancestries other than French-Canadian. Only 1 German patient was found to be heterozygous for the risk allele. Furthermore, 11 out of 180 French-Canadian hyperlipidemic cases were heterozygous for the P207L variant, whereas none of the 170 normolipidemic controls had the P207L variant.[29]

It is important to study the effect of sex on lipid traits to better understand the sex-specific differences in incidence of dyslipidemia and cardiovascular disease. The results of an earlier study demonstrated that ABCA1 has a sex-specific effect, because elevated levels of ABCA1 were observed in females,[37] which is in line with the higher HDL-C levels and the lower risk of females for coronary artery disease. In this study, we observed that functional mutations in ABCA1 affecting the cholesterol efflux[7,25,26] have a larger effect on HDL-C levels in male than female carriers of these variants. It is possible that the observed genotype–sex interaction results from the previously observed sex differences in ABCA1 expression levels,[37] because if males have lower baseline levels of ABCA1, the effect of the mutations could be even more profound in males. These interesting sex-specific mechanisms of ABCA1 may involve hormonal regulation of ABCA1, a hypothesis supported by our efflux experiment (Figure 2B),

demonstrating that exposure of fibroblasts of a male proband with the ABCA1 S1731C variant to increasing concentrations of 17β-estradiol led to a significantly increased efflux in the male proband with the ABCA1 variant. These intriguing findings warrant further investigation in future studies.

Our results demonstrate that 2 relatively rare functional ABCA1 and LPL variants contribute to the risk of low HDL-C in a unique combination involving a sex-effect in an extended family. We first identified a set of variants by filtering the variants shared by the exome-sequenced affected family members for variant type, frequency, and functional predictions. Because filtering has limitations caused by heterogeneity of complex traits,[4] we then used the extended family structure for statistical analysis exploring how much of the trait variance and heritability the 2 key ABCA1 and LPL variants explain. Thus, our study highlights the fact that the filtering strategy used in exome studies of Mendelian disorders[4] is not directly applicable for complex disorders, and that new methodologies that incorporate multiple susceptibility variants within a family are warranted. Because the 2 variants explain a major part of the variance in HDL-C and are shown to be functional,[7,25,26,29] they represent the key underlying HDL-C variants in this family, though other rare and common variants are likely to explain the remaining portion of the variance. Because both ABCA1 and LPL are known to affect HDL-C, our study did not reveal a novel HDL gene. However, our study does highlight the importance of exome sequencing of dyslipidemic families, because traditional linkage or haplotype analysis cannot detect complex segregation of several functional rare variants due to the inherent parametric restrictions of linkage analysis. This type of underlying biological complexity must have contributed to the low lod scores and weak success of linkage analysis in gene identification of complex lipid traits. In this study, we demonstrate for how family-based exome sequencing can successfully identify multiple rare variants to be followed up utilizing the effective cosegregation information available in extended dyslipidemic families. To the best of our knowledge, our study is the first described example of 2 functional rare variants conferring the susceptibility to low HDL-C in an extended family.

Acknowledgments

We thank the family members who participated in this study. We also thank Cindy Montes and UCLA core facilities for laboratory technical assistance.

Sources of Funding

This research was supported by the grants HL095056 and HL-28481 from the National Institutes of Health (Drs Pajukanta and Sinsheimer); grants MOP 97752 from the Canadian Institutes of Health Research (CIHR) and from the Heart and Stroke Foundation of Canada (Dr Genest); and by the American Heart Association grant 11POST7380028 (Dr Linga Reddy).

Disclosures

None.

References

1. Pajukanta P, Allayee H, Krass KL, Kuraishy A, Soro A, Lilja HE, et al. Combined analysis of genome scans of dutch and finnish families reveals a susceptibility locus for high-density lipoprotein cholesterol on chromosome 16q. *Am J Hum Genet*. 2003;72:903–917.
2. Soro A, Pajukanta P, Lilja HE, Ylitalo K, Hiekkalinna T, Perola M, et al. Genome scans provide evidence for low-HDL-C loci on chromosomes 8q23, 16q24.1-24.2, and 20q13.11 in Finnish families. *Am J Hum Genet*. 2002;70:1333–1340.
3. Teslovich TM, Musunuru K, Smith AV, Edmondson AC, Stylianou IM, Koseki M, et al. Biological, clinical and population relevance of 95 loci for blood lipids. *Nature*. 2010;466:707–713.
4. Ng SB, Buckingham KJ, Lee C, Bigham AW, Tabor HK, Dent KM, et al. Exome sequencing identifies the cause of a Mendelian disorder. *Nat Genet*. 2010;42:30–35.
5. Dastani Z, Quiogue L, Plaisier C, Engert JC, Marcil M, Genest J, et al. Evidence for a gene influencing high-density lipoprotein cholesterol on chromosome 4q31.21. *Arterioscler Thromb Vasc Biol*. 2006;26:392–397.
6. Brooks-Wilson A, Marcil M, Clee SM, Zhang LH, Roomp K, van Dam M, et al. Mutations in ABC1 in Tangier disease and familial high-density lipoprotein deficiency. *Nat Genet*. 1999;22:336–345.
7. Alrasadi K, Ruel IL, Marcil M, Genest J. Functional mutations of the ABCA1 gene in subjects of French-Canadian descent with HDL deficiency. *Atherosclerosis*. 2006;188:281–291.
8. Marcil M, Brooks-Wilson A, Clee SM, Roomp K, Zhang LH, Yu L, et al. Mutations in the ABC1 gene in familial HDL deficiency with defective cholesterol efflux. *Lancet*. 1999;354:1341–1346.
9. Marcil M, Yu L, Krimbou L, Boucher B, Oram JF, Cohn JS, et al Jr. Cellular cholesterol transport and efflux in fibroblasts are abnormal in subjects with familial HDL deficiency. *Arterioscler Thromb Vasc Biol*. 1999;19:159–169.
10. Langmead B, Trapnell C, Pop M, Salzberg SL. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol*. 2009;10:R25.
11. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al; 1000 Genome Project Data Processing Subgroup. The Sequence Alignment/Map format and SAMtools. *Bioinformatics*. 2009;25:2078–2079.
12. Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res*. 2010;38:e164.
13. International HapMap 3 Consortium, Altshuler DM, Gibbs RA, Peltonen L, Altshuler DM, Gibbs RA, et al. Integrating common and rare genetic variation in diverse human populations. *Nature*. 2010;467:52–58.
14. The 1000 Genomes Project Consortium. A map of human genome variation from population-scale sequencing. *Nature*. 2010;467:1061–1073.
15. Ramensky V, Bork P, Sunyaev S. Human non-synonymous SNPs: server and survey. *Nucleic Acids Res*. 2002;30:3894–3900.
16. Ng PC, Henikoff S. SIFT: Predicting amino acid changes that affect protein function. *Nucleic Acids Res*. 2003;31:3812–3814.
17. Lange K, Cantor R, Horvath S, Perola M, Sabatti C, Sinsheimer J, et al. Mendel version 4.0: A complete package for the exact genetic analysis of discrete traits in pedigree and population data sets. *Am J Hum Genet*. 2001;69:504.
18. Lange K, Sinsheimer JS, Sobel E. Association testing with Mendel. *Genet Epidemiol*. 2005;29:36–50.
19. Almasy L, Blangero J. Multipoint quantitative-trait linkage analysis in general pedigrees. *Am J Hum Genet*. 1998;62:1198–1211.
20. Marcil M, Bissonnette R, Vincent J, Krimbou L, Genest J. Cellular phospholipid and cholesterol efflux in high-density lipoprotein deficiency. *Circulation*. 2003;107:1366–1371.
21. Denis M, Haidar B, Marcil M, Bouvier M, Krimbou L, Genest J Jr. Molecular and cellular physiology of apolipoprotein A-I lipidation by the ATP-binding cassette transporter A1 (ABCA1). *J Biol Chem*. 2004;279:7384–7394.
22. Clark MJ, Chen R, Lam HY, Karczewski KJ, Chen R, Euskirchen G, et al. Performance comparison of exome DNA sequencing technologies. *Nat Biotechnol*. 2011;29:908–914.
23. Weissglas-Volkov D, Pajukanta P. Genetic causes of high and low serum HDL-cholesterol. *J Lipid Res*. 2010;51:2032–2057.
24. Clee SM, Zwinderman AH, Engert JC, Zwarts KY, Molhuizen HO, Roomp K, et al. Common genetic variation in ABCA1 is associated with altered lipoprotein levels and a modified risk for coronary artery disease. *Circulation*. 2001;103:1198–1205.
25. Brunham LR, Singaraja RR, Pape TD, Kejariwal A, Thomas PD, Hayden MR. Accurate prediction of the functional significance of single nucleotide polymorphisms and mutations in the ABCA1 gene. *PLoS Genet*. 2005;1:e83.
26. Cohen JC, Kiss RS, Pertsemlidis A, Marcel YL, McPherson R, Hobbs HH. Multiple rare alleles contribute to low plasma levels of HDL cholesterol. *Science*. 2004;305:869–872.

27. Brisson D, Ledoux K, Bossé Y, St-Pierre J, Julien P, Perron P, et al. Effect of apolipoprotein E, peroxisome proliferator-activated receptor alpha and lipoprotein lipase gene mutations on the ability of fenofibrate to improve lipid profiles and reach clinical guideline targets among hypertriglyceridemic patients. *Pharmacogenetics*. 2002;12:313–320.
28. Normand T, Bergeron J, Fernandez-Margallo T, Bharucha A, Ven Murthy MR, Julien P, et al. Geographic distribution and genealogy of mutation 207 of the lipoprotein lipase gene in the French Canadian population of Québec. *Hum Genet*. 1992;89:671–675.
29. Ma Y, Henderson HE, Murthy V, Roederer G, Monsalve MV, Clarke LA, et al. A mutation in the human lipoprotein lipase gene as the most common cause of familial chylomicronemia in French Canadians. *N Engl J Med*. 1991;324:1761–1766.
30. Weeks DE. SLINK: a general simulation program for linkage analysis. *Am J Hum Genet*. 1990;47:204.
31. Dastani Z, Pajukanta P, Marcil M, Rudzicz N, Ruel I, Bailey SD, et al. Fine mapping and association studies of a high-density lipoprotein cholesterol linkage region on chromosome 16 in French-Canadian subjects. *Eur J Hum Genet*. 2010;18:342–347.
32. Oram JF, Lawn RM. ABCA1. The gatekeeper for eliminating excess tissue cholesterol. *J Lipid Res*. 2001;42:1173–1179.
33. Albrecht C, Baynes K, Sardini A, Schepelmann S, Eden ER, Davies SW, et al. Two novel missense mutations in ABCA1 result in altered trafficking and cause severe autosomal recessive HDL deficiency. *Biochim Biophys Acta*. 2004;1689:47–57.
34. Ruel IL, Gaudet D, Perron P, Pascot A, Després JP, Bergeron J, et al. Determinants of HDL particle size in patients with the null (P207L) or defective (D9N) mutation in the lipoprotein lipase gene: the Québec LipD Study. *Atherosclerosis*. 2002;162:269–276.
35. Peterson J, Ayyobi AF, Ma Y, Henderson H, Reina M, Deeb SS, et al. Structural and functional consequences of missense mutations in exon 5 of the lipoprotein lipase gene. *J Lipid Res*. 2002;43:398–406.
36. Garenc C, Couillard C, Laflamme N, Cadelis F, Gagné C, Couture P, et al. Effect of the APOC3 Sst I SNP on fasting triglyceride levels in men heterozygous for the LPL P207L deficiency. *Eur J Hum Genet*. 2005;13:1159–1165.
37. Zhang Y, Klein K, Sugathan A, Nassery N, Dombkowski A, Zanger UM, et al. Transcriptional profiling of human liver identifies sex-biased genes associated with polygenic dyslipidemia and coronary artery disease. *PLoS ONE*. 2011;6:e23506.

CLINICAL PERSPECTIVE

It is important to elucidate the genetic background of low levels of high-density lipoprotein cholesterol (HDL-C), because low serum HDL-C predisposes to coronary artery disease, the major cause of mortality. Recent genome-wide association studies have successfully identified common DNA sequence variants for HDL-C. However, these common variants explain only a small proportion of variation in the HDL-C levels, indicating that currently still undiscovered rare variants contribute to HDL-C levels as well. Although exome sequencing has been successfully applied to discover rare mutations for Mendelian disorders, not much is known about its feasibility to identify genes for complex traits. We hypothesized that exome sequencing can be used to identify rare variants in unexplained familial cases with low HDL-C because the family-based approach allows the examination of the co-occurrence patterns in the entire family. We sequenced the exomes of subjects with low HDL-C from a large multigenerational family and identified 2 rare nonsynonymous variants in ABCA1 and lipoprotein lipase. The family members carrying either variant had low HDL-C levels, and the individuals carrying both variants had the lowest HDL-C values. Our data also demonstrated a sex effect for the ABCA1 variant. This type of complex combination of 2 rare variants causing low HDL-C in an extended family would not have been identified using traditional linkage analysis, emphasizing the conclusion that exome sequencing can effectively discover rare variants in familial unexplained forms of complex lipid traits.

SUPPLEMENTAL MATERIAL

Supplemental Methods

Library construction and sequencing

The genomic DNA was sheared (sonication) with Covaris S2 to achieve a uniform distribution of fragments with a mean size of 200 bp. The sheared DNA was purified using Agencourt AMPure XP Solid Phase Reversible Immobilization paramagnetic bead (SPRI) and the quality of DNA was tested with the Agilent 2100 Bioanalyzer. The end repair was done by removing the 3′ overhangs followed by the addition of a single "A" base to the 3′ end of the DNA fragments using Klenow fragment (3′ to 5′ exo minus). Specialized adaptors that have a T-base overhang at their 3′ ends were ligated. Following ligation, the samples were purified (using SPRI beads), PCR amplified and the quality was checked by the Agilent 2100 Bioanalyzer.

After hybridization the captured DNA was purified and amplified. The quality of the library was evaluated using the Agilent Bioanalyzer. Finally the 100-bp single end sequencing was performed on the Illumina Hiseq2000 platform with one sample per lane.

Mutation validation

Sanger sequencing was used to confirm the presence and genotype of variants in the candidate genes identified via exome sequencing and to screen the variants in additional family members.

Cellular cholesterol efflux assays

Human skin fibroblasts were seeded in 12-well plates and at mid-confluence labelled with 2 μCi/ml [^{3}H]-cholesterol (Perkin-Elmer Life Sciences) for 48 hours. Cells were subsequently stimulated, or not, with 2.5 μg/ml 22(R)-hydroxycholesterol (22OH) and 10μM 9-cis-retinoic acid (9CRA) for 17 hours and then incubated, or not, with 15 μg/ml lipid-free apolipoproteinA-I (ApoA-I) (Meridian Life Sciences) for 5 hours. Radioactivity was counted in both the medium and the cells. Cellular cholesterol efflux was determined as follows: ^{3}H cpm in medium / (^{3}H cpm in medium + ^{3}H cpm in cells); the results were expressed as percentage of total radiolabeled cholesterol. For the cholesterol

efflux assays in the presence of 17β-estradiol (Sigma-Aldrich), fibroblasts were labeled with 2 μCi/ml [^{3}H]-cholesterol (Perkin-Elmer Life Sciences) for 24 hours, stimulated, or not, with 2.5 μg/ml 22OH and 10μM 9CRA for 17 hours and subsequently incubated, or not, with 15 μg/ml lipid-free apolipoproteinA-I (ApoA-I) (Meridian Life Sciences) for 4 hours. During the 17 hours incubation with 22OH/9CRA, cells were simultaneously treated with increasing concentrations of 17β-estradiol. As above, assays were performed in 22OH/9CRA stimulated fibroblasts (to induce ABCA1 expression), as well as in unstimulated cells, in the presence or absence of lipid free ApoA-I. Cellular cholesterol efflux was determined as described above, but in order to specifically assess the effect of estradiol on the ABCA1 variant, we adjusted for background basal conditions of passive diffusion of cellular cholesterol. Student t-test and non-parametric two sample Wilcoxon rank sum test were used to assess differences between cholesterol efflux of a S1731C male proband and a healthy male control. A non-parametric Spearman trend test in R was used in-order to test whether increasing concentrations of 17β-estradiol has a significant influence on the cholesterol efflux of the carrier proband and wild-type control. The triplicate data for each concentration was utilized by setting the number of observations per unit equal to 3. All functional experiments were performed three times independently, involving triplicate sample measurements from individual wells for each experiment. Figure 2A-B presents one such replicate, representative of all three experiments performed where values represent the mean ± S.D. from triplicate wells.

Parametric linkage analysis

Two-point parametric linkage analysis of the low HDL-C status was performed in the extended family using the `Location-Score' option of the Mendel software.[1] We utilized an affecteds-only strategy, coding the family members as either "affected" or "unknown" based on the age- and sex-specific population 10th percentiles for HDL-C[2] to avoid problems of incomplete penetrance and ambiguity of the "unaffected" disease status. We used a dominant mode of inheritance, with gene frequencies set to 0.4% as described previously.[2] The genome scan was executed using 553 genome-wide microsatellite markers with an average density of 6 cM.[3] Genotyping and quality control

procedures of the microsatellite markers were explained in detailed previously.[3] The SLINK program[4] was utilized to approximate the maximum possible lod score of the extended family under the assumption of homogeneity within the pedigree. We used linkage parameters as given above and a marker with 4 alleles with equal frequencies. Based on 100 replicates the maximum lod score at θ=0.05 was 4.34.

Genotype by sex interaction

We included the extended family together with 10 additional families with previously identified mutations in ABCA1[5-7] in a gene-sex interaction analysis, comprising 200 individuals and 9 different mutations in ABCA1 (DelED1893, G616V, K776N, N1800H, Q2210H, R1851X, R2084X, R909X and S1731C). Genotype by sex interaction was tested by the SOLAR program[8] using variance-component analysis for discrete traits. We compared models with and without the gene-sex interaction term while keeping the ABCA1 genotypes in both the null and interaction model. We assumed a dominant genetic inheritance, classifying carriers of a mutation as 1 and 0 otherwise, and a multiplicative interaction term, multiplying the genotype score by sex (men=1 and women=0). We also coded a sex-interaction term in which men and post-menopausal women ($\geq$50 years) were coded as 1 and pre-menopausal women (<50 years) were coded as 0. Subjects with HDL-C levels < the age-sex specific 10th percentiles were classified as affected and subjects with HDL-C levels > the age-sex specific 20th percentiles as unaffected. P-values were generated by comparing the two models using a likelihood ratio statistic with one degree of freedom. Since the affection status is adjusted for gender, the inclusion of the main effect of sex in the model was no longer necessary. The binary HDL-C affection was tested because the variance of HDL-C levels in these ascertained families is reduced and thus limited for effective quantitative analysis.[9] SOLAR uses a liability threshold model in the variance-component analysis to handle discrete traits, assuming that the logarithm of the odds of being affected is a function of the effects of a major gene, polygenetic background, covariates, and residual environmental components.

Supplementary table 1. Lipid levels and other clinical characteristics of the 75 genotyped family members.

IND ID	ABCA1	LPL	TC	TG	HDL-C	%HDL	BMI	LDL-C	AGE	SEX
1	G/C	C/C	3.68	1.06	0.95	9	NA	2.24	88	F
3	C/C	C/T	2.58	6.36	0.36	<5	24.3	1.03	56	M
4	G/G	C/C	5.44	1.61	1.75	66	30.5	2.95	59	F
5	C/C	C/T	4.61	1.26	0.67	<5	22	3.36	66	M
6	G/G	C/C	6.04	1.04	2.04	84	30.7	3.53	68	F
7	G/C	C/C	3.57	1.8	0.85	<5	24.1	1.9	66	F
9	G/G	C/T	3.86	0.95	0.99	5	18.1	2.44	62	F
10	G/G	C/C	6.17	2.61	1.06	30	26.6	3.92	69	M
11	G/C	C/C	4.71	3.78	0.9	<5	23	1.92	64	F
12	G/G	C/C	3.65	2.11	0.87	11	23.4	1.82	75	M
13	C/C	C/C	4.68	0.73	1.16	15	24.7	NA	59	F
14	G/C	C/T	4.6	6.05	0.62	<5	25.8	NA	55	M
15	G/G	C/C	6.84	3.27	1.45	40	23.1	3.9	55	F
16	G/C	C/C	5.56	1.32	1.26	23	23	3.7	57	F
17	G/C	C/T	3.6	1.89	0.79	<5	21.9	1.94	31	F
18	G/C	C/T	3.74	4.8	0.72	<5	22.8	0	29	M
19	G/C	C/T	2.43	2.43	0.62	<5	27.2	0.7	21	M
20	G/C	C/C	8.7	1.62	0.77	<5	23.9	7.19	43	F
23	G/C	C/C	4.01	1.4	1.04	11	18.1	2.33	42	F
24	G/G	C/C	5.13	0.61	1.03	31	22.9	3.82	45	M
25	G/C	C/C	3.35	0.88	1.06	17	17.6	1.89	38	F
27	G/C	C/T	6.2	1.29	0.73	<5	20.2	4.88	35	F
29	G/C	C/C	4.95	1.16	0.8	<5	20.1	3.4	39	F
31	G/C	C/C	7.58	1.31	0.59	<5	21.1	5.97	45	F
32	G/G	C/C	4.68	1.94	1.22	59	22.7	2.58	48	M
34	G/G	C/C	6.61	0.84	1.6	92	26.9	4.62	45	M
35	G/G	C/C	4.45	3.29	0.54	<5	29.7	2.04	47	M
36	G/G	C/C	5.26	1.41	1.62	63	23.7	3	48	F
37	G/C	C/C	6.4	0.98	0.91	22	19.9	5.06	38	M
38	G/C	C/C	4.56	1.69	0.79	<5	26.4	2.83	45	F
40	G/G	C/C	4.54	1.9	0.64	<5	20.2	2.86	33	F
41	G/C	C/C	9.66	0.79	0.88	5	23.9	8.42	39	F
43	G/C	C/C	6.67	0.57	1.19	31	21.1	5.22	36	F
44	G/G	C/C	5.06	2.52	0.91	22	26.1	3	37	M
45	G/G	C/T	6.26	4.45	0.72	5	24.6	3.5	43	M
46	G/G	C/C	4.63	2.64	0.98	8	32.9	2.45	44	F
47	G/G	C/C	5.1	2.68	0.79	9	21.5	3.09	44	M
48	G/G	C/C	5.85	2.04	1.08	13	29.8	3.84	40	F
49	G/G	C/T	3.53	1.69	0.76	<5	20.8	1.75	43	F
50	G/G	C/C	5.58	1.41	1.19	61	28.1	3.75	43	M
51	G/G	C/C	5.37	2.31	0.82	13	24.2	3.49	37	M

52	G/G	C/C	4.79	0.8	1.82	85	18.8	2.61	37	F
55	G/G	C/C	3.24	2.46	1.07	39	19.3	1.04	27	M
56	G/G	C/T	2.84	0.83	0.85	12	18.8	1.61	24	M
57	G/G	C/C	2.72	1.03	1.1	49	24.4	1.15	43	M
58	G/C	C/C	3.92	1.37	0.67	<5	24.2	2.63	42	M
59	G/G	C/C	4.83	1.2	1.23	69	23	3.05	36	M
60	G/C	C/T	4.22	1.36	0.62	<5	28.2	2.62	35	F
61	G/G	C/C	4.56	3.44	1.35	52	20.4	1.65	29	F
62	G/G	C/C	6.11	3.02	1.46	64	25.2	3.28	24	F
63	G/C	C/C	6.51	0.7	0.98	6	17	5.21	13	F
64	G/C	C/C	4.57	1.66	1.24	39	19.5	2.58	21	F
65	G/C	C/C	5.16	1	1.41	59	17.6	3.3	19	F
66	G/G	C/C	3.22	0.74	1.26	43	17.1	1.62	16	F
67	C/C	C/C	2.85	0.71	0.73	<5	14.3	1.67	11	F
69	G/G	C/C	3.63	0.58	1.15	20	17.8	2.22	12	M
70	G/C	C/C	3.44	0.47	1.08	15	NA	2.15	10	M
71	G/G	C/C	3.51	1.23	1.31	48	18.4	1.64	21	F
73	G/C	C/C	2.71	1.01	0.75	<5	17.4	1.16	15	F
74	G/G	C/C	9.08	1.69	0.61	<5	28.1	7.29	25	M
75	G/G	C/C	6.18	1.59	1.4	58	21.5	4.06	23	F
76	G/G	C/C	4.23	0.77	1.36	80	20.3	2.52	20	M
77	G/G	C/C	5.57	1.66	0.68	<5	0	3.82	14	F
78	G/G	C/C	3.8	1.03	0.61	<5	22	2.28	15	F
79	G/C	C/C	3.18	0.72	0.83	<5	23.9	1.83	27	F
80	G/G	C/C	4.87	0.51	0.9	12	18.2	3.74	17	M
81	G/C	C/C	3.77	0.63	1.24	41	14.8	2.24	16	F
82	G/C	C/C	5.77	0.98	0.8	<5	16.5	4.52	10	F
87	G/G	C/C	3.86	0.91	1.07	36	23.1	2.38	21	M
89	G/G	C/T	3.81	2.19	0.7	<5	26.6	2.11	22	M
90	G/G	C/C	3.59	0.51	1.11	42	21.2	2.25	20	M
91	G/G	C/T	3.18	2.99	0.66	<5	23.5	0.94	16	M
92	G/G	C/C	4.04	0.55	1.39	79	18.4	2.4	16	M
93	G/G	C/C	3.98	0.87	0.93	15	19.7	2.65	15	M
94	G/G	C/C	4.21	0.42	1.43	53	17.6	2.59	13	M

The lipid levels are shown in millimoles per liter.

Supplementary table 2. List of 41 variants shared by the three exome sequenced individuals after filtering.

Chr no	Position no	rs number	Gene name	PolyPhen*	SIFT†
2	42990225	New	OXER1	Probably	Damaging
2	160993949	New	ITGB6	Stop	Stop
3	15477933	New	EAF1	Probably	Damaging
3	47047500	New	NBEAL2	Probably	Damaging
3	196529902	New	PAK2	Probably	Damaging
4	1388675	New	CRIPAK	Probably	Damaging
5	140784743	New	PCDHGA9	Probably	Tolerated
8	145094836	New	SPATC1	Possibly	Tolerated
8	145112971	New	OPLAH	Probably	Tolerated
9‡	**107558635**	**New**	**ABCA1**	**Probably**	**Damaging**
10	34606158	New	PARD3	Possibly	Tolerated
11	57076419	New	TNKS1BP1	Probably	Tolerated
12	124362332	New	DNAH10	Probably	Damaging
15	59139625	New	FAM63B	Benign	Damaging
16	2003016	New	RPL3L	Benign	Damaging
17	45234303	New	CDC27	Possibly	Tolerated
17	44144993	New	KIAA1267	Benign	Damaging
19	14675764	New	TECR	Probably	Damaging
22	45821982	New	RIBC2	Probably	Damaging
22	39069227	New	CBY1	Benign	Damaging
1	115537367	rs61730058	SYCP1	Probably	Tolerated
1	144852390	rs61804988	PDE4DIP	Stop	Stop
2	11943082	rs4669781	LPIN1	Possibly	Tolerated
3	49162583	rs35713889	LAMB2	Probably	Tolerated
5	35753763	rs79487218	SPEF2	Benign	Damaging
5	140255119	rs114654172	PCDHA12	Possibly	Tolerated
8‡	**19811790**	**rs118204060**	**LPL**	**Probably**	**Damaging**
8	144995494	rs76803079	PLEC	Probably	Damaging
10	43871158	rs41307500	FXYD4	Probably	Tolerated
10	127697954	rs1666	FANK1	Possibly	Damaging
11	68174189	rs4988321	LRP5	Probably	Tolerated
11	56310356	rs17547284	OR5M11	Stop	Stop
11	36458997	rs62621409	PRR5L	Probably	Damaging
15	45491082	rs80131405	SHF	Benign	Damaging
16	28488943	rs77595156	CLN3	Probably	Tolerated
16	1537693	rs61734779	PTX4	Possibly	Tolerated
17	37224211	rs75117355	PLXDC1	Probably	Tolerated
19	42341407	rs35476281	LYPD4	Probably	Tolerated
19	23545516	rs112713994	ZNF91	Probably	Tolerated
19	49445774	rs10423255	DHDH	Stop	Stop
19	41235167	rs112628847	ITPKC	Benign	Damaging

* PolyPhen-2 was used to predict the possible impact of an amino acid substitution on the structure and function of the protein. A score larger than 0.85 is considered as probably damaging, a score smaller than 0.15 as benign, and a score between 0.85 and 0.15 as possibly damaging, respectively.

† SIFT predicts the amino acid substitution to be damaging if the score is less than 0.05, and tolerated if the score is greater than 0.05.

‡ The ABCA1 (S1731C) and LPL (P234L) variants are highlighted in bold.

REFERENCES

1. Lange K, Cantor R, Horvath S, Perola M, Sabatti C, Sinsheimer J, Sobel E. Mendel version 4.0: A complete package for the exact genetic analysis of discrete traits in pedigree and population data sets. *Am J Hum Genet.* 2001;69:504.

2. Soro A, Pajukanta P, Lilja HE, Ylitalo K, Hiekkalinna T, Perola M, et al. Genome scans provide evidence for low-HDL-C loci on chromosomes 8q23, 16q24.1-24.2, and 20q13.11 in Finnish families. *Am J Hum Genet.* 2002;70:1333-1340.

3. Dastani Z, Quiogue L, Plaisier C, Engert JC, Marcil M, Genest J, et al. Evidence for a gene influencing high-density lipoprotein cholesterol on chromosome 4q31.21. *Arterioscler Thromb Vasc Biol*. 2006;26:392-397.

4. Weeks DE, Ott J, Lathrop GM. SLINK: a general simulation program for linkage analysis. *Am J Hum Genet.* 1990;47:204.

5. Brooks-Wilson A, Marcil M, Clee SM, Zhang LH, Roomp K, van Dam M, et al. Mutations in ABC1 in Tangier disease and familial high-density lipoprotein deficiency. *Nat Genet.* 1999;22:336-345.

6. Alrasadi K, Ruel IL, Marcil M, Genest J. Functional mutations of the ABCA1 gene in subjects of French-Canadian descent with HDL deficiency. *Atherosclerosis.* 2006;188:281-291.

7. Marcil M, Brooks-Wilson A, Clee SM, Roomp K, Zhang LH, Yu L, et al. Mutations in the ABC1 gene in familial HDL deficiency with defective cholesterol efflux. *Lancet.* 1999;354:1341-1346.

8. Almasy L, Blangero J. Multipoint quantitative-trait linkage analysis in general pedigrees. *Am J Hum Genet.* 1998;62:1198-1211.

9. Pajukanta P, Allayee H, Krass KL, Kuraishy A, Soro A, Lilja HE, Mar R, et al. Combined analysis of genome scans of Dutch and Finnish families reveals a susceptibility locus for high-density lipoprotein cholesterol on chromosome 16q. *Am J Hum Genet.* 2003;72:903-917