Journal of Taibah University Medical Sciences (2015) 10(3), 266–270

Taibah University

Journal of Taibah University Medical Sciences

www.sciencedirect.com

Review Article

Usefulness of genome-wide association studies to identify novel genetic variants underlying the plasma lipoprotein metabolism as risk factors for CAD

Naif A.M. Almontashiri, PhD [a,b,c,d,*] and Mohammed Hannan, PhD [a]

[a] *Center for Genetics and Inherited Diseases, Taibah University, Almadinah Almunawwarah, KSA*
[b] *University of Ottawa Heart Institute, Ottawa, Canada*
[c] *Department of Biochemistry, Microbiology and Immunology, University of Ottawa, Ottawa, Canada*
[d] *Brigham and Women's Hospital, Harvard Medical School, Harvard University, Boston, United States*

Received 14 December 2014; revised 24 January 2015; accepted 25 January 2015; Available online 4 March 2015

الملخص

تتفاعل العوامل البيئية والصحية والجينية لتتسبب في ظهور أمراض القلب التاجية. وتعتبر أمراض القلب التاجية من الأسباب الرئيسة للوفيات حول العالم. وتشمل العوامل التقليدية المسببة لأمراض القلب التاجية عوامل كثيرة من أهمها العمر، وارتفاع ضغط الدم، والسكري، والتدخين، وارتفاع نسبة الدهون في الدم. إن الحفاظ على الدهون في مستوياتها الطبيعية يخفض وبشكل كبير من نسبة الإصابة بأمراض القلب وتصلب الشرايين. وقد صممت دراسات الارتباط على مستوى الجينوم بطريقة غير منحازة للتعرف على الاختلافات في تسلسل الجينوم البشري، التي قد تضع الإنسان في خطر (أو تحميه من) الإصابة بأمراض معقدة، ومتعددة الأسباب الجينية كأمراض القلب التاجية والسكري. في هذه المراجعة لما تم عمله من دراسات حتى الآن، نسلط الضوء ونناقش المستجدات لما أسفر عنه استخدام هذه التقنية في اكتشاف اختلافات متعددة في تسلسل الجينوم التي أظهرت ارتباطا بمستويات الدهون في الجسم (نقص أو زيادة)، وكيف يمكن أن تساعد هذه النتائج في التعرف على أهداف علاجية، قد تستخدم مستقبلا في المساعدة في تخفيض مستويات الكولسترول والدهون المرتفعة في الدم.

الكلمات المفتاحية: دراسات الارتباط على مستوى الجينوم؛ بروتين شحمي؛ أمراض القلب التاجية؛ التصلب العصيدي؛ الكولسترول

Abstract

Coronary artery disease (CAD) is a major killer across the world. The pathogenesis of CAD is a construct of multiple predisposing elements, including environmental, health and genetic factors. Traditional risk factors for CAD include age, hypertension, diabetes, smoking, and dyslipidaemia. Optimizing the lipid levels to within the normal range significantly and drastically reduces the risk of coronary atherosclerosis. Genome-wide association studies (GWASs) promise to accurately identifying the variants that increase or decrease the risks of multiple and complex disorders. In this review, we shed light on and discuss the recent GWASs of lipoprotein genetics and how such studies have provided new pathways and pharmacological targets that might enable the control the pathological plasma cholesterol levels.

Keywords: Atherosclerosis; CAD; Cholesterol; GWAS; Lipoprotiens; SNPs

Introduction

Coronary artery disease (CAD) is the leading cause of death in the western world[1] and is also a major cause of mortality in other countries, including Saudi Arabia.[2] CAD is an age-progressive disease; although fewer than 10% of individuals exhibit symptoms of CAD before the age of 50, intravascular ultrasound examinations have revealed that 1 in 6 adolescents and 85% of persons over the age of 50 have

* Corresponding address: Center for Genetics and Inherited Diseases, Taibah University, Almadinah Almunawwarah, KSA.
E-mail: nmontashri@taibahu.edu.sa (N.A.M. Almontashiri)
Peer review under responsibility of Taibah University.

1658-3612 http://dx.doi.org/10.1016/j.jtumed.2015.01.012

measurable coronary atherosclerosis.[3] Using angiograms, the burden of atherosclerosis can be scored by the number of affected vessels (i.e., 0, 1, 2 and 3-vessel disease) and used to measure the severity of CAD.[3] Moreover, CAD is a silent killer because the majority of people who experience myocardial infarction (MI) do not experience symptoms prior to the infarction despite having extensive coronary atherosclerosis. Several modifiable factors increase the risk of CAD and MI, including smoking, diabetes mellitus, circulating cholesterol levels, hypertension, and obesity. Treating these risk factors reduces the mortality due to CAD by 30–40%.[4] In contrast, studies of twins have revealed that up to 50% of the risk for CAD is genetic,[5] which suggests that the identification of the responsible genetic factors will be useful for predicting individual risks for CAD. Understanding the molecular pathways that are affected by these genetic factors might also aid the identification of novel targets for therapeutic treatments for the disease. The traits associated with cardiovascular disease risk factors, such as coronary artery and plasma lipoprotein levels, exhibit complex inheritances that are suggestive of interactions between multiple genes and non-genetic factors such as the environment.[6] To unravel and dissect these complexities, extensive data collection and analyses from large samples are required to genetically map these complex traits. One such approach that allowed us to map these complex traits is the genome wide association study (GWAS).

In 2007, using microarrays of single nucleotide polymorphisms (SNPs) to genotype large numbers of cases and controls, McPherson et al. identified the first common genetic variants on chromosome 9p21.3 that increase CAD risk[7] using a GWAS. Several other large GWAS investigations have confirmed this association with CAD and/or extended it to MI.[8–10] Dandona et al. recently showed that this locus promotes coronary atherogenesis rather than plaque rupture because the number of risk alleles is strongly correlated with the severity of CAD but not with acute coronary syndrome among CAD cases defined by coronary angiography.[11] However, the mechanism by which the 9p21.3 locus increases the risk of CAD remains elusive.[12] To date, more than 30 loci have been identified by GWAS, including 13 novel loci that were discovered by a large international consortium called the CARDIoGRAM.[13] Importantly, using only 12 of these loci, Davies et al. showed cardiovascular risk prediction could be improved beyond traditional risk factors analysis.[14]

Factors that increase the risk for coronary atherosclerosis include high serum cholesterol concentrations, high levels of low density lipoprotein (LDL) and very low density lipoprotein (VLDL), and low levels of high density lipoprotein (HDL).[15–17] Intervention trials have shown that reducing the levels of LDL-C in particular can reduce CAD risk[18] and the rate of progression of coronary artery stenosis as demonstrated angiographically.[19] In contrast, HDL-C levels are inversely related to CAD risk.[20] In dyslipidaemia, the levels of lipids or lipoproteins are abnormal due to genetic and/or environmental conditions that alter the biogenesis or the metabolism of the lipoproteins in the circulation.

GWASs have allowed major advances the discovery of numerous risk loci that affect blood lipoprotein levels and increase the risk of coronary atherosclerosis. Based on their locations within or in close proximity to known[21,22] or previously unknown genes that affect lipoproteins levels,[23] these variants have increased our knowledge about the control and involvement of lipoproteins in the process of atherosclerosis. We next review examples, in chronological order, of the successes of GWASs in identifying novel loci that are associated with CAD risk and that affect lipoprotein levels and how some of these loci might be involved in new pathways that lead to CAD. Notably, the majority of the relevant studies used cohorts defined by coronary angiography to measure atherosclerosis.

The unbiased power of GWASs in identifying new variants

In 2008, Kathiresan and colleagues used genotype imputation and meta-analysis to combine three genome-wide surveys to scan for and identify loci that influence lipid concentrations and the risk of coronary artery disease.[24] These authors identified several loci that are associated with lipoprotein levels. Interestingly, the 11 independent common variants that were found to be associated with increased LDL cholesterol concentrations in their study also exhibited significant association with the risk of CAD. These loci are involved in many aspects of lipid metabolism. For example, PCSK9 is required for the clearance of the LDL receptor, and the LDLR is required for LDL uptake and clearance. PCSK9 has been shown to be elevated during acute myocardial infarction in two independent studies.[32] Collectively, these results suggest that the cumulative and the lifetime effects of these multiple common variants contribute to polygenic dyslipidaemia. These findings have been replicated by others.[22,25]

In 2009, Robert Clarke and colleagues identified the independent relevance of two common variants (rs10455872 and rs3798220) of the *LPA* gene to CAD.[21] Together, these two variants explained 36% of the total variation in Lp(a) lipoprotein levels. These authors found that the effects of these variants on CAD risk and Lp(a) lipoprotein levels are correlated. This linear dose-response relationship supports a pathogenic role of elevated Lp(a) lipoprotein levels in CAD risk. This study is consistent with earlier studies that demonstrated a clear association between Lp(a) and CAD severity.[26,27]

In 2010, Kathiresan and colleagues performed a large scale GWAS searching for common variants associated with the plasma lipids in European and non-European populations.[25] First, these authors performed a meta-analysis of 46 lipid GWASs that comprised more than 100,000 individuals of European descent. They were able to identify 95 significantly associated loci that reached the genome wide significance levels of $<5 \times 10^{-8}$. Of these loci, 59 were found to exhibit genome-wide significant associations with lipid traits for the first time. Among these 59 novel loci, 22 exhibited genome-wide significant associations with LDL-C, 31 with HDL-C, and 16 with TG. To identify additional SNPs that exhibited independent associations at each of the identified loci, these authors performed conditional association analyses for each of the four lipid traits. Using this approach, they identified 26 additional secondary loci.

When these secondary SNPs were combined with the primary SNPs, the overall mapped variants explained 12.1% (HDL-C), 12.2% (LDL-C), and 9.6% (TG) of the total variances in each lipid trait in the Framingham Heart Study, which represented 25−30% of the genetic variance for each lipoprotein trait. To gain further insight into how the newly identified variants affected plasma lipid concentration, these authors examined the correlations between each of the primary SNPs at the 95 loci and expression levels of the transcripts located within 500 kilobases of the SNPs in three different tissues that are relevant to these lipoproteins. They were able to identify tissue-specific effect of these SNPs on the lipids. These SNPs were relevant to other non-European populations (i.e., East Asians, South Asians and African Americans). To gain further insight into the clinical importance and relevance of these loci, they examined which of the lipid loci were associated with the risk of CAD. Four novel CAD-associated loci (i.e., IRS1, C6orf106, KLF14 and NAT2) that are related to TG and/HDL-C were identified. These findings might indicate that these variants exert their effects on CAD by affecting lipoproteins or triglycerides (CAD risk factors) via direct or indirect pleiotropic effects. Next, these authors examined whether the GWAS hits could provide biological insight into how the plasma lipoproteins and lipids are affected by these novel loci. In a mouse model, they validated the associations of selected genes (i.e., Galnt2, Ppp1r3b and Ttc39b) with previously unknown or unclear mechanistic links to lipid metabolism. The abilities of the GWASs to discover SNPs that affect lipid traits and the observation that only a small fraction of these SNPs affect coronary atherosclerosis suggest that the lipid traits have different end-point diseases and do not work through the same pathological mechanisms when they are perturbed.

In 2013, to increase the power to detect variants that occur at low frequencies and exert major effects on lipid levels, the Global Lipids Genetics Consortium conducted the largest GWAS.[28] This consortium combined 60 studies for a total of more than 188,000 participants of European ancestry to search for variants that affect the lipid traits and their associated phenotypes. The authors were able to identify 62 new loci in protein-coding regions (functional variants) that affect blood lipid levels, which brought the total number of loci that have been associated with lipid traits via GWAS investigations to 157 variants. These loci are associated with increased risks of CAD or metabolic disorders. One of the interesting variants they discovered is near the human orthologue of the GPR146 gene and is associated with high levels of total cholesterol and thus provided a new pharmacological target for reducing cholesterol and possibly improving CAD control.

From GWAS variants to new biological insights and functions

In another biological validation of the top GWAS hits, the same group functionally characterized a non-coding variant at the *SORT1* gene.[23] This locus on chromosome 1p13 has been shown to be strongly associated with LDL-C[25] and the risks of CAD and MI.[10,25] However, before its discovery in a GWAS, the functional linkages of *SORT1* with LDL-C and metabolism were not known. To provide further information about the association identified by the GWAS, these authors used a series of studies of human cohorts and human-derived hepatocytes to show that the common noncoding polymorphism at this locus, rs12740374, alters the hepatic expression of the SORT1 gene. Furthermore, using knockdown and over-expression experiments in mouse livers, they showed that Sort1 alters the plasma levels of LDL-C and VLDL by modulating hepatic VLDL secretion. This work provided a novel regulatory pathway of lipoprotein metabolism and a new avenue for therapeutic approaches to reducing the increased risk of MI that results from alterations to this pathway. Collectively, these two studies indicated the power of GWASs for identifying variants that might alter gene function and increase disease risk through novel pathways.

In 2011, The IBC 50K CAD Consortium conducted a large and systematic of CAD risk candidate gene study that included analyses of many rare and functional variants.[29] The authors identified novel variants at LIPA, TRIB1, and ABCG5/ABCG8. The functional significances of these variants are supported by gene expression data or demonstrated effects on lipid levels. These are variants of novel genes that are associated with the risk of CAD and thus might help to improve our understanding of the aetiopathologies of cardiovascular disease and identify new targets and pathways for therapeutic approaches that address the functional aspects of those aetiopathologies.

Around the same time, a large meta-analysis (14 GWAS investigations of CAD and MI) comprising 22,233 CAD cases and 64,762 controls[13] was conducted by the CARDIoGRAM. Ten loci that had previously been associated with the risk of CAD at the genome-wide significance level reached the same threshold of significance in the initial meta-analysis performed by the CARDIoGRAM (i.e., *SORT1, MIA3, WDR12, MRAS, PHACTR1, LPA, 9p21, CXCL12, LDLR, and MRPS6*). At least 4 of these loci are related to LDL-C. Most importantly, the CARDIoGRAM identified 13 novel loci that are related to CAD. Two of these novel loci, 9p34.2 (near ABO) and 11q23.3 (near ZNF259, APOA5, and A4-C3-A1), are associated with plasma LDL-C levels.

Many GWAS studies suffer the limitation of a lack of functional information about the identified lipid-associated loci. However, GWAS investigations provide a fertile clinical background for functional studies of promising variants that exert substantial effect on lipids and associated cardiovascular events. A remarkable large-scale functional study was performed by Blattman et al. to functionally analyse 133 candidate genes in close proximity to 56 loci that had previously been reported to be strongly associated with lipids and CAD-related factors, such as atherosclerosis and MI, using the RNA interference (RNAi) approach.[30] The knockdowns revealed that more than 40% of these genes were involved in LDL-C metabolism. Furthermore, these authors demonstrated previously unknown lipid regulatory roles for several genes (e.g., FAM174A, CXCL12, TBL2, PAFAH1B1, and SEZ6L). Unbiased yet resourceful functional studies are required to follow-up on these GWAS findings. Such studies will help to establish the foundation for the development of new pharmaceutical interventions

that target the genes that are involved in pathological lipid regulation and increased cardiovascular burden.

Conclusion

In conclusion, it is notable that seven of the 36 risk loci that have thus far been found to be associated with CAD or MI exert their effects by increasing cholesterol levels,[31] which indicates the importance of lipoproteins in the process of atherosclerosis. The independent associations of some of these loci with either CAD or MI indicate that specific genetic loci might promote the development of coronary atherosclerosis, whereas others might lead to MI in the presence of CAD. These GWASs have identified risk loci that are near or within genes, such as *PCSK9* and *LDLR*, which are known to affect lipoprotein metabolism, which proves the validity and authenticity of GWASs. In contrast, some GWAS investigations have identified novel CAD and lipid metabolism genes, such as *SORT1,* that were not previously known to affect lipid metabolism. Such findings illustrate the innovative and powerful characteristics of GWASs. Additional approaches, such as whole-exome sequencing, are required to identify additional independent variants in genes that are known to affect plasma lipoprotein levels and increase the risk of CAD through GWASs. Such approaches in combination with GWAS findings will help us to better calculate the risk scores conferred by these loci. However, extensive functional and biological characterizations are required to decode the importance of many of these risk loci to maximize their usefulness in the identification of new regulatory pathways and consequently new targets for the treatment of dyslipidaemia and the reduction of the risks for heart diseases.[23,25]

Recommendations

In view of the increasing incidence of CAD in the Saudi population, in addition to replicating these GWAS finding, it would be worthwhile to apply the GWAS approach to identify novel genetic risk factors and analyse the results in relation to the dietary/metabolic data of the patients in the country to better understand the gene-environment interactions that lead to CAD. Projects that involve exome and candidate gene sequencing for Mendelian and familiar disorders will be required to identify consanguinity-related diseases and mutations in the Saudi population that will not be accounted for by GWAS. Such knowledge will certainly aid the identification of individuals at risk and the planning of preventive strategies.

Conflict of interest

The authors have no conflict of interest to declare.

References

1. Murray CJ, Lopez AD. Alternative projections of mortality and disability by cause 1990-2020: Global Burden of Disease Study. **Lancet 1997**; 349: 1498–1504.
2. Al-Nozha MM, Arafah MR, Al-Mazrou YY, et al. Coronary artery disease in Saudi Arabia. **Saudi Med J 2004**; 25: 1165–1171.
3. Tuzcu EM, Kapadia SR, Tutar E, et al. High prevalence of coronary atherosclerosis in asymptomatic teenagers and young adults: evidence from intravascular ultrasound. **Circulation 2001**; 103: 2705–2710.
4. Shepherd J, Cobbe SM, Ford I, et al. Prevention of coronary heart disease with pravastatin in men with hypercholesterolemia. West of Scotland Coronary Prevention Study Group. **N Engl J Med 1995**; 333: 1301–1307.
5. Marenberg ME, Risch N, Berkman LF, Floderus B, de Faire U. Genetic susceptibility to death from coronary heart disease in a study of twins. **N Engl J Med 1994**; 330: 1041–1046.
6. Kathiresan S, Srivastava D. Genetics of human cardiovascular disease. **Cell 2012**; 148: 1242–1257.
7. McPherson R, Pertsemlidis A, Kavaslar N, et al. A common allele on chromosome 9 associated with coronary heart disease. **Science 2007**; 316: 1488–1491.
8. Wellcome_Trust_Case_Control_Consortium. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. **Nature 2007**; 447: 661–678.
9. Helgadottir A, Thorleifsson G, Manolescu A, et al. A common variant on chromosome 9p21 affects the risk of myocardial infarction. **Science 2007**; 316: 1491–1493.
10. Samani NJ, Erdmann J, Hall AS, et al. Genomewide association analysis of coronary artery disease. **N Engl J Med 2007**; 357: 443–453.
11. Dandona S, Stewart AFR, Chen L, et al. Gene dosage of the common variant 9p21 predicts severity of coronary artery disease. **J Am Coll Cardiol 2010**; 56: 479–486.
12. Almontashiri NA, Fan M, Cheng BL, Chen HH, Roberts R, Stewart AF. Interferon-gamma activates expression of p15 and p16 regardless of 9p21.3 coronary artery disease risk genotype. **J Am Coll Cardiol 2013**; 61: 143–147.
13. Schunkert H, Konig IR, Kathiresan S, et al. Large-scale association analysis identifies 13 new susceptibility loci for coronary artery disease. **Nat Genet 2011**; 43: 333–338.
14. Davies RW, Dandona S, Stewart AFR, et al. Improved prediction of cardiovascular disease based on a panel of single nucleotide polymorphisms identified through genome-wide association studies. **Circ Cardiovasc Genet 2010**; 3: 468–474.
15. Burke AP, Farb A, Malcom GT, Liang YH, Smialek J, Virmani R. Coronary risk factors and plaque morphology in men with coronary disease who died suddenly. **N Engl J Med 1997**; 336: 1276–1282.
16. Phillips NR, Waters D, Havel RJ. Plasma lipoproteins and progression of coronary artery disease evaluated by angiography and clinical events. **Circulation 1993**; 88: 2762–2770.
17. Shestov DB, Deev AD, Klimov AN, Davis CE, Tyroler HA. Increased risk of coronary heart disease death in men with low total and low-density lipoprotein cholesterol in the Russian Lipid Research Clinics Prevalence Follow-up Study. **Circulation 1993**; 88: 846–853.
18. Tyroler HA. Overview of clinical trials of cholesterol lowering in relationship to epidemiologic studies. **Am J Med 1989**; 87: 14S–19S.
19. Waters D, Lesperance J. Regression of coronary atherosclerosis: an achievable goal? Review of results from recent clinical trials. **Am J Med 1991**; 91: 10S–17S.
20. Eder HA, Gidez LI. The clinical significance of the plasma high density lipoproteins. **Med ClinNorth Am 1982**; 66: 431–440.
21. Clarke R, Peden JF, Hopewell JC, et al. Genetic variants associated with Lp(a) lipoprotein level and coronary disease. **N Engl J Med 2009**; 361: 2518–2528.
22. Kathiresan S, Voight BF, Purcell S, et al. Genome-wide association of early-onset myocardial infarction with single nucleotide polymorphisms and copy number variants. **Nat Genet 2009**; 41: 334–341.

23. Musunuru K, Strong A, Frank-Kamenetsky M, et al. From noncoding variant to phenotype via SORT1 at the 1p13 cholesterol locus. **Nature 2010**; 466: 714–719.
24. Kathiresan S, Willer CJ, Peloso GM, et al. Common variants at 30 loci contribute to polygenic dyslipidemia. **Nat Genet 2009**; 41: 56–65.
25. Teslovich TM, Musunuru K, Smith AV, et al. Biological, clinical and population relevance of 95 loci for blood lipids. **Nature 2010**; 466: 707–713.
26. Dahlen GH, Guyton JR, Attar M, Farmer JA, Kautz JA, Gotto Jr AM. Association of levels of lipoprotein Lp(a), plasma lipids, and other lipoproteins with coronary artery disease documented by angiography. **Circulation 1986**; 74: 758–765.
27. Danesh J, Collins R, Peto R. Lipoprotein(a) and coronary heart disease. Meta-analysis of prospective studies. **Circulation 2000**; 102: 1082–1085.
28. Willer CJ, Schmidt EM, Sengupta S, et al. Discovery and refinement of loci associated with lipid levels. **Nat Genet 2013**; 45: 1274–1283.
29. Large-scale gene-centric analysis identifies novel variants for coronary artery disease. **PLoS Genet 2011**; 7: e1002260.
30. Blattmann P, Schuberth C, Pepperkok R, Runz H. RNAi-based functional profiling of loci from blood lipid genome-wide association studies identifies genes with cholesterol-regulatory function. **PLoS Genet 2013**; 9: e1003338.
31. Roberts R, Stewart AF. The genetics of coronary artery disease. **Curr Opin Cardiol 2012**; 27: 221–227.
32. Almontashiri NAM, Vilmundarson RO, Ghasemzadeh N, et al. Plasma PCSK9 levels are elevated with acute myocardial infarction in two independent retrospective angiographic studies. **PLoS ONE 2014**; 9(9): e106294.