Editorial

Searching for the Causal Variants of the Association Between Hypertriglyceridemia and the Genome-Wide Association Studies–Derived Signals?
Take a Look in the Native American Populations

Carlos A. Aguilar-Salinas, MD, PhD; Magdalena del Rocío Sevilla González, MSc;
María Teresa Tusie-Luna, MD, PhD

Genome-wide association studies (GWAS) have made possible the identification of >175 loci that affect plasma lipid levels. Its results have been crucial to identify roles of new players in lipid metabolism (ie, apolipoprotein A5) or even to postulate potential treatment targets (ie, apolipoprotein C-III). However, a large proportion of the GWAS results has not been translated in clinically useful information because a large proportion of responsible single nucleotide polymorphisms (SNPs) are located either in noncoding regions or in genes without an obvious participation in any metabolic pathway.[1] This is the case of the association between rs964184 and plasma triglycerides concentrations. This highly significant association has been a constant finding in the GWAS reports, regardless the sample size or the study sample composition.[2] The frequency of the risk allele (G) varies between populations from 12% in whites to 27% in Mexicans. Also, it is common in Japanese (31%) and in Native American communities located in Central Mexico (≈50%). This SNP is nearby the 3′ untranslated region of the zinc finger gene (*ZPR1*) and close to *APOA5*. However, the identification of the functional variant that explains the association has not been possible.

See Article by Hsueh et al

The study of ethnic groups not included in the discovery cohorts, some methodological approaches (ie, admixture mapping or inception by descent [IBD] mapping), and deep genotyping are among the potential approaches to identify the variants responsible for a GWAS signal.[3] The study of the Native American populations is an option because this group has not included in the majority of the lipids-related GWAS. The greater susceptibility of the populations derived from the Native American founding groups for having hypertriglyceridemia is well documented.[4] These groups include the Amerindian communities living in Canada and the United States, mestizos living in the Latin American countries, and Hispanic residents living outside their countries. Native American populations are heterogeneous groups that have remarkable differences in admixture with other ethnic groups. The Amerindians have experienced biological and human crises that have reshaped several times their environment, lifestyle, and the size of the population.[5] Thus, this group has been under a selection process for centuries that may alter linkage disequilibrium patterns and change the frequency of rare variants with major effects. Interethnic studies have been successful approach to narrow large loci in which associations have been detected.

Mexican mestizos is an admixed group that have ≈50% of their genetic background from Native American origin. Two independent groups have attempted to identify the ethnic-specific genetic variants associated with plasma lipid levels in Mexican mestizos.[6–8] Both groups found that the strongest triglycerides signal is associated with rs964184. First, Weissglas-Volkov et al[6] published a 2-stage GWAS report in which only Mexicans cases and controls were included. A joint analysis of stages 1 and 2 data (n=4361) was performed. The main signal for triglycerides was rs964184 ($P=5.5\times10^{-35}$). The different disequilibrium pattern found in Mexicans allowed the refinement of the length of the loci. This SNP is in high linkage disequilibrium with 26 other SNPs in whites but not in Mexicans. The *APOA5* SNP rs3135506 was discarded as the functional variant because the statistical significance was lost when rs964184 was taken into account in the multivariate model. The authors proposed that this SNP is the responsible variant of the association by modulating *APOA5* expression. To prove it, serum ApoA5 protein level during a fat tolerance test was compared between the rs964184 genotype groups. Homozygotes for the risk allele (G/G) was the group with the lowest ApoA5 postprandial response. Thus, rs964184 may have a regulatory effect of *APOA5* expression. This observation was extended with a cross-ethnic mapping (cross-population allele screen) in which the strong association with rs964184 was replicated ($P=6.08\times10^{-33}$).[7] An independent signal in SIK family kinase 3 gene (*SIK3*) was found also.

Second, Parra et al[8] applied the admixture mapping approach, a strategy based on the association between a phenotype and locus ancestry. They included 2 samples composed of 1310 and 1787 individuals with a Native American heritage ≈64%. The strongest association with triglycerides levels was found in a large region in chromosome 11 in which the lead SNP was rs964184 (P=5.32×10^{-37}). Fine mapping of the region confirmed that there are no variants in strong linkage disequilibrium with rs964184 in Mexican mestizos. Using as the 1000 Genomes Project Phase 1 reference panels to do the imputations, the authors concluded that rs964184 is the causal SNP. The authors proposed that this association could be explained by 3 potential mechanisms: modification of enhancer activity, regulation of the expression of several genes, or modification of the methylation patterns of the promoter of *APOA5*. In summary, evidence obtained in Mexican mestizos with two analytic approaches applied to unrelated individuals suggest that rs964184 is the causal SNP for the strong and consistent association between the chromosome 11 locus and triglycerides levels

In this issue of the *Circulation: Cardiovascular Genetics*, Hsueh et al[9] inform the results of an IBD mapping analyses in a population-based sample obtained in Pimas in which plasma lipid values were available. This approach differs from the studies above described. In contrast to the GWAS, IBD mapping is based on data of related individuals. IBD mapping is a powerful tool for the identification of high-risk variants in a founder population,[10] such as the Pimas. Several reasons make IBD mapping a suitable approach in the Pima population: (1) Pimas communities could be considered a founder population because of the minimal admixture that they have with other ethnicities[11]; (2) Conventional genotyping platforms for GWAS still lack population-specific and low-frequency variants; (3) With the advent of methods for IBD mapping in population-based cohorts, there is no need to collect pedigree information, therefore, it uses all possible matching pairs within the study population; and (4) IBD mapping increases the power to identify genes or loci containing multiple disease susceptibility variants, providing the availability of whole sequencing in a fraction of the studied individuals.[12] The authors used as discovery sample the information from 1024 individuals in which a population-based genome-wide linkage study was available. Results were supplemented with fine-mapping analyses, replication association studies, and the fitting of a final linkage-and-association model.

They replicate the strong association between triglycerides and the chromosome 11q23 loci in which rs964184 is located (P=4×10^{-8}). The independence of associations of all SNPs in this region was lost if the rs964184 was included in the model. Fine mapping was performed using data from 296 Pimas as the reference pattern. The authors identified 3 variants. The strongest association was found with rs147210663 explaining 6.9% of the variance (P=1.6×10^{-13}). This *APOC3* SNP is an Ala→Thr substitution (A43T). The minor allele codes for the Thr residue and has a frequency of 2.6% in genome-wide linkage study. This variant is associated with lower triglycerides concentrations, and it is rare in whites. The SNP with the second strongest association was rs2072560 after adjusting for the effect of rs147210663 (P=0.00028). This SNP resides in intron 3 of *APOA5* but has

no known function. The third variant with a significant association was rs11357208 after adjusting for the effects of two previous SNPs (P=0.0049). This insertion–deletion variant resides in intron 5 of *SIK3* without any known function. The replication studies included a different set of Pima samples (n=4668) and a group of urban Amerindians included in the FIND study (Family Investigation of Nephropathy and Diabetes; n=2793). Only the *APOC3* A43T association was replicated. The authors look for associations between rs964184 and *APOA5* functional variants with a minor allele frequency ≥1%. Four variants fulfilled the requested criteria and were genotyped. Three *APOA5* SNPs (rs651821, rs3135506, rs2266788) remarkably diminish the strength of the rs964184 signal (P=0.015). These three SNPs have a greater minor allele frequency in Pimas compared with whites. This analysis was complicated by the high linkage disequilibrium of the variants. The authors did haplotype analyses to solve this limitation. The *ApoA5*2* haplotype (which contains rs2266788 and rs651821, 2 of the 3 *APOA5* functional variants) was strongly associated with triglycerides levels. Authors concluded their report looking if the *APOC3* A43T SNP and the 3 *APOA5* functional variants explain the chromosome 11q23 linkage signal in the discovery sample. The *APOC3* 43T SNP had a significant effect; it reduces the LOD score from 9.32 to 2.24. The remaining linkage signal was lost when the three *APOA5* SNPs were added to the model. They conclude that instead of being the causal variant, rs964184 is a marker for the aggregate effect of the 3 *APOA5* SNPs.

The main contribution of this report is the identification of the strong contribution of *APOC3* loss-of-function variant on the chromosome 11q23 signal. The analytic approach and the genetic background of the Pima population made possible this finding. The low frequency of the A43T variant would not be captured in a GWAS study. In Amerindian-derived populations, rare variants with large effects have been previously identified for different metabolic traits, including triglyceride and high-density lipoprotein levels.[13] However, the relatively small sample size, the low-density genome-wide data (400 000 SNPs) to refine chromosome 11 locus, and the lack of replication of the effect of the *APOA5* SNPs to the chromosome 11q23 signal in two Native American samples leave open the question if rs964184 is a marker or a causal variant. Additional experimental evidence will be required to discern if other genes besides *APOA5* contribute to the rs964184 signal.

Exome sequencing is another approach to identify the variants responsible for a GWAS signal. Recently, an exome-wide association was reported with the information of 300 000 individuals.[14] A multiethnic sample was included; <5000 were Hispanics. This report adds 75 novel loci to the 175 previously implicated regions. In addition, this enormous amount of data are useful to refine previously reported associated signals (n=131 loci). The rs964184 signal was among them. The *APOC3* SNP rs138326449 was identified as the putative causal variant (P=3.57×10^{-138}). Thus, the approach applied by Hsueh et al[9] was capable to disclose the prominent role of *APOC3* loss-of-function variants on the main chromosome 11q23 signal as the exome sequencing did. However, the variant may differ between ethnic groups.

In summary, Native American populations should be considered as a source of information to identify the causal variants

of GWAS signals reported in whites or other ethnic groups. Furthermore, because next-generation sequencing becomes more affordable, whole-genome and -exome sequencing data will be soon available for many Native American populations. As observed with other phenotypes (ie, type 2 diabetes mellitus[15,16]), it is likely that ethnic-specific variants will be found, especially, those with low frequency with a large effect. More new questions will come from these reports.

Disclosures

None.

References

1. Dron JS, Hegele RA. Genetics of triglycerides and the risk of atherosclerosis. *Curr Atheroscler Rep*. 2017;19:31. doi: 10.1007/s11883-017-0667-9.
2. Braun TR, Been LF, Singhal A, Worsham J, Ralhan S, Wander GS, et al. A replication study of GWAS-derived lipid genes in Asian Indians: the chromosomal region 11q23.3 harbors loci contributing to triglycerides. *PLoS One*. 2012;7:e37056. doi: 10.1371/journal.pone.0037056.
3. Wang QL, Tan WL, Zhao YJ, Shao MM, Chu JH, Huang XD, et al. Data analysis in the post-genome-wide association study era. *Chronic Dis Transl Med*. 2016;2:231–234. doi: 10.1016/j.cdtm.2016.11.009.
4. Aguilar-Salinas CA, Tusie-Luna T, Pajukanta P. Genetic and environmental determinants of the susceptibility of Amerindian derived populations for having hypertriglyceridemia. *Metabolism*. 2014;63:887–894. doi: 10.1016/j.metabol.2014.03.012.
5. Hünemeier T, Amorim CE, Azevedo S, Contini V, Acuña-Alonzo V, Rothhammer F, et al. Evolutionary responses to a constructed niche: ancient Mesoamericans as a model of gene-culture coevolution. *PLoS One*. 2012;7:e38862. doi: 10.1371/journal.pone.0038862.
6. Weissglas-Volkov D, Aguilar-Salinas CA, Nikkola E, Deere KA, Cruz-Bautista I, Arellano-Campos O, et al. Genomic study in Mexicans identifies a new locus for triglycerides and refines European lipid loci. *J Med Genet*. 2013;50:298–308. doi: 10.1136/jmedgenet-2012-101461.
7. Ko A, Cantor RM, Weissglas-Volkov D, Nikkola E, Reddy PM, Sinsheimer JS, et al. Amerindian-specific regions under positive selection harbour new lipid variants in Latinos. *Nat Commun*. 2014;5:3983. doi: 10.1038/ncomms4983.
8. Parra EJ, Mazurek A, Gignoux CR, Sockell A, Agostino M, Morris AP, et al. Admixture mapping in two Mexican samples identifies significant associations of locus ancestry with triglyceride levels in the BUD13/ZNF259/APOA5 region and fine mapping points to rs964184 as the main driver of the association signal. *PLoS One*. 2017;12:e0172880. doi: 10.1371/journal.pone.0172880.
9. Hsueh W-C, Nair AK, Kobes S, Chen P, Göring HHH, Pollin TI, et al. Identity-by-descent mapping identifies major locus for serum triglycerides in Amerindians largely explained by an *APOC3* founder mutation. *Circ Cardiovasc Genet*. 2017;10:e001809. doi: 10.1161/CIRCGENETICS.117.001809.
10. Powell JE, Visscher PM, Goddard ME. Reconciling the analysis of IBD and IBS in complex trait studies. *Nat Rev Genet*. 2010;11:800–805. doi: 10.1038/nrg2865.
11. Schulz LO, Chaudhari LS. High-risk populations: the Pimas of Arizona and Mexico. *Curr Obes Rep*. 2015;4:92–98. doi: 10.1007/s13679-014-0132-9.
12. Fu W, Browning SR, Browning BL, Akey JM. Robust inference of identity by descent from exome-sequencing data. *Am J Hum Genet*. 2016;99:1106–1116. doi: 10.1016/j.ajhg.2016.09.011.
13. Aguilar-Salinas CA, Canizales-Quinteros S, Rojas-Martínez R, Mehta R, Rodriguez-Guillén R, Ordoñez-Sanchez ML, et al. The non-synonymous Arg230Cys variant (R230C) of the ATP-binding cassette transporter A1 is associated with low HDL cholesterol concentrations in Mexican adults: a population based nation wide study. *Atherosclerosis*. 2011;216:146–150. doi: 10.1016/j.atherosclerosis.2010.10.049.
14. Liu DJ, Peloso GM, Yu H, Butterworth AS, Wang X, Mahajan A, et al. Exome-wide association study of plasma lipids in > 300,000 individuals. *Nat Genet*. 2017;49:1758–1766. doi: 10.1038/ng.3977.
15. Williams AL, Jacobs SB, Moreno-Macías H, Huerta-Chagoya A, Churchhouse C, et al; SIGMA Type 2 Diabetes Consortium. Sequence variants in SLC16A11 are a common risk factor for type 2 diabetes in Mexico. *Nature*. 2014;506:97. doi: 10.1038/nature12828.
16. Estrada K, Aukrust I, Bjørkhaug L, Burtt NP, Mercader JM; SIGMA Type 2 Diabetes Consortium. Association of a low-frequency variant in HNF1A with type 2 diabetes in a Latino population. *JAMA*. 2014;311:2305. doi: 10.1001/jama.2014.6511.

Searching for the Causal Variants of the Association Between Hypertriglyceridemia and the Genome-Wide Association Studies–Derived Signals?: Take a Look in the Native American Populations

Carlos A. Aguilar-Salinas, Magdalena del Rocío Sevilla González and María Teresa Tusie-Luna

The online version of this article, along with updated information and services, is located on the
World Wide Web at:
http://circgenetics.ahajournals.org/content/10/6/e002010