Genome-Wide Studies of Type 2 Diabetes and Lipid Traits in Hispanics

Jennifer E. Below[1] · Esteban J. Parra[2]

Abstract Although disproportionately affected by increasing rates of type 2 diabetes and dyslipidemias, Hispanic populations are underrepresented in efforts to understand genetic susceptibility to these disorders. Where research has been undertaken, these populations have provided substantial insight into identification of novel risk-associated genes and have aided in the ability to fine map previously described risk loci. Genome-wide analyses in Hispanic and trans-ethnic populations have resulted in identification of more than 40 replicated or novel genes with significant effects for type 2 diabetes or lipid traits. Initial investigations into rare variant effects have identified new risk-associated variants private to Hispanic populations, and preliminary results suggest metagenomic approaches in Hispanic populations, such as characterizing the gut microbiome, will enable the development of new predictive tools and therapeutic targets for type 2 diabetes. Future genome-wide studies in expanded cohorts of Hispanics are likely to result in new insights into the genetic etiology of metabolic health.

Keywords Type 2 diabetes · Lipids · Genome-wide studies · Hispanics

✉ Jennifer E. Below
jennifer.e.below@uth.tmc.edu

1 The Human Genetics Center, University of Texas School of Public Health, Houston, TX, USA

2 Department of Anthropology, University of Toronto at Mississauga, Mississauga, ON, Canada

Introduction

Type 2 diabetes (T2D) and dyslipidemias are two major health concerns in Hispanic populations. When referring to Hispanics, we use the broad US census definition that states that "Hispanic or Latino refers to a person of Cuban, Mexican, Puerto Rican, South or Central American, or other Spanish culture or origin regardless of race," so it is important to note that within this group, there is substantial heterogeneity with regard to cultural, geographic, economic, environmental, lifestyle, and genetic factors that are known to impact diabetes and dyslipidemia susceptibility. In 2012, diabetes was the 5th leading cause of death among Hispanics in the USA (7th in the overall population) [1]. This disease has substantial economic costs (direct medical costs and reduced productivity), which were estimated to be $245 billion in 2012 [2], as well as profound societal costs in terms of reduced quality of life. The impact of diabetes is particularly strong among Hispanics, given its high prevalence in this group: during 2010–2012, the age-adjusted percentage of people aged 20 years or older with diagnosed diabetes was estimated as 12.8 % in Hispanics vs. 7.6 % in non-Hispanic whites [3]. Substantial variation in rates of diabetes has been reported for different Hispanic groups: 8.5 % for Central and South Americans, 9.3 % for Cubans, 13.9 % for Mexican Americans, and 14.8 % for Puerto Ricans [3]. Dyslipidemias are a major risk factor for cardiovascular disease (CVD), which was the 2nd leading cause of death in Hispanics (first in the overall US population) [1] in 2010. The prevalence of dyslipidemias is very high in Hispanics. For example, in a recent study based on a large sample ($N > 169,000$) from Northern California [4], individuals of Mexican ancestry had some of the highest prevalences of high low-density lipoprotein cholesterol (LDL-C) (56.8 %), low high-density lipoprotein cholesterol (HDL-C) (50.9 %), and high triglyceride (TG)

concentrations (45.4 %) of all the groups included in the study. Similarly to diabetes, there is a heterogeneous pattern of dyslipidemias among Hispanics. In a recent report of the Hispanic Community Health Study/Study of Latinos (HCHS-SOL), including Hispanic individuals of diverse origins (Cubans, Dominicans, Mexicans, Puerto Ricans, Central Americans, and South Americans) [5], 65 % of the participants had some form of dyslipidemia, but distinctive patterns were found between groups for some lipid traits. The highest prevalence of elevated LDL-C was found among Cubans (44.5 vs. 36 % in the full sample), and the highest prevalence of high TG was found among Central Americans (18.3 vs. 14.8 % in the full sample). It should be noted that the threshold used to define high TG concentrations in this study, >200 mg/dL, was higher than the threshold used in the previous reference, >130 mg/dL, and this explains the difference in prevalence between the datasets.

The substantial variation in the prevalence of diabetes and dyslipidemias reported in different Hispanic groups is not surprising, given the broad geographic origins and the diverse genetic and cultural backgrounds of individuals classified as Hispanic. Most of the populations of the Americas are primarily the result of an admixture process involving indigenous Americans, Africans, and Europeans, which started approximately 500 years ago. However, the relative genetic contributions from each of these parental groups show substantial geographic variation [6–8]. This can be clearly seen in Fig. 1, which depicts a principal component (PC) plot of four American samples (Mexican Americans, Puerto Ricans, Colombians, and Peruvians) from the 1000 Genomes Project (Phase 3) [9], together with European, African, and indigenous American samples. Differences in continental admixture proportions are evident in the plot defined by the first two PC axes (Fig. 1a). The Mexican American and Peruvian individuals primarily have indigenous American and European ancestry, although some Peruvian individuals also show evidence of African ancestry. The Peruvian sample on average has higher indigenous American ancestry than the Mexican American sample. In contrast to the Mexican American and Peruvian samples, the Colombian and Puerto Rican samples have higher African and lower indigenous American contributions (particularly the Puerto Rican sample). Other interesting patterns emerge when evaluating lower PC axes. For example, in the plot of the first and fourth PC axes (Fig. 1b), it is evident that all the American samples align with the Southern European samples (Spain and Italy), but not with the samples from Central and Western Europe (England, CEU sample) or Finland, in agreement with historical data. Finally, the plot of the first and seventh PC axes reveals that the Mexican American sample aligns with the indigenous samples from Meso-America (Nahua and Maya), and the Peruvian sample aligns with the indigenous samples of South America (Quechua from Peru and Aymara from Bolivia), as expected

given the geographic locations of the admixed and indigenous American samples. In contrast, neither the Puerto Ricans nor the Colombians align with the indigenous American samples. This is presumably because the indigenous samples included in this plot are not the proper representatives of the indigenous American groups involved in the admixture process that gave rise to the current Puerto Rican and Colombian populations. These PC plots are relevant because they show that (1) there is a broad range of individual admixture proportions in the American samples, which may require appropriate modelling in genetic association tests in order to avoid false positives, and (2) in addition to potential differences in average continental admixture proportions, there may also be differences between admixed populations at the intra-continental level. The variation in genetic structure, particularly at the continental level, in admixed samples throughout the Americas can has practical consequences for genome-wide association (GWA) studies. It is well known that linkage disequilibrium (LD) patterns differ between continental populations [10]. Populations of African ancestry tend to have lower levels of LD than non-African populations, which have undergone bottlenecks and serial founder effects during and after the out-of-Africa migrations. Additionally, allele frequencies may differ across populations [11]. When performing GWA studies, differential allele frequency or LD patterns between populations may have analytical consequences in multiple study phases including discovery, replication, and fine mapping [12••, 13].

Sociocultural and economic factors are also key to understanding the higher prevalence of diabetes and dyslipidemias in Hispanics as a group, and the differences in prevalence observed between Hispanics of diverse origins. In this respect, significant differences in education level and income exist between Hispanics and non-Hispanic whites [14]. Differences in income, education, and several variables used to assess acculturation (e.g., migration, language use and ability, ethnic identity, acculturative distress, and family experience) have also been described between Hispanic groups [15]. In the HCHS-SOL study focused on type 2 diabetes in Hispanics [16], the investigators reported that the number of years living in the USA, education, and household income had a significant effect on diabetes risk. Similarly, in the HCHS-SOL studies focused on dyslipidemia [5], language preference, income, and education were associated with dyslipidemia prevalence in Hispanics.

There have been numerous candidate gene studies in Hispanics focused on type 2 diabetes and lipid traits. However, this review will only focus on the more recent studies undertaken at the genome-wide level. We will describe the main results of linkage and GWA studies in Hispanics and discuss these results in the context of the findings reported in other population groups. We also highlight other genomic approaches that offer great promise to uncover risk factors for type 2 diabetes and dyslipidemias, such as metagenomic analyses.

a

b

c
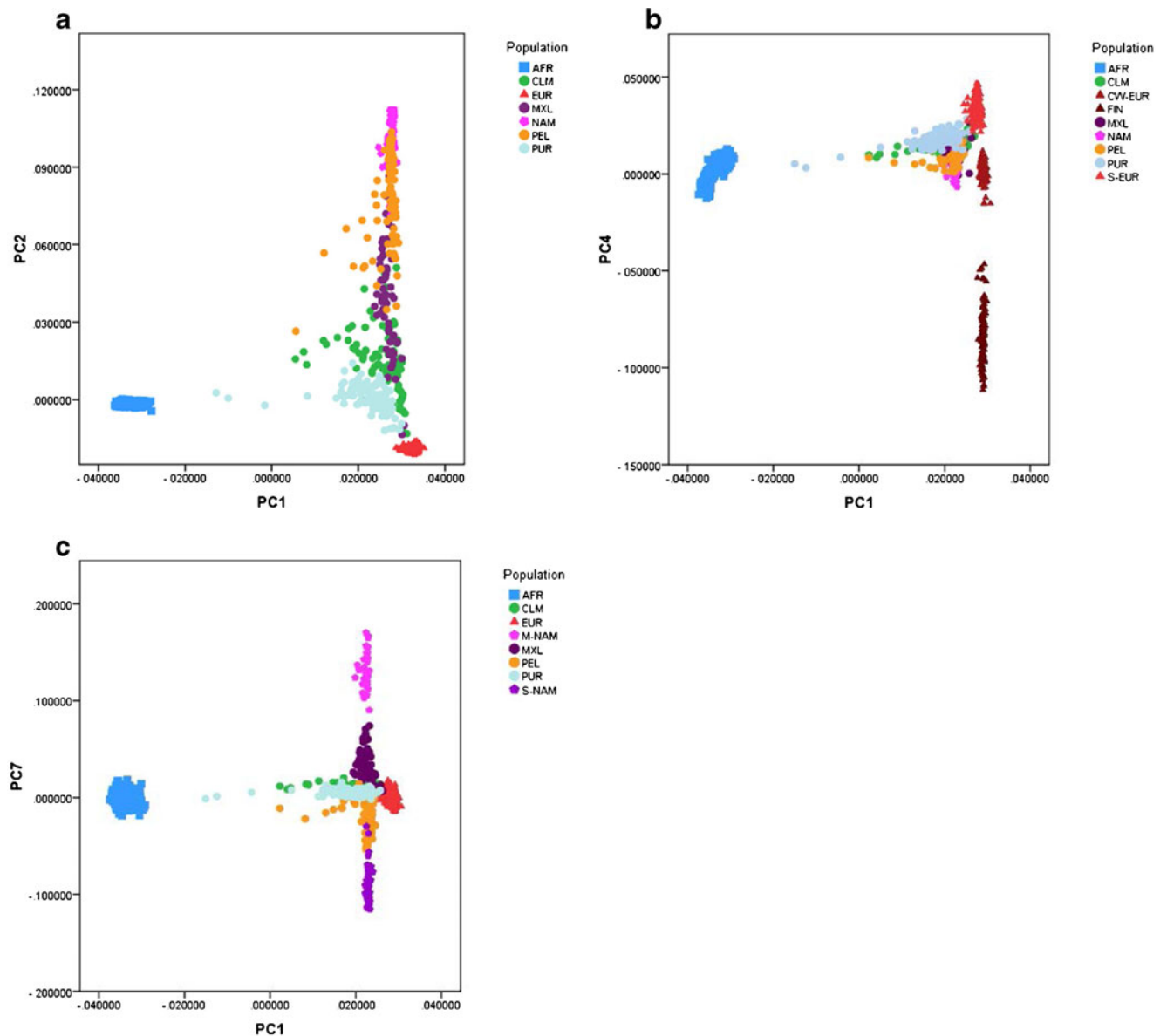
Fig. 1 Principal component (PC) plot of the four admixed American populations of the 1000 Genomes Project Phase 3 panel (Colombians, Mexican Americans, Peruvians, and Puerto Ricans), as well as African (Esan from Nigeria, Gambians, Luhya from Kenya, Mende from Sierra Leone, and Yoruba from Nigeria), European (Finnish, British, CEU, Italians, and Spaniards), and indigenous American (Nahua and Maya from Mexico, Quechua from Peru, and Aymara from Bolivia) samples. The PC plot was done with the program Eigenstrat, after pruning the markers based on linkage disequilibrium ($r^2 > 0.1$) and removing four genomic regions on chromosomes 2, 6, 8, and 17. The total number of markers used for the PC plot was ~98,500. The eigenvectors were inferred using the individuals in the parental populations (African, European, and indigenous American), and then all the individuals were projected onto those eigenvectors. **a** Plot of PC axes 1 and 2. *AFR* African samples (*blue squares*), *EUR* European samples (*red triangles*), *NAM* indigenous American samples (*pink pentagons*). *CLM* (Colombians, *green circles*), *MXL* (Mexican Americans, *purple circles*), *PEL* (Peruvians, *orange circles*), *PUR* (Puerto Ricans, *light blue circles*). **b** Plot of PC axes 1 and 4. In this plot, the European samples are separated in three groups: Southern Europeans (Spaniards and Italians), Central and Western Europeans (British and CEU), and Finnish. **c** Plot of PC axes 1 and 7. In this plot, the indigenous American samples are separated in two groups: Mesoamericans (Nahua and Maya) and South Americans (Quechua and Aymara)

Genome-Wide Analyses of Type 2 Diabetes in Hispanic Populations

In the mid-1990s, a critical nexus was reached: the combination of sufficiently efficient computational methodologies [17–19] and comprehensive, well-mapped interrogations of the human genome [20, 21] gave rise to an era of genome-wide searches for variation contributing to genetically complex traits. A genome-wide linkage scan for type 2 diabetes in Mexican Americans from Starr County, TX, utilized affected sib-pair approaches and successfully identified the *NIDDM1* region on chromosome 2 in 1996 [22]. The effects of

NIDDM1 were further found to be interacting with variation on chromosome 15 to increase risk of type 2 diabetes in Mexican Americans [23]. In 2000, a gene in the *NIDDM1* region, *calpain10*, was identified as a diabetes susceptibility gene [24], and although *calpain10* has not been replicated by GWA studies, Capn10 knockout experiments in mice have demonstrated that the gene plays an important role in diabetes and obesity [25]. Concurrently, genome-wide linkage analyses of type 2 diabetes and age at onset in 27 extended Mexican American pedigrees identified a significantly linked region on the q arm chromosome 10, which was later found to harbor *TCF7L2* [26, 27]. The highest reported prevalence of diabetes in the world is in the indigenous Pima population in Arizona [28]. Genome-wide linkage scans on sibling pairs from this population in the mid-1990s identified a region on 11q23-24 significantly linked to BMI and showing evidence of pleiotropic effects on type 2 diabetes [29, 30]. Since this time, efforts to understand the complex underlying genetic etiology of type 2 diabetes have shifted to ever larger-cohort analyses of ever-increasing amounts of genetic variation.

Today, more than 100 loci for type 2 diabetes and glycemic traits have been identified through numerous GWA studies of common and rare variation in populations of diverse ancestral origins [31]; however, to date, very few GWA studies have been published in cohorts of Mexican ancestry. The first GWA study performed in a non-European cohort was published in 2007 and comprised 561 Mexican American type 2 diabetes cases and controls drawn from the Starr County Health Studies [32]. Although no loci reached genome-wide significance, several loci identified in prior GWA studies in Europeans were replicated [32]. This analysis was subsequently expanded ($N = 1273$) and meta-analyzed with a cohort from Mexico City ($N = 1310$) in 2011 [33, 34]. The most significant variants observed in this meta-analysis included known regions near *HNF1A* and *KCNQ1*. Top association signals were then meta-analyzed with the DIAGRAM and DIAGRAM+ datasets of European ancestry individuals, resulting in two regions reaching genome-wide significance: *HNF1A* and *CDKN2A/CDKN2B* (Table 1). Top association signals in both studies were annotated to explore their roles as expression quantitative trait loci (eQTL) in both adipose and muscle tissues, revealing a marked excess of trans-acting eQTL in top signals in both tissue types.

Recently, using chip-based and imputed genotypes, the Slim Initiative in Genomic Medicine for the Americas (SIGMA) Type 2 Diabetes Consortium identified a novel risk haplotype defined by five SNPs in the gene *SLC16A11* in a Mexican and Latin American cohort ($N = 8214$) [35••] (Table 1). The 73-kb haplotype, which is common in indigenous American and East Asian populations, but rare in European and African individuals, was simultaneously identified in a Japanese cohort [36]. Investigations into the origin of this haplotype suggested that based on length, similarity to

Neandertal sequence, and an estimated coalescence that post-dates the estimated modern human and Neandertal evolutionary split, it entered modern human lineages through introgression via admixture with Neandertals. *SLC16A11* is expressed in the liver, and expression in HeLa cells resulted in significant increases in triacylglycerol levels, suggesting that *SLC16A11* may increase risk to type 2 diabetes through hepatic lipid metabolism [35••].

Several groups are now publishing analyses of low frequency and rare variant effects on type 2 diabetes risk in Hispanic and indigenous American populations [37•, 38], and more can be expected in the near future. The SIGMA consortium recently performed whole-exome sequencing on a Hispanic sample comprising 3756 individuals and identified one rare missense variant in *HNF1A* which, based on attempted replication in multiple ancestral groups, appears to be private to Hispanic populations. Characterization in experimental assays confirmed reduced activation of the mutant HNF1A protein on its target promoter. The SIGMA studies highlight the advantages of investigating type 2 diabetes in large cohorts drawn from understudied populations, such as Hispanic groups, in which allele frequency differences may provide increased power to detect novel risk variants [35••].

In an effort to explore the effects of genes influencing glucose homeostasis on type 2 diabetes risk in Hispanic populations, the Genetics Underlying Diabetes in Hispanics (GUARDIAN) Consortium was assembled, including seven Mexican-origin cohorts without type 2 diabetes in a discovery phase and six Mexican-origin cohorts including diabetes cases and controls in a translation phase, for a total sample of 19,871 [39••]. This strategy of utilizing measures closer to gene action may result in increased power to detect genes relevant to type 2 diabetes risk. Genomic regions associated with measures of glucose homeostasis in the discovery group were carried forward and evaluated for association with the clinical outcome of type 2 diabetes in the translation cohorts. Using this approach, GUARDIAN identified both novel and replicated known risk variants (Table 1). The most significant findings localized to known genes *KCNQ1*, *CDKAL1*, *CDKN2A/B*, and *MTNR1B*. *MTNR1B*, a gene previously implicated in type 2 diabetes and fasting glucose in populations of European descent, was found in this analysis to be associated with acute insulin response. *MTNR1B* is expressed in human islet cells, co-localizes with insulin, and is thought to inhibit glucose-stimulated insulin secretion [40, 41]. These analyses demonstrate the complex, interacting, and pleiotropic nature of type 2 diabetes and quantitative measures that underlie it.

Despite comparatively limited cohort sizes, analyses of type 2 diabetes risk in Hispanic populations have driven diabetes gene discovery by leveraging high disease prevalence, population-specific haplotypic variation, and a private mutation spectrum. There is evidence that these findings are relevant across ancestry: effects of variation in Hispanic

Table 1 Genome-wide significant regions identified in GWA studies for type 2 diabetes in Hispanics

Gene	CHR	POS	LEAD SNP	P-META	Trait	Ref
FAF1	1	50682573	rs17106184	4.10×10^{-9}	Trans-ethnic meta-analysis	Mahajan et al. [12••]
LPP	3	189223217	rs6808574	5.80×10^{-9}	Trans-ethnic meta-analysis	Mahajan et al. [12••]
TMEM154	4	153739925	rs6813195	4.10×10^{-14}	Trans-ethnic meta-analysis	Mahajan et al. [12••]
ARL15	5	53307177	rs702634	6.90×10^{-9}	Trans-ethnic meta-analysis	Mahajan et al. [12••]
SSR1/RREB1	6	7235436	rs9502570	1.40×10^{-9}	Trans-ethnic meta-analysis	Mahajan et al. [12••]
SLC35B3/TFAP2A	6	8998811	rs2064197	2.56×10^{-8}	SG	GUARDIAN [39••]
CDKAL1	6	20694884	rs2206734	1.11×10^{-8}	AIRg	GUARDIAN [39••]
POU5F1/TCF19	6	31244432	rs3132524	4.20×10^{-9}	Trans-ethnic meta-analysis	Mahajan et al. [12••]
CDKN2A/CDKN2B	9	22126489	rs1333051	9.6×10^{-10}	Type 2 diabetes (~BMI)	Parra et al.* [33]
TCF7L2	10	114758349	rs7903146	2.47×10^{-17}	Type 2 diabetes liability (~BMI)	SIGMA [35••]
KCNQ1	11	2858546	rs2237897	1.24×10^{-21}	DI	GUARDIAN [39••]
KCNQ1	11	2858546	rs2237897	4.94×10^{-16}	Type 2 diabetes liability (~BMI)	SIGMA [35••]
KCNQ1	11	2824776–2824783	rs139647931	5.30×10^{-8}	Type 2 diabetes liability (~BMI, known_var)	SIGMA [35••]
MTNR1B‖	11	92708710	rs10830963	5.86×10^{-9}	AIRg	GUARDIAN [39••]
HNF1A	12	121402932	rs7305618	1.9×10^{-8}	Type 2 diabetes (~BMI)	Parra et al.* [33]
MPHOSPH9	12	122013881	rs1727313	9.50×10^{-9}	Trans-ethnic meta-analysis	Mahajan et al. [12••]
SLC16A11/SLC16A13	17	7042968	rs13342692	2.40×10^{-12}	Type 2 diabetes liability (~BMI)	SIGMA [35••]

Parra et al.: Type 2 diabetes (~BMI) meta-analysis of logistic regression, adjusted for BMI; *markers with combined p values of $p < 10^{-4}$ in the Hispanic and 2008 DIAGRAM or DIAGRAM+ datasets

SIGMA: Type 2 diabetes liability (~BMI) type 2 diabetes liability based on prevalence of type 2 diabetes (type 2 diabetes) at different intervals of age and body mass index (BMI)

SIGMA: type 2 diabetes liability (~BMI, known var), rs139647931 adjusted for rs2237897, rs11564732 adjusted for rs2237897 and rs139647931

GUARDIAN: Top regions from GWAS of type 2 diabetes-related quantitative traits with translation to type 2 diabetes in Mexican-origin cohorts: Insulin sensitivity (SI from FSIGT or glucose infusion rate M from clamp as well as a meta-analysis combining these, denoted as SI + M), metabolic clearance rate of insulin (MCRI), acute insulin response (AIRg), disposition index (DI), and glucose effectiveness (SG)

Mahajan et al.: Novel findings from a trans-ethnic "discovery" GWAS meta-analysis of 26,488 cases and 83,964 controls of European, East Asian, South Asian, and Mexican and Mexican American ancestry; reported p values are from follow-up in a "validation" meta-analysis of an additional 21,491 cases and 55,647 controls of European ancestry

populations are significantly directionally consistent with analyses in European ancestry, even at fairly modest levels of significance ($p < 0.01$) [12••, 42, 43]. Furthermore, due to differential LD structure, inclusion of Hispanic populations in trans-ethnic fine mapping and meta-analyses provides an opportunity to narrow windows of association and localize causal alleles [12••].

As noted by Below et al. and others [34, 44], there is a significant enrichment of eQTLs among top type 2 diabetes-associated loci. Genetic heritability estimates for type 2 diabetes are markedly higher than can be explained by the variation identified to date; to characterize this "missing" heritability, Torres et al. composed multiple SNP subsets by partitioning interrogated maker sets into groups by status as eQTL in several insulin-responsive peripheral tissues [45•]. They discovered that these subsets explain a greater portion of type 2 diabetes risk than expected by chance, suggesting a significant role of regulatory variation in diabetes susceptibility. Several reasons have been suggested as to why so much of

the genetic heritability of type 2 diabetes remains unmapped to risk loci [46]. Conclusive identification of less common (0.5–5 % MAF) variation of modest effect will require investments in extremely large sample sizes. The heterogeneous nature of Hispanic populations increases the challenge because to detect variation or effects specific to groups or environments may require sample sizes beyond what exist to be collected. There is evidence that parent of origin may influence effects of variants on type 2 diabetes risk [47]. Studies in mouse models also demonstrate that some genetic effects on type 2 diabetes and related traits are modified by sex, diet, and epigenetic effects, indicating that careful environmental modelling and stratification will be necessary to identify some loci subject to interaction effects [48]. Genetic characterizations of larger Hispanic samples are underway, but especially in the case of extremely rare or private variation, a return to family-based study designs will improve power through enrichment of allelic observations and increased environmental and genetic homogeneity [49].

Genome-Wide Analyses of Blood Lipid Levels in Hispanic Populations

As described for type 2 diabetes, genome-wide studies focused on lipid traits in Hispanic populations have been scarce, particularly when compared with studies in European populations. The initial genome-wide studies were linkage studies in Hispanic families. In one of the earliest studies, Rainwater et al. [50] identified two quantitative trait loci (QTL) on chromosomes 3 and 4 that influenced variation in cholesterol concentrations of small LDL particles. Subsequent studies reported evidence of linkage on chromosome 11 for TG/HDL-C ratio [51], chromosome 2 and 16 for total cholesterol (TC) [52, 53], chromosome 12 and 15 for TG [54, 55], and chromosome 7 for LDL-C/HDL-C ratio [53]. The aforementioned studies used a few hundred highly polymorphic microsatellites spanning the human genome, and the linkage signals were quite broad. Two of the studies followed up the linkage signals with association analyses, reporting associations of markers in the gene *ADIPOR2* with TG on chromosome 12 [54], and markers near the gene *FLJ45974* with LDL-C/HDL-C ratio [53]. More recently, Hellwege et al. [56•] published a much denser family-based linkage analysis using markers of the Illumina HumanExome Beadchip. This study described 104 regions with LOD scores greater than 3 in Hispanic families. The region showing the strongest evidence of linkage was the known *CETP* gene for HDL-C, and multiple variants within this gene also showed strong associations with this trait [56•].

GWA studies for lipid traits in Hispanics using dense microarrays have only appeared in the literature relatively recently. Table 2 reports the genome-wide association signals that have been described in studies focused on Hispanic populations. Comuzzie et al. [57] carried out a family-based GWA study for obesity-related traits in children participating in the VIVA LA FAMILIA study, and reported a significant association of variants in the *APOA5-ZNF259* region with TG concentrations.

In 2013, Coram et al. [58•] performed a GWA study for lipid traits in a cohort of 3642 Hispanic participants from the Women's Health Initiative SNP Health Association Resource (WHI-SHARe) and reported genome-wide significant signals within or near the genes *GCKR*, *LPL*, and *APOA/APOC* for TG and *CETP* and *APOA/APOC* for HDL-C. These authors also showed that there is a substantial overlap in the genes associated with lipid traits in different population groups. When testing the markers showing genome-wide significance or suggestive evidence of association ($p \leq 10^{-5}$) in European GWA studies in the Hispanic and African American WHI cohorts, a strong enrichment of small p values was observed in both cohorts. Additionally, there was a significant correlation of the allelic effects of markers with $p \leq 10^{-5}$ identified in Europeans in the Hispanic and African American cohorts. The genomic regions showing association in Europeans accounted for a disproportionate amount of variance in both cohorts.

Weissglas-Volkov et al. [59•], also in 2013, performed a GWA study for HDL-C and TG in a sample from Mexico ($N = 4361$). Genome-wide significant signals were observed in seven regions previously described in European populations: *APOA5*, *GCKR*, and *LPL* for TG and *ABCA1*, *CETP*, *LIPC*, and *LOC55908* for HDL-C. A novel genome-wide significant region was also found near the *NPC1* gene for TG. In agreement with the findings of Coram et al.[58•], Weissglas-Volkov et al. also observed a highly significant concordance excess in the directionality of effects in Europeans and Hispanics for 100 SNPs that had been identified in European populations. However, the authors also noted that for 82 of the 100 loci, the SNP with the strongest evidence of association in the Mexican sample was not the same as the lead SNP/best-proxy SNP reported in the European sample, and the LD between the European lead SNP and the SNP showing the strongest association in the Mexican sample was low ($r^2 < 0.3$), suggesting different haplotypic structures or contributing variants underlying risk at these regions in their sample. Using cross-ethnic mapping, the authors were able to narrow several credible regions identified in the European studies (*APOA5*, *MLXIPL*, and *CILP2*). In particular, in the *APOA5* region, the number of likely susceptibility variants was reduced to only one, rs964184.

In 2014, Ko et al. [60•] used a different analytical approach from that used in previous efforts, by including only variants showing differences in frequency between European and Native American populations in the GWA study. This study comprised 3323 individuals in the discovery effort and 6159 additional samples in the replication stage, and reported genome-wide significant signals in regions already described in European studies (*CETP* for HDL-C, *LPL* and *ZNF259* for TG, and *CELSR* and *CETP* for TC). Novel signals were identified within or near the genes *UGT8* and *RORA* (for HDL-C) and *SIK3* (for HDL-C and TG). The signal observed in *SIK3* was close to the *APOA5/ZNF259* region on chromosome 11 and corresponds to an intronic SNP that is common in Mexicans but not observed in a Finnish sample [60•]. An interesting finding of this study was that in two regions associated with TG (*LPL* and *APOA5/ZNF259*), there was an excess of indigenous American ancestry in Mexican individuals with high TG concentrations with respect to individuals with low TG concentrations. In order to carry out this analysis, the authors used the program LAMP-LD [61] to infer locus ancestry in the relevant genomic regions.

Very recently, Below et al. [62•] published the results of a meta-analysis of lipid traits in 4383 individuals of Mexican ancestry. The genome-wide significant and suggestive associations were followed up in three additional Hispanic samples comprising 7876 individuals. In this study, genome-wide significant signals were identified in or near *CELSR2*,

Table 2 Genome-wide significant regions identified in GWA studies for lipid traits in Hispanics

Gene	CHR	POS[a]	LEAD SNP	P-META	Ref
TCHOL					
CELSR2	1	109817192	rs7528419	9.05×10^{-27}	Below et al. [62•]
	1	109819296	rs3902354	1.16×10^{-8}	Ko et al. [60•]
FN1	2	216304384	rs1250229	1.04×10^{-8}	Below et al. [62•][d]
ZNF259/APOA5	11	116648917	rs964184	4.88×10^{-13}	Below et al. [62•]
CETP	16	57000938	rs118146573	3.79×10^{-10}	Ko et al. [60•]
KANK2/DOCK6	19	11300365	rs138534124	7.86×10^{-9}	Below et al. [62•]
NCAN	19	19336608	rs2238675	1.23×10^{-13}	Below et al. [62•]
SAMM50	22	44377999	rs2235776	2.85×10^{-8}	Below et al. [62•][d]
HDL					
UGT8/NDST4	4	115638601	rs78557978	4.09×10^{-8}	Ko et al. [60•]
LOC100996634	6	109574095	rs884366	1.40×10^{-8}	Below et al. [62•][d]
LPL	8	19891227	rs28526159	8.41×10^{-9}	Below et al. [62•]
ABCA1	9	107601541	rs2472386	8.37×10^{-11}	Below et al. [62•]
	9	107589134	rs4149310	5.54×10^{-8}	Weissglas-Volkov et al. [59•][b]
	9	107620835	rs9282541	6.4×10^{-26}	Weissglas-Volkov et al. [59•][b]
COPB1	11	14504463	rs7121538	2.11×10^{-8}	Below et al. [62•][d]
ZNF259/APOA5	11	116581641	rs2367970	8.42×10^{-10}	Below et al. [62•]
	11	116648917	rs964184	2.81×10^{-12}	Coram et al. [58•]
SIK3	11	116884789	rs11216230	3.26×10^{-10}	Ko et al. [60•]
LIPC	15	58726744	rs261334	9.40×10^{-13}	Below et al. [62•]
	15	58723426	rs1077835	2.1×10^{-14}	Weissglas-Volkov et al. [59•]
RORA	15	61244884	rs148533712	3.41×10^{-8}	Ko et al. [60•]
CETP	16	57006590	rs7499892	3.80×10^{-49}	Below et al. [62•]
	16	56990716	rs247617	3.48×10^{-16}	Coram et al. [58•]
	16	57005479	rs1532624	1.39×10^{-24}	Weissglas-Volkov et al. [59•]
	16	56997349	chr16:56997349:I	6.75×10^{-20}	Ko et al. [60•]
DOCK6	19	11350488	rs2278426	3.44×10^{-9}	Weissglas-Volkov et al. [59•]
LDL					
CELSR2	1	109817838	rs660240	1.25×10^{-28}	Below et al. [62•]
APOB	2	21217490	rs13392272	2.16×10^{-12}	Below et al. [62•]
PFAS	17	8161149	rs4791641	1.13×10^{-8}	Below et al. [62•][d]
NCAN	19	19336608	rs2238675	2.33×10^{-8}	Below et al. [62•]
TRIG					
GCKR	2	27742603	rs780093	2.82×10^{-15}	Below et al. [62•]
	2	27741237	rs780094	7.35×10^{-9}	Coram et al. [58•]
	2	27730940	rs1260326	2.2×10^{-13}	Weissglas-Volkov et al. [59]
LPL	8	19848080	rs17410962	7.35×10^{-9}	Coram et al. [58•]
	8	19844222	rs12678919	2.7×10^{-10}	Weissglas-Volkov et al. [59•]
	8	19860460	rs79236614	3.79×10^{-8}	Ko et al. [60•]
TRIB1	8	126491733	rs2954031	3.26×10^{-13}	Below et al. [62•]
ZNF259/APOA5	11	116648917	rs964184	2.79×10^{-83}	Below et al. [62•]
	11	116648917	rs964184	3.66×10^{-33}	Coram et al. [58•]
	11	116648917	rs964184	5.5×10^{-35}	Weissglas-Volkov et al. [59•]
	11	116648917	rs964184	6.08×10^{-33}	Ko et al. [60•][c]

Table 2 (continued)

Gene	CHR	POS[a]	LEAD SNP	P-META	Ref
	11	116657561	rs3741298	2.47×10^{-8}	Comuzzie et al. [57]
SIK3	11	116807343	rs139961185	1.15×10^{-12}	Ko et al. [60•][c]
TMEM241	18	20879217	rs9949617	2.43×10^{-8}	Weissglas-Volkov et al. [59•]
MAU2	19	19455750	rs8102280	3.38×10^{-18}	Below et al. [62•]

[a] Position corresponds to GRCh37 assembly

[b] Two signals at ABCA1 for HDL were described as independent by Weissglas-Volkov et al.

[c] Two signals on chromosome 11 for TG were described as independent by Ko et al.

[d] Identified in a meta-analysis of the Hispanic and European datasets

ZNF259/APOA5, KANK2/DOCK6, and NCAN/MAU2 for TC; LPL, ABCA1, ZNF259/APOA5, LIPC, and CETP for HDL-C; CELSR2, APOB, and NCAN/MAU2 for LDL-C; and GCKR, TRIB1, ZNF259/APOA5, and NCAN/MAU2 for TG. All these regions have been previously described in GWA studies in European populations. However, based on LD and conditional analyses, the authors reported that the lead SNPs observed in the samples of Mexican ancestry for some of the regions (ABCA1 and LIPC for HDL-C, and NCAN/MAU2 for TG) seem to be independent of the lead SNPs described in the European studies. Combining the Mexican dataset with the European Global Lipids Genetics Consortium (GLGC) [63] dataset, five novel genome-wide significant regions were identified in or near the genes FN1 and SAMM50 for TC, LOC100996634 and COPB1 for HDL-C, and LINC00324/CTC1/PFAS for LDL-C. In agreement with the findings of Coram et al. [58•] and Weissglas-Volkov et al. [59•], Below et al. [62•] also observed strong concordance in direction of effects and little evidence of heterogeneity in effect sizes when comparing lead SNPs in the European and Mexican samples. Finally, the authors also carried out genome-wide association analyses focused on eQTLs reported for four different cell types (lymphoblastoid cell lines (LCL), human liver, adipose, and muscle). These analyses uncovered genome-wide significant signals that were not observed in the original GWA study (LPL, LOC102467079, and LIPC for HDL-C; DOK6 for LDL-C; and TRIB1 for TG). Interestingly, several of the significant eSNPs identified in the meta-analysis were the lead SNPs or had p values close to those of the lead SNP in the meta-analysis (e.g., SNPs in the NCAN, CETP, and CELSR2 regions), suggesting that at least some of the top signals observed in the meta-analysis may have a functional effect through regulation of gene expression. Finally, SNPs showing low p values in the meta-analysis were enriched for eQTLs in most of the cell types (except for LDL-C and TC in LCL, which showed no evidence of enrichment), but the levels of enrichment were tissue dependent. These findings emphasize the importance of using a wide variety of tissue types to understand and interpret the possible functional relevance of genetic effects, and suggest tissue-specific roles for some of these eQTLs.

Metagenomic Approaches in Hispanic Populations

In the previous sections, we highlighted the main findings of genome-wide studies for type 2 diabetes and lipid traits focusing on the human genome. However, there are other genomes that are also relevant to study in order to discover potential risk factors for type 2 diabetes and dyslipidemias. In this respect, in the last 5 years, researchers have uncovered interesting associations of metagenomic profiles with these traits. Most of these studies have focused on the gut microbiota. There is substantial evidence indicating that the microbiome may influence the development of diabetes (for relevant examples, see Musso et al. [64]; Amar et al. [65]; Cani et al. [66]; Qin et al. [67]; Karlsson et al. [68••]; Cai et al. [69]). Importantly, it has been shown that the metagenomic markers for type 2 diabetes may differ between populations. Karlsson et al. [68••] reported that the most discriminatory metagenomic clusters identified in a European sample were different to those of a Chinese cohort. These authors highlighted the need to develop metagenomic predictive tools for type 2 diabetes specifically for the geographical location of the populations studied. More recent research has also demonstrated the involvement of the gut microbiome in lipid levels. Fu et al. [70] identified associations of bacterial taxa with blood lipid concentrations and reported that microbiota explained 6 % of the TG variance and 4 % of the HDL variance, independently of sex, age, and genetic risk factors. Recently, Ross et al. [71] characterized the gut microbiome of 63 individuals from the Cameron County Hispanic Cohort (CCHC) by sequencing the 16S rRNA gene. No associations were observed between any of the taxa identified and diabetes status, cholesterol, or TG. However, this study identified significant differences for several taxa (at the phyla, family, and genus level) between the CCHC and the Human Microbiome Project (HMP) stool data,

which was used as a sample representative of a healthy Western microbiome. Interestingly, the shifts in the microbial community identified in CCHC vs. HMP are similar to those observed in studies focused on obese individuals and those with type 2 diabetes. Other gut microbiome studies are currently under way in Hispanic populations, and they offer great promise to develop new predictive tools and potential new targets for type 2 diabetes treatment [72, 73].

Conclusion

In this review, we have highlighted the major findings reported in genome-wide studies in Hispanic populations. Although the number of studies has been limited, and the sample sizes are still relatively small in comparison to published studies in European populations, these efforts have provided important insights about the architecture of type 2 diabetes and lipid traits. GWA studies in Hispanics have identified novel susceptibility regions for type 2 diabetes and lipid traits, which require replication in subsequent studies with larger sample sizes to confirm effects. In this respect, it is important to highlight that replication efforts in Hispanics are complicated by the known heterogeneity in ancestral background in this group. Additionally, these studies have uncovered independent signals in some of the regions previously described in other populations. Some of these signals, such as SLC16A11 [35••], RORA, and SIK3 [60•], may be population specific. Although there is clear evidence of allelic heterogeneity for some regions, in general, a substantial concordance has been described in the direction and size of effects between Hispanics and European populations, both for type 2 diabetes and lipid traits. This is something that has been also reported for other population groups [12••, 74, 75]. In particular, the recent trans-ethnic meta-analysis of type 2 diabetes, which included samples from European, East Asian, South Asian, and Hispanic populations, has provided interesting insights about the disease [12••], some of which may also be applicable to lipid traits. In the trans-ethnic meta-analysis of type 2 diabetes, a significant directional concordance in effect was observed between ethnicities for markers with $p \leq 0.001$ or $p \leq 0.01$ in the European meta-analysis, but this was not the case for markers showing no association ($p > 0.05$). Additionally, there was no evidence of heterogeneity of allelic effects at the majority of established loci for the disease. As pointed out by the authors, this suggests that (1) many other common variants with small effects remain to be discovered, but this will require larger sample sizes, and (2) it is improbable that the signals observed in most of the susceptibility regions described so far are due to rare variants (the "synthetic association hypothesis"), because these variants are typically not shared between population groups and the patterns of

LD with rare variants would be expected to vary between ethnicities (in addition, the largest whole-exome sequencing dataset for type 2 diabetes published to date, that of the SIGMA Consortium [37•], has failed to find supportive evidence for that hypothesis). The trans-ethnic meta-analysis of type 2 diabetes also showed that, for regions showing shared signals between populations, combining samples from different population groups may result in better fine-mapping resolution, due to the differences in LD patterns between population groups, in agreement with findings for other traits, including lipids [76–78]. Finally, the analyses focused on gene expression data in Hispanics have shown that this is an insightful approach from different perspectives: (1) statistical power is increased due to reduced penalty for multiple testing with respect to standard GWA studies, (2) the enrichment in eQTLs observed for genetic markers showing associations with type 2 diabetes or lipid levels suggests that some of the susceptibility regions may be due to variants having an effect on transcriptional regulation, and (3) it highlights the importance of studying tissue-specific effects in order to understand the potential mechanisms behind some of the observed associations.

It is important to note that most of the signals described in this review correspond to common markers. An exception has been the recent whole-exome sequencing study that identified the association of a rare missense variant in HNF1A with type 2 diabetes [37•]. The limited sample sizes used in the discovery efforts in Hispanic populations, typically smaller than 5000 individuals, render these studies underpowered to identify low-frequency variants (variants with a frequency between 0.5 and 5 %). Elucidating the role that low-frequency variants have on complex traits and diseases in Hispanics will require studies based on much larger sample sizes and very dense datasets (e.g., whole-genome or whole-exome sequencing studies, or studies using datasets imputed with dense reference panels). Some recent large-scale studies in European populations are providing important insights into the impact of low-frequency and rare variants on type 2 diabetes [79, 80] and lipid traits [81]. Ongoing meta-analytic efforts based on dense imputed data (e.g., data imputed with Phase 3 panel of the 1000 Genomes Project [82] or the dense haplotype panels created by the Haplotype Reference Consortium (http://www. haplotype-reference-consortium.org)) with the participation of many consortia and tens of thousands of samples are currently under way in Hispanics, not only for type 2 diabetes and lipid traits but also for anthropometric traits (body mass index, height, and waist-hip ratio) and blood pressure. These expanded efforts, in combination with those in other population groups, and future trans-ethnic meta-analyses will bring a better understanding of the genetic architecture of these traits, by uncovering more novel common and low-frequency variants, identifying population-specific effects and allelic and locus heterogeneity, and highlighting causal

variants taking advantage of better fine-mapping resolution and better annotation of the human genome.

Compliance with Ethics Guidelines

Conflict of Interest   Jennifer E. Below and Esteban J. Parra declare that they have no conflict of interest.

Human and Animal Rights and Informed Consent   This article does not contain any studies with human or animal subjects performed by any of the authors.

References

Papers of particular interest, published recently, have been highlighted as:
• Of importance
•• Of major importance

1. Heron M. Deaths: leading causes for 2012. Natl Vital Stat Rep. 2015;64(10):1–94.
2. American Diabetes Association. Economic costs of diabetes in the U.S. in 2012. Diabetes Care. 2013;36(4):1033–46.
3. Centers for Disease Control and Prevention: National Diabetes Statistics Report: estimates of diabetes and its burden in the United States, 2014. 2014, U.S. Department of Health and Human Services: Atlanta, GA.
4. Frank AT, Zhao B, Jose PO, et al. Racial/ethnic differences in dyslipidemia patterns. Circulation. 2014;129(5):570–9.
5. Rodriguez CJ, Daviglus ML, Swett K, et al. Dyslipidemia patterns among Hispanics/Latinos of diverse background in the United States. Am J Med. 2014;127(12):1186–94. e1.
6. Galanter JM, Fernandez-Lopez JC, Gignoux CR, et al. Development of a panel of genome-wide ancestry informative markers to study admixture throughout the Americas. PLoS Genet. 2012;8(3), e1002554.
7. Bryc K, Velez C, Karafet T, et al. Colloquium paper: genome-wide patterns of population structure and admixture among Hispanic/Latino populations. Proc Natl Acad Sci U S A. 2010;107 Suppl 2:8954–61.
8. Wang S, Ray N, Rojas W, et al. Geographic patterns of genome admixture in Latin American Mestizos. PLoS Genet. 2008;4(3), e1000037.
9. The 1000 Genomes Project Consortium, Abecasis GR, Auton A, et al. An integrated map of genetic variation from 1,092 human genomes. Nature. 2012;491(7422):56–65.
10. Jakobsson M, Scholz SW, Scheet P, et al. Genotype, haplotype and copy-number variation in worldwide human populations. Nature. 2008;451(7181):998–1003.
11. Adeyemo A, Rotimi C. Genetic variants associated with complex human diseases show wide variation across multiple populations. Public Health Genomics. 2010;13(2):72–9.
12.•• DIAbetes Genetics Replication Meta-analysis Consortium, Asian Genetic Epidemiology Network Type 2 Diabetes Consortium, et al. Genome-wide trans-ancestry meta-analysis provides insight into the genetic architecture of type 2 diabetes susceptibility. Nat Genet. 2014;46(3):234–44. **This paper described the results of the most ambitious genome-wide trans-ancestry meta-analysis for type 2 diabetes carried out to date, including samples of East Asian, European, Hispanic, and South Asian ancestry**.
13. Carlson CS, Matise TC, North KE, et al. Generalization and dilution of association results from European GWAS in populations of non-European ancestry: the PAGE study. PLoS Biol. 2013;11(9), e1001661.
14. Centers for Disease Control and Prevention: CDC health disparities and inequalities report—United States, 2013, in *MMWR Surveill Summ.* 2013. p. 1–187.
15. Guarnaccia PJ, Pincay IM, Alegria M, et al. Assessing diversity among Latinos: results from the NLAAS. Hisp J Behav Sci. 2007;29(4):510–34.
16. Schneiderman N, Llabre M, Cowie CC, et al. Prevalence of diabetes among Hispanics/Latinos from diverse backgrounds: the Hispanic Community Health Study/Study of Latinos (HCHS/SOL). Diabetes Care. 2014;37(8):2233–9.
17. Risch N. Linkage strategies for genetically complex traits. II. The power of affected relative pairs. Am J Hum Genet. 1990;46(2):229–41.
18. Risch N. Linkage strategies for genetically complex traits. I. Multilocus models. Am J Hum Genet. 1990;46(2):222–8.
19. Curtis D, Sham PC. Model-free linkage analysis using likelihoods. Am J Hum Genet. 1995;57(3):703–16.
20. Gyapay G, Morissette J, Vignal A, et al. The 1993–94 Genethon human genetic linkage map. Nat Genet. 1994;7(2):246–339.
21. Dib C, Faure S, Fizames C, et al. A comprehensive genetic map of the human genome based on 5,264 microsatellites. Nature. 1996;380(6570):152–4.
22. Hanis CL, Boerwinkle E, Chakraborty R, et al. A genome-wide search for human non-insulin-dependent (type 2) diabetes genes reveals a major susceptibility locus on chromosome 2. Nat Genet. 1996;13(2):161–6.
23. Cox NJ, Frigge M, Nicolae DL, et al. Loci on chromosomes 2 (NIDDM1) and 15 interact to increase susceptibility to diabetes in Mexican Americans. Nat Genet. 1999;21(2):213–5.
24. Horikawa Y, Oda N, Cox NJ, et al. Genetic variation in the gene encoding calpain-10 is associated with type 2 diabetes mellitus. Nat Genet. 2000;26(2):163–75.
25. Cheverud JM, Fawcett GL, Jarvis JP, et al. Calpain-10 is a component of the obesity-related quantitative trait locus Adip1. J Lipid Res. 2010;51(5):907–13.
26. Duggirala R, Blangero J, Almasy L, et al. Linkage of type 2 diabetes mellitus and of age at onset to a genetic location on chromosome 10q in Mexican Americans. Am J Hum Genet. 1999;64(4):1127–40.
27. Lehman DM, Hunt KJ, Leach RJ, et al. Haplotypes of transcription factor 7-like 2 (TCF7L2) gene and its upstream region are associated with type 2 diabetes and age of onset in Mexican Americans. Diabetes. 2007;56(2):389–93.
28. Knowler WC, Bennett PH, Hamman RF, et al. Diabetes incidence and prevalence in Pima Indians: a 19-fold greater incidence than in Rochester Minnesota. Am J Epidemiol. 1978;108(6):497–505.
29. Hanson RL, Ehm MG, Pettitt DJ, et al. An autosomal genomic scan for loci linked to type II diabetes mellitus and body-mass index in Pima Indians. Am J Hum Genet. 1998;63(4):1130–8.
30. Baier LJ, Hanson RL. Genetic studies of the etiology of type 2 diabetes in Pima Indians: hunting for pieces to a complicated puzzle. Diabetes. 2004;53(5):1181–6.
31. Mohlke KL, Boehnke M. Recent advances in understanding the genetic architecture of type 2 diabetes. Hum Mol Genet. 2015;24(R1):R85–92.
32. Hayes MG, Pluzhnikov A, Miyake K, et al. Identification of type 2 diabetes genes in Mexican Americans through genome-wide association studies. Diabetes. 2007;56(12):3033–44.
33. Parra EJ, Below JE, Krithika S, et al. Genome-wide association study of type 2 diabetes in a sample from Mexico City and a meta-analysis of a Mexican-American sample from Starr County. Texas Diabet. 2011;54(8):2038–46.
34. Below JE, Gamazon ER, Morrison JV, et al. Genome-wide association and meta-analysis in populations from Starr County, Texas,

and Mexico City identify type 2 diabetes susceptibility loci and enrichment for expression quantitative trait loci in top signals. Diabetologia. 2011;54(8):2047–55.

35.•• Sigma Type 2 Diabetes Consortium, Williams AL, Jacobs SB, et al. Sequence variants in SLC16A11 are a common risk factor for type 2 diabetes in Mexico. Nature. 2014;506(7486):97–101. **Large genome-wide association study that identified a common haplotype in Hispanics that strongly influences diabetes risk and appears to have introgressed into modern humans through admixture with Neandertals**.

36. Hara K, Fujita H, Johnson TA, et al. Genome-wide association study identifies three novel loci for type 2 diabetes. Hum Mol Genet. 2014;23(1):239–46.

37.• Sigma Type 2 Diabetes Consortium, Estrada K, Aukrust I, et al. Association of a low-frequency variant in HNF1A with type 2 diabetes in a Latino population. JAMA. 2014;311(22):2305–14. **The first exome-wide study of type 2 diabetes in a Hispanic population**.

38. Majithia AR, Flannick J, Shahinian P, et al. Rare variants in PPARG with decreased activity in adipocyte differentiation are associated with increased risk of type 2 diabetes. Proc Natl Acad Sci U S A. 2014;111(36):13127–32.

39.•• Palmer ND, Goodarzi MO, Langefeld CD, et al. Genetic variants associated with quantitative glucose homeostasis traits translate to type 2 diabetes in Mexican Americans: The GUARDIAN (Genetics Underlying Diabetes in Hispanics) Consortium. Diabetes. 2015;64(5):1853–66. **Largest study to date of Hispanic variants influencing type 2 diabetes risk, focused on risk via quantitative measures of metabolic health**.

40. Lyssenko V, Nagorny CL, Erdos MR, et al. Common variant in MTNR1B associated with increased risk of type 2 diabetes and impaired early insulin secretion. Nat Genet. 2009;41(1):82–8.

41. Prokopenko I, Langenberg C, Florez JC, et al. Variants in MTNR1B influence fasting glucose levels. Nat Genet. 2009;41(1):77–81.

42. Haiman CA, Fesinmeyer MD, Spencer KL, et al. Consistent directions of effect for established type 2 diabetes risk variants across populations: the population architecture using genomics and epidemiology (PAGE) consortium. Diabetes. 2012;61(6):1642–7.

43. Waters, KM, Stram, DO, Hassanein, MT, et al.: Consistent association of type 2 diabetes risk variants found in Europeans in diverse racial and ethnic groups. PLoS Genet 2010; 6(8).

44. Nicolae DL, Gamazon E, Zhang W, et al. Trait-associated SNPs are more likely to be eQTLs: annotation to enhance discovery from GWAS. PLoS Genet. 2010;6(4), e1000888.

45.• Torres JM, Gamazon ER, Parra EJ, et al. Cross-tissue and tissue-specific eQTLs: partitioning the heritability of a complex trait. Am J Hum Genet. 2014;95(5):521–34. **By subsetting the complete set of interrogated markers into those that act as eQTLs in insulin-responsive peripheral tissues, this group identified a significant enrichment of regulatory variation in the genetic heritability of type 2 diabetes**.

46. Manolio TA, Collins FS, Cox NJ, et al. Finding the missing heritability of complex diseases. Nature. 2009;461(7265):747–53.

47. Hanson RL, Guo T, Muller YL, et al. Strong parent-of-origin effects in the association of KCNQ1 variants with type 2 diabetes in American Indians. Diabetes. 2013;62(8):2984–91.

48. Lawson HA, Zelle KM, Fawcett GL, et al. Genetic, epigenetic, and gene-by-diet interaction effects underlie variation in serum lipids in a LG/JxSM/J murine model. J Lipid Res. 2010;51(10):2976–84.

49. Wijsman EM. The role of large pedigrees in an era of high-throughput sequencing. Hum Genet. 2012;131(10):1555–63.

50. Rainwater DL, Almasy L, Blangero J, et al. A genome search identifies major quantitative trait loci on human chromosomes 3 and 4 that influence cholesterol concentrations in small LDL particles. Arterioscler Thromb Vasc Biol. 1999;19(3):777–83.

51. Malhotra A, American Diabetes Association, GSG, Wolford JK. Analysis of quantitative lipid traits in the genetics of NIDDM (GENNID) study. Diabetes. 2005;54(10):3007–14.

52. Bielinski SJ, Tang W, Pankow JS, et al. Genome-wide linkage scans for loci affecting total cholesterol, HDL-C, and triglycerides: the Family Blood Pressure Program. Hum Genet. 2006;120(3):371–80.

53. Dong C, Beecham A, Wang L, et al. Genetic loci for blood lipid levels identified by linkage and association analyses in Caribbean Hispanics. J Lipid Res. 2011;52(7):1411–9.

54. Coletta DK, Schneider J, Hu SL, et al. Genome-wide linkage scan for genes influencing plasma triglyceride levels in the Veterans Administration Genetic Epidemiology Study. Diabetes. 2009;58(1):279–84.

55. Duggirala R, Blangero J, Almasy L, et al. A major susceptibility locus influencing plasma triglyceride concentrations is located on chromosome 15q in Mexican Americans. Am J Hum Genet. 2000;66(4):1237–45.

56.• Hellwege JN, Palmer ND, Raffield LM, et al. Genome-wide family-based linkage analysis of exome chip variants and cardiometabolic risk. Genet Epidemiol. 2014;38(4):345–52. **This paper reported the results of a genome-wide family-based linkage analysis using a dense panel of exome variants. In Hispanic families, the strongest evidence of linkage was observed in the gene CETP for HDL-cholesterol, and variants within this gene explained 4.5% of the variance for this trait**.

57. Comuzzie AG, Cole SA, Laston SL, et al. Novel genetic loci identified for the pathophysiology of childhood obesity in the Hispanic population. PLoS One. 2012;7(12), e51954.

58.• Coram MA, Duan Q, Hoffmann TJ, et al. Genome-wide characterization of shared and distinct genetic components that influence blood lipid levels in ethnically diverse human populations. Am J Hum Genet. 2013;92(6):904–16. **This GWA study reported strong concordance in direction and magnitude of genetic effects for lipid traits in African American, Hispanic, and European populations. However, there was also evidence of heterogeneity within shared loci, either due to population-specific rare variants, or shared variants with different allele frequencies**.

59.• Weissglas-Volkov D, Aguilar-Salinas CA, Nikkola E, et al. Genomic study in Mexicans identifies a new locus for triglycerides and refines European lipid loci. J Med Genet. 2013;50(5):298–308. **This was the first GWA study of lipid traits in Hispanics. The authors described a novel locus for triglycerides in the gene NPC1 and refined the signals at some loci previously described in European samples based on the different patterns of LD between Hispanic and European populations**.

60.• Ko A, Cantor RM, Weissglas-Volkov D, et al. Amerindian-specific regions under positive selection harbour new lipid variants in Latinos. Nat Commun. 2014;5:3983. **This study used a different approach than conventional GWA studies, by restricting the association tests to variants differing in frequency between European and Native American populations. Novel signals for lipid traits were reported in the genes RORA and SIK3. Native-American-specific signatures of risk variants were described for LPL and APOA5**.

61. Baran Y, Pasaniuc B, Sankararaman S, et al. Fast and accurate inference of local ancestry in Latino populations. Bioinformatics. 2012;28(10):1359–67.

62.• Below, JE, Parra, EJ, Gamazon, ER, et al.: Meta-analysis of lipid-traits in Hispanics identifies novel loci, population-specific effects, and tissue-specific enrichment of eQTLs. Sci Reports. 2016;6: 19429. **This is the largest GWA for lipid traits carried out to date in Hispanics. Association analyses were also carried out for eSNPs reported for different tissues. The authors described substantial concordance of genetic effects in European and**

Hispanic populations. However, for some genes, the signals observed in Hispanics were different from those reported in European studies.

63. Global Lipids Genetics Consortium et al. Discovery and refinement of loci associated with lipid levels. Nat Genet. 2013;45(11):1274–83.

64. Musso G, Gambino R, Cassader M. Interactions between gut microbiota and host metabolism predisposing to obesity and diabetes. Annu Rev Med. 2011;62:361–80.

65. Amar J, Serino M, Lange C, et al. Involvement of tissue bacteria in the onset of diabetes in humans: evidence for a concept. Diabetologia. 2011;54(12):3055–61.

66. Cani PD, Osto M, Geurts L, et al. Involvement of gut microbiota in the development of low-grade inflammation and type 2 diabetes associated with obesity. Gut Microbes. 2012;3(4):279–88.

67. Qin J, Li Y, Cai Z, et al. A metagenome-wide association study of gut microbiota in type 2 diabetes. Nature. 2012;490(7418):55–60.

68.•• Karlsson FH, Tremaroli V, Nookaew I, et al. Gut metagenome in European women with normal, impaired and diabetic glucose control. Nature. 2013;498(7452):99–103. **This study described a mathematical model to identify type 2 diabetes based on metagenomic profiles. The authors also reported that the discriminant metagenomics markers for type 2 diabetes were different between European and Chinese cohorts**.

69. Cai L, Wu H, Li D, et al. Type 2 diabetes biomarkers of human gut microbiota selected via iterative sure independent screening method. PLoS One. 2015;10(10), e0140827.

70. Fu J, Bonder MJ, Cenit MC, et al. The gut microbiome contributes to a substantial proportion of the variation in blood lipids. Circ Res. 2015;117(9):817–24.

71. Ross MC, Muzny DM, McCormick JB, et al. 16S gut community of the Cameron County Hispanic Cohort. Microbiome. 2015;3:7.

72. Delzenne NM, Cani PD, Everard A, et al. Gut microorganisms as promising targets for the management of type 2 diabetes. Diabetologia. 2015;58(10):2206–17.

73. Serino M, Blasco-Baque V, Burcelin R. Microbes on-air: gut and tissue microbiota as targets in type 2 diabetes. J Clin Gastroenterol. 2012;46(Suppl):S27–8.

74. Teslovich TM, Musunuru K, Smith AV, et al. Biological, clinical and population relevance of 95 loci for blood lipids. Nature. 2010;466(7307):707–13.

75. Dumitrescu L, Carty CL, Taylor K, et al. Genetic determinants of lipid traits in diverse populations from the population architecture using genomics and epidemiology (PAGE) study. PLoS Genet. 2011;7(6), e1002138.

76. van de Bunt M, Cortes A, Consortium I, et al. Evaluating the performance of fine-mapping strategies at common variant GWAS loci. PLoS Genet. 2015;11(9), e1005535.

77. Li YR, Keating BJ. Trans-ethnic genome-wide association studies: advantages and challenges of mapping in diverse populations. Genome Med. 2014;6(10):91.

78. Wu Y, Waite LL, Jackson AU, et al. Trans-ethnic fine-mapping of lipid loci identifies population-specific signals and allelic heterogeneity that increases the trait variance explained. PLoS Genet. 2013;9(3), e1003379.

79. Steinthorsdottir V, Thorleifsson G, Sulem P, et al. Identification of low-frequency and rare sequence variants associated with elevated or reduced risk of type 2 diabetes. Nat Genet. 2014;46(3):294–8.

80. Wessel J, Chu AY, Willems SM, et al. Low-frequency and rare exome chip variants associate with fasting glucose and type 2 diabetes susceptibility. Nat Commun. 2015;6:5897.

81. Surakka I, Horikoshi M, Magi R, et al. The impact of low-frequency and rare variants on lipid levels. Nat Genet. 2015;47(6):589–97.

82. Genomes Project C, Auton A, Brooks LD, et al. A global reference for human genetic variation. Nature. 2015;526(7571):68–74.