Association between triglycerides, known risk SNVs and conserved rare variation in SLC25A40 in a multi-ancestry cohort

Elisabeth A. Rosenthal1*, David R. Crosslin2, Adam S. Gordon3, David S. Carrell4, Ian B. Stanaway2, Eric B. Larson6, Jane Grafton4, Wei-Qi Wei5, Joshua C. Denny5,6, Qi-Ping Feng6, Amy S. Shah7, Amy C. Sturm8, Marylyn D. Ritchie9, Jennifer A. Pacheco3, Hakon Hakonarson10, Laura J. Rasmussen-Torvik11, John J. Connolly10, Xiao Fan12, Maya Safarova12, Iftikhar J. Kullo12,13 and Gail P. Jarvik1,14

Abstract

Background: Elevated triglycerides (TG) are associated with, and may be causal for, cardiovascular disease (CVD), and co-morbidities such as type II diabetes and metabolic syndrome. Pathogenic variants in APOA5 and APOC3 as well as risk SNVs in other genes [APOE (rs429358, rs7412), APOA1/C3/A4/A5 gene cluster (rs964184), INSR (rs7248104), CETP (rs7205804), GCKR (rs1260326)] have been shown to affect TG levels. Knowledge of genetic causes for elevated TG may lead to early intervention and targeted treatment for CVD. We previously identified linkage and association of a rare, highly conserved missense variant in SLC25A40, rs762174003, with hypertriglyceridemia (HTG) in a single large family, and replicated this association with rare, highly conserved missense variants in a European American and African American sample.

Methods: Here, we analyzed a longitudinal mixed-ancestry cohort (European, African and Asian ancestry, N = 8966) from the Electronic Medical Record and Genomics (eMERGE) Network. We tested associations between median TG and the genes of interest, using linear regression, adjusting for sex, median age, median BMI, and the first two principal components of ancestry.

Results: We replicated the association between TG and APOC3, APOA5, and risk variation at APOE, APOA1/C3/A4/A5 gene cluster, and GCKR. We failed to replicate the association between rare, highly conserved variation at SLC25A40 and TG, as well as for risk variation at INSR and CETP.

Conclusions: Analysis using data from electronic health records presents challenges that need to be overcome. Although large amounts of genotype data is becoming increasingly accessible, usable phenotype data can be challenging to obtain. We were able to replicate known, strong associations, but were unable to replicate moderate associations due to the limited sample size and missing drug information.

Keywords: Genetics, Cardiovascular disease, Triglycerides, Electronic health records

Background

Triglyceride (TG) levels are a marker for health outcomes. High TG, (defined as 200 < TG < 500 mg/dL in adults), are associated with, and may cause, cardiovascular disease (CVD) [1-4]. High TG is also associated with type II diabetes and metabolic syndrome [5, 6]. TG
levels are affected by environmental variables such as diet, alcohol consumption, smoking, and exercise [7-11]. Maintaining healthy levels of TGs may improve health outcomes [6, 12].

Both hypo- and hypertriglyceridemia (HTG) have major genetic components such as heterozygous variation at APOC3 and APOA5, respectively [13-15]. More recently, homozygous and compound heterozygous null variation at ANGPTL3 and ANGPTL4 have been shown to be associated with low TG [16, 17]. SNVs in the genes APOE, GCKR, APOA1/C3/A4/A5 gene cluster, INSR and CETP have smaller effects on TG [18-20].

We previously identified linkage and association of a rare, highly conserved missense SNV, rs762174003, in SLC25A40 with HTG in a large family, and replicated the association with rare missense SNVs at highly conserved sites in SLC25A40 in a European American (EA) and African American (AA) sample [21]. SLC25A40 encodes a mitochondrial transmembrane protein [22]. In drosophila, SLC25A40 loss-of-function was shown to contribute to mitochondrial damage due to excessive oxidation [23]. Here, we attempt to replicate the association between deleterious variation at SLC25A40, as well as other reported associations with variation at APOE, GCKR, APOA1/C3/A4/A5 gene cluster, INSR, and CETP, with TG in a multi-ancestry population from the Electronic Medical Records and Genomics (eMERGE) Network phase 3 study [24, 25].

Methods
Participants and phenotype data
Participants were ascertained from each of nine phase 3 eMERGE sites, a study tasked with investigating return of results from the sequencing of a panel of ~100 genes in a large (N=25,000), diverse cohort of participants [24]. Participants ages 18 years and older were included, as the distribution for TG levels in children differs from that of adults. We ascertained relevant patient data from the clinical records, including TG level measured in mg/dL (N=12,229 adult participants), age, sex, weight, height and body mass index (BMI), measured at the same clinical visit. We also ascertained relevant ICD9 and ICD10 codes for lipidemias, pancreatic cancer, and pancreatitis. Thirteen participants with chylomicronemia (ICD9 code 272.3, ICD10 code E78.3), 12 participants with Kaposi’s sarcoma (ICD9 code 176 and ICD10 code C46) and 2,539 participants with morbid obesity (ICD9 code 278.01 and ICD10 code E66.01) or bariatric surgery were removed as their TG values are likely associated with unaccounted for exposures. Quality control was performed within participants as well. Participants who had maximum TG (maxTG)<40 and had more than one record verifying this low TG were retained. However, one participant who had a single record with TG=19 mg/dL was removed. Records within 2 weeks prior or 4 weeks after a diagnosis of pancreatitis (ICD9 code 577.0, ICD10 code K85.9) for an individual were removed (382 participants), as high TG are correlated with pancreatitis and is likely due to an environmental exposure such as alcohol. Similarly, individual level records coincident and after a diagnosis of pancreatic cancer were removed (147 participants). Quality control of BMI was performed using height and weight data as described by Goodloe et. al [26]. For participants with only BMI data (i.e., no height or weight data), values more than 5 standard deviations from the participant specific median BMI were removed.

Genotype data
A subset of individuals (N=24,956) in eMERGE were genotyped using a custom capture target containing sequence for 788 genes with partial to full exon coverage, including 58 of the 59 genes considered actionable by the American College of Medical Genetics (ACMG) [27, 28], and detects 62,051 SNVs. Sequencing was performed at two sites: Partners Healthcare Laboratory for Molecular Medicine (LMM) and Baylor. Samples from dropped out participants, that failed quality control or had sex discrepancy were removed (N=229). After removing duplicate data for a single participant, sequence data for all participants was aligned to a merged target sequence map, using the Burrows-Wheeler Aligner version 0.7.10 [29]. Genotypes were jointly called on all participants using the Genome Analysis Toolkit version 3.5 [30]. 896 Variants with genotype missingness rate >5% were removed. Individual level genotyping rate was >98% for all participants. After cleaning the data, 8970 participants with both phenotype and genotype data remained in the analysis. Four additional participants were removed due to their influence on the residual model (see results).

Principal components (PC) of ancestry was calculated on a pruned set of 1571 SNVs (r2<0.7, MAF>0.05). Self-reported ancestry aligned well with the first two PCs.

Exon sequence data was available for the relevant genes SLC25A40, APOC3, and APOA5. Two individuals with known pathogenic variants in APOA5 (rs201079485, rs147528707) were removed from analysis. Exon sequence data was also available for known TG risk SNVs rs429358 and rs7412 (which determine the APOE ε2/3/4 genotype), rs7248104 (INSR), rs1260326 (GCKR), rs7205804 (CETP), and rs964184 (APOA1/C3/A4/A5 gene cluster).

Analysis
The phenotype of interest, medTGres, was calculated as the residuals from the median log10 transformed TG for each individual, adjusted for median age, sex, median
BMI, presence of hypertriglyceridemia (HTG, ICD9 code 272.1 and ICD10 code E78.1), presence of hyperlipidemia (HL, ICD9 codes 272.2, 272.4, 272.9 and ICD10 codes E78.2, E78.4, E78.9), APOE genotype, known risk SNVs, known APOC3 protective SNVs, the first two principal components (PCs) of ancestry, and site. The variable for median age was centered around age 50, as this is when TG levels tend to descend, and then squared, as the relationship between TG levels and age is parabolic. We log10 transformed TG as the distribution is highly skewed and the transformation makes it nearly Normal. We used the median transformed TG (medTG) value over all records for each individual because the median protects against outliers that could be due to non-fasting measurements. The resulting residuals from the above linear model (medTGres) were used as the primary phenotype. We also used a similar residual from the adjusted maximum log10 transformed TG (maxTGres), as using the maximum protects against errors introduced from treatment with Niacin, fibrates or binders for high TG. Given our more complete local data from the University of Washington and Kaiser Permanente Washington (UWKP) site, we expect about 3% of the participants to be receiving such treatment at some time point in their clinical records [31]. However, this data was not available to us study-wide.

In addition to assessing the significance of the known risk variation, we analyzed variants in SLC25A40 that are likely to affect TG levels [21]. This included rare (maximum population MAF < 0.005) variants that either caused an early termination in the first 90% of the coding region, or changed the amino acid sequence. Missense variants were further constrained to be evolutionarily conserved (Genomic Evolutionary Rate Profiling (GERP) score > 4.8) [32]. The maximum population MAF was derived from populations in the gnomAD database that have not experienced a bottleneck (i.e., ignored MAF from Ashkenazi Jewish or Finnish populations) [33, 34]. We performed four separate gene-wise tests for association with medTGres and maxTGres using 1. Genotype at single SNVs with more than 10 heterozygotes, 2. All other SNVs collapsed into a single indicator variable, 3. All other rare missense SNVs collapsed into a single indicator variable and 4. All other early terminations collapsed into a single indicator variable.

**Results**

**Demographics**

Demographics of the participants included in these analyses are given in Table 1. There were slightly more females than males at most sites. The distribution of age varied widely among the sites, with younger participants from Children’s Hospital of Philadelphia (CHOP) and Cincinnati Children’s Hospital Medical Center. Median BMI tended to be above normal (> 25), but there were some participants with low BMI (< 18.5) (Columbia, Mayo). There were 51 participants with very low median BMI < 11, all from Mayo, with at least 3 records supporting this measurement. The median TG ranged between 19 and 1371 mg/dL over all participants, but the majority of participants had median TG < 150 mg/dL, which is defined as normal (Fig. 1). The number of TG records available per participant varied widely, with a median of 6 overall. Most participants were of European ancestry.

| Site        | Total | Nfemale (%F) | Median age (min, max) | Median BMIa Avg (min, max) | Median TGb Avg (min, max) | Number of records |
|-------------|-------|--------------|-----------------------|---------------------------|---------------------------|------------------|
|             |       |              |                       |                           |                           |                  |
| CHOP        | 35    | 15 (43)      | 19 (18, 21)           | 27 (17, 48)               | 94 (33, 232)              | 1                |
| Cinc        | 287   | 150 (52)     | 21 (18, 56)           | 27 (14, 51)               | 130 (27, 610)             | 2                |
| Colu        | 618   | 341 (55)     | 50 (18, 87)           | 27 (11, 45)               | 146 (24, 857)             | 5                |
| Geis        | 1278  | 736 (58)     | 49 (18, 89)           | 30 (16, 58)               | 134 (25, 988)             | 5                |
| Harv        | 932   | 533 (57)     | 53 (18, 89)           | 27 (15, 45)               | 121 (30, 601)             | 7                |
| Mayo        | 2023  | 1140 (56)    | 53 (21, 69)           | 27 (9, 45)                | 139 (32, 1038)            | 15               |
| Nwes        | 1962  | 1068 (54)    | 59 (18, 89)           | 28 (15, 52)               | 116 (19, 1371)            | 4                |
| UWKP        | 1685  | 955 (57)     | 57 (22, 89)           | 27 (15, 40)               | 137 (32, 636)             | 5                |
| Vand        | 146   | 50 (34)      | 64 (26, 86)           | 29 (18, 47)               | 157 (41, 484)             | 12.5             |
| Total       | 8966  | 4988 (56)    | 53 (18, 89)           | 28 (9, 58)                | 131 (19, 1371)            | 6                |

CHOP Children’s hospital of Philadelphia, Cinc Cincinnati, Colu Columbia, Geis Geisinger, Harv Harvard, Mayo Mayo Clinic, Nwes Northwestern, UWKP University of Washington and Kaiser Permanente Washington, Vand Vanderbilt

a Low BMI, and the associated heights and weights were manually checked for consistency within an individual, across records

b Low TG was manually checked for consistency within an individual, across records
(77%) followed by African ancestry (7%) and Asian ancestry (6%) (Table 2, Fig. 2). Most of the participants of Asian ancestry (82%) were from UWKP and most of the participants of African ancestry (62%) were from Northwestern and Cincinnati.

**RiskSNVs**

Genotype counts for *APOE* and risk SNVs at *APOC3, APOA5, APOA1/C3/A4/A5* gene cluster and *GCKR* are given in Tables 3 and 4. The distribution of *APOE* genotypes follows expectations, given known allele frequencies. Sixty-two participants were heterozygous for 1 of 4 protective *APOC3* SNVs (rs76353203, rs147210663, rs138326449, rs140621530). Eighty-three participants had at least one of two risk alleles at *APOA5* SNV rs2075291.

---

**Table 2** Distribution of ancestry by site

| Site  | AA | AS | EA  | EA/AA | EA/AS | UN |
|-------|----|----|-----|-------|-------|----|
| CHOP  | 17 | 1  | 16  | 0     | 1     | 0  |
| Cinc  | 124| 0  | 139 | 21    | 3     | 0  |
| Colu  | 88 | 29 | 238 | 215   | 48    | 0  |
| Geis  | 28 | 4  | 1200| 30    | 16    | 0  |
| Harv  | 46 | 18 | 766 | 68    | 34    | 0  |
| Mayo  | 5  | 14 | 1989| 1     | 14    | 0  |
| Nwes  | 213| 50 | 1512| 84    | 101   | 2  |
| UWKP  | 23 | 539| 933 | 19    | 166   | 5  |
| Vand  | 2  | 0  | 143 | 0     | 1     | 0  |
| All   | 546| 655| 6936| 438   | 384   | 7  |

Principal components of ancestry are given in Fig. 2.

Chop Children’s Hospital of Philadelphia, Cinc Cincinnati, Colu Columbia, Geis Geisinger, Harv Harvard, Mayo Mayo Clinic, news Northwestern, UWKP University of Washington and Kaiser Permanente Washington, Vand Vanderbilt, AA African American, AS Asian American, EA European American, UN unknown.
Twenty-one rare SNVs, consisting of 2 stop gains, 3 frame shifts and 16 evolutionarily conserved missense variants, were heterozygous in the data set for 32 participants (Table 5). There were 119 participants who were heterozygous at the missense rs724665 and this SNV is considered separately from the others. Each participant was heterozygous for at most one of

Table 3 APOE genotypes counts and percentages derived from rs429358 and rs7412, by site and ancestry

| Site      | ε2/ε2 | ε2/ε3 | ε2/ε4 | ε3/ε3 | ε3/ε4 | ε4/ε4 |
|-----------|-------|-------|-------|-------|-------|-------|
| CHOP      | 0     | 5 (14)| 0     | 19 (54)| 10 (29)| 1 (2.9)|
| Cinc      | 2 (0.7)| 41 (14)| 5 (1.7)| 146 (51)| 79 (28)| 14 (4.9)|
| Colu      | 3 (0.5)| 73 (12)| 13 (2.1)| 373 (60)| 143 (23)| 13 (2.1)|
| Geis      | 15 (1.2)| 171 (13)| 27 (2.1)| 769 (60)| 281 (22)| 15 (1.2)|
| Harv      | 1 (0.1)| 120 (13)| 30 (3.2)| 561 (60)| 195 (21)| 25 (2.7)|
| Mayo      | 14 (0.7)| 180 (9)| 41 (2.1)| 1170 (58)| 560 (28)| 58 (2.9)|
| Nwes      | 8 (0.41)| 227 (12)| 51 (2.6)| 1189 (61)| 451 (23)| 36 (1.8)|
| UWKP      | 10 (0.59)| 206 (12)| 22 (1.3)| 1059 (62)| 370 (22)| 18 (1.1)|
| Vand      | 4 (2.7)| 18 (12)| 6 (1.1)| 77 (53)| 34 (23)| 7 (4.8)|

| Ancestry | ε2/ε2 | ε2/ε3 | ε2/ε4 | ε3/ε3 | ε3/ε4 | ε4/ε4 |
|----------|-------|-------|-------|-------|-------|-------|
| AA       | 3 (0.55)| 93 (17)| 24 (4.4)| 235 (43)| 167 (31)| 24 (4.4)|
| AS       | 1 (0.15)| 87 (13)| 4 (0.61)| 444 (68)| 116 (18)| 3 (0.46)|
| EA       | 49 (0.71)| 776 (11)| 152 (2.2)| 4157 (60)| 1658 (24)| 144 (2.1)|
| EA/AA    | 4 (0.91)| 45 (10)| 14 (3.2)| 255 (58)| 105 (24)| 15 (3.4)|
| EA/AS    | 0     | 39 (10)| 1 (0.26)| 268 (70)| 75 (19.5)| 1 (0.26)|
| UN       | 0     | 1 (14)| 0     | 4 (57)| 2 (29)| 0     |
| Total    | 57 (0.64)| 1041 (12)| 195 (2.1)| 5363 (60)| 2123 (24)| 187 (2.1)|

Chop Children’s hospital of Philadelphia, Cinc Cincinnati, Colu Columbia, Geis Geisinger, Harv Harvard, Mayo Mayo Clinic, news Northwestern, UWKP University of Washington and Kaiser Permanente Washington, Vand Vanderbilt, AA African American, AS Asian American, EA European American, UN unknown

**SLC25A40 genotype data**

Twenty-one rare SNVs, consisting of 2 stop gains, 3 frame shifts and 16 evolutionarily conserved missense variants, were heterozygous in the data set for 32 participants (Table 5). There were 119 participants who were heterozygous at the missense rs724665 and this SNV is considered separately from the others. Each participant was heterozygous for at most one of
the SNVs. Three participants were heterozygous at the frame shifts and 2 participants were heterozygous at the stop gains, limiting the power of the tests.

Residual model

The final residual model, containing only significant covariates is shown in Table 6. Although centered-median-age-squared was significantly associated with log10(medTG), its effect is near zero ($\beta = -4e-05$). Males had higher log10(medTG) ($\beta = 0.02$). Increased BMI was associated with higher log10(medTG) ($\beta = 0.01$). Similarly, individuals with HTG or HL had higher log10(medTG) ($\beta = 0.3$ and 0.08, respectively) than other participants. SITE was also highly associated with log10(medTG) (F-test $p < 2e-16$). Neither the INSR or CETP SNVs were significantly associated with median or maximum TG ($p > 0.16$), and were not included in the model (data not shown). Four potentially influential outlying participants (as defined by their cook's distance [35]) in either model (medTGres or maxTGres) were removed, leaving a total of 8,966 participants with complete phenotype and genotype data for analysis. None of the participants with low BMI (median BMI < 15) appeared to influence the models.

Association with known TG risk SNVs

Previously published risk SNVs at APOE, GCKR, APOA5 and APOC3 were associated with TG (Table 6). The minor alleles at the APOE SNVs rs429358 and rs7412 were associated with increased log10(medTG) ($\beta = 0.01$ and 0.04, respectively) and are jointly significantly associated with log10(medTG) (F-test $p = 1.1e-11$). Under an additive model, the minor alleles at GCKR and APOA1/C3/A4/A5 gene cluster were also significantly positively associated with log10(medTG) ($\beta = 0.03$, $p < 2e-16$, APOA1/C3/A4/A5 gene cluster: $\beta = 0.04$, $p < 2e-16$). Under a dominant model, presence of either minor allele at APOA5 was positively associated with log10(medTG) ($\beta = 0.09$, $p = 2.7e-05$). Jointly, a single copy of any of the minor alleles at the four APOC3 SNVs were negatively associated with log10(medTG) ($\beta = -0.23$, $p = 1.3e-21$).

Primary analysis

None of the SLC25A40 SNVs were significantly associated with medTGres, either separately or in combination.

---

**Table 4 Risk SNV genotype counts by site and ancestry**

| Site  | APOC3 01 | APOA5 00/01/02 | APOA1/C3/A4/A5 00/01/11 | GCKR 00/01/11 | INSR 00/01/11 | CETP 00/01/11 |
|-------|----------|----------------|--------------------------|---------------|--------------|--------------|
| CHOP  | 0        | 35/0/0/0/0     | 31/4/0                   | 15/16/4       | 4/16/15      | 19/12/4      |
| Cinc  | 3        | 286/1/0/0/0    | 183/89/15                | 156/94/37     | 37/94/156    | 151/91/45    |
| Colu  | 8        | 611/7/0/0     | 416/180/22               | 276/263/79    | 79/263/276   | 300/238/80   |
| Geis  | 10       | 1273/4/0/0/0  | 924/317/37               | 414/640/224   | 224/640/414  | 412/640/226  |
| Harv  | 11       | 932/0/0/0/0   | 626/277/29               | 310/451/171   | 171/451/310  | 316/450/166  |
| Mayo  | 3        | 2021/2/0/0     | 1435/537/51             | 695/982/346   | 346/982/695  | 642/989/392  |
| Nwes  | 22       | 1951/9/2/0    | 1360/546/56             | 746/905/311   | 311/905/746  | 742/922/298  |
| UWKP  | 5        | 1628/54/3/0   | 1132/487/66             | 607/764/314   | 314/764/607  | 637/809/239  |
| Vand  | 0        | 146/0/0/0/0   | 110/34/2                | 53/64/29      | 29/64/53     | 36/77/33     |

Ancestry

|        | AA 7 | 542/4/0/0/0 | 343/183/20 | 269/227/50 | 13/120/413 | 407/129/10 |
|        | 0    | 595/5/5/0/0 | 376/245/34 | 306/274/75 | 149/322/184 | 315/286/54 |
|        | 53   | 6930/5/0/0/0 | 5006/1752/178 | 2362/3396/1178 | 1271/3384/2281 | 2181/3440/1315 |
|        | 2    | 438/0/0/0/0 | 278/143/17 | 188/201/49 | 38/176/224 | 218/180/40 |
|        | 0    | 371/13/0/0/0 | 208/147/29 | 156/165/63 | 44/172/168 | 131/190/63 |
|        | 0    | 7/0/0/0/0/0 | 6/1/0/0 | 3/4/0 | 0/5/2 | 3/3/1 |
| Total  | 62   | 8883/77/5/1 | 6217/2471/278 | 3272/4179/1515 | 1515/4179/3272 | 3255/4228/1483 |

APOC3 consists of 4 SNVs (rs76353203, rs147210663, rs138326449, rs140621530), APOA5 = rs2075291, APOA1/C3/A4/A5 gene cluster = rs964184, GCKR = rs1260326. INSR = rs1260326, CETP = rs7205804

Chop Children’s hospital of Philadelphia, Cinc Cincinnati, Colu Columbia, Geis Geisinger, Harv Harvard, Mayo Mayo Clinic, news Northwestern, UWKP University of Washington and Kaiser Permanente Washington, Vand Vanderbilt, AA African American, AS Asian American, EA European American, UN unknown

00 = common genotype, 01 = heterozygote, 11 = alternate homozygote.

* rs2075291 has three alleles. For this SNV, 02 = rare heterozygote (Allele 2 has frequency less than allele 1)
Rosenthal et al. BMC Med Genomics (2021) 14:11

Furthermore, the estimated effect of the variants was negative in all situations (rs724665 $\beta = -0.01$; all SNVs excluding rs724665 $\beta = -0.02$; all rare missense $\beta = -0.01$; rare early terminations $\beta = -0.07$). Similar results were observed for maxTGres and for analysis including the EA-only subset.

### Discussion

We replicated association between TG and risk SNVs at APOE, GCKR, APOA5 and APOC3, with the respective directions of effect as expected. Association between variation at APOA5 and APOC3 with TG has been well

| Table 5 SLC25A40 variants which are rare and cause a coding change |
|------------------|---|---|---|---|
| rs943387265      | 87466057 | C | T | MS |
| rs1022111508     | 87470978 | T | C | MS |
| rs369745713      | 87470986 | T | C | MS |
| rs148648460      | 87473070 | G | C | MS |
| rs140104130      | 87473143 | A | T | MS |
| NA              | 87473157 | T | TGCTAGTAATTGTC |
| rs74655065       | 87473158 | CAT | C | FS |
| rs147753823      | 87473175 | A | C | MS |
| rs77550958       | 87476272 | G | A | MS |
| rs1035790230     | 87476319 | C | T | SG |
| rs890753675      | 87476428 | A | G | MS |
| NA              | 87477248 | T | C | MS |
| rs742665a        | 87477257 | G | A | MS |
| rs200954020      | 87479214 | C | T | MS |
| rs1021091982     | 87483577 | C | T | MS |
| rs748627166      | 87483582 | A | T | SG |
| rs1443219471     | 87483607 | C | G | MS |
| NA              | 87487954 | G | A | MS |
| NA              | 87488022 | TC | T | FS |
| rs747192743      | 87488041 | A | G | MS |

**Table 6 Effect of covariates in linear model used to adjust median TG**

| Variable       | $\beta$  | $p$       |
|----------------|---------|----------|
| Age            | $-4.04e-05$ | $2.15e-07$ |
| Sex (Male)     | $0.02$  | $3.08e-09$ |
| BMI            | $0.01$  | $2.46e-132$ |
| APOE (rs429358) | $0.01$ | $1.98e-03$ |
| APOE (rs7412)  | $0.04$  | $1.65e-11$ |
| GCKR (rs1260326) | $0.03$ | $1.23e-27$ |
| APOA1/C3/A4/A5 gene cluster (rs964184) | $0.04$ | $1.58e-29$ |
| APOC3          | $-0.23$ | $1.32e-21$ |
| APOA5 (rs2075291) | $0.09$ | $2.65e-05$ |
| PC1            | $-2.84$ | $2.57e-09$ |
| PC2            | $-2.17$ | $8.62e-11$ |
| SITE           | NA      | $<2e-16$ |
| HTG            | $0.3$   | $3.34e-176$ |
| HL             | $0.08$  | $6.92e-72$ |

| Variable       | $\beta$  | $p$       |
|----------------|---------|----------|
| rs724665       | $0.01$  | $0.52$   |
| Model 2        | $0.02$  | $0.49$   |
| Model 3        | $0.01$  | $0.69$   |
| Model 4        | $0.07$  | $0.4$    |

Model 2 = all SNVs except for rs724665. Model 3 = all missense SNVs except for rs724665. Model 4 = all early termination SNVs

($p > 0.4$, Table 7, Fig. 3). Furthermore, the estimated effect of the variants was negative in all situations (rs724665 $\beta = -0.01$; all SNVs excluding rs724665 $\beta = -0.02$; all rare missense $\beta = -0.01$; rare early terminations $\beta = -0.07$). Similar results were observed for maxTGres and for analysis including the EA-only subset.
documented, and our ability to detect these effects supports the reliability of these TG data. Similarly, replication of an association between APOE and TG lends further support to the association of APOE ε2/3/4 and lipid traits in general. The association between GCKR and TG further supports a biological mechanism between diabetes and TG [36]. However, we did not replicate the reported association between TG and risk SNVs at INSR and CETP. The association between INSR rs7248104 and TG was first reported by a large GWAS with a sample size of 176,000 subjects [18]. It is possible that the sample size here is too small to detect any effect. Similarly, the association with CETP rs7205804 was discovered in a cohort of 95,000 subjects [37], and therefore it is possible that the sample size here is too small to detect any effect with TG.

Additionally, we failed to replicate an association between putatively deleterious variation at SLC25A40 and TG. Previous studies relied on linkage in a large family and data from the Exome Sequencing Project (ESP). It is possible that the original signal found in the family is due to another variant in linkage with rs762174003, but unrelated to the function of SLC25A40. Further research into this locus may benefit from whole genome sequencing of selected individuals from this family. However, this would not explain why an association was detected in the relatively small sample from the ESP. Results for maximum TG also showed no association.

It is also possible that there is a relationship between SLC25A40 and TG, but we are not able to detect it in this sample. One major caveat is that we are missing data on treatment for high TG. In our local data set, UWKP, where treatment data were available, we discovered approximately 3% of individuals were being treated for high TG with niacin, fibrates or binders [31]. When we accounted for these treatments in the local data we found evidence supporting an association between TG and rare early termination and conserved missense variation at SLC25A40 [31]. As the number of participants who harbor such a variant in SLC25A40 is small (1.7% of our sample), an unaccounted for 3% treatment rate could obscure any signal. Although we used median TG to reduce the
effect of unknown treatment, it may have been insufficient as we do not know when treatment would have begun. We did not use the mean as it can be influenced by outliers and TG have a right skewed distribution. Furthermore, this is secondary use of data, which is prone to a high error rate in the data. Although we attempted to clean the data using multiple strategies, it is possible that enough error remained in the data to obscure any true association with \( \text{SLC25A40} \). In addition, the small numbers of heterozygotes that we observe in this study limit the power of our tests.

Conclusion

Although we replicated known, strong associations with TG, these data do not replicate the previously reported association between \( \text{SLC25A40} \) variation and TG. Larger datasets with more complete data on fasting and medication use may be required to further explore this association. Furthermore, use of secondary data, such as EHR data, needs extensive quality control and would benefit from more comprehensive data extraction methods.

Acknowledgements

The authors would like to thank all participants in this study as well as individuals who contributed to the eMERGE network.

Authors’ contributions

EAR conceived the analysis plan, performed all analyses and wrote the manuscript. DRC, DSC, JG acquired, maintained and analyzed the data used in this study. ASG, WQW, JCD, QPF, AS, ACS, MDR, JAP, LJR, JJC provided editorial feedback. XF acquired, maintained and performed quality control on the data used in this study. All data is shared under a consortia-wide data use agreement. No administrative permissions and/or licenses were acquired in order to access the clinical/personal patient data used in this research.

Availability of data and materials

Data have been posted to dbGaP. Imputed SNV genotype data are available at Study Accession: phs001584.v1.p1 (https://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study_id=phs001584.v1.p1). Custom capture sequence data are available at Study Accession: phs001616.v1.p1 (https://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study_id=phs001616.v1.p1). Covariates are available at Dataset Accession: phs000072.v1.p1 (https://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/dataset.cgi?study_id=phs000072.v1.p1&phet=9072). Median and maximum TG is available upon request.

Ethics approval and consent to participate

Informed consent was obtained at each site, under each site’s Institutional Review Board (IRB)-approved protocols [24]. Informed consent was obtained for all participants as either written or electronic. Some informed consents were carried out over the phone at Vanderbilt, as part of their collection of a prospective cohort, and this was approved by their IRB. Phone and verbal consents were approved at Geisinger in order to increase study participation, and this was approved by their IRB. The two sequencing centers had IRB-approved protocols that deferred consent to the participating sites. Protocol numbers are as follows: Partners Healthcare (2015P000929), Baylor College of Medicine (BH-40455). The federal-wide assurance number is FWA0002344. Ethics committees and reference numbers for each site are as follows:

CHOP—The Committees for the Protection of Human Subjects, Federal Wide Assurance number FWA00000459, IRB Registration number IRB00000316; Cincinnati, Institution Review Board, Cincinnati Children’s Hospital Medical Center; Columbia, Human Research Protection Office and IRBs, AAAP7926, AAAC09205; Geisinger, Geisinger Institutional Review Board, IRB Registration number IRB00008345; Harvard, Partners Human Research Committee, Protocol number 2015P000983; Mayo, Mayo Clinic Institutional Review Board – IRB-C, DHHS registration numbers: IRB00000200, IRB000003294, IRB000003295, and IRB000005256. Mayo Clinic holds Federal wide Assurance (FWA) number 0005001 covering Mayo Clinic in Rochester, Minnesota, Mayo Clinic in Arizona, Mayo Clinic in Florida, and Mayo Clinic Health Systems, Northwestern, Northwestern University’s Institutional Review Board, IRB number 151299. The network data does not include HIPAA identifiers. Only adult participants are included in this study. All data is shared under a consortia-wide data use agreement. No administrative permissions and/or licenses were acquired in order to access the clinical/personal patient data used in this research.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interest.

Author details

1. Division of Medical Genetics, School of Medicine, University of Washington Medical Center, 1705 NE Pacific St, Box 357720, Seattle, WA 98195, USA. 2. Department of Biomedical Informatics Medical Education, School of Medicine, University of Washington, Seattle, WA, USA. 3. Center for Genetic Medicine, Feinberg School of Medicine, Northwestern University, Chicago, IL, USA. 4. Kaiser Permanente Washington Health Research Institute, Seattle, WA, USA. 5. Department of Biomedical Informatics, Vanderbilt University Medical Center, Nashville, TN, USA. 6. Department of Medicine, Vanderbilt University Medical Center, Nashville, TN, USA. 7. Department of Endocrinology, Department of Pediatrics, Cincinnati Children’s Hospital and the University of Cincinnati, Cincinnati, OH, USA. 8. Genomic Medicine Institute, Geisinger, Danville, PA, 17822, USA. 9. Department of Genetics, University of Pennsylvania, Philadelphia, PA, USA. 10. Center for Applied Genomics, The Children’s Hospital of Philadelphia, Philadelphia, PA, USA. 11. Department of Preventive Medicine, Northwestern University Feinberg School of Medicine, Chicago, IL, USA. 12. Department of Cardiovascular Medicine, Mayo Clinic, Rochester, MN, USA. 13. Gonda Vascular Center, Mayo Clinic, Rochester, MN, USA. 14. Department of Genome Sciences, University of Washington, Seattle, WA, USA.
1. National Cholesterol Education Program (NCEP) Expert Panel on Detection, Evaluation, and Treatment of High Blood Cholesterol in Adults (Adult Treatment Panel III). Third report of the National Cholesterol Education Program (NCEP) expert panel on detection, evaluation, and treatment of high blood cholesterol in adults (Adult Treatment Panel III) final report. Circulation. 2002;106:3143–42.

2. Stampfer MJ, Krauss RM, Ma J, Blanche PJ, Holl LG, Sacks FM, Hennekens CH. A prospective study of triglyceride level, low-density lipoprotein particle diameter, and risk of myocardial infarction. Jama J Am Med Assoc. 1996;276:882–8.

3. Triglyceride Coronary Disease Genetics Consortium and Emerging Risk Factors Collaboration, Sarvar N, Sandhu MS, Ricketts SL, Butterworth AS, Di Angelantonio E, Boekholdt SM, Owuahend W, Watkins H, Samani NJ, et al. Triglyceride-mediated pathways and coronary disease: collaborative analysis of 101 studies. Lancet. 2010;375:1634–9.

4. Hokanson JE, Austin MA. Plasma triglyceride level is a risk factor for cardiovascular disease independent of high-density lipoprotein cholesterol level: a meta-analysis of population-based prospective studies. J Cardiovasc Risk. 1996;3:213–9.

5. Alberti KG, Eckel RH, Grundy SM, Zimmet PZ, Cleeman JI, Donato KA, Fruchart JC, James WP, Loria CM, Smith SC, et al. Harmonizing the metabolic syndrome: a joint interim statement of the International Diabetes Federation Task Force on Epidemiology and Prevention; National Heart, Lung, and Blood Institute; American Heart Association; World Heart Federation; International Atherosclerosis Society, and International Association for the Study of Obesity. Circulation. 2009;120:1640–50.

6. Aboujallal G, Shatyngan N, Wetz R, Costantino T, Abela GS. Current concepts in triglyceride metabolism, pathophysiology, and treatment. Metabolism. 2010;59:1210–20.

7. Mensink RP, Katan MB. Effect of dietary fatty acids on serum lipids and lipoproteins. A meta-analysis of 17 trials. Arterioscler Thromb. 1992;12:911–9.

8. Sopko G, Leon AS, Jacobs DR, Foster N, Moy J, Kuba K, Anderson JT, Casal JC. Rare loss-of-function mutations in ANGPTL family members contribute to familial combined hypolipidemia. Circ Cardiovasc Genet. 2012;5:42–50.

9. Pownall HJ, Ballantyne CM, Kimball KT, Simpson SL, Yeshurun D, Gotto AM. Effect of moderate alcohol consumption on hypertriglyceridemia: a study in the fasting state. Arch Intern Med. 1999;159:981–7.

10. Chair A, Mancini M, February AW, Lewis B. Clinical and metabolic study of hypertriglyceridemia. Lancet. 1972;2:62–4.

11. Willett W, Hennekens CH, Castelli W, Rosner B, Fesken R, Taylor J, Kass EH. A prospective study of triglyceride level, low-density lipoprotein particle diameter, and risk of myocardial infarction. Jama J Am Med Assoc. 1996;276:882–8.

12. Alberti KG, Eckel RH, Grundy SM, Zimmet PZ, Cleeman JI, Donato KA, Fruchart JC, James WP, Loria CM, Smith SC, et al. Harmonizing the metabolic syndrome: a joint interim statement of the International Diabetes Federation Task Force on Epidemiology and Prevention; National Heart, Lung, and Blood Institute; American Heart Association; World Heart Federation; International Atherosclerosis Society, and International Association for the Study of Obesity. Circulation. 2009;120:1640–50.

13. Pennacchio LA, Olivier M, Hubacek JA, Krauss RM, Rubin EM, Cohen JC. Two independent apolipoprotein AS haplotypes influence human plasma triglyceride levels. Hum Mol Genet. 2002;11:3031–8.

14. Talmud PJ, Labracahe J, Deplanque D, Touboul PJ, Amarenco P, Fibrates effect on cardiovascular risk is greater in patients with high triglyceride levels or atherogenic dyslipidemia profile: a systematic review and meta-analysis. J Cardiovasc Pharmacol. 2011;57:267–72.

15. Pennacchio LA, Olivier M, Hubacek JA, Krauss RM, Rubin EM, Cohen JC. Two independent apolipoprotein AS haplotypes influence human plasma triglyceride levels. Hum Mol Genet. 2002;11:3031–8.

16. Romeo S, Yin W, Kozlitina J, Pennacchio LA, Boerwinkle E, Hobbs HH, Goff AM. Relative contribution of variation within the APOC3/APOC4/APOC5 gene cluster in determining plasma triglycerides. Hum Mol Genet. 2002;11:3039–46.

17. Johansen CT, Kathiresan S, Hegele RA. Genetic determinants of plasma triglycerides. J Lipid Res. 2011;52:189–206.

18. Piscotta L, Favaire E, Magnolo L, Simonelli S, Ardini MP, Sallo R, Fancellu T, Zavaroni I, Ardigo D, Bemini F, et al. Characterization of three kindreds with familial combined hyperlipidemia caused by loss-of-function mutations of ANGPTL4. Circ Cardiovasc Genet. 2012;5:42–50.

19. Global Lipids Genetics Consortium, Miller CJ, Schmidt EM, Sengupta S, Peloso GM, Gustafsson S, Kanoni S, Ganna A, Chen J, Buckovich ML, et al. Discovery and refinement of loci associated with lipid levels. Nat Genet. 2013;45:1274–83.

20. van de Worpsteyne AP, van der Graaf Y, de Bakker PI, Asselbergs FW, Spiering W, Visseren FL, SMART Study Group. Rs964184 (APOA5-A4-C3-A1) is related to elevated plasma triglyceride levels, but not to an increased risk for vascular events in patients with clinically manifest vascular disease. PLoS ONE. 2014;9:e101082.

21. Rosenthal EA, Ranchalis J, Crosslin DR, Burt A, Brunzell JD, Motulsky AG, Nickerson DA, NHLBI GO Exome Sequencing Project, Wijsman EM, Jarvik GP. Joint linkage and association analysis with exome sequence data implicates SLCA2A40 in hypertriglyceridemia. Am J Hum Genet. 2013;93(1035):1045.

22. Pagliarini DJ, Calvo SE, Chang B, Sheth SA, Vafi SB, Ong SE, Walford GA, Sugiana C, Boneh A, Chen WK, et al. A mitochondrial protein compendium elucidates complex I disease biology. Cell. 2008;134:112–23.

23. Slabbaert JR, Kuenen S, Swerts J, Maes I, Uytterhoeven V, Kasprowicz J, Fernandes AC, Blust R, Verstreek P, Shaw, the drosophila homolog of SLCA2A40/4, is a mitochondrial carrier that promotes neuronal survival. J Neurosci. 2016;36:1914–29.

24. Fossey R, Kochan D, Winkler E, Pacya JE, Olson J, Thibodeau S, Connolly JJ, Harr M, Behr MA, Prows CA, et al. Ethical considerations related to return of results from genomic medicine projects: the eMERGE network (phase III) experience. J Pers Med. 2018. https://doi.org/10.3390/jpm8100102.

25. Kullo I, Haddad R, Prows CA, Holm L, Sanderson SC, Garrison NA, Sharp RR, Smith ME, Kuvanini H, Bottinger EP, et al. Return of results in the genomic medicine projects of the eMERGE network. Front Genet. 2014;5:50.

26. Goodloe R, Farber-Eger E, Boston J, Crawford DC, Bush WS. Reducing clinical noise for body mass index measures due to unit and transcription errors in the electronic health record. AMIA Jt Summits Transl Sci Proc. 2017;2017:102–11.

27. Green RC, Berg JS, Grody WW, Kalia SS, Korf BR, Martin CL, McGuire AL, Nuissbaum RL, O’Daniel JM, Ormond KE, et al. ACMG recommendations for reporting of incidental findings in clinical exome and genome sequencing. Genet Med. 2013;15:565–74.

28. Kalia SS, Adelman K, Bale SJ, Chung WK, Eng C, Evans JP, Herman GE, Huynagel SB, Klein TE, Korf BR, et al. Recommendations for reporting of secondary findings in clinical exome and genome sequencing, 2016 update (ACMG SF v2.0): a policy statement of the American College of Medical Genetics and Genomics. Genet Med. 2017;19:249–55.

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

30. Poplin R, Ruano-Rubio V, DePristo MA, Fennell TJ, Carneiro MO, Van der Auweve GA, Kling DE, Gauthier LD, Levy-Moonshine A, Roazen D, et al. Scaling accurate genetic variant discovery to tens of thousands of samples. bioRxiv (2017).

31. Rosenthal EA, Gordon AS, Stanaway IB, Crosslin DR, Albertson-Junkans L, Larson EB, Carrell D, Jarvik GP. Replication of association between triglyceride level and SLC25A40 in a multi-ancestry sample. Poster presented at the American Society of Human Genetics (ASHG) Annual meeting. San Diego, CA (2018).

32. Cooper GM, Stone EA, Asimengos N, NISC Comparative Sequencing Program, Green ED, Batzoglu S, Sidow A. Distribution and intensity of constraint in mammalian genomic sequence. Genome Res. 2005;15:901–13.

33. Lek M, Karczewski KJ, Minikel EV, Samocha KE, Banks E, Fennell T, O’Donnell-Luria AH, Ware JS, Hill AJ, Cummings BB, et al. Analysis of protein-coding genetic variation in 60,706 humans. Nature. 2016;536:285–91.

34. Karceski J, Francioli LC, et al. Variation across 141,456 human exomes and genomes reveals the spectrum of loss of function intolerance across human protein coding genes. bioRxiv (2019).

35. Cook RD, Weisberg S. Residuals and influence in regression. New York: Chapman and Hall; 1982.
37. Teslovich TM, Musunuru K, Smith AV, Edmondson AC, Stylianou IM, Koseki M, Pirruccello JP, Ripatti S, Chasman DI, Willer CJ, et al. Biological, clinical and population relevance of 95 loci for blood lipids. Nature. 2010;466:707–13.

Publisher’s Note
Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.