

RESEARCH

Open Access



# Frequency and phenotype consequence of *APOC3* rare variants in patients with very low triglyceride levels

Dana C. Crawford<sup>1\*</sup>, Nicole A. Restrepo<sup>1</sup>, Kirsten E. Diggins<sup>2</sup>, Eric Farber-Eger<sup>3</sup> and Quinn S. Wells<sup>4</sup>

From The 7th Translational Bioinformatics Conference  
Los Angeles, CA, USA. 29 September - 01 October 2017

## Abstract

**Background:** High levels of triglycerides (TG  $\geq 200$  mg/dL) are an emerging risk factor for cardiovascular disease. Conversely, very low levels of TG are associated with decreased risk for cardiovascular disease. Precision medicine aims to capitalize on recent findings that rare variants such as *APOC3* R19X (rs76353203) are associated with risk of disease, but it is unclear how population-based associations can be best translated in clinical settings at the individual-patient level.

**Methods:** To explore the potential usefulness of screening for genetic predictors of cardiovascular disease, we surveyed BioVU, the Vanderbilt University Medical Center's biorepository linked to de-identified electronic health records (EHRs), for *APOC3* 19X mutations among adult European American patients (> 45 and > 55 years of age for men and women, respectively) with the lowest percentile of TG levels. The initial search identified 262 patients with the lowest TG levels in the biorepository; among these, 184 patients with sufficient DNA and the lowest TG levels were chosen for Illumina ExomeChip genotyping.

**Results:** A total of two patients were identified as heterozygotes of *APOC3* R19X for a minor allele frequency (MAF) of 0.55% in this patient population. Both heterozygous patients had only a single mention of TG in the EHR (31 and 35 mg/dL, respectively), and one patient had evidence of previous cardiovascular disease.

**Conclusions:** In this patient population, we identified two patients who were carriers of the *APOC3* 19X null variant, but only one lacked evidence of disease in the EHR highlighting the challenges of inclusion of functional or previously associated genetic variation in clinical risk assessment.

**Keywords:** Precision medicine, Triglycerides, Biobank, Electronic health records, *APOC3*

## Background

Personalized or precision medicine is meant to distinguish tailored treatment from trial and error. The concept of precision medicine is not new and has been in practice arguably since the dawn of modern medicine [1]. Health care providers have long collected detailed data on patients, ranging from basic personal histories to technical laboratory assays and diagnostic procedures,

to provide specific diagnoses and treatments. These tools ordered in the precision medicine setting are constantly evolving, for example, the evolution of myocardial infarction diagnosis [2], resulting in high resolution and, in some cases, highly predictive individualized data.

Today's concept of precision medicine has evolved to specifically include the genetic profile of a patient in the prevention, diagnosis, and treatment of disease [3, 4]. Previous proxies for genetic profiles such as sex, race/ethnicity, family history, and response to therapy are now being augmented by Clinical Laboratory Improvement Amendments-certified genotyping and sequencing

\* Correspondence: [dana.crawford@case.edu](mailto:dana.crawford@case.edu)

<sup>1</sup>Department of Population and Quantitative Health Sciences, Institute for Computational Biology, Case Western Reserve University, 2103 Cornell Road, Wolstein Research Building, Suite 2-527, Cleveland, OH 44106, USA  
Full list of author information is available at the end of the article



at both the targeted and whole genome level. Indeed, technological advances in high-throughput genomics coupled with their rapid decreases in costs have made generating the data almost trivial, and the emergence of electronic health records (EHRs) in part through the HITECH Act [5, 6] make it possible to effectively deliver personalized medicine to the patient.

A major challenge in the delivery of personalized medicine is not the collection of data, but the interpretation of the data. Large-scale population sequencing studies have demonstrated that potentially functional variants exist in all DNA samples sequenced, but the biological and statistical data lag in filtering the potential from truly functional, and even less data are available to direct the course of clinical action based on these genotypes. As such, major areas of active research focus on methods to translate genomic discoveries into clinical applications [7]. This broad area of investigation includes research on who to test, what to test, how to test, what to report, how to report, and how to measure its effectiveness, to name a few.

To begin to address some of these gaps in research, we have undertaken a pilot study of genotyping for a loss of function variant, *APOC3* R19X (rs76353203), in a targeted population of 184 European American adults (men >45 and women >55 year of age) with very low triglyceride levels in BioVU, the Vanderbilt University Medical Center (VUMC)'s biorepository linked to de-identified EHRs. Triglyceride levels (TG) are a common biomarker measured in the clinic, and patients with extreme triglyceride (TG) levels may be flagged for further evaluation for cardiovascular disease risk assessment (TG  $\geq$ 200 mg/dL). Recent studies have identified a loss-of-function variant in *APOC3* (R19X or rs76353203) associated with low triglyceride levels and diminished post-prandial lipemia in heterozygous carriers [8] and improved clearance of plasma TGs after a fatty meal in homozygous carriers [9]. Although rare in the general population [10], we hypothesized that the loss-of-function allele would be at an increased frequency in this extreme population. Based on previous reports, we also hypothesized that evidence of cardiovascular disease would be absent in EHRs of these *APOC3* R19X carriers with very low TG levels.

## Methods

### Study population

The study population presented here is from BioVU, the VUMC's biorepository linked to de-identified EHRs. BioVU operations [11] and ethical oversight [12] have been previously described. Briefly, DNA is extracted from discarded blood drawn for routine clinical care at Vanderbilt outpatient clinics in Nashville, Tennessee and surrounding areas. The DNA samples are linked to a

de-identified version of the patient's EHR. The data in this study were de-identified in accordance with provisions of Title 45, Code of Federal Regulations, part 46 (45 CFR 46); therefore, this study was considered non-human subjects research by the Vanderbilt University Internal Review Board.

### Phenotyping

The de-identified EHR contains both structured (International Classification of Diseases, Ninth Revision, Clinical Modification (ICD-9-CM) billing codes; current procedural terminology (CPT) codes; problems lists; labs) and unstructured (clinical free text) data accessible for electronic phenotyping. We extracted all available labs for triglyceride levels in September 2012 for European American adults (>45 years of age of men and >55 years of age for women) and examined individuals whose median TG levels constituted the lowest 1% of BioVU at the time of data extraction. A total of 262 individuals were identified. Manual review of the clinical text for 30 random patients with low TG levels prior to selection for genotyping failed to identify obvious documented diagnoses or notes that may have led to very low TG levels in these patients. From these 262 individuals, 184 were selected for Illumina HumanExome BeadChip genotyping based on DNA quality and quantity, and preference was given to individuals with more than one triglyceride level reflecting consistently low TG levels.

The de-identified EHRs for the 184 patients genotyped on the Illumina HumanExome BeadChip were re-examined in July 2015 for evidence of myocardial infarction, revascularization, and other heart disease. Myocardial infarction was defined by mention of ICD-9-CM codes 410 or 410.\* or a problem list mention of "MI" or "myocardial infarction." Revascularization was defined by CPT codes (Table 1). Other heart disease was defined as a problem list mention of "coronary artery disease," "CAD," "coronary heart disease," or "CHD."

### Genotyping and statistical methods

Vanderbilt Technologies for Advanced Genomics (VANTAGE) genotyped 184 BioVU samples and eight International HapMap reference samples using the Illumina HumanExome BeadChip (v1.0 for 48 samples and v1.2 for 144 samples). As recommended by other research groups with Exome BeadChip experience [13], genotypes for these 192 samples were called as part of larger, ongoing Exome BeadChip projects in BioVU genotyped by VANTAGE. *APOC3* R19X (rs76353203/exm957809) was directly assayed by the Exome BeadChip, and we extracted these called genotypes for further analysis.

We also extracted genotypes for SNPs previously associated with incident myocardial infarction in European Americans [14]. Of the 46 previously associated SNPs,

**Table 1** Current Procedural Terminology codes used to define revascularization among European American patients in BioVU with very low triglyceride levels

| CPT code                 | Description   |
|--------------------------|---|
| 33510–33514; 33516       | Coronary artery bypass grafting with venous grafting only (1–6 or more grafts)  |
| 33515                    | Coronary artery bypass (old code)   |
| 33517–33519; 33521–33523 | Coronary artery bypass grafting with venous and arterial grafting (1–6 or more vein grafts); billed in conjunction with 33533–33536 (33517–33523 cannot be billed alone). |
| 33520                    | Coronary artery bypass (old code)   |
| 33534–33536              | Coronary artery bypass grafting with arterial grafting only (2–4 or more grafts)  |
| 92980–92981              | Transcatheter placement of an intracoronary stent, percutaneous, with or without other therapeutic intervention, any method (single vessel and each additional vessel)    |
| 92982; 92984             | Percutaneous transluminal coronary balloon angioplasty (single vessel and each additional vessel)   |
| 92995; 92996             | Percutaneous transluminal coronary atherectomy, by mechanical or other method, with or without balloon angioplasty (single vessel and each additional vessel)             |

37 were directly assayed by the Exome BeadChip (Additional file 1: Table S1). We then calculated both unweighted and weighted genetic risk scores (GRS) based on the genotypes of these 37 SNPs. Unweighted GRS were calculated by counting the number of risk alleles per individual. Weighted GRS were calculated based on counts of risk alleles multiplied or weighted by odds ratios recently reported for European American cases of coronary artery disease in comparison with controls [15].

## Results

As detailed in the Methods section, 262 European American men > 45 and women > 55 years of age had the lowest TG levels in BioVU in 2012. Slightly less than half (44%) of the patient sample was female, and the average birth decade was the 1940s. Approximately half of the patients (53.1%) with the lowest TG levels had more than one TG available in the EHR. For those with more than one TG level, the median values were calculated. The overall median TG level in this lowest 1% was 36 mg/dL.

The genotyped study population characteristics are given in Table 2. Like the 262 patients considered, the patients genotyped were majority male born in the 1940s and 1950s. The mean body mass index was 24.8 kg/m<sup>2</sup>, which is considered within the normal range. The first mention of TG in the clinical record was on average 39.3 mg/dL. The median of all TG levels available for this patient population was 36 mg/dL, ranging from 13.5 to 61 mg/dL (Fig. 1).

Among the 184 patients with low TG levels in BioVU genotyped here, 14.13%, 8.15%, and 12.5% had evidence in the EHR of a myocardial infarction, revascularization, and other heart disease, respectively. Only six out of 184 patients (3.26%) had evidence of all three. The unweighted and weighted GRS for patients with events (myocardial infarction or revascularization) or evidence for other heart disease were higher compared with patients without

events or evidence for other heart disease, although these differences were not statistically significant (Table 3).

We next examined the frequency of *APOC3* R19X (rs76353203) in this patient sample of extremely low TG levels. Among 182 patients successfully genotyped for this variant, two heterozygotes were identified for a sample allele frequency of 0.55%. The allele frequency for the loss-of-function allele (T) in this sample of very low TG levels is 6.9 to 27.5 fold higher compared with allele frequency estimates for American adults drawn from the general population [10, 16] and three to 10 fold lower in frequency compared with isolated populations [8, 17, 18] (Table 4).

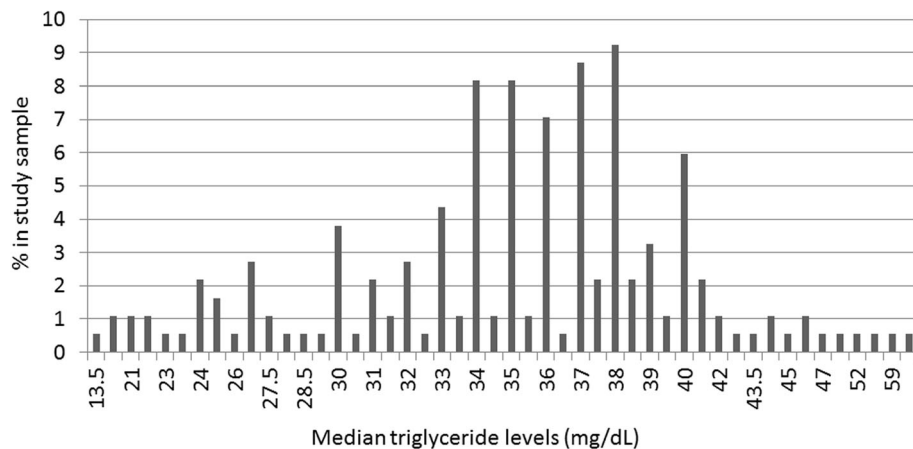
Previous studies in outbred [10] and isolated [8, 17] populations suggest that the 19X mutation arose once on a single haplotype background. We therefore examined the diplotypes spanning *APOA1/C3/A4/A5* gene cluster assayed by the Illumina HumanExome Beadchip for evidence of a single haplotype background containing the 19X mutation. Of the 78 SNPs assayed (spanning

**Table 2** Study population characteristics

|   |              |
|---|--------------|
| Female, %                                       | 43.5         |
| Decade of birth, %                              |              |
| 1910  | 0.5          |
| 1920  | 9.2          |
| 1930  | 14.7         |
| 1940  | 28.3         |
| 1950  | 37.5         |
| 1960  | 9.8          |
| Mean (±SD) body mass index (kg/m <sup>2</sup> ) | 24.8 (±4.7)  |
| Mean (±SD) first TG level (mg/dL)               | 39.3 (±18.4) |

Study population characteristics (sex, decade of birth, average body mass index, and average first mention of triglyceride levels in patient's electronic health record) are given for the *genotyped* 184 patients with the lowest 1% triglyceride levels in BioVU among European American men and women > 45 and > 55 years of age, respectively

Abbreviations: SD standard deviation, TG triglyceride



**Fig. 1** Distribution of low triglyceride levels among European American adults in BioVU. A total of 184 European American adults (men > 45 years and women > 55 years) with at least one triglyceride level in the lowest 1% of BioVU were genotyped on the Illumina HumanExome BeadChip for *APOC3* R19X. The frequency (expressed as percent in the study population) is given on the y-axis and the median triglyceride levels (in mg/dL) are given on the x-axis

chr11: 116619073–116,707,837), seven SNPs overlapped with variants used to infer haplotypes in non-Hispanic whites of the Third National Health and Nutrition Examination Survey (NHANES III): rs28927680, rs964184, rs12286037, rs5110, rs675, rs5104, and rs76353203. We found that one 19X carrier was homozygous at all variants in this region that passed quality control except for 19X, and that this variant-containing haplotype background was similar to 19X backgrounds in NHANES III (G-C-C-C-A-T-T). In contrast, the second 19X carrier was heterozygous at eight sites including rs76343203. Interestingly, one of these eight heterozygous sites included rs138326449 (IVS2 + 1G > A), another *APOC3* rare variant associated with decreased levels of TG [16, 19, 20]. A query for rs138326449 in the 184 patients with very low triglycerides revealed an additional carrier for this variant. Both carriers of rs138326449 are, with the inclusion of this rare variant, heterozygous at more than one site, and one carrier has missing data (26/78 sites) at this gene cluster; therefore, his or her haplotypes could not be unambiguously inferred with confidence. Much like R19X, the frequency of rs138326449 (0.57%) is higher in

this patient population of low TG levels compared with outbred population estimates [16, 19]. A third rare *APOC3* variant (rs147210663) associated with low TG levels [16] failed genotyping in this patient population.

As expected when stratified by *APOC3* R19X genotype (Fig. 2), the mean TG level in carriers (33 mg/dL; 2.83 standard deviation) was lower compared with the non-carriers (39.51 mg/dL; 18.52 standard deviation). However, the difference was not statistically significant ( $p = 0.11$ ) in this sample of adults with very low TG levels. Similarly, the mean TG level in IVS2 + 1G > A carriers was 32.5 mg/dL (2.12 standard deviation) for first-mentioned TG level.

Previous observational studies have suggested that TG levels are associated with risk of cardiovascular disease [21]. More recently *APOC3* 19X carrier status among other *APOC3* mutations was associated with lower risk of coronary heart disease [16]. Of the two *APOC3* 19X carriers, one had no evidence in the EHR of myocardial infarction, revascularization, or other heart disease. One *APOC3* 19X carrier had evidence in the EHR of all three. A more detailed review of the de-identified EHRs of the

**Table 3** Genetic risk scores, unweighted and weighted, by case status among European American patients with very low triglyceride levels

|                | All patients with very low TG levels (n = 184) | Patients with no evidence of MI, revascularization, or other heart disease (n = 144) | Patients with evidence of MI (n = 26) | Patients with evidence of revascularization (n = 15) | Patients with evidence of other heart disease (n = 23) | Patients with evidence of MI, revascularization, and other heart disease (n = 6) |
|----------------|--|--|---------------------------------------|--|--|--|
| Unweighted GRS | 38.97 (3.36)                                   | 38.78 (3.36)   | 39.46 (3.47)                          | 39.07 (3.95)   | 39.08 (3.57)   | 38.00 (4.38)   |
| Weighted GRS   | 41.97 (3.63)                                   | 41.77 (3.63)   | 42.54 (3.72)                          | 42.08 (4.32)   | 42.14 (3.82)   | 41.02 (4.84)   |

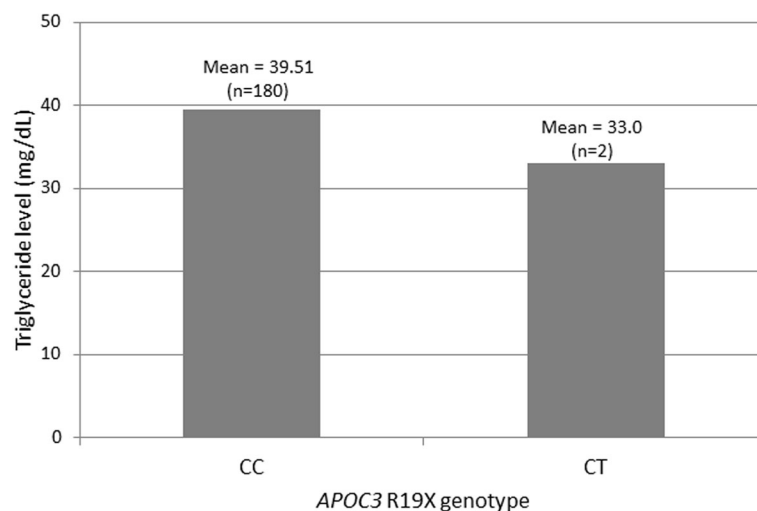
We calculated unweighted and weighted genetic risk scores (GRS) based on 37 SNPs and previous association estimates in European Americans with and without coronary artery disease (Additional file 1: Table S1). Unweighted GRS were calculated as counts of risk alleles per patient, and weighted GRS were calculated using the pooled odds ratios from CARDIoGRAM [15]. Shown are the means (standard deviations) for unweighted and weighted GRS for the total study population as well as by cases status. Although unweighted and weighted GRS were higher among all patient groups with events or other heart conditions compared with patients lacking evidence of events or other heart disease, these differences were not statistically significant (unpaired t-tests;  $p > 0.05$ )

**Table 4** *APOC3* R19X frequency, by population

| Population  | Sample size | Carriers identified (Overall allele frequency) | Allele frequency in European-descent populations | PubMed ID or website   |
|---|-------------|--|--|--|
| European American adults with very low triglyceride levels  | 184         | 2 (0.55%)                                      | 0.55%  | - (present study)  |
| Pennsylvania Amish  | 2503        | 140 (5.6%)                                     | 5.6%   | 19074352   |
| Greek isolate   | 1219        | 48 (1.9%)                                      | 1.9%   | 24343240   |
| Greek isolate   | 1087        | 34 (1.42%)                                     | 1.42%  | 27146844   |
| Americans regardless of health status   | 19,613      | 31 (0.08%)                                     | 0.20%  | 25363704   |
| Exome Aggregation Consortium (ExAC)   | 60,103      | 83 (0.07%)                                     | 0.046%   | 27535533 ( <a href="http://exac.broadinstitute.org/">http://exac.broadinstitute.org/</a> accessed May 2017)      |
| National Heart, Lung, and Blood Institute (NHLBI) Grand Opportunity (GO) Exome Sequencing Project (ESP) | 6495        | 3 (0.02%)                                      | 0.035%   | 24941081 ( <a href="http://evs.gs.washington.edu/EVS/">http://evs.gs.washington.edu/EVS/</a> accessed June 2015) |
| Ohio and Indiana Amish  | 1113        | 0 (-)  | -  | 25363704   |
| 1000 Genomes Project  | 2500        | 0 (-)  | -  | ( <a href="http://useast.ensembl.org/index.html">http://useast.ensembl.org/index.html</a> accessed June 2015)    |
| European Americans from Baltimore   | 214         | 0 (-)  | -  | 19074352   |

two *APOC3* 19X carriers was performed to identify other possible cardiovascular disease risk factors. The female 19X carrier was born in the 1940s, and her EHR contained a medical history significant for remote breast cancer treated with surgery and radiation, controlled hypertension, and overweight (BMI ~ 28 kg/m<sup>2</sup>). This female 19X carrier had never smoked cigarettes, and there was no evidence of coronary artery disease or myocardial infarction in her EHR. A single assessment of lipids was available for this female 19X carrier: low-density lipoprotein cholesterol (LDL-C) 116 mg/dL, high density

lipoprotein cholesterol (HDL-C) 52 mg/dL, and TG 35 mg/dL. The female 19X carrier had unweighted and weighted GRS of 33 and 35.39, respectively. The second *APOC3* 19X carrier was a male born in the 1920s. The male 19X carrier had a past medical history of uncontrolled hypertension, was overweight (BMI 27 kg/m<sup>2</sup>), and was a prior smoker. The male carrier had an extensive history of cardiac disease including atrial fibrillation, coronary artery disease with prior coronary artery bypass grafting, myocardial infarction, and ischemic cardiomyopathy (ejection fraction 30%) with heart failure. The



**Fig. 2** Mean triglyceride levels by *APOC3* R19X genotype. A total of 184 European American adults (men > 45 years and women > 55 years) with at least one triglyceride level in the lowest 1% of BioVU were genotyped on the Illumina HumanExome BeadChip for *APOC3* R19X. Two samples failed genotyping. The means for the first mentioned triglyceride level (y-axis) were calculated for the non-carriers (CC genotype) and carriers (CT genotype) at *APOC3* R19X (x-axis). Although the mean triglyceride level in carriers (33 mg/dL; 2.83 standard deviation) was lower compared with the non-carriers (39.51 mg/dL; 18.52 standard deviation), the difference between the two is not statistically significant (two-sided t-test assuming unequal variances;  $p = 0.11$ )



male 19X carrier was not treated with statins due to reported intolerance, and a single measurement of lipids was available: LDL-C 69 mg/dL, HDL-C 41 mg/dL, and TG 31 mg/dL. The male 19X carrier had unweighted and weighted GRS of 32 and 34.21, respectively. Neither *APOC3* 19X carrier has died as of 2015.

The male 19X carrier was also a carrier for *IVS2 + 1G > A*, the only potentially compound heterozygote described in the literature to date. The other *IVS2 + 1G > A* carrier was a female born in the 1950s. The female *IVS2 + 1G > A* carrier was normal weight (BMI 23 kg/m<sup>2</sup>) with two mentions of TG levels (34 and 23 mg/dL) in the EHR. This female *IVS2 + 1G > A* carrier has no evidence of myocardial infarction, revascularization, or other heart disease in the EHR.

## Discussion

We evaluated 184 adult European American patients with very low TG levels extracted from EHRs for the presence of the loss-of-function allele (19X) for *APOC3* rs76353203. Overall, we identified two carrier patients, and as hypothesized, the resulting allele frequency of *APOC3* 19X was higher in this extreme patient population compared with the general population [10]. Neither carrier patient had the lowest TG levels among this patient population. A review of the EHR revealed only one of the two *APOC3* 19X carriers was free of myocardial infarction, revascularization, and other heart disease. Coincidentally, the *APOC3* 19X carrier with evidence of cardiovascular disease was also a carrier of *IVS2 + 1G > A*, representing to our knowledge potentially the first compound heterozygote for these mutations in the literature [16].

Based on the current literature [16], we would expect that the addition of these genomic data to the EHR would assist a physician in the assessment of the carriers' risk of future cardiovascular disease. *APOC3* R19X was originally identified in a genome-wide association study (GWAS) of TG levels in the Pennsylvania Amish where a variant in linkage disequilibrium with R19X was significantly associated with decreased TG levels [8]. Follow-up sequencing revealed the loss of function mutation in *APOC3* likely responsible for the GWAS findings [8], and subsequent studies in both isolated [17, 18] and outbred [10, 16] populations have confirmed the strong association between lower TG levels and 19X carriers. More recently, prospective epidemiologic studies have demonstrated that 19X carriers have lower rates of cardiovascular disease compared with non-carriers [16].

Interestingly, one of the two *APOC3* 19X carriers identified here has evidence in the EHR of a myocardial infarction, revascularization, and other heart disease. Also, apart from sharing low TG levels, the two 19X carriers had different cardiovascular risk profiles. While the statistical evidence for the association between lower TG

levels and lower risk of coronary heart disease is strong at the population level, these data highlight the difficulty in translating a genetic association finding in the clinic for risk prediction at the patient level as envisioned for precision medicine. These data also highlight the genetic and environmental heterogeneity that drives cardiovascular disease risk. Thus, the addition of *APOC3* R19X in a clinical setting may contribute to the patient's risk assessment for cardiovascular disease, but it is not absolute and must be considered with other genetic and environmental risk factors [22].

We specifically targeted adult European Americans with low TG levels for *APOC3* R19X genotyping. The loss-of-function variant is common in isolated populations such as the Pennsylvania Amish [8] and Greek isolates [17, 18], but a study in NHANES confirmed the mutation is rare in the general population [10]. The frequency of 19X also varies by race/ethnicity. In NHANES [10] and in other studies [16], 19X is exceedingly rare in African Americans or African-descent populations compared with European-descent populations. In the present study of European Americans with low TG levels, we observed that the frequency of 19X was higher in this patient population compared with the general population, an observation consistent with the known genetic epidemiology of this loss-of-function variant. Although the carriers identified in this study had lower mean TG levels compared with non-carriers, we did not observe a statistically significant association most likely due to the fact that all patients genotyped already had low TG levels. Also, we only identified two carriers resulting in low statistical power.

The strategy of genotyping or targeting individuals with extreme phenotypes has been a popular and successful strategy in genetic epidemiology for gene discovery for many years [23]. Indeed, in the field of cardiovascular genetics, sequencing individuals with extreme LDL-C, HDL-C, and TG levels in multiple populations has identified several genetic variants and potential drug targets such as *PCSK9* [24]. An analogous strategy could be implemented in a clinical setting to augment the EHR with specific genotypes for an individual patient's risk assessment. For example, if a patient presents with an extreme lipid level, a panel of known functional variants (missense and loss-of-function) could be ordered for genotyping and added to the EHR to inform risk assessment for future cardiovascular events in that patient. A major advantage of this targeted approach is that the genetic assays would be ordered only on a fraction of the patients (e.g., patients with extreme labs) making it cost-effective compared with offering the panel to all patients regardless of lab results.

There are major disadvantages to a targeted approach to augmenting the EHR with genomic data. For most human traits and diseases, the known functional or

strongly associated variants were discovered in European-descent populations [25, 26]. As such, diverse populations such as African Americans and Hispanics may not benefit from a European-centric genotyping panel. And, even among European-descent populations the catalog of genotype-phenotype associations is far from complete or strongly predictive of clinical events [27]. As technology improves to sort functional from neutral variants [28, 29], all patients may benefit from whole genome sequencing.

## Conclusions

Further work is needed in developing appropriate tools for EHR integration and delivery of clinical decision support [30] for this and other clinically relevant genetic variants as envisioned in an era of precision medicine.

## Additional file

**Additional file 1: Table S1.** Genetic variants previously associated with risk of coronary artery disease in European Americans. (DOCX 27 kb)

## Abbreviations

BMI: Body mass index; CAD: Coronary artery disease; CHD: Coronary heart disease; CPT: Current procedural terminology; EHR: Electronic health record; GRS: Genetic risk score; GWAS: Genome-wide association study; HDL-C: High density lipoprotein cholesterol; ICD-9-CM: International Classification of Diseases, Ninth Revision, Clinical Modification; LDL-C: Low-density lipoprotein cholesterol; MAF: Minor allele frequency; NHANES: National Health and Nutrition Examination Survey; TG: Triglycerides; VANTAGE: Vanderbilt Technologies for Advanced Genomics; VUMC: Vanderbilt University Medical Center

## Funding

The cost of publication was funded by Case Western Reserve University' Institute for Computational Biology. The dataset (s) used for the analyses described were obtained from Vanderbilt University Medical Center's BioVU which is supported by institutional funding and by the Vanderbilt CTSA grant funded by the National Center for Research Resources, Grant UL1 RR024975-01, which is now at the National Center for Advancing Translational Sciences, Grant 2 UL1 TR000445-06.

## Availability of data and materials

The data that support the findings of this study are available from Vanderbilt University Medical Center but restrictions apply to the availability of these data, which were used under license for the current study, and so are not publicly available. Data are however available from the authors upon reasonable request and with permission of Vanderbilt University Medical Center.

## About this supplement

This article has been published as part of *BMC Medical Genomics* Volume 11 Supplement 3, 2018: Selected articles from the 7th Translational Bioinformatics Conference (TBC 2017): medical genomics. The full contents of the supplement are available online at <https://bmcmedgenomics.biomedcentral.com/articles/supplements/volume-11-supplement-3>.

## Authors' contributions

DCC designed the study. KED, EF-E, and QSW collected and prepared the data. DCC and NAR performed analyses. DCC drafted the manuscript. DCC, NAR, and QSW were major contributors in revising the manuscript critically for all important intellectual content. All authors gave approval to the final version of the manuscript and agreed to be accountable to all aspects of the work.

## Ethics approval and consent to participate

The data in this study were de-identified in accordance with provisions of Title 45, Code of Federal Regulations, part 46 (45 CFR 46); therefore, this study was considered non-human subjects research by the Vanderbilt University Internal Review Board.

## Consent for publication

Not applicable.

## Competing interests

The authors declare that they have no competing interests.

## Publisher's Note

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

## Author details

<sup>1</sup>Department of Population and Quantitative Health Sciences, Institute for Computational Biology, Case Western Reserve University, 2103 Cornell Road, Wolstein Research Building, Suite 2-527, Cleveland, OH 44106, USA. <sup>2</sup>Cancer Biology, Vanderbilt University School of Medicine, Nashville, TN, USA. <sup>3</sup>Vanderbilt Institute for Clinical and Translational Research, Vanderbilt University Medical Center, Nashville, TN, USA. <sup>4</sup>Departments of Medicine and Pharmacology, Vanderbilt University Medical Center, Nashville, TN, USA.

Published: 14 September 2018

## References

- Murray JF. Personalized medicine: been there, done that, always needs work! *Am J Respir Crit Care Med*. 2012;185(12):1251–2. <https://doi.org/10.1164/rccm.201203-0523ED>.
- Cervellin G, Lippi G. Of MIs and men—a historical perspective on the diagnostics of acute myocardial infarction. *Semin Thromb Hemost*. 2014; 40(5):535–43. <https://doi.org/10.1055/s-0034-1383544>.
- Hamburg MA, Collins FS. The path to personalized medicine. *N Engl J Med*. 2010;363(4):301–4. <https://doi.org/10.1056/NEJMp1006304>.
- Feero W, Guttmacher AE, Collins FS. The genome gets personal—almost. *JAMA*. 2008;299(11):1351–2. <https://doi.org/10.1001/jama.299.11.1351>.
- Blumenthal D. Launching HITECH. *N Engl J Med*. 2010;362(5):382–5. <https://doi.org/10.1056/NEJMp0912825>.
- Jensen PB, Jensen LJ, Brunak S. Mining electronic health records: towards better research applications and clinical care. *Nat Rev Genet*. 2012;13(6): 395–405. <https://doi.org/10.1038/nrg3208>.
- Green ED, Guyer MS. Charting a course for genomic medicine from base pairs to bedside. *Nature*. 2011;470(7333):204–13. <https://doi.org/10.1038/nature09764>.
- Pollin TI, Dammott CM, Shen H, Ott SH, Shelton J, Horenstein RB, et al. A null mutation in human APOC3 confers a favorable plasma lipid profile and apparent Cardioprotection. *Science*. 2008;322(5908):1702–5. <https://doi.org/10.1126/science.1161524>. PMC2673993
- Saleheen D, Natarajan P, Armean IM, Zhao W, Rasheed A, Khetarpal SA, et al. Human knockouts and phenotypic analysis in a cohort with a high rate of consanguinity. *Nature*. 2017;544(7649):235–9. <https://doi.org/10.1038/nature22034>. PMC5600291
- Crawford DC, Dumitrescu L, Goodloe R, Brown-Gentry K, Boston J, McClellan B Jr, et al. Rare variant APOC3 R19X is associated with cardio-protective profiles in a diverse population-base survey as part of the epidemiologic architecture for genes linked to environment (EAGLE) study. *Circ Cardiovasc Genet*. 2014; 7(6):848–53. <https://doi.org/10.1161/CIRCGENETICS.113.000369>. PMC4305446
- Roden DM, Pulley JM, Basford MA, Bernard GR, Clayton EW, Balsler JR, et al. Development of a large-scale De-identified DNA biobank to enable personalized medicine. *Clin Pharmacol Ther*. 2008;84(3):362–9. <https://doi.org/10.1038/clpt.2008.89>. PMC3763939
- Pulley J, Clayton E, Bernard GR, Roden DM, Masys DR. Principles of human subjects protections applied in an opt-out, De-identified Biobank. *Clin Transl Sci*. 2010;3(1):42–8. <https://doi.org/10.1111/j.1752-8062.2010.00175.x>. PMC3075971
- Grove ML, Yu B, Cochran BJ, Haritunians T, Bis JC, Taylor KD, et al. Best practices and joint calling of the HumanExome BeadChip: the CHARGE consortium. *PLoS One*. 2013;8(7):e68095. <https://doi.org/10.1371/journal.pone.0068095>. PMC3709915

14. Dehghan A, Bis JC, White CC, Smith AV, Morrison AC, Cupples LA, et al. Genome-wide association study for incident myocardial infarction and coronary heart disease in prospective cohort studies: the CHARGE consortium. *PLoS One*. 2016;11(3):e0144997. <https://doi.org/10.1371/journal.pone.0144997>. PMC4780701
15. Deloukas P, Kanoni S, Willenborg C, Farrall M, Assimes TL, Thompson JR, et al. Large-scale association analysis identifies new risk loci for coronary artery disease. *Nat Genet*. 2013;45(1):25–33. <https://doi.org/10.1038/ng.2480>. PMC3679547
16. TG and HDL Working Group of the Exome Sequencing Project, National Heart, Lung, and Blood Institute, Crosby J, Peloso GM, Auer PL, Crosslin DR, et al. Loss-of-function mutations in APOC3, triglycerides, and coronary disease. *N Engl J Med*. 2014;371(1):22–31. <https://doi.org/10.1056/NEJMoa1307095>. PMC4180269
17. Tachmazidou I, Dedoussis G, Southam L, Farmaki AE, Ritchie GRS, Xifara DK, et al. A rare functional cardioprotective APOC3 variant has risen in frequency in distinct population isolates. *Nat Commun*. 2013;4 <https://doi.org/10.1038/ncomms3872>. PMC3905724
18. Gilly A, Ritchie GR, Southam L, Farmaki A-E, Tsafantakis E, Dedoussis G, et al. Very low-depth sequencing in a founder population identifies a cardioprotective APOC3 signal missed by genome-wide imputation. *Hum Mol Genet*. 2016;25(11):2360–5. <https://doi.org/10.1093/hmg/ddw088>. PMC5081052
19. Timpson NJ, Walter K, Min JL, Tachmazidou I, Malerba G, Shin S-Y, et al. A rare variant in APOC3 is associated with plasma triglyceride and VLDL levels in Europeans. *Nat Commun*. 2014;5:4871. <https://doi.org/10.1038/ncomms5871>. PMC4167609
20. Drenos F, Davey Smith G, Ala-Korpela M, Kettunen J, Würtz P, Soininen P, et al. Metabolic characterization of a rare genetic variation within APOC3 and its lipoprotein lipase-independent effects. *Circ Cardiovasc Genet*. 2016; 9(3):231–9. <https://doi.org/10.1161/circgenetics.115.001302>. PMC4920206
21. Miller M, Stone NJ, Ballantyne C, Bittner V, Criqui MH, Ginsberg HN, et al. Triglycerides and cardiovascular disease: a scientific statement from the American Heart Association. *Circulation*. 2011;123(20):2292–333. <https://doi.org/10.1161/CIR.0b013e3182160726>.
22. Khara AV, Emdin CA, Drake I, Natarajan P, Bick AG, Cook NR, et al. Genetic risk, adherence to a healthy lifestyle, and coronary disease. *N Engl J Med*. 2016;375(24):2349–58. <https://doi.org/10.1056/NEJMoa1605086>. PMC5338864
23. Plomin R, Haworth CMA, Davis OSP. Common disorders are quantitative traits. *Nat Rev Genet*. 2009;10(12):872–8. <https://doi.org/10.1038/nrg2670>.
24. Cohen J, Pertsemliadis A, Kotowski IK, Graham R, Garcia CK, Hobbs HH. Low LDL cholesterol in individuals of African descent resulting from frequent nonsense mutations in PCSK9. *Nat Genet*. 2005;37(2):161–5. <https://doi.org/10.1038/ng1509>.
25. Rosenberg NA, Huang L, Jewett EM, Szpiech ZA, Jankovic I, Boehnke M. Genome-wide association studies in diverse populations. *Nat Rev Genet*. 2010;11(5):356–66. <https://doi.org/10.1038/nrg2760>. PMC3079573
26. Popejoy AB, Fullerton SM. Genomics is failing on diversity. *Nature*. 2016; 538(7624):161–4. <https://doi.org/10.1038/538161a>. PMC5089703
27. Assimes TL, Salfati EL, Del Gobbo LC. Leveraging information from genetic risk scores of coronary atherosclerosis. *Curr Opin Lipidol*. 2017;28(2):104–12. <https://doi.org/10.1097/mol.0000000000000400>.
28. Dewey FE, Grove ME, Pan C. Clinical interpretation and implications of whole-genome sequencing. *JAMA*. 2014;311(10):1035–45. <https://doi.org/10.1001/jama.2014.1717>. PMC4119063
29. Ramos EM, Din-Lovinescu C, Berg JS, Brooks LD, Duncanson A, Dunn M, et al. Characterizing genetic variants for clinical action. *Am J Med Genet C: Semin Med Genet*. 2014;166(1):93–104. <https://doi.org/10.1002/ajmg.c.31386>. PMC4158437
30. Shirts BH, Salama JS, Aronson SJ, Chung WK, Gray SW, Hindorff LA, et al. CSER and eMERGE: current and potential state of the display of genetic information in the electronic health record. *J Am Med Inform Assoc*. 2015; 22(6):1231–42. <https://doi.org/10.1093/jamia/ocv065>. PMC5009914

**Ready to submit your research? Choose BMC and benefit from:**

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

**At BMC, research is always in progress.**

Learn more [biomedcentral.com/submissions](https://biomedcentral.com/submissions)

