

<https://doi.org/10.1038/s41525-024-00417-9>

Rare variants at *KCNJ2* are associated with LDL-cholesterol levels in a cross-population study



Niccolò Rossi¹, Najeeb Syed², Alessia Visconti^{1,3}, Elbay Aliyev², Sarah Berry⁴, Mafalda Bourbon⁵, Tim D. Spector¹, Pirro G. Hysi^{1,7}, Khalid A. Fakhro^{2,6,7} & Mario Falchi^{1,7}✉

Leveraging whole genome sequencing data of 1751 individuals from the UK and 2587 Qatari subjects, we suggest here an association of rare variants mapping to the sour taste-associated gene *KCNJ2* with reduced low-density lipoprotein cholesterol (LDL-C, $P = 2.10 \times 10^{-12}$) and with a 22% decreased dietary trans-fat intake. This study identifies a novel candidate rare locus for LDL-C, adding insights into the genetic architecture of a complex trait implicated in cardiovascular disease.

Circulating lipids, and in particular LDL-cholesterol (LDL-C), have been extensively studied due to their role in atherosclerosis pathogenesis. While numerous common genetic loci associated with LDL-C have been described so far¹, the study of the contribution of rare variants to LDL-C variation remains an under-explored area, which has the potential to unveil novel therapeutic targets and improve risk profiles. Rare variants association studies represent a major challenge of genetic research, as they require sufficiently large sample sizes for a well-powered analysis, and they are prone to replication failure across diverse populations as a consequence of potential heterogeneity of populations in the genetic structure and environmental exposures^{2,3}. So far, only a few association studies^{4,5} have been successful in uncovering novel, rare loci associated with LDL-C. Identified associated variants fall in the coding portion of the genome due to the expected strong effect of coding variants on phenotypic variation.

The TwinsUK dataset⁶ consisted of high coverage (30×) WGS data of 1751 UK subjects, including 344 monozygotic and 425 dizygotic twin pairs and 213 singletons (Supplementary Table 1). After quality control, a total of 58,461,543 rare and common single nucleotide variants (SNVs) were identified. We first performed a single-point genome-wide association study (GWAS) for LDL-C on 7,859,673 common ($\text{MAF} \geq 1\%$) SNVs. Results confirmed the absence of population stratification or other confounding biases ($\lambda_{\text{GC}} = 1.02$; Supplementary Fig. 1a). The analysis identified 29 common variants significantly associated with LDL-C at the customary significance threshold of $P < 5 \times 10^{-8}$. These variants mapped to seven unique loci on chromosome 19: *APOC1*, *APOC1P1*, *APOE*, *BCAM*, *LDLR*, *NECTIN2*, and *TOMM40* (Supplementary Fig. 1b, Supplementary Table 3), whose associations with LDL-C were either previously reported or in high linkage disequilibrium (LD; $r^2 \geq 0.8$) with variants reported in the literature⁷.

We then carried out a rare variant region-based GWAS for LDL-C, using a SKAT-O method accommodating sample relatedness⁸, interrogating 1,341,836 sliding windows, each of a length of 4000 base-pair, comprising a total of 50,601,870 rare ($\text{MAF} < 1\%$) SNVs. We replicated previous rare-variants associations for LDL-C at the Mendelian dyslipidaemia genes *LDLR*, *APOB*, *APOE*, and *PCSK9* ($P < 1.27 \times 10^{-3}$), and identified novel candidate associations at 12 windows at a significance threshold for discovery of 6×10^{-9} (Supplementary Tables 4, 5). Eleven overlapping windows, spanning 26 kb of chromosome 17 (chr17:70,493,859–70,519,858, genomic coordinates GRCh38; $P_{\text{min}} = 1.03 \times 10^{-11}$), localised ~300 kb downstream of the *KCNJ2* gene, and one 4-kb window (chr16:67,304,097–67,308,096; $P = 4.16 \times 10^{-9}$) within the *KCTD19* gene (Fig. 1a, Supplementary Table 5). A multi-ancestry meta-analysis of lipid levels in 1.65 million individuals identified genome-wide significant associations between common variants at *KCNJ2* (lead SNV = rs9890133, $P = 6.17 \times 10^{-10}$, $\text{MAF}_{\text{EUR}} = 0.119$) and LDL-C¹⁰. Conditioning on rs9890133 or other common variants located within 500 kb and previously reported in association with lipids, adiposity measures, or heart conditions did not affect the significance of our associations (Supplementary Tables 6, 7).

Replication was conducted in 2587 individuals from the Qatari Biobank (QBB) cohort¹¹ for whom WGS data was available. Four out of 11 4-kb windows within the 26-kb region mapping to *KCNJ2* were replicated in QBB at a Bonferroni-adjusted threshold of 7.14×10^{-3} ($P_{\text{min}} = 2.72 \times 10^{-3}$), with the highest association from meta-analysis at chr17:70,493,859–70,497,858 ($P_{\text{combined}} = 2.10 \times 10^{-12}$; Supplementary Table 5). The region encompassing the *KCTD19* gene was not replicated ($P > 0.05$).

¹Department of Twin Research & Genetic Epidemiology, King's College London, London, UK. ²Department of Human Genetics, Sidra Medical and Research Center, Doha, Qatar. ³Center for Biostatistics, Epidemiology and Public Health, Department of Clinical and Biological Sciences, University of Turin, Turin, Italy. ⁴Department of Nutritional Sciences, King's College London, London, UK. ⁵Cardiovascular Research Group, Department of Health Promotion and Prevention of non-Communicable Diseases, Instituto Nacional de Saúde Dr. Ricardo Jorge, Lisbon, Portugal. ⁶Department of Genetic Medicine, Weill-Cornell Medical College, Doha, Qatar. ⁷These authors jointly supervised this work: Pirro G. Hysi, Khalid A. Fakhro, Mario Falchi. ✉e-mail: mario.falchi@kcl.ac.uk

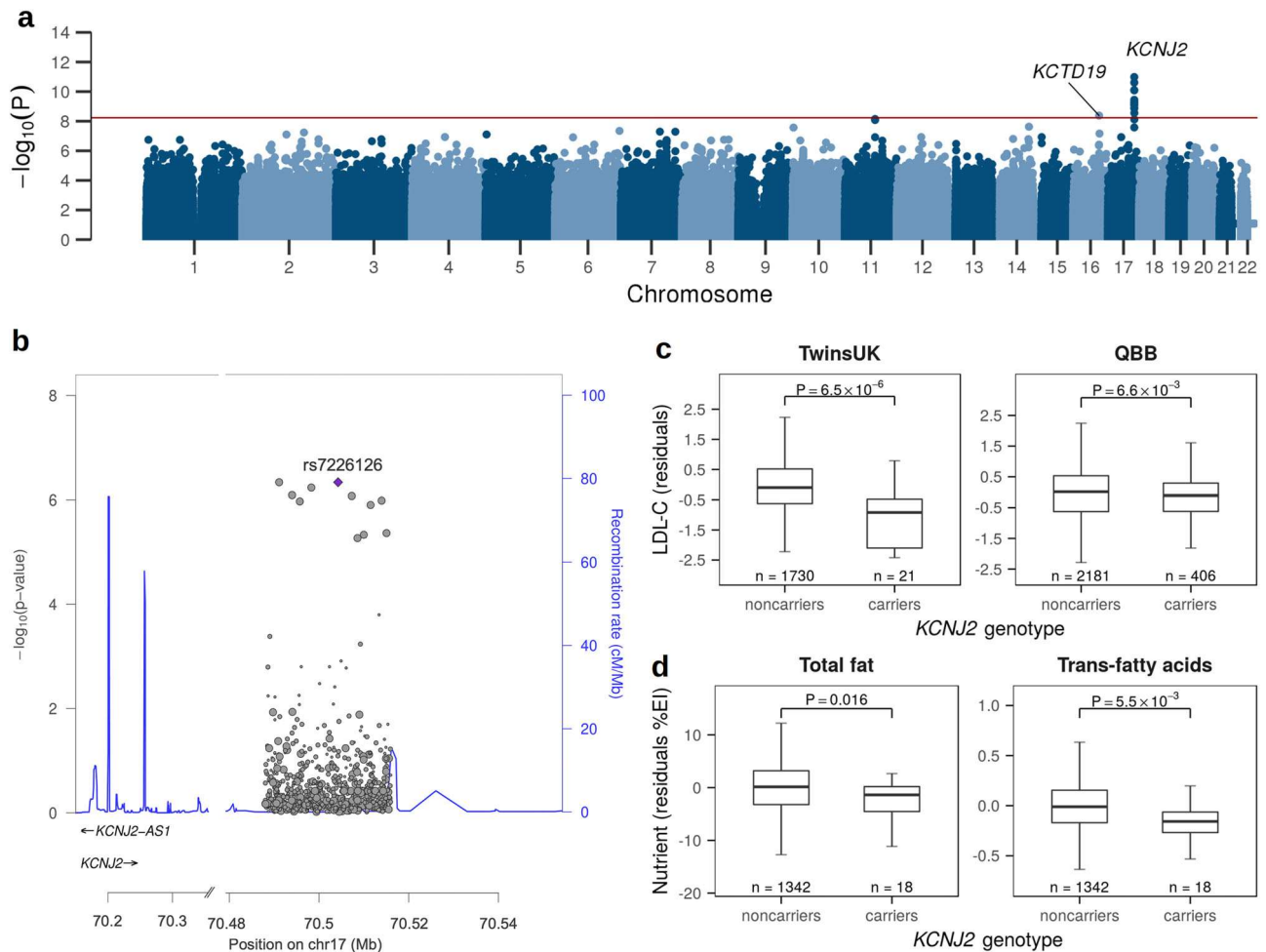


Fig. 1 | Association of rare variants at *KCNJ2* with LDL-C and fat intake.
a Manhattan plot of region-based rare (TwinsUK MAF < 1%) variants association for LDL-C in TwinsUK. Each dot corresponds to the p-value for a 4-kb window. Genome-wide significance threshold ($P < 6 \times 10^{-9}$) is indicated by the red horizontal line. For each significant locus, the closest gene is shown. **b** Regional Manhattan plot of the association with LDL-C on chromosome 17 downstream of *KCNJ2*. Dots represent the combined P-value in TwinsUK and QBB at each SNV, with the top SNV rs7226126 shown in purple. **c** Boxplots of the distribution of LDL-C (adjusted for age, sex, and BMI) in TwinsUK and QBB among carriers of any of the rare lead variants for LDL-C at *KCNJ2* compared to non-carriers. **d** Boxplots of the distribution of total fat and trans-fatty acid intake (expressed as a percentage of daily total energy intake, %EI, and adjusted for age and BMI) in TwinsUK among carriers of any of the rare lead variants for LDL-C at *KCNJ2* compared to non-carriers. The boxplots centre lines indicate the median and the limits the 25th and 75th percentile. The whiskers represent either 1.5 times the interquartile range or the maximum/minimum data point if they are within 1.5 times the interquartile range. Reported P-values are from Wilcoxon's tests.

To dissect the contribution of individual SNVs to the aggregate variants associated with LDL-C, we performed single point association of 92 rare (MAF_{TwinsUK} < 1%) variants included in the chr17:70,493,859–70,519,858 region, and present in both TwinsUK and QBB. Lead SNVs rs7226126 and rs80345346, 13 kb apart and in complete LD in both TwinsUK and QBB, were associated at $P = 1.71 \times 10^{-6}$ ($\beta = -1.19$; SE = 0.25; MAF = 0.5%) in TwinsUK and $P = 9.67 \times 10^{-3}$ in QBB ($\beta = -0.14$; SE = 0.05; combined $P = 4.70 \times 10^{-7}$; Fig. 1b), where they showed higher minor allele frequency (MAF ~ 0.07, Supplementary Table 8). Allele frequencies were comparable between females and males in the general population (Supplementary Table 9). Effect allele frequency breakdown by sex and ancestry is reported in Supplementary Table 9. The near 10-fold lower MAF observed in TwinsUK as compared to QBB at rs7226126 (and rs80345346) corresponded to approximately a ten-fold higher phenotypic effect size for LDL-C among carriers. Among carriers of at least one of these variants, average LDL-C (after accounting for sex, age, and BMI) was 0.97 (SD = 0.22) and 0.12 (SD = 0.05) mmol/L lower than non-carriers in TwinsUK and QBB, respectively (Fig. 1c).

The *KCNJ2* gene encodes for an acid-sensitive potassium channel ($K_{ir}2.1$) involved in the sour taste transduction cascade¹². $K_{ir}2.1$ is inhibited

by intracellular acidification resulting from hydrogen ions from acidic food entering sour-sensing taste cells¹². While *KCNJ2* is relatively ubiquitously expressed and contributes to the resting potassium potential in both sour and non-sour taste cells, the lower expression of *KCNJ2* in sour taste cells makes them more sensitive to intracellular acidification¹², suggesting that *KCNJ2* expression levels are determinant for sour taste transduction and sensitivity. The region chr17:70,493,859–70,519,858 encompasses expression quantitative trait loci (eQTLs) for *KCNJ2* and its antisense gene *KCNJ2-AS1* (Supplementary Fig. 2), and overlaps a putative transcription factor binding site for *KCNJ2* (chr17:70,503,507–70,503,756; OREG1689846¹³), suggesting that rare variants harboured within this region may affect *KCNJ2* expression and, in turn, sour taste sensitivity. This is supported by previous evidence that sour taste sensitivity is strongly (>50% of the variance) heritable¹⁴ and that common genetic variants with putative regulatory effects on *KCNJ2* expression (i.e., intronic or 3'UTR variants) are associated with sour food preference¹⁵. As taste sensitivity can affect food preference and intake¹⁶, with sour food consumption being itself (34–50% of the variance) heritable¹⁷, we hypothesised that the association between rare variants at *KCNJ2* and LDL-C could be explained, at least in part, by *KCNJ2* effect on food preference. We tested the association between aggregated rare variants

at chr17:70,493,859–70,519,858 and nutrient intake from food frequency questionnaires which were available for 1360 female subjects from TwinsUK (Supplementary Table 10). Multiple macronutrients (i.e., total fat, monounsaturated fatty acids, trans fatty acids, total sugars, and starch) were significantly associated with aggregated rare variants at chr17:70,493,859–70,519,858 at a Bonferroni-adjusted threshold of 5.56×10^{-3} (Supplementary Table 11), with total fat intake being the most significant ($P = 7.54 \times 10^{-4}$).

Particularly, among carriers of the lead rare variant rs7226126 (or rs80345346), a smaller percentage of the daily total energy intake was derived from fat ($\beta = -2.82\%$; SE = 1.28%, $P = 2.82 \times 10^{-2}$), and a larger percentage from total sugar and starch ($\beta = 4.91\%$, SE = 1.36%, $P = 2.98 \times 10^{-4}$ and $\beta = -3.56\%$, SE = 1.13%, $P = 1.67 \times 10^{-3}$, respectively; Supplementary Table 11) compared to non-carriers, corresponding to a relative reduction of 10% total fat intake and, more importantly, 22% trans fatty acids intake (Fig. 1d). While this is consistent with studies reporting that dietary reductions in trans fatty acids and their substitution by carbohydrates may have a beneficial effect on serum lipids¹⁸, the interaction between sour taste sensitivity and dietary fat intake remains to be elucidated. Interestingly, while a link between oral fat sensitivity, fat intake and obesity has long been established¹⁹, the existence of a fat-taste perception mechanism independent of other taste qualities is still debated, with a study showing that short-chain fatty acids stimulate a sensation similar to sour²⁰. This evidence suggests that individuals with a preference for sour-tasting foods may also favour a diet rich in fat. Notably, the *KCNJ2* synonymous variant rs173135-T, located ~300 kb upstream chr17:70,493,859–70,519,858 and reported for higher sour preference in 44 Canadian families from the Guelph Family Health Study¹⁵, has been associated with total energy density of diet (i.e., calories per food weight units) in the same cohort¹⁵, with fat having the highest energy density compared to other nutrients. The same allele has been reported for increased serum levels of the ketone body 3-hydroxybutyrate in 2482 Dutch genetic isolates from the Erasmus Rucphen Family (ERF) study²¹, with ketone bodies being produced on high-fat, low-carbohydrate (i.e., ketogenic) diets. Also, a genetic variation within chr17:70,493,859–70,519,858 has been associated with adiposity measures²². Interestingly, an association between increased BMI and lower sour taste sensitivity has been observed²³.

In summary, here we report the association results of genome sequencing data of a total of 4338 individuals from two populations. We suggest that a burden of rare variants mapping to the sour taste-associated gene *KCNJ2* contributes to lowered LDL-C levels despite population heterogeneity in terms of allele frequency and effect size. We propose that this finding may be due to the *KCNJ2* effect on food preference, as rare variant carriers from the UK had reduced dietary fat intake compared to non-carriers. The present study, while warranting further investigation, suggests *KCNJ2* as a novel candidate rare-variant locus for LDL-C.

Methods

Discovery cohort

Study subjects were 1751 twins, mainly females (97%; Supplementary Table 1), belonging to the TwinsUK cohort, a UK national register of volunteer adult twins unselected for any diseases or traits⁶. St. Thomas' Hospital Research Ethics Committee approved the study, and all twins provided informed written consent.

LDL-cholesterol data

Subjects from TwinsUK had data for longitudinal plasma biochemical indices of HDL-cholesterol (HDL-C), triglycerides (TG) and total cholesterol (TC). Fasting plasma levels were measured using an analysing device (Cobas Fara; Roche Diagnostics, Lewes, UK). TC, HDL-C, and TG were determined by a colorimetric enzymatic method²⁴. A median of four longitudinal measures was available per subject (the maximum number of measures was 8). The median duration of follow-up was 5 years (interquartile range: 2.87–7.51). Longitudinal data on medication usage indicated

that 292 subjects were using cholesterol-lowering drugs at the time of lipid measurement, and their effect was taken into account by dividing the measured total cholesterol by 0.8, as previously suggested²⁵. LDL-C values were estimated based on adjusted TC, HDL-C, and TG levels, according to the Friedewald equation²⁶. Longitudinal LDL-C data were averaged, inverse-normal transformed, and adjusted for sex, age, and BMI by linear regression analysis (using the *lm* function as implemented in the *stats* R package, v. 3.4.2). No outliers (individuals with any trait value further away than four standard deviations from the dataset mean) were detected.

Whole-genome sequencing and quality control

Whole genome sequencing was carried out at Human Longevity, Inc. (HLI). Details on sample preparation, library preparation, clustering and sequencing have been reported elsewhere²⁷. Reads were mapped to the UCSC GRCh37 human reference genome, and variants were called using Illumina Isaac Analysis Software²⁸ (v. 2.5.26.13). SNVs calls that did not pass all Isaac's filters (see Isaac Whole Genome Sequencing user guide for details: https://support.illumina.com/content/dam/illumina-support/documents/documentation/software_documentation/basespace/isaac-wgs-user-guide-15050954b.pdf) were discarded as low-quality calls. Multiallelic sites were removed, and missing genotypes were assumed to be reference homozygous calls. We further removed SNV with calls rate <95% and/or Hardy-Weinberg equilibrium (HWE) deviation $P < 1 \times 10^{-9}$, as calculated with PLINK²⁹ (v. 1.9). We finally removed variants mapping to the sex chromosomes, resulting in 58,461,543 biallelic SNVs.

To investigate population structure, we used PLINK to carry out a principal component analysis (PCA) on the genomic data. Here, we used all the common high-quality variants (i.e., missingness across variants = 0 and MAF > 5%) pruned with PLINK (option: --indep-pairwise) using the following parameters: window size of 50-kb, shift of 5 variants, and r^2 threshold of 0.05. No significant population stratification was observed (Supplementary Fig. 3).

Single-point association testing

SNVs with MAF $\geq 1\%$ were tested for association with sex, age, and BMI-adjusted plasma LDL-C levels using a linear mixed model as implemented in GEMMA³⁰ (v. 0.97), which uses the matrix of expected genetic sharing to model the non-independence of the twin data. Manhattan and Q-Q plots were generated using the *ggplot2* R package. Significant SNVs at conventional genome-wide significance threshold ($P < 5 \times 10^{-8}$) were annotated with Ensembl VEP web interface (v. 94) using the NCBI Reference Sequence Database (release 2015-01).

Identification of known LDL-C loci

We interrogated the NHGRI-EBI GWAS catalog⁷ (v. 1.0, release: 2019-01-11; association $P < 5 \times 10^{-8}$) to identify overlap between SNVs identified in our study (or in high linkage disequilibrium with them; $r^2 \geq 0.8$) and previously reported associations for LDL-C levels. Linkage disequilibrium statistics were calculated with PLINK.

Region-based associations testing

Region-based testing was carried out using MONSTER⁸ (v. 1.3), which uses a generalised version of the SKAT-O method for non-independent samples to perform robust region-based rare-variant association testing. Briefly, this programme uses a mixed effects model that accounts for covariates and additive polygenic effects, accounting for relatedness among individuals and adaptively estimating the correlation structure of variant effects to maximise the statistical power. Coefficients of the relationship were assumed to be 1 for monozygotic twins and 0.5 for dizygotic twins. We assessed associations by collapsing rare (MAF < 1%) variants within fixed-size sliding windows. Specifically, 1,341,836 sliding windows of 4 kb (overlapping by 2 kb) were generated beginning at position 1 bp for each chromosome, as described previously^{31–33}. The median number of SNVs per window was 48 (interquartile range: 39–65). We opted for a stringent genome-wide significance threshold of 6×10^{-9} , following suggestions from previous simulation

studies for WGS-based analytic strategies combining individual common variants testing and aggregated rare variants tests using the sliding window approach³⁴).

Region-based conditional analysis

We used secondary conditional analyses to assess the independence of signals of association from previously known GWAS signals arising from common variation near the chr17:70,493,859–70,519,858 and chr16:67,304,097–67,308,096 regions. Specifically, we retrieved from the NHGRI-EBI GWAS catalog⁷ (v. 1.0, release: 2019-01-11) any reported SNV associated with blood lipids, adiposity or cardiovascular risk within 500 kb on either side of the studied regions. Then, we fitted a new region-based regression model for LDL-C as implemented in MONSTER, including the reported SNVs as a covariate.

Replication cohort

The Qatar Genome Programme (QGP) is a national population-based initiative aiming at combining a comprehensive analysis of Qatari genomes with the extensive phenotypic data collected at the QBB¹¹. All subjects are enrolled in the Qatar Biobank via informed consent. The Qatar Biobank study was approved by the Institutional Review Board from the Hamad Medical Corporation Ethics Committee (IRB protocol E/2017/RES-ACC-0032/0002). Whole genome sequencing was performed on 2935 blood samples using Illumina X10 Sequencing machine, with a 150-base paired-end single-index-read format with an average coverage of 30×. Raw WGS data were converted from the native BCL format to paired-end FASTQ format using bcl2fastq (v. 2.16). The quality of the raw data was assessed using fastqc (<http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>). Sequence reads were aligned to the human reference genome version GRCh37 using bwakit (v. 7.12) (<https://github.com/lh3/bwa/tree/master/bwakit>). Variant calling was performed using GATK haplotype caller³⁵ (v. 3.3). SNVs with call rate <95% or Hardy–Weinberg equilibrium (HWE) deviation $P < 1 \times 10^{-9}$ were discarded from the analysis. We leveraged genotypes at 48 ancestry-informative SNVs to differentiate the three major Qatari subpopulations (Bedouin Arabs, Persians, and East Africans) in the QBB cohort, as previously described³⁶. We excluded from the analysis individuals of African ancestry due to the small sample size ($n = 63$), as well as 78 individuals, mainly ($n = 75$, 96%) admixed individuals, separated from the main cohort core based on PC2 value ($PC2 < -0.03$). Coefficients of the relationship between subjects were modelled using the kinship matrix between all individuals, as evaluated by PLINK (v. 1.9), to account for both familial and more distant relatedness.

LDL-C was measured from blood samples at the Hamad Medical Centre Laboratory, Doha, inverse-normal transformed and adjusted for sex, age and BMI. Outliers (measurements further away than four standard deviations from the dataset mean) were removed from subsequent analyses, which were ultimately conducted on 2587 individuals. The phenotypic details of the study subjects are summarised in Supplementary Table 2.

Food frequency data was not available for the QBB cohort.

Meta-analysis

Meta-analysis of window-based association results was carried out by combining P -values across TwinsUK and QBB using Fisher's combined probability test, as implemented in *metap* R package (v. 1.1, *sumlog* function). A Bonferroni-adjusted threshold of 7.14×10^{-3} was obtained by dividing a conventional α -threshold of 0.05 by the number of non-overlapping 4-kb windows tested for replication (i.e., six windows mapping to *KCNJ2* and one window to *KCTD19*).

Individual SNVs values were meta-analysed using a sample-size weighted Z -score-based analysis, as implemented in METAL³⁷ (release 2011-03-25), and their sex- and ancestry-specific allele frequencies were downloaded from the gnomAD browser³⁸ (v. 3.1.2). The regional association plot was generated using LocusZoom³⁹. Reported eQTLs for *KCNJ2* and *KCNJ2-AS1* (Ensembl gene identifiers ENSG00000123700.4 and

ENSG00000267365.1, respectively) were retrieved from GTEx⁴⁰ (v. 8) and eQTLGen⁴¹ (phase I) portals.

Food frequency data

Overall, 1360 female subjects from the TwinsUK cohort with WGS data available completed a 131-item validated food frequency questionnaire⁴² (FFQ) between 1993 and 2015 and had BMI data available within ± 1 year of completing the FFQ. These included 345 dizygotic pairs, 218 monozygotic pairs, and 234 singletons (Supplementary Table 10). Details on quality control, subject exclusion criteria and methods for nutrient determination from FFQ data can be found elsewhere⁴³. Briefly, study participants reported food intake frequencies for the past year of average serving sizes for 131 foods and beverages on a 9-point scale (ranging from never or less than once per month to 6+ times per day). The 131 food items were aggregated into food groups, defined by similarity in nutrient content and culinary use, resulting in 54 different nutrients. We considered 12 macro-nutrients for association testing, namely: the four major classes (alcohol, protein, total fat, and total carbohydrate) and eight subclasses of macronutrients (starch, total sugar, fibre, saturated fatty acids, monounsaturated fatty acids, polyunsaturated fatty acids, trans fatty acids and cholesterol), based on their relevance to LDL-C metabolism⁴⁴. Two or more longitudinal measures were available for 53% of the study subjects (the maximum number of measures was 4). The median duration of follow-up was 9 years (interquartile range: 3.57–10.22 years). When longitudinal data were available, nutrients calculated at each timepoint were averaged. Outliers (values further away than four standard deviations from the nutrient mean) were removed.

Association between *KCNJ2* and nutrients

We tested the association of aggregated rare variants at *KCNJ2* with nutrient intake, adjusted for daily total energy intake, age, and BMI, using MONSTER. Coefficients of relationship were assumed to be 1 for monozygotic twins, and 0.5 for dizygotic twins. To correct for multiple comparisons, we used Li's method⁴⁵ to estimate the effective number of independent tests in the nutrient dataset. The derived P -value threshold for statistical significance was $0.05/9 = 5.56 \times 10^{-3}$.

Subsequently, we used linear mixed model regression analysis (*lmer* function from *lme4* R package, v. 1.1-21) to test for association between lead rare variants at *KCNJ2* and selected nutrients, expressed as percentages of daily total energy intake, modelling family structure as random effect, and including age and BMI as covariates. Average nutrient intakes in the TwinsUK cohort (Supplementary Table 10) fell within the interquartile range of the UK adult population⁴⁶.

Ethical approval

The TwinsUK study was approved by the National Research Ethics Service London-Westminster, the St. Thomas' Hospital Research Ethics Committee (EC04/015 and 07/H0802/84). The Qatar Biobank study was approved by the Institutional Review Board from the Hamad Medical Corporation Ethics Committee (IRB protocol E/2017/RES-ACC-0032/0002). Written informed consent was obtained from all participants. All research therefore carried out in accordance with the ethical standards laid down in the 1964 Declaration of Helsinki and its later amendments.

Data availability

Data on TwinsUK twin participants are available to bona fide researchers under managed access due to governance and ethical constraints. Raw data should be requested via our website (<http://twinsuk.ac.uk/resources-for-researchers/access-our-data/>) and requests are reviewed by the TwinsUK Resource Executive Committee (TREC) regularly. QBB data should be requested by filling out the access application at www.qatarbiobank.org.qa, to be submitted to the research access office accessofficeqbb@qf.org.qa. Requests are reviewed and approved by QBB IRB and Access Committee.

Code availability

Analyses were carried out using open-source published pieces of software which are referenced in the manuscript.

Received: 17 July 2023; Accepted: 3 May 2024;

Published online: 28 June 2024

References

1. Willer, C. J. et al. Discovery and refinement of loci associated with lipid levels. *Nat. Genet.* **45**, 1274–1283 (2013).
2. Bomba, L., Walter, K. & Soranzo, N. The impact of rare and low-frequency genetic variants in common disease. *Genome Biol.* **18**, 77 (2017).
3. Lu, X. et al. Exome chip meta-analysis identifies novel loci and East Asian-specific coding variants that contribute to lipid levels and coronary artery disease. *Nat. Genet.* **49**, 1722–1730 (2017).
4. Lange, L. A. et al. Whole-exome sequencing identifies rare and low-frequency coding variants associated with LDL cholesterol. *Am. J. Hum. Genet.* **94**, 233–245 (2014).
5. van Leeuwen, E. M. et al. Genome of The Netherlands population-specific imputations identify an ABCA6 variant associated with cholesterol levels. *Nat. Commun.* **6**, 6065 (2015).
6. Verdi, S. et al. TwinsUK: the UK Adult Twin Registry Update. *Twin Res. Hum. Genet.* **22**, 523–529 (2019).
7. Welter, D. et al. The NHGRI GWAS Catalog, a curated resource of SNP-trait associations. *Nucleic Acids Res.* **42**, D1001–D1006 (2014).
8. Jiang, D. & McPeck, M. S. Robust rare variant association testing for quantitative traits in samples with related individuals. *Genet. Epidemiol.* **38**, 10–20 (2014).
9. Natarajan, P. et al. Deep-coverage whole genome sequences and blood lipids among 16,324 individuals. *Nat. Commun.* **9**, 3391 (2018).
10. Graham, S. E. et al. The power of genetic diversity in genome-wide association studies of lipids. *Nature* **600**, 675–679 (2021).
11. Al Kuwari, H. et al. The Qatar Biobank: background and methods. *BMC Public Health* **15**, 1208 (2015).
12. Ye, W. et al. The K⁺ channel KIR2.1 functions in tandem with proton influx to mediate sour taste transduction. *Proc. Natl Acad. Sci. USA* **113**, E229–E238 (2016).
13. Lesurf, R. et al. ORegAnno 3.0: a community-driven resource for curated regulatory annotation. *Nucleic Acids Res.* **44**, D126–D132 (2016).
14. Wise, P. M., Hansen, J. L., Reed, D. R. & Breslin, P. A. S. Twin study of the heritability of recognition thresholds for sour and salty taste. *Chem. Senses* **32**, 749–754 (2007).
15. Chamoun, E. et al. The relationship between single nucleotide polymorphisms in taste receptor genes, taste function and dietary intake in preschool-aged children and adults in the Guelph Family Health Study. *Nutrients* **10**, 990 (2018).
16. McCrickerd, K. & Forde, C. G. Sensory influences on food intake control: moving beyond palatability. *Obes. Rev.* **17**, 18–29 (2016).
17. Törnwall, O. et al. Genetic contribution to sour taste preference. *Appetite* **58**, 687–694 (2012).
18. Mensink, R. P., Zock, P. L., Kester, A. D. M. & Katan, M. B. Effects of dietary fatty acids and carbohydrates on the ratio of serum total to HDL cholesterol and on serum lipids and apolipoproteins: a meta-analysis of 60 controlled trials. *Am. J. Clin. Nutr.* **77**, 1146–1155 (2003).
19. Stewart, J. E. et al. Oral sensitivity to fatty acids, food consumption and BMI in human subjects. *Br. J. Nutr.* **104**, 145–152 (2010).
20. Running, C. A., Craig, B. A. & Mattes, R. D. Oleogustus: the unique taste of fat. *Chem. Senses* **40**, 507–516 (2015).
21. Demirkan, A. et al. Insight in genome-wide association of metabolite quantitative traits by exome sequence analyses. *PLoS Genet.* **11**, e1004835 (2015).
22. Graff, M. et al. Genome-wide physical activity interactions in adiposity —a meta-analysis of 200,452 adults. *PLoS Genet.* **13**, e1006528 (2017).
23. Bertoli, S. et al. Taste sensitivity, nutritional status and metabolic syndrome: Implication in weight loss dietary interventions. *World J. Diabetes* **5**, 717–723 (2014).
24. Middelberg, R. P. S., Spector, T. D., Swaminathan, R. & Snieder, H. Genetic and environmental influences on lipids, lipoproteins, and apolipoproteins: effects of menopause. *Arterioscler. Thromb. Vasc. Biol.* **22**, 1142–1147 (2002).
25. Cholesterol Treatment Trialists' (CTT) Collaboration. et al. Efficacy and safety of LDL-lowering therapy among men and women: meta-analysis of individual data from 174,000 participants in 27 randomised trials. *Lancet* **385**, 1397–1405 (2015).
26. Friedewald, W. T., Levy, R. I. & Fredrickson, D. S. Estimation of the concentration of low-density lipoprotein cholesterol in plasma, without use of the preparative ultracentrifuge. *Clin. Chem.* **18**, 499–502 (1972).
27. Long, T. et al. Whole-genome sequencing identifies common-to-rare variants associated with human blood metabolites. *Nat. Genet.* **49**, 568–578 (2017).
28. Racz, C. et al. Isaac: ultra-fast whole-genome secondary analysis on Illumina sequencing platforms. *Bioinformatics* **29**, 2041–2043 (2013).
29. Purcell, S. et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* **81**, 559–575 (2007).
30. Zhou, X. & Stephens, M. Genome-wide efficient mixed-model analysis for association studies. *Nat. Genet.* **44**, 821–824 (2012).
31. Morrison, A. C. et al. Whole-genome sequence-based analysis of high-density lipoprotein cholesterol. *Nat. Genet.* **45**, 899–901 (2013).
32. Yu, B. et al. Whole genome sequence analysis of serum amino acid levels. *Genome Biol.* **17**, 237 (2016).
33. Kim, T. & Wei, P. Incorporating ENCODE information into association analysis of whole genome sequencing data. *BMC Proc.* **10**, 257–261 (2016).
34. Xu, C. et al. Estimating genome-wide significance for whole-genome sequencing studies. *Genet. Epidemiol.* **38**, 281–290 (2014).
35. DePristo, M. A. et al. A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat. Genet.* **43**, 491–498 (2011).
36. Rossi, N. et al. Ethnic-specific association of amylase gene copy number with adiposity traits in a large Middle Eastern biobank. *NPJ Genom. Med* **6**, 8 (2021).
37. Willer, C. J., Li, Y. & Abecasis, G. R. METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics* **26**, 2190–2191 (2010).
38. Chen, S. et al. A genomic mutational constraint map using variation in 76,156 human genomes. *Nature* **625**, 92–100 (2024).
39. Pruim, R. J. et al. LocusZoom: regional visualization of genome-wide association scan results. *Bioinformatics* **26**, 2336–2337 (2010).
40. GTEx Consortium The Genotype-Tissue Expression (GTEx) project. *Nat. Genet.* **45**, 580–585 (2013).
41. Vösa, U. et al. Large-scale cis- and trans-eQTL analyses identify thousands of genetic loci and polygenic scores that regulate blood gene expression. *Nat. Genet.* **53**, 1300–1310 (2021).
42. Bingham, S. A. et al. Nutritional methods in the European Prospective Investigation of Cancer in Norfolk. *Public Health Nutr.* **4**, 847–858 (2001).
43. Teucher, B. et al. Dietary patterns and heritability of food choice in a UK female twin cohort. *Twin Res. Hum. Genet.* **10**, 734–748 (2007).
44. Lichtenstein, A. H. Thematic review series: patient-oriented research. Dietary fat, carbohydrate, and protein: effects on plasma lipoprotein patterns. *J. Lipid Res.* **47**, 1661–1667 (2006).
45. Li, J. & Ji, L. Adjusting multiple testing in multilocus analyses using the eigenvalues of a correlation matrix. *Heredity* **95**, 221–227 (2005).

46. Whitton, C. et al. National Diet and Nutrition Survey: UK food consumption and nutrient intakes from the first year of the rolling programme and comparisons with previous surveys. *Br. J. Nutr.* **106**, 1899–1914 (2011).

Acknowledgements

TwinsUK is funded by the Wellcome Trust, Medical Research Council, European Union, Chronic Disease Research Foundation (CDRF), the National Institute for Health Research (NIHR)-funded BioResource, Clinical Research Facility and Biomedical Research Centre based at Guy's and St. Thomas' NHS Foundation Trust in partnership with King's College London. QBB study was funded in part by the Qatar National Research Fund's Path towards Precision Medicine award (PPM1-1229-15002). The authors acknowledge the use of the research computing facility at King's College London, Rosalind (<https://rosalind.kcl.ac.uk>). PGH acknowledges support from a BrightFocus Foundation award (G2021011S).

Author contributions

N.R., A.V. and M.F. designed the study. N.R. conducted the computational experimentation in TwinsUK. N.S., E.A. and K.F. conducted the replication study in QBB. N.R., A.V., P.H., K.F. and M.F. drafted the manuscript. N.R., A.V., S.B., M.B., T.D.S. P.H. and M.F. contributed to the interpretation. All authors have read and approved the final version of the manuscript.

Competing interests

T.D.S. is a consultant for Zoe Global Ltd. The other authors declare that they have no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41525-024-00417-9>.

Correspondence and requests for materials should be addressed to Mario Falchi.

Reprints and permissions information is available at <http://www.nature.com/reprints>

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

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2024