RESEARCH NOTE

Genome-wide association study of vegetarianism in UK Biobank identifies association with *VRK2* [version 1; peer review: 3 approved with reservations]

Georgina K. Fensom^{1,2}, Karl Smith-Byrne³, Colm D. Andrews¹, Tim J. Key¹, Ruth C. Travis¹

¹Nuffield Department of Populations Health, University of Oxford, Oxford, OX3 7LF, UK

²Oxford Department of International Development, University of Oxford, Oxford, OX1 3TB, UK

³Genetic Epidemiology Group, International Agency for Research on Cancer, Lyon, 69008, France

V1 First published: 18 Dec 2020, 5:291
<https://doi.org/10.12688/wellcomeopenres.16396.1>
 Latest published: 18 Dec 2020, 5:291
<https://doi.org/10.12688/wellcomeopenres.16396.1>

Abstract

Prospective studies have observed differences in risks for several health outcomes when comparing meat-eaters and vegetarians, but the mechanisms underlying these differences remain uncertain. Identifying genetic factors related to vegetarianism may be valuable for assessing causality. We report a genome-wide association study (GWAS) of vegetarianism in 367,198 participants from UK Biobank. We identified one locus, rs10189138, near the vaccinia related kinase 2 (*VRK2*) gene, significantly associated with vegetarianism ($\beta=0.153$, $p=3 \times 10^{-8}$). The associations between rs10189138 and 40 traits were calculated, and the rs10189138 T allele (MAF=0.12) was found to be significantly associated with greater height, after controlling the false discovery rate (FDR). Correlations between genetically predicted vegetarianism and 855 other genetically predicted traits were also calculated, and vegetarianism had significant positive genetic correlations with fluid intelligence and age at menarche, after controlling the FDR. Future research on an independent sample is needed to see if this GWAS result can be replicated.

Keywords

Vegetarian, Diet, GWAS, *VRK2*

Open Peer Review

Approval Status ? ? ?

	1	2	3
version 1	?	?	?
18 Dec 2020	view	view	view

1. **Joanne Burnette Cole** , Massachusetts General Hospital, Boston, USA
The Broad Institute of MIT and Harvard, Cambridge, USA
Boston Children's Hospital, Boston, USA
2. **Robert Karlsson** , Karolinska Institutet, Stockholm, Sweden
3. **Natalia Pervjakova**, Institute of Genomics, University of Tartu, Tartu, Estonia

Any reports and responses or comments on the article can be found at the end of the article.

Corresponding author: Georgina K. Fensom (georgina.fensom@qeh.ox.ac.uk)

Author roles: **Fensom GK:** Conceptualization, Data Curation, Formal Analysis, Methodology, Software, Visualization, Writing – Original Draft Preparation, Writing – Review & Editing; **Smith-Byrne K:** Formal Analysis, Methodology, Software, Writing – Review & Editing; **Andrews CD:** Formal Analysis, Software, Writing – Review & Editing; **Key TJ:** Conceptualization, Funding Acquisition, Supervision, Writing – Review & Editing; **Travis RC:** Conceptualization, Funding Acquisition, Methodology, Project Administration, Supervision, Writing – Review & Editing

Competing interests: The authors declare they have no conflicts of interest. Where authors are identified as personnel of the International Agency for Research on Cancer / World Health Organization, the authors alone are responsible for the views expressed in this article and they do not necessarily represent the decisions, policy or views of the International Agency for Research on Cancer / World Health Organization.

Grant information: This work was supported by Wellcome (205212), Cancer Research UK (C8221/A29017) and the Medical Research Council (MR/M012190/1).

The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Copyright: © 2020 Fensom GK *et al.* This is an open access article distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

How to cite this article: Fensom GK, Smith-Byrne K, Andrews CD *et al.* **Genome-wide association study of vegetarianism in UK Biobank identifies association with *VRK2* [version 1; peer review: 3 approved with reservations]** Wellcome Open Research 2020, 5:291 <https://doi.org/10.12688/wellcomeopenres.16396.1>

First published: 18 Dec 2020, 5:291 <https://doi.org/10.12688/wellcomeopenres.16396.1>

Introduction

Vegetarian diets, defined by abstention from consuming meat, meat products, poultry and fish, have been gaining popularity in Western countries. Previous prospective cohort studies have reported lower risks for vegetarians compared to meat-eaters for several conditions, including obesity¹ and ischemic heart disease², but higher risks have been reported for haemorrhagic stroke² and some fractures³.

Understanding factors contributing to a person's decision to follow a vegetarian diet could be important for understanding the mechanisms that underlie differences in disease risks. Further, identifying genetic factors related to this behavioural trait may aid in assessing causality through Mendelian randomization methods. The genetic contribution to vegetarianism is largely unknown due to the lack, until recently, of large studies of diet and genetic variants. However, previous studies have reported moderate to strong heritability for dietary patterns, food preferences, and appetite⁴. We aimed to identify genetic loci associated with vegetarianism through a genome-wide association study (GWAS) using the UK Biobank cohort.

Methods

UK Biobank data

The UK Biobank is a prospective cohort of approximately 500,000 people aged 40-69 who were recruited from 2006-2010 across the United Kingdom. A [description of the study protocol and design](#) can be found online. Briefly, participants were identified from National Health Service registers and lived within 25 miles of one of the 22 assessment centres. At recruitment participants completed a touchscreen questionnaire, had physical measurements and blood samples taken, and provided informed consent. [Details of the blood sample assay methods and quality procedures](#) can be found online. The study was approved by the North West Multi-Centre Research Ethics Committee (reference number 06/MRE08/65).

We used information from the touchscreen 'food frequency' questionnaire (FFQ) to classify participants as either vegetarian or non-vegetarian. The FFQ included 29 questions on the frequency of consumption of foods, with answers ranging from 'never' to 'once or more daily', and has been shown to be highly reliable for meat and fish intake⁵. Participants were defined as vegetarian if they answered 'never' to all questions related to the consumption of meat (including processed meat, beef, lamb or mutton, pork, chicken, turkey or other poultry) and fish (including oily fish and other fish). Participants were excluded if they did not answer the relevant questions ($n=11,557$).

From April 2009, participants who were recruited were additionally invited to complete the Oxford WebQ, a 24-hour dietary assessment tool. If participants provided an email address then they were invited to complete the WebQ up to four additional times between February 2011 and June 2012. We excluded WebQ records where energy intake was below or above pre-specified sex-specific limits ($<3,349$ kJ or $>16,747$ kJ for men and $<2,093$ kJ or $>14,654$ kJ for women; $n=9,157$), or where a record was non-representative due to illness or fasting

($n=2,971$), and WebQ data were only used if a participant completed three or more questionnaires ($n=57,915$). Estimated mean weekly intakes of total, red, and processed meat, and poultry were calculated by combining responses from both the FFQ and WebQ.

Processing and analysis of genotype data

Genome-wide genotype data are available for the whole cohort. We excluded participants based on non-white British ancestry ($n=78,422$), mismatch between genetic versus self-reported sex ($n=308$), sex chromosome aneuploidy ($n=407$), and relatedness ($n=32,827$). We also excluded single-nucleotide polymorphisms (SNPs) with $MAF < 0.001\%$ ($n=76,570,322$), those not in Hardy-Weinberg equilibrium ($n=63,308$), and those with imputation $R^2 \leq 0.9$ if $0.001 < MAF \leq 0.005$ ($n=4,080,151$), $R^2 \leq 0.8$ if $0.005 < MAF \leq 0.01$ ($n=221,076$), $R^2 \leq 0.6$ if $0.01 < MAF \leq 0.03$ ($n=42,101$) and $R^2 \leq 0.3$ if $MAF > 0.03$ ($n=5,855$).

The additive association between the binary vegetarian phenotype and each SNP was tested using a logistic regression model, adjusted for the first twenty principal components, genotype array, sex, age, and age-squared.

Linear regression models, with the same covariates as listed above, were used to estimate the per-allele difference for selected baseline characteristics, blood biomarker concentrations, and meat intakes (in non-vegetarians). These estimates were used to calculate the percent change per-allele in comparison to the mean value in participants with GG genotype. A false discovery rate (FDR) controlling procedure (Benjamini-Hochberg) was used to account for multiple testing.

We used linkage disequilibrium (LD) score regression to estimate genetic correlation between genetically predicted vegetarianism and publicly available GWAS summary-level results for 855 heritable traits on LDHub⁶.

Results and discussion

A total of 367,198 participants, including 5,642 vegetarians, and 12,112,810 SNPs remained for our GWAS. We identified one locus, rs10189138 near the vaccinia related kinase 2 (VRK2) gene on chromosome 2, which was significantly associated with vegetarianism ($p=3 \times 10^{-8}$, $MAF=0.12$). Compared to the GG genotype, each additional T allele was associated with a 16.5% (95% confidence interval (CI): 10.4%, 23.0%; $\beta=0.153$) increased likelihood of following a vegetarian diet. The heritability of vegetarianism was low ($H^2=0.012$; 95% CI: 0.009-0.015).

VRK2 is a serine/threonine kinase that is important in several signal transduction cascades⁷. Previous studies have reported an association at this locus with differences in the overall intake of beef, lamb/mutton, pork, and processed meat⁸, higher risk of schizophrenia, major depressive disorder and genetic generalized epilepsy⁹ and longer sleep duration⁷.

Using linear regression models, we found that the rs10189138 T allele was significantly associated with increased height (0.10 cm; 95%CI: 0.05 to 0.14) (Figure 1). No other significant associations were observed.

Figure 1. Association between variant rs10189138 and selected baseline characteristics, biomarker concentrations and meat intakes (in non-vegetarians) in UK Biobank. ¹ Denotes associations that were statistically significant after applying a False Discovery Rate (FDR) controlling procedure to account for multiple test. ² Analysis only includes participants who are non-vegetarian.

Using LD score regression to estimate genetic correlation, we found that fluid intelligence ($r_g=0.12$; 95%CI 0.06 to 0.18) and age at menarche ($r_g=0.09$; 95%CI 0.04 to 0.14) were significantly correlated with vegetarianism (Figure 2) after controlling the FDR. These results are in keeping with previous epidemiological evidence that has shown a

significant association between vegetarianism and higher IQ in childhood and adulthood, as well as higher attainment of academic qualifications¹⁰. Additionally, a small prospective study of American schoolgirls previously reported a 6 month later age at menarche among vegetarians compared to non-vegetarians¹¹.

Figure 2. Genetic correlations of vegetarianism with selected heritable traits¹. ¹ Results are shown for 21 out of 855 heritable traits tested, where $p < 0.01$. ² Denotes correlations that were statistically significant after applying a False Discovery Rate (FDR) controlling procedure to account for multiple testing.

Conclusion

In summary, we report one novel SNP, rs10189138, near *VRK2*, that is significantly associated with a 16.5% per-allele increased likelihood of following a vegetarian diet. Future research is needed on an independent sample to see if this result can be replicated.

Data availability

We use data from the UK Biobank resource under Application Number 3248 for this work. All bona fide researchers can apply

to use the UK Biobank resource for health-related research that is in the public interest. Further information on the application process is available from the UK Biobank website (<https://www.ukbiobank.ac.uk/register-apply/>).

Acknowledgements

This work has been conducted using the UK Biobank Resource under Application Number 3248 and we express our gratitude to the participants and those involved in building the resource.

References

1. Davey GK, Spencer EA, Appleby PN, *et al.*: **EPIC-Oxford: lifestyle characteristics and nutrient intakes in a cohort of 33 883 meat-eaters and 31 546 non meat-eaters in the UK.** *Public Health Nutr.* 2003; **6**(3): 259–69.
[PubMed Abstract](#) | [Publisher Full Text](#)
2. Tong TYN, Appleby PN, Bradbury KE, *et al.*: **Risks of ischaemic heart disease and stroke in meat eaters, fish eaters, and vegetarians over 18 years of follow-up: results from the prospective EPIC-Oxford study.** *BMJ.* 2019; **366**: l4897.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
3. Thorpe DL, Knutsen SF, Beeson WL, *et al.*: **Effects of meat consumption and vegetarian diet on risk of wrist fracture over 25 years in a cohort of peri- and postmenopausal women.** *Public Health Nutr.* 2008; **11**(6): 564–72.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
4. van den Berg L, Henneman P, van Dijk KW, *et al.*: **Heritability of dietary food intake patterns.** *Acta Diabetol.* 2013; **50**(5): 721–6.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
5. Bradbury KE, Young HJ, Guo W, *et al.*: **Dietary assessment in UK Biobank: an evaluation of the performance of the touchscreen dietary questionnaire.** *J Nutr Sci.* 2018; **7**: e6.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
6. Zheng J, Erzurumluoglu AM, Elsworth BL, *et al.*: **LD Hub: a centralized database and web interface to perform LD score regression that maximizes the potential of summary level GWAS data for SNP heritability and genetic correlation analysis.** *Bioinformatics.* 2017; **33**(2): 272–79.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
7. Jones SE, Tyrrell J, Wood AR, *et al.*: **Genome-Wide Association Analyses in 128,266 Individuals Identifies New Morningness and Sleep Duration Loci.** *PLoS Genet.* 2016; **12**(8): e1006125.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
8. Cole JB, Florez JC, Hirschhorn, JN: **Comprehensive genomic analysis of dietary habits in UK Biobank identifies hundreds of genetic associations.** *Nat Commun.* 2020; **11**(1): 1467.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
9. Li M, Yue W: **VRK2, a Candidate Gene for Psychiatric and Neurological Disorders.** *Mol Neuropsychiatry.* 2018; **4**(3): 119–33.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
10. Gale CR, Deary IJ, Schoon I, *et al.*: **IQ in childhood and vegetarianism in adulthood: 1970 British cohort study.** *BMJ.* 2007; **334**(7587): 245.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
11. Sanchez A, Kissinger DG, Phillips RI: **A hypothesis on the etiological role of diet on age of menarche.** *Med Hypotheses.* 1981; **7**(11): 1339–45.
[PubMed Abstract](#) | [Publisher Full Text](#)

Open Peer Review

Current Peer Review Status: ? ? ?

Version 1

Reviewer Report 24 May 2021

<https://doi.org/10.21956/wellcomeopenres.18036.r43653>

© 2021 Pervjakova N. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Natalia Pervjakova

Estonian Genome Center, Institute of Genomics, University of Tartu, Tartu, Estonia

As the third reviewer of this article, I would like to note that I fully agree with the revision comments from Dr. Cole and Dr. Karlsson. I believe that addressing their questions and concerns would substantially improve this article.

In addition to their comments, my comments are as follows:

- 1) The title of the article must clearly state that this is preliminary finding.
- 2) The introduction part should be longer. An expanded review of the previous research and the potential substantial value added by the author's research would provide context for the hypothesis. The introduction part is not consistent yet.
- 3) Methods section must contain all details (might be moved to supplementary material if too big). Please state clearly the methods you have used, including QC procedures, software, adjusting for PC etc. This would give the reviewers and readers of the paper an understanding of its strengths and limitations.

Please describe the case and control selection. The strength of this paper is in combining different questionnaires for consistent selection of the cases. However there is a lack of details in description of the final numbers of cases and controls.

Please provide a full description of statistical analysis of association. Please mention the imputation panel used for these samples.

Please provide a more detailed description of LDSC analysis. Please provide all QC steps including the software used. I have the same concern as other reviewers - this analysis might be underpowered.

My main concern is that the identified variant is an artifact from uncorrected population structure

or some other issue with the data. Further research and combining this data with other cohorts may eliminate significant results from this study

4) Results and Discussion. Please provide Manhattan and LocusZoom plots first.

Please provide more details for additional tested traits:

Do they come from UKBB? How were they selected? Are there any overlaps between samples?

As mentioned above - please provide all the details about methodology, use other reviewers comments and take them step by step, then it would be easier to understand the data reported in the results section. So far I cannot comment on the heritability analysis results due to the lack of methodology described in the methods section.

Please list all strengths and limitations of your study. Please provide a more detailed description of associated gene and locus. Please bring your findings to the context of the study.

6) Please make conclusion longer.

I have listed only the very major comments. When this would be fixed, I would be happy to move to another round of reviewing with more detailed analysis. This paper can be a good example of a concise cohort report with preliminary findings. However all the comments and concerns raised by reviewers must be addressed first. Thank you!

Is the work clearly and accurately presented and does it cite the current literature?

Partly

Is the study design appropriate and is the work technically sound?

Partly

Are sufficient details of methods and analysis provided to allow replication by others?

No

If applicable, is the statistical analysis and its interpretation appropriate?

No

Are all the source data underlying the results available to ensure full reproducibility?

No

Are the conclusions drawn adequately supported by the results?

No

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: GWAS, genomic imprinting, reproduction, diabetes.

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have

significant reservations, as outlined above.

Reviewer Report 18 May 2021

<https://doi.org/10.21956/wellcomeopenres.18036.r43654>

© 2021 Karlsson R. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Robert Karlsson

Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden

This research note had already been thoroughly reviewed when I was invited to review it, and as far as I can see the manuscript has not yet been updated to address the concerns raised by reviewer 1 (Dr. Cole). I agree with all of the comments already given by Dr. Cole, and just want to add the following mostly minor remarks:

The methods subsection concerning genotype data analyses is a little short on details and references, specifically:

- what software was used for the quality control and management of genome-wide genotype data?
- what software was used for genome-wide association analysis?
- what software was used for other association tests?

The authors report a single marker with p-value just below the conventional genome-wide significance threshold of 5×10^{-8} , which may be a true association. However, it would increase confidence in the finding somewhat if nearby markers in LD with this variant showed similar levels of association. Could the authors please provide a Manhattan plot, and a regional association plot ("LocusZoom" style) as supplementary materials? As already suggested by reviewer 1, the full GWAS summary statistics should also be made available if at all possible.

Could the authors contrast their curated phenotype based on specific food items to self-reported adherence to a [vegetarian and/or vegan diet](#)? What is the correlation between the authors' derived phenotype and self-reported vegetarian diet? What is the effect of the reported lead SNP on self-reported vegetarian diet?

Is the work clearly and accurately presented and does it cite the current literature?

Partly

Is the study design appropriate and is the work technically sound?

Yes

Are sufficient details of methods and analysis provided to allow replication by others?

Partly

If applicable, is the statistical analysis and its interpretation appropriate?

Yes

Are all the source data underlying the results available to ensure full reproducibility?

Yes

Are the conclusions drawn adequately supported by the results?

Yes

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Genetic epidemiology, genome-wide association studies of complex traits.

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.

Reviewer Report 01 February 2021

<https://doi.org/10.21956/wellcomeopenres.18036.r42083>

© 2021 Burnette Cole J. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Joanne Burnette Cole

¹ Diabetes Unit and Center for Genomic Medicine, Massachusetts General Hospital, Boston, MA, USA

² Programs in Metabolism and Medical and Population Genetics, The Broad Institute of MIT and Harvard, Cambridge, MA, USA

³ Division of Endocrinology and Center for Basic and Translational Obesity Research, Boston Children's Hospital, Boston, MA, USA

Fensom *et al.* conduct a genome-wide association study of vegetarianism in individuals of white British ancestry from UK Biobank. They identify one locus reaching genome-wide significance, SNP rs10189138, in an intron of the *VRK2* gene, encoding a serine-threonine kinase. Follow-up analysis includes testing for association between this SNP with additional phenotypes in UK Biobank and overall genetic correlation analysis. Though this locus has already been identified to be associated with a wide variety of dietary habits including meat and fish intake and several empirical dietary patterns in individuals of European ancestry in UK Biobank ((Cole, Florez, & Hirschhorn, 2020) Supplemental Table 4)¹, this is the first reporting to my knowledge of its association with a manually curated composite phenotype of vegetarianism, providing more insight into its overall association with dietary intake and potentially the function of *VRK2*.

The primary strength of the study is the use and combination of multiple touchscreen questionnaire answers related to meat and fish consumption to curate a moderately sized 'case/control' analysis of vegetarianism (>5K) versus controls (>360K), as well as testing the lead

SNP's association with a wide variety of outcomes in UK Biobank to better understand pleiotropy.

Below are my comments, in order of significance

Major Comments

1. Overall the Conclusion section is quite brief, and I believe the reader would benefit greatly from a discussion on both the limitations and interpretations of the findings in this study.
2. UK Biobank may be subject to population structure that is uncorrected using principal components in white British individuals alone (Haworth *et al.*, 2019)². Given population and geographical differences are strongly associated with dietary intake and height (Sohail *et al.*, 2019)³, it remains possible that the lead signal is an artifact of uncorrected sample structure. This issue merits either further analytical investigation or at minimum attention in the Conclusion section. Of note, the follow-up analyses I performed in Cole *et al.* using linear mixed models and adjustment for assessment center found limited residual confounding via LD score regression (Cole *et al.*, 2020)¹, but again, I believe readers should be made aware of its potential impact in the Conclusion section.
3. Calculating genetic correlation (shared heritability) between a trait with ~1% heritability seems underpowered. I am surprised by the ability to identify significant genetic correlations, can you please A) describe how you calculated heritability, B) describe the original studies for these additional tested traits C) describe your results in the context of what would appear limited power. Also, do most of these traits come from UK Biobank, and if so, how would sample overlap affect your findings?
4. There is no description of how any of the additional phenotypes in Figure 1 are derived in UK Biobank. Given UK Biobank provides raw measurements, some level of quality control should be taken on these variables, For example, BMI and several biomarkers have largely skewed distributions, how did you handle those for linear regression? It may be easier to obtain summary statistics of this SNP's association with additional traits using PheWAS servers (<https://biobankengine.stanford.edu/>; <http://big.stats.ox.ac.uk/>), or look-ups in publicly available GWAS (e.g. GIANT consortium), or the recently published GWAS on UK Biobank biomarkers (Sinnott-Armstrong *et al.*, 2021)⁴. These resources performed at least minimal phenotypic pre-processing.

Minor Comments

1. It is unclear how the 24-hour recall questionnaire from 57K individuals was combined with FFQ data to calculate weekly meat intake for all individuals (>360K). Does this mean that the majority of data is from FFQ, and a subset of contributing individuals have their weekly intake value derived from both questionnaires? If this is the case, why limit to just the 57K individuals that took the questionnaire three times, when ~200K took it at least once. Understandably, this data is more reliable for individuals that participated in the 24HR questionnaire more than once, but at minimum a clear description of how the 24HR questionnaire supplemented the FFQ data is necessary to understand this phenotype.
2. Can you please clarify the % difference per-allele calculation in Figure 1. It appears the figure reports effect estimates in units of the predictor under an additive model at this SNP, but I am unclear how the percent change calculation is made.

3. Given the manuscript reports methods before results, the mention of a “GG genotype” in the linear regression models section has no context.
4. Is the effect allele, T, also the minor allele?
5. The SNP appears to be in an intron of VRK2, making it a slightly more compelling candidate gene than using the term “near”.
6. Will the GWAS be made publicly available for download? It should be shared at least by request to the authors.

References

1. Cole J, Florez J, Hirschhorn J: Comprehensive genomic analysis of dietary habits in UK Biobank identifies hundreds of genetic associations. *Nature Communications*. 2020; **11** (1). [Publisher Full Text](#)
2. Haworth S, Mitchell R, Corbin L, Wade K, et al.: Apparent latent structure within the UK Biobank sample has implications for epidemiological analysis. *Nature Communications*. 2019; **10** (1). [Publisher Full Text](#)
3. Sohail M, Maier R, Ganna A, Bloemendal A, et al.: Polygenic adaptation on height is overestimated due to uncorrected stratification in genome-wide association studies. *eLife*. 2019; **8**. [Publisher Full Text](#)
4. Sinnott-Armstrong N, Tanigawa Y, Amar D, Mars N, et al.: Genetics of 35 blood and urine biomarkers in the UK Biobank. *Nat Genet*. 2021. [PubMed Abstract](#) | [Publisher Full Text](#)

Is the work clearly and accurately presented and does it cite the current literature?

Yes

Is the study design appropriate and is the work technically sound?

Yes

Are sufficient details of methods and analysis provided to allow replication by others?

Partly

If applicable, is the statistical analysis and its interpretation appropriate?

Yes

Are all the source data underlying the results available to ensure full reproducibility?

Partly

Are the conclusions drawn adequately supported by the results?

Partly

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Genetics of complex traits, particularly that of dietary intake and its relationship with metabolic traits.

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.