## PEER REVIEW HISTORY



## ARTICLE DETAILS

| | |
|---|---|
| **TITLE (PROVISIONAL)** | Cohort Profile: Genomic data for 26,622 individuals from the Canadian Longitudinal Study on Aging (CLSA) |
| **AUTHORS** | Li, Rui; Forgetta, Vince; Darmond-Zwaig, Corinne; Belisle, Alexandre; Balion, Cynthia; Roshandel, Delnaz; Wolfson, Christina; Lettre, Guillaume; Pare, Guillaume; Paterson, Andrew; Griffith, Lauren; Verschoor, Chris; Lathrop, Mark; Kirkland, Susan; Raina, Parminder; Richards, Brent; Ragoussis, Jiannis |

## VERSION 1 – REVIEW

| | |
|---|---|
| **REVIEWER** | Overbey, Jessica<br>Mount Sinai Health System, Department of Population Health Science and Policy |
| **REVIEW RETURNED** | 23-Nov-2021 |

| | |
|---|---|
| **GENERAL COMMENTS** | In this manuscript, authors describe data available from the CLSA study. This dataset offers many exciting opportunities to explore the impacts of genetic and environmental factors on a variety of issues in an aging Canadian population. I do question whether this manuscript is appropriate for the audience of a medical journal. The CLSA has the potential to address and explore numerous research questions; however, this manuscript reads as an advertisement for the data, with no specific research questions or issues addressed. I'm not sure the detailed methods described on pages 7-16 are a good fit for a medical journal. Many of these technical details seem more suitable for a study protocol or a dataset users guide.<br><br>Other comments:<br>1. In the final point of the "strengths and limitations of this study" section, authors state: "Potential limitations may include the relatively lower genotyping coverage in participants with non-European ancestry and inadequate power to discover very rare predisposition variants. Such limitations associated with this type of data can be overcome by imputation and meta-analysis." It is unclear how imputation could overcome this limitation – can the authors clarify.<br>2. The grammar and general writing needs significant editing/proofreading. |

| | |
|---|---|
| **REVIEWER** | Ajnakina, Olesya<br>King's College London, Department of Biostatistics & Health Informatics, Institute of Psychiatry, Psychology and Neuroscience |
| **REVIEW RETURNED** | 07-Dec-2021 |

| | |
|---|---|
| **GENERAL COMMENTS** | This manuscript aims to present a Canadian longitudinal cohort of 26622 older adults who have also been genotyped leading to identification of 794,409 common marker. The manuscript describes the steps undertaken to obtain bloods for the genetic analyses, carry out the quality control of the genome-wide genotypes and imputation of the untyped markers.<br><br>This is certainly a very large longitudinal cohort and the study leads ought to be praised for making these data available to researchers. However, the authors' claim that it is unique is unsubstantiated as there are multiple longitudinal cohorts of older adults with the genome-wide genotypes available across the globe. Please have a look at the CLOSER Discovery (https://discovery.closer.ac.uk) to see the magnitude of data available to study older adults across their lifespan with GWAS data.<br><br>The manuscript as it stands offers a detailed documentation of the technical processes involved in preparing the data. It does not however demonstrate why this cohort is unique and what more it offers compared to other cohorts, or why its description warrants a publication.<br><br>The authors mentioned that gene expression is being prepared. I would argue that once the gene expression data become available, this is when the authors would have a stronger case to make arguing the unique features of this study. |

## VERSION 1 – AUTHOR RESPONSE

Reviewer #1
Dr. Jessica Overbey, Mount Sinai Health System
Comments to the Author:
In this manuscript, authors describe data available from the CLSA study. This dataset offers many exciting opportunities to explore the impacts of genetic and environmental factors on a variety of issues in an aging Canadian population. I do question whether this manuscript is appropriate for the audience of a medical journal. The CLSA has the potential to address and explore numerous research questions; however, this manuscript reads as an advertisement for the data, with no specific research questions or issues addressed. I'm not sure the detailed methods described on pages 7-16 are a good fit for a medical journal. Many of these technical details seem more suitable for a study protocol or a dataset users guide.
Authors' response: Thank you for the comments. The purpose of this manuscript is to make this resource available to the community. It is not to ask and answer a specific scientific question. Such resources are vital to modern scientific efforts and the current document serves to describe this resource and will consequently be well-cited. As it is submitted to the "Cohort profile" article type, the manuscript is formatted according to the article guidelines.
Other comments:
1. In the final point of the "strengths and limitations of this study" section, authors state:
"Potential limitations may include the relatively lower genotyping coverage in participants with non-European ancestry and inadequate power to discover very rare predisposition variants. Such limitations associated with this type of data can be overcome by imputation and metaanalysis."
It is unclear how imputation could overcome this limitation – can the authors clarify.
Authors' response: The coverage in participants with non-European ancestry will be improved by using imputation reference panels with high genetic diversity. The text has been revised to

state the strengths and limitations more clearly.
2. The grammar and general writing needs significant editing/proofreading.
Authors' response: The manuscript has been revised carefully according to the reviewer's suggestion.

Reviewer: 2
Dr. Olesya Ajnakina, King's College London
Comments to the Author:
This manuscript aims to present a Canadian longitudinal cohort of 26622 older adults who have also been genotyped leading to identification of 794,409 common marker. The manuscript describes the steps undertaken to obtain bloods for the genetic analyses, carry out the quality control of the genome-wide genotypes and imputation of the untyped markers.
This is certainly a very large longitudinal cohort and the study leads ought to be praised for making these data available to researchers. However, the authors' claim that it is unique is unsubstantiated as there are multiple longitudinal cohorts of older adults with the genomewide genotypes available across the globe. Please have a look at the CLOSER Discovery (https://discovery.closer.ac.uk) to see the magnitude of data available to study older adults across their lifespan with GWAS data.
The manuscript as it stands offers a detailed documentation of the technical processes involved in preparing the data. It does not however demonstrate why this cohort is unique and what more it offers compared to other cohorts, or why its description warrants a publication.
The authors mentioned that gene expression is being prepared. I would argue that once the gene expression data become available, this is when the authors would have a stronger case to make arguing the unique features of this study.
Authors' response: Thank you for the comments. The CLSA is an ongoing prospective study. The main purpose of this manuscript is to profile the genomic data and make this resource available to the community. Recently, the metabolomic data on 1,314 biochemicals are available in approximately 9,500 blood samples collected from CLSA participants. In addition, the CLSA has initiated a subcohort to investigate cognitive aging in 6,000 participants. Longitudinal data will be collected from magnetic resonance imaging of the brain and microbiome of the gut. This information has been added to the manuscript.