Skoltech
Skolkovo Institute of Science and Technology

Thesis Changes Log

Name of Candidate: Vita Stepanova
PhD Program: Life Sciences
Title of Thesis: Metabolic variations of modern and ancient human populations
Supervisor: Professor Philipp Khaitovich, Skoltech
Chair of PhD defense Jury: Associate professor Georgii Bazykin , Skoltech
***Email**:* vita.stepanova@skolkovotech.ru
Date of Thesis Defense: 19/12/19

The thesis document includes the following changes in answer to the external review process.

I would like to thank all the reviewers for the time they took reading my manuscript and for all of their helpful comments and corrections. As a result of their kind feedback, I have made a number of changes. I have corrected many spelling and grammar errors, standardised the formatting throughout and ensured consistency with the referencing style and bibliography. In addition, I have slightly expanded the Introduction section to highlight the importance of the work and the power of the metabolomics approach. I have included the dedication and acknowledgements. A picture of the ADSL protein has been added in Chapter 2. Following the suggestion of Prof. Kharchenko, I have slightly expanded the section on the autistic metabolome in Chapter 3.

Reviewer: Prof. Georgii Bazykin

Comment 1:

In pp. 26–27, two distinct ways are used to calculate the number of lipids differing between the HC and the other two populations: one yields 90 lipids (p. 26), while the other yields 395 lipids (p. 27). Similarly, the two analyses for metabolites yield 93 or 166 HC-specific differences. I was confused by the differences between these approaches, and the causes of these 2- or 3-fold differences between types of analyses; please clarify them.

Answer:

Initially, we subsampled the same number of individuals per population in every comparison to equalize the statistical power, which gave us the figures of 90 lipids and 93 metabolites. These numbers reflect the abundance levels specific to WE and AA in the average of 100 sample subsets. When we do not use subsamples and do not reduce the number of individuals 4-5 times, the t-test given all samples yielded 395 lipids and 166 metabolites. I added an explanation of this difference to the text in Chapter 3.

Reviewer: Dr. Vasily Ramensky

Comment 1:

In the section «Statistical analysis of lipid and metabolite differences among populations» the samples were separated into two subsets: DS:0-4 (n = 74, ages less than five years), and DS:5-71 (n = 229, ages from five to 71 years). For this specific type of analysis based on direct comparison of populations, the author shows that the observed specific lipidome behavior of Han Chinese population in DS:5-71 was not due to the difference in statistical power between the age-based subsets. My question is why the age was not used as one of predictor variables in the subsequent population classification using machine learning. Instead, the classification procedure was applied to the DS:5-71 only.

Answer:

That's true, we haven't used the age as a predictor. In particular, we detected a difference in age intervals and conducted the following analysis for two datasets separately. However, we estimated the accuracy of the logistic regression model which is low for the samples younger than two years of age and high for any samples older than two years of age (Fig. 1).

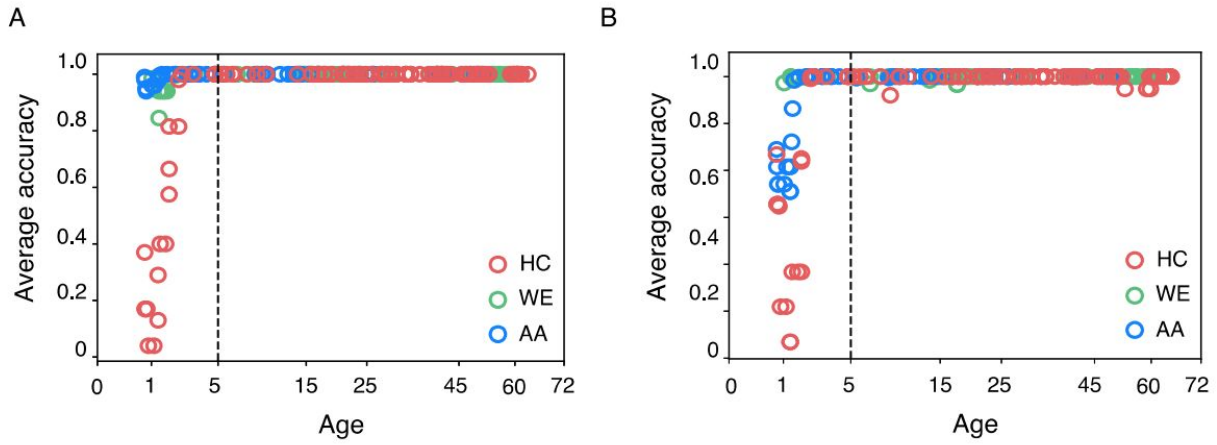

Fig.1 **Accuracy of logistic regression HC classifier depending in different age intervals.** The median performance estimates of the logistic regression model calculated using sliding age window and each population: (A) – lipids, (B) – metabolites.

Comment 2:

It is hypothesized that the Ala429Val substitution does not affect the kinetics of the murine ADSL enzyme, but instead destabilizes the secondary structure of the protein. Taking in view that the spatial structure for ADSL is available, I feel like the study would clearly benefit from the most basic bioinformatic analysis of the role of the residue on the protein function and structure. In particular, I seem to miss a figure of the protein with designated binding and substitution sites. Another thing is the prediction of the free energy change upon mutation with FoldX or related tool, although the performance of these tools is far from perfect.

Answer:

I added a figure of the ADSL protein with marked binding site and A429V substitution to Chapter 2. The destabilising effect of A429V substitution was previously shown *in vitro* by Van Laer *et al.* and is mentioned in Chapter 4. To support this observation we have estimated the ΔΔG values of V429A substitution in human ADSL using MAESTROweb (https://biwww.che.sbg.ac.at/maestro/web). The analysis yielded negative ΔΔG values equal to -0.26 for both 2j91 and 2vd6 ADSL protein structures, confirming the stabilising effect of V429A mutation.