UNIVERSITY
OF WARSAW

Faculty of Biology
Institute of Genetics and Biotechnology
prof. Paweł Golik

Warsaw, 16.11.2023

Review of the PhD thesis "Investigating the role of genetic and epigenetic variation in facial skin and scalp hair aging and DNA-based prediction of age-related human appearance traits" by Ms. Rezvan Noroozi

Understanding the relationship between genotypic and phenotypic variation is the fundamental question of genetics, one that is still not easy to fully resolve, despite enormous progress of the discipline in the last decades. Particularly in the case of complex multifactorial traits, commonly encountered in human genetics, understanding their genetic basis remains a significant challenge. The topic of project described in the reviewed thesis is one of the most challenging in the entire field of human biology. Aging is a complex and multifaceted dynamic phenotype resulting from a dense network of interactions between genetic, epigenetic, and environmental factors. The goal of the reviewed project – to shed more light on the underlying biological aspects of age-related traits (mostly related to facial and hair aging) and develop new models to predict age-related phenotypes – was therefore ambitious and required the application of a very wide variety of methods, both laboratory (genomic nucleotide variation and epigenetic methylation assays) and statistical. I find it particularly impressive that the Candidate – a young researcher – was able to master and successfully apply such a wide array of complex techniques encompassing both experimental genomics and epigenomics and very involved statistical calculations. One can therefore conclude that the described work is a successful application of appropriate scientific methodology to solve a valid and novel problem, and thus fulfills the main requirements of a PhD project.

The thesis is presented as a 115-page manuscript in English, describing unpublished results. All the sections required in a scientific manuscript are present, and the data are adequately presented in the text, tables, and figures (some minor issues are discussed in detail below). The very few minor errors and technical problems that are unavoidable in any text of this complexity do not affect the clarity of the message, and the manuscript is complete and easy to follow. Clear and concise Abstract

Pawińskiego 5A, 02-106 Warsaw, POLAND
tel.: +48 22 592 22 44, fax: +48 22 658 41 76
e-mail: p.golik@uw.edu.pl
http://www.igib.uw.edu.pl

and Conclusions sections are well written, and particularly useful in summarizing the most important points of the work. The references to relevant literature are properly cited and formatted.

The Glossary at the beginning of the manuscript (p. 1) is a very useful idea, as the main text uses multiple abbreviations. Unfortunately, this glossary is not complete, for example there is no explanation of the acronym MAE (mean absolute error) or DMP (differentially methylated position) – they are explained in the text, but their presence in the glossary would make reading easier. Also, the formatting of the glossary is not uniform: sometimes the full name is given with the acronym in parentheses (e.g., "Epigenetic age acceleration (EAA)", in other cases we have acronym followed by the full name (e.g., "SNP: Single nucleotide polymorphism", which I believe to be the preferable format for readability). EVC means external visible characteristic, not "extremely visible" (the main text uses the correct term, only the glossary has this rather amusing mistake). Finally, the glossary entries should be ordered alphabetically.

The manuscript begins with a brief Introduction which is in fact closer to a summary and statement of the main goals of the project, followed by a more detailed Theoretical background section. The background section is a complete and very well written introduction to the multiple aspects of aging-related phenotypes and using genomic and epigenetic markers to predict chronological and physiological age.

The Materials and Methods section is comprehensive and well written, particularly the sections describing the statistical methods and the multitude of recorded variables are truly impressive. The use of image analysis for detection of external phenotypes is also very interesting. I have a few questions and comments regarding this chapter:

- The Candidate states that she collected 962 samples, of which 741 were blood samples and the rest buccal swabs. But already in Table 1 we see that N = 735. What led to the elimination of 227 samples at this early stage, and how many of the remaining 735 were blood samples or buccal swabs.
- Later, describing GWAS and EWAS studies, the Candidate states that they were performed on blood samples (e.g., on p. 30 "The final imputed SNPs on the autosomal chromosomes of 719 blood samples were used for the genome-wide association analysis"). Were buccal swab samples of insufficient quality for GWAS or EWAS? I can understand why mixing different tissues would not be a good idea in EWAS, but for GWAS it should not matter. In section 4.1 the Candidate does compare buccal swab and

blood sample performance in DNAm age clocks, so there was sufficient material for analysis.

- On p. 30 the Candidate states that she performed GWAS "to find significant SNPs underlying epigenetic aging and to assess the heritability of epigenetic age acceleration". It is possible to estimate GWAS heritability (which is not the same as general heritability) from such studies, but in the remaining parts of the thesis I could not find any mention of estimating heritability (see e.g., Zaitlen & Kraft *Hum Genet.* 2012, 131(10): 1655–1664 for a review of possible approaches).
- What reference population was used for the imputation of missing genotypes in the GWAS analysis?
- Did the Candidate examine possible associations of mtDNA variants (including mitochondrial haplogroups) with age-related variables? This should be possible with Global Screening Array data.
- In section 3.2 we read that "wrinkles detection and quantification was done using MATLAB software". Was the MATLAB code for this written by the Candidate, a local collaborator, or was it obtained from literature? In the former case, some information about the availability of the source code (or reasons for withholding it) would be useful, and in the latter case a reference would be necessary.

The main body of the work is presented in the Results and Discussion chapters. These chapters are very well written, the data are of high quality, and their presentation is excellent. The only minor problem that made reading the Results chapter challenging was that any conclusions were presented only later in the Discussion chapter. While this is technically correct, given the enormous number of different variables and correlations studied, adding brief one-sentence conclusions to each section of the Results would make following the work much easier. For example, following all the tables and figure of section 4.1, the Author could simply state that the Skin&Blood clock with ssNoob normalization was found to be the most accurate for prediction of chronological age (we need to go many pages ahead to the Discussion to find this conclusion).

The first part of the study explores multiple correlations between epigenetic age acceleration (EAA) measures, environmental (lifestyle) factors and external aging-related phenotypes. This section presents many valuable results and conclusions, including some unexpected findings, such as the association of skin aging phenotypes with the male, not female gender. Another unexpected result of the study was the positive correlation between physical activity and increased Glogau Photoaging

scale value. The Author explains this result referring to the oxidative stress resulting from excessive training. Is it possible, however, that this could be explained simply by the amount of time spent outdoors (likely higher in active people)?

In addition to the already interesting results, the collection of very valuable data gathered by the Candidate can be a treasure trove for further research. I can only hope that following the publication of results at least some of these data will become available to the research community. I could even envision this collection of genomic data associated with an impressive number of medical and lifestyle parameters becoming the first step towards creating a comprehensive database similar to the UK Biobank project, but relevant for the Polish population.

The following sections describe the results of genomic (GWAS) and epigenomic association studies. Searching for variants associated with externally visible hallmarks of aging, the Candidate found only few SNPs significantly associated with wrinkle area and perceived age. My main question for this part is as follows: does this mean that for the other EVCs that were studied (the Glogau scale, alopecia, hair graying, etc.) no significant associations were found? In Discussion the Author describes a number of significantly associating SNPs found for these phenotypes in previous studies. Were the SNPs that associated with age-related characteristics in published studies not significant in the Polish population, or did the Author filter out previously known associating variants to concentrate on novel findings? In either case, this should be discussed. Interestingly, the SNPs associated with wrinkle area and perceived age are different variants in different genes. Isn't wrinkle area strongly correlated with the perceived age? If yes, why the variant significant for wrinkle area does not correlate with perceived age?

Similarly, the GWAS study for epigenetic age acceleration (EAA) measures resulted in the identification of a single variant associated with the GrimAgeAccel measure. Were there no significant associations for the other EAA measures? And how would the Author interpret the apparent lack of association with SNPs previously identified as significant in the literature (like the TERT variants mentioned in the Discussion)? As in the previous comment, this should be discussed. For example, were those previously identified SNPs completely insignificant in the present study, or perhaps they were just slightly below the significance threshold? Similarly, in the EWAS studies the Candidate found significant differential methylation for some aging-related phenotypes, but not for others. The negative findings also merit some discussion, for example, is blood a good material for studying epigenetic hallmarks of aging? Remarkably, all the significant GWAS and EWAS

associations found in the study were related to skin aging phenotypes, none were reported for scalp hair aging – I'm curious how the Candidate would explain this pattern.

The final sections of the Results are devoted to developing new models for prediction of characteristics such as perceived age, wrinkle area, and full-face wrinkles using the discovered correlations. Again, I'm impressed by the Candidate's proficiency in advanced statistical techniques.

The Discussion chapter is comprehensive, not only summarizing the obtained results, but also placing them in the wide context of published literature, demonstrating that the Candidate has extensive in-depth knowledge of her field.

To summarize, Ms. Rezvan Noroozi proved to be an independent, competent, and productive researcher. The reviewed thesis is an original solution of a valid and novel scientific problem and demonstrates her broad theoretical knowledge, capability of conducting scientific research, and impressive mastery of a very broad range of diverse methods. The text of the manuscript is complete and contains all the required elements. I can therefore conclude that the reviewed thesis fulfills the legal requirements necessary to obtain the PhD degree as stated by the Law on Higher Education and Science (PL: Ustawa z dnia 20 lipca 2018 r. Prawo o szkolnictwie wyższym i nauce (Dz. U. 2018 poz. 1668 z późn. zm.), as well as the conventional requirements expected of scientific work at the highest level. I thus recommend that the Scientific Council of Biological Sciences of the Jagiellonian University allows Ms. Rezvan Noroozi to proceed to the subsequent stages of the doctoral proceedings and defend the reviewed thesis in public.

prof. Paweł Golik