

Jury Member Report – Doctor of Philosophy thesis.

Name of Candidate: Ilia Kurochkin

PhD Program: Life Sciences

Title of Thesis: Comparative analysis of human brain based on mass-spectrometry data

Supervisor: Prof. Philipp Khaitovich

Chair of PhD defense Jury: Prof. Mikhail Gelfand

Email: m.gelfand@skoltech.ru

Date of Thesis Defense: 26 October 2018

Name of the Reviewer: Prof. Mikhail Gelfand

<p>I confirm the absence of any conflict of interest</p> <p>(Alternatively, Reviewer can formulate a possible conflict)</p>	<p>Signature:</p>  <p>Date: 26-09-2018</p>
---	---

The purpose of this report is to obtain an independent review from the members of PhD defense Jury before the thesis defense. The members of PhD defense Jury are asked to submit signed copy of the report at least 30 days prior the thesis defense. The Reviewers are asked to bring a copy of the completed report to the thesis defense and to discuss the contents of each report with each other before the thesis defense.

If the reviewers have any queries about the thesis which they wish to raise in advance, please contact the Chair of the Jury.

Reviewer's Report

Reviewers report should contain the following items:

- Brief evaluation of the thesis quality and overall structure of the dissertation.
- The relevance of the topic of dissertation work to its actual content
- The relevance of the methods used in the dissertation
- The scientific significance of the results obtained and their compliance with the international level and current state of the art
- The relevance of the obtained results to applications (if applicable)
- The quality of publications

The summary of issues to be addressed before/during the thesis defense

The thesis presents several interesting studies with biologically significant results, clearly relevant to the topic of the dissertation. The structure of the thesis follows the authors' papers; this is OK, but a better attempt in integration might increase the overall impact. One section that specifically suffers is the introductory review. The methods are solid and relevant. The results are of international level. One possible practical application is in medical diagnostics, however, this aspect is not specifically addressed in the thesis.

The publications are one shared-first author paper in *Mol. Biol. Evol.*, middle author paper in *Molecular Psychiatry* and a shared-first author paper in preparation. The results have been reported by the author at international conferences *Metabolomics-2016* (Ireland) and at *SMBE-2018* (Japan); both are high-level conferences, so all the formal requirements are met.

Hence, overall the presented PhD thesis reports valid, scientifically significant findings reported in a convincing manner. However, I have a number of specific comments that are listed below.

The review part of the thesis is good, although sometimes the line of reasoning blurs, and a list of facts are given instead of a logical discourse. On such example where this effect is especially pronounced is section 2.1. One more is description of the oligogenic and major gene models of ASD development: "Oligogenic model claims that ASD is caused by a relatively small number of genetic variants, each having a large risk of ASD development. Major gene model claims that ASD development can be caused by genetic variants, each having a large risk" – what is the difference? – and then: "Major gene model and polygenic model are not mutually exclusive". What dictates the choice of ADNP and ANK2 for specific discussion at the end of section 2.2? The selection of discussed metabolites in section 2.4.1 looks completely spurious (glucose and ATP are important, but why specifically these two?). Metabolome alterations in sections 2.4.2 and 2.4.5 would look much better in a tabular form instead of lists in the text or, even better, as Venn diagrams: that would allow a reader to assess the consistency of findings. In 2.5.4 it is not clear whether the observed changes are due to schizophrenia or to the fact that strong drugs have been taken, naturally leading to changes in the lipid content. At that, one would expect a review not merely to repeat the findings but to have a critical component as well.

In 3.1.1, 3.1.2 it is not explicit that the author did not participate in the preparation of samples (although this is implicitly explained in section 1.4).

Description of the methods needs some enhancement. In particular, what are "enzymes directly linked to metabolites" on p. 31? More importantly, in a situation of highly intercorrelated features it is not clear that it is feasible to assume that the best predictors in a machine learning have any special biological meaning (the authors mention instability of feature selection even with L_1 regularization; btw, in section 3.2.3 the same procedure is called lasso regularization). In the first paragraph of page 32, when defining human-specific metabolite changes, did the authors require the change to be significant? in what sense? At the bottom of the same page, was $|\log_2 \text{fold change}| > 0.2$ the only condition, or was there an additional condition on statistical significance? (At low expression levels one may observe high, but insignificant fold change.) At that, the applied threshold looks very weak: are the biological conclusions robust with regards to the threshold selection? If Fig. 3.3e is based on 500-fold resampling, it might be a good idea to present the ROC curve with error bars or as a distribution. In the last paragraph of section 3.2, is the observed differences in the fraction of human-specific metabolites in the autism-related modules mirrored by chimpanzee-specific differences, or are some modules enriched in human-specific differences, and other modules, in chimp-specific ones?

In the discussion (section 3.3) it would be instructive to concentrate not on similarities, but on

differences between the brain and the blood and urine metabolic changes in ASD, as the latter (unlike the former) may serve as diagnostic markers.

In section 4.1.5 it is not explained what clustering algorithm has been applied, is it the complete linkage as in chapter 3? Further, why only autism has been analyzed using linear regression?

Finally, the conclusions chapter would be much more interesting if the author had not just listed the findings of three studies but attempted to integrate them. In particular, precursors of lipids are metabolites: are there any concerted differences in the metabolite and lipid concentrations in various conditions or between species? (Indirectly same links may be observed via transcriptome analysis.) Given that the metabolite study identified some correlation between metabolite differences in autism and between primates, were similar correlations observed for lipids (at that, an integrative analysis of the results of chapters 4 and 5 would be instructive).

The English style and spelling need to be improved and misprints need to be corrected, below is a list of examples:

- Investigate metabolome alteration in prefrontal cortex in autism patients and identification of human-specific metabolome changes.
- We were the first who address
- all analysis
- It was demonstrated that complex traits are mainly driven by noncoding variants.
- have provided a detailed understanding
- we define metabolites as hydrophilic (polar) fraction ...lipids represent the hydrophobic fraction ("the" is needed in both cases)
- This classification is introduced due to the methodology
- metabolites are commonly spread in human organism
- One of the typical primary metabolite is glucose, that animal tissues utilize
- Despite the numbers of secondary metabolites are produced by microorganisms
- uniform structures
- functions covered by the metabolites spread
- The average concentration of each metabolite vary
- studied the most
- The diversity of glycerolipids to number, length and type of fatty acid chains
- They have an inositol group that can regulate its binding activity by phosphorylation/dephosphorylation process
- sense membrane property
- To test overrepresentation autism-related metabolites in metabolic pathways
- compare autism-related metabolites with genes differentially expressed in autism
- Lipidome alterations in human prefrontal cortex during and cognitive disorders (This misprint has occurred in the title of Chapter 4!)
- lipids constituting this 2% cluster in specific functions

Provisional Recommendation

I recommend that the candidate should defend the thesis by means of a formal thesis defense

I recommend that the candidate should defend the thesis by means of a formal thesis defense only after appropriate changes would be introduced in candidate's thesis according to the recommendations of the present report

The thesis is not acceptable and I recommend that the candidate be exempt from the formal thesis defense