

## 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: Dr Amaury Cazenave Gassiot

I confirm the absence of any conflict of interest.	Signature:  Date: 25 September 2018
--	--

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

## General appreciation

- The thesis is generally well written and structured. It is clear and understandable. It summarises a rather great amount of work that is more than adequate for a PhD thesis.
- The topic of the dissertation is highly relevant and of great interest to the field.
- The mass spectrometric analyses were conducted using state-of-the-art instrumentations and software. Some limitations in the lipid annotations could maybe have been addressed (see comments below) but it is a well-known caveat in even the most recent lipidomics workflows. The bioinformatics methods are very elaborate. The number and diversity of the samples used in the studies is a major strong point of the work.
- The results are worthy of publication in well-rated peer-reviewed journals, as demonstrated by the already published articles.
- More stressed may possibly have been put on the possible application of the metabolomics finding with regards to diagnostic tests (see comments below)

## Comments to be addressed:

1. It is almost a minor point, but the presentation and classification of metabolites and lipids may be a little too general in the introductory paragraphs 2.4.1 and 2.5.1. Specifically, the author rounds up all vitamins as metabolites, although vitamins A, D, E, K are fat soluble and generally classified as lipids. On page 15, it is stated that “lipids are classified into three major classes”, this seems like an oversimplification that hardly represents the chemical diversity of lipids. The author recognises already at the end of the same paragraph that TG constitute a fourth class, while the LipidMaps consortium catalogues at least eight classes of lipids.
2. Figure 2.2, a summary of pathway enrichment in ASD in twelve published studies, highlights the limitations of Chapter 2. The authors did a good job at summarising a number of studies but little perspective is given on these results. The differences between the studies are striking, taking the extreme: Ming et al. find all 49 pathways (!) involved in ASD while Cozzolino et al. only find one (and that one is common to only four out of twelve studies). It would have been good to add a final part to the chapter to discuss how the studies overlap (or do not), and what metabolic pathways are most likely to be of practical interest (those found by all studies?). The same comment is valid for Chapter 3 and the various lipid pathways identified in the studies described in Chapter 4.
3. On page 14, the summary of an article by Graham et al. highlights one of the main caveats of untargeted metabolomics and lipidomics studies: over- or even mis-annotation of molecular species. In that specific example, Graham et al. identified significant changes in the concentration of phosphatidylglycerophosphate (16:1 $\omega$ 7/18:0). However, for that study, the authors seem to have used a high-resolution single stage MS approach. Such an approach yields an accurate mass measurement, which can inform on the sum composition of a lipid (in this case 34:1) but cannot inform in any way about the on the actual fatty composition (16:1/18:0) unless MS/MS is undertaken, and even less yield information on the double bond position (16:1 $\omega$ 7). At best the authors should thus have identified the lipid as phosphatidylglycerophosphate (34:1). If one wants to dig deeper, one may also wonder why a lipid bearing two phosphate group, and therefore likely to yield negative [M-H]<sup>-</sup> or possibly [M-2H]<sup>2-</sup> ions was identified as positive ionisation... the same is true for the other lipid identified in the same study: CDP-diacylglycerol (18:1/18:2) that should probably be only reported as 36:3.

Of course, these issues are with the study by Graham *et al.* and not due to a mistake by the author of the present work. However, they highlight a common problem encountered when using single stage MS together with databases searches; an issue that should be mentioned somewhere in the dissertation as it has a direct effect on pathway identification.

4. Another example of the above issue, directly relevant to this dissertation, is shown in Figure 5.4 D. Here the author has chosen to show the concentration of monogalactosyldiacylglycerol (MGDG).

A few questions come to mind:

- MGDG is an entire class of lipid, the reference given (LMGL05010014) corresponds to MGDG (18:3/18:4). Can one assume that this is the molecular species that was identified?
- If so, how was the fatty composition elucidated? If only single stage MS was conducted the best annotation for that lipid should be MGDG (36:7). Defined fatty acyl moieties can only be described if MS/MS was done and the respective fragments found.
- Assuming the lipid is indeed MGDG (18:3/18:4), it seems a little odd. MGDG is a plant lipid, I don't think it has ever been described in mammalian kidney. In addition, 18:3 and 18:4 fatty acids might well be present in kidney but their concentration would be very low (traces level for 18:4).
- To support the identification of this lipid, the author could provide MS info. For example: what ions were detected, was the identification validated in both positive and negative ionisation, is there any MS/MS data?
- In addition, why choose to highlight that specific lipid in the figure while it is not discussed at all in the text?

5. In relation to the above comment, a better description of how lipids were annotated (included maybe the use of internal standards comparison and retention time information) could be added in paragraph 4.1.4. Maybe supplementary data could be made accessible, including the list of identified species (for example what species were used for generating Figure 4.4).

6. In the field of lipidomics, normalisation is widely done using class-specific internal standards (IS). In paragraph 4.1.2, the IS 1,2-diheptadecanoyl-sn-glycero-3-phosphocholine is mentioned, but in paragraph 4.1.3, another way of normalisation is mentioned. Was the IS used at all (as it is mentioned in Paragraph 5.1.2)? If so, how? If not, why? Is the IS mentioned in 5.1.2 the same IS?

7. Randomisation of samples is not mentioned in paragraph 3.1.4, also it is later in paragraph 3.2.1. Add a description of randomisation steps in the method section.

8. The following figures have already been published: 4.1, 4.3, 4.4, 5.2, 5.3, 5.4, 5.5, and 5.6. Should this be acknowledge either in the figures' legend or in section 1.5 where it could be described which publication corresponds to which chapter.

9. This comment is somewhat aligned with comment 2 above. In Chapter 3 and in the conclusions, it is stated: "Remarkably, many of these alterations, both at the pathway level and at the level of individual metabolites, coincide with the differences reported in urine and blood of ASD patients." This is a potentially very interesting point for potential diagnostic applications, but it is only mentioned in passing. What are these specific pathways and individual metabolites? Have they been validated? If not, that would be a perspective for future work.

Minor corrections:

- Page ii: rephrase "...the brain's compounds composition, their concentrations and the compound differences between the brain and other tissues." One way to rephrase could be "the brain's metabolome and lipidome, and their differences between the brain and other tissues".

- Pages iv to vi and throughout the thesis: in some titles, each word is capitalized (e.g. “Lipidome Evolution in Mammalian Tissues”), while in some others it is not (e.g. “Metabolome signature of autism in the human prefrontal cortex”).
- Pages vii and 53: change “Autism associated” to “Autism-associated”.
- Page viii: change “arachydonoil” to “arachidonoyl”.
- Page viii: change “Odd-carbon” to “Odd-chain”.
- Page ix: no hyphen required between “Trans” and “fatty”.
- Page 3: change “Most of bioinformatics analyses” to “Most of the bioinformatics analyses” and “all analysis” to “all analyses”.
- Page 6: line 1, change “Alzheimers” to Alzheimer’s”; line 14, change “on independent Canadians cohort” to “on an independent Canadian cohort”; line 15, change “authors” to “the authors”; line 23, delete “and” before “commonly”; line 25-26, change “in recent study” to “in recent study authors assessed” to “in a recent study, the authors assessed”.
- Page 7, line 11: reference(s) could be added with regards to the statement “Other explanations were offered as well”.
- Page 8: line 1: add “have” between “and” and “risen”, delete “the’ in “the meiosis”; last line: add a full stop after “social memory”, and start the following sentence by “Besides, knockout of ADNP...”.
- Page 9, first line: replace “Gene ANK2” by “ANK2 gene”; line 6: replace “at the other levels” by “at other levels”; line 8: replace “unpolar compounds, and metabolome” by “mostly non-polar compounds, while the metabolome”; line 16: change to “Overview of the metabolome”; line 21: add “molecular” between “small” and “weight”.
- Page 10, line 2: replace “Additional” by “In addition” and add a comma after “function”; line 4: add a comma after the reference and delete “and” before “amino acids”; line 6: replace “of plenty” by “with plenty”; line 7: replace “Being the crucial” by “A crucial” and “the energetic source” by “the main cellular energy source”; line 10: add commas after “metabolites” and after “primary ones”, delete “unusual and”; line 11: replace “numbers” by “diversity”, delete “are” between “metabolites” and “produced”, and a comma after “plants”; line 14: replace “are the typical” by “are a typical”; line 27: replace “The tissue metabolic” by “A tissue metabolic”; line 28: delete “the” between “in” and “urine”.
- Page 11: line 3: inverse “weight” and “molecular”; line 4-5: replace “were studied the most using a systematic metabolomics” by “were mostly studied using systematic metabolomics”; line 13: delete “the” between “in” and “pairwise”; line 26: replace “using the combination” by “using a combination”.
- Page 14, line 7: add “a” between “with” and “21-sample”; line 19: hyphenate “sphingosine-1-phosphate”; line 27: change “diacylglycerol” to “CDP-diacylglycerol”.
- Page 16, lines 1-2: the sentence “the most common among them are omega-3 ( $\omega$ 3) and omega-6 ( $\omega$ 6), in which double bond is at 3<sup>rd</sup> or 6<sup>th</sup> carbon atom in the tail of the chain.” is slightly misleading. It could be read as stating that those PUFA are the most common fatty acids, while 18:1 $\omega$ 9 is likely to be the most abundant fatty acid in nature. Please try to rephrase it. In addition, “rd” and “th” should be in superscript not subscript.
- Page 18, line 4: delete “the” before “lipid functionality”; figure legend: add an “s” to “interaction”
- Page 20, line 14: change “decrease of concentration levels in most PUFA” by “decrease in the concentration of most PUFA”; line 28” delete “both”.
- Page 24, line 19: change “In the pioneering study” to “In a pioneering study”.
- Page 25, line 1: delete “technique”.

- Page 29, line 20: delete extra space between “methanol:” and “methyl-tert...”; line 21: change “13” to “1/3”; line 21: delete full stop in “mg.”; line 24: delete full stop in “mm.”; line 28: change “were represented by” to “consisted of”.
- Page 30, line 19: delete “in water”; lines 22-23: The sentence “After a 3 minute wash with 100% buffer A, the column was re-equilibrated with 100% buffer B.” is redundant as this step is already described in the previous sentence.
- Page 31, line 10: subscript “4” in “[M+NH4]”; line 11: change “[MH]” to “[M-H]”; line 11: subscript “2” in “[M-H2O-H]”.
- Page 32: add a space between “3.3” and “B” in “Figure 3.3B”.
- Page 33, line 5: hyphenate “BH-corrected”; line 13: add “of” between “overrepresentation” and “ASD-related”.
- Page 36, line 16: change “each” to “both”.
- Page 38, line 4: add “the” before “glutathione”; line 26: add a space between “Fisher test,” and “p<0.001”.
- Page 43, line 7: add a comma after “It is noteworthy that,”.
- Pages 43-44: possibly add a reference(s) about “Purine metabolism is particularly interesting, as purinergic signaling is involved in neurodevelopment processes, including cell proliferation, differentiation and neuron-glia cross-talk.”.
- Page 46, line 6: hyphenate “spectrometry-based”.
- Page 47, line 16: no need to capitalise “The”; line 23: change “(3:1(v/v))” to “3:1 (v/v)”; line 24: is it 14,000 g or 14,000 rpm? ; line 25: change “was” to “were”; line 27: change “concentrated to complete dryness” to “dried”; last line: “th” should be superscript, not subscript.
- Page 48, line 4: change “NH4Acetate” to “ammonium acetate”; line 5: change “contained” to “consisted of”; line 5: change “NH4Acetate” to “ammonium acetate”; line 13: change “60.000” to “60,000”.
- Page 56, line 13: change “arachydonoil” to “arachidonoyl”.
- Page 59, line 20: add a space between “in” and “>50%”; line 21: delete the comma after “individuals”; line 28: add a space between “correlation” and “>0.7” and between “of” and “<0.05”.
- Page 60, line 28: change “Figure 5.3 a” to “Figure 5.3 A”.
- Page 61, line 12: add a space between “in” and “>50%”.
- Page 64, line 2: add a space between “in” and “>50%”.
- Page 65, figure title: Add “Lipidome” before “Data overview” to be consistent with the list of figures.
- Page 66, line 1: delete “an” between “on” and “average”.
- Page 69, line 23: add a space between “smallest” and “in”.
- Page 72, line 14: change “in three lineages the human, mouse, and chinchilla ones as well as a marginally significant trend in the two bat lineages.” to “in the human, mouse, and chinchilla lineages, as well as a marginally significant trend in the two bat lineages.”
- Page 77: The [Bloomberg et al., 2003] reference is the only one for which the title is in all capital letters.
- Page 84, [Lussu et al., 2017]: delete the space between “1” and “H-”.

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