RNDr. Dominik Filipp, Csc.
Laboratory of Immunobiology
Institute of Molecular Genetics AS CR
Videnska 1083
CZ-142 20 Prague 4
Czech Republic

Tel. (+420) 241.063.158
Mobil: (+420) 774.889.410
Email: dominik.filipp@img.cas.cz

Re: Examiner's Report on Klára Horáčková's Master thesis:

Mechanisms of immune dysregulation leading to inflammatory bowel disease

This study, describes the investigation into the identification of genes which are involved, or likely involved, in the Very Early Onset of Inflammatory Bowel Disease (VEO-IBD), i.e. in <13 year old pediatric patients. Since, in the early life, the impact of environment on the onset of VEO-IBD is assumed to be overwhelmingly dominanted by genetic factors, its early manifestation is largely considered to have a monogenic basis. Thus, the discovery path for genes potentially causing the onset of VEO-IBD uses a staightforward protocol starting with clinical anamnesis of IBD, blood sample collection, DNA isolation and library preparation, whole-exome sequencing, bioinformatical sequencing data processing, variant filtering, identification of causal variants and their validation. Thus, while this aproach is rather simple, the scope of the work is obviously complex and methodologically challenging with the final results dependent on many bioinformatical factors which are during analysis taken into considerations. Overall, the diploma thesis of Klara Horackova is without any doubt of the highest quality one can expect from a student completing the university studies. I would even dare to declare that this diploma thesis represents a new trend in rapidly emerging era of immuno-bioinformatics which in my opinion starting to dominate the field of clinical and translational research. In this context, Klara exemplifies a prototyp of long-awaited new type of students who have gained skills not only to conduct "wet experiments" in the lab, but foremost to "destile and extract" clinically and biologically essential information from big data collecting databases. This "*in-silico*" analysis will play a dominant part in designing future protocols for immunointervention therapies. From this point of view, I very much appreciate this approach whereby the immunology students gain this type of advanced training and expertise in clinical setting. Specifically, in this case, one must appreciate the high quality of the supervisor and world renowned organization CLIP in fostering and educating our students in this field.

The thesis is written up in a standard format, in English. It consists of 9 standard chapters, the Introduction, Literature overview, Diagnosis of VEO-IBD, Aim of study, Material and Methods, Results, Discussion, Conclusions and References. Thesis also contains a short chapter on "Online resources" and a three-part supplement.

In general, the work brings about several very interesting results. First, analyzing 20 pediatric patients diagnosed with VEO-IBD using WES approach, it identified and validated 5 causal variants in 4 genes, DUOX, FOXP3, NLRP3 and NOD2, with the first three being newly described. In addition, 6 more identified variants in 5 genes need further validations. I believe, that once these validations in combination with immuno and phenotypic analyses of these mutations are completed, these discoveries are publishable in internationally well-recognized journals. Second, and very important conclussion from this study is that VEO-IBD cases seem to be mostly related to primary immunodeficiencies with gastrointestinal manifestations. This can have far-reaching consequences for further delineation of future approaches for accelarating the rate of discovery in this field. Biological models which can test the predictive value of newly identified mutations in VEO-IBD will certainly follow, allowing the authors to ask additional and even more complex questions within the frame of this topic.

While I feel that the thesis is of excellent quality, described data are original and valuable for a broad research and clinical community, there are several suggestions and questions that should be further addressed and discussed.

First, I have four **formal concerns and technical questions**:

1/ As explicitly referred to above, the methodology used in this study involves several clinical, biological and bioinfomatical approaches. From the presented thesis, it is impossible to grasp which experiments and analyses were performed by the author herself and which by her colleagues. Also, in the text describing the sequencing data analysis, filtering and variant verification, the author refers to a specific day when such bioinformatical operation(s) took place (for example, on page 50, IVA, accessed Mar 25, 2020). Can you specify your contribution to this complex research task and highlight the time course/pipeline of your work? Did the analysis occur in the sequential mode, i.e. patient by patient, or you collected all sequencing data first and only then you subjected a huge chunk of data to bioinformatical analysis?

2/ There is a a formal problem with the Tab.9. After inspecting the dicrepancies between the text and the content of this Table, it seems that the table lines got shifted towards the top of the Table, commencing with the patient #8. Thus, the patients #7 is listed as being impaired in the gene PSTPIP1. Similarly, the patients #9, 12, 15, 18 are in the Table indicated as being carriers of other gene variants.

4/ Since the thesis do not deal with the “Mechanisms of immune dysregulation leading to IBD” *per se*, but rather with the identification of gene variants which are, or can be involved in the pathogenic process leading to IBD and other symptoms and diseases, I would opt for a slightly different title.

3/ Even though, overall, the thesis is written with a very good command of english, occasionally, there are several imperfections which make it hard to understand the meaning of given sentence. For example, on page 65: “Mefv$^{-/-}$ knockout mice with induced colitis presented with highly expressed Mefv in the inflamed gut...”. If Mefv is KO, how it can be highly expressed?

Questions for discussion:

1/ on page 18 you referred to the fact that there are so far about 240 susceptibility loci associated with IBD. However, your IBD and expanded IBD2 gene lists acounted only for 50 and 113 genes, respectively, which were described in the literature. Your Closest Disease-Causing Genes (CDG) gene list includes 425 genes. However, the original and full list of CDG (Requena et al, 2018) includes 5430 genes reported, and 13005 genes not curently reported to be disease associated. Given that you were able to identify causal mutations in only 20% of your pediatric patients (and perhaps 45% if counting also those genes which were not validated so far) it is clear that this IBD2 list has to be signifiicantly expanded to cover other possible variants. Can you explain why not to filter the variants using a much wider panel of genes, for example those already reported to be disease associated (5430 genes), or perhaps all of them, i.e. reported or not reported. Despite the fact that it would be a time-consuming operation, it would most likely provide very effective and robust results.

2/ FoxP3 variant mutation p.H400L in the patient #13 is very interesting because of the general impact of FoxP3 mutation on immunity. Can you explain how you modelled or predicted its damaging and/or causing loss of function properties? Do you or your colleauges test for its suppresive capacity, FoxP3 expression and cellularity of Tregs? How the biologiocal/immune aspects of suspicious gene variants are further evaluated?

3/ On page 94 you referred to Brodin et al (2015) that immunity is driven by genetics up to 40%, leaving a tengible room for environmental factors. How is this notion reconciled in case of pediatric IBD patients considering that VEO-IBD has a monogenic basis which largely excludes environmental factors in its onset? Can for example the microbiota enhance or attenuate such a strong genetic impact? Can breastfeeding or the use of antibiotics by mothers during pregnancy affect such interplay between microbiota and immune system defects?

4. On page 88, you stated that WES nowadays is a golden standard, but WGS will likely take over in the future. Under what conditions such transition from WES to WGS can happen? What is still missing to accomplish such technological upgrade?

Conclusions and recommendations

I have identified both the strengths and weaknesses of the thesis, although I have concentrated mainly upon the latter as it is expected from such report. However, I want to emphasize, that the above listed concerns in no way diminish the high quality of work presented in this thesis with significant overlap with translational and clinical research. Based on this, I recommend this thesis to be accepted as the fulfilment of the requirement for awarding the Master degree to the candidate.

Best regards,