Hi Shicheng, as Mary Sienko sets up your flight/hotel for Dec 15-16, I wanted to touch base on your presentation.  Please prepare a 40-min talk, allowing 15-20 min for QnA (seminar will be 1 hr total).

For your talk, I recommend you tailor it to how you would use statistical methods to analyze population based cohort data (eg. UK Biobank, FinnGen, AllofUs, MCRI) with the aim of novel target identification.

Right now, we have access to UK Biobank data - take a look at the UK Biobank Data Showcase, which provides a detailed listing (and some summary stats) of the types of data that are available.  There have also been numerous publications that provide this information.  I know you’ve worked with WES data from Marshfield Clinic, so you may have some familiarity with the complexities of EMR-based phenotype data, and linking it to large-scale genomic data for statistical analysis.

Please take time to carefully consider the phenotype data - what considerations must be made when using these data for analysis?  What statistical tools can you apply?  What considerations must be made (and downstream analyses must be done) if you were to identify a novel variant-disease relationship?

Provide examples from your own research that demonstrate your knowledge (if based on target discovery for drug development and/or UK Biobank data, that would be great, but if not, feel free to give other examples and explain how you could extend your work at Marshfield Clinic to achieve our goals).

Please include examples of novel approaches that are particularly relevant (eg. your approach that leveraged compound heterozygosity to identify novel targets in iron overload/ hemochromatosis).

Also keep in mind our therapeutic areas include oncology and immunology, so examples from your work in those areas would also be good.

Please let me know if you have any questions.  Looking forward to meeting you onsite next week!

Harold’s Comments:

I just have a quick look at the letter; I think MH provides you a very detailed thing that you could/should talk about during your talk, especially the ability to demonstrate

1) Statistical genetic methods you have developed to deal with genetic associations especially related to target discovery and validation

2) You current training and experiences with EMR data and domain knowledge of immunology or oncology which could be a broad working knowledge about these

3) Your thoughts about the analyses you might propose to help using UKBB data  I think you might have to pay attention to the recent papers that have published regarding to UKBB, most of them are in bioxriv and you could read them for free, Also, prepare to answer questions about the limitation of using UKBB and how you could deal with it.

PS: not all the people with your talk at JNJ would agree to use UKBB as the source of target discovery and validation, they might challenge you with these difficult questions, just be prepared.

I would suggest that you spend 2/3 of time to talk about your current project and prepare some of your thoughts for another 1/3, using your current method or knowledge to solve or provide some solutions you might have for using UKBB. You might include some current methods published in the papers and your method you think worth testing and trying and why.

By the way, prepare some questions about the position and also ask about other opportunities this position would have besides the UKBB project, usually, they expect you to have multiple projects at the same time.