Investigating Individual Specific Effects in Omics Datasets

A competent analysis of vast medical datasets requires multidisciplinary approach combining knowledge of statistics (data science), programming and algorithmics (computer science) with domain-knowledge of investigated field like genetics of disease.

I have a proven record of achievement in all those fields.

1) I possess a certification in Base SAS programming. SAS is a powerful domain-specific programming language used by statisticians throughout academia and industry for data analysis on large datasets. I earned this qualification during my 1-year long internship at SAS Scotland R&D.
2) As shown in my transcript I have obtained good, very good and excellent grades in algorithmics.
3) I have always obtained excellent grades in all programming courses, including C, Python and Java.
4) I'm a programming enthusiast and I've taught myself programming in a new programming language called Go. Samples of my projects written in this language can be found on my github profile, github.com/picrin
5) The topic of my undergraduate thesis focused on application of computer science to population genetics. The thesis is publicly available on https://github.com/picrin/naturalSelection
6) I possess basic domain knowledge in such sciences like chemistry and biology, as I have taken basic modules in these subjects, and performed very well in them.
7) I have collaborated on a publication in lipidomics published in a peer-reviewed journal (Journal of Analytical Chemistry), which indicates my ability to conduct scientific research.

Specific to the topic of research proposed by the team at Glasgow, I have a few ideas that could help in the investigation of DNA array data. Specifically, I believe that combining data on Protein-Protein Interaction Networks (PPIN) with microarray data could help to alleviate the curse of dimensionality which became a concern with regards to missing heritability. For example the following quotation comes from professor Leonid Kruglyak at Princeton University: "Taken to the extreme, practically every gene in the genome could have a variant that affects height, for example". I believe that using publicly accessible data on Protein-Protein Interaction Networks (PPIN) could help to rule out many genes, which have no chance of having effect at particular complex trait (such as heart disease), by the virtue of being far away in the PPIN graph from proteins encoded by the genes most affecting a given complex trait.